

# RESEARCH IN COMPUTING SCIENCE

ISSN: 1870-4069

## ADVANCES IN COMPUTING SCIENCE

NOVEMBER 2012, MEXICO, D.F.

Vol. 58

### VOLUME EDITORS:

E. Castillo Montiel

J.C. Chimal Eguía

A. Uriarte Arcia

L. Cabrera Rivera

# **Advances in Computing Science**

---

# Research in Computing Science

---

## Series Editorial Board

Comité Editorial de la Serie

### Editors-in-Chief:

Editores en Jefe

*Juan Humberto Sossa Azuela (Mexico)*  
*Gerhard Ritter (USA)*  
*Jean Serra (France)*  
*Ulises Cortés (Spain)*

### Associate Editors:

Editores Asociados

*Jesús Angulo (France)*  
*Jihad El-Sana (Israel)*  
*Jesús Figueroa (Mexico)*  
*Alexander Gelbukh (Russia)*  
*Ioannis Kakadiaris (USA)*  
*Serguei Levachkine (Russia)*  
*Petros Maragos (Greece)*  
*Julian Padget (UK)*  
*Mateo Valero (Spain)*

### Editorial Coordination:

Coordinación Editorial

*Blanca Miranda Valencia*

### Formatting:

Formación

*Evelia Araceli Santos Camacho*  
*Luis Cabrera Rivera*  
*Sabino Miranda Jiménez*  
*Alejandro Gómez Conde*  
*Julio César Rangel Reyes*  
*Imelda Escamilla Bouchán*  
*Carlos Vladimir Luna Soto*  
*Iliac Huerta Trujillo*  
*Juan Manuel Martínez Hernández*

**Research in Computing Science** es una publicación trimestral, de circulación internacional, editada por el Centro de Investigación en Computación del IPN, para dar a conocer los avances de investigación científica y desarrollo tecnológico de la comunidad científica internacional. **Volumen 58**, Noviembre, 2012. Tiraje: 500 ejemplares. *Certificado de Reserva de Derechos al Uso Exclusivo del Título* No. 04-2005-121611550100-102, expedido por el Instituto Nacional de Derecho de Autor. *Certificado de Licitud de Título* No. 12897, *Certificado de licitud de Contenido* No. 10470, expedidos por la Comisión Calificadora de Publicaciones y Revistas Ilustradas. El contenido de los artículos es responsabilidad exclusiva de sus respectivos autores. Queda prohibida la reproducción total o parcial, por cualquier medio, sin el permiso expreso del editor, excepto para uso personal o de estudio haciendo cita explícita en la primera página de cada documento. Impreso en la Ciudad de México, en los Talleres Gráficos del IPN – Dirección de Publicaciones, Tres Guerras 27, Centro Histórico, México, D.F. Distribuida por el Centro de Investigación en Computación, Av. Juan de Dios Bátiz S/N, Esq. Av. Miguel Othón de Mendizábal, Col. Nueva Industrial Vallejo, C.P. 07738, México, D.F. Tel. 57 29 60 00, ext. 56571.

**Editor Responsible:** *Juan Humberto Sossa Azuela*, RFC SOAJ560723

*Research in Computing Science* is published by the Center for Computing Research of IPN. **Volume 58**, November, 2012. Printing 500. The authors are responsible for the contents of their articles. All rights reserved. No part of this publication may be reproduced, stored in a retrieval system, or transmitted, in any form or by any means, electronic, mechanical, photocopying, recording or otherwise, without prior permission of Centre for Computing Research. Printed in Mexico City, November, 2012, in the IPN Graphic Workshop – Publication Office.

---

**Volume 58**

Volumen 58

---

# Advances in Computing Science

## Volume Editors:

Editores del Volumen

*Erandi Castillo Montiel*  
*Juan Carlos Chimal Eguía*  
*Abril Uriarte Arcia*  
*Luis Cabrera Rivera*

Instituto Politécnico Nacional  
Centro de Investigación en Computación  
México 2012



**ISSN: 1870-4069**

---

Copyright © 2012 Instituto Politécnico Nacional  
*Copyright © 2012 Instituto Politécnico Nacional*

Instituto Politécnico Nacional (IPN)  
Centro de Investigación en Computación (CIC)  
Av. Juan de Dios Bátiz s/n esq. M. Othón de Mendizábal  
Unidad Profesional “Adolfo López Mateos”, Zacatenco  
07738, México D.F., México

<http://www.ipn.mx>  
<http://www.cic.ipn.mx>

Indexed in LATINDEX and PERIODICA  
*Indexada en LATINDEX y PERIODICA*

**Printing: 500**  
Tiraje: 500

**Printed in Mexico**  
Impreso en México

## Preface

The purpose of this volume is to present the most recent advance in selected areas of computer science. Works included in this volume were carefully selected by the editors on the basis of the blind reviewing process. The main criteria for selection were originality and technical quality.

This issue of Research in Computing Science will be useful for researchers and students working in the different areas of computer science, as well as, for all reader interested in this field of knowledge.

In total, we received 39 papers by 97 authors that were submitted for evaluation; see Table 1. Each submitted paper was reviewed by 2 independent members of the editorial board of the volume or additional reviewers. This volume contains revised version of 30 accepted papers by 81 authors. The acceptance rate is 76.92 %.

**Table 1** Statistics of submissions and accepted papers by topic

<b>Topic</b>	<b>Submitted</b>	<b>Accepted</b>
<b>Artificial Intelligence</b>	13	11
<b>Communications &amp; Computer Networks</b>	2	2
<b>Databases &amp; Software Technology</b>	4	3
<b>Digital Signal Processing</b>	1	1
<b>Real Time &amp; Automation</b>	3	2
<b>Intelligent Processing of Geospatial Information</b>	2	1
<b>Micro technologies &amp; Embedded Systems</b>	4	3
<b>Natural Language Processing</b>	5	2
<b>Neural Networks &amp; Unconventional Computation</b>	2	2
<b>Simulation &amp; Modelling</b>	3	3

We would like to express our gratitude to all people who help to elaborate this volume. First to the authors of the papers for the technical excellence of their works, that guarantees the quality of this publication. We also want to thank the members of the editorial board for their hard work in evaluating and selecting the best's papers out of many submissions that we received. We express sincerely our gratitude to the Mexican Society of Artificial Intelligence (SMIA) for its collaboration in elaborates this publication. Also we want to give special recognition to the Center for Computing Research of the National Polytechnic Institute for facilities given in order to achieve the success in the publication of this volume. The submission, reviewing, and selection process was supported for free by the EasyChair system, [www.EasyChair.org](http://www.EasyChair.org).

November 2012

Erandi Castillo Montiel  
Juan Carlos Chimal Eguía  
Abril Uriarte Arcia  
Luis Cabrera Rivera



# Table of Contents

## Índice

Page/Pág.

### Artificial Intelligence

An Adaptive-local and Geometric Approach for Automatic Chromosome Image Segmentation .....	3
<i>Viridiana Rubí Calzada-Navarrete, César Torres-Huitzil</i>	
BPSNR Blind Peak Signal-to-Noise Ratio .....	17
<i>Jesús Jaime Moreno Escoba, Beatriz Adriana Jaime Fonseca</i>	
Efficient Multi-dimensional Data Clustering using Particle Swarm Optimization .....	33
<i>Edgar A. Garcia-Martinez, Salvador Godoy-Calderon, Ricardo Barron-Fernandez, Javier Arellano-Verdejo</i>	
Detection and Denoising of Impulsive noise using Vector-Marginal filter applied to color images .....	45
<i>Agustín Tortolero, Alberto Rosales, Francisco Gallegos, Jean Marie Kinani</i>	
Image Recognition System for the Mexican Sign Language .....	57
<i>Fausto Pavel Priego Pérez, Jesús Manuel Olivares Ceja, José Félix Serrano Talamantes, Diana Naim Rivera Aguilar</i>	
Intelligent Tutoring System: An approach to the model of the student using multiple intelligences .....	69
<i>Rosa María Rodríguez-Aguilar, Jorge Manuel Alejandro Sánchez de Antuñano, Ana Lilia Laureano-Cruces</i>	
Refined Achromatic Zone Segmentation in Color Images .....	79
<i>Rodolfo Alvarado-Cervantes, Edgardo M. Felipe-Riveron</i>	
Retinal Image Biometric Coding using a Logarithmic Spiral and a Time Series Representation .....	91
<i>Fabiola M. Villalobos-Castaldi, Edgardo M. Felipe-Riverón</i>	
Reconstruction of Gaussian Realizations by a Non-Optimal Algorithm Based on the Clipping .....	107
<i>Y. Olvera, V. Kazakov</i>	
Segmentación del disco óptico en imágenes de retina mediante la transformada de Hough y los contornos activos .....	117
<i>José Abel de la Fuente-Arriaga, Edgardo Felipe-Riverón, Eduardo Garduño-Calderón</i>	
Un filtro de imágenes mediante correlación espacial .....	131
<i>Carlos D. Estrada Chávez, Edgardo M. Felipe-Riverón, Mariano Rivera Meraz y Johan Van Horebeek</i>	

## Communications & Computer Networks

Frequency Planning with an Optimized Frequency Reuse Distance for Fixed WiMAX Networks .....	149
<i>Miguel Sanchez Meraz, Carlos Sosa Paz, Alizari Duarte Calderon</i>	
Management model for availability and interconnection service in cloud computing .....	163
<i>Juan Ernesto Chávez Pacheco, Chadwick Carreto Arellano, Salvador Álvarez Ballesteros</i>	

## Databases & Software Technology

Integration of the POI API with Java for Information Processing from Heterogeneous Sources.....	181
<i>Pedro Galicia Galicia</i>	
Opportunistic Reasoning in DL Prolog .....	195
<i>José Oscar Olmedo-Aguirre</i>	
Towards a Multi-Agent System Architecture for Supply Chain Management .....	205
<i>Carlos R. Jaimez-González, Wulfrano A. Luna-Ramírez</i>	

## Digital Signal Processing

Sensorless speed control current for a Brushless DC motor .....	221
<i>Luis G. González García, Domingo Cortés</i>	

## Real Time & Automation

Computational model for video surveillance, dual detection and processing of relevant events .....	237
<i>Pamela Araceli Rangel Tirado, Luis Pastor Sánchez Fernández, Oleksiy Pogrebnyak</i>	
Control of a Lego Robot with App Inventor for Android.....	251
<i>Sergio Sandoval-Reyes, Alfonso Gutierrez-Aldana</i>	

## Intelligent Processing of Geospatial Information

A Multi-Agent System for Integrating Information from Heterogeneous Data Sources .....	269
<i>Carlos R. Jaimez-González, Wulfrano A. Luna-Ramírez, Luis A. Ramírez-Colín</i>	

## Micro technologies & Embedded Systems

A Hardware/Software Co-Execution Model Using Hardware Libraries for a SoPC Running Linux .....	285
<i>Alejandro Gómez-Conde, José de Jesús Mata-Villanueva, Marco A. Ramírez-Salinas, Luis A. Villa-Vargas</i>	
Real Time Stereo Vision with a modified Census transform and fast tracking in FPGA .....	295
<i>Juan Manuel Xicotencatl-Perez, Arturo Lezama-León, José Miguel Liceaga-Ortiz-De-La-Peña, Rubén O. Hernández-Terrazas</i>	
Towards the improvement of the linearity-efficiency trade-off in TXs for MedRadio .....	309
<i>Luis A. Sánchez-Gaspariano, Carlos Muñoz-Montero, Salvador A. Arroyo-Díaz, José M. Trevera-Rodríguez and Victor H. Ponce-Ponce</i>	

## Natural Language Processing

A Proposal of Automatic Error Correction in Text.....	321
<i>Wulfrano A. Luna-Ramírez, Carlos R. Jaimez-González</i>	
Recognizing Textual Entailment with Similarity Metrics .....	337
<i>Miguel Rios, Alexander Gelbukh</i>	

## Neural Networks & Unconventional Computation

Image transform based on an alpha-beta convolution model.....	349
<i>Antonio Alarcón-Paredes, Elías Ventura-Molina, Oleksiy Pogrebnyak, Amadeo Argüelles-Cruz</i>	
Security Token for Web Bank Applications Using a Linear and Congruential Random Number Generator.....	361
<i>Luis Orantes, Marco Ramírez, Pablo Manrique, Victor Ponce, Aniceto Orantes, Victor Salazar, Antonio Montes, Carlos Hernández, Eric Gómez</i>	

## Simulation & Modeling

Classical Realization of Grover's Quantum Search Algorithm using Toffoli gates.....	373
<i>Manuel-Iván Casillas-Del-Llano, Álvaro-Lorenzo Salas-Brito</i>	
Modelado y simulación de una barra actuada por dos rotores con movilidad espacial restringida .....	389
<i>O. Sánchez-Rodríguez, J. C. Chimal-Eguía and R. Barrón-Fernández</i>	
Modelo de Autómata Celular 1-dimensional para una EDP hiperbólica .....	407
<i>I. Huerta-Trujillo, J. C. Chimal-Eguía, S. V. Flores-Cortes, N. Sanchez-Salas</i>	

<b>Author Index</b> .....	423
Índice de autores	
<b>Editorial Board of the Volume</b> .....	425
Comité editorial del volumen	

# **Artificial Intelligence**

---



# An Adaptive-local and Geometric Approach for Automatic Chromosome Image Segmentation

Viridiana Rubí Calzada-Navarrete and César Torres-Huitzil  
vcalzada@tamps.cinvestav.mx, ctorres@tamps.cinvestav.mx

Information Technology Laboratory  
CINVESTAV-Tamaulipas, Mexico

**Abstract.** Karyotyping is a cytogenetic technique used to detect genetic aberrations. For many years, different computational techniques have been proposed to automate or assist in this task. One of the main stages in Karyotyping is image segmentation which consists of 3 main subtasks: object-background separation, object discrimination and union/overlap solver. This paper proposes a method towards an automatic chromosome image segmentation, specially for the two first subtasks. First a re-threshold process based on Sauvola local adaptive technique is applied to extract objects of interest from background. In the discrimination subtask, object geometric characteristics are used to discriminate between single and cluster of chromosomes. Experiments show encouraging results with 93% of precision in the re-threshold process and 93.5% in the chromosome discrimination subtask.

**Keywords:** Karyotyping, segmentation of cromosomas, re-threshold

## 1 Introduction

Chromosome karyotyping analysis is an important technique used to detect abnormalities and diagnosis. By analyzing the number and shape of chromosomes found in an image, it is possible to detect abnormalities.

A chromosome is comprised of two chromatids joined at a point called centromere and each chromatid turn consists of DNA strings that are positioned in different ways giving different structures and size variety for each chromosome. Chromosomes are commonly textured as light and dark strips called banding patterns.

Chromosomes are stained from cells of tissue, blood or amniotic fluid samples. Then a microscope is used to acquire an image of the cell division process.

An automatic karyotyping system involves a number of different steps: segmentation, feature extraction, and classification.

The segmentation stage resolves the problems that presents the image as: lighting problems that affect the extraction of the actual area of the chromosome which is caused by two factors, the amount of incident light from the source on the scene, such as interference with elements of capture, type of microscope the intensity of light from its lamp, leave stains on the background of the image that

affect the outline of objects to perform the extraction and the other, the amount of light reflected by the stained chromosomes.

Low contrast, the band patterns comprising the chromosome and the centromere have different gray levels causing disadvantages in the identification of the actual shapes of the chromosome.

Union and overlapping chromosomes, is the most complex problem to solve because it depends on the position and number of chromosomes involved. The characteristics of each chromosome allow to identify and implement appropriate processes to separate and reconstruct them in the cases of union and overlapping.

For the lighting and low contrast problems, it is proposed to work with local adaptive thresholding techniques, allowing to find a threshold that adapts to changes in the gray level image according to the neighborhood of each pixel. For discrimination of single and clusters of chromosomes, a technique based on geometric and morphological features objects is proposed

The rest of the work is organized as follows. Section 2 describes the related work and justify the proposed idea. In section 3, "re-threshold" and discrimination techniques are presented. Section 4 shows results and finally in section 5, conclusions and future work are presented.

## 2 Related work

There are different proposed solutions to chromosome image segmentation; some of them need the intervention of an expert, however the research continues working on the improvement of techniques and results.

The first step in analyzing a chromosome image is the segmentation of individual chromosomes and chromosome clusters from the image background. Unfortunately, the high variability in chromosome and background fluorescence intensities makes the utilization of a global threshold inappropriate.

Some works like [1,2,8] suggest that to remove background objects is not sufficient to apply a global technique such as Otsu. Because of banding patterns and that centromere has a low contrast, chromosomes become confused with the background of the image resulting in poor extraction. Works like [4] use a global threshold, guided only with the most representative gray level in the image according to its histogram.

In [2,7], authors show that local adaptive techniques provide good results providing a technique that divides the original image into fixed windows and then apply Otsu thresholding in each of those windows. Then the threshold matrix is interpolated in order to reach the original image size. L.Ji [6] proposes a scheme based on a "re-threshold" with two global thresholds. Allowing a little more adjusted parameters in the second threshold.

For the discrimination of chromosomes L.Ji [6] proposes 3 features to discriminate objects including size, roundness and gray level. The drawback is that these feature values do not maintain a pattern, as they vary according to the sizes and gray levels between images.

Geometric features are considered for discrimination in [2,4] to separate single objects according to small or large sizes. While for the rest, skeleton endpoints are used to assign clusters of chromosomes for those objects with more than 2 endpoints, in other case they are singles.

### 3 Proposed solution

The method that is proposed in this paper includes the first two steps of the segmentation stage: extraction, whose aim is to separate from the background single and cluster of chromosomes, which are then passed on to the discrimination, where each object (either a single chromosome or cluster) is analyzed to test whether it is composed by a single chromosome or many of them, these latter chromosomes are separated from the rest because they need to be disentangled.

The proposed approach is depicted in figure 1. The first 2 steps of the segmentation are enclosed in boxes.

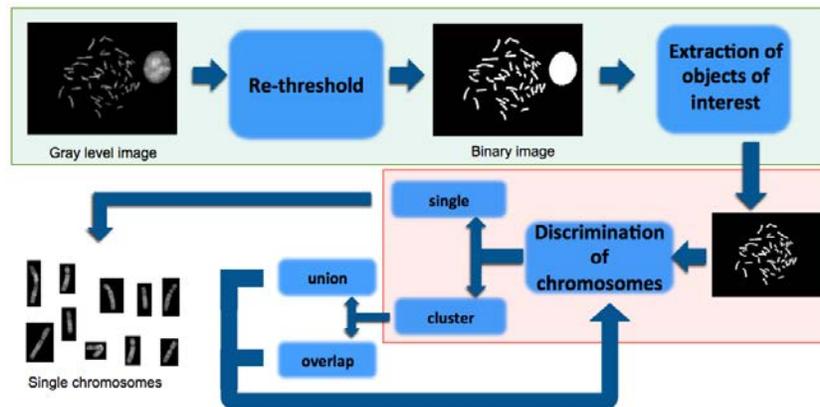


Fig. 1. Chromosomes segmentation process.

#### 3.1 Extraction of objects of interest

According to the characteristics of the image, it is proposed to work with a local adaptive technique and not with a modified global technique such as Otsu. For that reason, the Sauvola's technique is proposed, which assigns a threshold to each pixel according to its neighborhood, returning a binary image highlighting the edges of objects in the image.

Sauvola's method is the result of an improvement to the method of Niblack, since it effectively reduces the effect of poor lighting (stain) in the image, besides being less sensitive to the parameter  $k$  which is a constant that adjusts the amount

of pixels from the edge of the objects that are taken as part of it, and also appends a constant  $R$  which is the dynamic range of the standard deviation (see equation 1).

Sauvola's method adapts the threshold value according to the mean  $m(x, y)$  and local standard deviation  $\sigma(x, y)$ , which are calculated in a window of size  $b \times b$ . Calculating the threshold for each pixel as:

$$T(x, y) = m(x, y) \cdot \left[ 1 + k \cdot \left( \frac{\sigma(x, y)}{R} - 1 \right) \right] \quad (1)$$

where for binarization several authors determined that  $k = 0.2$ ,  $R = 128$  and  $b = 15$  work for most of the cases.

The local adaptive threshold can properly identify chromosomes when they have a good contrast. But in general, the image needs a contrast enhancement to detect all the chromosomes. It is important to consider that the contrast enhancement highlights the chromosomes and partially solves lighting problems in the image.

Therefore it is proposed a Re-threshold, which is carried out with a double binarization to extract all the chromosomes without being affected by the background.

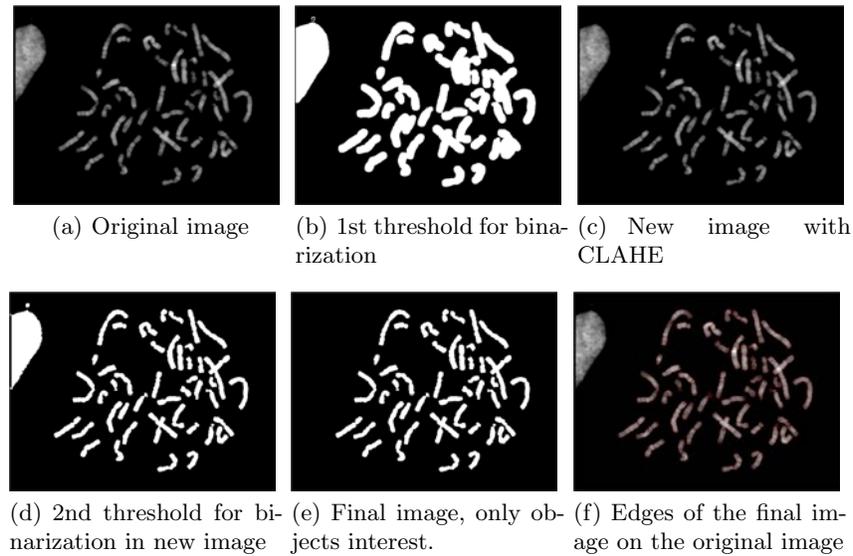
**Re-threshold** A first threshold is applied to the image in order to rescue areas that have a high probability of containing the objects of interest, thereby eliminating the background zones containing stains. Then the contrast on that area is enhanced with a CLAHE (Contrast Limited Adaptive Histogram Equalization) allowing to highlight the edges, bands and centromeres of chromosomes, and finally a new threshold to correctly extract the chromosomes is applied.

Thus, the proposed method for re-thresholding has the following steps:

1. The first threshold for the image  $I(x, y)$  is obtained with the Sauvola technique and the parameters are those recommended in the literature except the window size that was chosen smaller for this case in order to remove the area of interest without much detail,  $k = 0.2$ ,  $R = 128$  and  $b = 9$ .
2. The resulting image  $BW1(x, y)$  is used as a mask to extract the interest region from the original image.
3. In the new image  $I_m(x, y)$ , a contrast enhancement with (CLAHE) is applied.
4. The second threshold is obtained with Sauvola technique and the parameters are those recommended in the literature,  $k = 0.2$ ,  $R = 128$  and  $b = 15$ .

**Removal of unwanted objects** At this stage the binary image  $I1(x, y)$  of the previous step undergoes a procedure to identify and label each individual object. Therefore, each object can be selected and manipulated separately to remove undesirable objects with an erosion. If the object is not removed then it is a nuclei, otherwise the number of pixels and average gray level are evaluated. Then If the object is very small and dark, then it is eliminated.

In figure 2 is depicted how to perform the extraction of the interest objects.



**Fig. 2.** Object extraction from background

### 3.2 Discrimination

Once interest objects have been extracted and labeled, the second step is analyze their features for discriminating between singles and clusters of chromosomes. The discrimination is performed with 4 geometric characteristics of chromosomes: area, solidity, eccentricity and skeleton endpoints.

The evaluation of the geometric characteristics discriminate chromosomes that are in a group. A threshold for each feature is estimated for each image, so that the threshold will depend on the number of predominate objects (individual or clusters) in the evaluation. The discrimination is performed with a decision tree using these characteristics. This indicates that after evaluating the threshold for each feature, a set of three images named individuals, clusters or suspects will be generated. The last image moves to the next assessment in the decision tree.

The discrimination stage receives labeled image as input. The evaluation sequence is displayed on figure 3

The first evaluation is performed with the solidity which is the ratio between the area of the object and its convex polygon[3].

$$S = \frac{Area}{Convex\_poligon\_area} \quad (2)$$

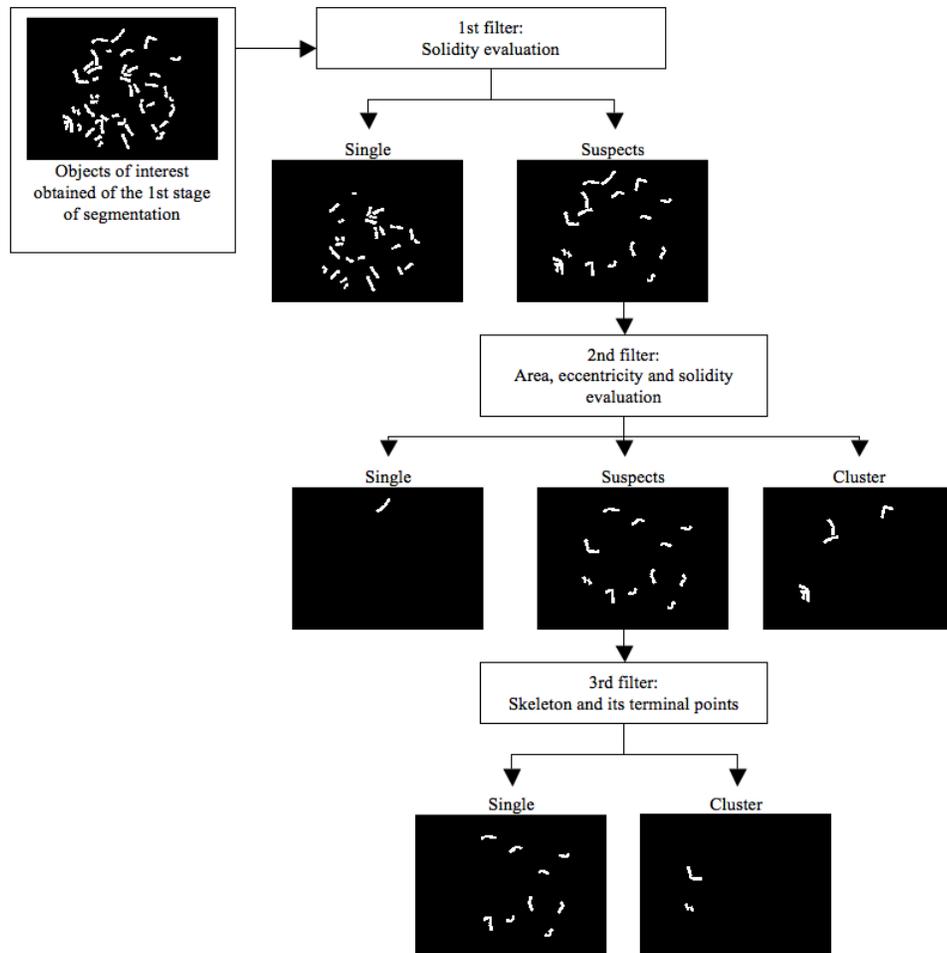


Fig. 3. Decision tree of chromosomes discrimination

where in 2: Area is the total number of pixels that form the object  $obj$ . For the binary image  $BW$  its area  $A$  is given by[5]:

$$A = \sum_{i=1}^n \sum_{j=1}^m BW[i, j] \quad (3)$$

and convex polygon area is the number of pixels that form the smallest convex polygon that can contain the selected object region[3].

Solidity  $S$  value is between 0 and 1. Objects with solidity close to 0 are those individual chromosomes whose form is completely straight, nearly straight or its size is small. While the rest are suspect chromosomes because still it is not possible to assign to a class. As there is no case where all chromosomes are overlapping, then the mean indicates the ratio between the number of group and individual chromosomes present in the image. Objects that are above the average of solidity are individual objects.

$$thS = \left( \frac{1}{n} \sum_{i=1}^n S_{obj_i} \right) - e \quad (4)$$

$$obj_i = \begin{cases} single & \text{if } S_{obj_i} \geq thS \\ suspect & \text{other case} \end{cases} \quad (5)$$

where  $e$  (equation 4) is a factor determined heristically to lower the threshold and rescue some individual chromosomes.

The second evaluation is the object area  $A$  and eccentricity  $E$ . It begins with the object area which also takes the mean of all objects in the image as a base to obtain the threshold. The mean indicates that if the object is very large (above the mean) is a group or very small (below the mean) is an individual object and small. Therefore, the thresholds for the area are as follows (see equations 6 - 8):

$$thA_s = 2.5\mu_A \quad (6)$$

$$thA_i = 0.5\mu_A \quad (7)$$

$$obj_i = \begin{cases} cluster & \text{if } A_{obj_i} > thA_s \\ single & \text{if } A_{obj_i} < thA_i \\ suspect & \text{other case} \end{cases} \quad (8)$$

For eccentricity which is the value obtained by dividing the minor axis between the major axis of the ellipse surrounding the object, it gives values between 0 and 1[3]. If the eccentricity is close to 1, it indicates that the object is a group of chromosomes while if it is close to zero corresponds to a single chromosome. There are cases in which the chromosome is curved in a "C" or "S" and could be confused with a group when it is not, for that reason there are two thresholds, one higher than the average eccentricity and very close to 1 which confirms that

it is a group and the other below the average and very close to 0 which ensures that it is a single well are the thresholds (see equations 9 - 10):

$$thE_s = \mu_E + \frac{3}{4}(max(E_{obj_i}) - \mu_E) \quad (9)$$

$$thE_i = \mu_E - \frac{4}{5}(\mu_E - min(E_{obj_i})) \quad (10)$$

Again, solidity of the object used to help to the eccentricity define the discrimination.

For this second evaluation, the result is divided into 3 new images, one containing individual chromosomes, other cluster chromosomes and other chromosomes suspects.

A final evaluation is assigned the last image chromosome suspects, for these objects their skeletons are calculated and counted the endpoints.

Endpoints of skeleton are those pixels where each ends the branches of the skeleton of the object. If the skeleton has more than two endpoints indicates that have 2 or more chromosomes in the object it is a group.

The skeleton is the result of thinning the object to obtain a line or sequence of pixels to retain the original shape of the object[4,5]. This is achieved through morphological operators of erosion ( $\epsilon$ ) and aperture ( $\gamma$ ). Then, the skeleton of an object O is defined as (see equations 11 - 14):

$$S(O) = \bigcup_{k=1}^K S_k(O) \quad (11)$$

with

$$S_k(O) = \epsilon_B^k(O) - \gamma_B[\epsilon_B^k(O)] \quad (12)$$

where B is the structuring element and  $\epsilon_B^k(O)$  indicates k successive erosions of O:

$$\epsilon_B^k(O) = \epsilon_B^k(\dots(\epsilon_B^2(\epsilon_B^1(O)))) \quad (13)$$

and K is the last iteration before O is the empty set:

$$K = max \{k \mid \epsilon_B(O) \neq \emptyset\} \quad (14)$$

Here there is the case where chromosomes with a centromere subtelocentric have a little separation between their chromatids therefore they could be confused with a cluster. In order to avoid such cases, endpoints are evaluated as follows: if the object has 2 endpoints is claimed to be single, if the object has more than 4 points is said to be a cluster, in other cases, it performs a trimming to the skeleton, which means that removes the branches having a certain number of pixels of the endpoint at the point of intersection of its main branch. To finally have the new endpoints and assign the chromosomes with 2 points to single and the rest to clusters.

The figure 4 depicted the values of the geometric characteristics of some chromosomes.

Type	Object	Solidity	Eccentricity	Area	Endpoints
Cluster		0.5727	0.6657	1091	4
Cluster		0.6150	0.8079	872	3
Single		0.7219	0.3366	553	2
Single		0.8853	0.3340	386	2
Single		0.9136	0.2597	370	2

Fig. 4. Geometric characteristics

## 4 Result

The dataset used in this work is obtained from <http://bioimlab.dei.unipd.it>. It is composed of 162 images with  $768 \times 576$  pixels resolution. The images do not necessarily contain a whole set of 46 chromosomes because the image may be spread over different images[2].

Image Q-band are obtained by staining the chromosomes with quinacrine, a fluorescent dye that concentrates in different regions of the chromosomes, giving rise to the characteristic banding patterns that identify the different chromosome types. The images, thus, appear as a dark background onto which the chromosomes stand out with a bright and dark banding.

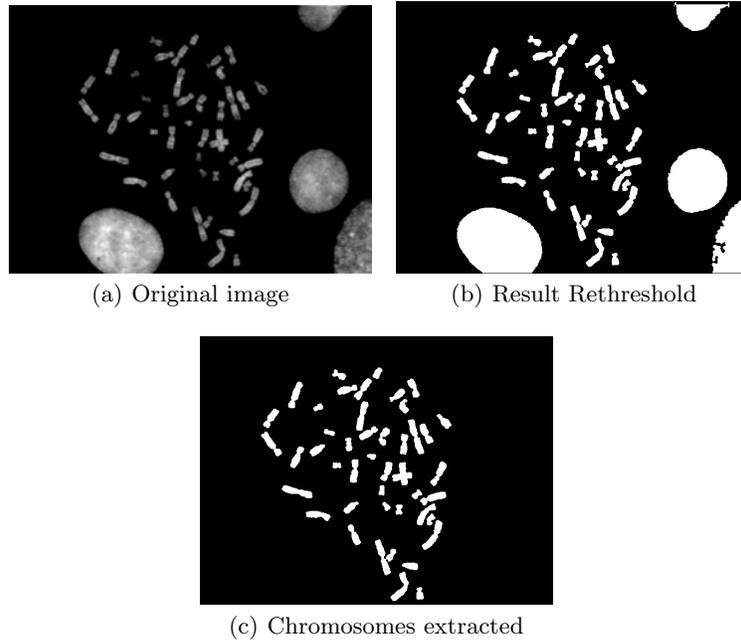
### 4.1 Extraction of objects of interest

Figure 5 shows the result of applying the "Re-threshold" and perform the extraction of objects of interest on an image containing nuclei.

It has two evaluations: 1) evaluates the successful extraction of total images and 2) evaluates the area error of each image.

Evaluation 1: on the the dataset, the method has successfully extracted chromosomes from 148 images , in 3 images was not possible to completely eliminate the unwanted objects and in 11 images some chromosomes were lost (see figure 6).

Evaluation 2: For each image, the are error was obtained with respect to a reference image. This is possible if the result of the proposed method (M) is



**Fig. 5.** First step of segmentation

compared with a reference (R) which includes all pixels of interest. View all 4 area error measurements in table 1.

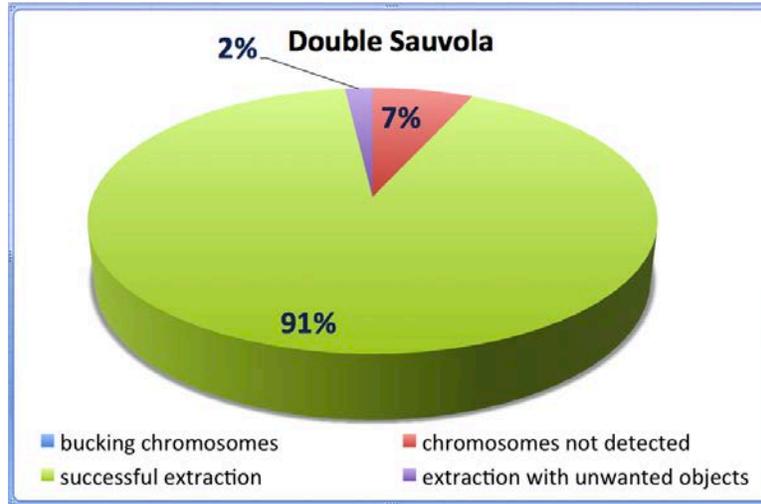
Measure	pixel value
True Positive (TP)	$M_{ij} = 1 \wedge R_{ij} = 1$
True Negative (TN)	$M_{ij} = 0 \wedge R_{ij} = 0$
False Positive (FP)	$M_{ij} = 1 \wedge R_{ij} = 0$
False Negative (FN)	$M_{ij} = 0 \wedge R_{ij} = 1$

**Table 1.** area error measurements

The percentage of area successfully extracted by the method is calculated using the above measures in the following indicators [7]:

$$precision = \frac{TP}{TP + FP} \quad (15)$$

$$accuracy = \frac{TP + TN}{TP + TN + FP + FN} \quad (16)$$



**Fig. 6.** Result of success of the first segmentation stage of chromosomes.

$$recall = \frac{TP}{TP + FN} \quad (17)$$

$$Fscore = \frac{2 \times precision \times recall}{precision + recall} \quad (18)$$

The dataset has no reference images to evaluate the error area. So we take randomly 34 images which are manually binarized to create reference images. The results obtained by the proposed method and adaptive Otsu are evaluated with the reference images. The results are shown in table 2.

Method	precision	recall	accuracy	Fscore
Proposed	0.9611	0.8849	0.9924	0.9200
adaptative Otsu	0.9793	0.7778	0.9887	0.8659

**Table 2.** Evaluation on 20% of the total dataset

The table 2 shows the comparison of the proposed method with adaptive Otsu. The recall indicates that our proposal obtains a greater number of pixels of interest than adaptive Otsu, therefore, the Fscore indicates that the proposed approach has a higher success rate.

## 4.2 Discrimination

For this step, the resultant binary images of the first stage were used in the discrimination process (93% of images of Q-banding).

For the 151 images that pass this second stage of segmentation has been obtained a script which contains the number of chromosomes in each cluster having one of the images according to the results yielded by the first stage of segmentation. Each image is evaluated and compared with this script and the results obtained are shown in table 3.

<b>Total of clusters</b>	508
<b>Clusters detected</b>	473
<b>Clusters undetected</b>	35
<b>percentage of success</b>	93.5%

**Table 3.** Result of chromosomes clusters found in the stage of discrimination in Q-banding images.

Groups of chromosomes that correspond to 6.5% were not detected, those who are connected at one end forming a single chromosome, some examples can be seen in figure 7.



**Fig. 7.** Examples of clusters of chromosomes undetected.

## 5 Conclusions and future work

According to Poletti et al.[7], adaptive techniques provides the best results in the segmentation as confirmed in this work by applying the proposed technique

based on Sauvola thresholding. Obtained results are competitive with others reported in the literature.

The proposed technique of "Double Sauvola" permits to find the appropriate threshold for each pixel, also it avoids cutting the chromosomes in the centromere or any of their bands and it reduces the number of chromosomes together.

The former evidence indicating that the segmentation from the perspective of a local adaptive method gives good results in segmenting images of chromosomes, has been confirmed by obtaining the 93% of successful images of 162 image.

The discrimination method allows to identify unions and overlaps with a 93.5% of success. Although the results are already competitive, we continue working on parameters tuning to rescue the chromosomes corresponding to the FP (chromosomes with separated chromatids) and FN (chromosomes connected at one end) in discrimination of chromosomes.

As future work the separation of cluster of chromosomes will be explored.

## References

1. Gady Agam and Its'hak Dinstein. Geometric separation of partially overlapping nonrigid objects applied to automatic chromosome classification. *IEEE Transactions on pattern analysis and machine intelligence*, 19(11), november 1997.
2. Enrico Grisan, Enea Poletti, and Alfredo Ruggeri. Automatic segmentation and disentangling of chromosomes in q-band prometaphase images. *IEEE Transactions on Information Technology in Biomedicine*, 13(4), july 2009.
3. Enamul Huque. Shape analysis and measurement for the hela cell classification of cultured cells in high throughput screening. Master's thesis, School of Humanities & Informatics University of Skvde, Sweden, February 2006.
4. Sahar Jahani, S. Kamaledin Setarehdan, and Emadedin Fatemizadeh. Automatic identification of overlapping/touching chromosomes in microscopic images using morphological operators. *Machine Vision and Image Processing (MVIP)*, 2011.
5. Ramesh Jain, Rangachar Kasturi, and Brian G. Schunck. *Machine vision*. McGraw-Hill, 1995.
6. L. Ji. Fully automatic chromosome segmentation. *Cytometry*, 17:196–208, 1994.
7. Enea Poletti, Francesca Zappelli, Alfredo Ruggeri, and Enrico Grisan. A review of thresholding strategies applied to human chromosome segmentation. *Computer Methods and Programs in Biomedicine*, 2012.
8. Gunter Rittery and Le Gao. Automatic segmentation of metaphase cells based on global context and variant analysis. *Pattern Recognition*, 41:38–55, 2008.



# **$\mathcal{B}$ PSNR**

## **Blind Peak Signal-to-Noise Ratio**

Jesús Jaime Moreno Escobar and Beatriz Adriana Jaime Fonseca

Superior School of Mechanical and Electrical Engineering,  
National Polytechnic Institute, Mexico City, Mexico.  
jmorenoe@ipn.mx

**Abstract.** The aim of this work is to define a no-referenced perceptual image quality estimator applying the perceptual concepts of the Chromatic Induction Model. The approach consists in comparing the received image, presumably degraded, against the perceptual versions (different distances) of this image degraded by means of a Model of Chromatic Induction, which uses some of the human visual system properties. Also we compare our model with a original estimator in image quality assessment, PSNR. Results are highly correlated with the ones obtained by PSNR but this proposal does not need an original image or a reference one in order to give an estimation of the quality of the degraded image.

*Keywords:* Human Visual System, Contrast Sensitivity Function, Perceived Images, Wavelet Transform, Peak Signal-to-Noise Ratio, No-Reference Image Quality Assessment.

## **1 Introduction**

The early years of the 21st century have witnessed a tremendous growth in the use of digital images as a means for representing and communicating information. A significant literature describing sophisticated theories, algorithms, and applications of digital image processing and communication has evolved. A considerable percentage of this literature is devoted to methods for improving the appearance of images, or for maintaining the appearance of images that are processed. Nevertheless, the quality of digital images, processed or otherwise, is rarely perfect. Images are subject to distortions during acquisition, compression, transmission, processing, and reproduction. To maintain, control, and enhance the quality of images, it is important for image acquisition, management, communication, and processing systems to be able to identify and quantify image quality degradations. The development of effective automatic image quality assessment systems is a necessary goal for this purpose. Yet, until recently, the field of image quality assessment has remained in a nascent state, awaiting new models of human vision and of natural image structure and statistics before meaningful progress could be made.

Nowadays, Mean Squared Error (MSE) is still the most used quantitative performance metrics and several image quality measures are based on it, being

Peak Signal-to-Noise Ratio (PSNR) the best example. But some authors like Wang and Bovik in [1, 2] consider that MSE is a poor algorithm, to be used in quality assessment systems. Therefore it is important to know what is the MSE and what is wrong with it, in order to propose new metrics that fulfill the properties of human visual system and keeps the favorable features that the MSE has.

In this way, let  $f(i, j)$  and  $\hat{f}(i, j)$  represent two images being compared and the size of them is the number of intensity samples or pixels. Being  $f(i, j)$  the original reference image, which has to be considered with perfect quality, and  $\hat{f}(i, j)$  a distorted version of  $f(i, j)$ , whose quality is being evaluated. Then, the MSE and the PSNR are, respectively, defined as:

$$MSE = \frac{1}{NM} \sum_{i=1}^N \sum_{j=1}^M [f(i, j) - \hat{f}(i, j)]^2 \quad (1)$$

and

$$PSNR = 10 \log_{10} \left( \frac{\mathcal{G}_{max}^2}{MSE} \right) \quad (2)$$

where  $\mathcal{G}_{max}$  is the maximum possible intensity value in  $f(i, j)$  ( $M \times N$  size). Thus, for gray-scale images that allocate 8 bits per pixel (bpp)  $\mathcal{G}_{max} = 2^8 - 1 = 255$ . For color images the PSNR is defined as in the Equation 2, whereas the color MSE is the mean among the individual MSE of each component.

An important task in image compression systems is to maximize the correlation among pixels, because the higher correlation at the preprocessing, the more efficient algorithm postprocessing. Thus, an efficient measure of image quality should take in to account the latter feature. In contrast to this, MSE does not need any positional information of the image, thus pixel arrangement is ordered as a one-dimensional vector.

Both MSE and PSNR are extensively employed in the image processing field, since these metrics have favorable properties, such as:

1. A convenient metrics for the purpose of algorithm optimization. For example in JPEG2000, MSE is used both in Optimal Rate Allocation [3, 4] and Region of interest [5, 4]. Therefore MSE can find solutions for these kind of problems, when is combined with the instruments of linear algebra, since it is differentiable.
2. By definition MSE is the difference signal between the two images being compared, giving a clear meaning of the overall error signal energy.

## 2 Image Quality Assessment

### 2.1 Full Reference (FR)

**Bottom-Up Approaches** Psychological and physiological studies in the past century have gained us a tremendous amount of knowledge about the human visual system (HVS). Still, although much is known about the mechanisms of

early, front-end vision, much more remains to be learned of the later visual pathways and the general higher level functions of the visual cortex. While the knowledge is far from complete, current models of visual information processing mechanisms have become sufficiently sophisticated that it is of interest to explore whether it is possible to deploy them to predict the performance of simple human visual behaviors, such as image quality evaluation.

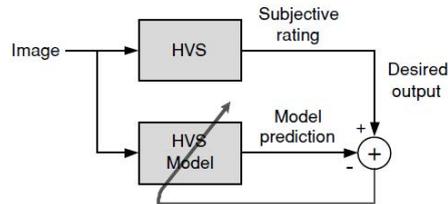
Bottom up approaches to image quality assessment are those methods that attempt to simulate well modeled functionalities of the HVS, and integrate these in the design of quality assessment algorithms that, hopefully, perform similar to the HVS in the assessment of image quality. In this chapter we begin with a brief description of relevant aspects of the anatomy and psychophysical features of the HVS. This description will focus on those HVS features that contribute to current engineering implementations of perceptual image quality measures.

Most systems that attempt to incorporate knowledge about the HVS into the design of image quality measures use an error sensitivity framework, so that the errors between the distorted image and reference image are perceptually quantized according to HVS characteristics.

**Top-Down Approaches** The bottom-up approaches to image quality assessment described in the last subsection (2.1) attempt to simulate the functional components in the human visual system that may be relevant to image quality assessment. The underlying goal is to build systems that work in the same way as the HVS, at least for image quality assessment tasks. By contrast, the top-down systems simulate the HVS in a different way. These systems treat the HVS as a black box, and only the input output relationship is of concern. A top-down image quality assessment system may operate in a manner quite different from that of the HVS, which is of little concern, provided that it successfully predicts the image quality assessment behavior of an average human observer.

One obvious approach to building such a top-down system is to formulate it as a supervised machine learning problem, as illustrated in Fig. 1. Here the HVS is treated as a black box whose input/output relationship is to be learned. The training data can be obtained by subjective experimentation, where a large number of test images are viewed and rated by human subjects. The goal is to train the system model so that the error between the desired output (subjective rating) and the model prediction is minimized. This is generally a regression or function approximation problem. Many techniques are available to attack these kinds of problems.

Unfortunately, direct application of this method is problematic, since the dimension of the space of all images is the same as the number of pixels in the image. Furthermore, subjective testing is expensive and a typical extensive subjective experiment would be able to include only several hundred test images—hardly an adequate coverage of the image space. Assigning only a single sample at each quadrant of a ten dimensional space requires a total of 1024 samples, and the dimension of the image space is in the order of thousands to millions. An excellent example of the problem of dimensionality.



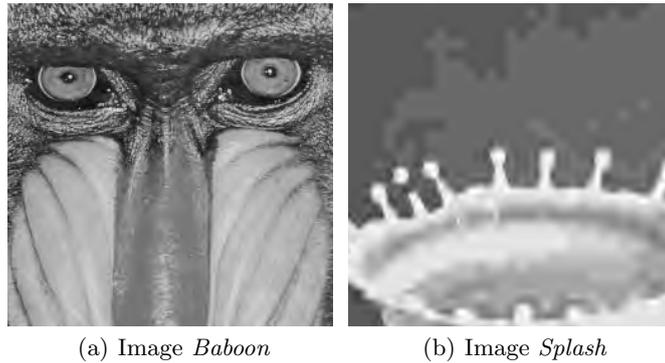
**Fig. 1.** Learning HVS.

One method that might be useful to overcome this problem is by dimension reduction. The idea is to map the entire image space onto a space of much lower dimensionality by exploiting knowledge of the statistical distribution of typical images in the image space. Since natural images have been found to exhibit strong statistical regularities, it is possible that the cluster of typical natural images may be represented by a low dimensional manifold, thus reducing the number of sample images that might be needed in the subjective experiments. However, dimension reduction is no trivial task. Indeed, no dimension reduction technique has been developed to reduce the dimension of natural images to 10 or less (otherwise, extremely efficient image compression techniques would have been proposed on the basis of such reduction). Consequently, using a dimension reduction approach for general purpose image quality assessment remains quite difficult. Nonetheless, such an approach may prove quite effective in the design of application specific quality assessment systems, where the types of distortions are fixed and known and may be described by a small number of parameters.

## 2.2 No-Reference

No-reference (NR) image quality assessment is, perhaps, the most difficult (yet conceptually simple) problem in the field of image analysis. By some means, an objective model must evaluate the quality of any given real world image, without referring to an original high quality image. On the surface, this seems to be a mission impossible. How can the quality of an image be quantitatively judged without having a numerical model of what a good/bad quality image is supposed to look like? Yet, amazingly, this is quite an easy task for human observers. Humans can easily identify high quality images versus low quality images, and, furthermore, they are able to point out what is right and wrong about them without seeing the original. Moreover, humans tend to agree with each other to a pretty high extent. For example, without looking at the original image, probably every reader would agree that the noisy, blurry, and JPEG2000 compressed images in Fig. 2 have lower quality than the luminance shifted and contrast stretched images.

Before developing any algorithm for image quality assessment, a fundamental question that must be answered is what source of information can be used to evaluate the quality of images. Clearly, the human eye/brain system is making use



**Fig. 2.**  $256 \times 256$  patches (cropped for visibility) of Images *Baboon* and *Splash* distorted by means of JPEG2000 compression, although both images have the same objective quality (PSNR=30dB), their visual quality is very different.

of a very substantial and effective pool of information about images in making subjective judgments of image quality.

Three types of knowledge may be employed in the design of image quality measures: knowledge about the original high quality image, knowledge about the distortion process, and knowledge about the human visual system (HVS). In FR quality assessment, the high quality original image is known a priori. In NR quality assessment, however, the original image is absent, yet one can still assume that there exists a high quality original image, of which the image being evaluated is a distorted representation. It is also reasonable to make a further assumption that such a conjectured original image belongs to the set of typical natural images.

It is important to realize that the cluster of natural images occupies an extremely tiny portion in the space of all possible images. This potentially provides a strong prior knowledge about what these images should look like. Such prior knowledge could be a precious source of information for the design of image quality measures. Models of such natural scenes attempt to describe the class of high quality original images statistically. Interestingly, it has been long conjectured in computational neuroscience that the HVS is highly adapted to the natural visual environment, and that, therefore, the modeling of natural scenes and the HVS are dual problems.

Knowledge about the possible distortion processes is another important information source that can be used for the development of NR image quality measures. For example, it is known that blur and noise are often introduced in image acquisition and display systems and reasonably accurate models are sometimes available to account for these distortions. Images compressed using block based algorithms such as JPEG often exhibit highly visible and undesirable blocking artifacts. Wavelet based image compression algorithms operating at low bit rates can blur images and produce ringing artifacts near discontinuities.

Of course, all of these types of distortions are application dependent. An application specific NR image quality assessment system is one that is specifically

designed to handle a specific artifact type, and that is unlikely to be able to handle other types of distortions. The question arises, of course, whether an application specific NR system is truly reference free, since much information about the distorted image is assumed. However, nothing needs to be assumed about the original image, other than, perhaps models derived from natural scene statistics or other natural assumptions. Since the original images are otherwise unknown, we shall continue to refer to more directed problems such as these as application specific NR image quality assessment problems.

Of course, a more complex system that includes several modes of artifact handling might be constructed and that could be regarded as approaching general purpose NR image quality assessment. Before this can happen, however, the various components need to be designed. Fortunately, in many practical application environments, the distortion processes involved are known and fixed. The design of such application specific NR quality assessment systems appears to be much more approachable than the general, assumption free NR image quality assessment problem. Very little, if any, meaningful progress has been made on this latter problem.

Owing to a paucity of progress in other application specific areas, this chapter mainly focuses on NR image quality assessment methods, which are designed for assessing the quality of compressed images. In particular, attention is given to a spatial domain method and a frequency domain method for block based image compression, and a wavelet domain method for wavelet based image compression.

### 3 The $\mathcal{B}$ PSNR Algorithm

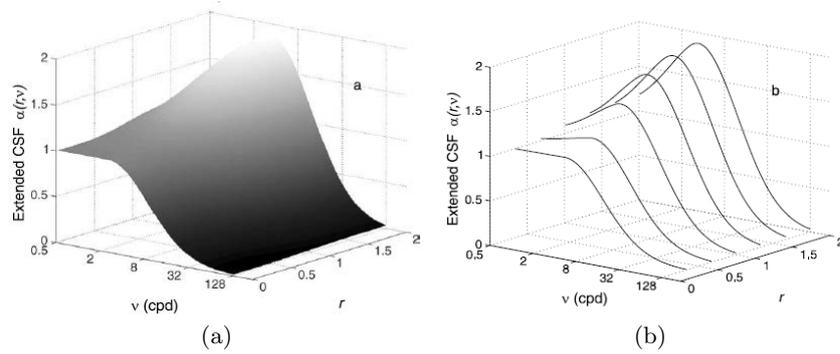
#### 3.1 Chromatic Induction Wavelet Model

The *Chromatic Induction Wavelet Model* (CIWaM) [6] is a low-level perceptual model of the HVS. It estimates the image perceived by an observer at a distance  $d$  just by modeling the perceptual chromatic induction processes of the HVS. That is, given an image  $\mathcal{I}$  and an observation distance  $d$ , CIWaM obtains an estimation of the perceptual image  $\mathcal{I}_p$  that the observer perceives when observing  $\mathcal{I}$  at distance  $d$ . CIWaM is based on just three important stimulus properties: spatial frequency, spatial orientation and surround contrast. This three properties allow to unify the chromatic assimilation and contrast phenomena, as well as some other perceptual processes such as saliency perceptual processes [7].

The CIWaM model takes an input image  $\mathcal{I}$  and decomposes it into a set of wavelet planes  $\omega_{s,o}$  of different spatial scales  $s$  (i.e., spatial frequency  $\nu$ ) and spatial orientations  $o$ . It is described as:

$$\mathcal{I} = \sum_{s=1}^n \sum_{o=v,h,dgl} \omega_{s,o} + c_n, \quad (3)$$

where  $n$  is the number of wavelet planes,  $c_n$  is the residual plane and  $o$  is the spatial orientation either *vertical*, *horizontal* or *diagonal*.



**Fig. 3.** (a) Graphical representation of the e-CSF ( $\alpha_{s,o,i}(r, \nu)$ ) for the luminance channel. (b) Some profiles of the same surface along the Spatial Frequency ( $\nu$ ) axis for different centersurround contrast energy ratio values ( $r$ ). The psychophysically measured CSF is a particular case of this family of curves (concretely for  $r = 1$ ).

The perceptual image  $\mathcal{I}_p$  is recovered by weighting these  $\omega_{s,o}$  wavelet coefficients using the *extended Contrast Sensitivity Function* (e-CSF, Fig. 3). The e-CSF is an extension of the psychophysical CSF [8] considering spatial surround information (denoted by  $r$ ), visual frequency (denoted by  $\nu$ , which is related to spatial frequency by observation distance) and observation distance ( $d$ ). Perceptual image  $\mathcal{I}_p$  can be obtained by

$$\mathcal{I}_p = \sum_{s=1}^n \sum_{o=v,h,dgl} \alpha(\nu, r) \omega_{s,o} + c_n, \quad (4)$$

where  $\alpha(\nu, r)$  is the e-CSF weighting function that tries to reproduce some perceptual properties of the HVS. The term  $\alpha(\nu, r) \omega_{s,o} \equiv \omega_{s,o;\rho,d}$  can be considered the *perceptual wavelet coefficients* of image  $\mathcal{I}$  when observed at distance  $d$  and is written as:

$$\alpha(\nu, r) = z_{ctr} \cdot C_d(\dot{s}) + C_{min}(\dot{s}). \quad (5)$$

This function has a shape similar to the e-CSF and the three terms that describe it are defined as:

$z_{ctr}$  Non-linear function and estimation of the central feature contrast relative to its surround contrast, oscillating from zero to one, defined by:

$$z_{ctr} = \frac{\left[ \frac{\sigma_{cen}}{\sigma_{sur}} \right]^2}{1 + \left[ \frac{\sigma_{cen}}{\sigma_{sur}} \right]^2} \quad (6)$$

being  $\sigma_{cen}$  and  $\sigma_{sur}$  the standard deviation of the wavelet coefficients in two concentric rings, which represent a center-surround interaction around each coefficient.

$C_d(\dot{s})$  Weighting function that approximates to the perceptual e-CSF, emulates some perceptual properties and is defined as a piecewise Gaussian function [8], such as:

$$C_d(\dot{s}) = \begin{cases} e^{-\frac{\dot{s}^2}{2\sigma_1^2}}, & \dot{s} = s - s_{thr} \leq 0, \\ e^{-\frac{\dot{s}^2}{2\sigma_2^2}}, & \dot{s} = s - s_{thr} > 0. \end{cases} \quad (7)$$

$C_{min}(\dot{s})$  Term that avoids  $\alpha(\nu, r)$  function to be zero and is defined by:

$$C_{min}(\dot{s}) = \begin{cases} \frac{1}{2} e^{-\frac{\dot{s}^2}{2\sigma_1^2}}, & \dot{s} = s - s_{thr} \leq 0, \\ \frac{1}{2}, & \dot{s} = s - s_{thr} > 0. \end{cases} \quad (8)$$

taking  $\sigma_1 = 2$  and  $\sigma_2 = 2\sigma_1$ . Both  $C_{min}(\dot{s})$  and  $C_d(\dot{s})$  depend on the factor  $s_{thr}$ , which is the scale associated to 4 cycles per degree when an image is observed from the distance  $d$  with a pixel size  $l_p$  and one visual degree, whose expression is defined by Equation 9. Where  $s_{thr}$  value is associated to the e-CSF maximum value.

$$s_{thr} = \log_2 \left( \frac{d \tan(1^\circ)}{4 l_p} \right) \quad (9)$$



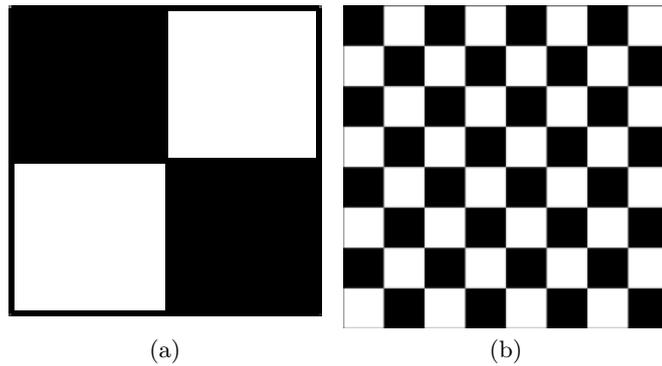
(a) Original image    (b)  $d=30$  cm.    (c)  $d=100$  cm.    (d)  $d=200$  cm.

**Fig. 4.** (a) Original color image *Lenna*. (b)-(d) Perceptual images obtained by CIWaM at different observation distances  $d$ .

Fig. 4 shows three examples of CIWaM images of *Lenna*, calculated by Eq. 4 for a 19 inch monitor with 1280 pixels of horizontal resolution, at  $d = \{30, 100, 200\}$  centimeters.

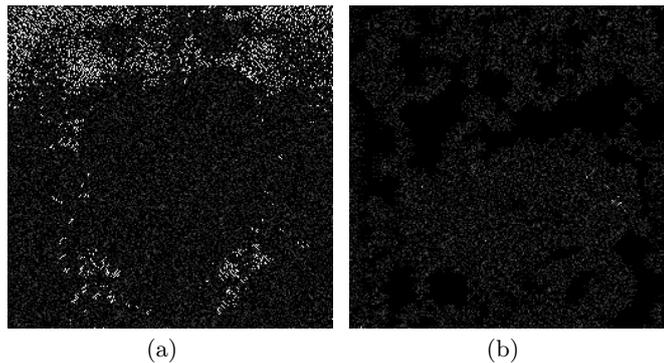
### 3.2 Basics

In the no-referenced image quality issue, there is only a distorted version  $\hat{f}(i, j) = A[f(i, j)]$  that is compared with  $f(i, j)$ , being  $A$  a distortion model and the unknown original image  $f(i, j)$  is considered a pattern  $\mathcal{T}([0,1;1,0])$  like a chessboard (Figs. 5) with the same size of  $\hat{f}(i, j)$ . The difference between these two images depends on the features of the distortion model  $A$ . For example, blurring, contrast change, noise, JPEG blocking or JPEG2000 wavelet ringing.



**Fig. 5.** (a) Pattern  $[0,1;1,0]$  or  $\mathcal{T}$ . (b) Pattern  $\mathcal{T}$  repeated sixteen times.

In Fig. 2, the images *Baboon* and *Splash* are compressed by means of JPEG2000. These two images have the same PSNR=30 dB when compared to their corresponding original image, that is, they have the same numerical degree of distortion (i.e. the same objective image quality PSNR). But, their subjective quality is clearly different, showing the image *Baboon* a better visual quality. Thus, for this example, PSNR and perceptual image quality has a small correlation. On the image *Baboon*, high spatial frequencies are dominant. A modification of these high spatial frequencies by  $A$  induces a high distortion, resulting a lower PSNR, even if the modification of these high frequencies are not perceived by the HVS. In contrast, on image *Splash*, mid and low frequencies are dominant. Modification of mid and low spatial frequencies also introduces a high distortion, but they are less perceived by the HVS. Therefore, correlation of PSNR against the opinion of an observer is small. Fig. 6 shows the diagonal high spatial frequencies of these two images, where there are more high frequencies in image *Baboon*.



**Fig. 6.** Diagonal spatial orientation of the first wavelet plane of Images (a) *Baboon* and (b) *Splash* distorted by JPEG2000 with PSNR=30dB.

If a set of distortions  $\hat{f}_k(i, j) = \Lambda_k[f(i, j)]$  is generated and indexed by  $k$  (for example, let  $\Lambda$  be a blurring operator), the image quality of  $\hat{f}_k(i, j)$  evolves while varying  $k$ , being  $k$ , for example, the degree of blurring. Hence, the evolution of  $\hat{f}_k(i, j)$  depends on the characteristics of the original  $f(i, j)$ . Thus, when increasing  $k$ , if  $f(i, j)$  contains many high spatial frequencies the PSNR rapidly decreases, but when low and mid frequencies predominated PSNR slowly decreases.

Similarly, the HVS is a system that induces a distortion on the observed image  $\hat{f}(i, j)$ , whose model is predicted by CIWaM. Hence, CIWaM is considered a HSV particular distortion model  $\Lambda \equiv \text{CIWaM}$  that generates a perceptual image  $\hat{f}_\rho(i, j) \equiv \mathcal{I}_\rho$  from an observed image  $f(i, j) \equiv \mathcal{I}$ , i.e  $\mathcal{I}_\rho = \text{CIWaM}[\mathcal{I}]$ . Therefore, a set of distortions is defined as  $\Lambda_k \equiv \text{CIWaM}_d$ , being  $d$  the observation distance. That is, a set of perceptual images is defined  $\mathcal{I}_{\rho,d} = \text{CIWaM}_d[\mathcal{I}]$  which is considered a set of perceptual distortions of the hypothetical image  $\mathcal{I}$ .

When image  $\hat{f}(i, j)$  is observed at distance  $\bar{d}$  and this distance is reduced, the artifacts, if this possesses, are better perceived. In contrast,  $\hat{f}(i, j)$  is observed from a far distance human eyes cannot perceive their artifacts, in consequence, the perceptual image quality of the distorted image is always high. The distance where the observer can perceive the best image quality of image  $\hat{f}(i, j)$  is considered as the distance  $D$ .

### 3.3 Methodology

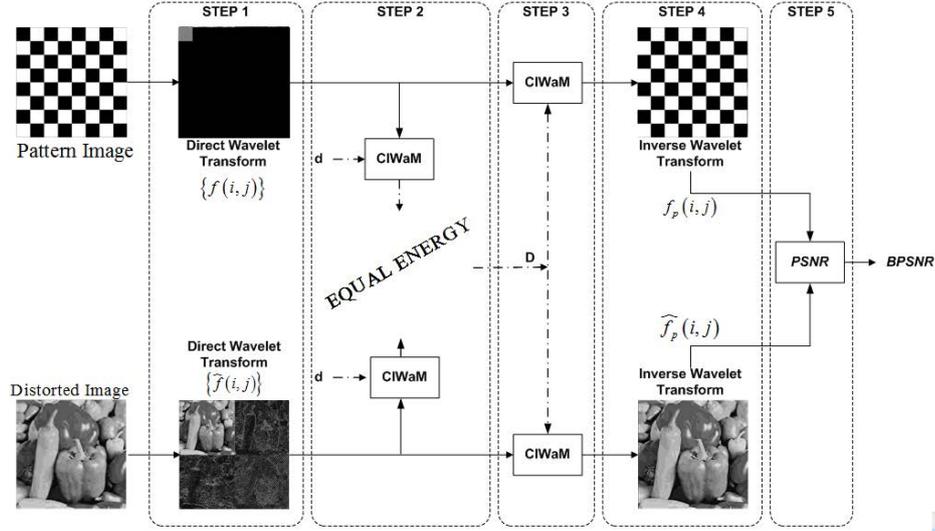
Let  $f(i, j)$  and  $\hat{f}(i, j) = \Lambda[f(i, j)]$  be an pattern image and a distorted image, respectively.  $\mathcal{BPSNR}$  methodology is based on finding a distance  $D$ , where there is no perpetual difference between the wavelet energies of the images  $f(i, j)$  and  $\hat{f}(i, j)$ , when an observer observe them at  $d$  centimeters of observation distance. So measuring the PSNR of  $\hat{f}(i, j)$  at  $D$  will yield a fairer and blind perceptual evaluation of its image quality.

$\mathcal{BPSNR}$  algorithm is divided in five steps, which is summarized by the Figure 7 and described as follows:

**Step 1: Wavelet Transformation** Forward wavelet transform of images  $f(i, j)$  and  $\hat{f}(i, j)$  is performed using Eq. 3, obtaining the sets  $\{\omega_{s,o}\}$  and  $\{\hat{\omega}_{s,o}\}$ , respectively. The employed analysis filter is the Daubechies 9-tap/7-tap filter (Table 1).

**Table 1.** 9/7 Analysis Filter.

Analysis Filter		
i	Low-Pass Filter $h_L(i)$	High-Pass Filter $h_H(i)$
0	0.6029490182363579	1.115087052456994
$\pm 1$	0.2668641184428723	-0.5912717631142470
$\pm 2$	-0.07822326652898785	-0.05754352622849957
$\pm 3$	-0.01686411844287495	0.09127176311424948
$\pm 4$	0.02674875741080976	



**Fig. 7.** Methodology for No-Reference PSNR weighting by means of CIWaM. Both Pattern and Distorted images are wavelet transformed. The distance  $D$  where the energy of perceptual images obtained by CIWaM are equal is found. Then, PSNR of perceptual images at  $D$  is calculated, obtaining the  $\mathcal{B}$ PSNR metrics.

**Step 2: Distance  $D$**  The total energy measure or the *deviation signature*[9]  $\bar{\varepsilon}$  is the absolute sum of the wavelet coefficient magnitudes, defined by [10]

$$\bar{\varepsilon} = \sum_{n=1}^N \sum_{m=1}^M |x(m, n)| \quad (10)$$

where  $x(m, n)$  is the set of wavelet coefficients, whose energy is being calculated, being  $m$  and  $n$  the indexes of the coefficients. Basing on the traditional definition of a calorie, the units of  $\bar{\varepsilon}$  are wavelet calories (wCal) and can also be defined by Eq. 10, since one wCal is the energy needed to increase the absolute magnitude of a wavelet coefficient by one scale.

From wavelet coefficients  $\{\omega_{s,o}\}$  and  $\{\hat{\omega}_{s,o}\}$  the corresponding perceptual wavelet coefficients  $\{\omega_{s,o,\rho,\tilde{d}}\} = \alpha(\nu, r) \cdot \omega_{s,o}$  and  $\{\hat{\omega}_{s,o,\rho,\tilde{d}}\} = \alpha(\nu, r) \cdot \hat{\omega}_{s,o}$  are obtained by applying CIWaM with an observation distance  $\tilde{d}$ . Therefore, Equation 11 expresses the relative wavelet energy ratio  $\varepsilon\mathcal{R}(\tilde{d})$ , which compares how different are the energies of the reference and distorted CIWaM perceptual images, namely  $\varepsilon_\rho$  and  $\hat{\varepsilon}_\rho$  respectively, when these images are watched from a given distance  $\tilde{d}$ .

$$\varepsilon\mathcal{R}(\tilde{d}) = 10 \cdot \left| \log_{10} \frac{\varepsilon_\rho(\tilde{d})}{\hat{\varepsilon}_\rho(\tilde{d})} \right| \quad (11)$$

Thus, the main goal of this step is to find  $\varepsilon\mathcal{R}(D)$ , namely, at  $D$   $\varepsilon_\rho$  is equal to  $\hat{\varepsilon}_\rho$ , where the energy of the distorted images are the same than the energy of the pattern.

**Step 3: Perceptual Images** Getting the perceptual images  $\{f_p(i, j)\}$  and  $\{\hat{f}_p(i, j)\}$  from the  $\{f(i, j)\}$  and  $\{\hat{f}(i, j)\}$  images watched at  $D$  centimeters, using Equation 4.

**Step 4: Inverse Wavelet Transformation** Perform the Inverse Wavelet Transform of  $\{\omega_{s,o;\rho,D}\}$  and  $\{\hat{\omega}_{s,o;\rho,D}\}$ , obtaining the perceptual images  $f_{\rho(i,j),D}$  and  $\hat{f}_{\rho(i,j),D}$ , respectively. The synthesis filter in Table 2 is an inverse Daubechies 9-tap/7-tap filter.

**Table 2.** 9/7 Synthesis Filter.

Synthesis Filter		
i	Low-Pass Filter $h_L(i)$	High-Pass Filter $h_H(i)$
0	1.115087052456994	0.6029490182363579
$\pm 1$	0.5912717631142470	-0.2668641184428723
$\pm 2$	-0.05754352622849957	-0.07822326652898785
$\pm 3$	-0.09127176311424948	0.01686411844287495
$\pm 4$		0.02674875741080976

**Step 5: PSNR between perceptual images** Calculate the PSNR between perceptual images  $f_{\rho(i,j),D}$  and  $\hat{f}_{\rho(i,j),D}$  using Eq. 2 in order to obtain the CIWaM weighted PSNR i.e. the  $\mathcal{B}$ PSNR.

## 4 Experimental Results

It is important to mention that  $\mathcal{B}$ PSNR estimates the degradation, thus, the smaller the better. In this section,  $\mathcal{B}$ PSNR performance is assessed by comparing the statistical significance of the images *Lenna* and *Baboon*, in addition to the Pearson correlation between  $\mathcal{B}$ PSNR and PSNR data.

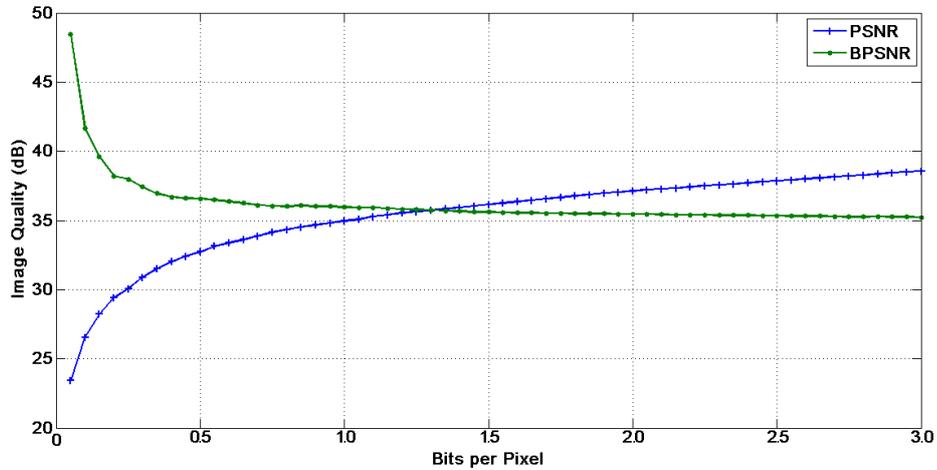
Figure 8 depicts three JPEG2000 distorted versions of the image *Lenna* with 0.05(8(a)), 0.50 (8(b)) and 1.00 (8(c)) bits per pixel. PSNR estimates 23.41, 32.74 and 34.96 dB, respectively. While  $\mathcal{B}$ PSNR computes 48.42, 36.56 and 35.95 dB, respectively. Thus, both PSNR and  $\mathcal{B}$ PSNR estimate that image at 1.00 bpp has lower distortion.

When this experiment is extended computing the JPEG2000 distorted versions from 0.05 bpp to 3.00bpp (increments of 0.05 bpp, depicted at Figure 9), we found that the correlation between PSNR and  $\mathcal{B}$ PSNR is 99.32 %, namely for image *Lenna* for every 10,000 estimation  $\mathcal{B}$ PSNR misses only in 68 assessments.

Figure 10 depicts three JPEG2000 distorted versions of the image *Baboon* with 0.05(10(a)), 0.50 (10(b)) and 1.00 (10(c)) bits per pixel. PSNR estimates 18.55, 23.05 and 25.11 dB, respectively. While  $\mathcal{B}$ PSNR computes 43.49, 30.07 and 28.71 dB, respectively. Thus, both PSNR and  $\mathcal{B}$ PSNR estimate that image at 0.05 bpp has higher distortion.



**Fig. 8.** JPEG2000 Distorted versions of color image *Lenna* at different bit rates expressed in bits per pixel (bpp). (a) High Distortion, (b) Medium Distortion and (c) Low Distortion.



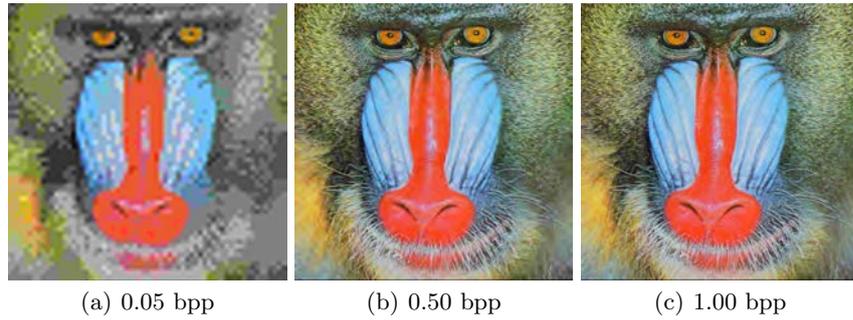
**Fig. 9.** Comparison of PSNR and  $\mathcal{B}$ PSNR for the JPEG2000 distorted versions of image *Lenna*.

When this experiment is extended computing the JPEG2000 distorted versions from 0.05 bpp to 3.00bpp (increments of 0.05 bpp, depicted at Figure 11), we found that the correlation between PSNR and  $\mathcal{B}$ PSNR is 96.95 %, namely for image *Baboon* for every 10,000 estimation  $\mathcal{B}$ PSNR misses only in 305 assessments.

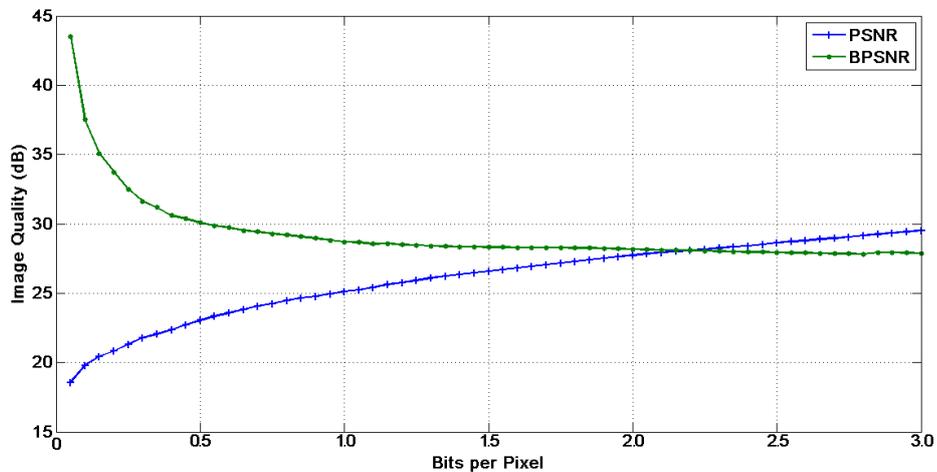
## 5 Conclusions

$\mathcal{B}$ PSNR is a new metric for no-reference or blind image quality based on perceptual weighting of PSNR by using a perceptual low-level model of the Human Visual System (CIWaM model). The proposed  $\mathcal{B}$ PSNR metrics is based on five steps.

The  $\mathcal{B}$ PSNR assessment was tested in two well-known images, such as *Lenna* and *Baboon*. It is a well-correlated image quality method in these images for



**Fig. 10.** JPEG2000 Distorted versions of color image *Baboon* at different bit rates expressed in bits per pixel (bpp). (a) High Distortion, (b) Medium Distortion and (c) Low Distortion.



**Fig. 11.** Comparison of PSNR and  $\mathcal{B}$ PSNR for the JPEG2000 distorted versions of image *Baboon*.

JPEG2000 distortions when compared to PSNR. Concretely,  $\mathcal{B}$ PSNR correlates with PSNR, on the average in 98.13%.

## Acknowledgment

This work is supported by The National Polytechnic Institute of Mexico by means of a granted fund by the Committee of Operation and Promotion of Academic Activities (COFAA).

## References

- [1] Wang, Z., Bovik, A.: Mean squared error: Love it or leave it? a new look at signal fidelity measures. *Signal Processing Magazine, IEEE* **26**(1) (jan. 2009) 98 –117

- [2] Wang, Z., Bovik, A.C.: Modern Image Quality Assessment. 1 edn. Morgan & Claypool Publishers: Synthesis Lectures on Image, Video, & Multimedia Processing (February 2006)
- [3] Auli-Llinas, F., Serra-Sagrasta, J.: Low complexity JPEG2000 rate control through reverse subband scanning order and coding passes concatenation. *IEEE Signal Processing Letters* **14**(4) (april 2007) 251 –254
- [4] Taubman, D.S., Marcellin, M.W.: *JPEG2000: Image Compression Fundamentals, Standards and Practice*. ISBN: 0-7923-7519-X. Kluwer Academic Publishers (2002)
- [5] Bartrina-Rapesta, J., Auli-Llinas, F., Serra-Sagrasta, J., Monteagudo-Pereira, J.: JPEG2000 Arbitrary ROI coding through rate-distortion optimization techniques. In: *Data Compression Conference*. (25-27 2008) 292 –301
- [6] Otazu, X., Párraga, C., Vanrell, M.: Toward a unified chromatic induction model. *Journal of Vision* **10**(12)(6) (2010)
- [7] Murray, N., Vanrell, M., Otazu, X., Parraga, A.: Saliency estimation using a non-parametric low-level vision model. In: *Proceedings of IEEE Conference on Computer Vision and Pattern Recognition (CVPR'2011)*. (2010) 433 –440
- [8] Mullen, K.T.: The contrast sensitivity of human colour vision to red-green and blue-yellow chromatic gratings. *The Journal of Physiology* **359** (February 1985) 381–400
- [9] van de Wouwer, G., Scheunders, P., van Dyck, D.: Statistical texture characterization from discrete wavelet representations. *IEEE Transactions on Image Processing* **8**(4) (April 1999) 592 –598
- [10] Wilson, B.A., Bayoumi, M.A.: A computational kernel for fast and efficient compressed-domain calculations of wavelet subband energies. *IEEE Transactions on Circuits and Systems II: Analog and Digital Signal Processing* **50**(7) (July 2003) 389 – 392



# Detection and Denoising of Impulsive noise using Vector-Marginal filter applied to color images

Agustín Tortolero,<sup>1</sup> Alberto Rosales<sup>1</sup>, Francisco Gallegos<sup>1</sup>, Jean Marie Kinani<sup>1</sup>

ESIME Zacatenco, Instituto Politécnico Nacional, México D.F., C.P. 07738.

agustin\_tortolero@hotmail.com, arosales23@ipn.mx

**Abstract:** we present a vector-marginal filter for impulsive noise detection based on: 1) Fuzzy rank-ordered differences static (FROD) working on the color components and 2) Fuzzy detection phase on the separate components of the image using the City-block ( $L_1$ ) distance. Also we present a denoising scheme for color images using angular criterion and the resulting detected noise.

**Keywords.** Color image processing, Vector filter, Marginal filter, Fuzzy Logic, Impulsive Noise.

## 1 Introduction

Digital image enhancement and analysis have played an important role in scientific, industrial, and military applications. In addition to these applications, image enhancement and analysis are increasingly being used in consumer electronics. Internet Web users, for instance, not only rely on built-in image processing protocols such as JPEG and interpolation, but they also have become image processing users equipped with powerful yet inexpensive software such as Photoshop. Image enhancement refers to processes seeking to improve the visual appearance of an image. Image enhancement is important because of its usefulness in virtually all image processing applications [1]. Color image processing techniques are commonly divided into two classes [2]: the first one Marginal (component-wise) methods that operate on each color channel separately, due to the independent processing ignores the correlation that exists between the color channels giving in some cases, the projection of the separate outputs into the color image usually results in perceivable color artifacts; and the second one Vector methods, that process the input samples as a set of vectors. Because no new color are introduced into the image, this kind of filtering is much more effective and adequate in color image processing applications as suggested in [2].

The design of the proposed filter is focused in the reduction of the impulsive noise in color images. This type of noise is generally caused by the acquisition or transmission of digital images through sensors or communication channels [3,4]. Image transmission noise is caused by various sources, among others; there are manmade

phenomena, such as car ignition systems, industrial machines in the vicinity of the receiver, switching transients in power lines, and various unprotected switches. In addition, natural causes, such as lightning in the atmosphere and ice cracking in the Antarctic region, can also affect the transmission process. This kind of noise can be modeled as impulsive noise. Two most common impulsive noises are fixed-value impulsive noise and random-value impulsive noise [5, 3, 6]. For example, if  $F_i^{col}$  denotes the input noisy image and  $O_i$  the original noise-free image in  $i$ -th pixel position, then we can express the impulsive noise as in Eq.(1)[7]:

$$F_i^{col} = \begin{cases} O_i^{col}, & \text{with probability } 1 - \delta \\ \zeta_i^{col}, & \text{with probability } \delta \end{cases} \quad (1)$$

The  $x$  and  $y$  indexes indicate 2-D pixel position and  $col$  the color component, respectively, with  $col = \{R, G, B\}$ . In this paper we consider two types of impulsive noise, random-value and fixed-value. The characteristic of the Random-value type is that  $\zeta_i^{col}$  is an identically distributed, independent random process with an arbitrary underlying probability density function. For fixed-value,  $\zeta_i^{col}$  has value of 0 or 255. .

This paper is organized as follows. In Section 2, the vector-marginal impulsive noise detector and the proposed filter are detailed. In Section 3, we present experimental results for the random and fixed valued impulsive noise. Finally, conclusions are presented in Section 4.

## 2 Proposed Vector-Marginal Impulsive Noise Detector and Filter (FDDEMV)

The proposed filter is divided in two, the noise detection which in turn is divided in two stages, namely, *marginal* and *vectorial* noise detection under which a fuzzy rule is implemented, and a fuzzy filter which uses the noise level results obtained from the detection phase above along with the angle computed between each pixel and its neighbors inside the sliding window. Detailed methodology is described in next section.

### 2.1 Vector-Marginal Impulsive Noise Detector

The detector is divided in two stages, *vectorial* and *marginal*. The *vectorial* stage is based on the Fuzzy rank-ordered differences static (FROD) [10]. The design of the *marginal* stage is based on the  $L_1$ (City-Block) distance followed by a fuzzy stage that uses fuzzy similarity between the central pixel and its neighbors.

### The Vectorial stage

The fuzzy rank-ordered differences static (FROD) uses the fuzzy version of the Chess-board metric ( $L_\infty$ ), the  $M_\infty$  fuzzy metric. This fuzzy metric is defined as:  $X = \{0,1, \dots, 255\}$ , and  $K > 0$  which denote an 8-bit pixel according to equation (2). Let define  $\mathbf{x}_i \in X^3$ . In this way let define the  $M_\infty$  metric on  $X^3$  as in Eq. (2)

$$M_\infty(\mathbf{x}_a, \mathbf{x}_b) = \min_{l=1}^3 \frac{\min\{x_a(l), x_b(l)\} + K}{\max\{x_a(l), x_b(l)\} + K}, \quad (2)$$

we use the value parameter of  $K = 1024$  according to [10].  $x_a$  and  $x_b$  are the pixels under consideration.

Consider a  $n \times n$  window denoted as  $\Omega_x$  over an image pixel  $\mathbf{x}$  where each pixel  $\mathbf{x}_j \in \Omega_x$  is associated to its RGB color components  $(x_R, x_G, x_B) \in X^3$ . Let denote by  $\Omega_x^0$  the set of neighbors of  $\mathbf{x}$  in  $\Omega_x$  that is  $\Omega_x^0 = \Omega_x - \{\mathbf{x}\}$ . Considering the usage of the  $M_\infty$  fuzzy metric to obtain the fuzzy distances  $d_{x, x_i} = M_\infty(\mathbf{x}, \mathbf{x}_i)$ ,  $\mathbf{x}_i \in \Omega_x^0$ , and we order the fuzzy distances in descending sequence so that  $s_j(\mathbf{x})$  is the  $j$ -th largest  $d_{x, x_j}$  value ( $s_1(\mathbf{x}) \geq s_2(\mathbf{x}) \geq \dots \geq s_{n^2-1}(\mathbf{x})$ ), the FROD static is given by Eq. (3) as:

$$FROD(\mathbf{x}) = \prod_{j=1}^m s_j(\mathbf{x}). \quad (3)$$

With the  $m$ -th nearest neighbors the FROD static provides the degree in which the vector  $\mathbf{x}$  is similar to its  $m$  most similar neighbors in  $\Omega_x^0$ , we selected  $m = 3$ . An impulsive noise pixel is not expected to be similar to its neighbors respect to its fuzzy metric, which implies a low value of  $FROD(\mathbf{x})$ , whereas noise-free pixels, even those close to image edges, are expected to have a  $FROD(\mathbf{x})$  value close to the unity.

### The Marginal stage

This algorithm is applied for each color channel in an independent way: measure the  $L_1$  distance given in the equation (4) between the central pixel and its neighbors, where  $\mu$  stands for each RGB channel,  $j = 0, \dots, n^2 - 1$ , with  $n = 3$ ,  $P_c^\mu$  and  $P_j^\mu$  are the central pixel and a neighbor pixel respectively:

$$\Delta_j^\mu = \text{dist}^\mu(P_c^\mu, P_j^\mu) = |P_c^\mu - P_j^\mu|. \quad (4)$$

Obtain the degree of membership of each  $\Delta_j^\mu$  within the Large fuzzy set using Eq. (5), obtaining in this manner  $r^\mu(\Delta_j^\mu)$ , which indicates the level of similarity between the central pixel and its neighbors, if the pixels are similar, Large fuzzy set will be close to the unity, otherwise it should be close to zero.

$$r^\mu(\Delta_j^\mu) = \begin{cases} 1, & \text{if } \Delta_j^\mu \leq \alpha \\ 1 - 2 \left( \frac{\Delta_j^\mu - \gamma}{\gamma - \alpha} \right)^2, & \text{if } \alpha < \Delta_j^\mu < \frac{\alpha + \gamma}{2} \\ 2 \left( \frac{\Delta_j^\mu - \gamma}{\gamma - \alpha} \right)^2, & \text{if } \frac{\alpha + \gamma}{2} < \Delta_j^\mu \leq \gamma \\ 0, & \text{if } \Delta_j^\mu > \gamma \end{cases}, \quad (5)$$

with  $\alpha = 20$  and  $\gamma = 100$  are parameters of the membership function obtained under experimentation, that indicate the level in which  $r^\mu(\Delta_j^\mu)$  has a maximum value ( $\approx 1$ ) or minimum value ( $\approx 0$ ).

Next, we arrange these degrees of membership in a descending order then name this sequence  $Z_j^\mu$ , doing this, we can obtain a fuzzy similarity function Eq. (6) that indicates the level of similarity between the central pixel and its most similar neighbors.

$$NF_\mu = \prod_{j=1}^B Z_j^\mu, \quad (6)$$

where  $B$  is the number of pixels taken into account in the product to obtain  $NF_\mu$ .

#### Respect to Fuzzy Rule

After having obtained the detected noise by means of vectorial and marginal processing, the next step is to combine these two stages of the detection scheme to obtain a noise level for the vectorial stage. This is done by using the *fuzzy rule 1* (Eq. (7)):

**fuzzy rule 1:**  
*if  $NF_R$  is large and  $NF_G$  is large and  $NF_B$  is large and  $FROD(x)$  is large then  $TNF$  is large* (7)

where  $TNF$  is the Total Noise-Free detected value produced by combination of the vectorial and marginal stages and represents the level of noise from the pixel under processing.  $NF_R$ ,  $NF_G$  and  $NF_B$  are the Noise-Free values resulting of the marginal processing on the Red, Blue and Green channels.

## 2.2 Proposed filtering scheme

Let introduce a filtering switching scheme that uses the detected noise. The Fig. 1 shows a block diagram of the algorithm, applied to each pixel in an image. The first is a conditional block, if  $TNF$  value is less than a threshold, it is performed a filtering over the pixel been processed, otherwise, no operation is done and the pixel under processing remains intact.

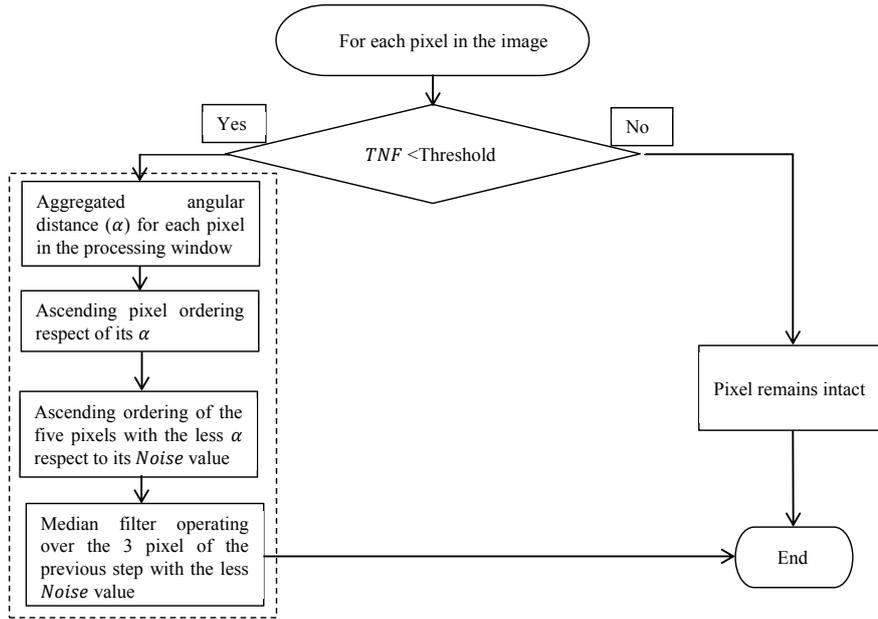
In the first block of the filtering, it is obtained an aggregated angular distance  $\alpha_i$  [8] given in Eq. (8), this parameter indicates how much a given pixel  $i$  present in the

sliding window is similar to all neighbors. This is done by calculating the angles between each of the pixels agree to Eq. (9).

$$\alpha_i = \sum_{j=1}^9 \theta(\text{Pixel}_i, \text{Pixel}_j), \quad (8)$$

$$\theta(\text{Pixel}_i, \text{Pixel}_j) = \cos^{-1} \left( \frac{R_i \cdot R_j + G_i \cdot G_j + B_i \cdot B_j}{\sqrt{R_i^2 + G_i^2 + B_i^2} \sqrt{R_j^2 + G_j^2 + B_j^2}} \right). \quad (9)$$

In the next block, an ascending ordering is done with respect to the aggregated angular distance of each pixel. After, ascending ordering of the five pixels with the less  $\alpha$  respect to its  $NF$  value, another ascending ordering is done but respect to the *Noise* value of the five pixels with the less aggregated angular distance. Finally, in the last block, median filter operating over the 3 pixel of the previous step with the less *Noise* value, a median filter is done over the three pixels with the less  $NF$  value. By experiments, we saw that taking into account five pixels for the ascending ordering respect to  $\alpha$  and three pixels for the median filter, we get the best results.



**Fig. 1.** Block diagram of the proposed filtering scheme.

### 3 Experimental Results

In this Section, we present the experimental results for the proposed filter. We compare several filters found in literature (Table 1) against the proposed filter using objective parameters. The parameters are: Peak Signal to Noise Ratio (PSNR) [1] Eq. (10) indicating noise suppression level, Mean Absolute Error (MAE) [1] Eq. (12), to measure fine details and edges preservation, and Mean Chromaticity Error (MCRE) [12] Eq. (13), characterizing the chromaticity error between the original and the filtered image.

$$PSNR = 10 \cdot \log \left[ \frac{255^2}{MSE} \right], \quad (10)$$

$$MSE = \frac{1}{3MN} \sum_{x=0}^{M-1} \sum_{y=0}^{N-1} [(|R(x, y) - R'(x, y)|^2 + |G(x, y) - G'(x, y)|^2 + |B(x, y) - B'(x, y)|^2)], \quad (11)$$

$$MAE = \frac{1}{3MN} \sum_{x=0}^{M-1} \sum_{y=0}^{N-1} [(|R(x, y) - R'(x, y)| + |G(x, y) - G'(x, y)| + |B(x, y) - B'(x, y)|)], \quad (12)$$

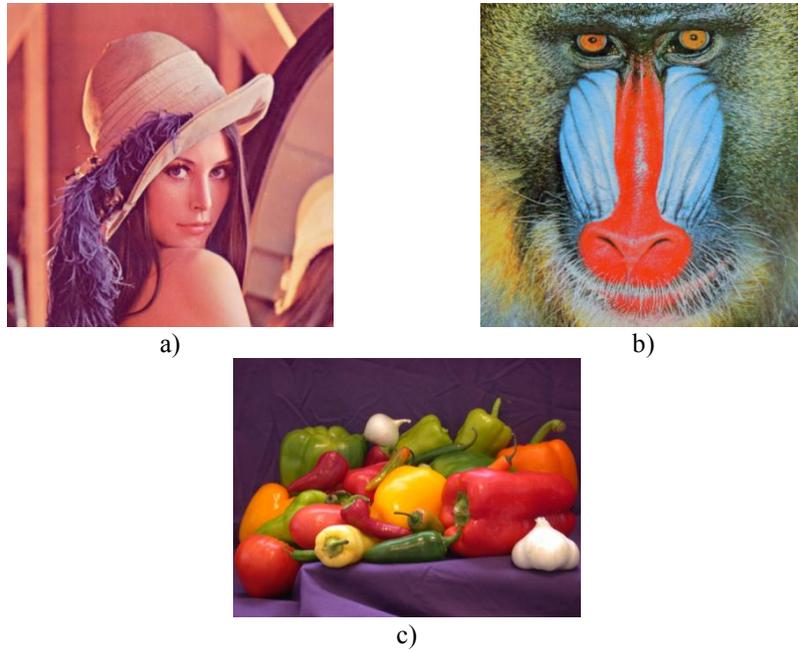
$$MCRE = \frac{\sum_{x=0}^{M-1} \sum_{y=0}^{N-1} C[f(x, y), f'(x, y)]}{MN}, \quad (13)$$

where  $M, N$  are the dimensions of the image,  $R'(x, y), G'(x, y), B'(x, y)$  are the pixel position on each channel of the filtered image and  $C[f(x, y), f'(x, y)]$  is the chromatic error between the vectors  $f(x, y)$  and  $f'(x, y)$ .

**Table 1.** Filters found in the literature

Filter	Reference
$\alpha$ -trimmed	[9]
Basic Vectorial Directional filter (BVDF)	[8]
Fuzzy Two Step Color filter (FTSCF)	[11]
Fuzzy ROD Filter (FRF)	[10]
Vector Median filter VMF	[4]

We have used several standard images to contaminate with different densities and the two types of noise. The standard images that we have used are: Lenna, Mandrill (512x512 pixels) and peppers (512x384 pixels) all them shown in Fig. 2.



**Fig. 2.** Standard test images used: a) Lenna, b) Mandrill and c) Peppers.

In Fig. 3, is displayed the PSNR results for Lenna contaminated with random-value impulsive noise, in Fig. 4 the MAE results for Mandrill contaminated fixed-value impulsive noise and in Fig. 5 are the MCRE results for the peppers image contaminated with fixed-value noise.

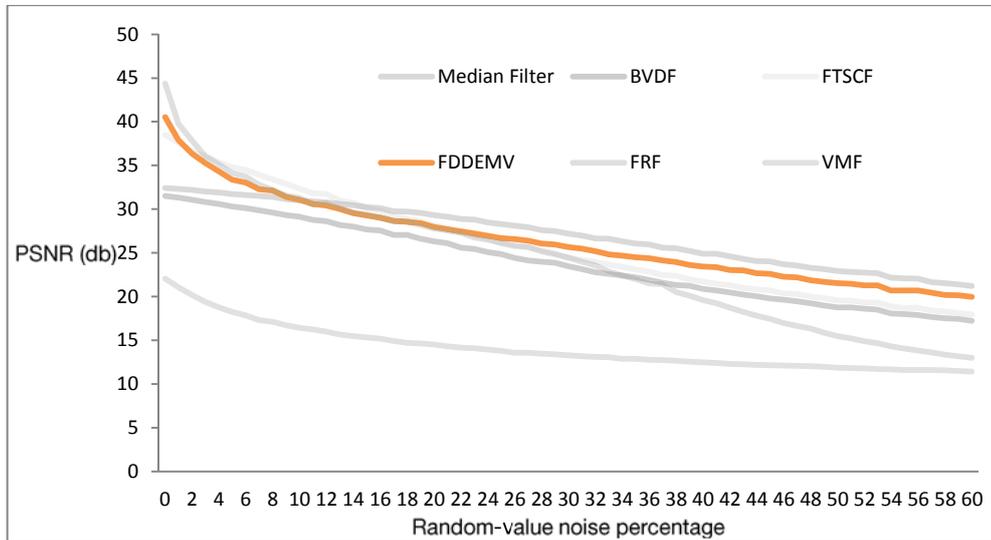


Fig. 3. PSNR results for Lenna contaminated with random-value impulsive noise

In the graphic in Fig. 3, we can note that the filter has a good performance respect to the others filters found in the literature in PSNR terms for low densities of random-value impulsive noise (<25%) and for high densities of noise (from 25 to 60%) is the second best.

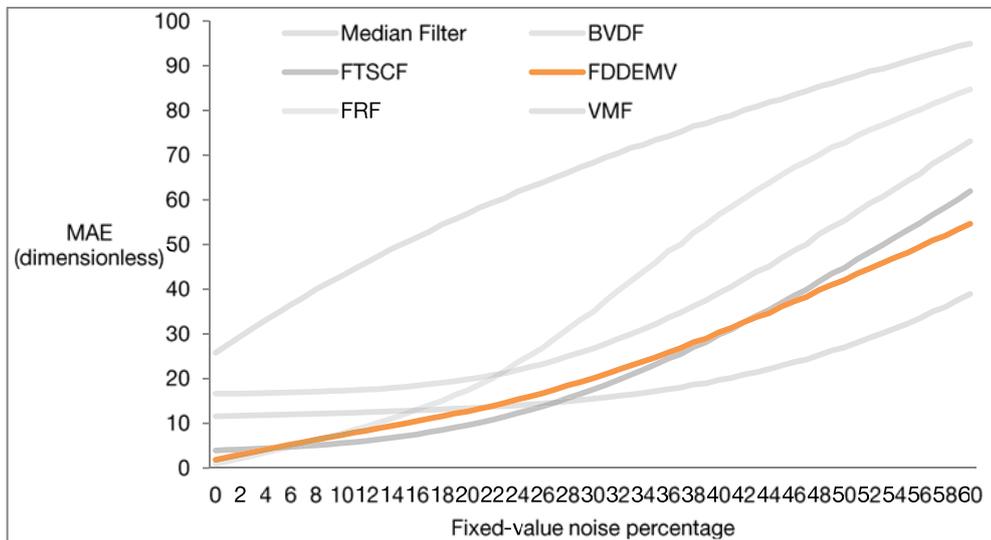


Fig. 4. MAE results for Mandrill image contaminated with “fixed-value” noise.

In the graphic in Fig. 4, we can note that the filter has a good performance respect to the others filters found in the literature in MAE terms for low densities of fixed-value impulsive noise (<10%) and for high densities of noise (from 40 to 60%) is the second best.

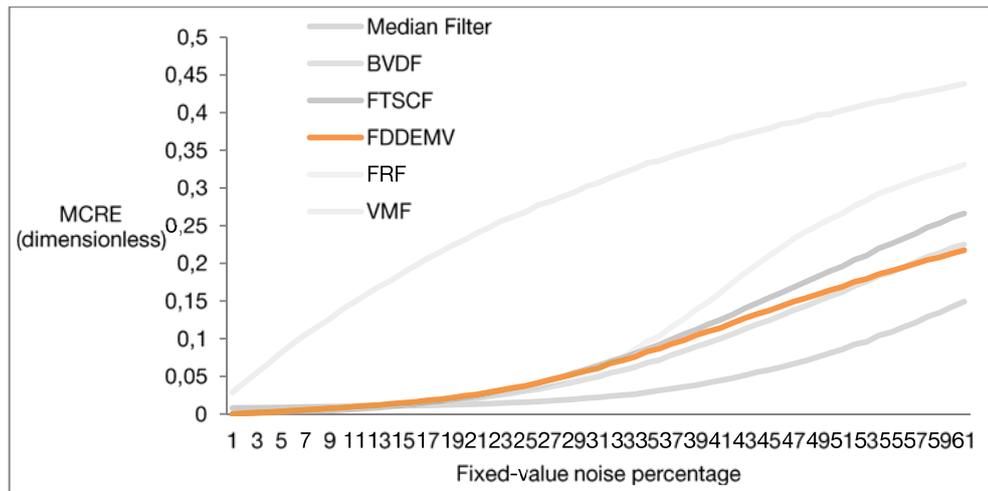


Fig. 5. MCRE results for Peppers image contaminated with "fixed-value" noise.

In the graphic in Fig. 5, we can note that the filter has a good performance respect to the others filters found in the literature in MCRE terms for low densities of fixed-value impulsive noise (<30%) and for high densities of noise (from 30 to 60%) maintains good performance.

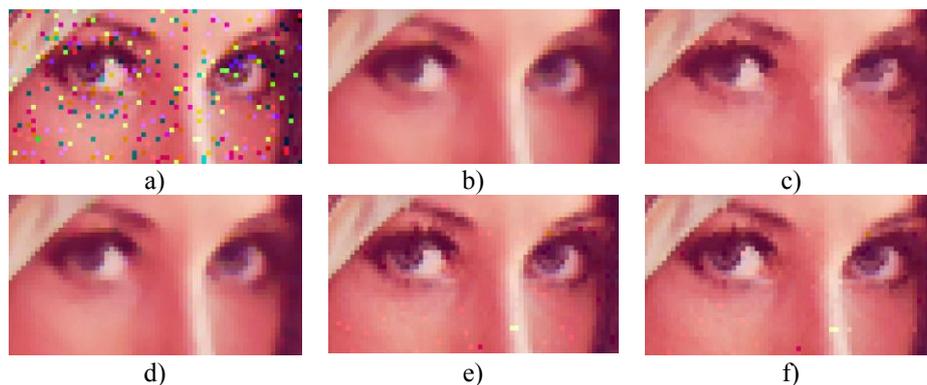
Table 2 shows the results for PSNR, MAE and MCRE of Peppers using random-value impulsive noise.

**Table 2.** results for PSNR, MAE and MCRE of Peppers using random-value impulsive noise

% noise	Parameter	BVDF	VMF	Median Filter	FTSCF	FRF	<b>FDDEM</b>
0	PSNR	35.37	26.46	38.94	47.69	50.54	46.02
	MAE	2.05	5.83	1.30	.03	.13	.03
	MCRE	.007	.029	.008	.000	.000	.000
5	PSNR	34.16	19.93	37.94	39.18	37.20	36.61
	MAE	2.34	9.83	1.46	.35	.54	.57
	MCRE	.0089	.0524	.0091	.0027	.0035	.0037
10	PSNR	32.55	17.58	36.56	34.86	33.95	33.63
	MAE	2.68	13.55	1.67	.73	1.04	1.8
	MCRE	.010	.074	.01	.005	.006	.007
15	PSNR	30.55	16.09	34.82	31.60	32.20	31.53
	MAE	3.14	17.13	1.91	1.20	1.50	1.58
	MCRE	.0127	.0931	.0115	.0090	.0090	.0103
20	PSNR	28.11	15.12	32.68	28.47	30.51	<u>29.65</u>
	MAE	3.87	20.20	2.26	1.90	2.01	2.20
	MCRE	.016	.11	.013	.013	.011	.014
25	PSNR	25.67	14.34	30.40	25.80	28.33	27.78
	MAE	4.86	23.11	2.72	2.84	2.63	2.96
	MCRE	.0213	.1255	.0164	.0199	.0144	.0186
30	PSNR	23.52	13.79	28.31	23.70	25.81	25.89
	MAE	6.22	25.63	3.35	4.04	3.52	3.98
	MCRE	.028	.138	.021	.027	.018	.024

We can see in the table that respect to PSNR, MAE and MCRE, the proposed filter has a competitive performance respect to the others filters found in the literature.

In Fig. 6, is displayed for subjective comparison Lenna corrupted with 5% fixed-value noise and the results from the filters found in the literature and the proposed filter.



**Fig. 6.** a) Lenna corrupted with 5% fixed-value noise and the results of b)VMF, c)BVDF, d) Median Filter, e) FRF and f) proposed FDDEM.

We can see in Fig. 6 that the proposed filter has good noise suppression that the rest of the filters, preserves better the image characteristics, also no color artifacts were introduced.

## 4 Conclusions

From the graphs and the tables, we can see that the proposal filter present good detail preservation, noise suppression and color preservation quality respect the other filters for the two types of noise, random-value and fixed-value at low densities of noise (<20%). With highest noise levels, the proposed filter is competitive respect to the others filters found in the literature.

## Acknowledgements

We would like to thank the National Polytechnic Institute of Mexico for promoting our research spirit and to the accomplishment of this work.

## References.

1. A. Bovik, Handbook of Image & Video Processing, First Edition Academic Press, 2000.
2. R. Lukac, Color image Processing, CRC Press, 2007.
3. K. N. Plataniotis and A. N. Venetsanopoulos, Color Image Processing and Applications. Berlin, Germany: Springer, 2000.
4. R. Lukac, B. Smolka, K. Martin, K. N. Plataniotis, and A. N. Venetsanopoulos, "Vector filtering for color imaging," *IEEE Signal Process. Mag.*, vol. 22, no. 1, pp. 74–86, Jan. 2005.
5. S. Schulte, M. Natchgael, V. De Witte, D. Van der Weken, and E. E. Kerre, "A Fuzzy impulse noise detection and reduction method", *IEEE Trans. Image Process.*, vol.15, no.5, pp. 1153-1162, May 2006.
6. S. Schulte, M. Natchgael, V. De Witte, D. Van der Weken, and E. E. Kerre, "Fuzzy two-step filter for impulse noise reduction from color images", *IEEE Trans. Image Process.*, vol. 15, no.11, pp. 3367-3578, Nov. 2006.
7. S. Schulte, S. Morillas, V. Gregori, E. Kerre «A New Fuzzy Color Correlated Impulse Noise Reduction Method», *IEEE Transactions On Image Processing*, Vol.16 , No.10 , 2565-2575, 2007.
8. Trahanias P.E, «Vector Directional Filters, A New Class of Multichannel Image Processing Filters», *IEEE Transactions On Image Processing*,528-534, vol. 2, 1993.
9. V. Zlokolica , «Robust Non-Linear Filtering for Video Processing», *Digital Signal Processing*, 2002. DSP 2002. 2002 14th International, 571 – 574, vol.2 .
10. J. Camarena, «Two-Step Fuzzy Logic-Based method for impulse noise detection in colour images», *Pattern Recognition Letters*, 1842-1849 vol.31, 2010.
11. V. Ponomaryov, «Fuzzy Directional (FD) Filter to Remove Impulse Noise from Colour Images», 570-572, VOL.E93-A, 2010.



# Efficient Multi-dimensional Data Clustering using Particle Swarm Optimization

Edgar A. Garcia-Martinez, Salvador Godoy-Calderon,  
Ricardo Barron-Fernandez, and Javier Arellano-Verdejo

Laboratorio de Inteligencia Artificial,  
Centro de Investigación en Computación (CIC),  
Instituto Politécnico Nacional (IPN), México<sup>1</sup>  
eagarciam@ipn.mx, sgodoyc@cic.ipn.mx, rbarron@cic.ipn.mx,  
jarellanob10@sagitario.cic.ipn.mx

**Abstract.** Clustering is a very important process that plays a key role in data analysis and data mining that has been effectively used for diverse procedures in artificial intelligence. On the other hand, Particle Swarm Optimization is a family of high-performance meta-heuristic techniques successfully used for solving multi-objective optimization problems. This paper presents a modified particle swarm algorithm for optimizing the localization of cluster centroids. The results yielded by several experiments show this algorithm to be reliable and efficient.

**Keywords:** Particle Swarm Optimization, Pattern Recognition, Clustering

## 1 Introduction

Clustering is a non-supervised technique used for uncovering the underlying structure on a dataset [6]. A precise knowledge about the structure of a particular dataset turns out to be essential for image segmentation [7]-[19], data mining [13] and data compression [1] problems among others. Since there is no supervision information available when clustering data, the only reliable source of information is the similarity or dissimilarity among patterns. For measuring that similarity or dissimilarity a problem-specific user-proposed metric called Pattern Analogy function is generally used. Each cluster is defined by its centroid or by the pattern with the biggest average similarity sometimes called medoid or holotype [16]. Clustering a dataset is a difficult problem since clusters can show a great diversity of sizes and shapes [12]. Besides, the optimum number of clusters for the dataset is not known a priori and automatically finding it is still an open research problem [11], so a great majority of algorithms require it as a parameter supplied by the user.

<sup>1</sup> The authors wish to thank the financial support of COFAA-IPN, SIP-IPN, CIC-IPN, ICyT-DF, and SNI-CONACYT; particularly through grants of projects SIP-20121154, SIP-20120820 and ICyT-PICCO10-113

Particle Swarm Optimization (PSO) is a meta-heuristic algorithm, designed by Kennedy and Eberhart in 1995, modeled following the social behavior of bird flocks and fish schools [14]-[15]. In this algorithm, a group (called swarm) of candidate solutions for an optimization problem (called particles) moves across the problem space, searching for a global optimum solution. Each particle's position is updated considering its fitness (its knowledge about the environment), the position of its neighborhood's best fitted particle (social influence) and the best position that particle itself has visited (cognitive influence). PSO's ability to search huge spaces with very few particles, and the rich set of possibilities it offers for defining inter-neighborhood and inter-particle collaborative search has contributed to consider it as one of the most efficient evolutionary algorithms available today.

Some partitional and density-based clustering algorithms (like K-Means [20] and DBSCAN [10]) strongly rely upon a set of initial problem conditions. When these conditions are not sufficiently satisfied the algorithm may converge to sub-optimal solutions. PSO algorithms, because of their particle updating dynamics, are believed to be much less sensitive to initial problem conditions, so using PSO algorithms for solving clustering problems is a sound strategy with higher probabilities of finding global optimum solutions. This paper shows a modified PSO-based algorithm that efficiently clusters data patterns by optimizing centroid or holotype positions of each cluster.

## 2 Theoretical Background

### 2.1 PSO Algorithm

This algorithm initializes its swarm with randomly generated solutions and a randomly selected speed factor for each particle. Each particle includes a field called *pbest* for storing its best previous value and its position in the search space. Globally known to all particles is the best global value found so far by any particle (*gbest*) and its position. *Pbest* and *gbest* are used for changing a particles' speed after each iteration (or generation) of the swarm search.

PSO is a stochastic and iterative process operating on a particle swarm. Each particle is composed by three vectors and two fitness values as described below:

- Vector  $x_i = \langle x_{i1}, x_{i2}, \dots, x_{in} \rangle$  stores the current particle's position in the search space.
- Vector  $pBest_i = \langle p_{i1}, p_{i2}, \dots, p_{in} \rangle$  stores the position of the particle's best solution.
- Vector  $v_i = \langle v_{i1}, v_{i2}, \dots, v_{in} \rangle$  stores the direction gradient (speed) for regulating the particle's movement.
- Value  $fitness_{x_i}$  is the fitness of the particle's current solution.
- Value  $fitness_{pBest_i}$  is the fitness of the particle's best solution.

The PSO algorithm starts by setting the initial position and speed of each particle in the swarm. The initial position of each particle can be generated

randomly according to a homogeneous distribution or by using some specific initialization heuristic. Once the initial positions are set, the fitness value of each particle is calculated and the  $fitness_{x_i}$  and  $fitness_{pBest_i}$  values are updated.

The speed of each particle is randomly selected with each component within the interval  $[-v_{max}, v_{max}]$ , where  $v_{max}$  will be the maximum speed that each moving particle can adopt. Empirical results have shown that zero-valued speeds almost never promote good final results. Particles should move in an iterative fashion by adding their speed vector  $v_i$  to their position vector  $x_i$ , thus finding the new position vector:  $x_{i+1} \leftarrow x_i + v_i$ .

Once in its new position, the particle's fitness is calculated and the  $fitness_{x_i}$  value is updated. If the current fitness value is better than the fitness of the best found solution, then  $fitness_{x_i}$  and  $fitness_{pBest_i}$  values are also updated.

The speed vector of each particle is updated using its previous speed, the value of a cognitive component, and also the value of a social component. The resulting mathematical model is at the core of the PSO algorithm:

$$v_i^{k+1} = \omega v_i^k + \phi_1 rand_1(pBest_i - x_i^k) + \phi_2 rand_2(gBest - x_i^k) \quad (1)$$

$$x_i^{k+1} = x_i^k + v_i^{k+1} \quad (2)$$

Equation 1 shows the speed-vector upgrade mechanism for particle  $i$  during iteration  $k$ . The *cognitive* component of this movement is set by the term  $\phi_1 rand_1(pBest_i - x_i^k)$  which represents the distance between the current particle's position and the position of the best known solution for the same particle. The social component is set by the term  $\phi_2 rand_2(gBest - x_i^k)$ , the distance between the current particle's solution and the best known solution for any particle in the neighborhood.

## 2.2 PSO-based clustering algorithms

The first PSO-based clustering algorithm was introduced by Omran et al. in [18]. The experimental results shown by Omran et al. [17]-[18] depicted that the PSO-based clustering method outperformed other well known algorithms like k-means, Fuzzy C-Means (FCM) and a few other state-of-the-art clustering algorithms.

Van der Merwe and Engelbrecht hybridized this approach with the k-means algorithm for clustering general datasets [21]. In their approach, a single particle of the swarm is initialized with the result of the k-means algorithm and the rest of the swarm is randomly initialized. In 2003, Xiao et al. used a new approach based on the synergism of the PSO and the Self Organizing Maps (SOM) [22] for clustering gene expression data.

Cui et al. [8] proposed a PSO based hybrid algorithm for classifying the text documents. They applied the PSO, K-means and a hybrid PSO clustering algorithm on four different text document datasets.

### 3 Proposed Algorithm

The proposed clustering algorithm uses a global PSO process to find the optimum position for the  $k$  cluster centroids, using a global variance measure as the fitness function. Each particle is represented by the position of one cluster centroid. For each given centroid position all patterns are assigned to the cluster defined by their nearest centroid, so that the global variance of the resulting clustering can be measured.

---

**Algorithm 1** Pseudocode of the proposed PSO algorithm
 

---

```

Population ← initialize_population();
while not_stop_condition() do
  for  $i \leftarrow 1$  to size(Population) do
    Evaluate each particle  $x_i$  in Population
    if fitness( $x_i$ ) is better than fitness(pBest $i$ ) then
      pBest $i$  ←  $x_i$ 
      fitness(pBest $i$ ) ← fitness( $x_i$ )
    end if
    if fitness(pBest $i$ ) is better than fitness(gBest) then
      gBest ← pBest $i$ 
      fitness(gBest) ← fitness(pBest $i$ )
    end if
  end for
  for  $i \leftarrow 1$  to size(Population) do
     $v_i \leftarrow \omega v_i + \phi_1 \text{rand}_1(\textit{pBest}_i - x_i) + \phi_2 \text{rand}_2(\textit{gBest} - x_i)$ 
     $x_i \leftarrow x_i + v_i$ 
  end for
end while

```

---

#### 3.1 Particle representation

Each particle  $p_i$  is represented as:

$$p_i = \begin{bmatrix} C_1 \\ C_2 \\ \dots \\ C_K \end{bmatrix}$$

Where each component  $C_j = \langle x_1, x_2, \dots, x_D \rangle$  in the particle corresponds to the position for the centroid of cluster  $j$ .  $D$  is the search-space dimension and  $K$  the number of clusters.

### 3.2 Fitness Function

Several cluster validation indexes are available to be used as the fitness function for a PSO-based clustering algorithm. The source code library used for experimentation with the proposed algorithm was carefully parametrized so that any suitable function can be used as the fitness function for each run. Although the choice depends on the topologic properties of the dataset, a few of the most widely used indexes are available, like the *I - index* [2], Davies-Bouldin [9] and Xie-Beni [23]. For the sake of clarity all the experimental results shown in this paper were achieved using a relatively simple global variance expression as fitness. So the fitness function for the experimental results section below is given by:

$$f(x) = \frac{1}{K} \sum_{j=1}^K \sum_{i=1}^N D_E(o_i, C_j) \quad (3)$$

Where  $K$  is the number of clusters,  $N$  the number of patterns in the  $j$  cluster,  $D_E$  is a standard similarity/dissimilarity metric (Euclidian distance for the experiments shown in this paper),  $o_i$  is the  $i$ th pattern in the  $j$  cluster, and  $C_j$  is the centroid for cluster  $j$ .

## 4 Experimental Results

To test the proposed algorithm several experiments were performed using widely known synthetic dataset. The results of proposed PSO algorithm were compared with the VGA-clustering algorithm proposed by Bandyopadhyay and Maulik in 2001 [2]. The fitness function used in VGA-clustering algorithm is computed through the *I - index* described in [2]. The datasets, labeled *Data\_5\_2* [2]-[3]-[5], *Data\_4\_3* [2]-[3]-[5], *Data\_6\_2* [3], *Data\_9\_2* [2]-[3]-[5], *Data\_10\_2* [4]-[5] were taken from <http://www.isical.ac.in/~sanghami/data.html>. Each experiment was performed 100 times to reliably test its performance and accuracy. As is traditional when testing meta-heuristic procedures, configuration parameters for each experiment were empirically set and are shown in tables 1, 2, 3, 4 and 5. The considered performance metrics were execution time and number of iterations. All experiments were run on a AMD Athlon <sup>TM</sup> II X2-220 processor at 2.8 Ghz, with 6 GB Ram. Table 6 shows the average result from the 100 executions of each kind of experiment using the proposed PSO and VGA-clustering algorithms. Lastly, figures 1, 2, 3, 4 and 5 show the best clustering found for each dataset and each algorithm.

Dataset *Data\_5\_2* contains 250 patterns in  $\mathbb{R}^2$  and 5 clusters. The configuration parameters for the PSO procedure used during this experiment were:

**Table 1.** Configuration parameters for experiments with the *Data\_5\_2* dataset

Parameter	Value
Swarm size (in particles)	5
Search-space dimensions	2
Inertia factor	0.9
Cognitive factor weight	1.8
Social factor weight	1.6
Search-space limits	[5.0 16.0]
Maximum speed	1.5

The *Data\_4\_3* dataset contains 400 patterns in  $\mathbb{R}^3$  and 4 clusters. Configuration parameters for experiments with this data set were:

**Table 2.** Configuration parameters for experiments with the *Data\_4\_3* dataset

Parameter	Value
Swarm size (in particles)	20
Search-space dimensions	3
Inertia factor	0.9
Cognitive factor weight	1.8
Social factor weight	1.2
Search-space limits	[-2.0 17.5]
Maximum speed	2.0

The *Data\_6.2* dataset contains 300 patterns in  $\mathbb{R}^2$  and 6 clusters. Configuration parameters for experiments with this data set were:

**Table 3.** Configuration parameters for experiments with the *Data\_6.2* dataset

Parameter	Value
Swarm size (in particles)	20
Search-space dimensions	2
Inertia factor	0.9
Cognitive factor weight	1.2
Social factor weight	1.4
Search-space limits	[-2.0 21.0]
Maximum speed	1.2

The *Data\_10.2* dataset contains 500 patterns in  $\mathbb{R}^2$  and 10 clusters. Configuration parameters for experiments with this data set were:

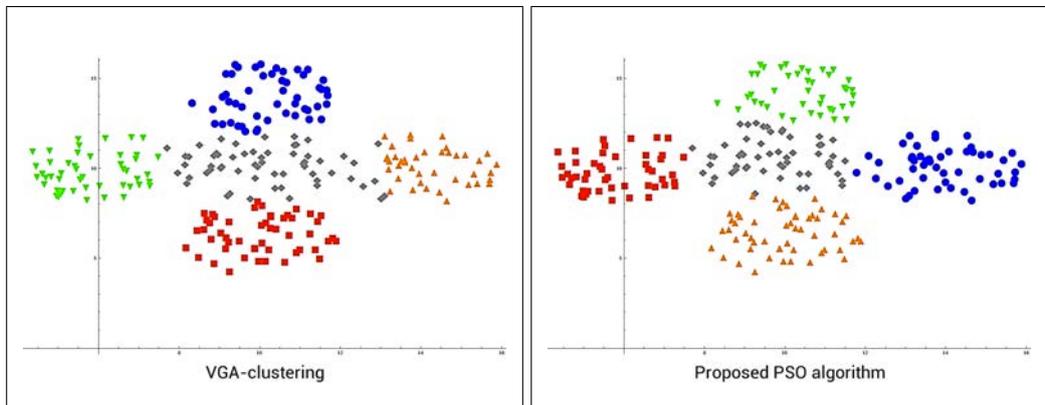
**Table 4.** Configuration parameters for experiments with the *Data\_10.2* dataset

Parameter	Value
Swarm size (in particles)	15
Search-space dimensions	2
Inertia factor	0.9
Cognitive factor weight	1.8
Social factor weight	1.2
Search-space limits	[-18.0 18.0]
Maximum speed	2.0

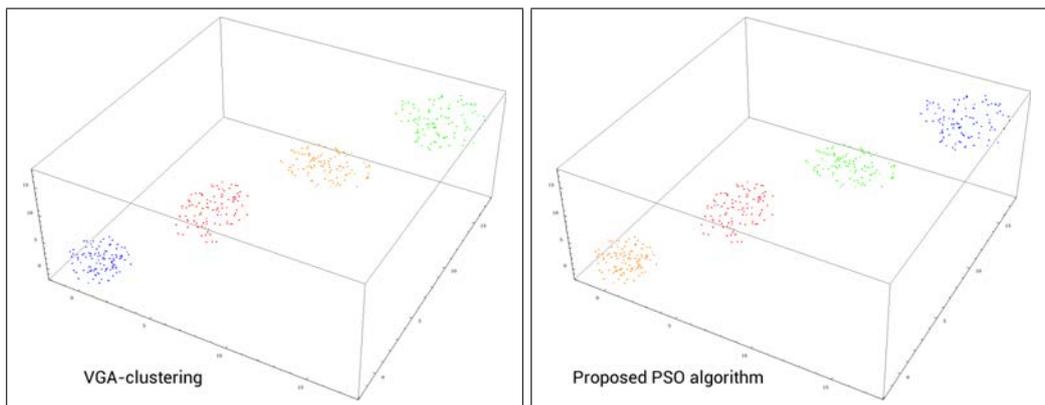
The *Data\_9.2* dataset (sometimes referred to as the *st900.2.9* dataset) contains 900 patterns in  $\mathbb{R}^2$  and 9 clusters. Configuration parameters for experiments with this data set were:

**Table 5.** Configuration parameters for experiments with the *Data\_9\_2* dataset

Parameter	Value
Swarm size (in particles)	25
Search-space dimensions	2
Inertia factor	0.9
Cognitive factor weight	1.8
Social factor weight	1.2
Search-space limits	$[-3.5 \ 3.5]$
Maximum speed	1.5



**Fig. 1.** Best clustering results of each algorithm for all *Data\_5\_2* experiments.



**Fig. 2.** Best clustering results of each algorithm for all *Data\_4\_3* experiments.

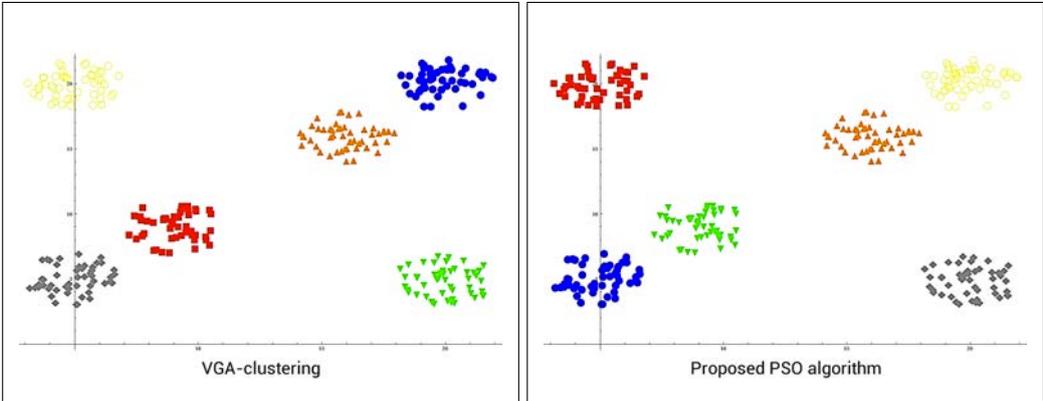


Fig. 3. Best clustering results of each algorithm for all *Data\_6\_2* experiments.

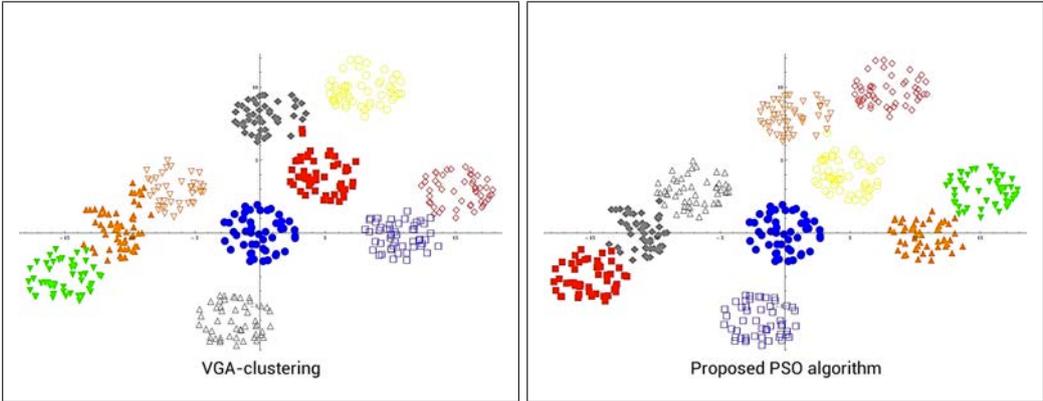
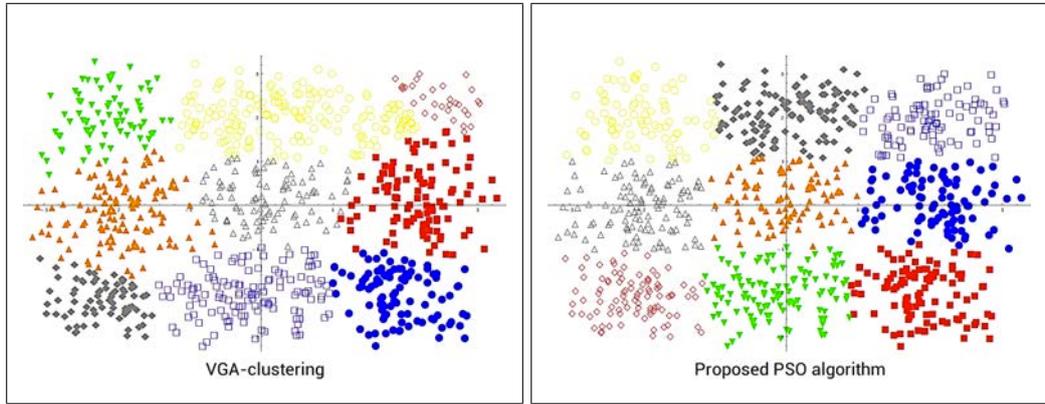


Fig. 4. Best clustering results of each algorithm for all *Data\_10\_2* experiments.



**Fig. 5.** Best clustering results of each algorithm for all *Data\_9\_2* experiments.

**Table 6.** Average execution time and number of iterations of the proposed PSO and VGA-clustering algorithms.

Dataset	Algorithm	Number of patterns	Execution time	Number of iterations
<i>Data_5_2</i>	<i>Proposed PSO</i>	250	2.70 seconds	194
<i>Data_5_2</i>	<i>VGA – clustering</i>	250	4.61 seconds	395
<i>Data_4_3</i>	<i>Proposed PSO</i>	400	12.74 seconds	196
<i>Data_4_3</i>	<i>VGA – clustering</i>	400	21.36 seconds	274
<i>Data_6_2</i>	<i>Proposed PSO</i>	300	55.85 seconds	912
<i>Data_6_2</i>	<i>VGA – clustering</i>	300	68.74 seconds	954
<i>Data_10_2</i>	<i>Proposed PSO</i>	500	52.88 seconds	497
<i>Data_10_2</i>	<i>VGA – clustering</i>	500	66.02 seconds	519
<i>Data_9_2</i>	<i>Proposed PSO</i>	900	54.57 seconds	196
<i>Data_9_2</i>	<i>VGA – clustering</i>	900	101.98 seconds	273

## 5 Conclusions

Particle Swarm Optimization offers some particular advantages over other similar meta-heuristics. It requires only a small number of particles and yet it can explore very large search-spaces, therefore it poses as a good candidate for data-mining and huge data volume problems. Besides, PSO is not vulnerable to poorly set initial conditions which make it ideal for clustering and classification tasks.

This paper presented a modified PSO algorithm which uses a global variance measure for solving multivariate data clustering problems. Five different and widely known datasets were used for testing the proposed algorithm. Each experiment was run 100 times and average results (in table 6) show it to be precise and efficient. In all datasets the proposed PSO algorithm shows better clustering results than the VGA-clustering algorithm.

Experimental results show how a small number of particles (or candidate solutions) can successfully solve any data clustering problem. Also, as shown by those results, even a dramatic increase in the dataset size can only induce a very small increase in the number of particles needed for a specific problem, and the observed variation between the execution time with the smaller and the bigger dataset is non-relevant.

## References

- [1] Abbas, H.M., Fahmy, M.M.: Neural Networks for Maximum Likelihood Clustering, *Signal Processing*, vol. 36(1), 111-126 (1994).
- [2] Bandyopadhyay, S., Maulik, U.: Nonparametric genetic clustering: Comparison validity indices, *IEEE Transactions on Systems, Man and Cybernetics, Part C*, vol. 31, no. 1, 120-125 (2001).
- [3] Bandyopadhyay, S., Maulik, U.: Genetic Clustering for Automatic Evolution of Clusters and Application to Image Classification, *Pattern Recognition*, vol.35, 1197-1208 (2002).
- [4] Bandyopadhyay, S., Murthy, C. A., Pal, S. K.: Pattern Classification Using Genetic Algorithms, *Pattern Recognition Letters*, vol. 16, 801-808 (1995).
- [5] Bandyopadhyay, S., Pal, S. K.: *Classification and Learning Using Genetic Algorithms: Applications in Bioinformatics and Web Intelligence*, Springer, Heidelberg (2007).
- [6] Cai, W., Chen, S., Zhang, D.: A simultaneous learning framework for clustering and classification, *Pattern Recognition* 42 (7), 1248-1259 (2009).
- [7] Coleman, G.B., Andrews, H.C.: Image Segmentation by Clustering, *Proc.IEEE*, vol. 67, 773-785 (1979).
- [8] Cui, X., Potok, T.E.: Document clustering analysis based on hybrid PSO + K-means algorithm. *Journal of Computer Sciences (Special Issue)*, 27-33 (2005) ISSN 1549-3636.
- [9] Davies, D.L., Bouldin, D.W.: A cluster separation measure. *IEEE Transactions on Pattern Analysis and Machine Intelligence* 1, 224-227 (1979).
- [10] Ester, M., Kriegel, H.P., Sander, J., and Xu, X.: A density-based algorithm for discovering clusters in large spatial databases with noise. *Proc. 2nd Int. Conf. on Knowledge Discovery and Data Mining*. Portland, OR, 226-231 (1996).

- [11] Hamerly, G., Elkan, C.: Learning the K in K-means, 7th Annual Conference on Neural Information Processing Systems (2003).
- [12] Jain, A.K., Duin, R., Mao, J.: Statistical Pattern Recognition: A Review, *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 22 (1), 4-37 (2000).
- [13] Judd, D. , Mckinley, P., Jain, A.K.: Large-scale Parallel Data Clustering, *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 20 (8), 871-876 (1998).
- [14] Kennedy, J., Eberhart, R.: Particle Swarm Optimization, *Proceedings of IEEE International Conference on Neural Networks, Perth, Australia*, vol. 4, 1942-1948 (1995).
- [15] Kennedy, J., Eberhart, R.: *Swarm Intelligence*, Morgan Kaufmann (2001).
- [16] Lee, C.-Y., Antonsson, E.K.: Dynamic Partitional Clustering Using Evolution Strategies, In *The Third Asia-Pacific Conference on Simulated Evolution and Learning* (2000).
- [17] Omran, M., Engelbrecht, A.P., Salman, A.: Particle swarm optimization method for image clustering. *International Journal of Pattern Recognition and Artificial Intelligence* 19(3), 297-322 (2005).
- [18] Omran, M., Salman, A., Engelbrecht, A.P.: Image classification using particle swarm optimization. In: *Conference on Simulated Evolution and Learning*, vol. 1, pp. 370-374 (2002).
- [19] Ray, S., Turi, R.H.: Determination of Number of Clusters in K-Means Clustering and Application in Colour Image Segmentation, *Proceedings of the 4th International Conference on Advances in Pattern Recognition and Digital Techniques (ICAPRDT'99)*, Calcutta, India, 137-143 (1999).
- [20] Tou, J.T., Gonzalez, R.C.: *Pattern Recognition Principles*, Addison-Wesley, Reading, MA (1974).
- [21] Van der Merwe, D.W., Engelbrecht, A.P.: Data clustering using particle swarm optimization. In: *Proceedings of the 2003 IEEE Congress on Evolutionary Computation*, Piscataway, NJ, pp. 215-220 (2003).
- [22] Xiao, X., Dow, E.R., Eberhart, R.C., Miled, Z.B., Oppelt, R.J.: Gene clustering using self-organizing maps and particle swarm optimization. In: *Proc. of the 17th International Symposium on Parallel and Distributed Processing (PDPS 2003)*. IEEE Computer Society, Washington (2003).
- [23] Xie, X., Beni, G.: Validity measure for fuzzy clustering. *IEEE Trans. Pattern Anal. Machine Learning* 3, 841-846 (1991)

# Image Recognition System for the Mexican Sign Language

Fausto Pavel Priego Pérez<sup>1</sup>, Jesús Manuel Olivares Ceja<sup>1</sup>,  
José Félix Serrano Talamantes<sup>2</sup>, Diana Naim Rivera Aguilar<sup>3</sup>

<sup>1</sup>Centro de Investigación en Computación del Instituto Politécnico Nacional (CIC-IPN)  
Av. Juan de Dios Bátiz esq. Othon de Mendizabal S/N  
CP 07738, Ciudad de México

<sup>2</sup>Escuela Superior de Computo del Instituto Politécnico Nacional (ESCOM-IPN)  
Av. Juan de Dios Bátiz esq. Othon de Mendizabal S/N  
CP 07738, Ciudad de México

<sup>3</sup>Centro de Innovación y Desarrollo Tecnológico en Cómputo del Instituto Politécnico Nacional  
(CIDETEC-IPN)  
Av. Juan de Dios Bátiz esq. Othon de Mendizabal S/N  
CP 07738, Ciudad de México

pavelpriego@gmail.com, jesuso@acm.org,  
jfserrano@ipn.mx, dr\_euro7@hotmail.com

**Abstract.** It has been found that people with severe hearing loss problems are also mutes in many cases and consequently marginalized in many social activities, including academic activities and other jobs.

This article offers an alternative way to solve the communication problem of people with severe hearing loss. A proposed information system is developed to recognize images of human hands representing signs that belong to the Mexican Sign Language (LSM) which correspond to the ISO 639-2 sgn-MX international standard.

The system contains two main modules: learning and recognition. Recognition is applied constantly to obtained images with a Kinect device by comparing them with patterns in the knowledge base. The learning module is activated when a new pattern is detected (when the similarity between an image and a pattern is less than 90 %).

Kinect is used with a sampling frequency of one image per second. The sampling frequency is based on the time required to process and recognize each image that represent an LSM sign.

The results obtained allow us to propose our system as an alternative for communication among people with severe hearing loss and ordinary people.

**Keywords:** image recognition, learning, evolutionary system, LSM, kinect

## **1 Introduction**

Currently in many places of the world, millions of people with any disability, face marginalization from various human activities, including academic opportunities and some jobs.

In Mexico in 2010 the estimated population of people with any disability was 5 million 739 thousand 270 people (according to the INEGI [1] [2]) representing 5.1% of the total population. Approximately 39% of this population has no access to elementary education because of their situation. Among these people are those with severe hearing loss.

Some disabilities can be overcome using prostheses, Braille code or animal aids among others. We are interested in developing a solution using a computerized system as an alternative way to help mute people produce sounds based on hand sign recognition.

An interesting alternative is that sign languages for severe hearing loss people can be taught from elementary schools as part of the basic culture to ease interaction among mute and ordinary people.

Traditional approaches to image recognition consider a vector of attributes that describe image properties. In [3] [4] recognition based on image pattern is proposed and used in this work.

The rest of the paper is organized as follows. In section 2 the state of art related with aids for deaf and mute people is presented. Section 3 shows elements of the Mexican Sign Language. Section 4 describes the system proposed in this paper. Section 5 presents tests and results. Finally we present conclusions and future work.

## **2 State of art**

In recent decades the computer has become a popular tool to aid severe hearing loss and mute people. We found that developments take into consideration the varieties of sign languages developed: American, Indian, Australian, Spanish, Mexican, Japanese, Chinese, Arabic, among others.

In [5] a system to enable human-computer interaction for the Indian Sign Language is presented. In [6] a system to recognize static signs belonging to the Spanish Sign Language is presented by image processing and obtains a minimum recognition percentage of 55% with the pattern representing the letter H and a maximum of

99.7 % for the one representing B. In [7] a system that transforms text into a sequence of colloquial Russian Sign Language is developed to help deaf people.

In [8] a video game that uses the Kinect as input is presented. The goal is to integrate deaf people with ordinary people by playing together.

In [9] a virtual classroom that operates via the Web to provide advice to people with severe hearing loss by a teacher who explains to the students using the sign language in one window and in another window he interacts using text and images.

In [10] we found an information system for people with hearing problems to alert them in case of severe disasters.

In [11] a module is presented that was added to the portal of the Greek government to keep people informed with expressions using sign language.

In [12] a system that recognizes the static signs of the Mexican Sign Language alphabet is developed.

### **3 Mexican Sign Language (LSM)**

Gestures and body language was probably a form of communication among humans previous to spoken communication forms. [13] and [14] states that in 1620, Juan Pablo de Bonet in Spain wrote the first book for teaching sign language to deaf people. In 1755 Charles Michele de L'Epee in Paris, France, established the first school for deaf people in the world. In 1778 Samuel Heinicke of Leipzig, Germany established the first school for deaf people in Germany. In United States ancient natives had already a system of gesture for intertribal communication rather than for deaf people. Thomas Hopkins Gallaudet in order to help his neighbor Alice Cogswell traveled to France in 1815 to study methods of communication with deaf people and as a consequence in 1817 Gallaudet founded in Hartford, Connecticut the first school for deaf people in USA. Therefore the American sign system is mostly based on the French sign system. Mexican Sign Language (LSM) like the American Sign Language also has its roots in the proposal of Charles Michele de L'Epee. LSM captures specific linguistic differences found in Mexico.

Figure 1 shows an example of the letters of the LSM alphabet; the signs are classified in static and dynamic. A dynamic sign has an arrow indicating the movement direction starting from an initial position.

Mexican Sign Language consists of the alphabet, numbers and words classified in classical categories: verbs, nouns, adjectives, conjunctions, prepositions [15].

In this paper we consider only static symbols of the Mexican Sign Language alphabet although our proposal will evolve taking into account dynamical signs also.



Fig. 1. Examples of Mexican Sign Language [15]

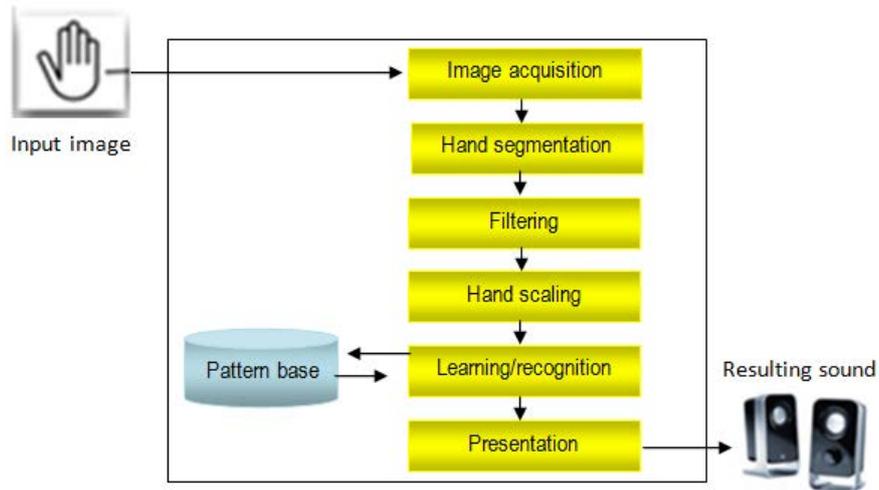
## 4 Proposed system

In this section we present the architecture and description of the main modules of the proposed system to recognize the Mexican Sign Language alphabet.

### 4.1 Architecture

The modules that are part of the system are presented in figure 2. The modules are explained in subsections below:

- a) image acquisition,
- b) hand segmentation,
- c) filtering,
- d) hand scaling,
- e) learning/recognition, and
- f) presentation.



**Fig. 2.** Main modules of the proposed system

#### 4.2 Image acquisition

Images to be recognized are obtained from a Kinect device using a sample frequency of 30 frames per second with a resolution of 640 x 480 pixels each image. Two matrices are obtained with each sample; one is a RGB image and the other is a depth matrix. Depth sampling was calibrated from 40 cm. up to 60 cm.

#### 4.3 Hand segmentation

Depth matrix is obtained with the IR (Infrared) camera included in the Kinect and this is used for image segmentation using 20 cm. interval for sampling and discriminating the image background to isolate the hand. The IR camera provides tolerance to indirect sun light, artificial light and darkness conditions. The result as observed in figure 3 is the isolated work area in which the depth isolated area is mapped on top of the RGB image to help the user to identify hand signs.

Depth segmentation represents an advantage with respect to image transformation from RGB to HSV color code, originally used for skin detection in our first experiments to enable hand image recognition.



Fig. 3. Example of depth segmentation to isolate the hand

#### 4.4 Filtering

As a result of the segmentation process using depth matrix with the IR camera we obtain several Regions Of Interest (ROI). For each image Hu moments  $m00$ ,  $m01$  and  $m10$  are calculated to obtain its centroid and the parallelogram that surrounds each ROI. If no ROI exists in the image no calculation is made. The image result is generated with a binary code with the background set to zero and informative pixels are set to one. Figure 4 shows an image of the B sign obtained with this module. Noisy white points are filtered using the mass obtained with the Hu moments.

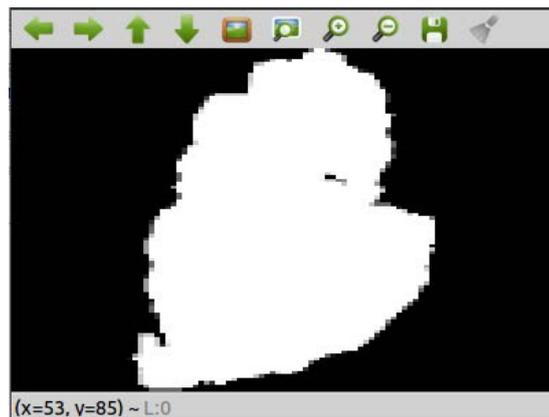


Fig. 4. Example of image segmented using a binary representation

#### 4.5 Hand scaling

Filtered images may result in an enlarged horizontal or vertical shape. In this module the first process is to transform the image to another squared shape by considering the largest vertical (Y) or horizontal (X). The squared shape has  $\max\{X, Y\}$  dimensions and stored in a matrix.

The second process in this module is hand scaling by transforming the squared matrix to other 100 x 100 binary images. Tests were carried out with values of 100 x 100, 200 x 200, 250 x 250, and 300 x 300 and from those 100 x 100 was the best result to find the patterns with enough useful details. Values above 200 x 200 pixels result in an increase of time processing.

#### 4.6 Learning/recognition

During the learning/recognition module, each scaled image is compared with the images contained in a pattern base. Table 1 shows an example of the pattern base formed by tuples with a pattern and the label that describes the pattern. Patterns are stored as vectors of 10,000 cells obtained by placing matrix rows one in front of the other. Figure 5 shows an example of the usual stored sign patterns. Each label describing a pattern is a string.

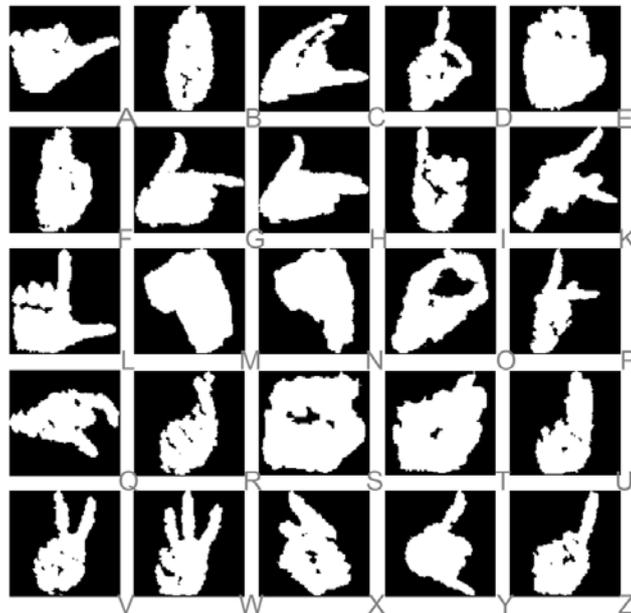


Fig. 5. Usual stored sign patterns example

Table 1 shows an example of the pattern formed by a vector  $y_i, i=0, \dots, 99$  and the label that represent the meaning of the vector.

**Table 1.** Example of the pattern base

$Y_0$	$Y_1$	$Y_2$	$Y_3$	$Y_4$	$\dots$	$Y_{99}$	Label
1	1	0	1	0		0	A
1	1	1	0	0		0	B
1	1	0	0	0		1	C
							.
							.
							.

Image recognition employs a pixel for pixel comparison between the image processed and each of the patterns. Images are formed by pixels  $x_i$ , with  $i = 0, \dots, 99$  while each pattern is formed by  $y_i$ , with  $i = 0, \dots, 99$  and a similarity percentage is calculated with a formula presented in expression (1). Table 2 shows the distance between patterns.

If the similarity percentage is equal or higher than an established threshold, the image is set as a candidate image. The response image is that with the highest value in the candidate images calculated. If no image is above the threshold then a learning process is started by asking the user for the label of the actual pattern that is different from the ones in the pattern base. The user also has the chance to ignore the pattern and indicate that no pattern with a label needs to be stored. The pattern database is stored in the main memory (RAM) to make the recognition process more efficient.

$$S = \frac{1}{100} \sum_{i=0}^{99} compare(x_i, y_i) \tag{1}$$

where:

- S: similarity percentage in range [0, 1]
- $x_i$ : each pixel in the image to be recognized
- $y_i$ : each pixel in the pattern contained in the pattern base
- compare(x, y): function that returns 1 if  $x = y$ , 0 otherwise

4.7 Presentation

The result obtained in the Recognition/Learning module when a hand sign is detected is a string that is used as a filename to produce the corresponding sound.

Table 2. Distance between patterns

	a	b	c	d	e	f	g	h	i	k	l	m	n	o	p	q	r	s	t	u	v	w	x	y	z
a	0	0.35	0.31	0.33	0.35	0.31	0.31	0.28	0.33	0.34	0.32	0.40	0.40	0.32	0.26	0.22	0.39	0.45	0.35	0.43	0.35	0.33	0.33	0.33	0.39
b	0.35	0	0.30	0.18	0.22	<u>0.11</u>	0.28	0.32	<u>0.16</u>	0.30	0.33	0.30	0.30	0.33	0.23	0.29	<u>0.16</u>	0.41	0.34	<u>0.16</u>	<u>0.16</u>	<u>0.16</u>	0.20	0.24	0.21
c	0.31	0.30	0	0.28	0.29	0.30	0.23	0.22	0.31	0.19	0.30	0.41	0.41	0.28	0.29	0.30	0.31	0.41	0.29	0.32	0.32	0.31	0.29	0.24	0.31
d	0.33	0.18	0.28	0	0.27	<u>0.14</u>	0.24	0.27	0.17	0.25	0.26	0.41	0.40	0.32	0.18	0.34	0.17	0.46	0.34	0.18	0.17	<u>0.16</u>	0.22	0.21	0.17
e	0.35	0.22	0.29	0.27	0	0.19	0.26	0.28	0.26	0.31	0.31	0.23	0.26	0.30	0.32	0.36	0.28	0.23	0.22	0.28	0.29	0.29	0.22	0.28	0.32
f	0.31	<u>0.11</u>	0.30	<u>0.14</u>	0.19	0	0.22	0.26	0.12	0.29	0.28	0.32	0.32	0.31	0.19	0.31	0.18	0.40	0.32	0.19	<u>0.16</u>	<u>0.14</u>	<u>0.16</u>	0.24	0.19
g	0.31	0.28	0.23	0.24	0.26	0.22	0	<u>0.11</u>	0.26	0.25	0.24	0.43	0.45	0.25	0.26	0.38	0.27	0.38	0.30	0.32	0.26	0.27	0.26	0.25	0.26
h	0.28	0.32	0.22	0.27	0.28	0.26	<u>0.11</u>	0	0.27	0.25	0.28	0.42	0.44	0.27	0.30	0.36	0.31	0.40	0.31	0.34	0.33	0.32	0.28	0.25	0.28
i	0.33	<u>0.16</u>	0.31	0.17	0.26	0.12	0.26	0.27	0	0.29	0.33	0.36	0.37	0.36	0.21	0.30	0.15	0.44	0.33	0.18	0.17	0.17	0.18	0.25	<u>0.16</u>
k	0.34	0.30	0.19	0.25	0.31	0.29	0.25	0.25	0.29	0	0.32	0.40	0.40	0.25	0.25	0.36	0.29	0.44	0.31	0.30	0.30	0.29	0.31	0.22	0.27
l	0.32	0.33	0.30	0.26	0.31	0.28	0.24	0.28	0.33	0.32	0	0.48	0.51	0.36	0.29	0.40	0.36	0.41	0.32	0.37	0.31	0.28	0.30	0.34	0.33
m	0.40	0.30	0.41	0.41	0.23	0.32	0.43	0.42	0.36	0.40	0.48	0	<u>0.11</u>	0.39	0.43	0.33	0.38	0.30	0.32	0.36	0.39	0.39	0.30	0.32	0.44
n	0.40	0.30	0.41	0.40	0.26	0.32	0.45	0.44	0.37	0.40	0.51	<u>0.11</u>	0	0.38	0.42	0.30	0.38	0.34	0.34	0.36	0.39	0.39	0.31	0.33	0.43
o	0.32	0.33	0.28	0.32	0.30	0.31	0.25	0.27	0.36	0.25	0.36	0.39	0.38	0	0.32	0.38	0.35	0.33	0.28	0.35	0.33	0.32	0.37	0.32	0.32
p	0.26	0.23	0.29	0.18	0.32	0.19	0.26	0.30	0.21	0.25	0.29	0.43	0.42	0.32	0	0.31	0.25	0.52	0.39	0.28	0.21	0.21	0.27	0.29	0.23
q	0.22	0.29	0.30	0.34	0.36	0.31	0.38	0.36	0.30	0.36	0.40	0.33	0.30	0.38	0.31	0	0.34	0.47	0.33	0.35	0.34	0.33	0.30	0.30	0.35
r	0.39	<u>0.16</u>	0.31	0.17	0.28	0.18	0.27	0.31	0.15	0.29	0.36	0.38	0.38	0.35	0.25	0.34	0	0.43	0.33	<u>0.11</u>	<u>0.16</u>	0.21	0.25	0.22	<u>0.14</u>
s	0.45	0.41	0.41	0.46	0.23	0.40	0.38	0.40	0.44	0.44	0.41	0.30	0.34	0.33	0.52	0.47	0.43	0	0.21	0.41	0.44	0.45	0.40	0.46	0.44
t	0.35	0.34	0.29	0.34	0.22	0.32	0.30	0.31	0.33	0.31	0.32	0.32	0.34	0.28	0.39	0.33	0.33	0.21	0	0.31	0.34	0.32	0.30	0.31	0.31
u	0.43	<u>0.16</u>	0.32	0.18	0.28	0.19	0.32	0.34	0.18	0.30	0.37	0.36	0.36	0.35	0.28	0.35	<u>0.11</u>	0.41	0.31	0	<u>0.16</u>	0.20	0.25	0.24	<u>0.13</u>
v	0.35	<u>0.16</u>	0.32	0.17	0.29	<u>0.16</u>	0.26	0.33	0.17	0.30	0.31	0.39	0.39	0.33	0.21	0.34	<u>0.16</u>	0.44	0.34	<u>0.16</u>	0	<u>0.11</u>	0.24	0.26	<u>0.11</u>
w	0.33	<u>0.16</u>	0.31	<u>0.16</u>	0.29	<u>0.14</u>	0.27	0.32	0.17	0.29	0.28	0.39	0.39	0.32	0.21	0.33	0.21	0.45	0.32	0.20	<u>0.11</u>	0	0.25	0.30	<u>0.15</u>
x	0.33	0.20	0.29	0.22	0.22	<u>0.16</u>	0.26	0.28	0.18	0.31	0.30	0.30	0.31	0.37	0.27	0.30	0.25	0.40	0.30	0.25	0.24	0.25	0	0.22	0.26
y	0.30	0.24	0.24	0.21	0.28	0.24	0.25	0.25	0.25	0.22	0.34	0.32	0.33	0.32	0.29	0.30	0.22	0.46	0.31	0.24	0.26	0.30	0.22	0	0.27
z	0.39	0.21	0.31	0.17	0.32	0.19	0.26	0.28	<u>0.16</u>	0.27	0.33	0.44	0.43	0.32	0.23	0.35	<u>0.14</u>	0.44	0.31	<u>0.13</u>	<u>0.11</u>	<u>0.15</u>	0.26	0.27	0

Table 3. Example of the recognition process

Input image	Detected pattern	% Similarity	Response
		none	Learns "A"
		92 % "A"	"A"
		90% "A"	"A"
		60 % "A"	Learn "B"
		68 % "A" 70% "B"	Learn "C"

## **5 Test and results**

In this section we show a table with result examples that can be obtained from the system proposed in this paper.

Table 2 contains a column with a person and the hand isolated from the background, and then a column with the detected pattern; the third column indicates a sign with its highest similarity percentage.

We assume in table 3 that the pattern base is empty. The first row shows that sign “A” is learned because no pattern matches with a similarity value higher than 90%. The threshold was determined after several experiments. When a different person shows the same pattern “A”, a percentage of 92 % is obtained with respect to the already known pattern. Row 3 shows again “A” pattern with a slightly different position and it is also recognized properly. Row 4 shows that “B” sign is compared with known patterns; in this case “A” with a similarity of 60 % and therefore the system determines that learning is required. Learning of “C” sign occurs in row 5.

## **6 Conclusions and future work**

In this paper a computerized system as an aid for severe hearing loss people is presented as part of aids to sign language recognition developments.

The system takes as input, visual information obtained from the Kinect device in IR (infrared) and visible spectra. IR information has advantages because no skin detection is necessary, and is also used as an alternative to isolate the interest region of the image (ROI).

Results promote our efforts to consider this approach as a promising alternative to help deaf and mute people, enabling them to interact with audible expressions. One possible use is to report emergencies using phones. Future work deals with learning dynamic signs.

## **7 Acknowledgements**

Pavel Priego Pérez and Diana Naim Rivera Aguilar, are supported by CONACyT scholarships. Authors thank to Fernando Galindo Soria, Roberto Velázquez Cabrera, Manuel Mera Hernández, Elizabeth López Romero, José Luis Arevalo and special thanks to Pavel’s thesis reviewers and Migdaled López Juárez for her aid during tests.

## 8 References

1. INEGI Las personas con discapacidad en México: una Visión censal, 2011, [http://www.inegi.org.mx/prod\\_serv/contenidos/espanol/bvinegi/productos/censos/poblacion/2000/discapacidad/discapacidad2004.pdf](http://www.inegi.org.mx/prod_serv/contenidos/espanol/bvinegi/productos/censos/poblacion/2000/discapacidad/discapacidad2004.pdf)
2. INEGI Características de las personas con discapacidad auditiva, 2011,
3. Fernando Galindo Soria, Sistemas Evolutivos in *Boletín de Política Informática*, México 1986
4. Fernando Galindo Soria, Sistemas Evolutivos: nuevo paradigma de la informática, in *memorias del congreso TEC-COMP 91*, México 1991
5. Subha Rajam, G. Balakrishnan, Real time Indian Sign Language Recognition System to aid deaf-dumb people, 2011
6. Isaac Garcia, Jaime Gómez García-Bermejo, Eduardo Zalama Casanova, Gesture Recognition for Deaf People Interfacing in *18<sup>th</sup> International Conference on Pattern Recognition (ICPR'06)*, IEEE, 2006
7. Mikhail G. Grif, Yelena A. Demyanenko, Development of computer sign language translation technology for deaf people in *2011 6<sup>th</sup> International Forum on Strategic Technology (IFOST)*, Dept. of Autom. Control Syst. Novosibirsk State Tech. University, Russia, 2011
8. Fakhteh Soltani, Fatemeh Eskandari, Shadan Golestan, Developing a Gesture-Based Game for Deaf/Mute People Using Microsoft Kinect, 2012 Sixth International Conference on Complex, Intelligent, and Software Intensive Systems, IEEE
9. Chowdhuri, Debabrata, Virtual classroom for deaf people, in *2012 IEEE International Conference on Engineering Education: Innovative Practices and Future Trends (AICERA)*, 19-21 July 2012, India
10. Atsushi Ito, Hitomi Murakami, Yu Watanabe, et. al., Information Delivery System for Deaf People at a Larger Disaster in *2010 Fifth International Conference on Broadband and Biomedical Communications (IB2Com)*, 15-17 dec 2010, Tokyo Japan
11. Stavroula-Evita Fotinea, Eleni Efthimiou, Tools for Deaf Accesibility to an eGOV Environment, Lecture Notes on Computer Science 5105, ICCHP, Springer-Verlag Berlin Heidelberg, 2008
12. Laura Jeanine Razo Gil, *Sistema para el reconocimiento del alfabeto dactilológico* (Master degree thesis), Centro de Investigación en Computación del Instituto Politécnico Nacional, México, 2009
13. Guillermo Searle Hernández, Historia de la informática (1), los primeros números de la aritmética antropomórfica a la cibernética antropocentrista, *ASTIC, Boletín 37*, España, marzo 2006
14. History of Sign Language [http://www2.uic.edu/stud\\_orgs/cultures/daa/ASLHistory.html](http://www2.uic.edu/stud_orgs/cultures/daa/ASLHistory.html)
15. Luis Armando López García, Rosa María Rodríguez Cervantes, María Guadalupe Zamora Martínez, Susana San Esteban Sosa, *Mis Manos Que Hablan, Lengua de señas para sordos* 1era. Reimpresión, editorial Trillas, 2010

# Intelligent Tutoring System: An approach to the model of the student using multiple intelligences

Rosa María Rodríguez-Aguilar <sup>1</sup>, Jorge Manuel Alejandro Sánchez de Antuñano <sup>2</sup>,  
Ana Lilia Laureano-Cruces <sup>3</sup>.

<sup>1</sup>Universidad Autónoma del Estado de México, Unidad Académica Profesional Nezahualcóyotl. Departamento de Ingeniería en Sistemas Inteligentes. Avenida Bordo de Xochiaca S/N Colonia Benito Juárez, Ciudad Nezahualcóyotl, Estado de México. México. C.P. 57000.

<sup>1</sup>aguila\_rosa@hotmail.com

<sup>2,3</sup>Universidad Autónoma Metropolitana-Azcapotzalco. Departamento de Investigación y Conocimiento para el Diseño. Av. San Pablo 180, Col. Reynosa Tamaulipas, Del. Azcapotzalco, México, D.F. {clc, jsab}@correo.azc.uam.mx

**Abstract.** An Intelligent Tutoring System (ITS) emulates human tutors to determine what to teach, when to teach and how to teach the students in an autonomous manner. The technique of Artificial Intelligence Reasoning Based in Cases is proposed, to be included into Student Modeling, thus achieving an efficient representation by means of modeling that incorporates the different abilities and intelligences (multiple intelligences proposed by Gardner) that students require for their learning, in our case study, Basic Algebra.

**1 Key Words:** Intelligent tutoring system, student model, case based system, multiple intelligences.

## 1 Introduction

The learning of Mathematics is one of the areas of utmost concern due to the low performance that students have demonstrated.

The main cause of the learning problems that have been identified is that the knowledge that should have been acquired in previous courses was actually never assimilated. [1].

The contents are not taught in depth and are oriented towards memorizing and repetitive exercise solving in order to mechanize the solving procedure, that is to say that previous learning is not used to build new knowledge [5].

The didactics of Mathematics has recently experienced a surge in research and development on teaching [3]. The inclusion of ICTs has acquired relevance through computational learning environments [8].

The development of tutoring systems has had a significant advance in the development of applications such as CAI (Computer Assisted Instruction), ICAI (Intelligent

adfa, p. 1, 2011.

© Springer-Verlag Berlin Heidelberg 2011

Computer Assisted Instruction) and ITS (Intelligent Teaching Systems) [7], [12]. Currently Virtual Learning Environments (VLE) are spaces where the right conditions are created so that the individual may seize new knowledge, new experiences and new elements that may suggest to him/her processes of analysis, reflection and appropriation of the same, according to Ávila [2].

### 1.1 Intelligent Tutoring Systems

The objective of ITSs is to imitate human tutors in their ability to determine—in an autonomous manner— what to teach, when to teach and how to teach in each case.

An ITS uses three types of knowledge: knowledge of the domain (topics of a particular subject), knowledge of strategies and teaching methods (e.g. the personal guidance or *coaching* method) and knowledge of the student (personal preferences and styles of learning). These three types of knowledge provide the system with the capacity to assist the student in his learning process.

Usually, an ITS should identify the strengths and weaknesses of a particular student in order to establish an instructional plan that will be consistent with the results. It should find all the relevant information on the learning process of that particular student (such as learning style) and apply the best teaching method according to his/her individual needs.

In order to support the learning process, the ITS proposes the implementing of Howard Gardner's (multiple intelligences) propositions where it has used his theories with good results in its applied tasks. Gardner considers all human beings to be different in learning capacities, without downplaying any particular one, but rather channeling them towards well defined objectives [9]. Due to the above, it is assumed that if the student counts with the necessary elements for his/her development within the ITS according to his/her specific needs, he/she will assimilate the structural knowledge from the expert by navigating the system in the way he/she considers to be fitting to his/her personal style of learning, the one that has to do with his/her preponderant intelligence.

### 1.2. Case Based Systems

Within the applications of Artificial Intelligence, we have that Case Based Reasoning (CBR) is a problem solving method based on remembering previous similar situations and reutilizing the knowledge and information about those situations [4].

Until now CBR has not been used to model the student within the ITSs, in spite of it being an AI technique that provides a cognitive model for the organization of memory, reasoning and human learning [16]. This approach provides a group of particular characteristics that can be used in the elaboration of ITSs such as [11]:

- Reasoning based on specific episodes. The students' cognitive state cannot be explained by experts such as generalized chains. It is therefore proposed that they be described by means of selected trait experiences.
- Knowledge maintenance does not require of expert intervention. The experience of new models can be evaluated and, if found convenient, it can be incorporated to the system by the very ITS user.

- Re-utilization of previous solutions. In addition to methodology, student models that are frequently repeated may be used in the teaching-learning process.
- Derived solutions based on real cases. This allows the justification of decisions taken facing the user during his/her performance within the system.

The CBR works by trying to imitate human behavior as it solves problems found on the execution of a plan, based on similar cases that have been recorded and given optimal solutions.

## **2 Multiple Intelligences**

Developmental psychologist Howard Gardner [9] propounds the theory of multiple intelligences, where he holds that human intelligence has eight differentiated dimensions, to each of which corresponds a different symbolic system and manner of representation: i.e. musical, bodily-kinesthetic, linguistic, logical-mathematical, spatial, interpersonal, intrapersonal and naturalistic.

Gardner holds that educational practice is basically centered on the mathematical and linguistic intelligences. However, due to the multiple character of human intelligence, the perspective ought to be widened in order to take into account the diverse abilities of the individuals. In this manner, projects that admit alternative ways of symbolic expression should be proposed to the students, in order to create group projects that invite the students to work with the language of the media as well as with those symbolic systems with which they feel the greatest affinity.

### **2.1 Types of Intelligence and their Characteristics**

Our model is based on the following considerations according to Gardner [8] and Rivière [18]:

- 1) We don't all have the same interests and capacities.
- 2) We don't all learn the same way.
- 3) Nowadays no one can learn all there is to learn.

Howard Gardner has identified eight different types of intelligences [8]:

- *Logical-mathematical* intelligence uses logical thinking to understand cause and effect, connections, relations between actions, objects and ideas
- *Linguistic* intelligence is the ability to think in words and use language to express and understand complex meanings.
- *Visual-spatial* intelligence is the ability to think and perceive the world in the form of a three dimensional mental model.
- *Auditory-musical* intelligence is the ability to think in terms of sounds, rhythms and melodies.
- *Bodily-kinesthetic* intelligence is the ability to use body movements as a means of self-expression.
- *Intrapersonal* intelligence is the ability to understand oneself.

- *Interpersonal* intelligence is relating and understanding other persons, harmonizing and recognizing the differences among persons and appreciating their perspectives, while being sensitive to their personal motives and intentions.
- *Naturalistic* intelligence is the understanding of the natural world including plants, animals and the scientific observation of nature.

### 3 Intelligent Tutoring Systems

In order to integrate new technologies and produce a new tool which allows us to adapt to a dynamic environment, which in turn is the cognitive state of the student, we propose to develop an ITS with reactive characteristics [11], [13].

The main components of an ITS are shown in Figure 1, the detailed description of which follows.

The *Student Model* is integrated by the data base, which represents the cognitive state of the student, and its environment, which we intend to change.

The *Expert Module* is the place where the expert knowledge is accumulated.

The *Tutoring Module* is in charge of determining the pedagogical and instructional strategies of the system.

The *Interface* may be considered as a simulation environment in the sense that it is the place where the inputs and outputs of the system take place. Its basic responsibility is the communication between the system and the student, though, since it is the output means of the ITS actions, it also has the didactic responsibility [13], [14] and [23].

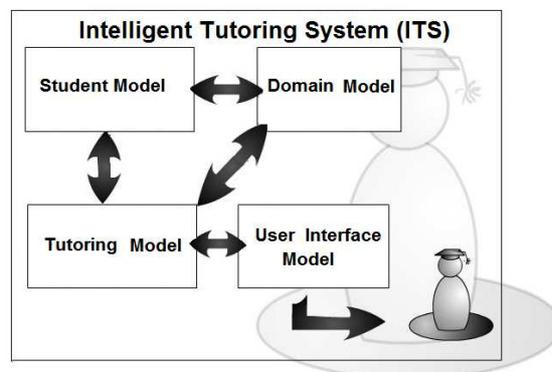


Figure 1. Components of an Intelligent Tutoring System.

ITSs approaches the learning process as co-operation between tutor and student. The tutor, based on the student's perception, decides at every moment which is the right strategy. These strategies are decided based on parameters such as mistakes made, learning style, acquired knowledge and so on.

All of the above, in order to decide what to explain, at which level of detail and when and how to interrupt the student. Research on ITSs has focused on different aspects, from the inspection of each of its constituent modules, to the elaboration of generic architectures as well as automatic learning and the construction of teaching system aide design.

In the case of educational resources, the contribution of ITSs refers to learning techniques that support the problem solving process and, thus are able to diagnose the student's mistakes. In the specific case study we refer to basic algebra. This dominion was determined based on the analysis of the diagnostic tests given to university students, where conflict areas of students from different semesters and studies that integrate the Unit were detected. We must take into account that the student should seize his knowledge and should do so in a step by step manner, and thus begin with the conflicting areas, increasing the complexity and based on his/her cognitive maturity. This is a corrective aspect due to academic history deficiencies. The *diagnosis of mistakes* allows us to know—based on the mistakes the student makes— whether he/she has understood, thus, to design remedial strategies. Learning techniques allow us to implement the most appropriate learning styles as well as instructional strategies and tactics. In this context, ITS are considered virtual environments, where AI techniques can be applied to its different modules in order to simulate the one-on-one teaching-learning process, so that it emulates human behavior during the interactive process [13], [14] and [23].

#### **4 Methodology for the Development of the Intelligent Tutoring System (ITS)**

The methodological aspect of this work started with a *first phase* of qualitative and quantitative study derived from a mathematics diagnostic test given to students of different areas, i.e. International Commerce, Education for Health and Intelligent Systems Engineering. The exam was given to 600 college students of the *Universidad Autónoma del Estado de México – Unidad Académica Profesional Nezahualcóyotl*.

The test included Basic Arithmetic, Algebra, Analytic Geometry and Trigonometry topics. The evaluation showed a series of deficiencies mainly due to a cognitive lack of maturity in the knowledge of basic Algebra. From the above, the ITS will be made up of questions about these topics, starting with Arithmetic, following with Basic Algebra, Analytic Geometry and finally Trigonometry. The order goes from the simple to the more complex.

The relations between the fundamental components of a Case Based System and an Intelligent Learning System are described in the architecture of the final ITS in Figure 2. For this, the definition given by Gutiérrez [11]—which states that all the Case Based System is basically conformed by the Knowledge or Case Base, the Case Recovery Module and the Solution Adaptation Module— is followed. The proposition that the three main components of all ITSs, i.e. contents knowledge, knowledge of the student and knowledge of the learning methodologies and strategies, thus generating the final architecture of the system, is taken from Sánchez.

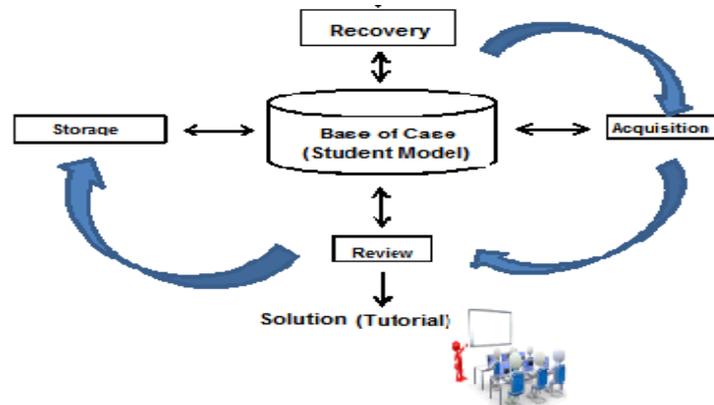


Figure 2. ITS Architecture.

The phases for the development of the ITS are specified in the following steps.

1. Design of the knowledge graph, is conformed by the different Micro-worlds of knowledge or topics that will constitute the ITS, taking as its reference the areas of knowledge on which the diagnostic test was structured (Arithmetic, Elementary Algebra, Analytic Geometry and Basic Trigonometry). See Figure 3.
2. Design of the cognitive model that allows the citing of the user characteristics according to Gardner's proposal (multiple intelligences). In our case study, in order to detect the type of intelligence that allows the adapting of the system to the student, only four of the eight propositions in his theory were considered, to be applied at the time of the study, since they are the most representative i.e. logical-mathematical, auditory-musical, verbal or linguistic, and visual. From here on the system will be denoted by the initials MALVI.
3. The case based teaching/learning model proposed is structured by means of a series of questions or reactivities —each of which is, in turn, pondered by means of a weighted system— in order to determine to what type of intelligence it belongs predominantly, i.e. logical-mathematical, auditory-musical, verbal or linguistic, and visual, where each of these pure types may include aspects of any of the other three types, which in turn may be used within the system. The classifying, according to their relative weight, was done by the expert in Pedagogy. In those instances in which the questions or reactivities were oriented in their majority towards a single type of intelligence, they were adjusted so that that same reactive could be adapted to the remaining types of intelligences in the project. This part of the modeling was backed by an expert in the Mathematics Area.
4. In agreement with the above analysis, different scenarios were designed (presented interfaces) where, according to the different weight that each question presented, it was oriented towards that particular type of intelligence. In some cases changes were made in order to adapt the reactivities to the four types of intelligence to be used in our system, with the support of our mathematician ex-

pert. If a student with a mainly visual intelligence was identified, (s)he was presented with reactivities oriented mainly towards a visual-spatial interface (supported by images and graphs). If a logical-mathematical student was identified, (s)he was presented with the normal scenarios (without adding any other component to the interface). In the case of the mainly linguistically oriented student, scenarios were designed so that the question or reactive was presented to him/her on screen and, additionally in an oral manner, by means of a recording which indicated what the question to be answered consisted in. Lastly, for the auditory-musical scenarios —depending on the degree of complexity of the proposed system, from beginning to advanced— Vivaldi's Four Seasons would be added within the interface. According to the expert in Pedagogy: Spring, the first season, for the first micro-world (Arithmetic, simple degree of complexity); Summer, the second season, for the second micro-world (basic Algebra); for the third micro-world (Analytical Geometry) the third season, Autumn; and for the fourth micro-world (Trigonometry) of a high degree of complexity, the fourth season, Winter, which consists of a soft melody, which is, in turn, considered adequate for the high degree of concentration needed to resolve the questions of this micro-world.

5. Design of the teaching-learning process from two perspectives: 1) the type of intelligence related to the type of knowledge, and 2) the type of intelligence detected in the user.

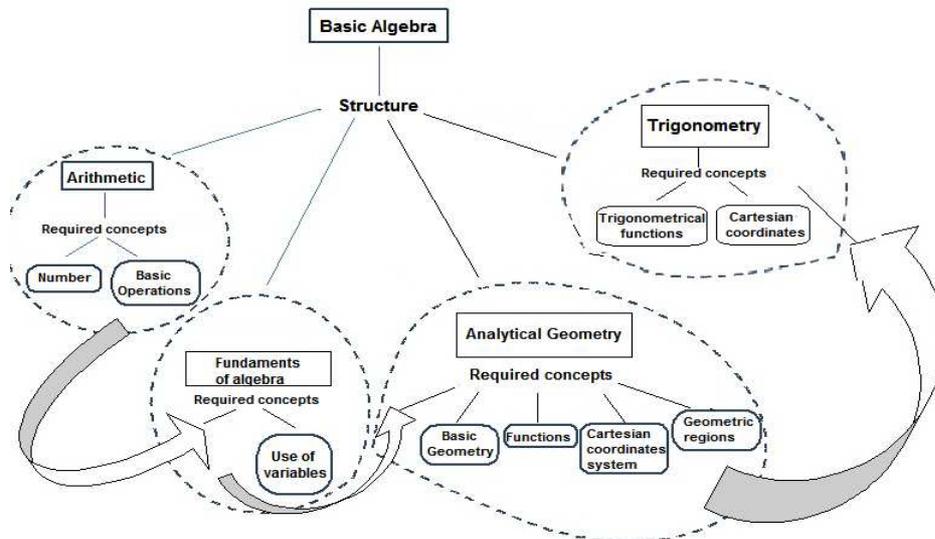


Figure 3. Graph of the Basic Algebra Knowledge ITS.

6. Adapting of the reactive inference motor that would be integrated to the teaching-learning process along the different scenarios.
7. Design of the interface that would integrate the techniques for the diagnosis, and instructional tactics and techniques.

## 5 Conclusions

In this work we propose a new model that may be able to take into account the Intelligent Tutorial Systems, which will be implemented in a Computational System (MALV I), thus intending to prove the previously stated hypothesis.

The proposed model maintains the fundamental components of all the Case Based System, the basis of knowledge or data base, where the cases describe models of students, and the teaching methodologies to be followed in order to achieve its personalization.

The case retrieval module and the solution adaptation module execute the diagnostics, which is —without any doubt— one of the most important processes within any ITS, since its quality will depend on the capacity of the system to adapt to the student.

The system is about to be finished and soon its performance will be tested in the student community of the *Unidad Académica Profesional – Nezahualcóyotl*, so that the results thus obtained may be shared.

## References

1. Andrade A., Gómez G., López M., Contreras-Castillo J.: Las tecnologías en el aprendizaje de las matemáticas universitarias. Año 2 Num. 4. Universidad de Colima. (2007).
2. Avila, P., Bosco M.: Ambientes virtuales de aprendizaje. Una nueva experiencia. 20th. International Council for Open and Distance Education. Düsseldorf, Alemania, 1-5 Abril 2001. (2001).
3. Batanero C., Godino J.: Perspectivas de la educación estadística como área de investigación: Líneas de investigación en Didáctica de las Matemáticas (pp. 203-226). Badajoz: Universidad de Extremadura. España. (2005).
4. Bregón A., Arancha S., Alonso C., Pulido B., Moro I., Rodríguez J.: Un sistema de razonamiento basado en casos para la clasificación de fallos en sistemas dinámicos. III Taller de Minería de Datos y Aprendizaje. Thomson. (2005).
5. Cuevas A., Martínez M.: El ECAEM, un recurso computacional alternativo para promover la comprensión de los conceptos del cálculo diferencial. CINVESTAV-UAEM. México. (2004).
6. De Arriaga, F., El Alami, M., Ugena A.: Acceleration of the Transfer of Novices into Experts: the Problem of Decision Making, Proceedings International Conference BITE'01. pp-245-258, University of Eindhoven. (2001).
7. Fernández R., Server P.: Aprendizaje con nuevas tecnologías paradigma emergente. ¿Nuevas modalidades de aprendizaje? Resolución de problemas matemática y TIC Pág. 1. Eductec. Revista Electrónica de Tecnología Educativa. Núm. 20/ Enero 06. Cuba. (2006).
8. Galaviz M.: Apoyo a la enseñanza de la Matemática utilizando un Ambiente de aprendizaje basado en Instructores Interactivos de Diversiones Matemáticas: Modelo para el Maestro. CICESE UABC. México. (2006).
9. Gardner, M.: La nueva Ciencia de la Mente: Historia de la Psicología Cognitiva. Barcelona, Paidós. (1987).
10. Gardner, H., Hatch, T.J]: Inteligencias Múltiples de ir a la escuela: implicaciones para la Educación de la teoría de las inteligencias múltiples. Investigador Educativo, 18(8), 4-9. (1989).
11. Gutiérrez I.: Modelo para la toma de Decisiones usando Razonamiento Basado en Casos en condiciones de incertidumbre. Villa Clara, Cuba, Departamento de Ciencias de la Computación. Universidad Central de Las Villas. (2003)

12. Laureano A., de Arriaga, F.: Reactive Agent Design For Intelligent Tutoring Systems. *Cybernetics and Systems, an International Journal* ISSN: 0196-9722. 31(1), pp 1-47. (2000).
13. Laureano A.: Agentes Pedagógicos. Memorias en CD, ISBN: 970-36-0155-3. En el XVII Congreso Nacional y III Congreso Internacional de Informática y Computación de la ANIEI. 20-22 de octubre, (Tepic, Nayarit). <http://delfosis.uam.mx/~ana/>. (2004).
14. Laureano A., Terán A., Rodríguez, R.: Cognitive and Affective Interaction in a Pedagogical Agent. *Memorias en CD, ISBN: 970-31-0528-9. XVIII Congreso Nacional y IV Congreso Internacional de Informática y Computación de la ANIEI, 26-28 de octubre*, (Torreón, Coah.), <http://delfosis.uam.mx/~ana/>. (2005).
15. Laureano A., Velasco P., Mora M., Acevedo-Moreno, D. Hacia Interfaces Inteligentes. En el *Libro Científico: Avances de las Mujeres en las Ciencias, las Humanidades y todas las disciplinas. Sustentabilidad, Aire, Agua, Fuego y Tierra*, pp. 237-246. ISBN: 978-607-477-212-8CYAD/UAM-A. (2009).
16. Martínez N., Ferreira G., García M., García Z.: El razonamiento Basado en Casos en el Ambito de la Enseñanza/Aprendizaje. *Revista de Informática y Medios Audiovisuales* Vol. 5(10), págs. 25-32. ISSN 1667-8338. (2008).
17. Newell, A.; Simon, H.: *Human problem solving*. Englewood Cliffs, NJ. Prentice-Hall. (1972).
18. Riviere A.: *El sujeto de la Psicología Cognitiva*. Editorial Alianza: Madrid. (1987).
19. Rodríguez R., Laureano A., Sánchez de Antuñano J.: Diseño de un Ambiente Virtual de Aprendizaje: un enfoque reactivo a las distintas inteligencias. XXIII Congreso Nacional y XI Congreso Internacional de Informática y Computación de la ANIEI, Puerto Vallarta 11-15. ISBN 978-607-707-097-9, pp. 494-505. (2010).
20. Rodríguez, M., Bazán P., Ungaro A.: Spaces virtual with educational value, research laboratory in new technology computer, Facultad de Informática, Universidad Nacional de La Plata, Buenos Aires, Argentina. 11 pp. (2006).
21. Rembold C, Watson DW.: Pos-test probability by weights: a simple form of Bayes's theorem. *Ann Intern Med.* Pags. 108: 115-20. (1988).
22. Sánchez, E., Lama M.: Monografía: Técnicas de la Inteligencia Artificial aplicadas a la Educación. *Inteligencia Artificial, Revista Iberoamericana de Inteligencia Artificial.* pp. 7-12. ISSN: 1137-3601. © AEPIA. (<http://www.aepia.org/revista>). No.33 (2007).
23. Velasco P., Laureano A. Mora M., Sánchez L.: La Importancia del Diseño de una Interfaz en el Proceso de Enseñanza-Aprendizaje. Memorias en CD, ISBN 978-970-15-14388-2. En XXI Congreso Nacional y VII Congreso Internacional de Informática y Computación de la ANIEI. 1-3 de octubre, (Monterrey) pp. 108-113. <http://delfosis.uam.mx/~ana/>. (2008).



# Refined Achromatic Zone Segmentation in Color Images

Rodolfo Alvarado-Cervantes and Edgardo M. Felipe-Riveron<sup>1</sup>

Center for Computing Research, National Polytechnic Institute, Juan de Dios Batiz w/n, Col.  
Nueva Industrial Vallejo, P.O. 07738, Mexico  
ateramex@gmail.com; edgardo@cic.ipn.mx

**Abstract.** We present an interactive, semiautomatic image segmentation method that processes the color information of each pixel as a unit, thus avoiding color information scattering. The color information of every pixel is integrated in the segmented image by an adaptive color similarity function designed for direct color comparisons. The border between the achromatic and chromatic zones in the HSI color model has been transformed in order to improve the quality of the pixels segmentation when their colors are very obscure and very clear. The color integrating technique is direct, simple and computationally inexpensive, and it has also good performance in low chromaticity and low contrast images. It is shown that segmentation accuracy is above 95% as average and that the method is fast. These results are significant when compared to other solutions found in the current literature.

**Keywords:** Color image segmentation; Adaptive color similarity function; Improved HIS color model; Achromatic zone definition.

## 1 Introduction

Image segmentation consists of partitioning an entire image into different regions, which are similar in some predefined manner. Segmentation is an important feature of human visual perception, which manifests itself spontaneously and naturally. All subsequent steps, such as objects recognition depend on the quality of segmentation. Without a good segmentation algorithm, objects of interest in a complex image are difficult (often impossible) to recognize using automated techniques [1] [2] [3].

At present, several segmentation techniques are available for color images, a good amount of them are monochromatic methods applied on the individual planes in different color spaces where the results are combined later in different ways [3]. A common problem to this approach is that when the color components of a particular pixel are processed separately the color information is so scattered in its components that most of the color information is lost [1] [3] [4].

In this work, an interactive, semiautomatic image segmentation method is presented that uses the color information for each pixel as a whole, thus avoiding color information scattering.

---

<sup>1</sup> Corresponding Author. Tel.: (52)-55-5729 6000 Ext. 56515  
**E-mail address:** edgardo@cic.ipn.mx

One weakness in the characterization of the achromatic region as presented in [5], [7] and [10] is its poor performance in the border regions of achromatic and chromatic zones, both in low and high brightness, due to the fact that the commonly used HSI color model does not take into account the human visual response at low and high brightness. Human vision has a nonlinear perceptual response to luminance [11]. This problem was overcome by modifying the saturation value of every pixel by a factor that reflects the human exponential response to brightness. We show this improvement in Section 3 comparing the achromatic zone obtained by both methods.

The results tabulated in Section 3 show that the segmentation method presented in this paper offers a useful and efficient alternative for the segmentation of objects with different colors in relatively complex color images with good performance in the presence of the unavoidable additive noise. It has also good performance in gray level and low contrast images [5].

## 2 Description of the Method

The segmentation method proposed in this paper relies on the calculation of a color similarity function for every pixel in a RGB 24-bit true color image to form what we call a Color Similarity Image (CSI), which is a gray level image. A true color image usually contains millions of colors and many thousands of them represent the same perceived color of a single object due to the presence of additive noise, lack of definition between color borders and regions, shadows in the scene, the spatial resolution of the human vision system, etc. [1] [6] [7]. The color similarity function allows the clustering of the many thousands of colors representing the same perceived color in a single gray output image. The CSI is then automatically thresholded and the output can be used as a segmentation layer, or it can be modified with morphological operators to introduce geometric enhancements if they are needed. The generation of a CSI only requires calculating Eq. 1 (below) for every pixel in the RGB input image. Thus the complexity is linear with respect to the number of pixels of the source image and for that reason inexpensive computationally.

Firstly, we compute the color centroid and color standard deviation of a small sample consisting of few pixels (less than 3-4 pixels per color). The computed centroid represents the desired color to be segmented using the technique we designed for that purpose.

Then, our color similarity function uses the color standard deviation calculated from the pixel sample to adapt the level of color scattering in the comparisons. The result of a particular similarity function calculation for every pixel and the color centroid (meaning the similarity measure between the pixel and the representative color value) generates the CSI. The generation of this image is the basis of our method and preserves the information of the color selected from the original color image. This CSI is a discrete representation in the range of [0 - 255] of a continuous function whose values are in the normalized range [0 - 1].

The CSI can be thresholded with any automatic thresholding method. To obtain the results presented in this work we used Otsu's method [8] [9].

To generate a CSI we need: 1. A color image in RGB 24-bit true color format and 2. A sample of arbitrarily located pixels forming a sample of the color desired to be segmented. From this sample of pixels we calculate the statistical indicators according to our HSI modified color model which can be consulted in [5]. This information is necessary to adapt the color similarity function in order to obtain good results. To obtain the CSI we calculate for every pixel  $(i, j)$  in the image the following empirical color similarity function  $S$  (Eq. 1).

$$S_{i,j} = e^{\left(\frac{-\Delta_h^2}{2\sigma_h^2}\right)} * e^{\left(\frac{-\Delta_s^2}{2\sigma_s^2}\right)} * e^{\left(\frac{-\Delta_i^2}{2\sigma_i^2}\right)} \quad (1)$$

where  $\Delta_h$  is the hue distance between  $hue(i,j)$  and the *average\_hue*;  $\Delta_s$  is the saturation distance between  $saturation(i,j)$  and the *average\_saturation*;  $\Delta_i$  is the intensity distance between  $intensity(i,j)$  and the *average\_intensity*;  $\sigma_h$  is the hue standard deviation of the sample;  $\sigma_s$  is the saturation standard deviation of the sample;  $\sigma_i$  is the intensity standard deviation of the sample. In Equation (1) the color information is integrated giving high importance to perceptual small changes in hue, as well as giving wide or narrow tolerance to the intensity and saturation values depending on the initial sample, which is representative to the desired color to be segmented. The use of Gaussians in the definition of  $S_{i,j}$  (Eq. 1) reflects our belief that the color model modifications proposed in this paper takes into account normal distributions of the color attributes in the modified HSI space.

The HSI color model represents colors similarly as humans describe colors in terms of hue, saturation and brightness. Hue is a color attribute that describes pure color whereas saturation gives a measure of the degree to which a pure color is diluted by white light. Brightness is a subjective descriptor that embodies the achromatic notion of intensity [6].

We use the intensity value but we believe that this choice is of minor importance because the achromatic information is much less important to discriminate colors than the chromatic one (mainly the hue). So the use of all this different spaces should give approximately the same results.

The common disadvantages attributed to the HSI color space such as the irremovable singularities of hue in very low saturations or the periodical nature of hue [7], which is lost in its standard representation as an *angle*  $\in [0^\circ, 360^\circ]$ , are overcome in our technique using vector representation in  $\mathfrak{R}^2$ , in the separation of chromatic and achromatic regions, and in the definition of the  $\Delta_h, \Delta_s$  and  $\Delta_i$  distances. We use the Euclidean distance to define saturation distance  $\Delta_s$  and intensity distance  $\Delta_i$  between two pixels or color centroids. We calculate the hue distance  $\Delta_h$  with vector difference. The statistical values needed in Eq. (1) are calculated from the pixel sample using the common statistics formulae adapted to our model. The details about all this can be consulted in [5] and [10]. Some modifications of the standard HSI color space were necessary in order to create a consistent model to represent colors and color centroids; they can be consulted for reference in [5] and [10].

## 2.1 Pixel Sample Selection

The pixel sample is a representation of the desired color(s) to be segmented from a color image. From this pixel sample we calculate two values to feed our segmentation algorithm: the color centroid and a measure of the dispersion from this centroid, in our case the standard deviation.

If we take only one pixel, its color would represent the color centroid, and would produce dispersion equal to zero, giving a Dirac delta in the calculation of Eq. (1). This means that the similarity function would be strictly discriminative to the pixel color. This is not the general intention of segmenting color images which usually contain millions of colors, many thousands of which are due to additive noise.

If we additionally take another pixel, we obtain then the centroid from both and the standard deviation of them to feed our algorithm. So when we look for this additional pixel, we should take it from a region which was not or poorly segmented when we used only the first pixel. That means that in order to obtain good results the pixels for the sample should be taken from the color regions in such a form that they must constitute a good representation of the colors of the region.

If we continue adding more and more pixels to the sample we find that the corresponding centroid of the area to be segmented increases in accuracy. Here we may have a relatively minimum representative sample of the color area to segment. Beyond a given number of well-selected pixels in the sample to increase them does not affect sensibly the segmentation quality because adding more pixels to the sample of approximately the same perceived color does not affect the statistical estimators to feed the algorithm in a sensible manner.

In order to obtain the average of the hue ( $H_m$ ) of several pixels from a sample, we take advantage of the vector representation of hue in  $\mathcal{R}^2$ . Vectors that represent the hue values of individual pixels are combined using vector addition. From the resulting vector we obtain the average hue corresponding to the angle of this vector with respect to the red axis.

## 2.2 The Achromatic Zone G

The achromatic zone G is the region in the HSI color space where no hue is perceived by the human. This means that color is perceived only as a gray level because the color saturation is very low or the intensity is either too low or too high.

In order to model better the human visual response in the abrupt corners regions near the union zones of the two cones with the cylinder of the singularity zone G, as presented in [5], [7] and [10], we found heuristically convenient to modify the characterization of the HSI color model for colors belonging to regions with very low or very high brightness. This step will permit us to differentiate the colors in these extreme regions, because the human response to brightness does not have this possibility. In order to adjust the zone G to a better model nearer to the human response to brightness we introduced an exponential function with three parameters to define the improved singularity of zone G: saturation threshold (st), inflection point (ip), and

Slope (Fig 1). Then, the saturation image will be affected now by a factor calculated from the inflection point (ip), and Slope for every pixel (P) as shown in Eq.(2):

$$New\_Saturation\_ (P) = \left(1 - Slope * (e)^{abs(ip-intensity(P))}\right) * Saturation\_ (P) \quad (2)$$

where Slope is a parameter representing the degree of affectation of the exponential factor to the saturation attribute for every pixel; abs( ) represents the absolute value function and intensity( ) is defined as an average of the RGB channels in Eq. (6) in [5]. After calculating the New\_Saturation values for the source image, it is thresholded with the saturation threshold (st) to obtain the effects of the improved achromatic zone G.

We found good performance with the parameters empirically determined in the following ranges: st ∈ [0.07, 0.1], ip ∈ [0.5, 0.6], and Slope ∈ [1, 1.5].

### 3 Experiments and Results

In this section we describe the experiments and present the results achieved from our segmentation method applied to two classical color images in RGB 24-bit true color format. These experiments consisted of segmenting color regions according to the following two steps:

Selection of the pixel sample. This is the only step to be left up to the user, and [5] can be consulted .as a guide. This step is automatic; its output is a gray image showing the similarity of each pixel of the RGB true color image to the color centroid formed with the chosen pixel sample taken from of the region of interest to be segmented, being white for 100% of similarity and black for 0%.

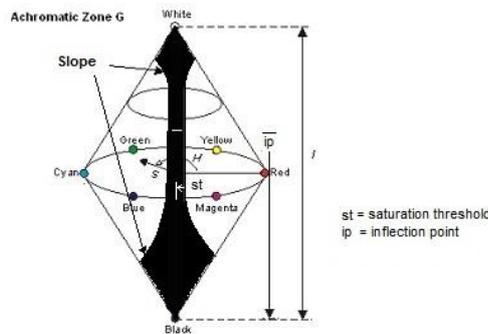


Fig. 1 The new achromatic zone G

The user can now threshold the CSI. This step could be necessary to obtain a template for a final segmentation of the desired color from the region of interest; it could be arranged as an automatic step by using, for example, the non-supervised Otsu thresholding method [8], [9].

We can also separate the achromatic area to obtain a better segmentation. This achromatic zone can be subdivided later using only intensity as a discriminative characteristic. In Figs. 2, 3 and Table 1 we show the improved results of our new definition of  $G$  in Eq. (2) with respect to the former results presented in [5], [7] and [10].

In all the composite images we used the XOR logical function to avoid the possibility that one pixel could belong simultaneously to two different color-segmented zones or regions.

Next, we demonstrate the effectiveness of the proposed color segmentation method in some relatively complex color images.

Figure 2 (left) shows the results of the application of our method to the RGB color image (sized 200 x 200 pixels and with 33 753 different colors) of the popular baboon image. In this image we can see four main hues of colors despite the many thousands of actual RGB values to represent them: The red part of the baboon's nose, the blue part of the nose, the orange eyes and the yellow-orange part of the fur.

In our experiments we do not use any preprocessing at all. We obtained 95.5% of pixels segmented properly (See Table 1). In figure 2 we show the results with the old (center) and new achromatic region  $G$  (right) for appreciation. The improvement in quality is significant.

As another example, consider the image of two horses with shadows shown in figure 3 (left). Figure 3 (right) shows the composite image of the different segmented areas. Non-segmented pixels are shown in pseudo red color. We obtained a 97.1% of pixels segmented (See Table 1). Figure 4 shows the segmentation of the achromatic zone with the former (left) and the new definition (right). The improvement in quality is appreciable.



**Fig. 2** Composite image of 5 segmented colors (left) and results obtained with the old (center) and the new achromatic zone  $G$  (right).

In order to evaluate the efficiency of the color segmentation method and due to the difficulty of obtaining a ground-truth for each complex image to which we applied the method or to compare the results from different methods, the evaluation was based on the number of pixels not segmented with respect to the total number of pixels in the corresponding image. This ratio combines the pixels not belonging to any color cluster and those selected by two or more clusters (obtained by means of the XOR operation) with respect to the total number of pixels in the image. It gives us a measure of the segmentation efficiency. Obviously, the number of different colors in each image

is always less than the total number of pixels in it. The maximum number of selected pixels related to each color sample was always ten or less. In general the number of selected pixels for the samples depends on the complexity of the image (texture of objects), on the diversity of colors and on their spatial distribution in the image.

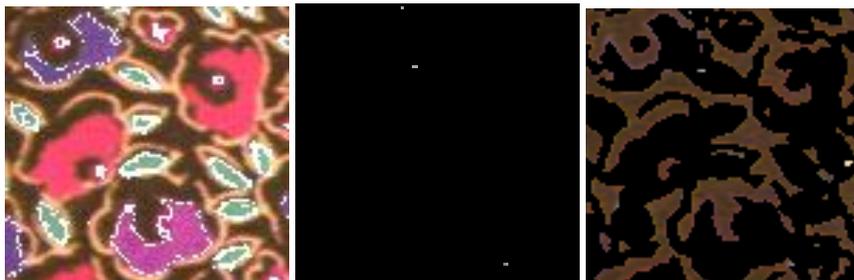
Table 1 show the results obtained in the segmentation of three images. The last column of Table 1 shows the percent of segmented pixels obtained in these images.. In all cases, the possibility that one pixel could belong to two different colors segmented zones or regions have been avoided by means of the application of the XOR logical function (of two or more partial segmentations). A result with many black pixels indicates that has been coincidence in segmented pixels from two or more partial segmentations considered in the XOR operation.



**Fig. 3** Composite image of 3 segmented colors (left) and results obtained with the old (center) and the new achromatic zone G (right).

As it can be observed from Table 1, the average accuracy of the color segmentation in both complex images, without an exhaustive selection of colors and a few numbers of pixels (4-5 as average) per color sample, is 96.55%.

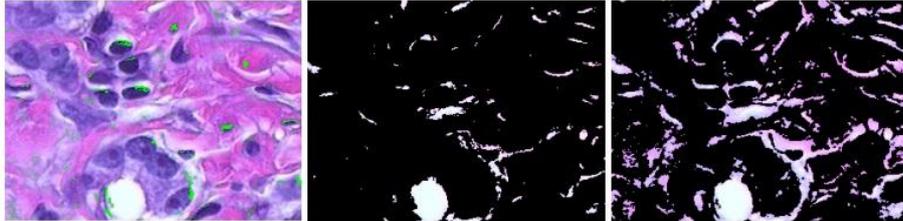
In Figure 4 (left) a composite image of 5 segmented colors (with the non-segmented pixels in pseudocolor) of a color fabric is shown; it contains an achromatic region with high saturation but low brightness that is very difficult to separate with the former G zone (Figure 4 center) but easily separated with the new one shown in Fig. 4 (right).. We obtained 95.3% of pixels segmented.



**Fig. 4** Composite image of 5 segmented colors (left) and results obtained with the old (center) and the new achromatic zone G (right).

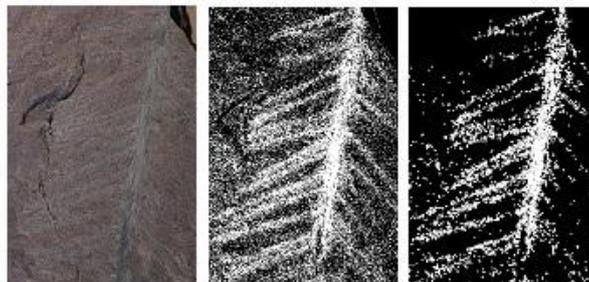
When we compare the new achromatic region G with the former one, the difference is astonishing whereas the new achromatic area seems correctly segmented the former one segmented only a couple of pixels.

Figure 5 (left) shows a composite image of 3 segmented colors of tissue stained with hemotoxylin and eosin (H&E) color image (which is a very popular staining method in histology and the most widely used stain in medical diagnosis) with the non segmented pixels in pseudocolor. The former and the new achromatic areas are shown in Figure 5 (center) and Figure 5 (right) respectively.



**Fig. 5** Composite image of 3 segmented colors (left) and results obtained with the old (center) and the new achromatic zone G (right).

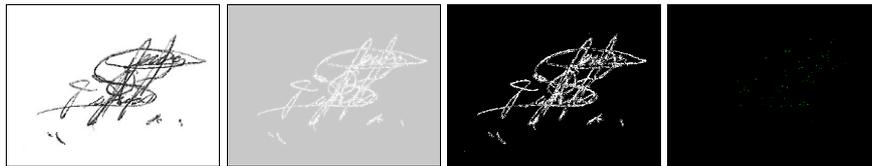
We will show the good results obtained by our method applied to low contrast color images and gray level images in the following three examples: Fig. 6 shows a fossil inserted in a rock, we took a small pixel sample of the fossil area from which we obtained its corresponding CSI (Fig. 6 center). Figure 6 (right) shows the thresholded image after applying the Otsu thresholding method and some morphological processing. Figure 7 shows a gray level image obtained with an infrared camera, we took a small pixel sample from the face area and obtained its correspondent CSI shown in Fig.7 (second left to right). The segmented face appears in Fig.7 (third left to right) after thresholding by Otsu method. Fig.7 (extreme right) shows a composite of the segmented pixels with the non-segmented pixels in green for easier appreciation. Figure 8 shows a gray level image of a person's signature. After taking a four pixels sample from the signature we calculated its CSI shown in Fig. 8 (second left to right). Fig.8 (third left to right) shows the final segmentation after thresholding by Otsu method. Fig.8 (extreme right) shows a composite of the segmented pixels with the non-segmented pixels in green.



**Fig. 6** Fossil embedded in rock. CSI of fossil Thresholded by Otsu and processed



**Fig. 7** Low contrast infrared Image CSI of face Segmented face Pixels not segmented in green



**Fig. 8** Signature CSI of signature CSI thresholded by Otsu Pixels not segmented in green

**Table 1** Results of the global segmentation per image.

No.	Image	Number of pixels in image	Number of colors (levels)	Number of colors (levels) selected	Total number of pixels used as samples	Number of non-segmented pixels	% of segmented pixels
1	Baboon	40000	33753	5	31	1803	95.5
2	Horses	154401	71727	3	14	4619	97.1
3	Fabric	9900	9349	5	23	465	95.3
4	Stained tissue	68026	27146	3	11	1171	98.3
5	Fossil	121128	40756	7	15	95 262	-
6	Gray level Infrared	81426	255	4	12	347	99.57
7	Signature	98 532	255	2	4	76	99.92

#### 4 Conclusions

The results in the previous section, demonstrate that the color segmentation method presented in this paper offers a useful and efficient alternative for the segmentation of objects with different colors in relatively complex color images with good performance in the presence of the unavoidable additive noise, and in images with low contrast, and with very low chromaticity.

The steps required to obtain a good segmentation of regions with different colors by using the proposed methodology are usually straightforward, simple and repetitive. If color is a discriminative characteristic in the layer of interest, only the selection of a given threshold to the color similarity image CSI is needed to obtain a good segmentation result. From many experiments we have observed that a good percentage of colors were segmented in a straightforward way only by thresholding the Color Similarity Image. The method discriminates whichever type of different color objects independently on their shapes and tonalities in a very straightforward way.

The new characterization of the achromatic region improves its performance, compared to that presented in [5], [7] and [10], due to the affectation of saturation by an exponential factor in an effort to better model the human visual response in case of very low or very high brightness. The improvement in quality of its results is significant and can be visually appreciated in examples of Section 3.

One of the inconveniences of the proposed color segmentation method is the degree of human intervention, summarized in the necessity of selecting a pixel sample. This sample has to be carefully chosen in order to obtain the best results. An associated problem of the use of a pixel sample is to obtain the number of data points necessary to capture the color distribution. These related problems will be treated in future developments of the algorithm enforced to obtain a complete automatic version.

**Acknowledgements.** The authors of this paper wish to thank the Computing Research Center (CIC), Mexico; Research and Postgraduate Secretary (SIP), Mexico, and National Polytechnic Institute (IPN), Mexico, and CONACyT, Mexico, for their support to this work.

## References

1. Alvarado-Cervantes R, 2006. Segmentación de patrones lineales topológicamente diferentes, mediante agrupamientos en el espacio de color HSI, M. Sc. Thesis, Center for Computing Research, National Polytechnic Institute, Mexico.
2. Felipe-Riveron, E.M., García-Ramos, M.E. Levachkine S.P., 2000. Problemas potenciales en la digitalización automática de los mapas cartográficos en colores. In: Proceedings of International Congress on Computation CIC IPN, Mexico City, Mexico.
3. Cheng H., Jiang X, Sun Y., Wang J., 2001. Color image segmentation: Advances and prospects, *Pattern Recognition*, 34(12): 2259-2281, Elsevier.
4. Hanbury, A., Serra, J. A., 2002. 3D-polar coordinate colour representation suitable for image analysis, Technical Report PRIP-TR-77, Austria.
5. Alvarado-Cervantes R. and Felipe-Riveron E.M., 2011. 16th Iberoamerican Congress on Pattern Recognition, CIARP 2011, An Adaptive Color Similarity Function for Color Image Segmentation, Rodolfo Alvarado-Cervantes y Edgardo M. Felipe-Riveron, C. San Martín and Sang-Woon Kim (Eds.), Pucón, Chile, 15-18/Nov/2011, LNCS 7042, pp. 113--124. Springer, Heidelberg.
6. Gonzalez, R.C., Woods, R.E., 2008. *Digital Image Processing*, third Edition, Prentice Hall, USA.

7. Plataniotis, K.N., Venetsanopoulos, A.N., 2000. *Color Image Processing and Applications*. First Edition, Springer, Germany.
8. Otsu N., 1979. A Threshold Selection Method from Gray-Level Histograms, *IEEE Transactions on Systems, Man and Cybernetics*, 9(1):62-66.
9. Sezgin M. and Sankur B., 2003. Survey over image thresholding techniques and quantitative performance evaluation. *Journal of Electronic Imaging* 13 (1): 146–165. doi:10.1117/1.1631315
10. Alvarado-Cervantes R., Felipe-Riveron E.M. and Sanchez-Fernandez L.P., 2010. 15th Iberoamerican Conference on Pattern Recognition, CIARP 2010, Color Image Segmentation by means of a Similarity Function, Rodolfo Alvarado-Cervantes, Edgardo M. Felipe-Riveron y Luis P. Sánchez-Fernández, I. Bloch and R.M. Cesar, Jr. (Eds.): Sao Paulo, Brasil, November 8-11, 2010, LNCS 6419, pp. 319--328. Springer, Heidelberg.
11. Poynton C., 2002. <http://www.poynton.com/PDFs/GammaFAQ.pdf>
12. Avinava D., Indrajit B. and Shantanu G., 2010. A Cluster-Level Semi-Supervision Model for Interactive Clustering, *Machine Learning And Knowledge Discovery In Databases, Lecture Notes in Computer Science*, 2010, Volume 6321/2010: 409-42



# Retinal Image Biometric Coding using a Logarithmic Spiral and a Time Series Representation

Fabiola M. Villalobos-Castaldi<sup>1</sup>, Edgardo M. Felipe-Riverón<sup>2</sup>

<sup>1</sup>, *Center of Research and Advanced studies of the National Polytechnic Institute  
Av. IPN # 2508 Col. Instituto Politécnico Nacional s/n, La Laguna Ticoman,  
Gustavo A. Madero, Mexico City, D.F., 07360, Phone: (52)- 55 5747 3800/6211*

<sup>2</sup>, *Center for Computing Research, National Polytechnic Institute,  
Av. Juan de Dios Batiz and Miguel Othon de Mendizabal, Mexico, D.F., P. O.  
07738. Mexico, Phone: (52)-55-5729 6000/56515.*

<sup>1</sup>*mvillalobos@cinvestav.mx*, <sup>2</sup>*edgardo@cic.ipn.mx*

**Abstract.** The retina has many desirable characteristics as a basis for authentication, including uniqueness, stability, and permanence. In this paper, we propose a new approach for retinal images feature extraction and template coding. The use of the logarithmic spiral in scanning and tracking the vascular network, and the time series technologies such as SAX representation conform the key to make this new approach simple, flexible and reliable. Experiments show that this approach can achieve the reduction of data dimensionality and of the required time to obtain the biometric code of the vascular network in a retinal image.

**Keywords:** Biometric code, logarithmic spiral scan and tracking, time series, retinal images.

## Introduction

It has been proposed a number of authentication methods based on the retina. These methods have focused primarily on selecting the appropriate features to represent the retina (bifurcations, ending points, etc.). But the final representation of the features has never been studied carefully. Biometric representation (*template*) is a machine readable and understandable form of a biometric trait. It influences the system's accuracy and the design of the rest of the system. The machine representation of a biometric is critical to the success of the matching algorithm. In a practical authentication system, the database can contain records of millions of people. Choosing an appropriate representation of the features in order to make the database smaller in size, having a rapid response search and while retaining high accuracy in the verification, is a vital task.

From previous studies, the retina's features used for authentication can be classified into three main categories: structural, statistical and algebraic features. Some typical structural features include main lines (centerlines), branching points,

crossing points, termination points, positions, angles, diameters, etc. Some statistical features are the texture moments and the random values. Likewise, the algebraic features, such as band-tree based radial partition, the ring method, etc.

The main contribution of this paper will be the implementation of a new biometric representation method based on coding blood vessel segments, through a new scanning and tracking algorithm of a logarithmic spiral type, from which is obtained a time series representation of the local features of the detected and analyzed segments. The time series have been studied extensively in data mining, bioinformatics and pattern recognition in biometrics [1].

### **Disadvantages of the actual sampling methods**

There are several disadvantages that the actual sampling methods have when they are applied to retinal images. [1]-[16]. These are:

- The retina has an average of 400 minutiae points, between bifurcation, intersection and ending points, thus the point to point comparison for all the retinal images contained in a database is impractical and computationally expensive.
- To encode a minutiae, it is necessary to perform an image preprocessing step, within which is included retinal network thinning, which presents a problem because most existing thinning methods, if not all, create false positives when converting an intersection in a double crossing point and breaking segments; this problem generates a greater number of ending points than those that actually exist in the original image [28-32].
- To make the method more robust, it is necessary to extract additional information from the detected minutiae such as angles, distances between the minutiae and the reference point (center of the optic disc), minutiae distance, etc., that increases the required processing time [33].
- The minutiae-based encoding method does not use the retinal vascular network structure properties that are essentially the most robust and stable of all characteristics [34-35].

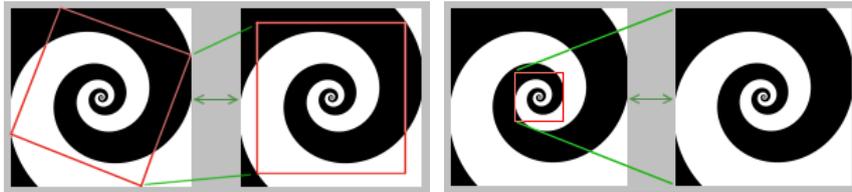
### **Logarithmic spiral**

A spiral is a curve that winds itself around a certain point. While not being a circle, the radius will vary along the angle [20], [21]. The logarithmic spiral is the spiral for which the radius grows exponentially with the angle. The logarithmic relation between radius and angle leads to the name of **logarithmic spiral**. In this curve the distances where a radius from the origin meets the curve increases in geometric progression.

The logarithmic spiral is a spiral whose polar equation is given by:

$$r = \quad \quad \quad (\text{Eq. 1})$$

where  $r$  is the distance from the origin,  $\theta$  is the angle from the  $x$ -axis, and  $a$  and  $b$  are arbitrary constants. The constant  $a$  is the rate of increase of the spiral. The sign of  $a$  determines the direction of rotation of the spiral. The logarithmic spiral is also known as the growth spiral, equiangular spiral, and spiral mirabilis. The logarithmic spiral is remarkable because of its unique self-similarity; it is invariant after a similarity transform. After any scaling (uniformly increasing or decreasing the size), logarithmic spirals can be rotated such that they match the original figure.



**Fig. 1** Self-similarity of the logarithmic spiral

The spiral boundaries can have any rate of twist, or pitch. The pitch is defined as the angle between a tangent to the curve and the tangent to the circle at that radius. The pitch is the same everywhere on a logarithmic spiral. Any such linear iterated function system (IFS) that includes scaling and rotation (an affine transformation) will contain logarithmic spiral patterns. Naturally, occurring algorithms (e.g. the expression of a genetic code, hurricane dynamics, and galaxy formation) commonly result in this shape. The shape also seems to be appealing to the eye, perhaps because our visual perception is tuned for interpreting similarity (scale and rotational invariance) of known objects.

#### **Advantages of the logarithmic spiral**

As sampling pattern method of the retinal image, the logarithmic spiral offers the following advantages:

1. The data dimensionality is reduced because the retinal vascular network is represented by a real valued data sequence [35].
2. The discontinuities are eliminated in the data caused by sampling, which occurs when concentric flattened circles or other sampling methods are used. The connectivity is a concept geometrically intuitive: A set is connected if it is composed of a single segment. Tuning the initial concept has come to define the connectedness paths: A set  $C$  is connected by road from any point  $x$  if the same can reach any point  $C$  and touring a "continuous path" that does not leave the set. We must recall that a connected component is a set of pixels such that for any pair of pixels on the whole, there is a digital path that connects them. Intuitively, a connected set is that composed of a single 'piece', which cannot be 'split' into parts [38].
3. Only the points of the vascular network detected by the spiral are encoded and not the entire region which includes the background; the features are extracted only

along the spiral path. For every point P in the spiral certain structural information at P is used as the feature [37].

4. A single sample for each point and a single sequence for each kind of coded data are provided [37].
5. Travel and distribute in the same way as the vascular network does: in geometric progression. The most problematic area is located within the optical disc which we had removed previously.
6. The most robust and stable structure of the image is coded, that is, the vessel segments of the vascular network and not to the branches, crossings and ending points. Bifurcations and crossings are eliminated in order to avoid coding errors [28-32].
7. It is not necessary to specify whether a given vessel is a vein or an artery, because the midpoint of every detected vessel segment is used as the feature descriptor.
8. The amount of coded information is increased because it is possible to extract some other features from the detected vessels along the spiral path [37].
9. The required size of the spiral or the number of turns is determined by the number of the necessary points that ensure the subject's individuality [36]. This provides invariance to small changes in the scale.

### **DRIVE Database**

In this paper we used the images included in the publicly available DRIVE database [17], to implement the proposed method and to assess its performance. It consists of 40 color retinal images of size  $565 \times 584$  pixels with 8 bits per color channel. The File of View (FOV) is circular with approximately 535 pixels in diameter. For each image, a mask image is provided that delineates the FOV. Hence, detection of the FOV border is not needed in this case. Images have been divided into 2 sets: a training set and a test set, each containing 20 images. The training set is useful to design supervised segmentation methods. Those images of the test set were segmented twice, resulting in a set A and a set B. In set A, 12.7% of pixels were marked as vessel, against 12.3% for set B. Performance is evaluated on the test set using the segmentations of set A as ground truth. All images were manually segmented.

### **The proposed method**

As depicted in Fig. 2, the proposed methodology is composed of two main processing stages. In the next section we discuss each stage in detail.

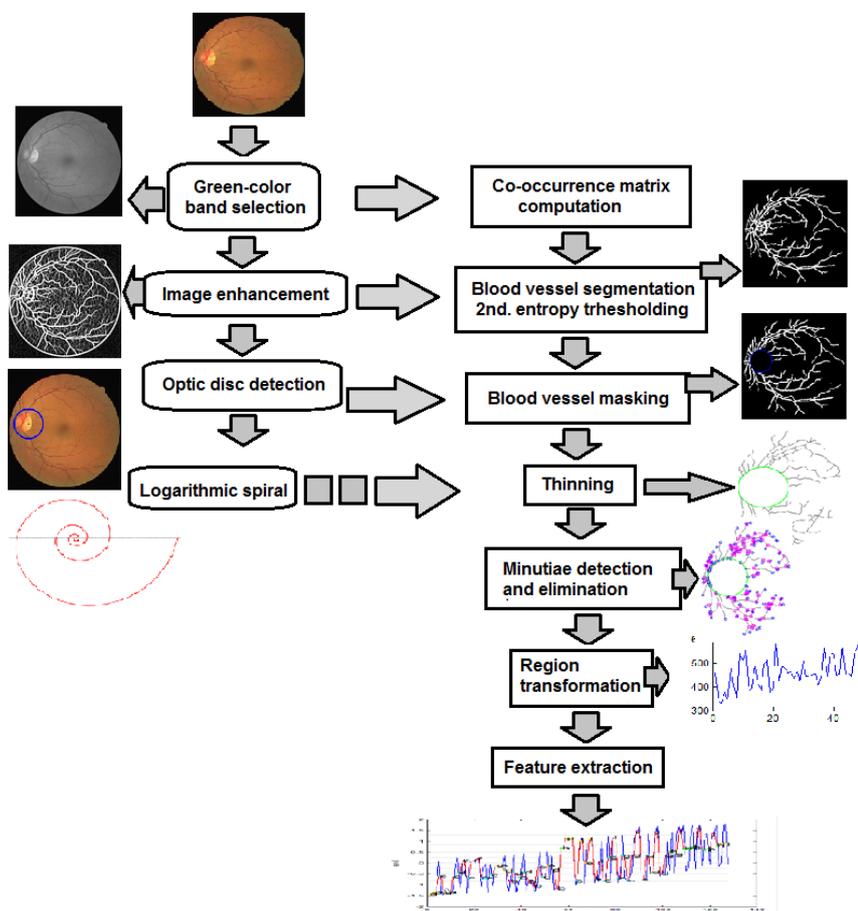


Fig. 2 Block diagram of the proposed method

**1. Preprocessing stage:** This step permits us to obtain the blood vessel image segmented from the original gray scale retinal image and thus, the minutiae feature extraction task is simplified. The preprocessing step consists of the following 4 stages: a) green-color band selection, b) image enhancement for vessel network detection, c) optic disc detection, and d) logarithmic spiral creation.

**2. Main processing stage:** This issue constitutes the essence of an automatic biometric authentication system design and has far reaching implications on the performance of the rest of the system. The main process consists of 7 stages: a) co-occurrence matrix computation, b) vessel segmentation by the second entropy thresholding technique, c) blood vessel masking in the vicinity of the optic disc, d) morphological thinning of the network, e) landmarks detection for their elimination, f) rectangular region transformation into sequential data, and g) time series creation that represent the features.

### **Preprocessing stage**

Preprocessing of the retinal image is a requirement after image acquisition due to the high variability of this process. The preprocessing stage is also required to obtain a better and a more homogeneous representation of the retinal image in order to achieve repeatability in the feature extraction stage [10]. Preprocessing of retinal image is necessary to: [15]

- Improve the contrast of blood vessel structures;
- Maintain their integrity;
- Avoid introduction of spurious structures or artifacts; and
- Retain the connectivity of the blood vessels while maintaining separation between them.

#### **2.0.1 Green-color band selection**

A gray-level image is obtained by extracting the green layer from the original RGB image. The green component has the blood vessels on a highly contrasted background (dark blood vessels on a relatively bright background). Hence, the green channel of image is preferred for the retinal vasculature detection [2] [16].

#### **2.0.2 Image enhancement**

Image denoising and contrast enhancement are needed before applying the vessel segmentation algorithm for landmark extraction. Uneven illumination (also called shading) is present in retinal images and must be suppressed in order to achieve more accurate segmentation of the blood vessels. In order to characterize the retinal features of interest, we use a Gaussian matched filter (GMF) to detect piecewise linear segments of blood vessels in retinal images [18].

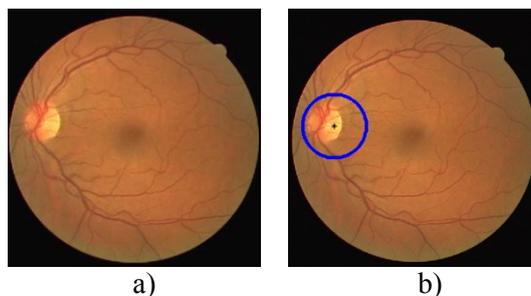
#### **2.0.3 Optic Disc Detection**

The optic disk is the brightest area in images that have not large areas of exudates and it is a slightly oval disk. It is the entrance region of vessels and its detection is very important since it works as a landmark for the other features in the retinal image. Hence, by targeting common structures such as the optic disc and the retinal vascular branches, a consistent source of readily identifiable, yet contrasting structures are available for digital imaging and processing. The entrance point of the optic nerve itself is taken as a point of reference. The distances and directions of the vein forks from this reference point provide coordinates which can be hooked together in a serial number for classified filing and quick comparison.

In the preprocessing technique used to detect the optic disc, the blood vessels are “erased” from the original retinal fundus image through the successive application of two morphological operations with a disk-like structuring element with N pixels radius. First, it is used a morphological opening operation. The circled structuring element is applied on the green color band of the fundus image. Next, it is used a

morphological closing operation employing the same circled structuring element, which is applied to the resulting image obtained from the previous step operation. Then, the resulting image becomes a binary mask by means of a thresholding scheme, which uses the Matlab function “IM2BW” found in the library of the image processing Toolbox. The threshold value  $h$  is computed by using the Matlab function “GRAYTHRESH” included in the same Toolbox. After this, the obtained binary image is labeled. The area of the labeled regions in the binary image is calculated and the region with the greatest area is isolated, which corresponds to the optic disc. Finally, we compute the coordinates of the centroid of this isolated region to be used as the reference point [16].

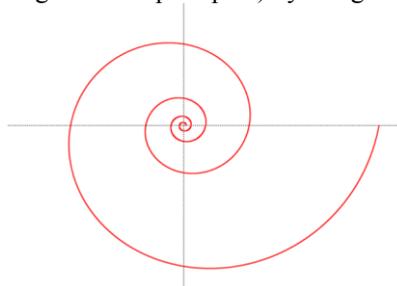
The experiment results of the preprocessing step of our algorithm are illustrated in Fig. 3. Fig. 3 a) shows the original image used to illustrate the accurate detection of the optic disc location. Fig. 3 b) illustrates the origin of the detected location of the optic disc indicated by a blue asterisk surrounded by a circle of the same color.



**Fig. 3** Experimental results of the optic disc detection step for a typical image. a) Original image, b) optic disc detected

#### 2.0.4 Logarithmic spiral creation

We create the logarithmic spiral (Fig. 4) starting from the center of the optic disc (used as the origin of the logarithmic spiral path) by using the Eq. 1.



**Fig. 4** Logarithmic spiral

### Main Processing stage

To characterize the searched retinal features, it is necessary a good image preprocessing. Therefore, the propose method detect vessels using the knowledge of their known gray level profile and the concept of the matched filter detection, which is used to detect piecewise linear segments of blood vessels in retinal images.

#### 2.1.1 Gaussian matched filter

A matched filter is constructed for the detection of the vessel edge segments searching in all possible directions [19]. A Bell-Shaped Gaussian matched filter (BSGMF) was developed to cover all 12 orientations where designed kernel is given by Eq. 2.

$$K_{x,y} = \pm \left( \begin{matrix} \cdot \\ \cdot \\ \cdot \end{matrix} \right) \quad (\text{Eq. 2})$$

The application of this method enhances individual vessels segments in the image. A proper thresholding scheme must be used to distinguish between the enhanced vessel segments and the background.

#### 2.1.2 Second-entropy thresholding segmentation method

The proposed segmentation thresholding method exploits the entropy of the distribution of the gray levels in the image. The maximization of the entropy (Eq. 3) of the thresholded image is interpreted as indicative of maximum information transfer [22], [23], [24]. In Figure 5 some examples of the segmented blood vessels, using the 2<sup>nd</sup> local entropy thresholding method are presented.

$$t_{LRE} = \max_{t \in [0, 255]} J_{LRE}(t) \quad (\text{Eq. 3})$$

#### 2.1.3 Blood vessel masking

Eventually, the edge of the optic disc could be misinterpreted by the segmentation method as a blood vessel, therefore, it is necessary to delete or hide this area to reduce the errors that its structure may cause.

In order to remove the blood vessels within the region of the optic disc, we superimposed a disc-shaped binary mask with a radius  $r_1$  10% greater than the radius of the optic disc centered at optic disc location. It is possible to do so when we can assure that all images are of the same size. Two examples of the blood vessel erased images are shown in Fig 5.

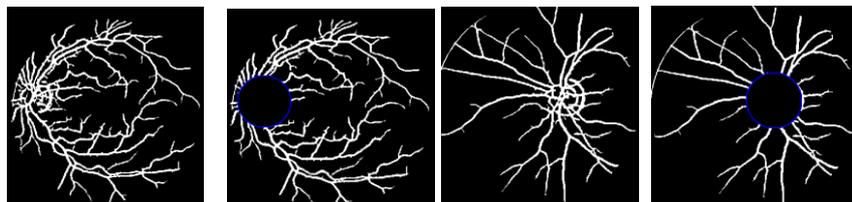


Fig. 5 Results of the erasing blood vessels in the vicinity of the optic disc

**2.1.4 Morphological thinning**

Vessels in the vicinity of the optic disc have different diameter size. To overcome this, retinal vessel thinning (through a skeletonizing method) is usually implemented using a morphological operator that reduces the width of vessels to a single pixel width line segments while preserving the extent and connectivity of the original shape. The thinned representations is typically easier to process in later stages producing savings in both time and storage complexity [25].

**2.1.5 Minutiae extraction for their elimination**

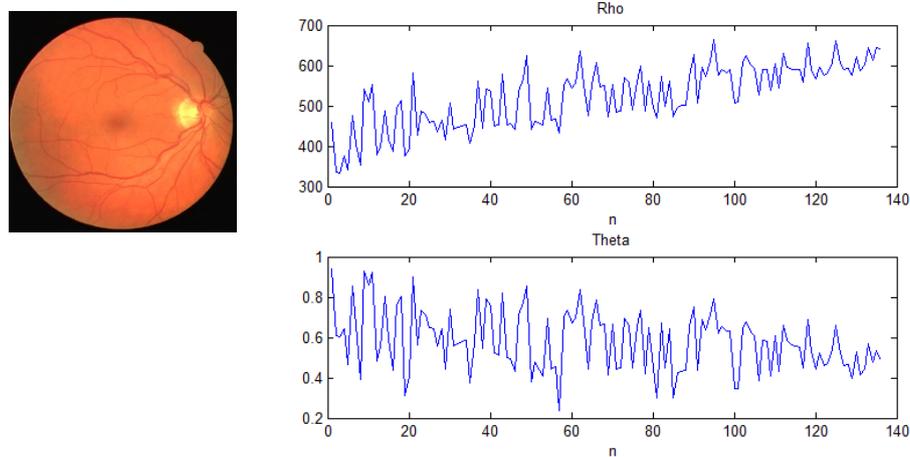
The bifurcation points in the thinned vessel tree are detected using the crossing number method [26] and then erased, thus only the vascular structure of the vessel segments is retained.

**2.1.6 Rectangular region transformation**

In order to represent the retinal vascular network using a time series, the rectangular region with the skeletonized vascular network must be transformed into sequential data. There are many possible ways of decomposing a 2D image into sequential data. In this work, we adopted a logarithmic spiral as the track for the decomposition. Thus, only this region is sampled with the logarithmic spiral, starting from the external edge of the masked optic disc towards the periphery of the vascular network.

**2.1.7 Feature extraction and representation**

For each point on the spiral, the position and angle of the midpoint of the blood vessel segments are encoded and a time series is created for each descriptor, one time series for the position and one time series for the angle. Figure 6 show both time series of a typical coded retinal image using a logarithmic spiral with a parameter  $a = 0.15942$ .



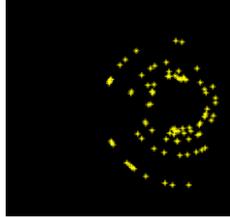
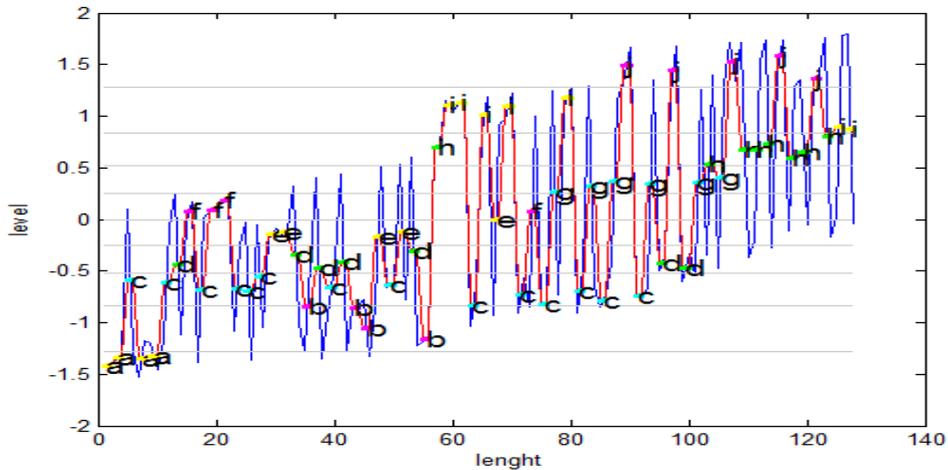


Fig. 6 Time series with  $a = 0.15942$

All the steps mentioned above are repeated using different parameter values of the logarithmic spiral equation ( $a$  and  $b$ ) in order to analyze the behavior of the generated biometric code.

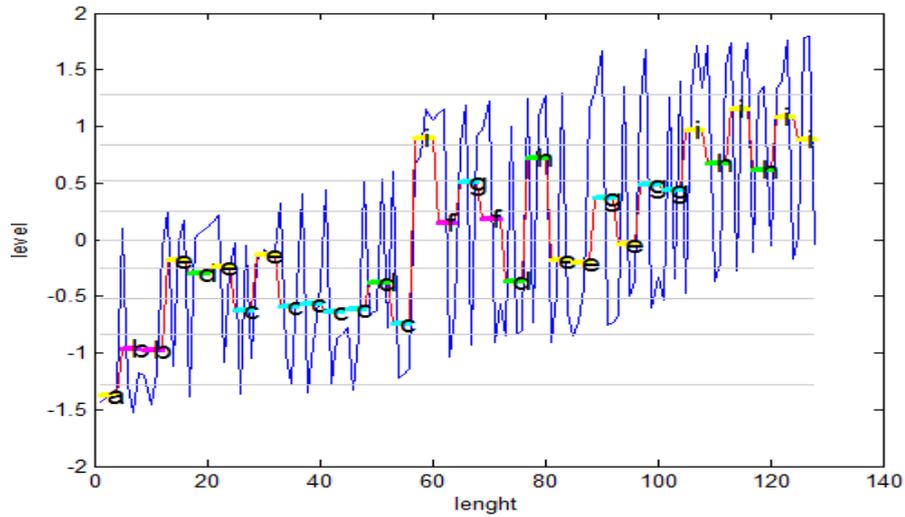
### 2.1.7.1 SAX Representation

Dimensionality reduction is achieved by converting time series into its SAX representation  $Q_{sax}$  [27]. We first transform the time series into the Piecewise Aggregate Approximation (PAA) representation and then symbolize it as a discrete string. The PAA representation is merely an intermediate step required to obtain the symbolic representation, to be transformed later in the SAX discrete sequence,  $Q_{SAX}$ . The number of segments and levels of the  $Q_{SAX}$  can be used to control the degree of the dimensionality reduction (See figure 7).



no. Seg = 64;  
data length = 10;

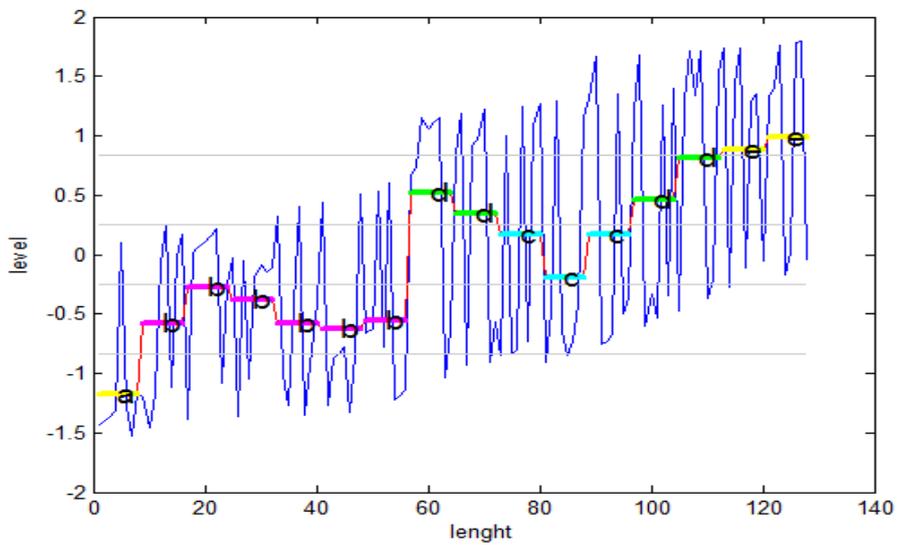
'a' 'a' 'c' 'a' 'a' 'c' 'd' 'f' 'c' 'f' 'f' 'c' 'c' 'c' 'e' 'e' 'd'  
'b' 'd' 'c' 'd' 'b' 'b' 'e' 'c' 'e' 'd' 'b' 'h' 'i' 'i' 'c' 'i' 'e'  
'i' 'c' 'f' 'c' 'g' 'i' 'c' 'g' 'c' 'g' 'j' 'c' 'g' 'd' 'j' 'd' 'g'  
'h' 'g' 'j' 'h' 'h' 'h' 'j' 'h' 'h' 'j' 'h' 'i' 'i'



No. seg = 32;

Data length = 10;

'a' 'b' 'b' 'e' 'd' 'e' 'c' 'e' 'c' 'c' 'c' 'c' 'd' 'c' 'i' 'f' 'g'  
 'f' 'd' 'h' 'e' 'e' 'g' 'e' 'g' 'g' 'i' 'h' 'i' 'h' 'i' 'i'



No. seg = 16;

Data length = 5;

'a' 'b' 'b' 'b' 'b' 'b' 'b' 'd' 'd' 'c' 'c' 'c' 'd' 'd' 'e' 'e'

Fig. 7 Three SAX representations and data sequences of the same retinal image with different parameters

Fig. 7 shows some examples of the extracted data sequence and the corresponding SAX representations with different lengths and levels illustrated in a typical retinal image. The SAX representation  $Q_{sax}$  is used as the template of the retinal vascular network.

## **Experimental results**

In this section we present our experimental results on the performance of the proposed blood vessel segmentation method and of the feature extraction scheme. The blood vessel segmentation and feature extraction approach introduced in the last stages of the system will be evaluated to verify the suitability of such proposal.

The first stage of the evaluation is the blood vessel segmentation method. The second part of the experiment will be focused on testing the feature points extraction step.

### **Blood vessel segmentation performance**

The proposed segmentation method was applied to all 20 images of the training set of the DRIVE database, and TPR (true positive rate) and FPR (false positive rate) metrics were computed using the mask images provided by the database. A pixel-by-pixel comparison between the outcome images and the ground truth was made. The algorithm in the presented form yielded a TPR for these 20 images of 0.95356726 with a FPR of 0.00352145. The experimental results show that the proposed method performs well in extracting vessels. There are several parameters of the algorithm that have effects in the performance of the vessel segmentation method. The most significant parameter is the thresholding value. Since the proposed segmentation method obtains automatically this value for each image, it is not necessary to establish a range of thresholding values, and also, the interaction of the user to adjust this value depending on the image case is not necessary.

### **Feature extraction performance**

To adjust the parameters and evaluate the methodology of using encryption as a logarithmic spiral sampling pattern, we implemented the sequence of steps presented above and analyzed the results obtained for each image and for each value used in the implementation parameters. Table 1 shows the minimum, maximum and average number of detected points and the required computational time depending on the rotation value of the logarithmic spiral in all analyzed database.

From the results it is possible to observe the inverse relationship between the logarithmic spiral turn and the amount of encoded points, i.e., when this value is small, the resulting spiral travels very gently, so that the number of turns is greater and therefore, the number of points that can be detected increases, but also greatly increases the required time to generate the time series.

The obtained results using the complete database allow us to establish a more realistic picture about the average number of detected points for a given rotation value of the

spiral. In this way we could estimate the better rotation value to ensure that the number of detected spots for all images is considered adequate to establish the individuality of the person. As a result, the subject could be positively identified using this biometric template without sacrificing the efficiency of the system and without increasing significantly the processing time.

For example, if we choose a rotation value of 0.05 for the logarithmic spiral, the average number of coded points will be 125, and the time that the system need to encode an image would be 2.22 s.

**Table 1.** Max, min and average number of detected points and the required computational time according to the rotation value of the logarithmic spiral

Rotation value (°)	Max	Min	Mean	Time (s)
0.01	347	212	293	4.5
0.02	272	173	228	4.2
0.03	232	132	178	3.9
0.04	208	99	150	2.56
0.05	166	89	125	2.22
0.06	154	81	112	2.12
0.07	139	60	94	1.09
0.08	123	60	84	1.01
0.09	127	56	80	1.01
0.1	104	49	75	1.01
0.2	66	33	46	1.01
0.3	38	15	25	1
0.4	39	16	26	1
0.5	22	11	17	1
1	13	4	7	1
2	11	3	6	1

On the other hand, if we want to increase the number of coded points to 178, we must select as the spin value to 0.03, for which the system will take 3.9 s. This allows us to establish a range of allowed spin values, without increasing the processing times while ensuring the efficiency of the system and the individuality of the biometric code based on the logarithmic spiral.

From this analysis, we decided to select as the spiral value twist of 0.04, which guarantees to encode an average of 150 points in a time of 2.56 s.

## Conclusions

We have proposed a novel approach for the retinal image coding using logarithmic spiral and time series technologies. Using a logarithmic spiral as the rectangular region decomposition method and the SAX tool for the template representation are the keys of this new approach. They have the following advantages: first, logarithmic spiral share the geometrical characteristics that retinal network has, it is simple to implement and the overall computational complexity is very low compared to previous works (for more details consult ref [37], Table 3); second, it is very flexible as the spiral parameters, the coded local features and the SAX parameters can be adjusted according to different system requirements; third, the SAX representation (essentially symbol string) makes it very convenient for the implementation of multi-biometrics using feature fusion; fourth, logarithmic spiral and SAX representation reduce the data dimensionality of the original retinal image to a real sequential data; fifth, reduce the computational time required for the template representation and the matching step.

**Acknowledgements.** The authors of this paper wish to thank the Center for Computing Research (CIC), Mexico; Research and Postgraduate Secretary (SIP), Mexico, National Polytechnic Institute (IPN), Mexico, and CONACyT, Mexico, for their support to this work.

## References

- [1] Keogh, E., Chakrabarti, K., Pazzani, M. & Mehrotra, S. (2001). Locally Adaptive Dimensionality Reduction for Indexing Large Time Series Databases. In *proceedings of ACM SIGMOD Conference on Management of Data*. Santa Barbara, CA, May 21-24. pp 151-162.
- [2] Jung E. and Hong K., (2006). Automatic Retinal Vasculature Structure Tracing and Vascular Landmark Extraction from Human Eye Image, *Proceedings of the International Conference on Hybrid Information Technology*, IEEE Computer Society.
- [3] Hill R. B., (1992). *Retina Identification*, Portland, OR, USA.
- [4] Hill R.B., (1978). Apparatus and method for identifying individuals through their retinal vasculature patterns, U.S. Patent No. 4109237.
- [5] Simon C., and Goldstein I., (1935). A new scientific method of identification. *New York State. J. Medicine*, 35(18):901-906.
- [6] Tower P., (1955). The fundus oculi in monozygotic twins: Report of six pairs of identical twins. *Arch. Ophthalmol.*, 54:225-239.

- [7] Marshall J. and Usher D., (2006). Method for generating a unique and consistent signal pattern for identification of an individual, U.S. patent No. 6993161.
- [8] Derakhshani R. and Ross A., (2007). A Texture-Based Neural Network Classifier for Biometric Identification using Ocular Surface Vasculature, Appeared in Proc. Of International Joint Conference on Neural Networks (IJCNN), Orlando, USA.
- [9] Golden B. L., Rollin B. E., Switzer JR. R. V., (2004). Apparatus and method for creating a record using biometric information, U.S. Patent No. 028343.
- [10] Ortega M., Gonzalez M.F., (2009). Automatic system for personal authentication using the retinal vessel tree as biometric pattern, PhD. Thesis, Department of Computer Science of the Faculty of Informatics of the University of Coruña downloaded from: <http://www.varpa.es/>. (Revised on June 10, 2012.)
- [11] [http://www.absoluteastronomy.com/topics/Retinal\\_scan](http://www.absoluteastronomy.com/topics/Retinal_scan). (Revised on June 10, 2012.)
- [12] Bevilacqua V., Cambó S., Cariello L., Mastronardi G., (2007), Retinal Fundus Hybrid Analysis Based on Soft Computing Algorithms, Communications To Simai Congress, ISSN 1827-9015, Vol. 2.
- [13] Usher D., Tosa Y. and Friedman M., (2007). Ocular Biometrics: Simultaneous Capture and Analysis of the Retina and Iris, Advances in Biometrics Sensors, Algorithms and Systems, pp. 133-155.
- [14] Usher D.B., (2003). Image analysis for the screening of diabetic retinopathy, PhD thesis, University of London.
- [15] Keisuke Fukuta; Toshiaki Nakagawa; Yoshinori, Hayashi; Yuji Hatanaka; Takeshi Hara; Hiroshi Fujita, ((2008), Personal identification based on blood vessels of retinal fundus images (Proceedings Paper), Medical Imaging, Image Processing, Proceedings Vol. 6914.
- [16] Lee S. S., Rajeswari M., Ramachandram D., Shaharuddin B., (2006). Screening of Diabetic Retinopathy - Automatic Segmentation of Optic Disc in Colour Fundus Images, Proc 2nd International Conference on Distributed Frameworks for Multimedia Applications, pp. 1-7.
- [17] <http://www.isu.uu.nl/Research/Databases>. (Revised on September 20, 2012)
- [18] Choe T.E., Cohen I., Lee M. and Medioni G., (2006). Optimal Global Mosaic Generation from Retinal Images, the 18<sup>th</sup> International Conference on Pattern Recognition.
- [19] Chaudhuri S., Chatterjee S., Katz N., Nelson N. and Goldbaum M., (1989). Detection of Blood Vessels in Retinal Images Using Two-Dimensional Matched Filters, IEEE Transactions on Medical Imaging, 8(3):263–269.

- [20] <http://mathworld.wolfram.com/LogarithmicSpiral.html>. (Revised on September 20, 2012)
- [21] <http://www.2dcurves.com/spiral/spirallo.html>. (Revised on September 20, 2012).
- [22] Zhang Y. F, and Zhang Y, (2006). Another Method of Building 2D Entropy to Realize Automatic Segmentation, International Symposium on Instrumentation Science and Technology; Journal of Physics: Conference Series 48, 303–307.
- [23] Kullback, S., (1977). Information theory and statistics, Communications and Information Theory, Vol. 1, Issue 4, pp. 417 – 528.
- [24] Pal N.R. and Pal S.K., (1989). Entropic thresholding, Signal Process, 16, 97-10.
- [25] Bevilacqua V., Cambo S., Cariello L. and Mastronardi G., (2007). Retinal Fundus Hybrid Analysis Based on Soft Computing Algorithms, Communications To Simai Congress, ISSN 1827-9015, Vol. 2.
- [26] Wu C., (2007). Advanced Feature Extraction Algorithms for Automatic Fingerprint Recognition Systems, a Dissertation submitted to the Faculty of the Graduate School of State University of New York at Buffalo in Particular fulfillment of the requirements for the degree of Doctor of Philosophy.
- [27] J. Lin, E. Keogh, S. Lonardi, B. Chiu, (2003). A Symbolic Representation of Time Series, with Implications for Streaming Algorithms., Proceedings of the 8th ACM SIGMOD Workshop on Research Issues in Data Mining and Knowledge Discovery, San Diego, CA. pp. 2 -11.
- [28] Zana F., Klein J.C., (1977), Robust Segmentation of Vessels from Retinal Angiography, International Conference on Digital Signal Processing, pages 1087–1091, Santorini, Greece.
- [29] Zhoue L., Rzeszotarski M., Singerman L., Cokreff J., (1994), The detection and quantification of retinopathy using digital angiograms, IEEE Transaction on Medical Imaging: 13-4, 619-626.
- [30] Matsopoulos G. K., Mouravliansky N. A., Delibasis K. K., Nikita K .S., (1999), Automatic retinal image registration Scheme using global optimization techniques, IEEE Trans. Information technology in biomedicine: 3.
- [31] Wang L., Bhalerao A., (2003), Model Based Segmentation for Retinal Fundus Images, Proc. of Scandinavian Conference on Image Analysis (SCIA).
- [32] H. Farzin, H. Abrishami-Moghaddam, M. Moin, (2008), "A Novel Retinal Identification System", EURASIP Journal on Advances in Signal Processing, Article ID 280635, 10 pages, doi:10.1155/2008/280635.
- [33] A. Arakala, J. S. Culpepper, J. Jeffers, A. Turpin, S. Boztas, K. J. Horadam, and A. M. McKendrick, (2009), Entropy of the Retina Template, ICB '09 Proceedings of the Third International Conference on Advances in Biometrics, Pages 1250–1259, ISBN: 978-3-642-01792-6.

- [34] T. Fuhrmann, J. Hammerle-Uhl, and A. Uhl, (2009), Usefulness of Retina Codes in Biometrics, *Advances in image and video technology, Lecture Notes in Computer Science*, Volume 5414/2009, 624-632, DOI: 10.1007/978-3-540-92957-4\_54.
- [35] M. Z. Che Azemin, Dinesh K. Kumar and Hong Ren Wu, (2009), Shape Signature for Retinal Biometrics, *2009 Digital Image Computing: Techniques and Applications, DICTA'09*, pgs. 382-386.
- [36] [http://www.biometricnewsportal.com/retina\\_biometrics.asp](http://www.biometricnewsportal.com/retina_biometrics.asp).
- [37] Jiansheng Chen, Yiu-Sang Moon, Ming-Fai Wong, Guangda Su, (2010), Palmprint authentication using a symbolic representation of images, *Image and Vision Computing* 28 (2010) 343–351.
- [38] [http://en.wikipedia.org/wiki/Connected\\_space#Path\\_connectedness](http://en.wikipedia.org/wiki/Connected_space#Path_connectedness).



# Reconstruction of Gaussian Realizations by a Non-Optimal Algorithm Based on the Clipping

Y. Olvera, V. Kazakov

Department of Telecommunications, SEPI-ESIME, IPN, México, D. F.  
{yairfom, vkaz41}@hotmail.com

**Abstract.** This document analyses a non-optimal algorithm for the Sampling-Reconstruction Procedure of Gaussian realizations. The algorithm is based on the clipping. It means that it knows just the zero crossings in the realization. To finding out its effectiveness, it is compared with an optimal algorithm, which considers some samples of the realization located at strategic points. The result is that the non-optimal algorithm does not give a correct reconstruction. So it is necessary to include a new parameter within this methodology to improve the performance. However, the application has some disadvantages, mainly reflected in the reconstruction error. For all this, it is possible to conclude that the non-optimal algorithms are just approximations to the optimal algorithms. The analysis is centered in the reconstruction of Markovian Gaussian realizations.

**Keywords:** Clipping, Conditional mean rule, Gaussian Markovian realization, Non optimal algorithm, Sampling-Reconstruction Procedure.

## 1 Introduction

A fundamental problem in communication theory is to establish a statistical description that defines the reconstruction of the realizations that compose a random process through the multitude of their samples. Throughout history there have been several investigations to try to resolve it. Perhaps the most important work was done by A. Balakrishnan [1]. But in recent years one methodology has been studied extensively by lots of people, it is called *conditional mean rule* (cmr). This rule is capable of reconstructing a random realization with the minimum error possible taking into account its main statistical characteristics. It gives to each random process its own *optimal reconstruction algorithm* and optimal reconstruction error algorithm (see for example [2-5]).

Despite having an optimal reconstruction algorithm, any realization of a random process has different ways to be reconstructed. That is, it can use an alternative methodology to the right methodology according to the characteristics of each process. This parallel technique can (or not) have the same amount of statistical parameters

adfa, p. 1, 2011.  
© Springer-Verlag Berlin Heidelberg 2011

than the optimal technique uses to make the reconstruction. Although these parameters may not belong to the random process. Because of this, many times the reconstruction is not adequate. For this reason, it is called *non optimal reconstruction algorithm*. Even the conditional mean rule can be considered as non-optimal algorithm if it does not consider the appropriate parameters. The purpose is to have a simpler methodology in its development and implementation, considering the range of allowable error and the processing time.

In many cases, there is not enough information about the random process (or their realizations) to reconstruct it, specifically the value of the samples. Then, it is necessary to saturate the realization by a clipper converter. Getting a binary signal of the realization, which gives as information the moments where it crosses by zero.

The research is focus on the creation and analysis of the *Sampling-Reconstruction Procedure* (SRP) of Gaussian Markovian realizations having optimal and non-optimal algorithms. The methodologies are based on sampling methods using the zero crossings by clipping.

## 2 The Optimal Reconstruction Algorithm

The Sampling-Reconstruction Procedure using the optimal algorithm is performed by the conditional mean rule. The mean idea of this methodology has been proposed in [2]. Firstly, we consider a random process  $x(t)$  characterized by its multidimensional probability functions  $w_m[x(t_1), x(t_2), \dots, x(t_m)]$ . One realization of this process is discretized in time instants  $T = \{T_1, T_2, \dots, T_N\}$ . Therefore, we form a set of samples  $X, T = x(T_1), x(T_2), \dots, x(T_N)$ , where the number of samples  $N$  and their times of occurrence  $T$  are arbitrary. It means that the initial and central moment functions and their probability densities are modified. Now they are conditional and depend on the value of each sample  $x(T_1), x(T_2), \dots, x(T_N)$ .

In this way, the conditional mean function  $\tilde{m}(t) = \langle x(t) | X, T \rangle$  is used as reconstruction function. The quality of the reconstruction is evaluated by the conditional variance function  $\tilde{\sigma}(t) = \langle [x(t) - \tilde{m}(t)]^2 | X, T \rangle$  or reconstruction error function. Both characteristics  $\tilde{m}(t)$  and  $\tilde{\sigma}(t)$  can be found on the basis of the conditional multidimensional probability density function  $w_{N+1}(x(t) | X, T)$  of the given process. With these parameters, it is possible to reconstruct a random realization and get the quality of it. It is clear that one cannot know exactly the sampled realization, but with this rule we obtain a statistical approach for each moment of time  $t$ . The rule also provides the minimum estimation reconstruction error for random realizations with an arbitrary probability density function. For all that, the conditional mean rule is called optimal algorithm.

Considering that the realization to reconstruct is Gaussian, the conditional characteristics are [6-7]:

$$\tilde{m}(t) = m(t) + \sum_{i=1}^N \sum_{j=1}^N K(t, T_i) a_{ij} [x(T_j) - m(T_j)], \quad (1)$$

$$\bar{\sigma}^2(t) = \sigma^2(t) - \sum_{i=1}^N \sum_{j=1}^N K(t, T_i) a_{ij} K(T_j, t), \quad (2)$$

where  $m(t)$  and  $\sigma^2(t)$  are respectively the mathematical expectation and the variance of the initial process  $x(t)$ .  $K(\cdot)$  is the covariance function, and  $a_{ij}$  represents the elements of the inverse covariance matrix. Assuming that the process is stationary, we consider  $m(t) = 0$  and  $\sigma^2(t) = 1$ . This is the complete information about the process or realization given. According the properties of a Gaussian process, just the reconstruction function depends on the value of the samples; the reconstruction error function does not.

However, if we saturate the realization and apply the clipping operation, we have only the zero crossings as information. These zero crossings are considered samples. Under these conditions it is not possible to apply the conditional mean rule, because the result is zero. So we need to modify the methodology. This is accomplished by adding a sample between two zero crossings. With this proposal, we have the case where the sample is located at the midway between two zero crossings. Its magnitude is equal to the magnitude of the realization at this point.

Once we have the samples, it is possible to apply the condition mean rule for making the reconstruction and get a different result to zero. Then, the reconstruction function and the reconstruction error function are obtained by (1) and (2) respectively.

### 3 The Non Optimal Reconstruction Algorithm

The Sampling-Reconstruction Procedure using the non-optimal algorithm is also performed by the conditional mean rule. The difference consists in the samples that it uses. Although in this methodology is also used an additional sample between two zero crossing, the value of this sample is not considered part of the realization. The information we have is only the moments where the realization crosses by zero. Then, the sample used is located at the midpoint between the two zero crossings. But its magnitude is equal to the distance that exist between the two zero crossings where the samples is. It means that there is not relation with the realization. As in this case the conditional mean rule does not take into account the samples that belong to the realization, it is considered as a non-optimal algorithm.

When the zero crossings are obtained, and therefore the value of the additional samples, we can use the conditional mean rule to calculate the reconstruction function  $\hat{m}(t)$  of the non-optimal algorithm by (1). But the use of the reconstruction error function represented by (2) is not appropriate. Because of the non-optimal algorithm does not depend on the value of the samples, the equation (2) does not reflect the real reconstruction error. Besides, the error is the same that the obtained in the optimal algorithm because both algorithms have and depend on the same parameters. However, it is possible to define the quality of the reconstruction by a reconstruction total error approximate function  $\varepsilon_T^2(t)$ , which is directly related with the optimal reconstruction algorithm for knowing the differences that exist in the reconstructions of the

realization. This function is realized in two parts. The first is a special deterministic part of the reconstruction error function generated by a relation between the reconstruction function of the optimal algorithm  $\tilde{m}(t)$  and the reconstruction function of the non-optimal algorithm  $\hat{m}(t)$ , that is [8-9]:

$$\varepsilon_d^2(t) = [\tilde{m}(t) - \hat{m}(t)]^2 . \quad (3)$$

Clearly when  $\tilde{m}(t) = \hat{m}(t)$  this error is equal to zero. The second is a random part of the reconstruction error. It is obtained on the basis of the reconstruction error function of the optimal algorithm  $\tilde{\sigma}(t)$ . Therefore, the reconstruction total error approximate function  $\varepsilon_T^2(t)$  is determined by [8-9]:

$$\varepsilon_T^2(t) = \varepsilon_d^2(t) + \tilde{\sigma}^2(t) . \quad (4)$$

Now the error depends on the value of the samples as the reconstruction. So, its curves change drastically.

#### 4 Comparison between both Reconstruction Algorithms

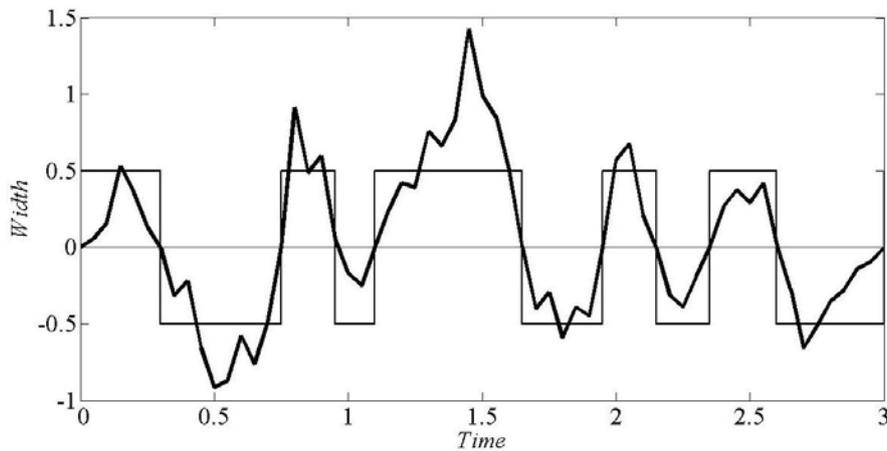
The random realization to reconstruct is obtained on the output of an one-stage RC filter when it is driven by Gaussian white noise. It means, it is a Gaussian Markovian realization. Its covariance function is:

$$K(\tau) = \exp(-\alpha|\tau|) . \quad (5)$$

Because we work with an unitary covariance time  $\tau_c$ , then  $\alpha = 1$ . As example we consider a Gaussian Markovian realization composed by 61 samples in a time of 3 seconds, this is  $N = 61$  and  $t = 3$ . The starting point is in zero, value of the mathematical expectation. The samples are separated periodically by 0.05 seconds, it means  $\Delta T = 0.05$ . The first step is to saturate the realization for knowing exactly where the zero crossings are located. In Fig. 1 are illustrated the realization to reconstruct and its saturated form.

For starting the reconstruction of the realization, both algorithms consider as samples the points where the zero crossings happen. They also include an additional sample located inside each pair of zero crossings. The optimal algorithm places the sample at the midpoint between the two zero crossings, with a magnitude that is equal to the magnitude of the realization at this point. That is, the samples are part of the realization. The non-optimal algorithm also places the sample at the midway between the

two zero crossings. But the magnitude changes, now it is equal to the distance between the two zero crossings where the sample is located. It means, the samples do not belong to the realization. This is the difference between both algorithms; the optimal algorithm considers more statistical parameters which are part of the realization to reconstruct.

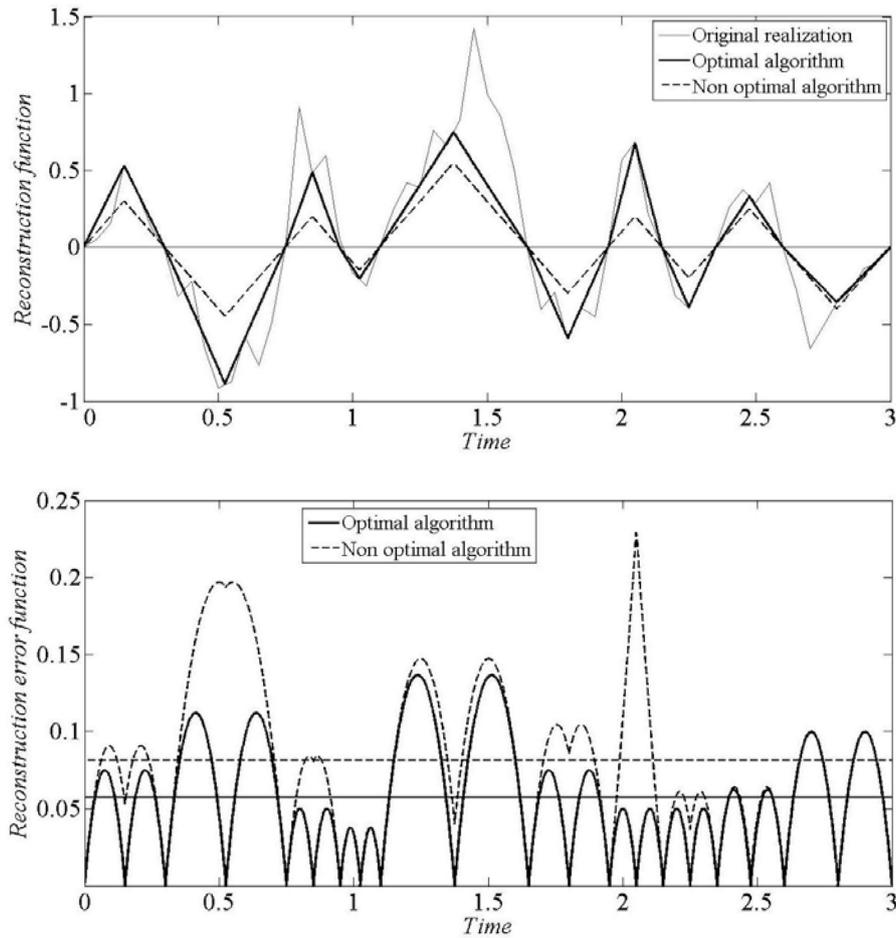


**Fig. 1.** Gaussian Markovian realization with  $N = 61$  and  $\Delta T = 0.05$ , and its saturated form.

With the samples defined, we can perform the Sampling-Reconstruction Procedure of the realization. Both algorithms are based on the conditional mean rule for making the reconstruction function by (1). The difference takes place in the reconstruction error function. The optimal algorithm continues using that rule for determining the reconstruction error by (2). But the non-optimal algorithm associates the reconstruction functions of both algorithms for obtaining the reconstruction error, which is defined by (4). The optimal case does not consider the value of the samples to get the reconstruction error, the non-optimal case does. Although a feature of the Gaussian process is that the magnitude of the reconstruction error does not depend on their samples. In Fig. 2 are showed the reconstruction functions and the error reconstruction functions of both algorithms.

It is clear that the reconstruction curve originated by the optimal algorithm is better than the reconstruction curve caused by the non-optimal algorithm. This is because the first curve covers almost all the realization. The chaotic behavior of the realization and the low number of samples used for making the reconstruction ( $N = 21$ ), prevent that the result could be better. Nevertheless, the curve in the optimal case is acceptable. Its error varies in each interval of time as a result of the different periods of separation between the samples. The maximum error is found in the middle of each interval, while the minimum error exists in the sampling points. The reconstruction curve of the non-optimal case does not cover a large part of the realization. One could even think that the behavior is another. This causes a bigger error. In some instants of time

it is very high. It is important to note that the error is not equal to zero at the sampling points which are located between the zero crossings. When the difference between the reconstruction curves grows in those intervals, the error curve of the non-optimal case is further from a similar behavior to the error curve of the optimal case.



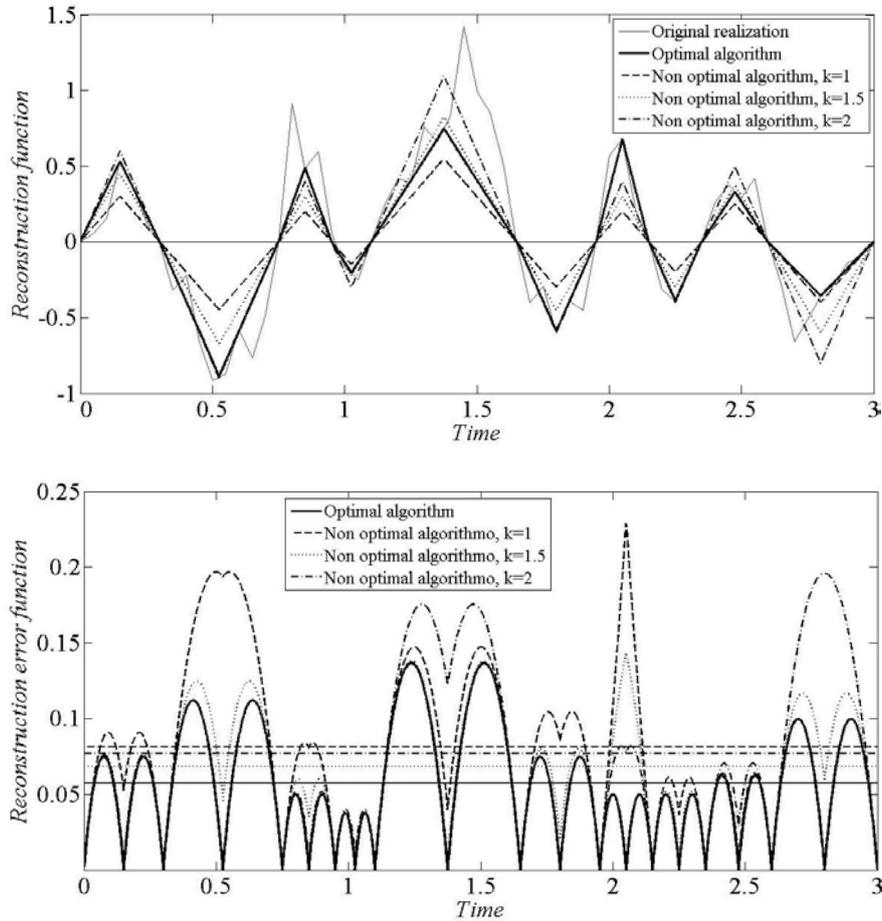
**Fig. 2.** Reconstruction and reconstruction error functions of a Gaussian Markovian realization.

Following the comparison, we include the average reconstruction error curves. They represent the area under the reconstruction error curves. They are obtained by:

$$\varepsilon_P^2(T) = \frac{1}{T} \int_0^T \rho^2(t) dt, \tag{6}$$

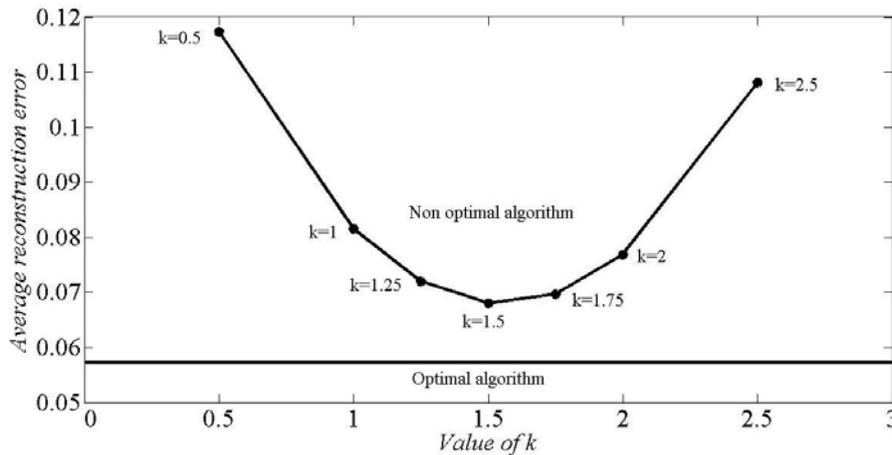
where  $\rho^2(t)$  represents the reconstruction error according to each algorithm. Obviously, the average error in the non-optimal algorithm is always bigger than the average error in the optimal algorithm.

For improving the deficient results, it is possible to introduce a new parameter represented by  $k$ . The purpose of this constant is to multiply all samples used in the reconstruction function of the non-optimal algorithm for forming a new set of samples  $X, T = k \times [x(T_1), x(T_2), \dots, x(T_N)]$ . The value of  $k$  is given according to the width of the realization, and it can change until finding the best curve that represents the behavior of the realization. With  $k$  we could obtain a width in the reconstruction curve similar to the width of the realization. In Fig. 3 are presented the reconstruction functions and the reconstruction error functions of the non-optimal algorithm with different values of  $k$ , it is also presented the functions of the optimal algorithm.



**Fig. 3.** Reconstruction and reconstruction error functions of a Gaussian Markovian realization with different values of  $k$ .

If we include the parameter  $k$ , the curves of the non optimal algorithm significantly improve. For example, when  $k = 1.5$ , the reconstruction curve maintains a close behavior to the realization. The new curve reaches to cover the realization much of the time. In many instants it is a curve very similar to the curve of the optimal algorithm. Hence, the reconstruction error is lower than when  $k = 1$  in almost all the time. The same thing happens when we consider  $k = 2$ . Although with this value, the reconstruction error is higher in the time intervals where the separation between the samples is big. This is because the magnitude of the samples increases considerably. In both cases the results are better than when we do not use the parameter  $k$ . This is reflected in the average reconstruction error curves. In Fig. 4 there is a graph that indicates how the average reconstruction error is if we consider different values of  $k$ . It is also showed the relation with the optimal average reconstruction error.



**Fig. 4.** Relation between the value of  $k$  and the magnitude of the average reconstruction error in the non-optimal algorithm.

One notes that for this example, the smallest reconstruction error is obtained when  $k = 1.5$ . But this may vary from one realization to another. Mainly because the characteristics change, such as: the number of zero crossings and the separation between them, the amplitude of the realization, the number of samples, and more. These cause that the graph of Fig. 4 also changes. Nevertheless after several experiments with many Gaussian realizations, we can conclude that using a value of  $k$  between 1.4 and 1.8, we get a similar reconstruction to the realization and therefore a smaller reconstruction error.

It is important to mention that the sampling procedure in the optimal algorithm is carried out at the transmitter. It means that on the output of the transmitter are sent three samples (two samples located at the zero crossings and one sample located at the midpoint between these zero crossings). But in the non-optimal algorithm the sampling procedure is performed at the receiver. Due to the transmitter gives only the

location of the zero crossings as information, then the receiver is responsible for estimating the magnitude of the samples.

The figures demonstrate the difference between the reconstruction algorithms under analysis. The precise estimates of the reconstruction error function in the optimal algorithm give a family of curves that depends on several specific parameters of the realization, as the value of the samples. While the non-optimal algorithm gives curves that do not clearly describe the reality of the realization, because it considers samples that do not have relation with the samples of the realization. This could have the objective of making an analysis and procedure simpler and faster, in order to save time and space. But as a consequence, the reconstruction error is higher. In this way, we need to introduce a new parameter for improving the reconstruction. Obviously, the magnitude of the reconstruction error in the non-optimal algorithm is always higher in any case. This is because it is composed by an alternative methodology, and it is an approximation only. So it is a natural effect.

## 5 Conclusions

Two different reconstruction algorithms are analyzed to describe the Sampling-Reconstruction Procedure of Gaussian Markovian realizations. Both principal characteristics, reconstruction function and reconstruction error function, are obtained. The algorithms are based on techniques that use the sampling zero crossings by clipping. This is performed by saturating the realization through a clipper converter. Getting a binary signal of the realization, which gives as information the moments when the realization crosses by zero.

To make the reconstruction, the optimal algorithm adds an additional sample at the midpoint between each pair of zero crossings, with a magnitude equal to the magnitude of the realization at this point. The non-optimal algorithm also put an extra sample at the midpoint between two zero crossings, but now the magnitude is equal to the distance that exists between these zero crossings. The results show that the curves obtained for the optimal algorithm are widely better in all the time. This is reflected in a big difference between the reconstruction errors.

However, if we introduce a new parameter  $k$  that multiplies all samples of the reconstruction function in the non-optimal algorithm, we can get better curves that represent the behavior of the realization. Considering various realization, we conclude that the best results are obtained when  $k$  has a value between 1.4 and 1.8.

Clearly, the methodology that uses the greatest number of statistical parameters of the random realization, gives the correct reconstruction. This does not mean that the non-optimal algorithms are incorrect. Simply, they must be declared as special cases, which can be used depending on the application.

## References

1. A. V. Balakrishnan. "A note on the sampling principle for continuous signals", *IRE Trans. On Inf Theory*, Vol. IT-3, 1957 pp.143-146.

2. V. A. Kazakov. "Regeneration of samples of random processes following non linear inertialess conversions", *Telecommunication and Radio engineering*, Vol. 43, No. 10, 1988, pp. 94-96.
3. V. A. Kazakov, M. A. Beliaev. "Sampling – Reconstruction Procedure for non Stationary Gaussian Processes Based on Conditional Expectation Rule", *Sampling Theory in Signal and Image Processing*, vol. 1, No. 2, May 2002, pp. 135-153.
4. V. Kazakov. "The Sampling – Reconstruction Procedure with a Limited Number of Samples of Stochastic Processes and Fields on The Basis of the Conditional Mean Rule", *Electromagnetic Waves and Electronic Systems*, Vol. 10, No. 1-2, 2005, pp. 98-116.
5. V. Kazakov. "Non Uniform Sampling and Reconstruction of Stochastic Processes with the Limited Number of Samples", *WSEAS Transactions on Systems*, Issue 8, Vol. 5, August 2006, pp. 1777-1784.
6. P. E. Pfeiffer. "*Probability for applications*", Springer Verlag, 1990.
7. R. L. Stratonovich. "*Topics in the Theory of Random Noise*", New York: Gordon and Breach, 1963.
8. V. Kazakov, Y. Olvera. "Optimal and Suboptimal Reconstruction of Non Gaussian Realizations on the Output of Non Linear Converters", *Recent Researches in Telecommunications, Informatics, Electronics and Signal Processing*, Vol. 2011, May 2011, pp. 59-64.
9. V. Kazakov, Y. Olvera. "Sampling – Reconstruction Procedures of Non Gaussian Processes by Two Algorithms", *International Journal of Circuits, Systems and Signal Processing*, Issue 5, Vol. 5, USA, July 2011, pp. 557-564.

# Segmentación del disco óptico en imágenes de retina mediante la transformada de Hough y los contornos activos

José Abel de la Fuente-Arriaga<sup>1</sup>, Edgardo Felipe-Riverón<sup>2</sup>, Eduardo Garduño-Calderón<sup>3</sup>

<sup>1</sup> Tecnológico de Estudios Superiores de Jocotitlán, Carretera Toluca-Atlaconulco Km. 44.8, Ejido de San Juan y San Agustín, Jocotitlán, C.P. 50700, Estado de México, México.

<sup>2</sup> Centro de Investigación en Computación, Instituto Politécnico Nacional, Juan de Dios Bátiz s/n, Col. Nueva Industrial Vallejo, C. P. 07738, México.

<sup>3</sup> Centro Oftalmológico, Libramiento Jorge Jiménez Cantú 1208, Atlaconulco, C. P. 50450, Estado de México, México

<sup>1</sup>abeldlfa@gmail.com, <sup>2</sup>edgardo@cic.ipn.mx, <sup>3</sup>edugarcal@yahoo.com

**Abstract:** This paper presents a new method for automatic segmentation of the optic disc in images of human retinas. The method consists of multiple steps: finding the region of interest (ROI), achieved with a combination of the Fourier transform (FT) and the method of thresholding *P-tile*, followed by removing of the venules and arterioles, and finally the optical disk segmentation by means of morphological operations and the combination of the circular Hough transform and active contours. By combining these two last steps yields a robust method for images with low contrast, high levels of noise and discontinuities. The method was successful with 74 images from a total of 80 images achieving 92.5% accuracy.

**Keywords:** Image segmentation, optical disc, Hough transform, active contours, *P-tile* thresholding.

**Resumen.** En este trabajo se presenta un nuevo método para la segmentación automática del disco óptico en imágenes de retinas humanas. El método consta de varios pasos: la búsqueda de la región de interés (ROI), logrado con la combinación de la transformada de Fourier (TF) y el método de umbralado *P-tile*, seguido por la eliminación de las vénulas y arteriolas y finalmente la segmentación del disco óptico mediante operaciones morfológicas y la combinación de la transformada de Hough circular y los contornos activos. Al combinar estos dos últimos pasos se obtiene un método robusto ante imágenes con bajo contraste, altos niveles de ruido y discontinuidades. El método tuvo éxito con 74 imágenes de un total de 80 imágenes lográndose el 92.5% de precisión.

**Palabras clave:** Segmentación de imágenes, segmentación del disco óptico, transformada de Hough, contornos activos, umbralado *P-tile*.

## 1 Introducción

Este trabajo presenta una nueva técnica para la localización automática del disco óptico en imágenes de retinas humanas.

El análisis de imágenes de retinas humanas es una de las técnicas médicas no invasivas más utilizadas en el análisis del sistema visual humano [1]. La información que puede obtenerse acerca de éstas puede ser usada para valorar la severidad de ciertas enfermedades oculares. Las enfermedades oculares más importantes y peligrosas por sus características asintomáticas y por el hecho de que si no son tratadas a tiempo conducen a la pérdida progresiva de la visión y la ceguera irreversible, son el Glaucoma y la Retinopatía Diabética.

Tomando en cuenta que en estas enfermedades el Glaucoma es responsable de un 12% a 15% [2], [4] y la Retinopatía Diabética un 5% [3] del total de la población mundial con ceguera, y que en México representan la segunda y tercera causa de ceguera irreversible [5], es de vital importancia que se trabaje en la búsqueda de métodos automáticos y exactos para el reconocimiento y valoración del disco óptico para la detección de estas enfermedades. Una de las ventajas de este trabajo, es que además de servir para la localización del disco óptico, también sirve para la identificación precisa y exacta de su contorno.

El algoritmo que se propone fue implementado en Matlab 2008, versión 7.6.0.324 y probado en 80 imágenes de retina de pacientes reales, con un tamaño de 720x576 píxeles, las cuales fueron tomadas con una cámara de fondo de ojo convencional [6], [7]. Las mismas forman parte de la colección del Centro de Investigación en Computación del Instituto Politécnico Nacional y de la clínica oftalmológica del Dr. Eduardo Garduño Calderón.

El algoritmo fue diseñado a partir de los mejores resultados, de algunos de los trabajos existentes de identificación automática del disco óptico [11], [16], [17], [18], [19], [20], [21], los cuales utilizan la transformada de Hough [8] y los contornos activos [9], lográndose un algoritmo novedoso y robusto.

## 2 Estado del arte

En el análisis de la retina existen diferentes procedimientos para analizar específicamente el disco óptico. La naturaleza de estos cobra un amplio espectro de técnicas y métodos en diferentes áreas.

Particularmente en la oftalmología los procedimientos médicos para diagnosticar algunas enfermedades son la identificación directa o indirecta del disco óptico partiendo de fotografías ópticas de retinas humanas [10].

Aunado a esto se comenzaron a desarrollar algoritmos de localización del centro del disco óptico, por medio del cálculo del pixel central denominado semilla [11], el cual después de haber sido identificado, es usado para la determinación de una subimagen o región de interés (ROI) que contiene el disco óptico aislado.

Algunos otros trabajos con el mismo objetivo son: [12] donde se hace uso de la varianza de los píxeles adyacentes al punto de análisis, para encontrar la semilla; este método dio resultados de un 99% de precisión. [13] que hace uso del método de

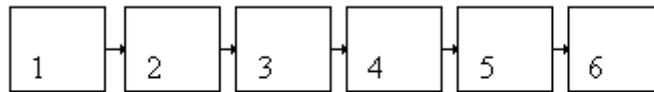
umbral de *P-tile* y la combinación de los canales de color, y [11], donde fueron usadas 3 semillas simultáneas, implementando los métodos de varianza, máximo contraste y transformada de Fourier (TF).

Otros autores enfocaron su estudio en la detección exacta del contorno del disco mediante la identificación del contorno por métodos morfológicos y la transformada de cuencas (*watershed*), con un resultado de 90% de precisión ([14]), y finalmente [15] que identificó el disco óptico haciendo uso de operaciones morfológicas y umbrales que dependen de la desviación estándar de la imagen.

También se puede apreciar un uso considerable de la transformada de Hough [11], [16], [17], [18] con resultados que van de 72% al 97.5% de precisión, y más recientemente el uso de contornos activos [19], [20], [21], [22] con precisiones entre 87% y 90.32%.

### 3 Método propuesto

El objetivo del método propuesto es la segmentación automática del disco óptico, aun cuando se aplique sobre imágenes de baja calidad (ruidosas y con bajo contraste). La Fig. 1 muestra el diagrama de bloques con la secuencia de operaciones para la segmentación del disco óptico con esta metodología.



- 1.- Imagen RGB
- 2.- Normalización del tamaño
- 3.- Localización de la región de interés (ROI)
- 4.- Eliminación de los vasos sanguíneos dentro del disco óptico
- 5.- Transformada de Hough y los contornos activos
- 6.- Segmentación del contorno del disco óptico

**Fig. 1.** Diagrama de bloques con la secuencia para la segmentación del disco óptico

#### 3.1 Imagen RGB

Para la implementación del algoritmo propuesto, se hace uso de imágenes de retinas humanas en colores en el modelo de color RGB, con un tamaño de 720x576 píxeles y formato gráfico JPG.

Una imagen que pertenece a esta colección se puede apreciar en la Fig. 2. No obstante, a pesar de que las pruebas del trabajo se realizan bajo imágenes de tamaño similar, el método tiene la capacidad de adaptarse al uso de imágenes con diferentes tamaños.

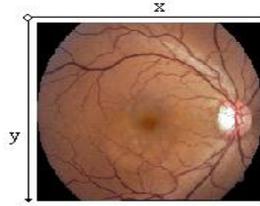


Fig. 2. Imagen de fondo de ojo en colores de la retina humana en el modelo de color RGB

### 3.2 Normalización del tamaño

El objetivo principal del proceso de normalización es el de garantizar el uso de las imágenes de retina aunque procedan de distintas fuentes; Sin embargo, también es útil para reducir el tiempo de procesamiento del algoritmo sin comprometer su precisión.

El método permite reducir proporcionalmente el tamaño de la imagen sin afectar su relación de aspecto original, mediante el establecimiento de una relación de proporcionalidad con cualquier tamaño de imagen, ya que la reducción (o aumento) se hace en el eje Y en proporción al tamaño asignado al eje X. Este proceso se realiza con una simple regla de tres valores conocidos y una incógnita, donde se establece una relación de interpolación lineal. La fórmula que lleva a cabo este proceso es la siguiente:

$$y' = x' * (y/x)$$

- y' Nuevo tamaño del eje y
- x' Nuevo tamaño del eje x
- y Tamaño actual del eje y
- x Tamaño actual del eje x

La fórmula está formada por 4 variables, donde  $x$  y  $y$  representan el tamaño de los lados de la imagen original y  $x'$ ,  $y'$  son los nuevos tamaños de los lados de la imagen normalizada.

En este trabajo se asigna un valor constante de 600 píxeles para la variable  $x'$ , dando como resultado que la imagen de 720x576 píxeles dé una imagen normalizada de tamaño 600x480 píxeles. La Fig. 3 muestra en A) la imagen de tamaño original y en B) la imagen reducida después de la normalización.

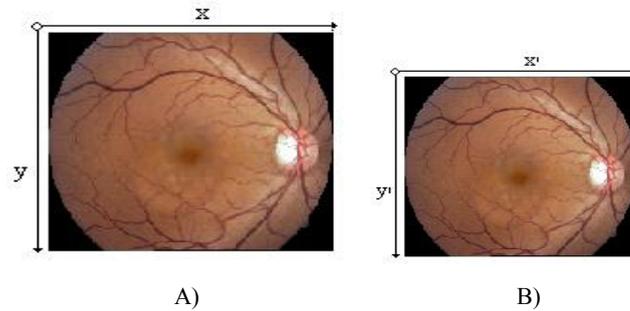


Fig. 3. Normalización del tamaño: A) Imagen de tamaño original; B) Imagen normalizada

### 3.3 Localización de la región de interés (ROI)

El objetivo de esta etapa es el de identificar en las imágenes de retina el centro del disco óptico para crear la subimagen que constituirá la región de interés (ROI) donde se encontrará el disco óptico aislado. El proceso completo para obtener la ROI se muestra en la Fig. 4.

El método para la detección del centro del disco óptico es aplicado sobre el canal verde del modelo de color RGB. Este canal fue seleccionado ya que es el que mejor contraste y más detalles brinda [11], tal como se puede observar en A) de la Fig. 4.

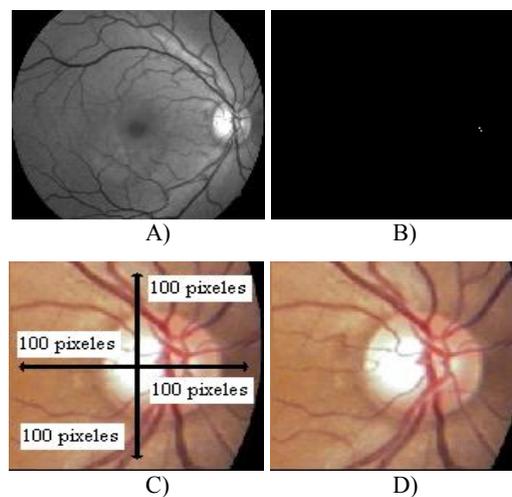


Fig. 4. Proceso para obtener la región de interés (ROI) del disco óptico: A) Imagen monocromática del canal verde del modelo RGB; B) búsqueda de las semillas y localización del pixel central del disco óptico; C) número de píxeles para establecer el tamaño de la ROI con base en el pixel central del disco; D) ROI del disco óptico aislado en el modelo RGB

El siguiente paso es la detección del pixel de referencia dentro del disco óptico. Para su determinación precisa, es necesario primero la detección de dos puntos individuales que denominaremos semillas con dos métodos diferentes, el método de la

Transformada de Fourier (TF) [11] y el método de umbralado *P-tile* [21]. Al combinar estos dos métodos se logra mejorar la precisión en la localización del pixel de referencia y brindar cierta tolerancia para evitar un resultado incorrecto por parte de alguno de los dos métodos.

En general, los pixeles del disco óptico son los que presentan un nivel de gris mayor en la imagen. No obstante, el pixel de mayor nivel de gris en la imagen completa puede ser que no pertenezca al disco óptico, pues otros puntos aislados pueden cumplir eventualmente con esta propiedad de máximo nivel de gris. Mediante el método de la TF se obtiene una nueva imagen en el dominio de las frecuencias; a la misma se le aplica un filtro pasa bajas y la transformada inversa de Fourier, donde el mayor nivel de gris en esta imagen sería la primera semilla que buscamos. Como filtro pasa bajas, en este trabajo se aplica un filtro de mediana con un *kernel* de tamaño 9x9.

El método de umbralado *P-tile* es una técnica de umbralado que utiliza el área bajo el histograma que ocupan los objetos que se quieren detectar; mediante esta partición del histograma, el umbral puede ser elegido asignando un porcentaje  $P$  de pixeles a los objetos. Por tanto,  $P$  es un valor que indica el porcentaje de imagen que ocupa el objeto u objetos.

En este trabajo se le asigna a  $P$  un valor de 99.99% para obtener un umbral que al ser aplicado a la imagen, da como resultado un conjunto de puntos en torno al disco. Con las coordenadas de estos puntos individuales y a su posición en la imagen, se calcula el punto central (centroide), que es la segunda semilla que buscamos.

Después de obtener los dos pixeles semilla por los métodos de la TF y *P-tile* en B) de la Fig. 4, para decidir cuál será el pixel central de referencia del disco, calculamos la distancia euclidiana entre las dos semillas y aplicamos las siguientes reglas [11]:

1.- Si la distancia entre las semillas es menor que 160 pixeles (diámetro máximo estimado para el tamaño del disco en este trabajo), la localización del pixel central de referencia del disco es la media de la posición de las dos semillas.

2.- Si la distancia entre las dos semillas es mayor que 160 pixeles, la localización del pixel de referencia del disco óptico que constituirá la semilla buscada, será la obtenida mediante la transformada de Fourier.

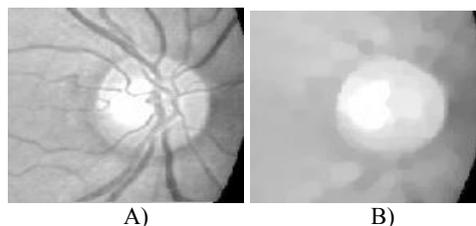
Con base en las reglas anteriores y después de obtener el pixel central de referencia del disco, para definir la ROI (que será una subimagen cuadrada dependiente del pixel hallado dentro del disco) asignamos 100 pixeles en las cuatro direcciones (arriba, derecha, abajo, izquierda) en los tres canales de color RGB, como se puede apreciar en los incisos C) y D) de la Fig. 4.

### 3.4 Eliminación de los vasos sanguíneos dentro del disco óptico

La eliminación de los vasos sanguíneos dentro de la ROI está basado en la secuencia de las operaciones morfológicas de clausura y apertura sobre el canal rojo del modelo RGB.

La operación clausura llena los vasos sanguíneos en el disco y la copa y suaviza los bordes exteriores, mientras que la apertura remueve las pequeñas manchas y picos claros presentes en la imagen [12], [14], [15], [22].

En la Fig. 5 se puede apreciar el procedimiento expuesto sobre el disco óptico. En A) se muestra el disco antes de la eliminación de los vasos y en B) se muestra el disco después de la eliminación.



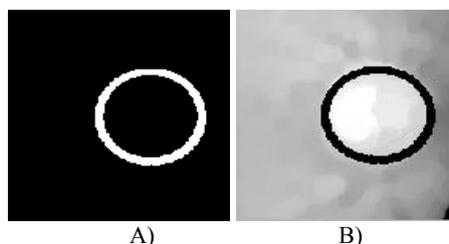
**Fig. 5.** Eliminación de los vasos sanguíneos dentro del disco óptico: A) Disco óptico antes de la eliminación de los vasos; B) Disco óptico después de la eliminación de los vasos

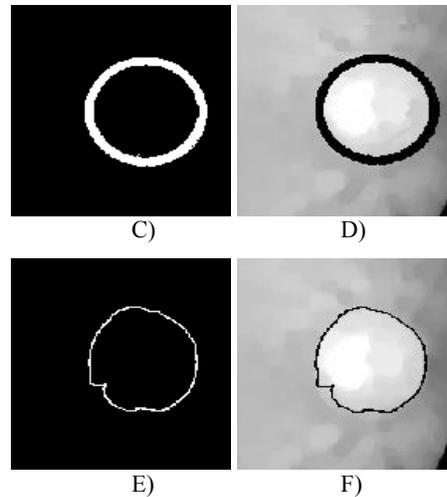
### 3.5 Transformada de Hough y los contornos activos

La transformada de Hough, es un algoritmo empleado en reconocimiento de patrones que permite encontrar ciertas formas dentro de una imagen, tales como líneas, círculos, etc. Explícitamente la transformada de Hough circular sirve para la detección de los círculos los cuales pueden ser descritos mediante su centro y su radio. En este trabajo se usa el detector de Prewitt para la detección de los bordes, seguido por la transformada de Hough circular sobre el canal rojo de la imagen en el modelo de color RGB.

Para evitar circunferencias espurias obtenidas con la transformada de Hough producto de otros objetos circulares en la imagen como puede ser el marco alrededor de la misma imagen de retina, la conjunción de exudados, etc., el tamaño de radios aceptados es de 10 a 80 píxeles; al multiplicar por dos el valor máximo posible del radio de 80 píxeles, nos da el diámetro de la circunferencia de 160 píxeles estimados como diámetro máximo del disco.

Otros trabajos que utilizan la transformada de Hough circular arrojan resultados con precisiones entre 72% y 97.5% [11], [18], lo que demuestra que es un método eficiente y robusto para la detección del disco óptico en imágenes de baja calidad, con ruido y discontinuidades en los bordes. De forma similar, los algoritmos de contornos activos tiene resultados que van de 87% a 90.32% de precisión, demostrado en los trabajos [9] y [22]. La mayor ventaja de los contornos activos es su habilidad para puentear las discontinuidades independientemente del ruido presente en las imágenes. La combinación de los efectos de estos dos algoritmos se aprecia en la Fig. 6.





**Fig. 6.** Combinación de la transformada de Hough y los contornos activos. A) Circunferencia de Hough; B) circunferencia del inciso A) sobre el disco óptico sin vasos; C) Circunferencia de Hough con el radio aumentado en 5 píxeles; D) Circunferencia del inciso C) sobre el disco óptico sin vasos; E) Contorno del disco óptico obtenido de la combinación del inciso C) y los contornos activos; F) Contorno resultado del inciso E) sobre el disco óptico sin vasos

Cómo primer paso para la combinación de estos dos algoritmos, se usa la transformada de Hough circular para obtener la circunferencia aproximada al contorno del disco óptico (Ver A de la Fig. 6).

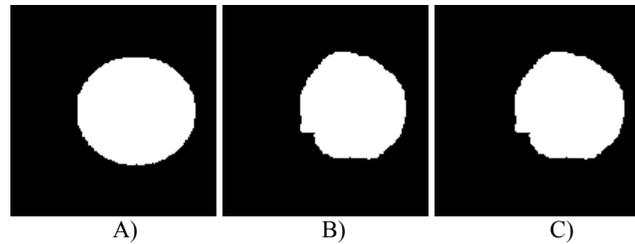
Sin embargo, aun cuando la circunferencia es muy aproximada al contorno del disco óptico, como se aprecia en B de la Fig. 6, éste sobresale ligeramente de la circunferencia en sus partes superior e inferior.

Por esta razón se modificó el resultado de la circunferencia obtenida de aplicar la transformada de Hough aumentando el radio de la misma en 5 píxeles como se aprecia en C de la Fig. 6. Este valor puede variar de acuerdo a la colección de imágenes que han de ser analizadas. Como resultado de este incremento en el radio, se puede apreciar que ahora el disco calza perfectamente dentro de la circunferencia, (Ver D en la Fig. 6).

Esto da la pauta para la implementación del método de contornos activos, usando como parámetro de máscara de inicio la circunferencia de Hough modificada. Los resultados de esta combinación se pueden apreciar en E y F de la Fig. 6.

No obstante, aun contando con la unión de estos métodos, el algoritmo de contornos activos puede fallar y cambiar la polaridad de la segmentación. Esto quiere decir que en lugar de ir cerrando paulatinamente el contorno claro y detectar el disco, puede abrirse y salirse del margen de la circunferencia resultado de la transformada de Hough modificada. Para esto se incorporó un pequeño agregado que permite asegurar el buen funcionamiento del algoritmo.

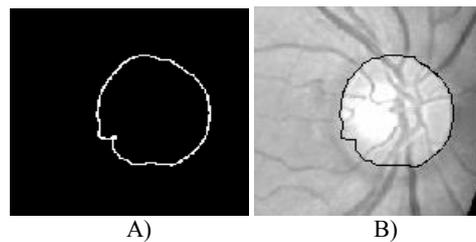
El proceso consiste en calcular la intercepción de la circunferencia de Hough modificada, que se muestra en A de la Fig. 7, y la región obtenida de los contornos activos que se muestra en B de la Fig. 7. El resultado  $A \cap B$  se muestra en el inciso C de la Fig. 7.



**Fig. 7.** Proceso de intersección entre la transformada de Hough modificada y los contornos activos: A) Circunferencia de Hough de radio modificado; B) Resultado de la aplicación de los contornos activos; C) Intersección de las imágenes A y B

### 3.6 Segmentación del contorno del disco óptico

Después de lograr la región exacta que ocupa el disco óptico, como se aprecia en C de la Fig. 7, se calcula su contorno mediante la diferencia de la dilatación y la erosión morfológicas al disco óptico obtenido. Este contorno se muestra en A de la Fig. 8, el cual se ve superpuesto sobre el disco óptico real en B.



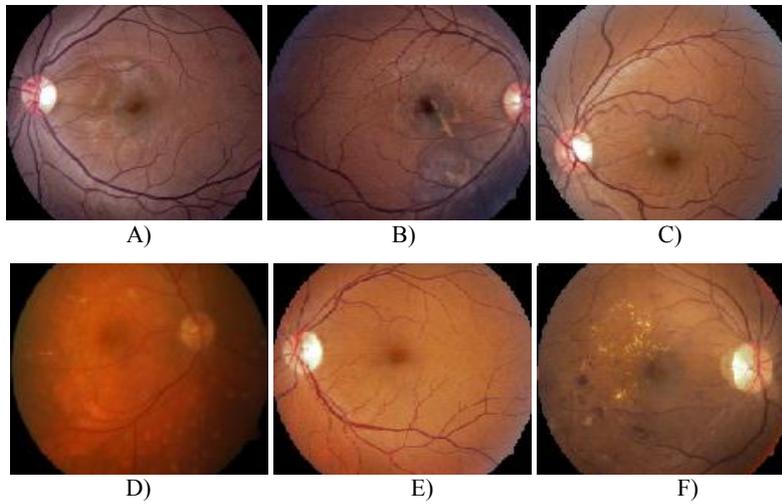
**Fig. 8.** Contorno del disco óptico obtenido como resultado del método propuesto: A) Contorno resultante de la implementación del método propuesto; B) Resultado de sobreponer el contorno mostrado en A sobre el disco óptico real

## 4 Discusión de los resultados

La Fig. 9 muestra algunas imágenes de la colección de 80 imágenes de retinas humanas que fueron tratadas bajo el algoritmo propuesto.

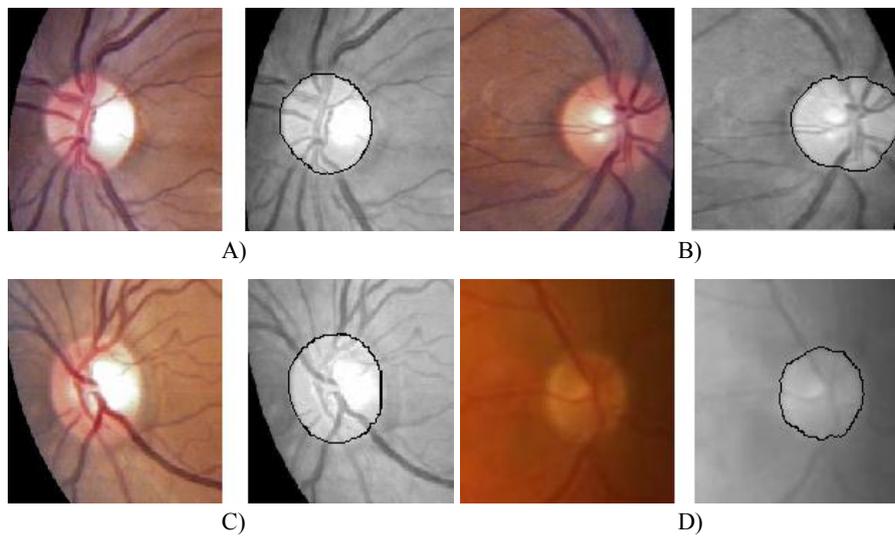
La población de imágenes para la discusión de los resultados fue conformada en función de imágenes con distintas calidades, ya que se pretendió probar la robustez del método ante imágenes con ruido aditivo, y bajo contraste.

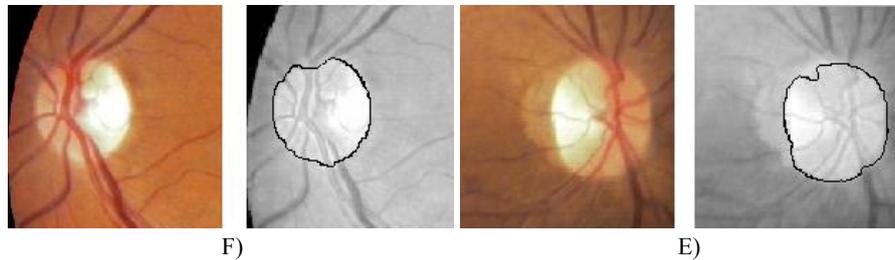
Los factores tomados en cuenta, para conformar la población de imágenes de prueba fueron: imágenes con iluminación homogénea (C y E); imágenes con una iluminación pobre o heterogénea (A y B), e imagen con poca iluminación, borrosa, y con manchas provenientes del lente de la cámara (D), e imagen con exudados en la retina y junto al borde del disco óptico (F).



**Fig. 9.** Colección de algunas de las imágenes usadas para la prueba del algoritmo propuesto: Con iluminación pobre o heterogénea (A, B); Imágenes con iluminación homogénea (C, E); Imagen con poca iluminación, borrosa y con manchas provenientes del lente de la cámara (D), e imagen con exudados en la retina y junto al borde del disco óptico (F)

En la Fig. 10 se muestran en detalle la región de interés ROI del disco óptico de las imágenes de la Fig. 9 y el contorno resultante al aplicar el algoritmo propuesto. En ella podrá apreciarse que en todos los casos se obtuvo un contorno del disco óptico preciso, aun en imágenes con ruidos y discontinuidades.





**Fig. 10.** Detalle de la región de interés ROI del disco óptico y el contorno resultante al aplicar el algoritmo propuesto a las imágenes de retina mostradas en A, B, C, D, E y F de la Fig. 9

En la Tabla 1 se aprecian los resultados de la segmentación del contorno del disco óptico usando la transformada de Hough circular, en función de la precisión y del número de imágenes de muestra utilizadas. Los resultados van desde un 72% a un 97.5% de precisión

Por otra parte, en la Tabla 2 se observan los resultados de la segmentación del contorno del disco óptico mediante los contornos activos, en función de la precisión y el número de imágenes de muestra usadas. Los resultados van de un 87% a un 90.32% de precisión.

**Tabla 1.** Trabajos y resultados de la segmentación del contorno del disco óptico, con transformada de Hough circular, en función de la precisión y el número de imágenes de muestra utilizadas

<b>Autores</b>	<b>Publicación</b>	<b>No. de muestras</b>	<b>Precisión</b>
R.A. Abdel Ghafar et al.	UMIST, 2004 [17]	90	72%
Arturo Aquino Martin et al.	Universidad de Huelva, 2007 [11]	120	97.5%
S. Sekhar et al.	EUSIPCO, 2008 [18]	60	90.85%
Jorge Ruda C et al.	Revista Avances en Sistemas e Informática, 2011 [16]	89	73%

**Tabla 2.** Trabajos y resultados de la segmentación del contorno del disco óptico, en función a la precisión y el número de imágenes de muestra usadas

<b>Autor</b>	<b>Publicación</b>	<b>No. de muestras</b>	<b>Precisión</b>
Alireza Osareh et al.	Universidad de Bristol, 2001 [19]	75	90.32%
Chisako Muramatsu et al.	Computer Methods and Programs in Biomedicine, 2011 [21]	128	87.5%
Chisako Muramatsu et al.	Journal of Biomedical Optics, 2011 [20]	80	87%

En la prueba del algoritmo propuesto fueron analizadas 80 imágenes de retina, obteniendo un resultado global de 92.5% de precisión, con 74 contornos correctos y 6 incorrectos. Con el 100% de las imágenes fue detectado correctamente el disco óptico. De acuerdo a lo indicado en las Tablas 1 y 2, se aprecia que la precisión del método que se propone está por encima en un 2.18% (92.50 – 90.32) de los mejores

resultados obtenidos por el método de contornos activos, y por debajo en un 5% (97.5 – 92.5) de los obtenidos con la transformada de Hough circular.

Si se analiza detenidamente, muchos de los trabajos donde se aplicó la transformada de Hough en la Tabla 1, estuvieron en función de la localización indirecta del disco donde no se obtiene su contorno detallado. Sobre esta base, podemos decir que el método presentado es eficiente en los objetivos que se propuso.

Finalmente, es oportuno exponer que dentro de las 80 imágenes analizadas, incluyendo aquéllas donde se obtuvieron resultados incorrectos, eran de muy mala calidad. Esto corrobora que el método propuesto para detectar el disco óptico y segmentar su contorno, es robusto ante el manejo de imágenes con bajos contrastes, altos niveles de ruido y discontinuidades.

## 5 Conclusiones

En este trabajo se ha presentado el algoritmo y los resultados del desarrollo de un método novedoso para la detección y la segmentación automática del contorno del disco óptico en imágenes de retinas humanas ópticas reales en colores. Este método consiste en la combinación de los algoritmos de la transformada de Hough circular y los contornos activos. Al combinar estos dos métodos, el algoritmo adquiere las características de cada uno, como es la eficiencia y robustez ante bajas calidades de las imágenes con bajos contrastes, alta inmunidad al ruido y la eventual presencia de discontinuidades en los bordes de los objetos.

Con la aplicación de la metodología propuesta se localizó el contorno del disco óptico en el 92.5% de los casos, aun con imágenes de retina de mala calidad. Este valor está entre los índices actuales logrados internacionalmente. Para garantizar mejores resultados e incrementar la eficacia del método, se sugiere capturar las imágenes de retina con alto contraste, al menos en la zona del disco óptico.

**Agradecimientos.** Los autores de este trabajo agradecen al Centro de Investigación en Computación (CIC), México, del Instituto Politécnico Nacional (IPN), México, por su apoyo en la realización y presentación de este trabajo.

## Referencias

1. Manzanaro, G.: Análisis digital de la papila con cámara de fondo de ojo convencional: estudio de variabilidad. *Arch. Soc. Esp. Oftalmol. de Madrid*, v. 79, n. 3, marzo (2004)
2. Elolia, R.; Stokes, J.: Monograph Series on Aging-related Diseases: XI. Glaucoma. *Chronic Diseases in Canada*, Vol. 19, No. 4 – (1998)
3. Barriavon-Bischhoffshausen, F.; Martínez-Castro, F.: Guía práctica clínica de retinopatía diabética para Latinoamérica, *Programa Visión 2020/LAPB Latinoamérica*, Asociación Panamericana de Oftalmología APAO, Christian Blind Mission. CBM, (2010)

4. Quigley, Ha.: Number of people with glaucoma worldwide. *Br J Ophthalmol*, 80, 389-393, (1996)
5. Felipe-Riverón, E.M.; Morales-Guitron, S.L.; Ortiz-Yáñez, S.: Detección automática de la normalidad de las retinas humanas. *CIC-IPN*, México, (2004)
6. Manzanaro, G.: Fotografía de fondo de ojo con filtros. *Boletín de a Soc. Oftalmol. de Madrid*, N.º 44, (2004)
7. Jackman, Wt.; Webster, J.D: On photographing the retina of the living human eye. *Philadelphia Photographer*, 23: 275, (1886)
8. C. Gonzalez, R.; Woods, E. R.: *Digital Image Processing*. Addison-Wesley, ISBN-10: 0201180758, (2002)
9. Kass, M.; Witkin, A.; Terzopoulos, D.: Snakes Active contour models. *Int. J. Computer vision*, vol. 1, pp. 321-331, (1988)
10. De la Fuente-Arriaga, J.A.; Garduño-Calderón, E., Cuevas-de la Rosa, F.J.: Estado actual de las técnicas computacionales para la valoración del disco óptico en glaucoma. *Memorias del Congreso Mexiquense CTS+I*, 1ra Edición 2011, ISBN: 978-607-707-382-6, pp. 155-161, México diciembre (2011)
11. Aquino-Martín, A., Gegúndez-Arias, M.E.; Andujar-Márqueza, J.M: Sistema experto para la detección precoz de la retinopatía diabética mediante análisis de imágenes digitales de retina: primeros resultados. *Universidad de Huelva*, (2007)
12. Sinthanayothin, C.; Boyce, J.F; Cook, H.F; Wiliamson, T.H: Automated localization of the optic disc, fovea, and retinal blood vessels from digital colour fundus images. *Br J Ophthalmol* 83:902-910, doi: 10.1136/bjo.83.8.902, (1999)
13. Hatanaka, Y.; Noudo, A.; Sawada, A.; Hara, T.; Yamamoto, T.; Fujita, H.: Automated Measurement of Cup to Disc Ratio Based on Line Profile Analysis in Retinal Images. *33<sup>rd</sup> Annual International Conference of the IEEE EMBS Boston*, Massachusetts USA, August 30 - September 3, (2011)
14. Thomas, W.; Jean-Claude, K.: Segmentation of Color Fundus Images of the Human Retina: Detection of the Optic Disc and the Vascular Tree Using Morphological Techniques. *ISMDA*, LNCS 2199, pp. 282-287, Springer-Verlag Berlin Heidelberg (2001)
15. Nayak, J.; Acharya, R.: Automated Diagnosis of Glaucoma Using Digital Fundus Images. *J Med Syst*, 33:337-346 doi: 10.1007/s10916-008-9195-z, (2009)
16. Rudas C., J.; Toscano C., R.; Sánchez Torres, G.: A technique for optical disk detection on retinal images. *Revista Avances en Sistemas e Informática*, Vol.8 No3, Medellín. ISSN 1657-7663. Diciembre de (2011)
17. Abdel-Ghafa, R. A.; Morris, T.; Ritchings, T.; Wood, T.: Detection and Characterization of the Optic Disk in Glaucoma and Diabetic Retinopathy. *Department of Computation, UMIST*, PO Box 88, Manchester, M60 1QD. To become School of Informatics, Manchester University, on October (2004)
18. Sekhar, S.; Al-Nuaimy, W.; Nandi, K.: Automated localisation of optic disk and fovea in retinal fundus images. *16<sup>th</sup> European Signal Processing Conference (EUSIPCO 2008)*, Laussane, Switzerland, copyright by EURASIP, August 26-29 (2008)
19. Osareh, A; Mirmehdi, M; Thomas, B; Markham, R.: Colour Morphology and Snakes for Optic Disc Localization. *Department of Computer Science*, University of Bristol, Bristol, BS8 1UB, U.K. (2001)
20. Muramatsu, C.; Nakagawa, T.; Sawada, A.; Hatanaka, Y.; Yamamoto, T.; Fujita, H.: Automated determination of cup-to-disc ratio for classification of glaucomatous and normal eyes on stereo retinal fundus images. *Journal of Biomedical Optics*, 16(9), 096009, September (2011)
21. Muramatsu, C.; Nakagawa, T.; Sawada, A.; Hatanaka, Y.; Hara, T.; Yamamoto, T.; Fujita, H.: Automated segmentation of optic disc region on retinal fundus

- photographs: Comparison of contour modeling and pixel classification methods. *Computer Methods and Programs in Biomedicine*, 101, 23-32, Elsevir Ireland Ltd, (2010)
22. Kavitha, S.; Karthikeyan, S.; Duraiswamy, Dr. K.: Neuroretinal rim Quantification in Fundus Images to Detect Glaucoma. *IJCSNS international Journal of Computer Science and Network Security*, Vol. 10 No.6, June (2010)

# Un filtro de imágenes mediante correlación espacial

Carlos D. Estrada Chávez<sup>1</sup>, Edgardo M. Felipe-Riverón<sup>2</sup>, Mariano Rivera Meraz<sup>1</sup> y Johan Van Horebeek<sup>1</sup>

<sup>1</sup> Centro de Investigación en Matemáticas, Jalisco S/N, Col. Valenciana, C.P. 36240 Guanajuato, Gto, México

<sup>2</sup> Centro de Investigación en Computación, Instituto Politécnico Nacional, Juan de Dios Bátiz s/n, Col. Nueva Industrial Vallejo, C. P. 07738, México  
cestradac@gmail.com; edgardo@cic.ipn.mx; mrivera@cimat.mx; horebeek@cimat.mx

**Abstract.** This paper proposes a novel algorithm for removing noise in gray-levels and color images. The core concept on which is based the image filtering algorithm is a measure of spatial correlation between colors: we say that two colors are spatially correlated if they appear nearby more often in the image, with respect to other pairs of colors. In the discrete case, this measure of spatial correlation is represented by an Adjacency matrix, which can be calculated efficiently. Firstly it is presented an iterative filter with a simple scheme of 4 neighbors and local weighted averages. Subsequently, we present a non-iterative autocalibrable filter and a strategy to estimate their parameters. Given its similarity in its implementation with the bilateral filter (FB), comparisons are made with this kind of filters using a current technique which estimates the parameters on the FB. It can be seen that the proposed filter generally produces better results than the FB.

**Keywords:** Image filter, spatial correlation, bilateral filter, adjacency matrix, auto-calibration.

**Resumen.** Se propone un algoritmo novedoso para la eliminación de ruido en imágenes en niveles de gris y color. El concepto central sobre el cual se basa dicho algoritmo de *filtrado de imágenes* es una medida de correlación espacial entre colores: diremos que dos colores están correlacionados espacialmente si estos aparecen cercanos en la imagen con mucha frecuencia, con respecto a los otros pares de colores. En el caso discreto, dicha medida de correlación espacial se representa mediante una matriz de Adyacencia, la cual puede calcularse de manera eficiente. Se presenta primero un filtro iterativo con un esquema simple de 4 vecinos y promedios locales ponderados. Posteriormente, se plantean un filtro no iterativo autocalibrable y una estrategia para estimar sus parámetros. Dada su similitud en su implementación con un filtro bilateral (FB), el cual se implementa mediante la iteración de promedios locales pesados, se realizan comparaciones con este tipo de filtros mediante una técnica actual que estima los parámetros en el FB. Se puede constatar que el filtro propuesto en general produce mejores resultados que el FB.

**Palabras clave:** Filtro de imágenes, correlación espacial, filtro bilateral, matriz de adyacencia, autocalibración.

## 1 Introducción

El problema de eliminar el ruido de las imágenes digitales ha sido abordado desde varias perspectivas [1] y se continúan proponiendo nuevos algoritmos con los que cada vez se obtienen mejores resultados [2]. En la práctica, sin embargo, filtros tradicionales como el bilateral [3] continúan siendo de uso común debido a su sencillez y relativa eficacia para eliminar el ruido.

Con este trabajo intentamos introducir un nuevo enfoque basado en una heurística simple: en una imagen, los píxeles que consideramos sin ruido corresponden a los que poseen colores más “correlacionados” entre sí. De forma más específica, definiremos una medida de correlación espacial que nos indique cuándo la presencia de un color implique la presencia de otro a una distancia determinada.

Con este trabajo brindamos los resultados que se obtienen cuando le aplicamos a una imagen digital en colores, deteriorada con ruido aditivo, un filtro basado en la correlación espacial que existe entre los valores de cada píxel y el de los vecinos fuertes que lo rodean.

## 2 Correlación espacial en imágenes

Representaremos a una *imagen digital*  $I$  mediante una función  $I: \Omega \rightarrow \Phi$  en donde tanto el dominio  $\Omega$  de *coordenadas espaciales* como el conjunto de *valores*  $\Phi$  son finitos y discretos. Llamaremos *posición* a cada elemento de  $\Omega$  e *intensidad* a cada elemento de  $\Phi$ , siempre y cuando este sea unidimensional, es decir, cuando se haga referencia a una imagen en escala de gris. De lo contrario, diremos que cada elemento de  $\Phi$  es un *color*, debido a que en una imagen digital en colores, este valor *codifica* el color de manera tal que pueda ser interpretado y realizado por algún dispositivo.

Una manera de codificar el color, es haciendo referencia a un *espacio de color*, donde los valores de  $\Phi$  representen coordenadas dentro de dicho espacio. De esta forma es posible calcular una *distancia entre los colores* en el espacio de color en el cual están representados. Sin embargo, en este trabajo hablaremos de una medida de distancia entre los colores, según el grado de asociación entre los colores de los píxeles con base en su distribución espacial en la imagen. Esto lo determinaremos mediante la *correlación espacial entre colores*. De esta manera diremos que dos colores están correlacionados espacialmente en una imagen digital, si estos aparecen cercanos espacialmente con mayor frecuencia, con respecto a otros pares de colores.

### 2.1 Imagen de etiquetas

Para poder calcular la correlación espacial entre cada par de colores en una imagen, es necesario identificar todas las parejas de colores similares posibles. Es necesario entonces que el número de parejas de colores similares sea adecuado a las capacidades de procesamiento. Para reducir el número de colores le asignamos a los colores similares una sola etiqueta. A este proceso de reducir el número de colores en la imagen, se le conoce en la literatura como cuantización (o cuantificación) del color.

Nos referiremos como etiquetas a un conjunto de índices (ej.  $\{1, 2, 3, 4, \dots\}$ ), que identifican la clase o estado de un pixel después de la cuantización.

Llamaremos imagen de etiquetas de una imagen digital, a un arreglo bidimensional del mismo tamaño que la imagen, pero cuyos valores son etiquetas que identifican a un color, o a varios colores similares. Esta manera de codificar una imagen de color mediante etiquetas es conocida como *indexado de color*.

## 2.2 Matriz de adyacencia (MA)

Dada la *imagen de etiquetas*  $\tilde{I}$ , con rango  $\Lambda = \{1, 2, \dots, N\}$ , su *matriz de adyacencia*  $\mathbf{A}$  de dimensión  $N \times N$ , es aquella donde cada una de sus componentes  $\mathbf{A}_{a,e}$ ,  $a, e \in \Lambda$  toma el valor del conteo de coincidencias entre los pixeles contiguos etiquetados con los números  $a$  y  $e$ , Ec. (1), i.e.:

$$\mathbf{A}_{a,e} = \sum_{\forall p \in \Omega} \mathbf{1}[\tilde{I}_p = a] \sum_{k=1}^4 \mathbf{1}[\tilde{I}_{(p-t_k)} = e] \quad (1)$$

donde  $\mathbf{1}$  es la *función indicadora*,  $t_1 = (1, 0)$ ,  $t_2 = (0, 1)$ ,  $t_3 = (-1, 0)$ ,  $t_4 = (0, -1)$ .

La matriz de adyacencia se puede formar considerando una vecindad *4-conectada* acumulando el número de veces que dos etiquetas son contiguas en la imagen. El siguiente algoritmo muestra lo económico que resulta calcular la MA para una imagen de  $m$  filas y  $n$  columnas.

---

### Algoritmo 1. Generación de la matriz de adyacencia.

---

*Entrada:* imagen de etiquetas  $\tilde{I}$ , número de etiquetas  $N$ .

*Salida:* matriz de adyacencia  $\mathbf{A}_{N \times N}$ .

Definir:

$m, n$ : alto y ancho de  $\tilde{I}$

$\mathbf{A} = \mathbf{0}_{N \times N}$

Para cada  $i=1, \dots, m$

  Para cada  $j=1, \dots, n$

    Si  $i > 1$

$$\mathbf{A}(\tilde{I}_{(i,j)}, \tilde{I}_{(i-1,j)}) = \mathbf{A}(\tilde{I}_{(i,j)}, \tilde{I}_{(i-1,j)}) + 1$$

    Fin

    Si  $j > 1$

$$\mathbf{A}(\tilde{I}_{(i,j)}, \tilde{I}_{(i,j-1)}) = \mathbf{A}(\tilde{I}_{(i,j)}, \tilde{I}_{(i,j-1)}) + 1$$

    Fin

  Fin  $j$

Fin  $i$

$$\mathbf{A} = \mathbf{A} + \mathbf{A}^T$$

Notación.  $\mathbf{A}(a, e) := \mathbf{A}_{a,e}$

La MA nos proporciona la correlación espacial entre las etiquetas de la imagen tomando una distancia de un pixel. La primera observación que se hace sobre la MA, es que generalmente las etiquetas correspondientes a colores similares resultan estar correlacionadas espacialmente; esto es debido a que, en una imagen comúnmente encontraremos regiones de pixeles con colores similares entre sí. Esto se aprecia en la figura 1, donde se toma la imagen en tonos de gris como la imagen de etiquetas. Las gráficas a la derecha de cada imagen resultan después de ordenar los renglones y las columnas de la MA con respecto a su tono de gris asociado.

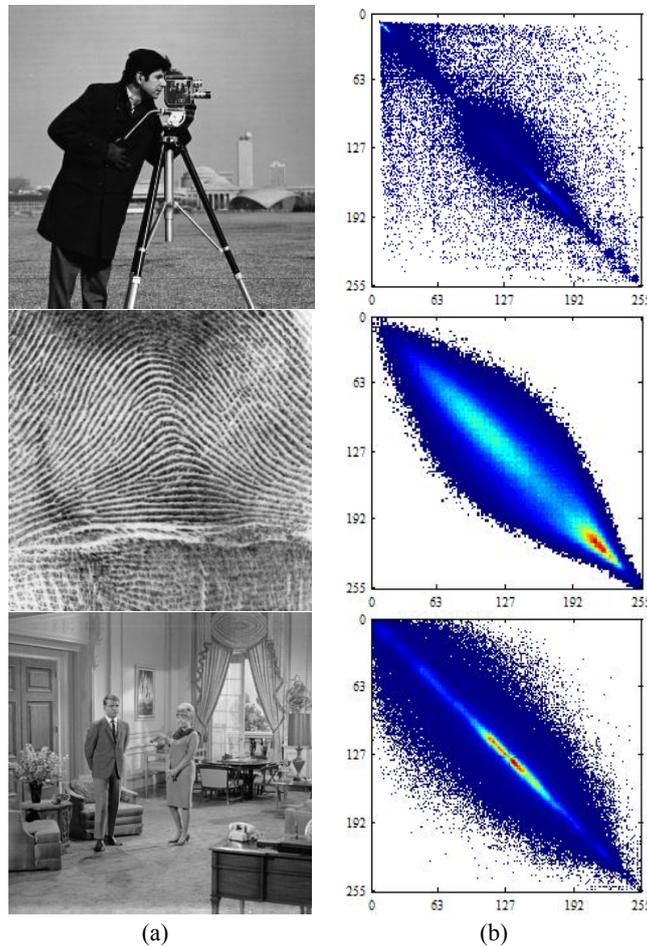


Fig. 1. (a) Imágenes en tonos de gris y (b) gráfica de su correspondiente MA

### 3 El filtro 4-conectado

Se presenta un filtro que estima el valor de cada pixel con base en los valores de los cuatro pixeles adyacentes fuertemente conectados, de manera que se esperaría que la escasa información local sea compensada al emplear los pesos calculados mediante la MA.

---

#### Algoritmo 2. Filtrado de imágenes con vecindad 4-conectada.

---

*Entrada:* imagen  $I$

*Salida:* imagen filtrada  $\hat{I}$

Calcular imagen de etiquetas  $\tilde{I}$  (ver sección 2.1): En imágenes en colores se recomienda emplear el algoritmo de cuantificación de Wu [4]. Para el caso de imágenes en escala de grises se sugiere usar cuantización uniforme sobre la imagen con el histograma ecualizado.

Calcular matriz de adyacencia  $\mathbf{A}$  (algoritmo 1)

Para cada  $p \in \Omega$ :

$$\hat{I}_p = \frac{\sum_{q \in V_p^4} \mathbf{A}_{a,e} I_q}{\sum_{q \in V_p^4} \mathbf{A}_{a,e}} \quad (2)$$

donde  $a = \tilde{I}_p$ ,  $e = \tilde{I}_q$  y  $V_p^4$  son los 4 vecinos fuertemente conectados de  $p$ .

Obsérvese que se recorre la imagen de etiquetas  $\tilde{I}$  al mismo que la imagen de entrada  $I$ .

---

El algoritmo anterior bien puede aplicarse utilizando una ventana de cualquier tamaño, tema que abordaremos más adelante. Por ahora, mostraremos el efecto de aplicar este filtrado con vecindad 4-conectada de forma iterativa.

La implementación se realizó en Matlab. Para generar la imagen de etiquetas correspondientes a las imágenes de color, se tomó la función `rgb2ind`<sup>1</sup> con la opción `nodither`, la cual implementa el algoritmo de cuantización de mínima varianza de Wu<sup>2</sup> [4]. Se pudo comprobar visualmente que con dicho algoritmo de cuantización se producen mejores resultados que con la cuantización uniforme. La cuantificación de Wu divide el cubo de color RGB en ‘cajas’ (donde cada caja representa una etiqueta), asignando más etiquetas en regiones del cubo de color (histograma) donde existe una

---

<sup>1</sup> Documentación de `rgb2ind` en Matlab:

<http://www.mathworks.com/help/matlab/ref/rgb2ind.html>  
Consultado 13-Nov-2012

<sup>2</sup> Código disponible en la página Web de Xiaolin Wu:

<http://www.ece.mcmaster.ca/~xwu/cq.c>  
Consultado 13-Nov-2012

mayor concentración de colores. Para las pruebas se utilizó un máximo de 255 etiquetas.

Lo primero que se pudo observar al suavizar las imágenes es que en general el filtro propuesto respeta los bordes después de varias iteraciones como puede verse en la figura 2.

Una explicación de esto podría ser que la cantidad de píxeles de los bordes de los objetos presentes en la imagen, por lo general es pequeña en relación con la cantidad de píxeles en las regiones suaves y por lo tanto la correlación espacial entre colores diferentes será baja con relación a la correlación con colores similares. También, como consecuencia de lo anterior, las imágenes con superficies granulares o con texturas finas presentan un mayor emborronamiento (ver ejemplo en la figura 3).



**Fig. 2.** Filtrado iterativo del filtro 4-conectado. De izquierda a derecha y de arriba abajo: imagen original e imágenes resultantes después de aplicar el filtro iterativamente 3, 10 y 25 veces respectivamente



**Fig. 3.** Efecto del filtrado después de 10 iteraciones en imágenes con superficies granulares o de texturas finas

Una característica que pudo observarse en las imágenes filtradas es que los colores con bajo valor de correlación espacial con ellos mismos tienden a desaparecer en la imagen filtrada como se observa en la figura 4.



**Fig. 4.** De izquierda a derecha: imagen con ruido de alta frecuencia e imagen filtrada en 3 iteraciones

## 4 Adaptabilidad al ruido

Los parámetros de configuración del filtro 4-conectado son solamente el número de iteraciones y los parámetros propios del algoritmo de cuantización empleado. Extenderemos dicha estrategia a un *filtro de ventana local*, para lo cual es necesario introducir otros parámetros. Los parámetros de calibración permiten adecuar el filtro a las diversas características de la imagen y a diversas aplicaciones o propósitos. Un valor adecuado para estos parámetros puede inferirse sólo después de observar la imagen, por lo que la efectividad del filtrado queda sujeta a la capacidad visual y experiencia de quien calibra la imagen. Desde luego, sería interesante tener un filtro autocalibrable, es decir, que se ajuste sin supervisión a las condiciones de la imagen. En esta sección se busca cumplir con los siguientes objetivos:

1. Mejorar los resultados del filtro 4-conectado presentado en la sección anterior.
2. Presentar un filtro autocalibrado que tome una región local más amplia.
3. Medir la efectividad del filtro resultante comparándolo contra el filtro bilateral.

### 4.1 El filtro bilateral y su calibración

Dada una *función de cercanía*  $f$  que evalúa si la distancia entre dos píxeles se encuentra dentro de cierta escala espacial, y una *función de similitud*  $g$  que indica el grado de semejanza entre dos valores del rango de la imagen  $I$ , el filtro bilateral [3] estima el nuevo valor de cada píxel mediante la Ec. (3):

$$\hat{I}_p = \frac{\sum_{q \in \Omega} f(p, q) g(I_p, I_q) I_q}{\sum_{q \in \Omega} f(p, q) g(I_p, I_q)} \quad (3)$$

La forma usual de definir  $f$  y  $g$  es mediante funciones gaussianas, Ec. (4) y Ec. (5):

$$g_{bil}(I_p, I_q) = \exp\left(-\frac{\|I_p - I_q\|^2}{2\sigma_r^2}\right) \quad (4)$$

$$f_{bil}(p, q) = \exp\left(-\frac{\|p - q\|^2}{2\sigma_d^2}\right) \quad (5)$$

donde los parámetros  $\sigma_r$  y  $\sigma_d$  controlan la apertura de la gaussiana en el rango y en el dominio respectivamente.

La calibración del filtro bilateral consiste en encontrar valores idóneos para  $\sigma_r$  y  $\sigma_d$  de tal forma que estos se adecuen a las características de la imagen que se quiere filtrar. El par de parámetros (más el tamaño de ventana) que controlan el comportamiento del filtro bilateral logran que este filtro sea muy flexible; sin embargo, de acuerdo a [5] no existía a la fecha un análisis teórico de cómo seleccionar de manera óptima los parámetros  $\sigma_d$  y  $\sigma_r$ . En el mismo trabajo citado se propone con base en pruebas realizadas, que una buena elección del valor de  $\sigma_d$  parece estar en el rango [1.5, 2.1], mientras que  $\sigma_r$  parece estar relacionado linealmente con la desviación estándar ( $\sigma_n$ ) del ruido blanco aditivo Gaussiano (AWGN - *Additive White Gaussian Noise*) agregado artificialmente a imágenes de prueba; en la práctica se desconoce el valor  $\sigma_n$  de la imagen que se desea filtrar.

Para la realización de nuestros experimentos se contaminaron imágenes con ruido AWGN; posteriormente se aplicó el filtro propuesto y se midió la diferencia con respecto a la imagen original (sin ruido). Finalmente, se compararon los resultados con los del filtro bilateral calibrándolo de la misma forma que en [5]:  $\sigma_d = 1.8$ ,  $\sigma_r = 2 \times \sigma_n$  con tamaño de ventana de  $11 \times 11$ . Adicionalmente, se obtuvo la calibración óptima por nivel de ruido mediante la función `fminsearch` de Matlab. El valor óptimo obtenido para los parámetros  $\sigma_d$  y  $\sigma_r$  para un ancho de ventana de  $11 \times 11$  se muestra en la Tabla 1.

**Tabla 1.** Calibración óptima del conjunto de imágenes de prueba para el filtro bilateral. Los valores  $\sigma_r$  se refieren a los valores de la intensidad tomados en el rango [0, 1]

$255\sigma_n; \sigma_n$	$\sigma_d$	$\sigma_r$
10; 0.0392	1.054225	0.109223
15; 0.0588	1.170315	0.169392

20; 0.0784	1.258260	0.232878
25; 0.0980	1.331346	0.302320
30; 0.1176	1.401339	0.376674
35; 0.1373	1.467100	0.460825
40; 0.1569	1.530953	0.561410
50; 0.1961	1.669907	0.822967
60; 0.2353	1.831355	1.159397
80; 0.3137	2.197954	2.002180
100; 0.3922	2.596337	3.022389

Para las pruebas que se presentan en las siguientes subsecciones se usaron 11 imágenes en tonos de gris descargadas del sitio [6]. Con respecto a la cuantización de las imágenes en tonos de gris, se efectuó una cuantización uniforme usando 255 bins sobre la imagen con el histograma ecualizado.

#### 4.2 Similitud contextual

Una manera de visualizar las relaciones entre los colores de una imagen es a través de un grafo, donde cada etiqueta corresponde a un nodo y donde las aristas representan la *probabilidad de transición* de una etiqueta a otra. A partir de la matriz de adyacencia  $\mathbf{A}$ , definiremos dichas probabilidades mediante la *matriz de transición* (MT)  $\mathbf{P}_{N \times N}$ .

$$\mathbf{P}_{a,e} = \frac{\mathbf{A}_{a,e}}{\sum_{i=1}^N \mathbf{A}_{a,i}} \quad (6)$$

donde  $\mathbf{P}_{a,e}$  es la probabilidad de pasar de un nodo con etiqueta  $a$  a otro con etiqueta  $e$ .

Nos gustaría entonces, medir la similitud entre dos etiquetas tomando en cuenta su contexto, es decir, incluir la influencia de trayectorias indirectas que conecten a los correspondientes nodos en el grafo de etiquetas. Con este propósito, considérese la matriz (Ec. 7):

$$\mathbf{Q} = \left( \frac{\mathbf{P} + \mathbf{P}^T}{2} \right) \quad (7)$$

Cada entrada de la matriz  $\mathbf{Q}^t$  indica la probabilidad de trasladarse en  $t$  pasos entre los dos nodos (promediando las probabilidades de transición de ida y vuelta). Podemos utilizar ahora la matriz  $\mathbf{Q}^t$  en vez de  $\mathbf{A}$  para filtrar la imagen con el algoritmo del filtro 4-conectado. Así se pudo determinar después de varios experimentos (omitidos aquí) que tomar  $t = 2$  arroja los mejores resultados para el conjunto de imágenes de prueba. La gráfica de la figura 5 brinda una comparación del filtro 4-conectado utilizando las matrices  $\mathbf{A}$  y  $\mathbf{Q}^2$ .

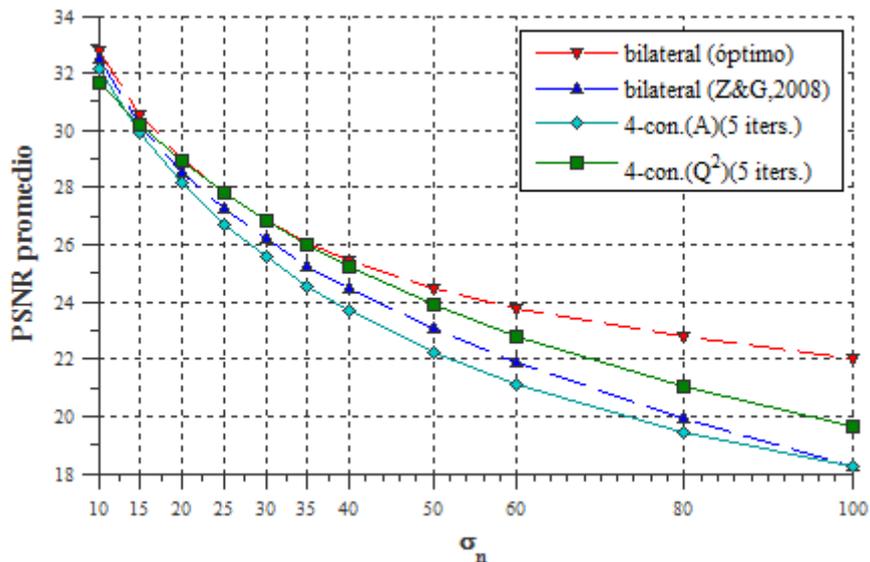


Fig. 5. Comparación del PSNR<sup>3</sup> promedio para las funciones de similitud A vs.  $Q^2$  utilizando el filtro 4-conectado

Como se puede apreciar en la gráfica anterior, con  $Q^2$  se mejoran significativamente los resultados del filtro 4-conectado.

### 4.3 Cercanía y tamaño de la vecindad

Dado un pixel de la imagen, la *función de cercanía* debe ajustarse a la región de pixeles similares al pixel dado en la imagen original idealmente sin ruido, es decir, a diferencia de la función de similitud, la función de cercanía no debería verse afectada por la magnitud del ruido presente en la imagen.

Sin embargo, en general se tendrá que a mayor nivel de ruido será necesaria una cantidad mayor de valores de muestra, i.e. una región local más amplia para así poder obtener una estimación más acertada del verdadero valor del pixel actual. Por lo tanto, podemos decir que una buena definición de la función de cercanía debe considerar un equilibrio entre dos criterios opuestos: incluir un número suficiente de pixeles (muestras) de entre los más cercanos, pero considerando una región local lo suficientemente pequeña de manera que se tomen en cuenta solo los más similares.

Por lo tanto, podemos conjeturar que dada una función de similitud  $g$ , existe una función de cercanía  $f$  que optimiza un *criterio de selección de región local* de manera que el filtro repare lo mejor posible una imagen. No está claro, sin embargo, hasta qué punto pueden ser útiles las herramientas con las que contamos para poder encontrar

<sup>3</sup> Del inglés *Peak Signal-to-Noise Ratio*. Medida asociada con el logaritmo del error cuadrático medio.

dicha  $f$  óptima. De momento, solo consideraremos un indicador que parece ser útil para calibrar  $f$ , que capte en esencia los criterios para definirla adecuadamente.

Sea pues,  $\boldsymbol{\theta}$  el vector correspondiente a la diagonal de la matriz  $\mathbf{P}$ .

$$\boldsymbol{\theta} = \text{diag}(\mathbf{P}) \quad (8)$$

Entonces,  $\theta_a$  es la probabilidad de transitar de un pixel con etiqueta  $a$  a otro pixel con la misma etiqueta. Sea  $S_{\sim a}$  una variable aleatoria que representa el número de veces que se transitaría a una etiqueta diferente de  $a$  en  $n$  intentos, i.e.:

$$S_{\sim a} \sim \text{Binomial}(n, 1 - \theta_a) \quad (9)$$

Un indicador del *tamaño de la vecindad* de la etiqueta  $a$  puede ser el número de transiciones esperadas a etiquetas distintas a la actual. Tomando esta cantidad normalizada quedaría:

$$\frac{E[S_{\sim a}]}{\sqrt{\text{Var}[S_{\sim a}]}} = \frac{n(1 - \theta_a)}{\sqrt{n\theta_a(1 - \theta_a)}} = \sqrt{\frac{n(1 - \theta_a)}{\theta_a}} \quad (10)$$

Para tener un solo indicador para todas las etiquetas, usaremos el promedio de los valores de  $\boldsymbol{\theta}$  ponderados con el histograma de la imagen de etiquetas  $\mathbf{h}$ :

$$\theta = \frac{\sum_{i \in \Lambda} \mathbf{h}_i \theta_i}{\sum_{i \in \Lambda} \mathbf{h}_i} \quad (11)$$

De esta forma, si  $f$  es definida mediante una gaussiana como en (5), una estimación para  $\sigma_d$  puede obtenerse con:

$$\hat{\sigma}_d = c \sqrt{\frac{(1 - \theta)}{\theta}} \quad (12)$$

En la Tabla 2 se muestra el valor del indicador de tamaño de vecindad para algunas de las imágenes de prueba con distintos niveles de ruido para  $c = 1$ .

Con esto, actualizamos el filtro 4-conectado proporcionado en el algoritmo 2 para usar  $\mathbf{Q}^2$  como función de similitud y  $f$  calibrada mediante el indicador de tamaño de vecindad. A este filtro lo identificaremos como *filtro por correlación espacial* (fce). La diferencia primordial con el algoritmo 4-conectado es que el fce no es un filtro iterativo. La gráfica de la figura 6 brinda una comparación del fce y el filtro bilateral.

**Tabla 2.** Indicador del tamaño de vecindad en imágenes en tonos de gris

Tamaño de la vecindad: $\sqrt{\frac{(1-\theta)}{\theta}}$								
$\sigma_n$	camarógrafo 256 × 256	Lena 512 × 512	Bárbara 512 × 512	bote 512 × 512	pareja 512 × 512	huella 512 × 512	colina 512 × 512	casa 256 × 256
10	5.33	5.76	6.00	5.24	5.34	6.41	5.43	5.42
15	6.07	6.58	6.62	5.98	6.14	6.82	6.13	6.38
20	6.86	7.14	7.05	6.52	6.75	7.25	6.69	7.14
25	7.32	7.55	7.44	6.96	7.20	7.60	7.07	7.77
30	7.81	7.91	7.73	7.32	7.64	7.84	7.43	8.23
35	8.10	8.15	7.96	7.63	7.99	8.18	7.73	8.41
40	8.36	8.32	8.11	7.79	8.20	8.38	7.98	8.54
50	8.70	8.73	8.48	8.15	8.63	8.60	8.34	9.04
60	8.76	9.00	8.70	8.49	8.88	8.78	8.67	9.37
80	9.33	9.32	9.02	8.92	9.35	9.09	9.08	9.72
100	9.32	9.53	9.28	9.12	9.54	9.20	9.28	9.68

---

**Algoritmo 3. Filtrado de imágenes por correlación espacial**

---

*Entrada:* imagen  $I$ , tamaño de ventana  $l$ , constante de vecindad  $c$ , exponente de la matriz de correlación  $t$ .

*Salida:* imagen filtrada  $\hat{I}$ .

Calcular la imagen de etiquetas  $\tilde{I}$  (ver sección 2.1): En imágenes en colores se recomienda emplear el algoritmo de cuantificación de Wu [4]. Para el caso de imágenes en escala de grises se sugiere usar cuantización uniforme sobre la imagen con el histograma ecualizado.

Calcular la matriz de adyacencia  $\mathbf{A}$  (algoritmo 1)

Calcular la matriz de transición  $\mathbf{P}$ , Ec. (6)

$$\mathbf{Q} = \left( \frac{\mathbf{P} + \mathbf{P}^T}{2} \right)$$

Para cada  $p \in \Omega$ :

$$\hat{I}_p = \frac{\sum_{q \in W_p^l} \mathbf{Q}_{a,e}^t f(p,q) I_q}{\sum_{q \in W_p^l} \mathbf{Q}_{a,e}^t f(p,q)} \tag{13}$$

$$\text{Con } f(p, q) = \exp\left(-\frac{\|p - q\|_2^2}{2\hat{\sigma}_d^2}\right), \hat{\sigma}_d = c \sqrt{\frac{(1-\theta)}{\theta}} \text{ para una constante estimada } c$$

y donde,  $a = \tilde{I}_p$ ,  $e = \tilde{I}_q$  y  $W_p^l$  son las posiciones de una ventana con centro en  $p$  y lado de longitud  $l$ .

Obsérvese que se recorre la imagen de etiquetas  $\tilde{I}$  al mismo que la imagen de entrada  $I$ .

#### 4.4 Discusión de los resultados del filtro por correlación espacial (fce).

La gráfica muestra los resultados del fce con  $c = 1/5$ ,  $t = 2$ .

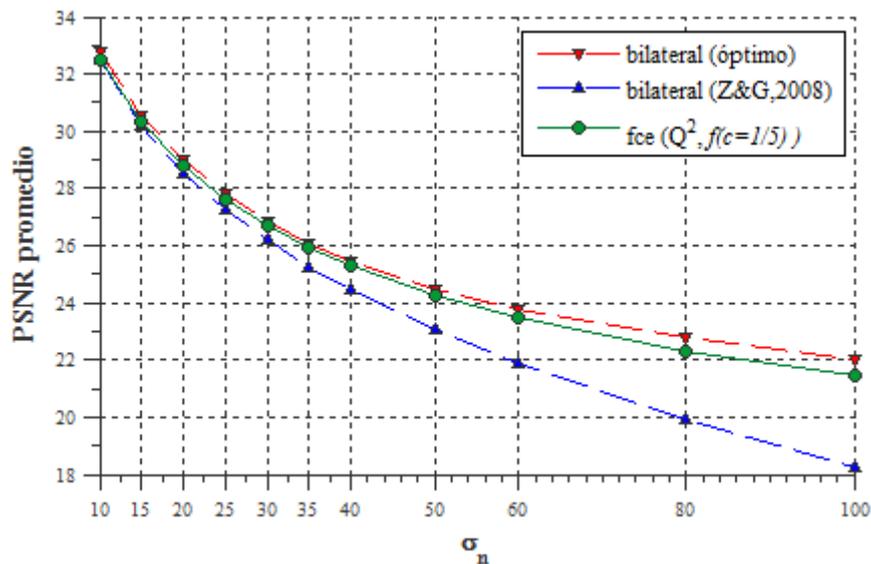


Fig. 6. Comparación del fce y el filtro bilateral

Puede observarse que el filtro propuesto obtiene mejores resultados que el filtro bilateral calibrado como en [6] y apenas por debajo de la calibración óptima obtenida para el conjunto de imágenes de prueba. Considerando que para calibrar el filtro bilateral en estos experimentos fue necesario conocer el valor de  $\sigma_n$ , la ventaja práctica del fce sobre el filtro bilateral es evidente, ya que como se había mencionado anteriormente, en la práctica se desconoce el valor  $\sigma_n$ , por lo que para una calibración automática del filtro bilateral se requeriría entonces una estimación de  $\sigma_n$ .

Si bien el fce requiere establecer el valor de los parámetros  $t$  y  $c$ , estos no cambian en función del nivel del ruido, por lo que con base en los experimentos realizados es evidente que el fce es adaptable al nivel del ruido presente en la imagen. Lo siguiente es comprobar que el mismo filtro funciona bien en imágenes de color, lo cual se validó visualmente al aplicarlo a varias imágenes en colores. La figura 7 muestra las imágenes

de color originales sobre las cuales se probó el algoritmo. La figura 8 muestra las imágenes contaminadas con ruido y los resultados obtenidos con el filtro propuesto.



Fig. 7. Imagen utilizada para la prueba de la figura 8 (izquierda) y acercamiento (derecha).

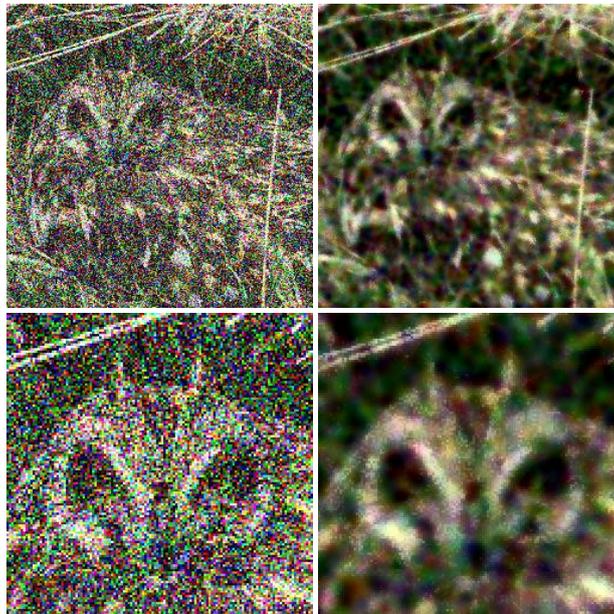


Fig. 8. Arriba izquierda: imagen contaminada con AWGN, a)  $\sigma_n = 100$  ; arriba derecha: imagen filtrada con el fce ( $\mathbf{Q}^2, f(c=1/5)$ ), tamaño de ventana  $11 \times 11$ . Abajo, acercamientos de las correspondientes imágenes anteriores

## 5 Conclusiones

El fce mostró ser eficaz al eliminar el ruido blanco aditivo gaussiano utilizando los parámetros propuestos. El fce representa una mejora al filtro bilateral al buscar adaptarse a las condiciones de ruido presentes en la imagen, además de que no hace uso de una medida de distancia entre los colores (relativa a un espacio de color), sino

de la correlación espacial entre los colores de la imagen. Debe tenerse en cuenta que la eficacia de la estrategia presentada depende en gran medida del algoritmo de cuantización empleado, lo que también repercutirá en el tiempo de cómputo.

**Reconocimientos.** Se agradece al Dr. José Luis Marroquín Zaleta y al Dr. Jean-Bernard Hayet, del CIMAT, quienes fueron revisores de la tesis de maestría que sirvió de base para el presente trabajo. Agradecemos también al Centro de Investigación en Computación (CIC) del Instituto Politécnico Nacional (IPN), por las facilidades brindadas para la presentación de este trabajo.

## Referencias

1. Buades, A., Coll, B., Morel, J.M.: A review of image denoising algorithms, with a new one. In: Proc. Multiscale Modeling and Simulation, Society for Industrial and Applied Mathematics (SIAM), vol. 4, n. 2, pp. 490-530 (2005)
2. Dabov, K., Foi, A., Katkovnik, V., Egiazarian, K.O.: BM3D image denoising with shape-adaptive principal component analysis. In: Proc. Workshop on Signal Processing with Adaptive Sparse Structured Representations, SPARS 2009 (2009)
3. Tomasi, C., Manduchi, R.: Bilateral filtering for gray and color images. In : IEEE International Conference on Computer Vision, pp. 839–846 (1998)
4. Wu, X.: Efficient Statistical Computations for Optimal Color Quantization. In : Arvo, J. (ed.) Graphics Gems, vol. 11, pp. 126–133 (1991)
5. Zhang , M., Gunturk, B. K.: Multiresolution Bilateral Filtering for Image Denoising. In: IEEE Transactions on Image Processing 17(12), 2324 - 2333 (2008)
6. Danielyan, A., Maggioni, M., Dabov, K. Foi, A., Katkovnik, V., Egiazarian, K.: Results of the proposed grayscale BM3D method. Available at: <http://www.cs.tut.fi/~foi/GCF-BM3D>. Accessed 6 Nov 2012



# **Communications & Computer Networks**

---



# Frequency Planning with an Optimized Frequency Reuse Distance for Fixed WiMAX Networks

Miguel Sanchez Meraz mmeraz@ipn.mx  
Carlos Sosa Paz csosa@ipn.mx  
Duarte Calderon Alizari irazila@hotmail.com

Departamento de Telecomunicaciones ESIME Zacatenco Instituto Politécnico  
Nacional  
México City, México.

**Abstract.** This paper propose a methodology based on Integer Linear Programming to perform the frequency assignment in WiMAX networks. In a first step the reuse frequency distance is calculated. The obtained results for the performance of WiMAX networks using the frequency assignments obtained with the proposed methodology are similar with those obtained with frequency assignment obtained from professional radio planning tools.

## 1 INTRODUCTION

At present broadband wireless networks are one of the best technological alternatives to meet the communication needs of end users located in suburban and rural areas with poor connectivity services. In the case of Mexico, there are many areas within its territory that result inaccessible to wired networks which are the main offering option for broadband connectivity in the country. These areas have lacked connectivity services due to the high complexity and cost of providing such services. Finally this situation results in a lag in the development of these areas.

In order to increase the broadband penetration level in Mexico, the federal government is currently deploying a nationwide network of broadband wireless access based on WiMAX technology. This network is called "Redes Estatales de Educación Salud y Gobierno (REESyG)". With the purpose of operate these networks the federal government has booked the frequency spectrum from 3300 to 3350 MHz (50 MHz bandwidth) in all the country [1].

One of the key factors that define the performance of a wireless broadband network is the level of interference present in the network. The level of interference directly affects the transmission rate and limits the performance. The interference present in a network depends on the used frequency assignment plan [2]. In a network with a reduced number of sites, assignment of frequencies to different sectors of the network can be done manually. But with a large number of sites the frequency assignment problem becomes a major design challenge. In the case of the REESyG it is estimated that may be taken into operation

around 2000 sites in the country. This will require the design of efficient frequency assignment plans in order to ensure the coexistence of many sites in the same geographical area without reducing the global network performance.

As shown, the problem of frequency assignment is critical in the design and optimization of wireless broadband networks such as WiMAX. This paper proposes a methodology based on Integer Linear Programming to generate frequency assignment plans. This methodology is tuned to the WiMAX technology specifications and considers a fixed assignment of one channel for each sector of the network. The obtained frequencies assignment plans ensure an efficient use of the available electromagnetic spectrum resources by reducing the levels of interference in the network. An efficient assignment of the frequencies results in benefits to the end-user and to the network itself. Currently there are several professional-level tools for radio network planning that can address the issue of design of frequency plans. However, the proposal of this paper seeks to provide a simpler alternative to generate efficient frequency plans adjusted to the particular needs of the REESyG.

This paper is organized as follows. In the second section it is presented a procedure for determining the frequency reuse distance based on different scenarios of interference. The third section presents the proposed methodology for frequency assignment in WiMAX networks. Finally the fourth section presents the performance results of a network that operates with the frequency plan obtained with the proposed methodology.

## 2 DETERMINING THE REUSE DISTANCE

### 2.1 $C/I$ ratio evaluation

Carrier to Interference ratio ( $C/I$ ) provides a measure of the relationship between the average power of the transmitted signal and the average power of interference [3]. Interference is a disturbance affecting the performance of a radio system due to the close operation of one or more links that use the same frequency (co-channel interference) or an adjacent frequency (adjacent channel interference) as shown on Figure 1. Both the adjacent interference as co-channel interference can severely limit the capacity of a network. The design of an efficient frequency assignment plan must offer reduced interference levels for the network.

In Figure 1 above, ED indicates the wanted link and IE indicates the interfering link. The received power depends on the propagation loss, the power transmitted by the site and the gains of the transmitting and receiving antennas. In this paper we consider that the site use sector antennas of 90 degrees. A site is composed of 4 sites to provide coverage of 360 degrees.

The received power of a wanted signal is determined as [3]:

$$PrD(dBm) = PtI + GTD + GRD - Lbd \quad (1)$$

The received power of an interfering signal is determined as:

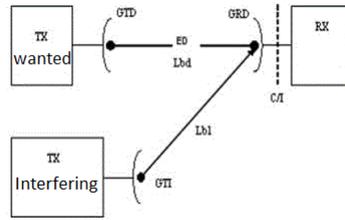


Fig. 1. Radio link operating with an interfering transmitter

$$PrI(dBm) = PtI + GTI + GRI - LbI \tag{2}$$

where:

$PtD$  is the wanted transmitter power

$PtI$  is interfering transmitter power

$GTD$  is the antenna gain of the wanted transmitter

$GTI$  is the gain of the interfering transmitter antenna

$GRD$  is the receiver gain on the wanted link

$GRI$  is the receiver gain in the interfering link

$Lbd$  are the propagation losses of wanted link

$LbI$  are de propagation losses of interfering link

$$\frac{C}{I}(dBm) = PrD - PrI \tag{3}$$

If multiple interference exists, the total interfering received power must be calculated, which is the result of the addition of partial interference sources. In this case the above equation can be expressed as:

$$\frac{C}{I}(dBm) = PrD - 10 \log \left( \sum_{i=1}^n PrI_i \right) \tag{4}$$

where:

$$PrI_i = 10^{0.1PrI} \tag{5}$$

In the case of the WiMAX technology the  $C/I$  ratio of a link defines the highest modulation that can be used on it. Table 1 shows an example of the minimum  $C / I$  ratio required by a manufacturer of WiMAX equipment for a link to operate with different modulation schemes. Different manufacturers offer this type of tables to use their equipment.

From the point of view of the network planning, a main objective is to ensure links with high level modulations. So the  $C/I$  ratio in the network must be maintained above a threshold to ensure the operation of links with the wanted modulation. To achieve this goal is very important to define a distance for the

Modulation	Required $C/(N + I)$ dB
BPSK 1/2	3
QPSK 1/2	9.375
QPSK 3/4	12
16QAM 1/2	15.75
16QAM 3/4	18.375
64 QAM 2/3	22.5
64 QAM 3/4	24

**Table 1.** Minimum  $C/I$  ratio required to operate different modulation schemes on the platform redmax [4]

reuse of frequencies. Based on this distance it is possible to determine those areas of a network that can share the same frequency. The operating frequency assignment to each sector of the network must be such as to ensure that the  $C/I$  ratio is kept at the wanted levels.

In this work the propagation losses area calculated using the SUI model of the Stanford University. The IEEE 802.16 standard working group in conjunction with Stanford University developed a WiMAX channel model for suburban environments. One of the most important achievements was the SUI model for propagation loss (Stanford University Interim), which is an extension of previous work developed by AT & T Wireless and the analysis of it made by Erceg [5]. This model was adopted by the IEEE 802.16 working group as the recommended model for fixed WiMAX applications. For the propagation loss through the Extended SUI model it is required to know three values: the height of the antennas, the operating frequency and link distance. The model proposes three different types of scenarios: Erceg A, applicable to mountainous terrain with medium / high density of trees and urban areas; Erceg B, applicable for mountainous areas with low tree density or flat terrain with moderate / high density of trees and suburban areas; and Erceg C, applicable to flat terrain with low tree density and rural areas. In this work the Erceg B model was used because of the conditions of the analysis scenario in a zone of the Mexico City. For a complete treatment of the SUI model refer to [5,6].

## 2.2 Reuse distance calculation

To calculate the reuse distance an analysis was performed on a scenario consisting of four interfering sites named I1, I2, I3 and I4, and the wanted Tx transmitter site, each with 4 sectors of 90 degrees. This scenario is representative of most real scenarios where network deployment consider the frequency reuse and an analysis of  $C/I$  ratio is required. The analysis is performed only with those sectors that may radiate to the area where the interfered receiver is located, as shown in Figure 2. In this figure the analysis area is a 5Km x 5Km box, divided into cells with 0.5 Km by side. A cell represents the location of a receiver that wants to establish a link to the wanted transmitter Tx. The analysis considered that

the interfering sectors were located at distances from 3Km up to 15Km from the transmitter.

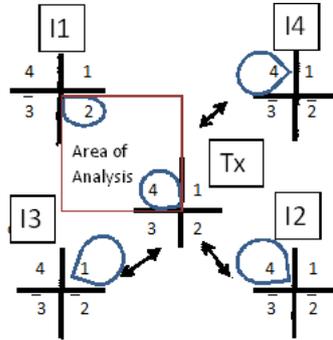


Fig. 2. Figure 2. Interference analysis scenario

For the  $C/I$  ratio evaluation were considered technical specifications of the transmitter and receiver, such as the sector antenna radiation pattern and transmitter power. Based on the equations 5 and 7 calculations of the  $C/I$  ratio were performed for each of the 100 cells in the analysis area. These calculation were performed for each interfering site in a separately way and also for the case when the four sites were interfering at the same time. The obtained results of  $C/I$  ratio let to establish the highest modulation that could be operated in a radio link for each of the 100 analyzed cells. Table 2 shows the percentage of cells with 64QAM 3/4 (the highest WiMAX modulation) for different distances of analysis, according to Figure 2, ranging from 3Km to the 15Km. This table presents the percentages of cells with 64QAM 3/4 modulation for the case of only one interfering site of type I2 and for the case of only one interfering site of the type I3 or I4 (because of their symmetry, interfering sites I3 and I4 affects in the same way).

These obtained results let to define the frequency reuse distance depending on the main wanted modulation in the operation of a WiMAX network. For example if it is wanted that more than 80% of the network operate with 64QAM 3/4 modulation, the frequency reuse distances should be of 9 Km for type I2 interfering sites and 5 Km for type I3 and I4 interfering sites.

### 3 CHANNEL ASSIGNMENT USING ILP

Once the frequency reuse distances were calculated, in this section the proposed methodology for the frequency or channels assignment in the WiMAX network is presented. The Assignment Problem for radio frequencies is complex. It requires sharing the available resources among all sites in the network efficiently and at the same time minimizing the interference within the network.

Analysis distance (Km)	Percentages of cells with 64QAM 3/4 modulation	
	I2 type interferent	I3, I4 types interferences
3	27.27	48.76
5	50.41	85.95
7	71.90	97.52
9	85.12	99.17
11	94.21	99.17
13	97.52	99.17
15	99.17	99.17

**Table 2.** Percentage of cells with 64qam 3/4 modulation for different distances and different interference analysis

Consider a wireless network system consisting of  $n$  sites each one is formed by  $m$  sectors. Let,  $R$  be the set of sites  $R = \{r_1, \dots, r_n\}$  and  $S$  be the set of sectors  $S = \{s_1, \dots, s_m\}$ . Therefore the universe of sites-sectors in the network is defined by the set  $N = R \times S = \{(r_1, s_1), (r_1, s_2), \dots, (r_1, s_m), (r_2, s_1), \dots, (r_n, s_m)\}$ .

There will be a possible electromagnetic connectivity between the sector  $s_j$  belonging to the site  $r_i$  and the sector  $s_k$  belonging to the site  $r_l$  when the distance between these sectors is less or equal than a predefined distance  $d_\gamma$ . This predefined distance  $d_\gamma$  will be estimated according to the calculations of the  $C/I$  ratio in the previous section.

### 3.1 Electromagnetic Compatibility Matrix

The Electromagnetic Compatibility Matrix  $C_{p \times p}$ , where  $p = |N|$  and  $|\cdot|$  is the cardinality, defines in a certain way the behavior of the communication network. It is a square matrix formed by  $p$  site-sectors in the network and establishes the restrictions required to assign channels.

Without loss of generality, the entries of the matrix  $c_{((r_i, s_j), (r_l, s_k))} = 1$  if the sector  $s_l$  of the site  $r_k$  is within the transmission range of a sector  $s_j$  belonging to the site  $r_i$ , and  $c_{((r_i, s_j), (r_l, s_k))} = 0$  in other case.

Let us define the set  $Ch = \{ch_1, ch_2, \dots, ch_q\}$ , where  $q$  represents available frequency channels which can be assigned to a sector.

### 3.2 Design Variables

A channel  $ch_\beta$  has to be assigned to a sector  $s_j$  of the site  $r_i$  this is denoted by the decision variable  $A_{((r_i, s_j), ch_\beta)}$ .

This decision variable is a dummy variable i.e.

$A_{((r_i, s_j), ch_\beta)} = 1$  if the channel  $ch_\beta$  is assigned to the sector  $s_j$  of the site  $r_i$  and  $A_{((r_i, s_j), ch_\beta)} = 0$  in other case. So a value equal to 1 in an entry of this matrix indicates that the pair of involved sectors can generate an unacceptable level of interference to the network if they use the same frequency channel. The proposed methodology of frequencies assignment takes the entries of this matrix as its first decision variable.

### 3.3 Set of Constraints

If according to the  $C$  matrix, two sectors may cause interference due to the use of the same channel, then a specific channel can be assigned to only one of those sectors. This is expressed as:

$$A_{((r_i, s_j), ch_\beta)} + A_{((r_l, s_k), ch_\alpha)} \leq 1 \quad \forall \{ch_\beta = ch_\alpha \mid c_{((r_i, s_j), (r_l, s_k))} = 1\}. \quad (6)$$

Only one channel can be assigned to one sector, i.e.:

$$\sum_{\beta=1}^q A_{((r_i, s_j), ch_\beta)} \leq 1 \quad \forall (r_i, s_j) \in R \times S. \quad (7)$$

The objective function is to minimize the use of channels, with respect to  $A$ ; it is defined as follows:

$$\min_A (A) := \sum_{\beta=1}^q 1 - \sum_{i=1}^n \sum_{j=1}^m A_{((r_i, s_j), ch_\beta)}$$

Now, we can define the integer-linear programming problem  $\Theta$ :

$$\min_A (A) := \sum_{\beta=1}^q 1 - \sum_{i=1}^n \sum_{j=1}^m A_{((r_i, s_j), ch_\beta)}$$

subject to (6)-(7).

The integer-linear programming problem that we present here is a convex problem therefore we warranty to find a global minimum, for more details please refer to [8].

## 4 TESTING AND RESULTS

### 4.1 Frequency reuse distance

In order to validate the estimated values for the frequency reuse distance some simulations were conducted. These simulations include the verification of the reuse distance as a function of the accepted  $C/I$  ratio levels or as a function of the wanted modulations. These simulations were performed using the professional network radio planning tool Mentum Planet 5.2 [7].

The simulation scenario includes three interfering transmitters I1, I2 and I3, all of them located at a distance of 9 Km from the transmitter Tx. An analysis of  $C/I$  ratio is performed on the 5Km x 5Km square, shown in Figure 3.

Using equations (4) and (5) calculations for the  $C/I$  ratio were performed, with  $N = 100$  radio links in each of the different positions of the receiver in the analysis area. With these results the better modulation in each link was defined in accordance with the requirements of minimum  $C/I$  ratio. Table 3 summarizes

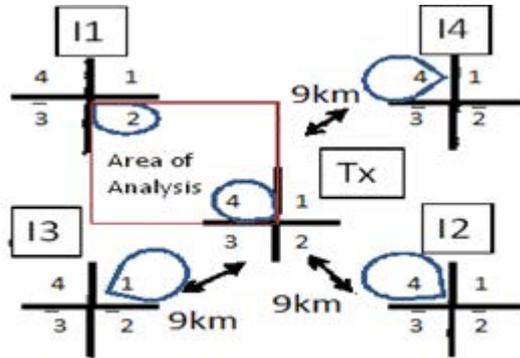


Fig. 3. Figure 3. Scenario for interference analysis.

	Modulation percentages for each modulation	
Modulation	Proposed estimation using SUI model	Mentum Planet
No link	12.39	1.89
BPSK 1/2	8.26	13.67
QPSK 1/2	4.13	8.48
QPSK 3/4	1.65	6.58
16QAM 1/2	0	2.92
16QAM 3/4	7.43	3.47
64 QAM 2/3	4.13	1.03
64 QAM 3/4	61.9	61.96

Table 3. Percentages of modulation usage in the analysis area.

the obtained results and presents also those results obtained with the Mentum Planet tool.

The obtained results confirm that there is high similarity between modulation levels that are estimated using the procedure described in Section 2 and the results obtained with Mentum Planet. It is important to point out that Mentum Planet uses different information layers in the propagation estimations. These layers include a digital terrain model and land usage layer (clutter). This additional information will allow more accurate propagation estimates. However it can be seen from Table 3 that the estimates for the higher modulations are very similar in both cases.

## 4.2 Frequencies assignment

Once the frequency reuse distance was defined, this information is fed to the process of frequency assignment to the sectors of the WiMAX network. In this case the defined reuse distances were set to 11 Km for type I2 interferers and to 9 Km for types I3 and I4 interferers. Thus it is expected that a percentage of about of 90% of the territory of the network coverage operate with a 64QAM 3/4 modulation.

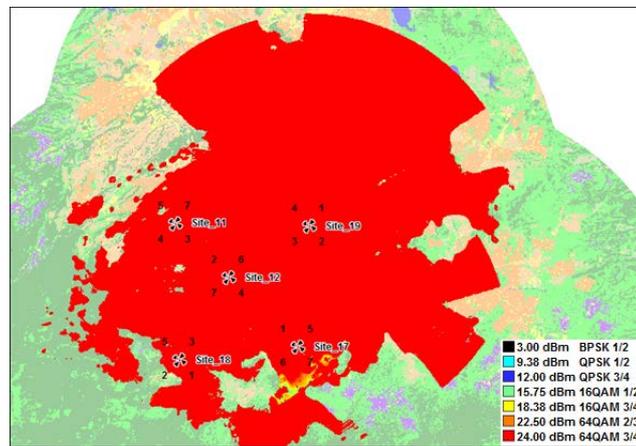
The following analysis is performed in order to validate the results obtained from the proposed methodology for frequency assignment. In this case the frequency plan obtained with the proposed methodology is applied to a network with 5 sites, each with 4 sector antennas of 90 degrees. It is considered that the network operates under the IEEE 802.16-2004 (fixed WiMAX). There are 7 available channels to be assigned to the sectors because of the booked 50 MHz bandwidth for the operation of the REESyG and because of the 7 MHz of bandwidth for the WiMAX channels. Table 4 presents the frequency assignment plan obtained with the proposed methodology and also the plan obtained with the Mentum Planet tool. The used frequency channels are named 1, 2, 3, 4, 5, 6 and 7 and they are distributed in the available 50 MHz to operate de WiMAX network.

In order to evaluate the performance of the network many parameters can be used, but the most representative one to assess the impact of the frequency assignment plan is the Carrier to Interference ratio  $C/I$  in the downlink. Figure 4 shows a Mexico City map, where the analysis scenario was placed, and the 5 sites are shown. This map shows the  $C/I$  ratio values in the downlink values over a clutter layer for Mexico City. The frequency assignment plan used to generate this analysis was that obtained with the proposed methodology. This map shows the different  $C/I$  ratio levels in the area with coverage from the WiMAX network.

As shown in Figure 4, much of the territory covered by the WiMAX network achieves high  $C/I$  ratios. However, in order to perform a more formal comparison between the frequency assignment plan resulting from the application of the proposed methodology and the frequency plan obtained with Mentum Planet, the Table 5 presents the percentages of different modulation obtained in the WiMAX network using both frequency plans.

Frequency assignment plans		
Sector	Assigned frequency channel with the Proposed Methodology	Assigned frequency channel with the Mentum Planet
1	4	7
2	2	3
3	5	4
4	3	5
5	3	6
6	7	4
7	4	7
8	2	2
9	4	5
10	2	7
11	3	6
12	1	1
13	5	3
14	4	1
15	1	2
16	4	5
17	1	1
18	4	2
19	5	3
20	3	4

**Table 4.** Frequency assignment using the proposed methodology and frequency assignment using mentum planet



**Fig. 4.** C/I ratio for the downlink channel for the analyzed WiMAX network for 5 sites.

	Percentage of modulation usage in the coverage area	
Modulation	Using Mentum Planet frequency plan	Using the frequency plan obtained with the proposed methodology
BPSK 1/2	0	0.01
QPSK 1/2	0	0.01
QPSK 3/4	0.01	0.04
16QAM 1/2	0.14	0.05
16QAM 3/4	0.3	0.14
64 QAM 2/3	0.14	0.05
64 QAM 3/4	99.41	99.7

**Table 5.** Percentage of different modulation levels for 5 sites.

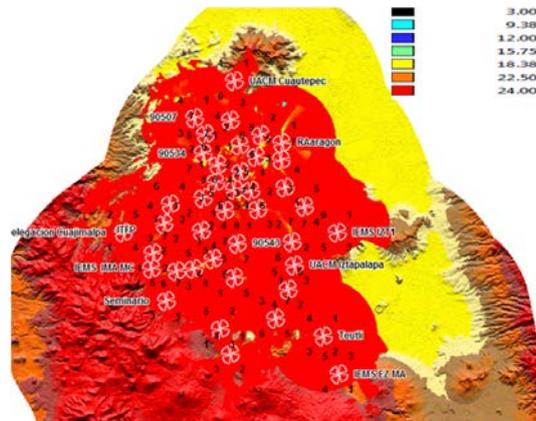
As can be seen from this table the percentages of use for the different modulations are very similar in both frequency plans. However it is important to note that in the frequency plan obtained from the proposed methodology uses only 6 channels as opposed to the proposal of Mentum Planet that uses all the 7 available channels. So in this case the proposed methodology offers a more efficient spectrum usage.

The proposed methodology was applied to a more complex network composed of 41 sites and 164 sectors. This network was designed to offer coverage to the majority of the populated zones of the Mexico City. Due to the number of sites, we decide to reduce the reuse distances to 9 Km for type I2 interferers and to 5 Km for types I3 and I4 interferers. The obtained results are shown in Table(6).

	Percentage of modulation usage in the coverage area	
Modulation	Using Mentum Planet frequency plan	Using the frequency plan obtained with the proposed methodology
BPSK 1/2	0.03	0
QPSK 1/2	0.06	0
QPSK 3/4	0.16	0.03
16QAM 1/2	0.22	0.14
16QAM 3/4	0.56	0.3
64 QAM 2/3	0.36	0.16
64 QAM 3/4	98.61	99.37

**Table 6.** Percentage of different modulation levels for 41 sites.

As we can see from Tables (5,6) in both cases, our proposed methodology offers a better efficient spectrum usage.



**Fig. 5.**  $C/I$  ratio for the downlink channel for the analyzed WiMAX network for 41 sites.

Figure 5 shows a Mexico City map, where the analysis scenario was placed, and the 41 sites are shown. The frequency assignment plan used to generate this analysis was that obtained with the proposed methodology. This map shows the different  $C/I$  ratio levels in the area with coverage from the WiMAX network for Mexico City.

## 5 Conclusions

The definition of the frequency reuse distance is very important to generate frequency plans to ensure that the network  $C/I$  ratio levels are maintained above a desired threshold. The frequency assignment methodology proposed offers similar performance level as the frequency plan generated with a professional tool for radio network design. However the proposed methodology achieves a more efficient use of the available spectrum.

## Acknowledgments

This work is carried out in the framework of the National Polytechnic Institute Project SIP 20120555 “Investigación de técnicas para mejorar el desempeño de sistemas de comunicación con canal directo y con canal de retorno”.

## References

- [1] [www.politicadigital.com.mx](http://www.politicadigital.com.mx)
- [2] Harry R. Anderson. Fixed Broadband Wireless System Design, Wiley. 2003.
- [3] Stavroulakis Peter, Interference Analysis and Reduction for Wireless Systems. Artech House. 2003

- [4] AN-80i Radio Platform <http://www.rdlcom.com/es/productos/an-80i-radio-platform>
- [5] V. Erceg et al., "Channel Models for Fixed Wireless Applications" IEEE802.16.3c-01/29r4, Broadband Wireless Working Group, IEEE P802.16, 2001.
- [6] M. Shahjahan, A. Q. Abdulla Hes-Shafi, "Analysis of Propagation Models for WiMAX at. 3.5 GHz," MS thesis, Blekinge Institute of Technology, Karlskrona. Sweden. 2009.
- [7] [www.mentum.com](http://www.mentum.com)
- [8] G. Sierksma "Linear and Integer Programming: Theory and Practice" Marcel Dekker, 2002.



# MANAGEMENT MODEL FOR AVAILABILITY AND INTERCONNECTION SERVICE IN CLOUD COMPUTING

Juan Ernesto Chávez Pacheco<sup>1</sup>, Chadwick Carreto Arellano<sup>2</sup> and Salvador Álvarez Ballesteros<sup>1</sup>

<sup>1</sup> Escuela Superior de Ingeniería Mecánica y Eléctrica, Unidad Profesional Adolfo López Mateos, Av. Instituto Politécnico Nacional s/n, Edificio Z-4, Tercer piso, Colonia Lindavista, Zacatenco, México D. F., C.P. 07738  
jchavezp0400@hotmail.com, salvarez@ipn.mx

<sup>2</sup> Escuela Superior de Computo, Unidad Profesional Adolfo López Mateos, Av. Juan de Dios Bátiz s/n esquina Miguel Othón de Mendizabal, Colonia Lindavista, Zacatenco, México, D. F., C.P. 07738  
ccarreto@ipn.mx

**Abstract.** This paper presents a model for the management and improvement of Quality of Service (QoS) Availability of Interconnection and Access in the Cloud Computing, it's an automatic mechanism of connectivity that allows all services provided by a "Public Cloud" are most of the time available to users when they need it. This model is based on a layered architecture to allow users to access their services in the cloud no matter if there is any problem of interconnection.

**Keywords:** Architecture, Mobile Computing, Cloud Computing, Communications, Availability, Networking and Services.

## 1 Introduction

One of the emerging trends that are observed at this time and that will mark the field of Information Technology and Communications (ICT) in the coming years is what has come to be known as cloud computing or cloud computing [1].

According to the Cloud Security Alliance (CSA) Cloud computing is a demand model for the allocation and consumption of computers, which describes the use of a number of services, applications, information and infrastructure composed of resource reserves computing, networking, and storage information [2].

The cloud computing is the convergence and evolution of several concepts related to information technologies, such as virtualization, distributed application design or the design of networks [3].

This allows the information is no longer necessarily in has to be stored on computing devices, but on the systems provided by their own "cloud" computing resources and that are shared by multiple users across different devices and can work together over the same content; achieving a greater agility and cost efficiency in the management of information.

While it is true that the availability is an advantage, currently is a concern because it is a responsibility that falls solely to the service provider as well as the quality of information management which includes operations as extraction, handling, treatment, purification, conservation, access, etc. For this we have proposed creating a model that will provide a solution to this problem.

The structure of this work consists firstly of a current panorama of the problem and subsequently showing the proposed model, its characteristics and a brief description of the layers that comprise it; in the following section describes the architecture that has been developed for model; immediately is found the design and installation of the cloud using ownCloud; subsequently are shown some preliminary tests and the results obtained; and finally presents the conclusions and future work of our model.

## 2 Connectivity Model

The system in "cloud" is designed to be used remotely, so any user will have access to most of the systems wherever you are. So in terms of the availability, this is negotiated with a supplier through a service level agreement that specifies a guarantee of continued access to services with a minimum quality [4].

In Fig. 1 shows the proposed model, which aims be fitted with a standard character to be implemented on different architectures of "cloud" that will optimize the services of availability, connectivity, reconnaissance and monitoring of the link status and of the users.



**Fig. 1.** Management model for cloud computing.

As you can see our proposed model is composed of five interconnected layers. Some of the most important features of the connectivity model for the cloud computing are:

- Must be suitable for different devices no matter what operating system platform beneath which these operate.
- Must operate with different wireless technologies permitting users access to services that are associated with it, regardless of the type of technology that uses the device.

- It should provide users freedom of movement within any environment.
- Must offer services to users in a transparent, without the user cares about the location of the service provider.
- It must be handled through a friendly interface that enables users to connect to the system to easily and simple.
- Based on your structure should be susceptible to changes or modifications to enable continuous improvement of this.
- Must be able to grow depending on the needs of the environment.

Then it describes the operation of model layers, which will manage the availability of the services offered in the cloud.

### **2.1 Interconnection Layer**

This layer is responsible for making the user's mobile device detects the Cloud Network existing in the area and have the possibility of entering the environment or domain of the cloud, for this, there will be a point of interconnection of the domain that will be responsible for allow access to the Cloud and therefore all services, applications and other resources that in provide it.

### **2.2 Validation and Identification Layer**

When a user performs the connection request, you will be asked to enter your name and password, this information along with the direction of the interconnection point where the user wishes to enter the network, should be sent to a server for authentication [5].

This server would be responsible for verifying that both the name and password sent coincide with those registered in its database, so you will be assigned an address to the user devices and can access the corresponding domain; while if the information provided by the user is incorrect will be asked to forward your name and password.

### **2.3 Layer Management of Information**

The management of services is provided according to the profile that users have within the environment, which caters to each according to your servicing requirements and information [6], and through a graphical interface the user selects one of the services and applications to which you're entitled, access to these is via a client - server mode.

The user can select from their list of the services and applications you need, once you select the application establishes a connection between the user and the server, the server processes the service and sends a response to the user through a graphical interface through its mobile device.

## **2.4 Layer Management of Information**

It aims to track the consumption of cloud resources for users through a database of services and user profiles in the type of service that is provided, when he began to use it and when finished, in addition to collecting and analyzing traffic. [7].

Instantly of arrive at this layer, take place a detailed analysis of: The number of users accessing and the services they need, knowing whether they are taking advantage of resources such as the bandwidth, know whether each service consumes the bandwidth you need, user or service that generates more traffic than they may require, if the power levels, of noise and interference are tolerable, etc.

It aims to track the consumption of cloud resources by users, through a database with the services and users profiles in the type of service that is provided, when he began to use it and when finished, in addition to collecting and analyzing traffic.

## **2.5 Layer of Correction Techniques**

Here is carry out the procedures and protocols to reverse the errors detected by the previous layer during transmission of the information, services and applications. It should establish a set of rules that allow adjustment of the signal power, tolerable levels of noise and interference, the bandwidth consumed by an application, the proper management of synchronization and data frames, and define procedures to determine when an error has occurred and how they should be corrected [7].

For the treatment of errors during data transmission can be followed two different strategies, such as including redundant information so that the user device requested by the retransmission of the information or allow you to deduct that was wrong and correct it, without asking for retransmission the information sent [8].

## **3 Architecture of the Model**

The Architecture of the Model, it is very important as it defines as interacting the elements that make up the model and they are supported by some standards, norms, protocols, processes, rules, etc., in order to make the model fully functional has been proposed [9].

The interaction you have with the end user is central to the objectives of the model, because here is where it the user feels comfortable and fully familiar with the management of services in addition to offer them in the time and manner in which the requested, the architecture proposed model shown in Fig. 2.

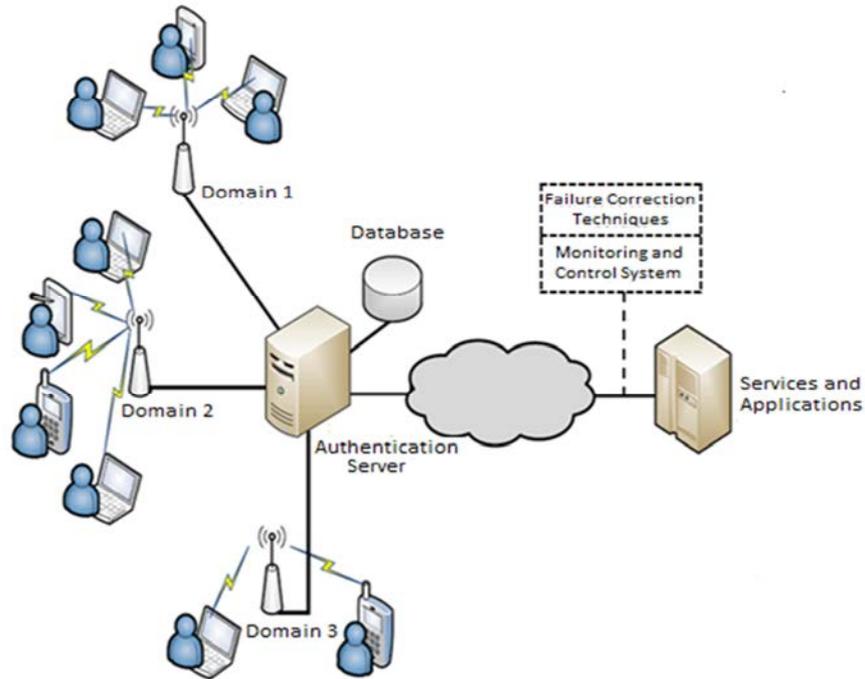


Fig. 2. Architecture of the model of connectivity.

In general the architecture of the model of connectivity operates as shown in the sequence diagram of the model of Fig. 3, and then explains the process to be followed by a service request in order to make the allocation of services and applications in the cloud to end users.

- At be within wireless coverage area of a station is necessary to use the Network recognition, where the user through their device and interconnect device from the domain of "Cloud" exchange information to establish a connection.
- Then proceed to the validation, identification and authentication, the user identifies itself to a server that authenticates against a database of users to be able to begin the Request for Services that offered by the environment.
- The Authentication Server shall be responsible for verifying that both the username and password match recorded in its database.
- If the information provided by the user is incorrect, it does not will have access to cloud, while if the authentication is successful, and will be assigned an IP address to the user device and you can access the domain.
- Once the user has connected and is know the location of this in the network, a Web server is responsible for providing a whole list of services associated with a database of the user profile.
- The submission of the applications and services is carried at the time the user opens the Web browser on your device.

- Returning to the provision of services, the user must select one of the services and applications that are available, for this establishes the connection with the services server where these are found so as to process it and deliver it to the user.
- Finally, the two upper layers are responsible for monitoring and control of the state of the network and users by the use of the corresponding Technical Testing and Performance Monitoring Techniques, and so then that should be required to implement the respective Technical Correction Fault.

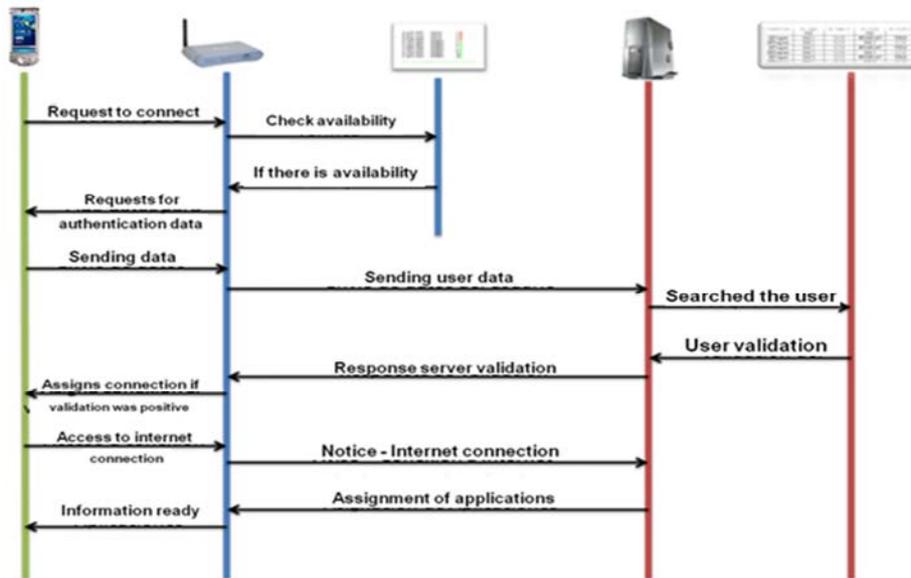


Fig. 3. Sequence diagram of the model.

Mastering the Cloud is open to anyone who is within their reach and should be sent periodically throughout the Interconnection Point a package Domain Identifier (SSID) that users can be detected through their personal devices. So then, the model is designed so that various Wireless Technologies can interoperate with each other, so this is not a barrier in communication [10].

#### 4 Design of the Cloud

The proposed connectivity model is developed within the Mobile Computing Lab of the School of Computing of the IPN, because in this exist the means and facilities to develop and conduct the necessary tests to define its operation and functionality.

So far we have installed a Server Cloud Computing using ownCloud, eliminating the limitations in terms of capacity and we will have all the free space we have on your hard disk to storing information and as the cloud administrators we can also assign disk quotas to other users.

#### 4.1 Cloud Server Installation with ownCloud

First you must start by starting our server under Ubuntu Server operating system, which uses a Linux kernel and based on Debian, to proceed to install Ubuntu Server, first you must download the ISO from the Ubuntu website.

<http://www.ubuntu.com/download>

After the computer boots from our CD, USB or ISO image as the case, select the language in which the program will be installed, a window appears immediately of welcome to installation wizard (Fig. 4), and then must complete all the steps indicated by the installation wizard.



**Fig. 4.** Menu ubuntu server installation.

During the installation of Ubuntu Server defines a user account and password for the system, the disk partitioning, configuration of updates, programs that accompany the base system, the GRUB boot loader and finally reboot.

Upon completion of the installation into the system, this will enter the login and password you chose during installation, if the data is correct must be displayed one home screen like Fig. 5.

The server automatically assigns IP address dynamically, so it is must configure the network card to assign a fixed IP. For this opens and edits the file "interfaces" using the following statement:

```
sudo nano /etc/network/interfaces
```

Is written fixed IP address, the subnet mask, the gateway address and the DNS (see Fig. 6), then save the changes and exit the editor.

```
Welcome to Ubuntu 12.04 LTS (GNU/Linux 3.2.0-23-generic-pae 1686)

* Documentation:  https://help.ubuntu.com/

System information as of Tue Jul  3 17:09:43 COT 2012

System load:  0.0          Processes:      62
Usage of /:   9.7% of 9.23GB Users logged in:  0
Memory usage: 5%          IP address for eth0: 148.202.159.7
Swap usage:  0%

Graph this data and manage this system at https://landscape.canonical.com/

57 packages can be updated.
26 updates are security updates.

The programs included with the Ubuntu system are free software;
the exact distribution terms for each program are described in the
individual files in /usr/share/doc/*/copyright.

Ubuntu comes with ABSOLUTELY NO WARRANTY, to the extent permitted by
applicable law.

cel@cel-proxu:~$
```

Fig. 5. Home screen of ubuntu server.

Finally, you must restart the card and using the command "ifconfig" will verify that the changes have been made.

```
sudo /etc/init.d/networking restart
ifconfig
```

```
GNU nano 2.2.4      Archivo: /etc/network/interfaces      Modificado
# This file describes the network interfaces available on your system
# and how to activate them. For more information, see interfaces(5).

# The loopback network interface
auto lo
iface lo inet loopback

# The primary network interface
auto eth0
iface eth0 inet static
    address 192.168.1.100
    netmask 255.255.255.0
    network 192.168.1.0
    broadcast 192.168.1.255
    gateway 192.168.1.254
    # dns-* options are implemented by the resolvconf package, if installed
    dns-nameservers 192.168.1.1
    dns-search server.com_

^G Ver ayuda  ^O Guardar  ^R Leer Fich ^V RePág.   ^K Cortar Tex ^G Pos actual
^X Salir     ^J Justificar ^W Buscar   ^U Pág. Sig. ^U PegarTxt  ^T Ortografia
```

Fig. 6. Configuring the network card.

Once installed Ubuntu Server, proceed to the installation of ownCloud, but first you must be install the following programs: apache2, php5, php5-mysql, libcurl3, php5-curl, php-pear, php-xml-parser, php5-sqlite, php5-json, sqlite, mp3info, curl, libcurl3-dev and zip; using the following statement:

```
sudo apt-get install [program name]
```

Once installed in the server root directory ownCloud downloaded, then unzip the package, once permission is given the server web directory ownCloud and finally apache restarts, using the following code:

```
sudo wget http://owncloud.org/releases/owncloud-2.tar.bz2
sudo tar -xvf owncloud-2.tar.bz2
sudo chown -R www-data:www-data owncloud
sudo /etc/init.d/apache2 restart
```

Now the browser from another computer on the network is keyed the command below.

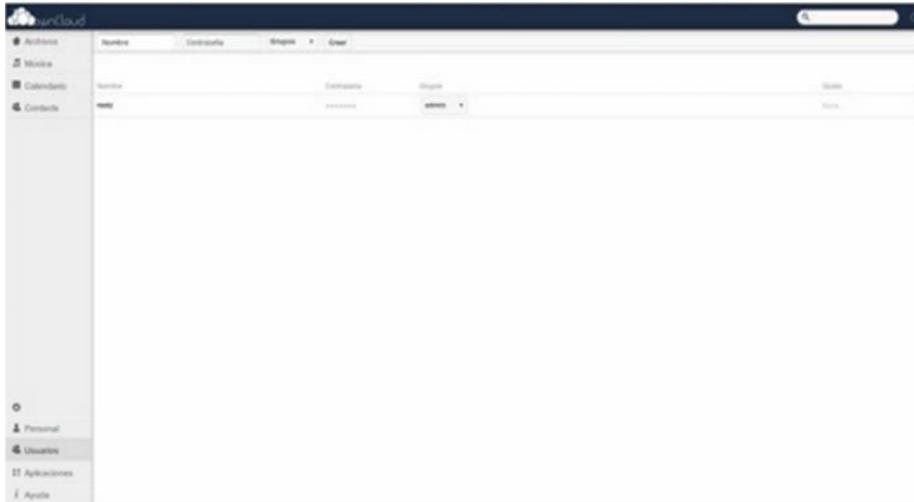
```
http://IP.SERVER/owncloud or http://DOMAIN/owncloud
```



Fig. 7. Screen owncloud.

You create an administrator account by entering a username, a password and select "Completing the installation". For access to the data is a little safer creates a user with fewer privileges than the administrator, this user will be the one used to connect remotely to the Cloud.

To create users open the settings menu located on the bottom left of the user manager, select the "Users", you enter a new username and password, select the "Groups", select the option "add Group" must be entered a group and finally pressed "Create".



**Fig. 8.** Window management owncloud users.

## 4.2 Support for WebDAV

To use WebDAV support the user created in the web environment is also created in the system using the command "adduser user" done you can proceed with the installation.

It installs and configures davfs2, select "Yes" and are given permissions to users so they can use the cloud:

```
sudo apt-get install davfs2
sudo dpkg-reconfigure davfs2
sudo usermod -aG davfr2 USER
```

Then edit the file "fstab" to include the user's mount point and cloud, web path is inserted if they have a domain pointing to our server or if the local IP, insert the mount point within the /home user and permissions, then save the changes and exit the editor.

```
sudo nano /etc/fstab
http://[IP or domain]/owncloud/files/webdav.php home/USER
/ownCloud davfs user,rw,noauto 0 0
```

Then create two directories, at one is edited and saved the key file, we left editor and assign permissions.

```
mkdir /var/www/.davfs2
mkdir /var/www/owncloud
sudo nano /var/www/.davfs2/secrets
http://[IP or domain]/files/webdav.php USER PASSWORD
chmod 600 var/www/.davfs2/secrets
```

With this has finally installed a Server Cloud Computing with ownCloud ready to be used, thus eliminating the limitations in terms of capacity because it has all the free space you have on your hard drive where everything is staying only under the administrator's control; managing their security, privacy and ensuring the confidentiality of data.

## 5 Tests and Results

Tests have been made file storage in the cloud with registered users to date via web connecting through the above address or also using the Nautilus file browser that comes installed by default in almost all GNU/Linux and default in Ubuntu. In the Nautilus file browser, select the File menu and choose the "Connect to Server", and a window appears like Fig. 8.



Fig. 8. Connection window.

We press the tab "Public FTP" and select "WebDAV (HTTP)" in the "Server" enter the server IP, port is left like this and "Folder" you enter the path /ownCloud/files/webdav.php (see Fig 9).



Fig. 9. Connecting to the cloud server.

Finally insert the username and password of the user who will be using the Cloud, press the button "Connect" and see that the Cloud is mounted as a network drive, as shown in Fig. 10.

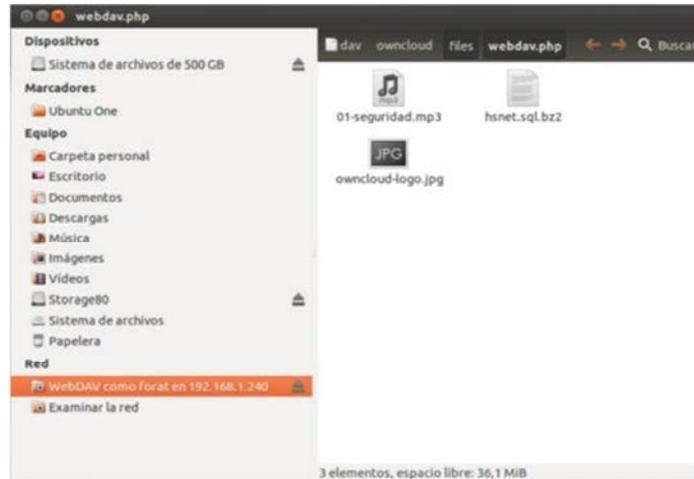


Fig. 10. Using cloud owncloud with nautilus in ubuntu.

At the same time we are working on developing an application for mobile devices with the purpose of provide the user through a graphical interface that provides a service the possibility to save, download and view certain types of digital files using the cloud which has been mounted thus will be more simple, flexible and dynamic access to information.

For the moment in Fig. 11 shows the GUI for Android application with file management locally and server, as well as uploading and downloading of files that you have so far.

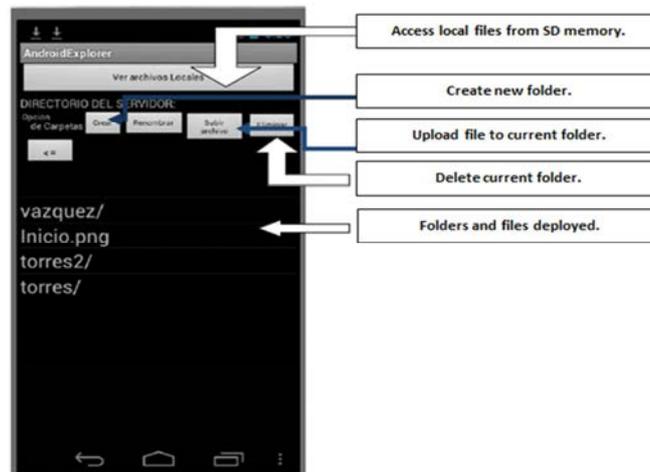


Fig. 11. Application for mobiles.

In Fig. 12, displays all functional tests conducted on the implementation of Android, which has been obtained so far is the ability to create and delete folders, also browse and select files from a folder, and upload, download and delete files in the cloud, in the same way you can delete the contents of an entire folder generally within Cloud Sever that has been mounted.

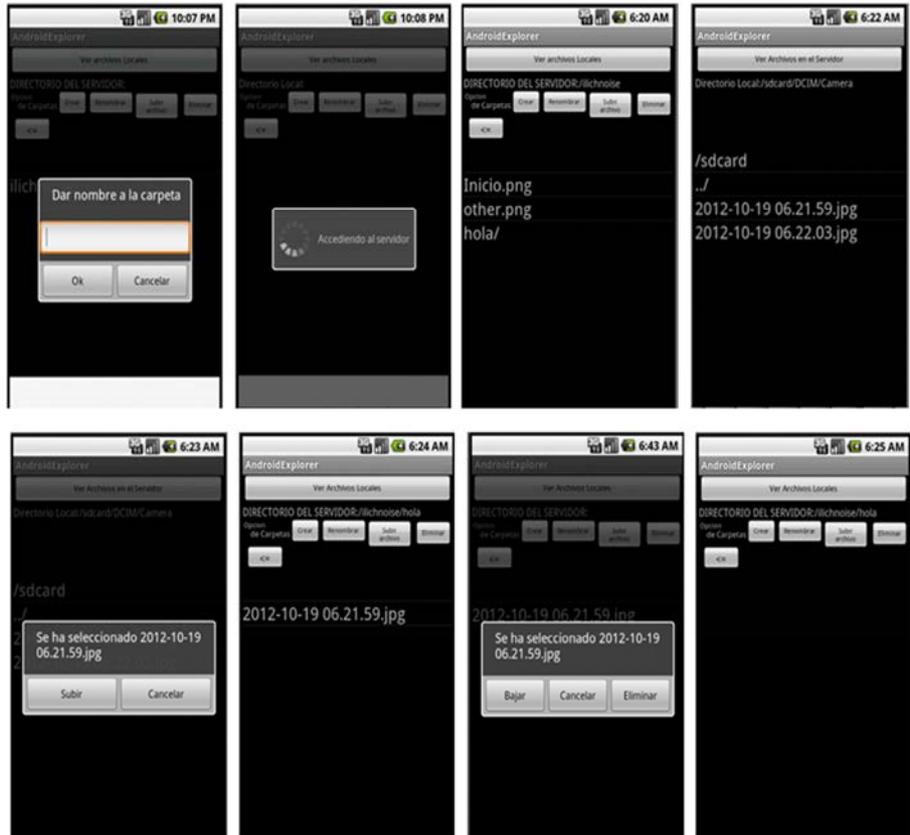


Fig. 12. Screens android application.

As seen before to consult or requesting the services, users must firstly contact the administrator of the cloud to register in the Cloud Server with ownCloud to give access to the related information according to their profile or they are can be provided a user name and password.

To this end one of the main concepts proposed model is accompanied on the prioritization, is to say that will have to give more importance to some connections than others.

Also, the model is not designed to work with a specific access technology, however for ease and comfort has chosen to work with a single wireless access technology that is WiFi.

## 6 Conclusions and Future Work

We can conclude that the results obtained so far are satisfactory since through ownCloud package we have create a storage area in the cloud, and thus get manage our own server and dispose of it in the cloud very easily.

Currently work is still in the stage of design improvements of the Cloud and preliminary tests of model we have proposed, while continuing with the proper documentation.

It also seeks to use some software or package with the functions of monitoring, detection and solution, to be in charge of analyzing the correct behavior of the communication between the cloud and users in various aspects, as well as the relevant decisions.

The next work that has obviously considered the implementation of the model will be run within some case study, which finally throw an idea of the functionality of the model and thus will issue the final conclusions.

In summary, the main reason for proposing and implementing this model, the purpose of which is convertible into a flexible and robust, which instantly intone be implemented within a public cloud can allow users to have a "Access "quality to the corresponding information at the place and time they need, in a more efficient, easy and fast, through their personal mobile devices and thus may be present in both labor and business environments.

## Acknowledgements

The authors of this paper thank the IPN and SIP support for the development of this research, and most especially to the Section of Postgraduate Studies and Research ESIME Zacatenco and ESCOM, as well as people and lab partners involved in this project for their time, their work and their valuable comments during the preparation of this document.

## References

1. L.J. Aguilar.: "La computación en Nube: El Nuevo Paradigma Tecnológico para Empresas", Revista cuatrimestral de las Facultades de Derecho y Ciencias Económicas y Empresariales. Mayo (2009).
2. Cloud Security Alliance, <http://www.cloudsecurityalliance.org/>
3. Oracle Corporation. "Architectural Strategies for Cloud Computing", Oracle White Paper in Enterprise Architecture. Agosto 2009. [Online]. Disponible en: <http://www.oracle.com/>
4. Cloud Expo, <http://cloudcomputingexpo.com/>
5. Xipeng Xiao.: Technical Commercial and Regulatory Challenges of QoS an Internet Service Model Perspective. Morgan Kauffman, Burlington (2008).

6. Farrell Adrián.: Network QoS: Know It All. Morgan Kauffman, Burlington (2009).
7. Wang, Henry H.: Telecommunications Network Management. McGraw-Hill, University of Michigan (1999).
8. Hegering, Heinz-Gerd.: Integrated Network and System Management Network Management. Addison -Wesley, University of Michigan (1994).
9. Oracle Corporation Architectural Strategies for Cloud Computing, <http://www.oracle.com/>
10. The National Institute of Standards and Technology, Information Technology Laboratory, Definition of Cloud Computing, <http://csrc.nist.gov/>



# **Databases & Software Technology**

---



# Integration of the POI API with Java for Information Processing from Heterogeneous Sources

Pedro Galicia Galicia

Computing Research Center. National Polytechnic Institute.  
México City, México  
pgalicia@cic.ipn.mx

**Abstract.** Nowadays, there is a growing need for people to search for specific information. Frequently, the relevant information is stored in different formats such as Excel, Access, Word, etc. An additional complication is that diverse data can be located in a single file, and pieces of the same information can be contained in different files. Often people that accomplish data processing tasks have to extract, transform, and load information from Excel format files to a database environment. The Apache POI Project offers APIs to access files in the Microsoft OLE 2 Compound Document format, including .XLS format files. To solve the problem of interoperability between Excel and databases, we propose to integrate this API with Java technology, obtaining a technological combination powerful and easy enough to be combined with existing applications. We consider the case of information management and dynamic retrieval in an enterprise, in a setting where several departments work together within an integrated information environment.

**Keywords:** Information process, ETL process, Java, EXCEL, Apache POI API, MySQL, XAMPP, TOMCAT, JDBC.

## 1 Introduction

Since many years ago enterprises have had an urgent need for data integration related with areas such as finance, management, production planning, production and sales. Currently there is a great demand from the users to obtain accurate and timely information for decision-making. Often people that accomplish data processing tasks have to extract, transform, and upload information from Excel format files to a database environment. All of this is because the data processing environments tend to be more complex, and decentralized, and have to incorporate flexibilities and the modularity, using new technologies. To focus these requirements is needed to innovate and generate new forms of data processing considering proper handling of the information, and Excel-Java-Database combination is another option. The extraction of data residing in Excel format files, as a result from business activities, to process it and upload it into a corporative database is the main goal.

adfa, p. 1, 2011.  
© Springer-Verlag Berlin Heidelberg 2011

## 2 The information analysis

Enterprises base their performance on the information management that different areas produces and for this reason have to process a lot of data contained in several files such as Excel format files, and as result of this, users have to manage many data in different ways. The problems arise when the information contained in Excel format files is required for several areas at the same time and the data into spreadsheets turns in very huge files to be interchanged.

Hence information that resides in those Excel format files is copied into different computers with the risk to be processed or updated in a wrong way. This can create a disagreement conflict between the enterprise areas about the decision making that can be solved with a corporative database; however there is a situation, the interoperability among Excel format files and the Databases.

To solve this problem we propose to integrate the POI API with Java technology and a Database manager as MySQL to obtain a technological powerful and easy enough combination jointly with existing environments as Apache Web server and TOMCAT application server for instance. Therefore to let's briefly describe the Excel files and features of POI API.

### 2.1 Excel Files

**Excel Files.** These are files based upon Microsoft format to files Compound Document format OLE 2 (OLE2CDF) [1], and OLE2CDF file contains a complete filesystem, laid out using nested Entries Directory, that contain Entries. We are interested in Entry elements of the Entries Document type and an Entry Document contains data structures of an application-specific (e.g. Excel) [2].

This kind of files are known as workbook type, and each workbook can include multiple spreadsheets either empty or with information. These document types are known as Excel format files, and have an extension type ".XLS" or "XLSX".

Excel files supports two basic types of spreadsheets to contain text or numerical data, and can be used to develop calculations and to contain formulas or graphs, also are used to display information by using multiple schemes charts (bar charts, pie, 3D, etc.).

**Spreadsheet Documents.** This type of document basically consists of rows and columns, which contain several data types, for instance the case of a normal spreadsheet that can contain values and formulas, graphs, images or macros.

These documents also contains a structure and its records are bounded mainly by the type BOF (Begin-Of-File) and EOF (End-Of-File), also includes information about dimensions, view, fonts list that contains, list of names and external references, and the format of the cells, the width of the rows and the height of the columns.

## 2.2 POI API

**POI API.** The Apache Software Foundation [3] includes into its Project Listing, the Apache POI's Project, the mission of this last one is to create and maintain JAVA APIs for manipulating various file formats based upon the Office Open XML standards (OOXML) and Microsoft's OLE 2.

The Apache POI's Project, contains the POI (Poor Obfuscation Implementation) API [4] this is based in a pure Java API and allows manipulate various file formats. This API facilitates the reading and writing Excel (HSSF), Word (HWPF) and PowerPoint (HSLF) files.

For purposes of this paper we will focus on the part of HSSF and its interaction with POIFS component, needed to create or read documents .XLS or .XLSX.

The module HSSF (Horrible Spread Sheet Format) provides several packages, as shown in Table 1, and allows interacting with Excel type documents and also creates other new ones.

**Table 1.** HSSF Packages

Package	Function
Org.apache.poi.hssf.eventmodel	Handles different events generated in the processes of reading and writing Excel documents
Org.apache.poi.hssf.eventusermodel	Provides classes for the process of reading Excel documents
Org.apache.poi.hssf.record.formula	Contains classes to handle FORMULA used in Excel documents
Org.apache.poi.hssf.usermodel	Contains classes for creating Excel documents
Org.apache.poi.hssf.util	Common environment to handle different kinds of documents attributes Excel

Figure 1 shows an excerpt from the HSSF module and its Usermodel class diagram.

Another use of the Apache POI API is for text extraction applications such as web spiders, index builders, and content management systems.

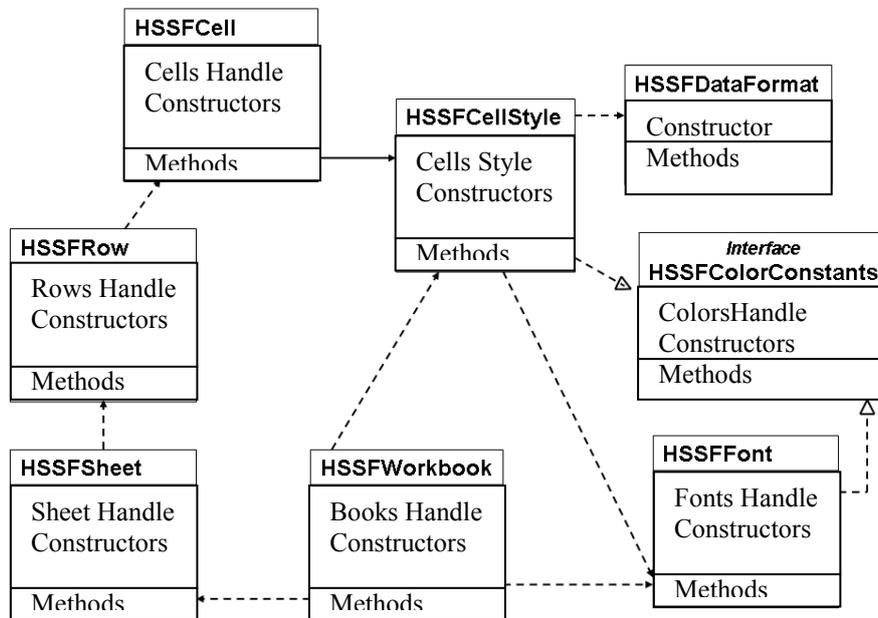


Fig. 1. Excerpt from the Usermodel Class Diagram HSSF<sup>1</sup>.

Important aspects:

- The workbook is represented by the HSSFWorkbook class.
- Each worksheet in the document is represented by the class and creates HSSFWorksheet from HSSFWorkbook class through: `object_HSSFWorkbook.createSheet ("Sheet name")`.
- Before making a cell reference must reference the row.
- The reference to a row is represented by the class HSSFRow.
- The row is obtained from the sheet (HSSFWorksheet).
- Cell HSSFCell is obtained from the row (HSSFRow).

Create a book involves several actions as create sheets, and each sheet creates rows, and cells, and settings to contain data, for instance such as the formula's type for a Total or a region for data capture, finally the created document must be saved.

<sup>1</sup> Usermodel Class Diagram by Matthew Young. [poi.apache.org/spreadsheet/diagram1.html](http://poi.apache.org/spreadsheet/diagram1.html)

Although there are other applications [5] to read and write Excel format files, for the problem posed in this article we have chosen the option to develop an application based in the combination of POI API with Java, jointly with MySQL and supported in a Web environment.

### **3 Design of the application**

So far we have presented an overview of the features of Excel format files and POI API, and we have mentioned others development environments based on the management of these format types. Now, let's set the initial specification of the application, how to get the input data, user profiles and technological support required as well as the objectives design and to identify the actors involved.

#### **3.1 Initial specifications**

The main goals of the application are: to retrieve data from Excel format files, to upload their references into a MySQL database, to process the retrieved data and update the relevant information into the test database, to perform information queries, to manage access according to the user profile, as well as the users registration and the Excel resource registration too.

Therefore, it is necessary to define and configure the test database to incorporate information from Excel files and from the users; as well as the design and develop the graphical interfaces for communication client/server, the reports, queries and the data validation. At this step the application development [6] considers: the requirements analysis, design, development, testing and install, additionally could be included the maintenance and operating manuals.

#### **3.2 Graphical User Interfaces (GUI)**

GUI should have:

- Security, users access the application via a user name and password respectively registered in the database, this will give them access according to the user profile.
- Users management, users have to register data such as user name, user profile identifier (login) and password (password).
- Interaction with the user must have Web interfaces in each modules of the application.
- Show the results of the updating, modifying, or deleting records process, respectively.
- Show options for the start and stop of a user session.
- Queries information.

### 3.3 Technological support

An application development like this considers the following elements: have a connection to a public or private network communication (Internet), a GUI scheme of access via a Web Browser, a test database MySQL, usage the POI API, the Java language programming [11] and its specifications JSP and JDBC, an a based pattern design architecture as the Model-View-Controller, as well as a MySQL database manager, a XAMPP Web Server, and a TOMCAT Application Server, and for this work the web and the application server will be installed at the same computer.

### 3.4 Application Architecture

When developing applications with Beans and JSPs, is required to identify and separate the presentation logic and business logic, for which we need to use a standard software architecture that separates the application data, the user interface, and the logic control into three distinct components, the architecture pattern is called Model-View-Controller (MVC) [7].

Generally, while the application is operating via GUI via the Web server [8], the application server [9] considers environment operating requirements, and business logic is provided by the application components. In fact, these last ones components perform user validation and where appropriate, the needed tasks to establish the connection to the MySQL [12] database manager to execute the respective operations in the database environment. Figure 2 depicts the operation of the architecture considered for this application.

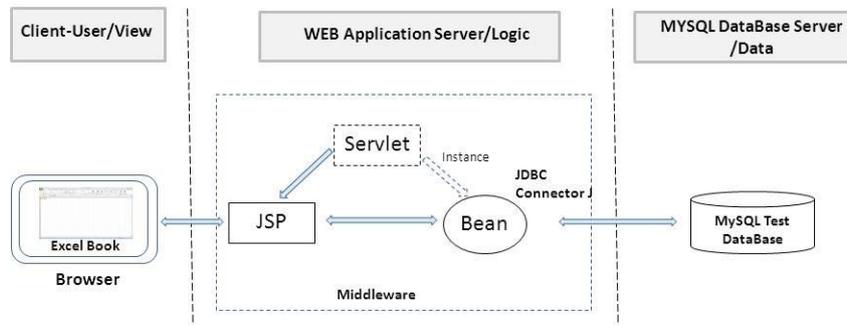


Fig. 2. Application Architecture

Modeling the Test Database [10] is considered at the corporate level as a study case and focused to defining appropriate information resources.

### 3.5 Application Packages

Several packages are part of this application for the management and dynamic information retrieval from heterogeneous sources (MDIR). These packages are composed of different classes that provide the necessary functionality to the application. The corresponding diagram is shown in Figure 3.

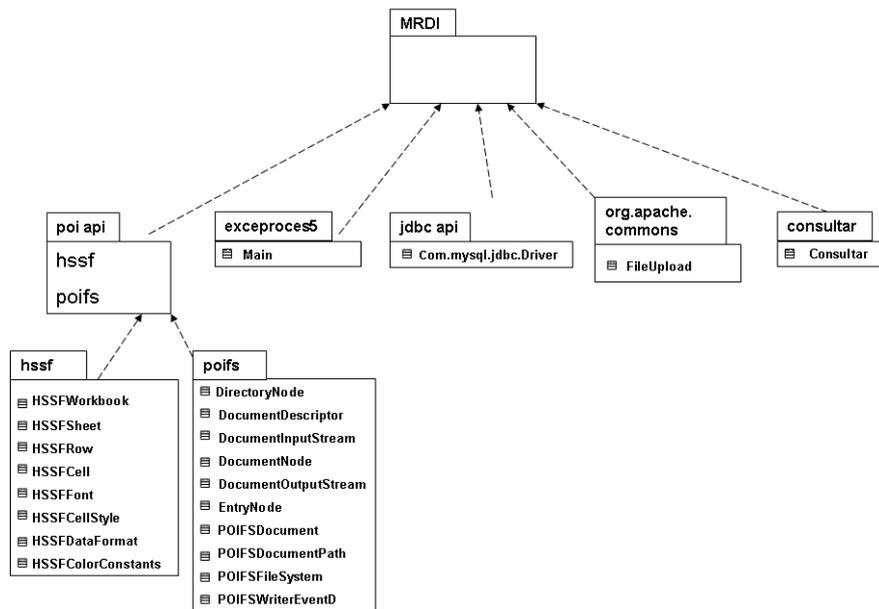
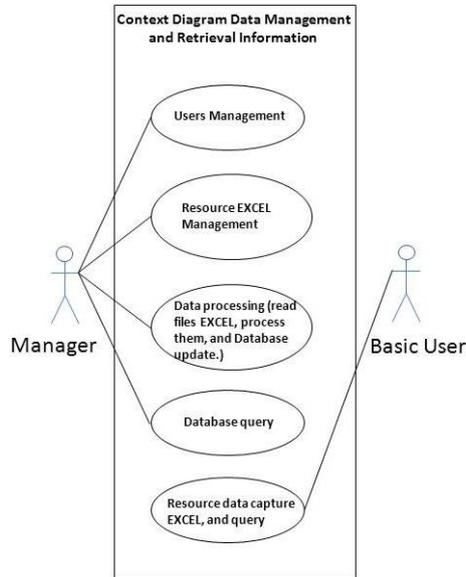


Fig. 3. Diagram packages that integrate the application

## 4 Application Development

The purpose is to solve the problem mentioned at Initials Specifications paragraph. For this application has defined two types of actors [13]. The first one is the basic user to use the Excel module resources, and the second one is the management user that can use the other modules that make up the application.

The following context diagram shows the use cases for, see Figure 4.



**Fig. 4.** Context Diagram Data Management and Retrieval Information.

**4.1 Outline of the operation**

The flow pattern information processing includes: the data collection process and data capture in Excel format files (events held outside the application), the access to the application via the Web browser, to perform user authentication via the login and password processes. When the user is a basic user only can access the Excel module resources to data capture and upload the information that resides in Excel files to the database and consult. When the user is a manager user can use the management, processing and queries modules, The Management module allows the users management and the Excel resource management, the Processing module allows the user to choose the Excel files that will be read and to execute the process for retrieving data and updating the relevant information into the test database, and the Queries module allows to generate the queries for each user.

The process operating diagram shows the dynamic part performed into a web environment in which to start a session, the user must be validated according to its level of access, see Figure 5.

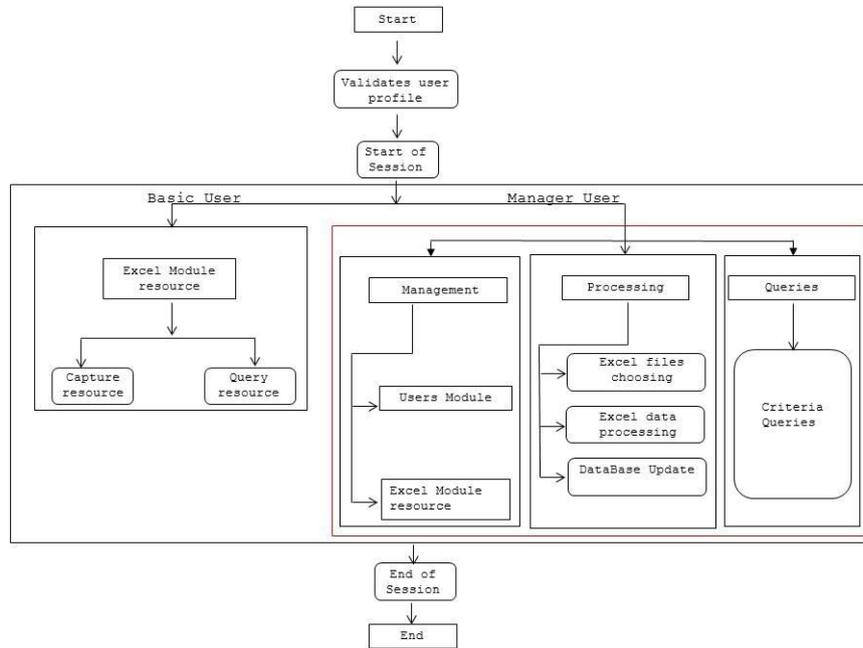
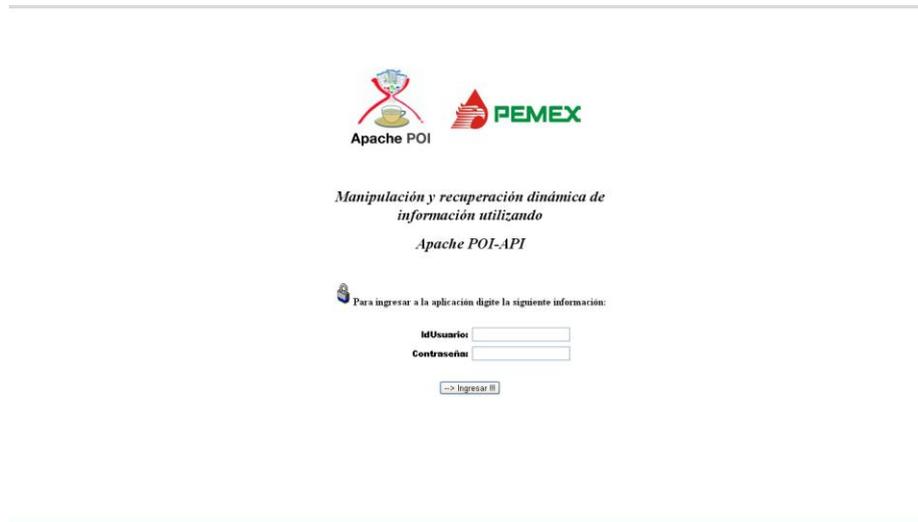


Fig. 5. The process operating diagram that shows dynamic part of the process

## 5 Test and Results

This section presents the graphical interface for access the application as well as some the modules that make up. Now let's see the operation tests and results assessment. Here we have a corporate testing database, which in fact has already been modeled, and which we have added new tables related to user management, Excel resource, and other components like budgets. Figure 6 shows the graphical interface to access the application, which validates the user name and password against the database.



**Fig. 6.** Graphical Interface Access

The upload of Excel files to database is important because this task has to be checked it up to verify the load time of the Excel file to the server. The upload time can be checked from both a local and remote environments that emulates the uploading time. For this test the respective code was implemented to the loading process to measure the size of the latter. Table 2 shows uploading results.

**Table 2.** Uploading time to the server of resource files Excel.

No.	File Name	Size (KB)	Local uploading time (seg.)	LAN uploading time (seg.)
1	Capturador_prueba1	1,851	0.015	0.016
2	Capturador_prueba2	1,979	0.016	0.016
3	Capturador_prueba3	1,995	0.016	0.018
4	Capturador_prueba4	2,110	0.032	0.035
5	Capturador_prueba5	1,853	0.015	0.016

Figure 7 shows the graphical interface for import Excel data, where the user choose files to use and the application displays the path the file as well the information retrieved, and the amounts of the global costs, which are shown in red color, to be upload into the database.

**Valores importados del capturador de Excel**

---

Fecha: 23/10/2012			
<b>Costos (en millones de pesos) y cantidad de eventos</b>			
Ruta del Portafolio que contiene la información en Excel: C:\Archivos de programa\Apache Software Foundation\Tomcat 7.0\w			
1. Inversión total de Sísmica 2D	24.0	8. Kilómetros totales en Sísmica 2D	90.0
2. Inversión Total de Sísmica 3D	36.0	9. Kilómetros cuadrados totales en Sísmica 3D:	66.0
3. Inversión Total en Sísmica 2D y 3D	60.0		
4. Inversión total de Estudios de Campo	30.0	10. Número de Estudios de Campo:	12.0
5. Inversión total de Estudios de Gabinete	40.0	11. Número de Estudios de Gabinete:	20.0
6. Inversión Total de Estudios de Campo y de Estudios de Gabinete	70.0	12. Número total de Estudios de Campo y de Gabinete	32.0
7. Inversión Operacional total	81.0	13. Número de eventos de Inversión Operacional:	18.0
<b>Monto de la Inversión Total</b>	<b>211.0</b>		

Para actualizar primero elija el Libro Excel y después continúe:

[continuar](#)

Fig. 7. Graphical interface for the import process of Excel files

The graphical interface to see the results is shown in Figure 8.

**Proceso de importaci3n a la base de datos**

---

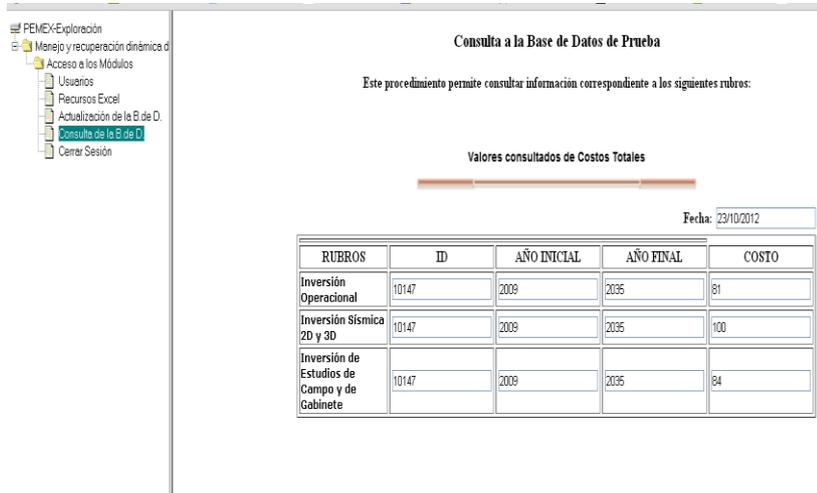
*Valores importados del portafolio de informaci3n (Capturador en Excel)*

Costos (en millones de pesos)		Fecha: 23/10/2012	
1. Inversi3n Total en Sísmica 2D y 3D	60.0	2. Inversi3n Total en Estudios de Campo y de Gabinete	70.0
3. Inversi3n Operacional Total	81.0	4. Monto Total de la Inversi3n	211.0

[continuar](#)

Fig. 8. Graphical interface to see the results of the import process

The graphical interface to see the results in the database is shown in Figure 9 in case of different update processes; the application can to accumulate each amount respectively.



**Fig. 9.** Graphical interface to query the database to see the results of the update

## 6 Conclusions

Enterprises are integrated of different areas, which perform several operations such as the consolidating operations; this involves its consolidation, through registration and dynamic management of data, by using appropriated tools.

This is why the data processing environments tend to be complex, decentralized and changing, and so are required to incorporate flexibility and modularity, using new technologies.

The development of this work is aimed to solve the problem before described, to establish an application environment that allows combining the use of tools to facilitate the Excel-Java-Database combination, and the POI API allows working this type of scheme at a lower cost supported in the use of free technologies.

The application presents an option to processing simulation at low cost, readily available resources and consistent with rapid and appropriate integration with existing processes, supported in an architecture based on widely recognized standards, such as the pattern called Model-View-Controller (MVC), allowing a greater versatility and facilitates interoperability with other applications.

The operation of the application is not at a disadvantage with the processes that could exist, since it is based on a test environment and easy scalability, supported by the integration of Java with other technologies such as Excel and MySQL, which together allow for greater versatility.

## **Acknowledges**

We appreciate the support of the Computing Research Center and of the National Polytechnic Institute, México, to carry out this work.

## **References**

1. Rentz, Daniel [et al], OpenOffice.org's Documentation of the Microsoft Excel File Format. <http://sc.openoffice.org/excelfileformat.pdf>.
2. Sengupta, Avik. Oliver Andrew. Opening Microsoft file formats to Java. <http://onjava.com/pub/a/onjava/2003/01/22/poi.html>
3. Apache Software Foundation. <http://www.apache.org>.
4. API POI (Poor Obfuscation Implementation). <http://poi.apache.org>.
5. Java Libraries to read/write Excel (.XLS) document files. [http:// Schmit.devlib.org/java/libraries-excel.html](http://Schmit.devlib.org/java/libraries-excel.html).
6. Metzger, Philip. Administración de un proyecto de programación. Trillas 1978.
7. Design Patterns Model-View-Controller. <http://java.sun.com/blueprints/patterns/MVC.html>.
8. XAMPP for WINDOWS. <http://www.apachefriends.org/en/xampp-windows.html>
9. Apache TOMCAT. <http://tomcat.apache.org>
10. James, Martin-1933-Computer data-base organization / James Martin -- 2a.Ed. -- Englewood Cliffs, N.J.: Prentice-Hall 1977.
11. Deitel, Harvey M. Java: how to program / H.M. Deitel, P. J. Deitel -- 2a.Ed. -- Upper Saddle River, NJ : Prentice Hall 1998.
12. MySQL. [http://www.webestilo.com/mysql\\_](http://www.webestilo.com/mysql_)
13. Schmuller, Joshep. Aprendiendo UML en 24 horas. PEARSON EDUCACIÓN. México, 2000. 1ª. Edición.



# Opportunistic Reasoning in DL Prolog

José Oscar Olmedo-Aguirre

Electrical Engineering, Cinvestav-IPN,  
Av. Instituto Politécnico Nacional 2508, 07300 Mexico City, MEXICO,  
oolmedo@cinvestav.mx

**Abstract.** The declarative programming language DL Prolog is currently being developed for coupling deduction and interaction for multi-agent system applications. The language design is addressed by providing a uniform programming model that combines the well-known refinements of resolution, SLD-resolution, UR-resolution and positive hyper-resolution, along with control strategies for problems dealing with stateless and state-based descriptions. The computational model of the language is shown to be correct with respect to resolution-based refutation. The main contribution of this work is in introducing a computational model that uses dynamic logic modalities in forward rules, leading to more efficient versions with respect to the reduced number of rules required to deal with the same problem.

**Keywords.** Logic programming, interaction, automated theorem proving, Horn clause theories, Dynamic Logic.

## 1 Introduction

Dynamic acquisition of interactive knowledge is an effective approach to deal with complex system design and development [7]. Blackboard systems [1, 2] were among the first interactive knowledge systems that imitates the opportunistic reasoning that arise in the brainstorming sessions of human experts while solving complex problems. Opportunistic reasoning allows to draw conclusions from a given set of facts placed in the blackboard, by reasoning in either forward or backward manner. The logic programming language DL Prolog [5] is being developed to combine efficiently both forms of reasoning. The computational model comprises SLD-resolution, UR-resolution and positive hyper-resolution [8, 9], to describe respectively stateless deduction and state-based transitions. The coordination model consists of a transactional global memory of ground facts along with a strategy for the theorem prover to control program execution by syntactically guided rule selection. In addition, the set of support restriction strategy coordinates the input and output of facts with the shared memory, maintaining the coherence of the current state of the computing agent.

Let us briefly explore other approaches that can be compared with ours: resolution theorem provers, constraint logic programming and coordination logic programming. The resolution-based theorem prover *OTTER* [8, 9] comprises a number of refinements of resolution along with a set of control strategies to

prune the explosive generation of intermediate clauses. However, OTTER does not account for interaction. The set of all instantaneous descriptions essentially corresponds to the set of support strategy. In OTTER, a clause is selected and removed from the set of support to produce a new set of clauses deduced from the axioms of the theory. Then, after simplifying a new clause by demodulation and possibly discarding it by either weighting, backward or forward subsumption, the new clause is placed back to the set of support.

*Concurrent Constraint Programming* (CCP) [6] proposes a programming model centered on the notion of constraint store that is accessed through the basic operations 'blocking *ask*' and 'atomic *tell*'. Blocking *ask*(*c*) corresponds to the logical entailment of constraint *c* from the contents of the constraint store: the operation blocks if there is not an enough strong valuation to decide on *c*. In this respect, the blocking mechanism is similar to the one used in DL Prolog to obtain the set of ground facts that match with the left-hand side of some rule. Besides, the constraint store shares some similarities with the global memory of ground facts. However, operation *tell*(*c*) is more restrictive than placing ground atoms in the global memory because constraint *c* must be logically consistent with the constraint store.

*Extended Shared Prolog* (ESP) [3] is a language for modeling rule-based software processes for distributed environments. ESP is based in the PoliS coordination model that extends Linda with multiple tuple spaces. The language design seeks for combining the PoliS mechanisms for coordinating distribution with the logic programming Prolog. Coordination takes place in ESP through a named multiset of passive and active tuples. They correspond to the global memory of facts in DL Prolog, although no further distinction between passive and active ground facts is made. ESP also extends Linda by using unification-based communication and backtracking to control program execution.

The paper is organized as follows. First we illustrate the forward and backward reasoning schemes that arise from the computational model with a programming example. Next, the syntax and the declarative semantics of the the DL Prolog programming language is presented. Finally, some remarks are given to conclude.

## 2 A programming example

As agents perceive the surrounding environment through *sensors* and act upon it through *effectors*, their interaction can effectively be decoupled by a shared coordination medium consisting of a multiset of ground facts that can be implemented by a blackboard system. By abstracting away interaction from deduction, the inherently complex operational details of sensors and effectors become irrelevant. The behavior of each individual agent is described by a set of backward and forward rules that describe the exchange of information through the coordination medium. As an example, consider the problem of parsing and evaluating simple arithmetic expressions. The parser uses the context free grammar (CFG):

$$E \rightarrow i \mid (E) \mid E + E \mid E \times E$$

where non-terminal  $E$  stands for a well-formed integer expression.

Table 1 shows theory *Natural* for the natural numbers written in DL Prolog, closely similar to those written in pure Prolog. This theory uses *backward rules* that have the general form  $P \Leftarrow P_1, \dots, P_n$  with  $n \geq 0$ . The logical propositions of the theory are built upon infix predicates  $=$ ,  $<$ , and  $\leq$ , whose recursive definitions are given by clauses  $N_1$  to  $N_5$ . *Natural* represents the deductive component of the interactive parser.

<b>theory</b> <i>Natural</i>	
<b>axioms</b>	
$N_1 :$	$0 + y = y \Leftarrow$
$N_2 :$	$(x + 1) + y = (x + y) + 1 \Leftarrow$
$N_3 :$	$0 \leq y \Leftarrow$
$N_4 :$	$(x + 1) \leq (y + 1) \Leftarrow x \leq y$
$N_5 :$	$x < y \Leftarrow (x + 1) \leq y$
<b>end</b>	

**Table 1.** Natural numbers using backward rules.

Table 2 shows a theory written in DL Prolog for a parallel *Parser* that extends *Natural*. This theory uses *forward rules* that have the general form  $P_1, \dots, P_n \mid C \Rightarrow [A] P$  with  $n \geq 0$ . The declarative reading of the forward rule is that, if appropriate predicates  $P_1, \dots, P_n$  have been placed in the common memory and their contents satisfy the condition  $C$ , then the action (i.e. imperative program)  $A$  is executed to obtain the values bound to the variables occurring in the postcondition  $P$ . The rules of *Parser* define a bottom-up parser for simple arithmetic expressions whose syntactic entities are represented by ground atoms.  $T(n, t)$  asserts that symbol  $t$  occurs at position  $n$ , while  $E(n_1, n_2, x)$ , with  $n_1 \leq n_2$ , asserts that the sequence of symbols from  $n_1$  to  $n_2$  forms a well-formed arithmetic expression whose evaluation is the integer value  $x$ . The forward rules are remarkably similar to those given in the CFG. Table 3 shows the forward rules  $P'_1$  through  $P'_3$  that respectively replace  $P_1$  through  $P_3$  from the theory *Parser* to produce a sequential parsing and evaluation of the number enclosed in brackets. The parsing goes from left to right, starting as soon as the left-bracket '[' is detected and terminating after the right-bracket ']' is detected. Table 4 shows the sole forward rule  $P''_{123}$  that replaces rules  $P_1$  through  $P_3$  from the theory *Parser* to produce an alternative sequential parser along with its evaluation. This version of the parser differs from the previous sequential version in that it only uses one forward rule for the parsing of the numbers, leading to a more efficient parsing and evaluation. In rule  $P''_{123}$ , the action resembles the usual sequential imperative program that parses and converts the sequence of digits to a number whose value is bound to the output variable  $z$  of the postcondition  $E(n_1, n_2, z)$ . The meaning of the action connectives is explained in section devoted to the formal description of the language.

```

theory Parser
extends Natural
rules
P1 : T(n, t) | digit(x) ⇒ [x := toInt(t)] N(n, n, x).
P2 : N(n1, n2, x), N(n3, n4, y)
      | n1 ≤ n2, n2 = n3, n3 ≤ n4
      ⇒ [z := x × 10n4-n3+1 + y] N(n1, n4, z).
P3 : T(n1, '['), N(n2, n3, x), T(n4, ')')
      | n1 + 1 = n2, n2 ≤ n3, n3 + 1 = n4
      ⇒ E(n1, n4, x).
P4 : T(n1, '('), E(n2, n3, x), T(n4, ')')
      | n1 + 1 = n2, n2 ≤ n3, n3 + 1 = n4
      ⇒ E(n1, n4, x).
P5 : E(n1, n2, x), T(n3, '+'), E(n4, n5, y)
      | n1 ≤ n2, n2 + 1 = n3, n3 + 1 = n4, n4 ≤ n5
      ⇒ [z := x + y] E(n1, n5, z).
P6 : E(n1, n2, x), T(n3, '×'), E(n4, n5, y)
      | n1 ≤ n2, n2 + 1 = n3, n3 + 1 = n4, n4 ≤ n5
      ⇒ [z := x × y] E(n1, n5, z).
end

```

Table 2. Bottom-up parallel parser for arithmetic expressions.

```

P'1 : T(n, '[') ⇒ N(n, n, 0).
P'2 : N(n1, n2, x), T(n3, t)
      | n1 ≤ n2, n2 + 1 = n3, digit(t)
      ⇒ [z := x × 10 + toInt(t)] N(n1, n3, z).
P'3 : N(n1, n2, x), T(n3, ')')
      | n1 ≤ n2, n2 + 1 = n3
      ⇒ E(n1, n3, x).

```

Table 3. Sequential parsing of numbers.

```

P''123 : T(n1, '[') ⇒ [ int x, t, n : (
  ( x, n := 0, n1 + 1;
    ( T(n, t)?; digit(t)?;
      x, n := 10 × x + toInt(t), n + 1; ) *;
    t = ')'?;
    z, n2 := x, n
  ) ] E(n1, n2, z).

```

Table 4. Sequential parsing of numbers with a dynamic logic modal action.

Table 5 sketches the interaction that produce both sequential parsers in the common memory, when dealing with the input  $([3141] + [79]) \times [2]$ . In the table, the time increases downwards with each row, whereas the symbols of the input are disposed horizontally. At the top of the table, the input consists of two rows: the first row corresponds to the indexes of the input, starting with 0, whereas the second row corresponds to all the symbols of the input occurring at the corresponding position given by the index. The entire input is not available immediately, but rather unpredictably and individually each symbol from another. When a sensor detects a symbol  $t$  at position  $n$ , it asserts the ground predicate  $T(n, t)$ , as shown in the first column of the table. For example the symbols '4' and '1', at positions 4 and 14, respectively, were early detected, while the symbol ']' at position 6 was detected soon after. When there are enough symbols in the memory to activate a forward rule, those symbols are enclosed in a box and the rule applied is shown in the first column of the table. For example, at the row 8, the rules  $P'_1$  and  $P'_2$  are applied one after the other. However due to there is no symbol at index position 3, the parsing stops there. The parsing resumes immediately after the symbol '1' at 3 becomes available, continuing the parsing of the entire number, from positions 1 to 6 with value 3141. The parsing and evaluation process continues until all the symbols of the input expression are eventually analyzed. Then an effector may inform to a client agent that the given expression is well-formed by placing the term  $E(0, 16, 6440)$  that include its evaluation.

### 3 DL Prolog formal description

An experimental system for DL Prolog has been built to evidence the viability of the approach. The system consists of a parser with integrated type inference to decide whether the program constructs are well-formed. The computational model is described as a structured-operational semantics interpreter that calculates the next state of the shared memory.

Let  $\Sigma = \bigcup_{\alpha} \Sigma_{\alpha}$  be a set of *constructor* (constant) names and let  $\Xi = \bigcup_{\beta} \Xi_{\beta}$  be a set of *variable* names, each partitioned by the basic types `bool`, `int`, and `act`, among others. The set  $T(\Sigma, \Xi)$  of *terms with variables* is the minimal set of phrases that is closed under composition of a constructor with a (possibly empty) previously constructed sequence of terms. The set  $T(\Sigma) = T(\Sigma, \emptyset)$  of *ground terms* consists of the terms with no variables. Type judgments are embedded in the grammar rules of the language to ensure that clauses and programs are well-formed. In particular, the grammar rule shown below describes the syntactic structure of a well-formed term  $T_{\beta}$  of type  $\beta$ :

$$T_{\beta} ::= x_{\beta} \mid c_{\beta} \mid c_{\beta_1 \dots \beta_n \rightarrow \beta}(T_{\beta_1}, \dots, T_{\beta_n})$$

The set  $P(\Sigma, \Xi)$  of *atomic predicates with variables* is the minimal set closed under composition of predicate symbols with (possibly empty) sequences of terms. The set  $P(\Sigma) = P(\Sigma, \emptyset)$  of *ground atoms* consists of all atoms with no variables.

	0	1	2	3	4	5	6	7	8	9	0	1	2	3	4	5	6	
	(	[	3	1	4	1	]	+	[	7	9	]	)	×	[	2	]	
$\Rightarrow T(4, '4'), \Rightarrow T(16, '')$					4										[			
$\stackrel{P'_1}{\Rightarrow} N(16, 16, 0)$															]			
$\Rightarrow T(6, '')$								]										
$\Rightarrow T(2, '3'), \Rightarrow T(10, '7')$			3							7								
$\Rightarrow T(12, '')$																		
$\Rightarrow T(10, '9'), \Rightarrow T(11, '')$											9	]	)					
$\Rightarrow T(1, '')$																		
$\stackrel{P'_1}{\Rightarrow} N(1, 1, 0), \stackrel{P'_2}{\Rightarrow} N(1, 2, 3)$	[	[	3															
$\Rightarrow T(3, '1'), \Rightarrow T(13, ' \times')$				1											×			
$\stackrel{P'_2}{\Rightarrow} N(1, 4, 314)$	[	[	3	1	4													
$\Rightarrow T(5, '1'), \Rightarrow T(18, '')$						1											]	
$\stackrel{P'_2}{\Rightarrow} N(1, 5, 3141)$	[	[	3	1	4	1												
$\stackrel{P'_3}{\Rightarrow} E(1, 6, 3141)$	[	[	3	1	4	1	]											
$\Rightarrow T(7, ' +'), \Rightarrow T(15, '2')$								+								2		
$\stackrel{P'_2}{\Rightarrow} N(14, 15, 2)$															[	[	2	
$\stackrel{P'_3}{\Rightarrow} E(14, 16, 2)$															]	[	2	
$\Rightarrow T(8, '')$																		
$\stackrel{P'_1}{\Rightarrow} N(8, 8, 0)$										[								
$\stackrel{P'_2}{\Rightarrow} N(8, 10, 79)$										[	7	9						
$\stackrel{P'_3}{\Rightarrow} E(8, 11, 79)$										[	7	9	]					
$\stackrel{P'_6}{\Rightarrow} E(1, 11, 3220)$		[	3	1	4	1	]	+		[	7	9	]					
$\Rightarrow T(0, '')$	(	[	3	1	4	1	]	+		[	7	9	]	)				
$\stackrel{P'_3}{\Rightarrow} E(0, 12, 3220)$	(	[	3	1	4	1	]	+		[	7	9	]	)				
$\stackrel{P'_6}{\Rightarrow} E(0, 16, 6440)$	(	[	3	1	4	1	]	+		[	7	9	]	)	×	[	2	]

Table 5. Interactive parsing.

A *literal* is an atom or a negated atom. A *clause* is a disjunction of literals. The set of clauses with variables is denoted  $C(\Sigma, \Xi)$ . Clauses are usually written in implication form  $P \Leftarrow Q$ , where  $P$ , called the *consequent*, is a disjunction of atoms and  $Q$ , called the *antecedent*, is a conjunction of atoms. A *unit clause* contains only one literal. A *positive clause* contains no negated atoms, whereas a *negative clause* contains no positive atoms. A *Horn clause* contains at most a positive atom. A *goal* consists only of negative atoms that can be represented by an implication with false as consequent. The *set of goals with variables* is denoted  $G(\Sigma, \Xi)$ .

**Terms**  $T ::= x \mid c \mid c(T_1, \dots, T_n)$

**Predicates**  $P ::= \text{false} \mid \text{true} \mid T_1 = T_2 \mid p(T_1, \dots, T_n)$

**Goals**  $G ::= P \mid G_0 \wedge G_1$

**Horn clauses (Backward rules)**  $B ::= P \mid P \Leftarrow G \mid \forall x.B$

**Events**  $E ::= P \mid E_1, E_2$

**Actions (Programs)**

$$A ::= \text{skip} \mid \text{fail} \mid G? \mid (A) \mid A_1; A_2 \mid A_1 \cup A_2 \mid A^* \mid \text{int } x_1, \dots, x_n : A \mid x_1, \dots, x_n := T_1, \dots, T_n$$

**Modal actions**  $M ::= P \mid [A] M \mid \langle A \rangle M$

**Modal clauses (Forward rules)**  $F ::= M \mid E \mid G \Rightarrow M \mid \forall x.F$

The DL modal extensions require the introduction of well-formed imperative programs in clauses, hereinafter called *actions*. The set  $A(\Sigma, \Xi)$  of *actions with variables* is the minimal set of phrases that is closed under composition of actions with action connectives. A basic action is either the null-action (`skipact`), the failure (`failact`), or the assignment of ground terms to variables (using the binary operator `:=int intact`). Action connectives are the postfix unary operator for testing a condition (`?act`), the postfix unary operator for iteration (`*actact`), the infix binary operators for sequential composition (`;actact`), and the infix binary operator for non-deterministic choice (`∪actact`). The mixfix modal connectives of modal necessity (`[]actbool`) and possibility (`\langle \rangleactbool`) compose actions along with their postconditions. The following definitions provide an interpretation of the usual imperative programming constructs in terms of DL actions:

$$\begin{aligned} \text{skip} &\equiv \text{true?} \quad , \quad \text{fail} \equiv \text{false?} \\ \text{if } F \text{ then } A \text{ fi} &\equiv F?; A \\ \text{if } F \text{ then } A_1 \text{ else } A_0 \text{ fi} &\equiv (F?; A_1) \cup (\neg F?; A_0) \\ \text{while } F \text{ do } A \text{ od} &\equiv (F?; A)^*; \neg F? \end{aligned}$$

The precedence in decreasing order among the action connectives is the following: test (?), iteration (\*), sequence (;) and non-deterministic choice (∪). Parenthetical expressions are allowed to modify the precedence order of the action connectives. Variables occurring in an action  $A$  are either *logical variables*

or *local imperative variables*. Logical variables occurring in a clause are universally quantified, whereas local variables are introduced by declaration within an action. A declaration of local variables  $\text{int } x_1, \dots, x_n : A$  creates new local imperative variables whose scope and duration are restricted to the block  $A$ . A simple assignment  $x := T$  evaluates the term  $T$  in the current state and the resulting constant value is assigned to  $x$ . Logical and imperative variables are compatible in assignments of the same type, so they can appear in both sides of the assignment. Note however that logical variables can be defined at most once, whereas imperative variables can be redefined. A multiple assignment  $x_1, \dots, x_n := T_1, \dots, T_n$  evaluates all the terms at the right-hand side in the current state and the resulting values are assigned to the corresponding variables at the left-hand side of the assignment. Modal necessity composition  $[A] P$  means that after executing action  $A$ , postcondition  $P$  is necessarily true.

In a signature  $(\Sigma, \Xi)$  with variables, a *substitution* is a partial function  $\sigma : \Xi \rightarrow T(\Sigma, \Xi)$ , where  $\sigma(x) \neq x$  for any variable  $x \in \Xi$ .  $\{\}$  denotes the empty substitution. A *ground substitution* is a substitution  $\sigma : \Xi \rightarrow T(\Sigma)$  valued on ground terms. For any variable  $x \in \Xi$  and any substitution  $\sigma$ , let  $x\sigma = \sigma(x)$  if  $x \in \text{dom}(\sigma)$  and  $x\sigma = x$  otherwise. For any term  $t \in T(\Sigma, \Xi)$ , let  $t\sigma$  be the term obtained by substituting any variable  $x$  appearing in  $T$  by  $x\sigma$ :

$$\begin{array}{ll} x\{\} = x & x\sigma = \begin{cases} x & \text{if } x \notin \text{dom}(\sigma) \\ (x\sigma) & \text{if } x \in \text{dom}(\sigma) \end{cases} \\ c\sigma = c & c(T_1, \dots, T_n)\sigma = c(T_1\sigma, \dots, T_n\sigma) \\ [A]p\{\} = [A]p & [A]p\sigma = [\sigma_{:=}; A]p \end{array}$$

where notation  $[\sigma_{:=}]$  stands for the multiple assignment  $x_1, \dots, x_n := T_1, \dots, T_n$  given the substitution  $\sigma = \{x_1 \mapsto T_1, \dots, x_n \mapsto T_n\}$ , for  $1 \leq n$ . Thus the substitution for a modal action  $A$  is defined as the initial value that the variables take before the action starts its execution. The *composition* of two substitutions  $\sigma_0, \sigma_1 \in \Xi \rightarrow T(\Sigma, \Xi)$ , written  $\sigma_0 \cdot \sigma_1$ , is defined as

$$\sigma_0 \cdot \sigma_1 : x \mapsto \begin{cases} (x\sigma_0)\sigma_1 & \text{if } x\sigma_1 \notin \text{dom}(\sigma_1) \\ x\sigma_1 & \text{if } x \in \text{dom}(\sigma_1) - \text{dom}(\sigma_0) \\ \text{failure} & \text{otherwise} \end{cases}$$

Besides the natural extension to terms  $T(\Sigma, \Xi) \rightarrow T(\Sigma, \Xi)$ , substitutions are also extended to predicates, goals, and both backward and forward rules.

### 3.1 Computational model

The *backward computation* relation  $\triangleleft \subset G(\Sigma, \Xi) \times (\Xi \rightarrow T(\Sigma, \Xi))$  consists of pairs relating goals and substitutions, where the substitutions are defined upon the variables occurring in a renamed variant of the rule. An *instantaneous description*  $I \subset P(\Sigma) \times (\Xi \rightarrow T(\Sigma))$  relates ground predicates and ground substitutions, describing a portion of the current state of the shared memory. The substitutions keep a track of the bindings for all the variables that occurred in

the renamed variant of each forward rule applied. The *forward computation* relation  $\triangleright \subset \mathcal{P}(I) \times \mathcal{P}(I)$  relates pairs of instantaneous descriptions. The transition relations are defined in Table 6.

<p><i>Backward computation</i></p> $\frac{P' \Leftarrow G' \in B(\Sigma, \Xi) \quad P\sigma' = P'\sigma'}{(\{P\} \cup G, \sigma) \triangleleft (G'\sigma' \cup G\sigma', \sigma\sigma')}$ <hr style="width: 50%; margin: 10px auto;"/> <p><i>Forward computation</i></p> $\frac{E_1, \dots, E_n \mid G \Rightarrow [A]P \in F(\Sigma, \Xi) \quad E_i\sigma_i = P_i\sigma_i, i \in 1, \dots, n \quad (G, \sigma_1 \cdots \sigma_n) \triangleleft^* (\{\}, \sigma)}{\{(P_1, \sigma_1), \dots, (P_n, \sigma_n)\} \cup I \triangleright \{(P_1, \sigma_1), \dots, (P_n, \sigma_n), ([\sigma :=; A]P, \sigma\sigma')\} \cup I}$
--

**Table 6.** Operational semantics of backward and forward computation.

The backward computation rule describes a *refutation step* from  $(\{P\} \cup G, \sigma)$  to  $(G'\sigma' \cup G\sigma', \sigma\sigma')$  by replacing the head  $P\sigma'$  with the body  $G'\sigma'$  of the instance of the backward rule  $P \Leftarrow G'$  under a suitable substitution  $\sigma'$  such that  $P\sigma' = P'\sigma'$ . The new goal is an instance under  $\sigma'$  of the body  $G'$  and the remaining goal  $G$ , along with the new answer substitution obtained from the composition of  $\sigma'$  with the previous one  $\sigma$ . In case that the application of the rule leads to a failure, another backward rule if any is selected and applied after backtracking to the previous goal and the previous substitution; otherwise, if no more rules can be selected, the backward computation terminates in failure.

The forward computation rule  $E_1, \dots, E_n \mid G \Rightarrow [A]P$ , with  $n > 0$ , can be selected for deducing the ground predicate  $P\sigma\sigma'$  only if the following three conditions hold: (i) there are  $n$  ground predicates  $P_1, \dots, P_n$  already asserted in the working memory, (ii) there are  $n$  ground substitutions  $\sigma_1, \dots, \sigma_n$  that makes syntactically identical the corresponding instances of each event  $E_i$  with an appropriate predicate  $P_i$ , i.e. equation  $E_i\sigma_i = P_i\sigma_i$  holds for  $1 \leq i \leq n$ , and (iii) the composition  $\sigma_1 \cdots \sigma_n$  of the  $n$  substitutions satisfies the goal  $G$ . Whenever these conditions are met, the forward rule can be applied. In the rule, because the variables occurring in any event  $E_i$  does not occur in any other  $E_j$  ( $i \neq j$ ), the composition of their corresponding ground substitutions simply consists of the union of all of them. The ground substitution  $\sigma$  produced by the backward computation rule may extend the composition  $\sigma_1 \cdots \sigma_n$  with bindings for the new variables that  $G$  may introduce. Hence, the equation  $E_i\sigma = P_i\sigma$  also hold for  $\sigma$  with  $1 \leq i \leq n$ . The instance under  $\sigma$  of the modal action  $[A]P$  is then executed following the standard interpretation of the action connectives [4]. Assuming that  $A$  terminates with the initial values given by  $\sigma$ , the postcondition  $P$  becomes

satisfied by the substitution  $\sigma\sigma'$ , i.e. by the values computed by  $A$  assigned to the output variables occurring in  $\sigma'$ . The truth of the ground predicate  $P\sigma\sigma'$  leads to the instantaneous description  $(P, \sigma\sigma')$ . However, if the guard  $G\sigma$  fails, another set of predicates asserted in the shared memory must be considered. If no more possible selections of predicates were possible for the forward rule, another rule is selected if any. If no more forward rules were applicable, the agent would appear as non-responsive until another predicate assertion were eventually produced in the shared memory.

The correctness of the computational model can be stated as follows:

**Proposition 1 (Correctness).**

$$\{(P_1, \sigma), \dots, (P_n, \sigma)\} \triangleright \{(P_1, \sigma), \dots, (P_n, \sigma), (P, \sigma\sigma')\} \text{ implies } P_1\sigma \wedge \dots \wedge P_n\sigma \Rightarrow P\sigma\sigma'$$

Note that a predicate asserted by the forward rule monotonically increases the content of the shared memory as the events are never retracted by the rule.

## 4 Conclusions

The problem of coupling interaction in a resolution theorem prover with syntactically guided selection of the control strategy to be used has been presented in this paper. The experimental programming language DL Prolog has been designed to deal with state-based descriptions using forward rules and stateless deduction using backward rules. The programming model allows to combine backward and forward rule chaining in a simple and more efficient manner.

## References

1. D. D. Corkill Collaborating Software: Blackboard and Multi-Agent Systems and the Future. In Proceedings of the International Lisp Conference, New York, New York, Oct, 2003.
2. D. D. Corkill GBBopen Tutorial. The GBBopen Project, March 2011. online: <http://gbbopen.org/hypertutorial/index.html>
3. P. Ciancarini Coordinating Rule-Based Software Processes with ESP, *ACM Trans. on Software Engineering and Methodology*, 2(3):203-227, July, 1993.
4. D. Harel, J. Tiuryn, and D. Kozen Dynamic Logic. Cambridge, MA, USA: MIT Press, 2000.
5. J. O. Olmedo-Aguirre and G. Morales-Luna. A Dynamic Logic-based Modal Prolog. In Proceedings MICAI 2012, San Luis Potosi, México, Oct, 2012. [To appear]
6. V.A. Saraswat Concurrent Constraint Programming. *Records of 17th ACM Symposium on Principles of Programming Languages*, 232-245. San Francisco, CA. 1990.
7. P. Wegner, Interactive Software Technology, *CRC Handbook of Computer Science and Engineering*, May 1996.
8. L. Wos, R. Overbeek, E. Lusk and J. Boyle, *Automated Reasoning. Introduction and Applications*, McGraw-Hill, Inc., 1992.
9. L. Wos and G. Pieper, *A Fascinating Country in the World of Computing: Your Guide to Automated Reasoning*, World Scientific Publishing Co., 1999.

# Towards a Multi-Agent System Architecture for Supply Chain Management

Carlos R. Jaimez-González, Wulfrano A. Luna-Ramírez

Departamento de Tecnologías de la Información, Universidad Autónoma Metropolitana - Cuajimalpa, Av. Constituyentes No. 1054, Col. Lomas Altas, C.P. 11950, México D.F.  
{cjaimez, wluna}@correo.cua.uam.mx

**Abstract.** Individual business processes have been changing since the Internet was created, and they are now oriented towards a more distributed and collaborative business model, in an e-commerce environment that adapts itself to the competitive and changing market conditions. This paper presents a multi-agent system architecture for supply chain management, which explores different strategies and offers solutions in a distributed e-commerce environment. The system is designed to support different types of interfaces, which allow interoperating with other business models already developed. In order to show how the entire multi-agent system is being developed, the implementation of a collaborative agent is presented and explained.

**Keywords.** Supply Chain Management, Agent Technology, Collaborative Agents, Multi-Agent Systems, E-Commerce.

## 1 Introduction

The great amount of information and the new technologies increase the client expectations about services, as well as the global competition makes enterprise leaders look for new ways of handling businesses. Companies are not in a position to trust in static business strategies, but they have to be capable of facing environments that are uncertain and change rapidly; such as changes in the bank currency values, delays in the delivery of production materials from suppliers, broken relationships with suppliers and clients, failure in production facilities, workers absence and illness, cancelations or changes of client orders, etc.

Companies have to carry out a series of activities, such as obtaining materials, manufacturing products, storage of products, sales and delivery of products, services to the client, among others. All these activities have to be carried out as if they were a dynamic process, in such a way that there is a balance between them. In a company, this is precisely the main task of the Supply Chain Management (SCM); which is in charge of negotiating with suppliers to obtain all the necessary materials for production, handling client orders, controlling the inventory, establishing the amount of time dedicated for manufacturing and delivering finished products, etc.

It is common that companies have distributed businesses, where suppliers and clients are located around the world. The development of the information technologies have produced that organizations use the Internet to participate in e-commerce activities; this way they reduce administrative and transactional costs, increase their profit, and interact with a greater number of business partners in different geographical locations. The Internet has produced the change of individual business processes, towards a more distributed and collaborative model. In order to be capable to handle this model, companies need a solution that allows them participate in e-commerce environments. This solution has to include a system that helps in the decision-making, and at the same time adapts to the changes; such system can collect and process information from a great number of heterogeneous sources, as well as help making decisions more precisely in competitive and changing market conditions. Due to the need of designing strategies for coordinating and integrating business entities within these environments, one of the objectives of this work is the development of strategies with agent technology for e-commerce. For the implementation and testing, it is needed to develop a system for supply chain management in the environments mentioned.

This paper presents a multi-agent system architecture for supply chain management, which is based on multiple collaborative agents, explores different strategies of the production process globally, and offers solutions for managing supply chains in distributed e-commerce environments. The system is designed to support different types of interfaces, which allows interacting with the existing business models of others participants, through the management or integration of the cooperative supply chain. An implementation of a collaborative agent is presented and explained, so that it can be observed how the entire system is being developed. Because the electronic market is a recently established business model and it is conceived as an activity for solving cooperative distributed problems, the design of a system for supply chain management has become more important than ever. It should be noticed that the work presented in this paper, is an extension of an initial proposal for developing a multi-agent system with collaborative agents, which is described in [1].

The rest of the paper is organized as follows. Section 2 provides preliminaries and state of the art for SCM, its progress, the multi-agent approach, and an experimentation platform. In section 3 it is presented the architecture of the system and its six collaborative agents. Some sales and production strategies for e-commerce agents are presented in section 4. Finally, we provide conclusions and future work in section 5.

## **2 Preliminaries and State of the Art**

This section provides some preliminaries, and summarizes the state of the art in the field of the SCM, and the decision support systems for SCM. In particular, the tendency in these areas is to move from the static business processes, towards dynamic and distributed models. The decision support systems for SCM are currently designed as multi-agent systems to support such models. Many research groups around the world dedicate their work to explore several problems in the domain of the SCM, and have carried out studies and experiments on the simulation platform called *Trading*

*Agent Competition – Supply Chain Management (TAC SCM)*, which is used widely in research, and it is described at the end of this section.

## 2.1 SCM Progress

SCM is a process that includes a variety of interrelated activities, such as negotiation with suppliers to obtain materials, competition for client orders, inventory management, production schedule, delivery of products to clients, etc. SCM concepts have been used by companies since 1950 [2]. In the 80's the idea of automate business processes through SCM was very popular, however the experts involved, isolated every entity of the supply chain as a static process, separated from the rest. Some of the work carried out towards the end of the 90's concentrated in solving specific separated areas of the SCM [3], [4]. In [5], it started to conceive the SCM problem as a dynamic environment and as an integrated process with constraints [6].

With the creation of the Web, e-commerce systems have become extremely popular, mainly in the last years. There are e-commerce integrated models, which include suppliers, clients, commercial partners, e-commerce agents, etc., within a global electronic environment [7]. There have been proposals for system architectures to support the participation in e-commerce [8]. Many researchers agree that the architecture of a decision support system for SCM must be agile in order to compete in the dynamics of electronic markets, as well as easily configurable, adaptable to several businesses, and support several protocols of different commercial environments. The multi-agent approach has demonstrated to be the most appropriate to fulfill these requirements [9], [10], [11], [12].

## 2.2 Multi-agent approach

Agent technology has become the most popular tool to design distributed systems for supply chain management, because it provides a dynamic and adaptable way of managing separately every piece of the chain. Agent-based supply chain management systems can respond rapidly to internal or external changes through decision-making mechanisms. A detailed review of some agent-based systems can be found in [13].

Designing the SCM solution as a multi-agent system brings another advantage, it allows separating the different tasks of the SCM, exploring them independently, and analyzing them as a whole. This feature is particularly important because it allows focusing on two important pieces of the supply chain: the demand, which refers to the sales to clients; and the production, which involves getting the raw materials and manufacturing the products. In the case of the sellers, the main problem they face when dealing with supply chains is to decide what offers they will place to their clients, what prices they will give them, when to sell their products, etc. The aim is to increase their profit; but the task is not easy to solve in the e-commerce context, where prices are established dynamically.

It is important to note that there have been proposals using different approaches to design agent-based decision-making systems for SCM [11], and also there have been developments of agent-oriented architectures for e-commerce [7]. There has been also

an integration of a multi-agent system for SCM in [14]. A multi-agent decision support system for supply chain management is described in [15], which also uses several agents to accomplish the tasks of the supply chain. A description of recent projects that use agents in the context of SCM is provided in [16].

### 2.3 The Experimentation Platform

Companies do not implement the proposed solutions immediately; instead they wait until solutions have been fully tested through a series of experiments with the production and sales strategies proposed. There have been many attempts to create simulation and experimentation tools and platforms for SCM, which allow studying and testing the different algorithms and strategies. Among these attempts, there is an experimentation platform called the *Trading Agent Competition Supply Chain Management* (TAC SCM) [17], which allows testing SCM systems, because it encapsulates many of the problems that can be faced in real SCM environments, such as time constraints, unpredictable opponents, limited production capacity, late deliveries, missing components or materials, competition for suppliers, etc. This platform will be used for the experimentation of the system proposed; specifically for testing the performance solution of the complete system, as well as for the production and sales strategies that will be implemented.

This platform was designed by the Carnegie Mellon University (CMU) and the Swedish Institute of Computer Science (SICS) in 2003, as part of the International Trading Agent Competition [18]. The platform simulates a game, in which trading agents compete in the context of SCM, such that they can evaluate their algorithms and proposals. The agents that participate in this competition are developed by different research groups around the world. The first author of this paper has already participated in this TAC-SCM competition in 2004, with a production-driven agent called Socrates [19], which was placed among the best 12 agents out of 31.

During the game there are six agents that compete among them in a simulation that lasts for 220 days, with a real duration of 55 minutes. Each agent is a manufacturer of computers, which assembles them from their basic components, such as CPUs, motherboards, RAM memories, and hard disks. The CPUs and motherboards are available in two families of different products: IMD and Pintel. A Pintel CPU only works with a Pintel motherboard, while an IMD CPU can only be incorporated to an IMD motherboard. The CPUs are available in two speed models: 2.0 and 5.0 GHz; the memories are produced in two sizes: 1 and 2 GB; and the hard disks are available in two sizes: 300 and 500 GB. There are ten different components in total, which can be combined in order to have a total of 16 different models of computers. All the agents in the game need to obtain these components from 8 different suppliers; and at the same time the agents need to secure client orders every day, assemble a number of computers of the required model, and subsequently deliver the finished products. All the agents start the game with no money in their bank accounts, with no components of finished components in their inventory, and with no pending client orders. Every computer model requires a different number of cycles to be assembled, and an agent has a limited production capacity of 2000 cycles a day.

The TAC server simulates the suppliers, clients and bank; and also provides production and storage services to all the six agents. The agent who accomplishes the greatest profit at the end of the game is the winner. The rules of the game, as well as the required Agentware software to carry out experimentation with this platform can be downloaded directly from the TAC Web site [18]. Figure 1 illustrates all the tasks that an agent has to carry out every day during a TAC SCM game, such as dealing with suppliers and customers, managing the inventory of components and products, programming the production and delivery schedules, etc.

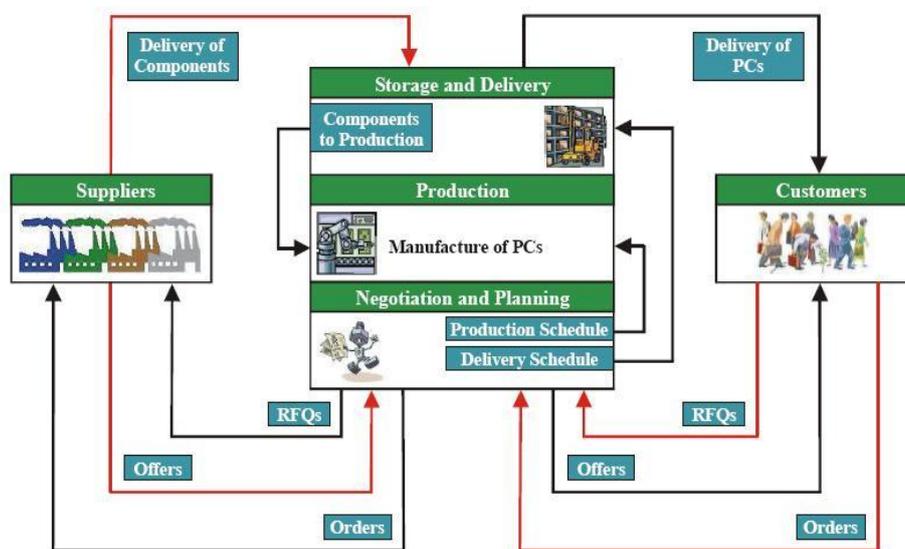


Fig. 1. Daily tasks of an agent in TAC SCM.

### 3 System Architecture

This section describes the system architecture, and the six agents that are part of it. The system is being developed using the principles of the multi-agent systems. Using this approach allows us studying several of the problems that can be faced in the SCM domain. Although we planned that the system competes in the game organized by the TAC SCM, its architecture will be generic and configurable, in such a way that can be adapted to similar environments, as well as modifiable to meet the particular requirements of a specific business. Tackling this problem with a multi-agent approach allows us exploring different problems of the SCM separately. We also propose to use a Web service based architecture, through which the multi-agent system can interoperate with other systems or agent-based systems implemented with other technologies or programming languages.

In a commercial environment, the system should produce certain decisions in the acquisition of materials, manufacture, sales and delivery of products. In order to make these decisions, it is needed to complete several tasks, which can be linked to the result of others or be independent. Because there are time constraints for decision-making, the independent tasks must be carried out in parallel to save time. Trying to solve this problem with a multi-agent approach is adequate to meet these requirements. Through the coordination and collaboration, the agents are capable of managing the distributed activities in the supply chain. The multi-agent approach is a “*natural way of modularizing complex systems*” [16], such as the systems for SCM.

The system is composed of six agents; five of them correspond to each of the entities of the supply chain: Sales Manager, Supply Manager, Inventory Manager, Production Manager, Delivery Manager, and the sixth agent is the Coordinator, who is in charge of coordinating the performance of the whole system, as well as communicating with the external environment. Figure 2 illustrates the multi-agent system architecture, created using the Prometheus Design Tool (PDT) [21], which is a graphical editor for specifying, designing and implementing intelligent agents.

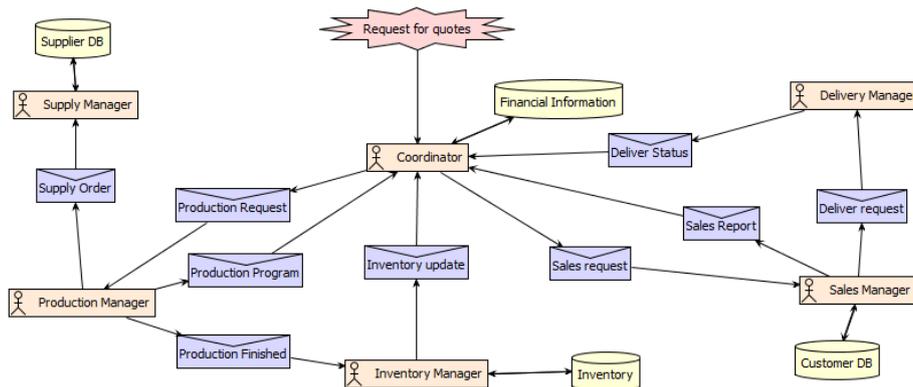


Fig. 2. Multi-agent system architecture.

Having separated agents allows reusing the system in parts, experiment with them in a wide range of real SCM applications, because every agent can be easily included or removed from the system without affecting the performance of the entire system. Although every agent has its own goals, the agents work in cooperation to reach the common global goal, which is to maximize the total profit. This goal can be divided in sub-goals, such as maximizing the profit in sales, minimizing the prices when buying materials, minimizing the costs of storage of materials and products, minimizing the penalizations for late deliveries, etc.

The following code fragment, written in Jason [22], shows the project where the multi-agent system is being implemented. The code includes the six collaborative agents previously mentioned.

```

/* Jason Project */
MAS cms {
  infrastructure: Centralised
  agents:
    coordinator;
    supply_Manager;
    production_Manager;
    inventory_Manager;
    delivery_Manager;
    sales_Manager;
}

```

This initial implementation was written in Jason, due to the possibility of using the internal actions implemented in Java to allow agents more autonomy to carry out their tasks, abilities for information management, and data processing. This also helps separating the abstraction level from the coordination, communication and deliberation. Additionally, this allows testing the system in a different implementation from the one that will be used for the TAC SCM experimentation platform, which uses a Java-based software called Agentware.

The following Jason code fragment implements part of a plan for the *Coordinator Agent*, in which it receives a request. The code shows the declaration of the initial beliefs and rules; goals; and the implementation of the plans, which will be used to carry out the agent's tasks; it will also try to accomplish the goals specified at the beginning, and those that the other agents request.

```

// Agent coordinator in project CMS.mas2j
/* Initial beliefs and rules */
...
/* Initial goals */
...
/* Plans */
...
+!attendRequest(RQST) : true <-
  .evalProgram(ProdProgram);
  .send(delivery_Manager, achieve, evalDelivery(ProdSchedule));
  .updateProdEschedule(ProdProgram, ProdSchedule, RQST);
  .send(production_Manager, achieve, evalProdRequest(RQST,
    ProdProgram, TimeProduction));
  .send(production_Manager, achieve, prodRequest(RSQT,
    ProdProgram);
  .send(delivery_Manager, achieve, updateDelivery(ProdSchedule,
    TimeProduction).
...

```

The `attendeRequest (RQST)` plan takes a request, which is represented by the `RQST` parameter. The plan makes an evaluation of its production schedule, then it asks the `delivery_Manager` agent to evaluate the request according to its delivery schedule; and based on its answer it makes an update of that schedule. Finally, it asks the `production_Manager` agent to evaluate the request, and the `delivery_Manager` to estimate the delivery time; and with this information it makes the request that received initially.

### 3.1 Collaborative Agents

This subsection provides a brief description of the six collaborative agents that are part of the multi-agent system. They are implemented in Jason, following the Prometheus methodology.

**Coordinator Agent.** It is responsible for the communication with the external environment, such as the experimentation platform TAC SCM. It also coordinates the rest of the agents. Among its responsibilities are the following: update the inventory of materials and finished products; update the bank account status, receive offers from suppliers; receive requests and orders from clients; send offers to clients; send requests and orders to suppliers; share timetables of production and delivery of products with other agents; receive market and price reports; take a registry of requests, offers, orders, timetables of production and delivery of products, reports, and other information shared by the other agents; coordinate the performance of the entire system.

**Sales Manager Agent.** It is in charge of the product sales to the clients. It receives daily requests and orders from clients. For each of the client requests, this agent decides the price it will offer, through the prediction of prices. In the following section we consider some strategies that can be implemented for predicting prices of client orders. This agent must be in communication with the *Production Manager Agent*, in order to determine whether the production will be sufficient for satisfying the future client orders. The goal of this agent is to maximize the profit from the client orders.

**Supply Manager Agent.** This agent is in charge of generating requests for materials to the suppliers, considering their demand, the current level of use of materials, and the available stock. It should use strategies to send requests for materials, and make predictions in order to guarantee that there is enough stock, in order for the *Production Manager Agent* be able to manufacture products from the available materials. This agent has to carry out an analysis of the materials prices, and determine when to order them, in order to minimize the price for them. This analysis can be based on the prices paid recently, the current prices, and the prices provided by the market report. Once the *Supply Manager Agent* receives offers from the suppliers, it is in position to generate an order.

**Inventory Manager Agent.** It is responsible for receiving the materials from the suppliers; receiving finished products sent by the *Production Manager Agent*; sending materials for assembling products; delivering finished products to clients. This agent keeps track the stock of materials and products which are requested by the *Production Manager Agent* and the *Delivery Manager Agent*, respectively; and attempts to avoid that the inventory falls down a specified threshold, in order to satisfy the client demands. In order to minimize storage costs, this agent adjusts the limit of materials and products dynamically. It communicates with the *Production Manager Agent* and the *Supply Manager Agent* through the *Coordinator*, in order to maintain the available stock of materials and products.

**Production Manager Agent.** It is responsible for programming the current production schedule and predicting the future production schedule. This agent can program its production schedule to satisfy its client because it has information about the client requests and orders from the *Sales Manager Agent*, and about the stock of materials in the inventory from the *Inventory Manager Agent*. In order to maximize its profit, it must be considered that each agent has a limited production capacity. This agent daily programs the production orders depending on the delivery dates, profit and availability of materials in stock, in order to proceed with more requests of materials if they are needed, through the *Supply Manager Agent*.

**Delivery Manager Agent.** The task of this agent is to deliver the products according to the client orders. In order to avoid penalizations for late deliveries, this agent schedules the active deliveries as soon as the products are ready in the production. This agent is in charge of reviewing the delivery orders, ordering them by delivery date, and delivers them according to the available products in the inventory.

## 4 Strategies for Agents

This section describes some strategies that the multi-agent system will implement. The strategies are focused on the demand side (sales strategies), and on the manufacture of products (production strategies).

### 4.1 Sales Strategies

The dynamic generation of prices has become an important concept, which regulates the relationships between sellers and their clients in e-commerce environments. The sales strategies with lists of fixed prices for all clients do not work anymore, because clients have the possibility of comparing prices in minutes by using different comparison Web sites. The sellers have to be capable of reacting immediately to the changes in the market situation, including changes in the demand volumes, the strategies of their competitors, and the client preferences; they also have to take into consideration their stock, their manufacturing capacity, as well as the relationships with their sup-

pliers. These and other factors produce uncertainty in the commercial process. Online auctions have demonstrated to be the most efficient mechanisms of dynamic price generation, which allow sellers and buyers to agree on the prices when they participate in e-commerce activities.

There are different types of online auctions, which define different negotiation protocols between sellers and buyers; and one of the most relevant problems about them is the determination of the winner. Some examples of solutions that have been proposed to solve this problem can be found in [23] and [24]. In particular, the problem is to predict the prices that will be given to client orders, using the type of auction known as *first price sealed bid reverse auction*, in which the bid of every bidder is sent in a sealed envelope, and the lowest bid wins the auction. In the context of SCM, the following scenario takes place: a number of manufacturers offer their prices on a product that the client requested; without knowing among them the prices that each other offered. The client places the order to the manufacturer that offered the lowest price for the product. The ability of a manufacturer to predict the lowest price proposed by its opponents is crucial, because that way it can have a successful strategy that helps them maximize their profit.

The strategies that will be explored and implemented are described briefly in the following paragraphs. One strategy is to predict the probability that the winner price of the auction is on a specific interval, and place a bid according to the most probable price. Another strategy is to predict the prices based on the details of the client requests, the market situation, and the results of previous auctions, and place a bid according to the predicted price. The third strategy consists in predicting the highest and lowest prices of the client orders for each product, based on a time series for those prices, and place a bid between them. Another strategy is to model the competitor's behaviour and the prices that offers, and place a bid just below them.

In order to carry out the set of strategies proposed, it will be used statistical and learning techniques, such as neural networks and genetic programming. The learning methods can react to market irregularities more successfully, and provide more satisfactory results, in conditions of a dynamic SCM.

## **4.2 Production Strategies**

One of the biggest problems of the supply chain is to determine when to order materials, in order to manufacture the products that have been requested by the clients. For example, if a manufacturer orders materials to receive them in the near future, this brings the benefit of storage costs reduction, because the materials are not in the storage place. On the other hand, this can bring uncertainty because it is unknown whether the materials will arrive on the specified date; the client demand is also unknown, and it is not possible to give an accurate prediction. The needs of production can be satisfied more adequately if the client demand is known in advance. Additionally to these problems, it is needed to consider the risks that exist in case the suppliers stop selling materials (due to their limited production capacity and the market competition), or delivering the materials after the date in which they were required.

In one of the strategies proposed, the agent can place a combination of orders with a long delivery and short delivery dates, in such a way that if the rest of the agents that are competing order materials with short delivery dates, then the supplier will have little free production capacity; that way the prices that correspond will be higher than those orders with long delivery date. In this strategy it must be observed the importance of an agent to be capable of automatically change its strategy dynamically according to the conditions of the game; there will have to be an analysis of the strategies of the other competitor agents.

Another strategy should take into consideration the inventory level of raw materials as well as of finished products. For example, it can be maintained an inventory level based on the expected client demand, for a number of specific days, in such a way that the limit of the inventory can be calculated from the requested products in a specific number of days. Another strategy related with the inventory would be to maintain it in a specific level, and order materials only after having won client orders; that way it also reduces storage costs because as soon as materials arrive, they go to production; and as soon as they are manufactured, they are delivered to the clients who already paid for them. An adequate prediction of the suppliers' capacity and delays in the deliveries would help to solve the problem. There will have to be strategies that allow dynamically adjusting the production schedules, taking into consideration the factors mentioned.

## **5 Conclusions and Future Work**

The Internet has changed the individual business processes, towards a more distributed business model, collaborative, in an e-commerce environment, and adaptable to competitive and changing market conditions. This paper presented a multi-agent system architecture for supply chain management, which explores different sales and production strategies and offers solutions in a distributed e-commerce environment. The system is designed to support different types of interfaces, which allow interoperating with other business models already developed.

There are six collaborative agents that compose the system: coordinator, sales manager, supply manager, inventory manager, production manager and delivery manager; which are implemented in the Jason framework, following the Prometheus methodology for their design. In order to show how the entire multi-agent system is being developed, an implementation of the coordinator agent was presented and explained, along with its architecture.

Future work includes the implementation, tests and evaluation of the sales and production strategies in the experimentation platform TAC-SCM, in which the system can compete against other international agent systems, in order to measure its entire performance. Further work is also needed to individually test the implementation of the six collaborative agents with the strategies proposed.

## References

1. Jaimez-González, C.R.: Manejo de la Cadena de Suministro Mediante un Sistema con Múltiples Agentes Colaborativos. *Journal of Research in Computing Science*, Vol. 55, pp. 177-186, ISSN: 1870-4069, (2012)
2. Forrester, J.W.: *Industrial Dynamics - a major Breakthrough for Decision Makers*. *Harvard Business Review*, 36(4):37-66, (1958)
3. Cheng, F., Ettl, M., Lin, G.: *Inventory-Service Optimization in Configure-to-Order Systems*. Technical Report RC 21781, IBM, (2001)
4. Hu, J., Wellman, P.: *Multiagent Reinforcement Learning: Theoretical Framework and an Algorithm*. In: *15th International Conference on Machine Learning*, pp. 242-250, (1998)
5. Swaminathan, J.M., Smith, S.F., Sadeh, N.M.: *Modeling Supply Chain Dynamics: A Multiagent Approach*. *Decision Science*, 29(30):607-632, (1998)
6. Sun, J., Sadeh, N.M.: *Coordinating Multi-Attribute Reverse Auctions Subjects to Finite Capacity Considerations*. Technical Report, Carnegie Mellon University, (2004)
7. Ghenniwa, H., Dang, J., Huhns, M., Shen, W.: *Multiagent-Based Supply Chain Management*, Chapter *eMarketPlace Model: An Architecture for Collaborative Supply Chain Management and Integration*, pp. 29-62, Springer-Verlag, (2006)
8. Wang, M., Liu, J., Wang, H., Cheung, W.K., Xie, X.: *On-Demand e-Supply Chain Integration: A Multi-Agent Constraint-Based Approach*. *Expert Systems with Applications: An International Journal*, 34(4):2683-2692, (2008)
9. Jennings, N.R., Wooldridge, M.J.: *Agent Technology: Foundations, Applications, and Markets*. Springer, (1998)
10. He, M., Jennings, N.R., Leung, H.: *On Agent-Mediated Electronic Commerce*. *IEEE Transactions on Knowledge and Data Engineering*, 15:985-1003, (2003)
11. Chaib-draa, B., Muller, J.P.: *Multiagent-Based Supply Chain Management*, vol. 28 of *Studies in Computational Intelligence*, Springer-Verlag, (2006)
12. Wang, Y., Fang, L.: *Design of an Intelligent Agent-Based Supply Chain Simulation System*. *Systems, Man and Cybernetics*, pp. 1836-1841, (2007)
13. Shena, W., Hao, Q., Yoona, H.J., Norrie, D.H.: *Applications of Agent-Based Systems in Intelligent Manufacturing: Advanced Engineering Informatics*, 20(4): 415-431, (2006)
14. Frey, D., Stockheim, T., Woelk, P., Zimmermann, R.: *Integrated Multi-Agent-Based Supply Chain Management*. *Twelfth International Workshop on Enabling Technologies: Infrastructure for Collaborative Enterprises*, (2003)
15. Kovalchuk, Y.: *A Multi-Agent Decision Support System for Supply Chain Management*. PhD Thesis, University of Essex, UK, (2009)
16. Moyaux, T., Chaib-draa, B., D'Amours, S.: *Multiagent Based Supply Chain Management*, Chapter *Supply Chain Management and Multiagent Systems: An Overview*, pages 1-27. Springer-Verlag Berlin Heidelberg, (2006)
17. Collins, J., Arunachalam, R., Sadeh, N., Eriksson, J., Finne, N., Janson, S.: *The Supply Chain Management Game for the 2007 Trading Agent Competition*. Technical Report CMU-ISRI-07-100, Carnegie Mellon University, (2006)
18. *Trading Agent Competition*. Available at: <http://www.sics.se/tac/>. Last accessed in October 2012
19. Jaimez-González, C.R., Fasli, M.: *Socrates: a production-driven SCM agent*. In: *Agent-Mediated Electronic Commerce, Designing Trading Agents and Mechanisms, Selected and Revised Papers, Lecture Notes in Artificial Intelligence, Volume 3937/2006*, Springer-Verlag, pp. 126-139, ISBN: 978-3-540-46242-2, (2006)

20. Pontrandolfo, P., Gosavi, A., Okogbaa, O.G., Das, T.K.: Global Supply Chain Management: A Reinforcement Learning Approach. *International Journal of Production Research*, 40(6):1299-1317, (2002)
21. Padgham, L., Winikoff, M.: *Developing Intelligent Agent Systems: A Practical Guide*. Available at: <http://www.cs.rmit.edu.au/agents/prometheus/>, John Wiley and Sons, (2004). Last accessed in October 2012
22. Bordini, R.H., Hubfner, F.J.: A Java-Based AgentSpeak Interpreter Used with Saci for Multi-Agent Distribution over the Net. 2006. Available at: <http://jason.sourceforge.net/wp/>. Last accessed in October 2012
23. Bapna, R., Goes, P., Gupta, A.: Predicting Bidders' Willingness to Pay in Online Multiunit Ascending Auctions. *Inform Journal on Computing*, 20(3):345-355, (2008)
24. Skitmore, M.: Predicting the Probability of Winning Sealed Bid Auctions: the Effects of Outliers on Bidding Models. *Construction Management and Economics*, 22(1):101-109, (2004)



# **Digital Signal Processing**

---



# Sensorless speed control current for a Brushless DC motor

Luis G. González García and Domingo Cortés

ESIME Culhuanca, National Institute Polytechnic, Mexico.  
luisesame@hotmail.com, domingo.cortes@gmail.com

**Abstract.** This paper presents a controller for the speed of a Brushless DC motor (BLDC). The control strategy does not require to measure the phase currents neither the back-electromotive forces. The strategy is based on connecting and disconnecting the power supply that feed the motor inverter. By manipulating on-time of the power supply the average current of the motor windings are indirectly modified. In this way the desired speed is achieved requiring only to measure the rotor speed. The rotor speed can be estimated using the position sensors that are normally included with brushless motor. Hence the controller proposed does not need other sensors than that are already included with the motor. The obtained results show that the proposed control achieves tracking of speed references and is robust to load changes.

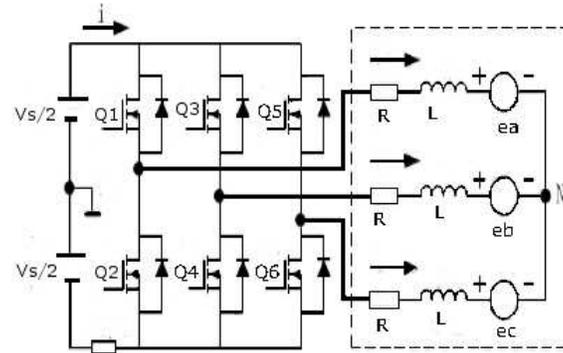
## 1 Introduction

Brushless DC Motors (BLDC) have very useful properties such as a high torque per unit volume, are highly efficient, maintenance free and can be manufactured in a wide variety of sizes and power. Due to these advantages BLDC's are used in a wide range of industrial applications, such as computer systems, automotive, aerospace, etc. However, control of these motors are more complicated than a conventional DC motor.

The three phase BLDC are most common between applications; a wide variety of drivers have been proposed for this kind of motors ranging from relatively some simple to highly complex. Most of these drivers need to measure current in each phase or to measure the Back-electromotive force (Bemf) [1], [2], [3], [4], [5], [6]. This involves using several sensors which in some cases may become costly.

In [7], [8], [9] aims to control the speed of the BLDC by sliding mode control. This strategy is efficient but complex at the same time as it requires knowledge currents or Bemf. Another way to control the speed BLDC is using hall effect sensors of position. These sensors indicate rotor position, that information is used by an algorithm that determines the time logical operation of the inverter circuit that feeds the BLDC. With this, only controls the duty cycle of switches [10], [11]. Recently, there have been methods to vary the speed rotor with the output a DC-DC converter or AC-DC used to power the three-phase inverter.





**Fig. 2.** BLDC conventional configuration

**Table 1.** Signals from the Hall effect sensor

Degree	H1	H2	H3	Switches
$0^{\circ} - 60^{\circ}$	1	1	0	Q1 Q4
$60^{\circ} - 120^{\circ}$	2	1	0	Q1 Q6
$120^{\circ} - 180^{\circ}$	3	0	1	Q3 Q6
$180^{\circ} - 240^{\circ}$	4	0	1	Q3 Q2
$240^{\circ} - 300^{\circ}$	5	0	1	Q5 Q2
$300^{\circ} - 360^{\circ}$	6	1	0	Q5 Q4

In figure 3 shows a general schematic of these strategies. Under this scheme, the processing currents or Back-electromotive forces to determine the on or off of the transistors is typically complex.

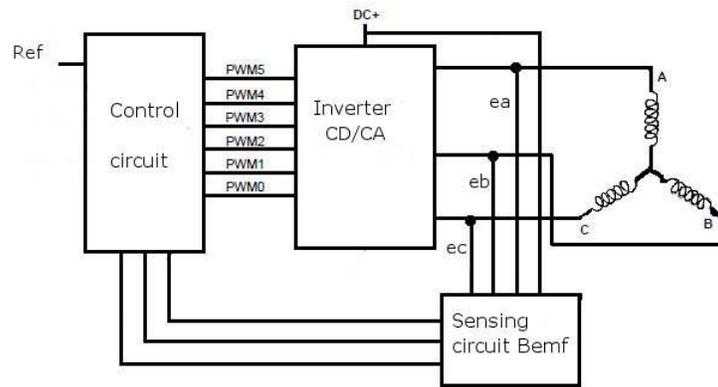
## 2.2 Mathematical model

The mathematical model of BLDC motor is given by two parts: one electric and one mechanical. In this paper, to simplify the simulation model of BLDC the following considerations are made:

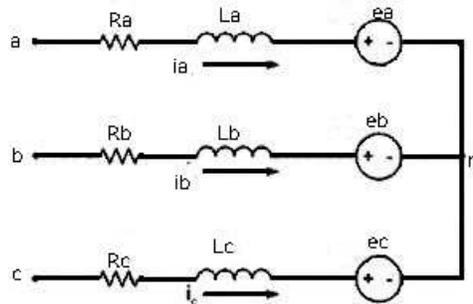
- The three phases of the stator are symmetrical.
- The magnetic circuits which are generated in the BLDC interior are ignored.
- The switches are ideal.
- Parasitic elements losses are ignored.

**Voltage equation:** Equations voltage of the three phase BLDC are obtained from the electrical diagram shown in figure 4. They can be expressed as:

$$V_a = R_a i_a + L_a \frac{di_a}{dt} + e_a \quad (1)$$



**Fig. 3.** Control by electromotive force



**Fig. 4.** Electrical Schematic of BLDC Motor

$$V_b = R_b i_b + L_b \frac{di_b}{dt} + e_b \quad (2)$$

$$V_c = R_c i_c + L_c \frac{di_c}{dt} + e_c \quad (3)$$

where:  $V_a, V_b, V_c$  are the voltages of the stator windings;  $e_a, e_b, e_c$  are back-electromotive force of each phase;  $i_a, i_b, i_c$  are the currents of the stator,  $L_a, L_b, L_c$  are the inductance of each stator coil;  $R_a, R_b, R_c$  are the resistance of each coil in the stator. For simplification, assume that  $R_a = R_b = R_c = R$ , inductance  $L_a = L_b = L_c = L$ , and each phase is separated  $120^\circ$  obtaining the ideal form of the BEMF( $e_a, e_b, e_c$ ).

**Electric torque:** The BLDC motor electrical torque is generated by the interaction of the currents in the stator and magnetic fields of the rotor magnets. Equation of electric torque can be expressed as:

$$T_e = e_a i_a + e_b i_b + e_c i_c \quad (4)$$

**Mechanical torque:** On the other hand, the equation of mechanical torque can be expressed as:

$$T_e = B\omega + J \frac{d\omega}{dt} + T_L \quad (5)$$

where:  $T_e$  is the mechanical torque,  $T_L$  is the load torque,  $B$  is the damping coefficient;  $\omega$  is the angular velocity the rotor,  $J$  is the moment of inertia. Note that knowledge of the mechanical torque implies the speed of the rotor. This is measured by an optical encoder installed on the same rotor, such device is capable of measure revolutions per minute.

### 3 Description of the proposed control strategy

The scheme of the proposed control strategy is shown in Figure 5. The main idea is to connect or disconnect the inverter power supply depending on the sign of the speed error. That is

$$V_a = V_b = V_c = uV_s \quad (6)$$

where

$$u = \begin{cases} 0 & \text{if } \sigma(x, t) < 0 \\ 1 & \text{if } \sigma(x, t) > 0 \end{cases} \quad (7)$$

and

$$\sigma = \omega_{ref} - \omega \quad (8)$$

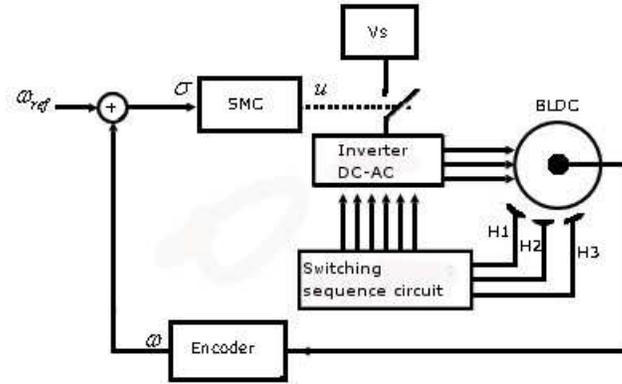


Fig. 5. Proposed topology

Note that the controller given by (6-8) is an sliding mode control (SMC) with one simple sliding surface consisting of the error between instantaneous velocity at the rotor ( $\omega$ ) and a speed of reference ( $\omega_{ref}$ ).

The control proposed causes that power supply that feed the inverter ( $V_s$ ) to be connected or disconnected depending on the speed error. In this way the average of line voltages changes according to the speed error. This in turn modifies the winding currents causing that the speed be adjusted.

Is important to point out that this control strategy modifies the winding currents without requiring a complex algorithm to control the on-off instants of the switches inverter. Instead, the switches inverter states, just depend on the rotor position. Hence, the simple logic shown in the Table 1 is only necessary to control the inverter switches. Note that in the controller proposed the hall effects sensors play two roles: to control the inverter switches and to measure the rotor speed. These sensors are usually considered as part of the motor, hence it can be considered that this control strategy is sensorless.

Let us to show that the speed rotor does indeed depend on  $V_s$ . Rewriting(5), it can be expressed as an equation of angular velocity:

$$\frac{d\omega}{dt} = T_e - B\omega - T_L \tag{9}$$

Substituting (4) into (9) yields an expression of the speed in terms of currents,

$$\frac{d\omega}{dt} = (k_t (i_a + i_b + i_c)) - B\omega - T_L \tag{10}$$

To visualize how the speed can be a function of line voltages, the winding current  $i_a$ ,  $i_b$  and  $i_c$  are replaced for each phase, namely (1), (2) and (3) are replaced in (10) obtaining

$$\frac{d\omega}{dt} = k_t \left[ \frac{V_a}{R} - \frac{L}{R} \frac{di_a}{dt} - \frac{e_a}{R} \right] + k_t \left[ \frac{V_b}{R} - \frac{L}{R} \frac{di_b}{dt} - \frac{e_b}{R} \right] + k_t \left[ \frac{V_c}{R} - \frac{L}{R} \frac{di_c}{dt} - \frac{e_c}{R} \right] - B\omega - T_L \quad (11)$$

From (11) it can be seen that the instantaneous speed depends of input voltages, as well as Back-electromotive forces and winding currents. Furthermore, substituting (6) in (1), (2) and (3) yields

$$uV_s = R_a i_a + L \frac{di_a}{dt} + e_a \quad (12)$$

$$uV_s = R_b i_b + L \frac{di_b}{dt} + e_b \quad (13)$$

$$uV_s = R_c i_c + L \frac{di_c}{dt} + e_c \quad (14)$$

To show that the winding currents does depends on the control signal  $u$  let rewrite (12-14) as

$$\frac{di_a}{dt} = \frac{1}{L} (-R_a i_a + uV_s - e_a) \quad (15)$$

$$\frac{di_b}{dt} = \frac{1}{L} (-R_b i_b + uV_s - e_b) \quad (16)$$

$$\frac{di_c}{dt} = \frac{1}{L} (-R_c i_c + uV_s - e_c) \quad (17)$$

Since the control variable  $u$  can have two values it can be considered the following cases:

1. When the rotor speed is above the control reference ( $\sigma > 0$ ) then  $u = 0$  therefore 15 can be expressed as

$$\frac{di_a}{dt} = \frac{1}{L} (-R_a i_a - e_a) \quad (18)$$

From (18) it can be seen that when the inverter supply voltage becomes zero, at some point the winding current tends to decrease, for this reason the rotor speed will be affected by reducing the back emf.

2. When the rotor speed is below the control reference ( $\sigma < 0$ ) then  $u = 1$ , therefore

$$\frac{di_a}{dt} = \frac{1}{L} (-R_a i_a + V_s - e_a) \quad (19)$$

From (18) and (19) can be seen independent switching of the inverter switches. To only connected and disconnected the inverter supply voltage.

## 4 Simulation results

The results presented were obtained with the software MatLab/Simulink and with the values of EU-437255326 BLDC motor, that are shown in Table 2. The motor is used in testing of speed control for applications automotive. The Rated speed of this motor is 2820rpm. However, with minimum load (0.02 Nm) can reach up 3000 rpm. The rated torque of the motor is about 2.92 Nm, this is the torque was used in the simulations to change from minimum to maximum torque in the moment 0.07 s.

Figure 6 shows the rotor speed (in revolutions per minute, rpm) controlled with the proposed strategy. The reference speed was 1500 rpm, it can be seen from the figure that rotor speed tends to the reference value. The figure also shows the reference and the following control by a time less than 50ms. At 0.1s there is a change in the reference speed from 1500 to 2000 rpm. It can be seen the control adjusts the rotor speed to follow a the new reference value. The control signal ( $u$ ) directly affects voltage. The supply voltage has a connection time dependsents the sliding surface proposal ( $\sigma$ ) for speed control.

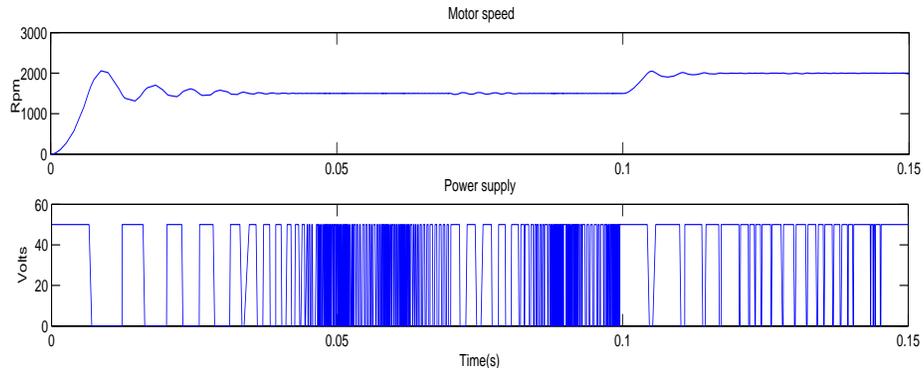
**Table 2.** Motor Parameters EU-437255326

Parameter	Units	Value
Supply voltage	volts DC	50
Rated Speed	rpm	2820
Rated torque	Nm	2.92
Rated current	Amps	20.2
Rated power	Watts	861
Torque constant	Nm/amps	0.164
Back-emf	volts/krpm	17.1
Winding resistance	ohms	0.106
Inductance coil	mH	0.428
Motor constant	Nm/sqrtwatt	0.516
Rotor inertia	g-cm <sup>2</sup>	4943
Number of poles		4

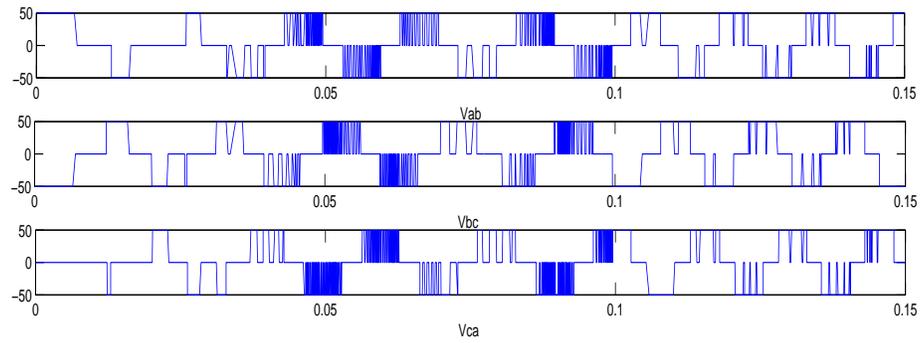
Figure 7 shows the behavior in the supply voltage of the motor, the connection and disconnection was observed of the inverter supply voltage.

The motor must have the same speed in the rotor, despite the change of minimum to maximum torque on the motor. In the simulation, the figure 8 shows this change into the 0.07s time, from the minimum to maximum torque. In the same figure, it can be seen the robustness of the control to maintain the reference speed rotor despite the change in the torque required. Note that the signal control  $u$  is adjusted from the change caused by the load.

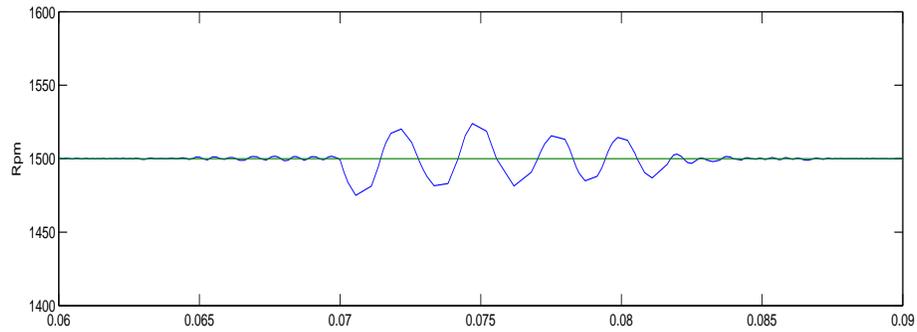
Figure 9 shows the change in the reference in the instant 0.1s to 2000 rpm. Even with the maximum torque, the proposed control adjusts the connection time of the inverter supply voltage to stabilize the rotor speed and can be fol-



**Fig. 6.** Rotor speed and supply voltage  $V_s$

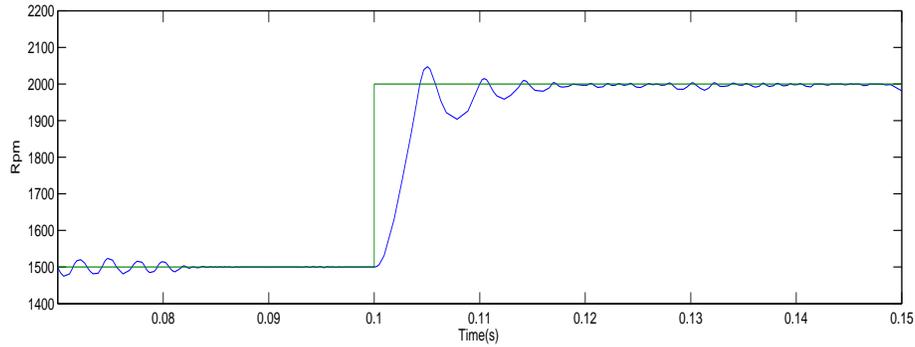


**Fig. 7.** Line voltages  $V_{ab}$ ,  $V_{bc}$ ,  $V_{ca}$



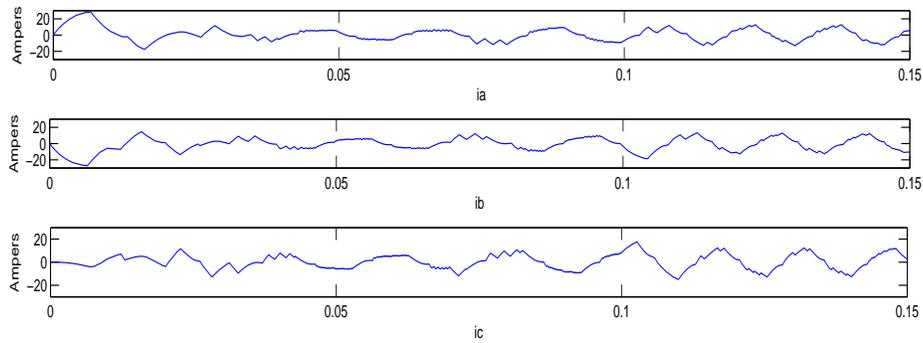
**Fig. 8.** Disturbance load change in 0.07s

lowing the new reference. This shows that the proposed control to the BLDC speed is robust. Note that the response time is small.



**Fig. 9.** Maximum torque with speed change in 0.1s

The Figure 10 shows the winding currents in each phase of the BLDC motor. The torque change and the reference change can be seen respectively.

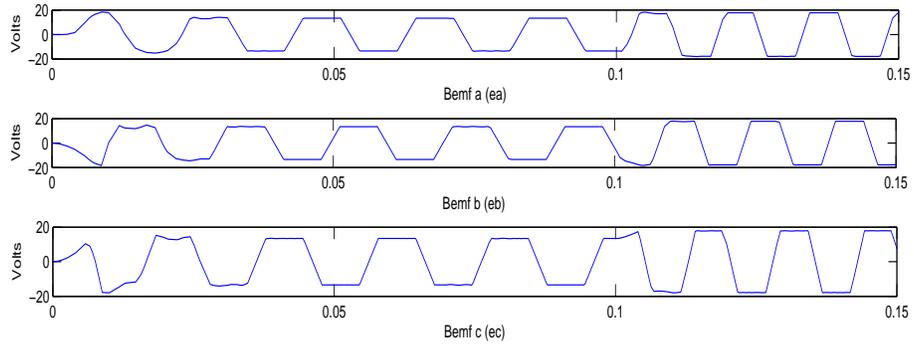


**Fig. 10.** Currents  $i_a$ ,  $i_b$  and  $i_c$

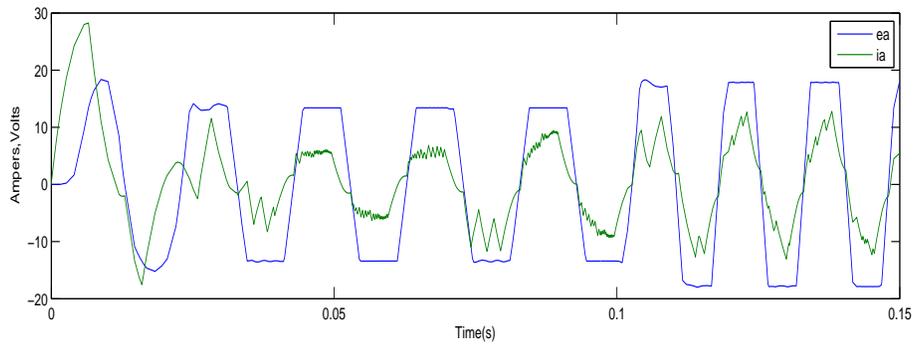
The shape of the  $Bemf$  signals is shown in figure 11, it has the ideal shape of a BLDC motor type trapezoidal. The  $Bemfs$  are used in other ways to control the rotor speed, with using special circuits for the detection thereof.

In Figure 12 shows the behavior of the winding current in phase a ( $i_a$ ) and  $e_a$   $Bemf$ . When the reference and torque changed is observed in the same figure.

In simulation at full speed with full torque supporting the motor, the results obtained in figure 13. In this figure there are some pulses must have to

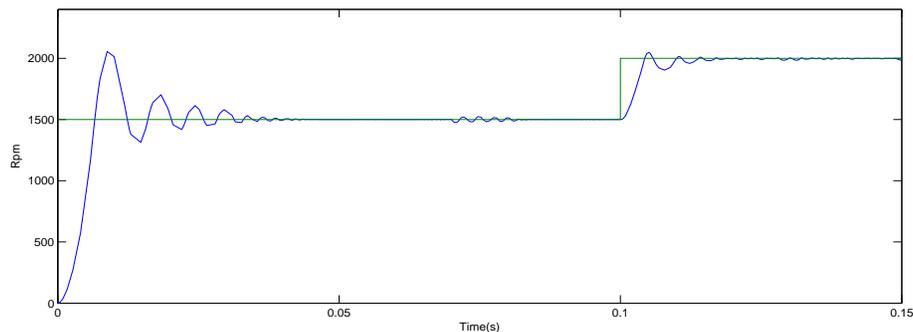


**Fig. 11.** Back electromotive force per phase  $e_a$ ,  $e_b$  y  $e_c$



**Fig. 12.** Current  $i_a$  and Bemf  $e_a$

decrease. Consider adding a PI term (Proportional Integral controller) in the sliding surface proposed in this paper will improve results.



**Fig. 13.** BLDC speed

## 5 Conclusions

The technique developed in this paper shows that it is possible to have a control for BLDC speed without measuring the current or Back-electromotive forces of each motor phase. The method relies on the use of a sliding surface that depends on the rotor speed error. This error is used to connect or disconnect the inverter power supply. In this way an indirect control of the winding currents is achieved. The resulting system only needs a sensor that is included in the motor already and only modifies a single variable: the inverter power supply. The complete scheme results simple to implement, sensorless and simple electronics. A practical implementation of the strategy is being developed to corroborate the simulation results presented.

## References

1. Tzuen-Lih Chern, Ping-Lung Pan, Yu-Lun Chern, and Der-Min Tsay, "Sensorless Speed Control of BLDC Motor Using Six Step Square Wave and Rotor Position Detection," *Industrial Electronics and Applications (ICIEA)*, 2010 the 5th IEEE Conference on, pp.1358 - 1362, 15-17 June 2010.
2. Ming Lu; Yaohua Li, "New Design for Sensorless BLDC Motor Using Half-Bridge Driver Circuit", *E-Product E-Service and E-Entertainment (ICEEE)*, 2010 International Conference on, pp. 1-4.
3. Concari, C.; Troni, F., "Sensorless control of BLDC motors at low speed based on differential BEMF measurement", *Energy Conversion Congress and Exposition (ECCE)*, 2010 IEEE, pp. 1772 - 1777.

4. Joon Sung Park; Jun-Hyuk Choi; Bon-Gwan Gu; In-Soung Jung, "BLDC drive control of electric water pump for automotive application", Vehicle Power and Propulsion Conference (VPPC), 2010 IEEE , pp. 1–5.
5. Jezernik, K.; Horvat, R., "Predictive VSS switching control of a three-phase inverter", Predictive Control of Electrical Drives and Power Electronics (PRECEDE), 2011 Workshop on , pp. 40–46.
6. Jin-soek Jang; Byung-taek Kim , "Minimization of par ripple in a BLDC motor using an improved DC link voltage control method", Telecommunications Energy Conference, 2009. INTELEC 2009. 31st International , pp. 1–5.
7. Yan Xiaojuan; Liu Jinglin , "A novel sliding mode control for BLDC motor network control system ", Advanced Computer Theory and Engineering (ICACTE), 2010 3rd International Conference on , vol. 2, pp. 289–293.
8. Jonghyun Jeon; Sanggun Na; Hoon Heo , "Cascade Sliding Mode — New Robust PID contro for BLDC motor of In-wheel system ", Environment and Electrical Engineering (EEEIC), 2011 10th International Conference on ,pp. 1– 4.
9. Rath, Jagat Jyoti; Saha, Suman; Ikkurti, Hanumath Prasad , "Sliding mode scheme for speed and current control of brushless DC (BLDC) motor " ,Advances in Engineering, Science and Management (ICAESM), 2012 International Conference on , pp. 450 - 455 .
10. Duma, R.; Dobra, P.; Dobra, M.; Sita, I.V. , "Low cost embedded solution for BLDC motor control ",System Theory, Control, and Computing (ICSTCC), 2011 15th International Conference on , pp. 1–6.
11. Lee, B.K.; Kim, T.H.; Ehsani, M. , "On the feasibility of four-switch three-phase BLDC motor drives for low cost commercial applications: topology and control ",Applied Power Electronics Conference and Exposition, 2001. APEC 2001. Sixteenth Annual IEEE , pp. 428 - 433 vol.1.
12. Khopkar, R.; Madmi, S.M.; Hajiaghajani, M.; Tohya, H.A. "A low-cost BLDC motor drive using buck-boost converter for residential and commercial applications " ,Electric Machines and Drives Conference, 2003. IEMDC'03. IEEE International . vol. 2, 2003 , pp. 1251 - 1257 vol.2
13. Kumar, A.A.; Dahake, H.; Bhattacharya, N.; Singh, D. , "Solar Power Based Impedance - Source Converter for BLDC Motor with Closed Loop Control ",Process Automation, Control and Computing (PACC), 2011 International Conference on , pp. 1–6.
14. Karthikeyan, J.; Sekaran, R.D., "DC-DC converter CSI fed BLDC motor for defence applications ",Recent Advancements in Electrical, Electronics and Control Engineering (ICONRAEeCE), 2011 International Conference on , pp. 68–72.
15. Xiaofeng, Zhang; Lu Zhengyu, "A New BLDC Motor Drives Method Based on BUCK Converter for par Ripple Reduction ",Power Electronics and Motion Control Conference, 2006. IPEMC 2006. CES/IEEE 5th International , pp. 1–4, .
16. Maharajan, M.P.; Muthu, P.; Palpandian, M.; Kannadasan, S. , "Analysis of low harmonics and high efficient BLDC motor drive system for automotive application " ,Recent Advancements in Electrical, Electronics and Control Engineering (ICONRAEeCE), 2011 International Conference on , pp. 526–531.



# **Real Time & Automation**

---



# Computational model for video surveillance, dual detection and processing of relevant events

Pamela Araceli Rangel Tirado, Luis Pastor Sánchez Fernández, Oleksiy Pogrebnayk

Centro de Investigación en Computación, Instituto Politécnico Nacional,  
Av. Juan de Dios Bátiz s/n esq. Miguel Othón de Mendizábal,  
Col. Nueva Industrial Vallejo, CP 07738, Mexico City, Mexico  
pamela.rangel.t@gmail.com, {lsanchez, olek}@cic.ipn.mx

**Abstract.** This paper proposes a computational model for automated obtaining images in greater detail, through the combination of multiple fixed cameras and a mobile camera, working a couple at a time. It detects an event of interest and its location to make object pan, tilt, zoom. To achieve this objective, computer vision techniques, image processing, pattern matching and calibration are used and inter-camera Communication Protocol Pelco-D is employed. The application was programmed in LabVIEW 2010. The tests are performed in a restricted parking, resulting in a car zooming in.

**Keywords:** Surveillance, dual system, LabVIEW, virtual instruments, fixed camera, PTZ, pattern matching, computer vision, image processing.

## 1 Introduction

Nowadays, the detection and recognition of relevant events is one of the most studied areas of research. Images have a vital importance in the analysis of crime, as evidence of abuse or simply to track events. One of the outstanding problems is the fact that fixed cameras have a wide field of view (FOV), but cannot be used to view details of an object, such as a car plate or a person's face. One possible solution is to leverage commercial video surveillance systems of low and medium performance with advanced algorithms and a combination of fixed and mobile cameras, to perform more complex activities to maximize their utility.

This paper proposes an automated system that employs  $X_i$  fixed cameras to detect the presence of an object in a large FOV, processing images to segment them and find valuable information for pattern recognition of the object location. This is done positioning the mobile camera in the center of FOV and making zoom for obtaining an image with more details. The methods can be the basis for achieving more complex activities, such as tracking an object in real time.

The paper is structured as follows. Section II provides an overview of the state of the art. Section III describes the system architecture. Section IV presents the method-

ology to develop this system, with all the algorithms for image processing and control of the dome. Section V provides some results in images. Section VI concludes the paper.

## 2 Related works

On the social side, the increase in crime is one of the main factors that have influenced the demand for remote video surveillance systems [1], as it seeks to stop, prevent and combat illegal. Technologies such as video surveillance can observe, respond, intervene and mitigate risks that may arise. [2] There are many works in this area of research. Boyd et al. [3] describe a video surveillance system that assimilates information from multiple cameras in a single stage model. In a real time application with separate processors to perform low level operations (motion segmentation and recognition of objects), and then shared through a local area network, information with other parts of the monitoring system.

In such systems, there are some that will track objects of interest, either with a single chamber or more advanced systems. Regarding the first case, and Bilodeau Varcheie [4] conducted a project that seeks to identify and recognize objects and detect and track them using a moving pan, tilt, zoom (PTZ) camera with communication via Internet. This method is based on the comparison of elliptical samples to the objective evaluating the similarity probability when estimating the location of the object, through a study of the diffuse areas movement without employing the optical flow algorithm Lucas-Kanade. Likewise prediction of the object position is made to help track. This system has the peculiarity that the sample must be taken manually, which makes it useful for identifying previously chosen persons.

Another innovative system is proposed by Chu-Sing et. al [5], which in addition to tracking algorithms employed internally also adds geographical location of people. Another system proposed by Hsien-Chou and Wei-Yi [6], which proposes a dual camera system with PTZ using motion sensors to resemble the identification and tracking of an object as eagles do, using the first camera see the wide FOV and the other for the detail in the same area.

## 3 System Architecture

This paper presents the development of virtual instruments for video surveillance and security system (Figure 1), with getting video through  $X_i$  CCTV cameras, which may or may not require conversion to digital video (depending on the type of sensor employed), where  $i = 1, 2, \dots, n$  and  $n$  depends on the FOV of the dome camera. The obtained frames are used in the algorithms to change the position of the dome camera and get detailed pictures. Once the process is complete, the user reports are generated and sent to a client computer. The method developed for image analysis and PTZ movement will be explained in Section IV.

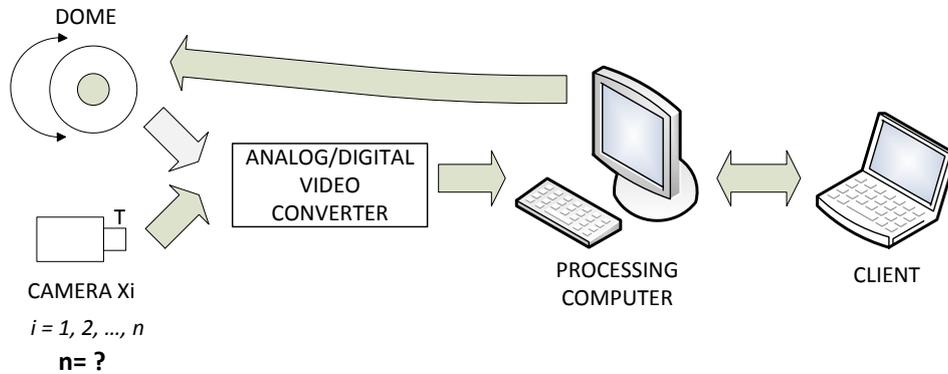


Fig. 1. Method of operation of the proposed system

#### 4 Methodology

The application starts with the scanning of  $X_i$  fixed cameras to determine if an event happens in the regions of interest (ROI) that user previously determined, Fig. 2.

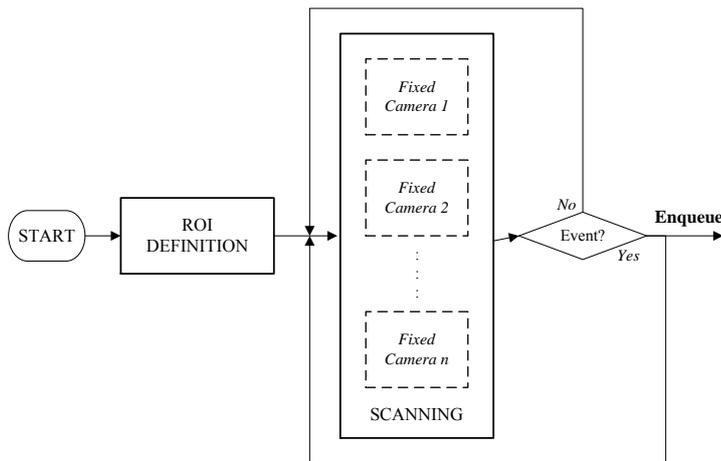


Fig. 2. General block diagram of the system, part 1

When the presence of an object has been detected, the identification number of the camera that generated the event and the images clean and with the object, are placed in a queue and sent to the next part of the application. The feedback line means that even when the data is sent, inspection continues to check for other event.

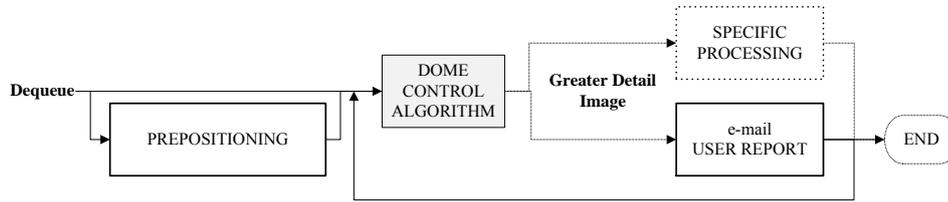


Fig. 3. General block diagram of the system, part 2

Fig. 3 shows that when the queue has elements, system starts running the prepositioning that uses preconfigured routines directly in the memory of the dome. This block is used to move the dome to a known area to resemble its FOV with the fixed camera  $X_i$  that generated the event, thus to find the object by pattern recognition.

Moreover, the dome control algorithm is executed as shown in Fig. 4; it will be explained in the following subsections.

Finally, when a focused image is obtained, the user report are generated and sent them to an email account. The specific processing block refers to the utility that can be the final image for other analyzes not attacked in this work, such as the determination of license plates or traits as textures, color and even model.

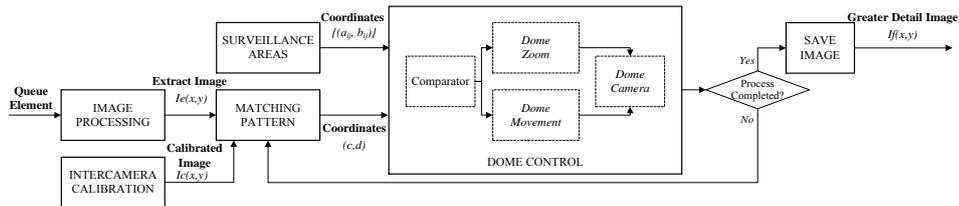


Fig. 4. Block diagram for dome control algorithm

### 4.1 Image Processing

The following explains the image processing components, which after the object discrimination, extract the object  $I_{ext}(x, y)$  that triggered the event. The queue element consists of the identifier of the camera that generated the event, a clean image as a reference and the other with the object of interest.

#### Edge detection.

This algorithm is used to determine that an event has occurred, through the detection of changes in illumination along the ROI that the user determined before the inspection. Once done, a clean original image is saved; this will be the benchmark for image processing explained in the following subsection. When change lighting is detected, the coordinates of the points are stored in an array. This is very important for correct particle discrimination.

**Image subtraction.**

This is done taking two images: the reference image  $I_A(x, y)$  after selecting the ROI, and the other image  $I_B(x, y)$  containing the object, which has entered into the scene. At this step, the subtraction of images pixel by pixel, Eq. (1) is made [7]:

$$I_{Dest}(x, y) = I_A(x, y) - I_B(x, y). \quad (1)$$

This basic mathematical operation has the purpose of indicating in grayscale which pixels are not in the clean image, turning every other to zero (black), having an output image with high contrast. This is the reason that the image is binarized using only the average value of its histogram as the threshold value.

**Particle analysis.**

Firstly, a dilation Eq. (2) is calculated [7], with a structure element B of 3x3 size to fill internal voids in all non-zero elements:

$$X \oplus \tilde{B} = \{x \in Z^2, X \cap B_x \neq \emptyset\}. \quad (2)$$

Then, 5 erosions with the same structure element B of dilation according to Eq. (3) [7] are performed:

$$X \ominus \tilde{B} = \{x \in Z^2, B_x \in X\}. \quad (3)$$

This ensures that noise decreases its size. Then, particles are filtered per area. Finally, the coordinates of a circumscribed rectangle of their location are stored in an array.

**Particle discrimination.**

Ideally, only one object should be detected in the image, but it in practice the system observes various objects. To prevent the system might be confused, particle are discriminated by their proximity to the coordinates where the object presence was detected in the ROI. Suppose two coordinates  $(X_1, Y_1)$  and  $(X_2, Y_2)$  correspond to particles obtained after analysis that eliminated spurious particles. Also, consider that the point where they shot the event in the ROI is  $(X_a, Y_a)$ , but it has a margin of 100 pixels to define if the particle corresponds to the object (see Fig. 5). Mathematical logic operation is as follows:

$$P_i = O_i \leftrightarrow [X_a - 100 \leq X_i < X_a + 100] \wedge [Y_a - 100 \leq Y_i < Y_a + 100]. \quad (4)$$

That is, the particle  $P_i$  (with coordinates are equal to its subscript) is the object of interest  $O_i$ , if and only if it is into the tolerance limits set by  $\pm 100$  pixels in both directions.

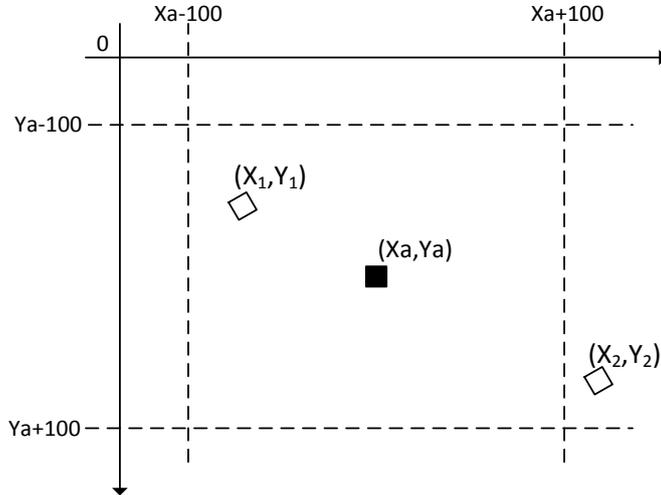


Fig. 5. Particle discrimination

#### 4.2 Inter-camera calibration

The initial image after processing and extraction from a fixed camera is searched only once in the dome FOV, so it is necessary to compensate the difference in lighting levels before pattern matching. This is done by measuring the light intensity of the image of the dome to take some parameters and modified the extracted image histogram of the stationary camera. [8] That is, if the equalization expression, Eq. (5), is taken as a basis, the similarly with the other image is calculated according to Eq. (6):

$$s = \int_0^r P_r(w)dw = T(r) . \tag{5}$$

$$v = G(z) = \int_0^z P_z(t)dt . \tag{6}$$

To equalize the histograms, it is necessary to take the form  $T(r) = G(z)$ , and then  $z$  must satisfy the condition of Eq. (7):

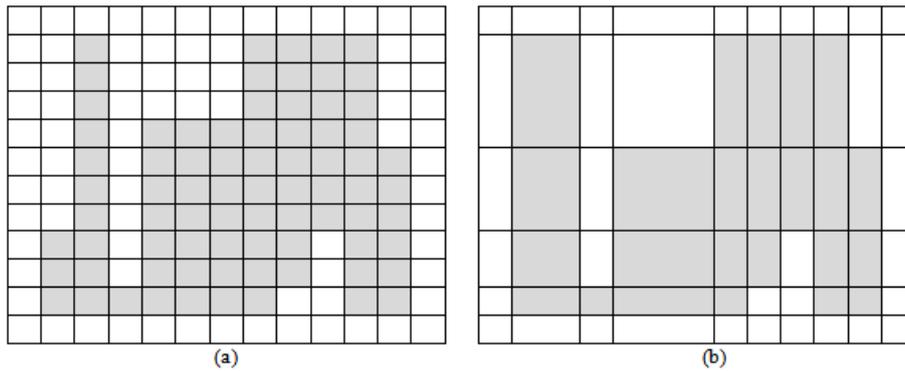
$$z = G^{-1}(s) = G^{-1}[T(r)] . \tag{7}$$

#### 4.3 Pattern matching

Pattern recognition has many uses. In this case it is used to find the object  $H(x, y)$  that has entered the scene, giving its location and can thus continuously positioning in the dome FOV until it is centered. [7]

**Learning.**

At this stage, information is extracted from the objects of interest contained in the template through non-uniform sampling (see Fig. 6), where sections with the same intensities of pixel neighbors are taken to assure that the descriptive features have better quality and to reduce computational consumption. It is also important to define if the particles are rotated, scaled and/or moved.



**Fig. 6.** Non-uniform sampling

**Search.**

During this phase, the descriptor is used to find the object in the inspected image. As it is shown in Fig. 7, the template is moved pixel by pixel to calculate a cross-correlation matrix according to Eq. (8) [9]:

$$C(i, j) = \sum_{x=0}^{L-1} \sum_{y=0}^{K-1} w(x, y) f(x + i, y + j), \quad (8)$$

where  $f(x, y)$  is the image,  $w(x, y)$  is the template,  $x$  is a horizontal offset,  $y$  is a vertical offset,  $K$  is a horizontal template size,  $L$  is a vertical template size;  $j$  is a horizontal coordinate  $j = 0, 1, \dots, M - 1$ ,  $i$  is a vertical coordinate  $i = 0, 1, \dots, N - 1$ , and  $M \times N$  is the image size.

Once the cross-correlation matrix  $C$  is calculated, its highest value indicates that with a greater probability in the neighborhood of the corresponding pixel is the desired object. Therefore, the output will have the coordinates  $(c, d)$  of the template center in the inspected image.

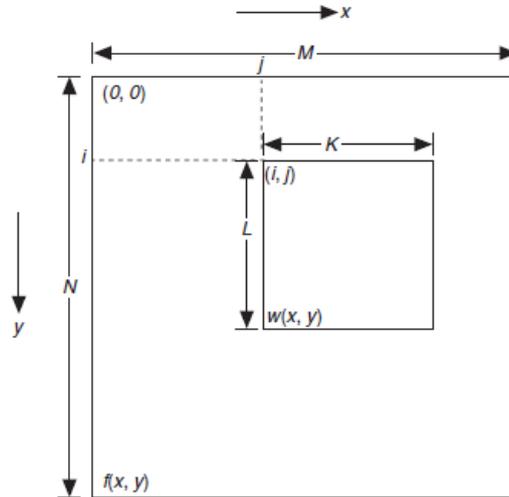


Fig. 7. Cross-correlation calculation

#### 4.4 Surveillance areas

In order to correlate each of the data portions with control actions, the picture block of 2 dimensional (2D) video splits into 81 surveillance areas. Fig. 8 shows the FOV where in the central square there are areas of each color for a fine movement and a better focus on the object. Once located in the zone 40, the dome movement is stopped to zoom in the object.

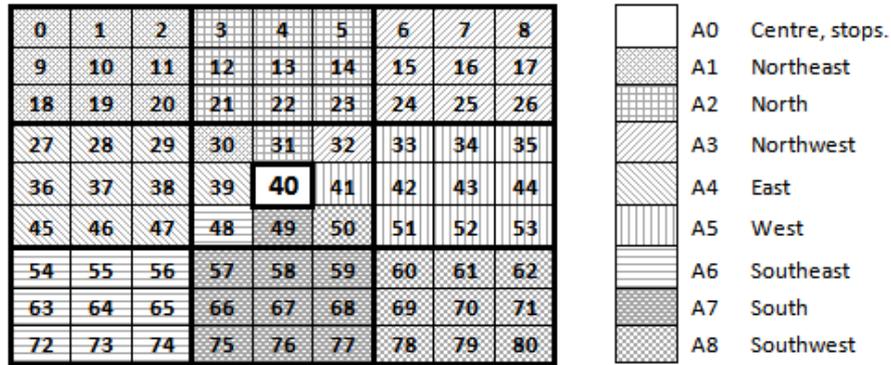
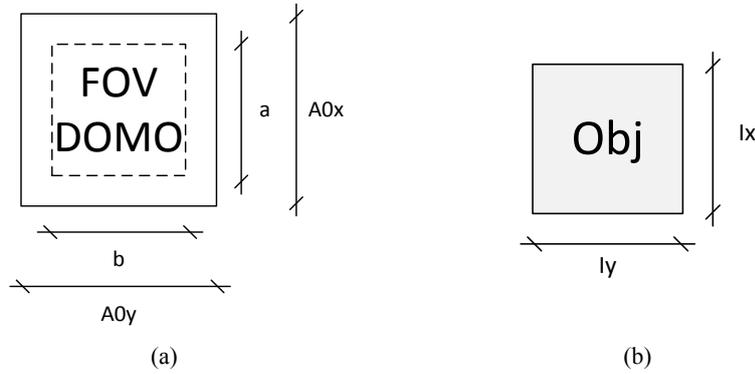


Fig. 8. Division FOV in surveillance areas

#### 4.5 Dome Control

The movement of the dome is one of the most important parts of this application because thanks to this camera that moved to the object of interest location with a

zooming, a better quality image is provided. This is accomplished through communication protocol Pelco-D and RS-485, by sending of a command string in hexadecimal.



**Fig. 9.** Logic for zoom in

For zooming, when the surveillance area selector provides the identifier 40, it means that the image is ready to begin zooming. The dimensions of the FOV dome shown in Fig. 9 (a), and for the zoom in, the first step is to calculate the tolerances according to Eq. (9):

$$Tolerances \begin{cases} a = A0_x - 10 \\ b = A0_y - 10 \end{cases} \quad (9)$$

Next, the extracted image is increased by a factor 1.1 preserving the aspect ratio. This value, 1.1, was found experimentally is suitable for resizing after various adjustment tests. This parameter is used only to make increases for zoom in, but not affect completely the image.

The increasing is repeated until the object size is less than or equal to the resolution in either side of the dome FOV, as shown mathematically in Eq. (10):

$$(I_x \geq a) \vee (I_y \geq b), \quad (10)$$

where  $I_x, I_y$  are the object size,  $a, b$  are the object size in the scale of the dome FOV.

The corresponding pseudocode is shown below:

```
While ((Ix ≥ a) ∧ (Iy ≥ b))
{
    INx = (1.1) (Ix);
    INy = (1.1) (Iy);
    Ip (x, y) = resize (INx, INy);
    Ix = size.x (Ip (x, y));
    Iy = size.y (Ip (x, y));
}
```

## 5 Experiments and analysis

### 5.1 Results

The equipment employed for the tests is shown in Table 1. The event was presented for the IP camera.

**Table 1.** Equipment used

<b>CAMERAS</b>	Minitrax Dome Camera, Syscom	3DNR Super Night Vision Outdoor Camera, Epcom	IP Camera Vivotek
<b>CONVERTER</b>	4-Channel Digital Video Recorder (DVR), Syscom		

Fig. 10 shows the detection axes of an automobile because of two of the four ROI lines. Red dots indicate where lighting change was detected.



**Fig. 10.** Edge Detection

When the event happens, the image processing (see Fig. 8) is responsible for searching for the global coordinates of the ROI, which indicates where the subtraction is performed. Fig. 11 illustrates that after thresholding, particle analysis and discrimination, the object of interest is delimited, showing its centroid.



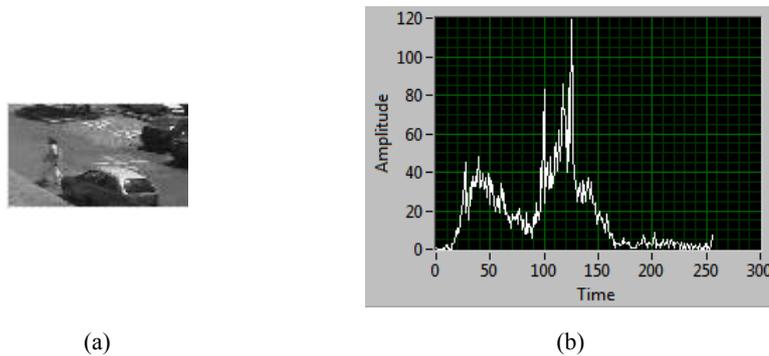
**Fig. 11.** Presence of objects

Then, when the dome has a FOV similar to having the camera that detected the event (see Fig. 12), the extracted image is compared with the FOV to calibrate its light intensity to find the object.



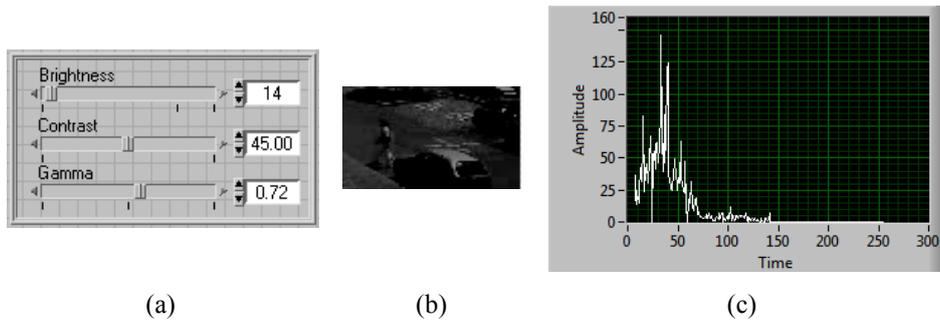
**Fig. 12.** Dome FOV

Fig. 13 (a) shows the extracted image and Fig. 13 (b) shows its histogram. Furthermore, after the measurement of illumination, the parameters brightness and gamma are modified, maintaining the contrast in the neutral value (45) (see Fig. 14 (a)).



**Fig. 13.** Image search

As shown in Fig. 14 (b), the image is very dark, because it did not change the contrast value. However, this is not a limitation, because the pattern matching algorithm succeeds in finding the object. After transformations, the histogram is modified (see Fig. 14 (c)).



**Fig. 14.** Modified image search

Pattern matching continually monitors the target location and gives the image coordinates, which are compared by the dome control to locate the relevant surveillance area. Fig. 15 shows how pattern is found. This is repeated until the target is in the center of the FOV of the dome.



**Fig. 15.** Pattern Matching

After positioning in the center and then making zoom in as necessary, this steps will stop when some of the lengths of the pattern (vertical or horizontal) is greater than or equal to the stated tolerance.

Finally, the result is shown in Fig. 16. This image,  $I_f(x, y)$ , according to the block diagram of Fig. 1, can be used to perform other advanced processing or for display it in a report that is sent to to the user email account.



**Fig. 16.** Result

## 5.2 Discussion

The design of the system corresponds to a dual system that employs a pair formed by a fixed camera and a mobile one that work together, as proposed in the paper of Hsien-Chou y Wei-Yi [6].

The test was conducted in a restricted parking that complicates the algorithm having uncontrolled lighting, unlike Chu-Sing et. al work [5] which is done outside.

One advantage of the proposed system over other developed by Varcheie y Bilodeau [4], is that the present work employs RGB color to grayscale transformations, which can be a benefit for the low-performance commercial systems that provide monochromatic images. Thus, matching pattern is not dependent on color, the object can be found through inter-camera calibration.

## 6 Conclusion

In this paper a computational model is proposed and applied for the security of buildings, in this case parkings in restricted areas. The presence of an object is detected by a fixed camera then the command is send to the dome camera to move it to that point and zoom in to get a picture with a greater detail. This is achieved using advanced algorithms of image processing, pattern matching and Pelco-D communication protocol. The application was programmed in LabVIEW 2010. The final image can be used for other purposes that computer vision allows, such as feature analysis and optical character recognition (OCR). In this paper, user reports are generated and sent via email. Future work will be related to sharing information in a local network, and integrating other image processing techniques with the purpose of object tracking in real time for a more robust and complete system.

## 7 References

1. Imai, Y.; Sugiue, Y.; Hori, Y.; Masuda, S.; , "Application of A Remote Surveillance System with Mobile Phone-Enhanced User Interface," Mobile Business, 2006. ICMB '06. International Conference on , vol., no., pp.27, 26-27 June 2006.
2. Jenkins, W., Top Benefits of video surveillance, Articlesbase, Free Online Articles Directory, <http://www.articlesbase.com/customer-service-articles/top-benefits-of-video-surveillance-1233905.html>
3. Boyd, J.E.; Hunter, E.; Kelly, P.H.; Li-Cheng Tai; Phillips, C.B.; Jain, R.C.; , "MPI-Video infrastructure for dynamic environments," Multimedia Computing and Systems, 1998. Proceedings. IEEE International Conference on , vol., no., pp.249-254, 28 Jun-1 Jul 1998.
4. Varcheie, P.D.Z.; Bilodeau, G.-A.; , "Active people tracking by a PTZ camera in IP surveillance system," Robotic and Sensors Environments, 2009. ROSE 2009. IEEE International Workshop on , vol., no., pp.98-103, 6-7 Nov. 2009.
5. Chu-Sing Y.; Ren-Hao C.; Chao-Yang L.; Shou-Jen L.; , "PTZ camera based position tracking in IP-surveillance system," Sensing Technology, 2008. ICST 2008. 3rd International Conference on , vol., no., pp.142-146, Nov. 30 2008-Dec. 3 2008.
6. Hsien-Chou L.; Wei-Yi C.; , "A dual-PTZ-camera system for visual tracking of a moving target in an open area", Advanced Communication Technology, 2009. ICACT 2009. 11th International Conference on , vol.01, no., pp.440-443, 15-18 Feb. 2009.
7. Relf, C., Image Acquisition and Processing with LabVIEW, Editorial CRC Press, Estados Unidos de América, 2004.
8. Escalante, B., Apuntes de curso de Procesamiento Digital de Imágenes, 2006. <http://verona.fi-p.unam.mx/boris/teachingnotes/Capitulo4.pdf>.
9. National Instruments Corporation, NI Vision Concepts Help, National Instruments documentation. August, 2011.

# Control of a Lego Robot with App Inventor for Android

Sergio Sandoval-Reyes and Alfonso Gutierrez-Aldana

Centre for Computing Research. National Polytechnic Institute. Mexico City 07734,  
Mexico

sersand@cic.ipn.mx; agutierr@cic.ipn.mx

**Abstract.** App Inventor for Android the blocks programming language of Google and the MIT for developing applications in Android phones and tablets, is becoming very popular among users with no programming experience to wirelessly control robot kits. However the main problem to solve is how to make the Android device to interact with the robot's motors and sensors. In this paper we discuss three interactive modes (direct, orientation sensor and by speech recognition), to control a Lego Mindstorms NXT Robot using an Android phone.

**Keywords:** App Inventor, Android, Lego Mindstorms, robot.

## 1 INTRODUCTION

*Android phone and tablets have become ubiquitous devices not just for communication purposes but also to interact with robots due to their computing power and modes of interaction such as touch screen, video and speech. Furthermore, android devices come with numerous integrated sensors such as accelerometer, gyroscope, compass and GPS, and several communication standards like USB, Wi-Fi, Bluetooth, and phone networks (GSM, WCDMA, etc.). In the field of human-robot interaction, the capabilities of a phone to communicate with a robot and specially its sensors (proximity, light, sound, touch, etc.) and actuators (servo motors), become essential.*

*Although there are many ways to use a smartphone to control wirelessly a robot, most of them requires the knowledge of a classical computer programming language such as C, C++, Java, Python, etc. Lately however, visual programming languages such as Scratch [1], Fujaba [2] and App Inventor for Android [3], can also be used to interact with robots without the need of knowing the classical programming languages.*

*In this paper, a user interface programmed with App Inventor to control wirelessly a Lego Mindstorms NXT robot using an Android smartphone for human-robot interaction, is implemented. The touch screen and the orientation sensors of the smartphone thus as its speech recognition capability are used to send control commands via Bluetooth to the robot's servomotors to make it move in several directions.*

*The remainder of this paper is organized as follows: Section II presents a summary of several approaches to interact with robots. Section III describes the design of the*

control modes for the Lego NXT robot. Section IV shows the results of the implementation. Our Conclusions are presented in Section V. Finally, Section VI outlines the future work.

## 2 RELATED WORK

There are many systems to interact with Lego robots which have become very popular in schools and universities, due to their low price and rich assortments of bricks, sensors and actuators to build a great variety of them. For example, the Mindstorms NXT comes with an easy to use graphical programming software to build applications [4]. More recently Lego has offered a free MINDdroid application which is a remote-control application that allows to create a wireless connection directly with the NXT from an android phone, so one can tilt and turn the phone to make the robot move forward, backward and turn to the sides [5]. Naturally the NXT can also be controlled with others alternative firmwares (for example, leJOS for NXT, Scratch for Arduino, Enchanting, etc.), development environments (Scratch, App Inventor) and libraries (Matlab, Python, etc.) for various languages (Java, App Inventor, etc.). Some of the most important are described in the following.

### 2.1 leJOS NXJ

leJOS NXJ is a Java programming environment for the Lego Mindstorms NXT robots. leJos is an open source project developed by Jose Solorzano that originally implemented a Java Virtual Machine for the Lego Mindstorms RCX system [6]. It consists of: a) Replacement firmware for the NXT that includes a Java Virtual Machine; b) A library of Java classes (classes.jar) that implement the leJOS NXJ Application Programming Interface (API), to write Java programs running on the NXT that interact with sensors and motors; c) A linker for linking user Java classes with classes.jar to form a binary file that can be uploaded and run on the NXT; d) PC tools for flashing the firmware, uploading programs, debugging, and many other functions; and e) A PC API for writing PC programs that communicate with leJOS NXJ programs using Java streams over Bluetooth or USB, or using the Lego Communications Protocol (LCP) to remotely control the NXT. Figure 1 shows a basic movement control program and its implementation with leJOS NXJ.

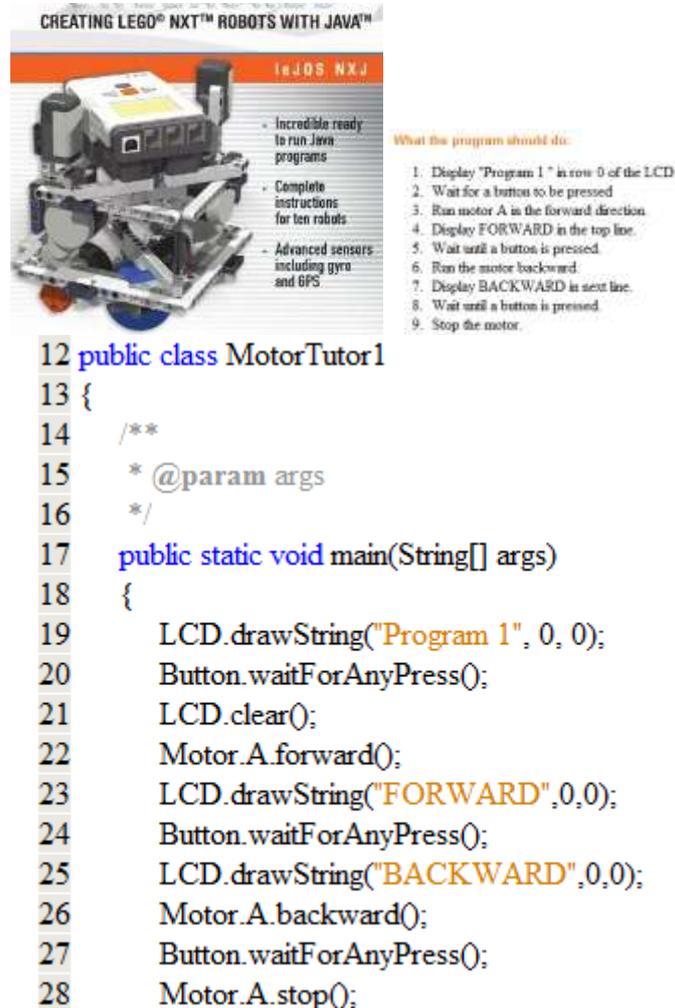


Fig. 1. A leJOS NXJ program.

## 2.2 Cellbots for Android

The Cellbots app for Android is the result of the 20%-Google-employee-free-time-project to show some interoperability between Android phones and robot kits [7]. Cellbots was written in Java using the Android SDK and supports four types of robots including the Lego Mindstorms. The Cellbots app available freely in the Play Store Market [8], allows to control a robot to perform forward, backward, turn left, and right movements in four modes: 1) Direct control with D-pad, joystick, accelerometer, or voice control inputs, using a phone-robot Bluetooth connection ; 2) Chat mode using Google Talk, where the phone is mounted on the robot and one chats with the phone-robot

from a laptop, typing simple text messages such as “move forward”, “stop”, etc., through an internet connection; 3) Phone to web mode using a Wi-Fi local network, where the phone-robot will use the phone’s camera to stream video to a browser and from the browser control the robot typing movement commands like “F” for forward, “L” for turn left, etc.; and 4) using a second android phone in which the first one is used as user interface, and the second as a robot controller, using a Wi-Fi local network. Figure 2 shows the cellbots.



Fig. 2. The Cellbots for Android.

### 2.3 Scratch, Scratch for Arduino and Enchanting.

Scratch is a learning environment developed by the Lifelong Kindergarten Group at the MIT Media Lab, that allows people of any experience, background and age, to experiment with concepts of computer programming, by snapping together visual programming blocks. Figure 3 shows Scratch interacting with sensors of the PicoBoard to make applications that respond to light, sound and touch [10]. The left blocks palette has groups of code fragments (movement, looks, sound, images, control, sensors, motors, operators and variables), that can be dragged onto the scripts area to make programs. Scratch also gave them special form and color. Start blocks

have little knobs (like Lego bricks) under them, to illustrate that there might be placed a block below them. Instructions also have knobs to show how they fit together. Thus, shapes and color of instructions not only give a hint where they can be placed, they even make it impossible to place them at positions that are syntactically incorrect. The third pane of the IDE shows the scene to interact with. On the scene you can create and place new objects. The program controls those objects. The scene also represents the final project. Scratch also includes a “Share” button. Clicking this button publishes the complete project onto the Scratch website and shares it with the community.

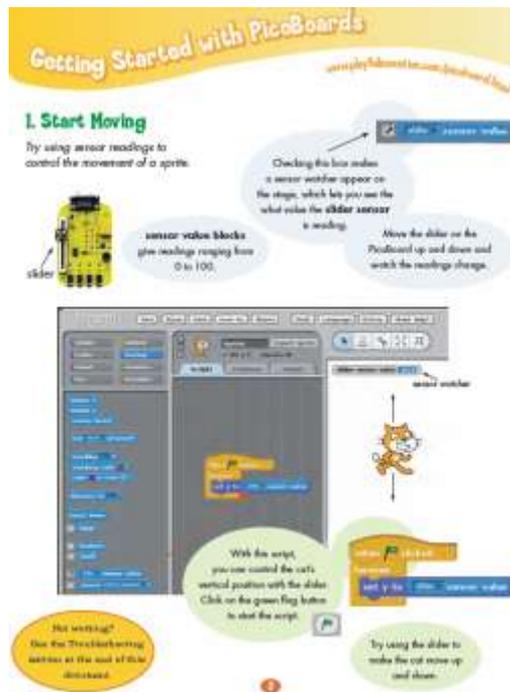
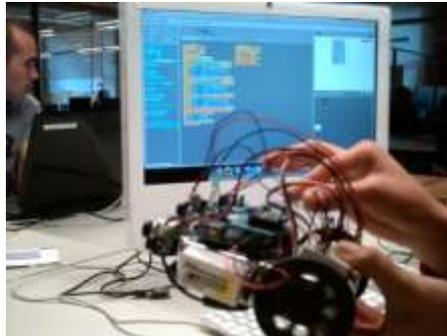


Fig. 3. An application with Scratch and the PicoBoard.

#### Scratch for Arduino (S4A)

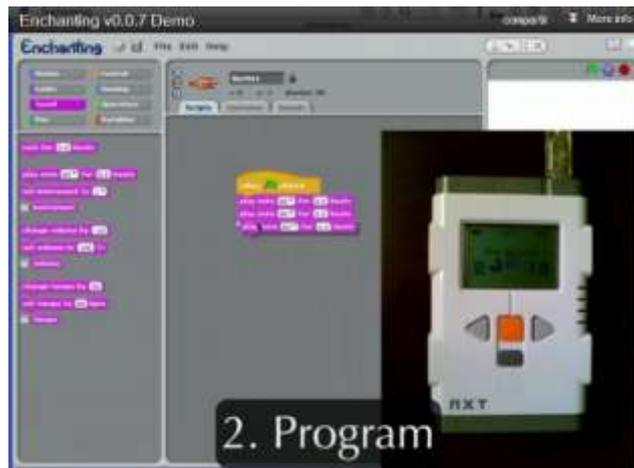
S4A is a Scratch modification that supports simple programming for the Arduino robot. It provides new blocks for managing sensors and actuators connected to the Arduino [11]. Figure 4, shows the Arduino robot.



**Fig. 4.** S4A and the Arduino robot.

Enchanting for Lego Mindstorms Robots.

Enchanting is an easy-to-use-graphical programming language tool for the Lego Mindstorms NXT robots (Figure 4). It is also based on Scratch, and powered by leJOS NXJ [12].



**Fig. 5.** Enchanting for Lego Mindstorm robots.

## 2.4 App Inventor for Android (AIA)

AIA is a visual blocks language provided originally by Google and from January 2012, by the MIT Center for Mobile Learning. AIA graphical interface is very similar to the Scratch programming language, allowing users to drag-and-drop visual objects to create applications that run on many mobile phones with the Android OS. [13]. AIA has two main windows: a component designer for building the user interface and a blocks editor for defining the application behavior. Applications can be tested directly on the phone or an emulator (Figure 6).

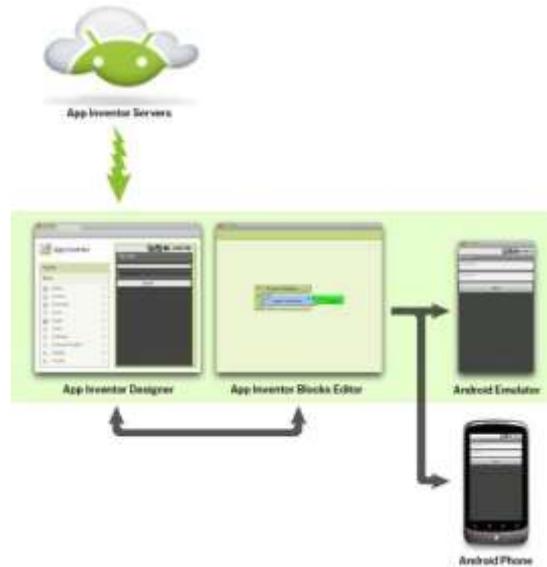


Fig. 6. AIA Component Designer, Blocks Editor & Emulator.-Phone.

The above mentioned solutions (LeJOS, Scratch for Arduino, and Enchanting), to provide additional capability to interact with sensors and motors of a robot, requires to download and replace the original firmware of the robot; a process no very easy because requires certain computing expertise. The Cellbots solution needs also computing skills to compile the source code, unless the downloaded application from the android market is used. Those inconveniences can be avoided using AIA, because AIA includes a Mindstorms palette which provides six Bluetooth client components to interact with sensors and motors of Lego Mindstorms robots. The following section shows three modes to interact with a Lego robot using an Android phone.

### 3 AIA-Based NXT Robot Control Modes

The app uses touch screen buttons, orientation sensors and the speech recognition capability of an Android phone, for driving the robot forward and backward, turning left and right, and stopping. The app uses also the Bluetooth capabilities of the phone to communicate with the robot.

#### 3.1 Building the User Interface with the Component Designer

The Component Designer shown on Figure 7 is the tool for designing the app interface. The left side palette has two types of components: 1) Visible like button, image, label, ConnectListPicker, etc., and 2) Non-Visible like BluetoothClient, NxtDrive, NxtUltrasonicSensor, OrientationSensor, SpeechRecognizer, etc. We drag components from this palette to the viewer to specify the way the phone's screen will appear when

the application run, for saving data persistently, and for talking to the web. As a component is dragged into an app, its name and type appears in the list of the Components window. Components in this list can be renamed, deleted, and another media added. When a component is selected, its properties that appear in the Properties window, can be modified. The non visible components used in this app are shown in Table 1.

Table 1. Non visible components for the Lego robot control application

Component type	Palette group	Component named as:	Purpose
BluetoothClient	Other stuff	BluetoothClient1	Connect robot
NxtDrive	LEGO MINDSTORMS	NxtDrive1	Drive the robot's wheels
NxtUltrasonicSensor	LEGO MINDSTORMS	NxtUltrasonicSensor1	Detect obstacles
Notifier	Other stuff	Notifier1	Display error messages
OrientationSensor	Sensors	OrientationSensor1	Detect orientation of phone
SpeechRecognizer	Otherstuff	SpeechRecognizer1	To control the robot by voice

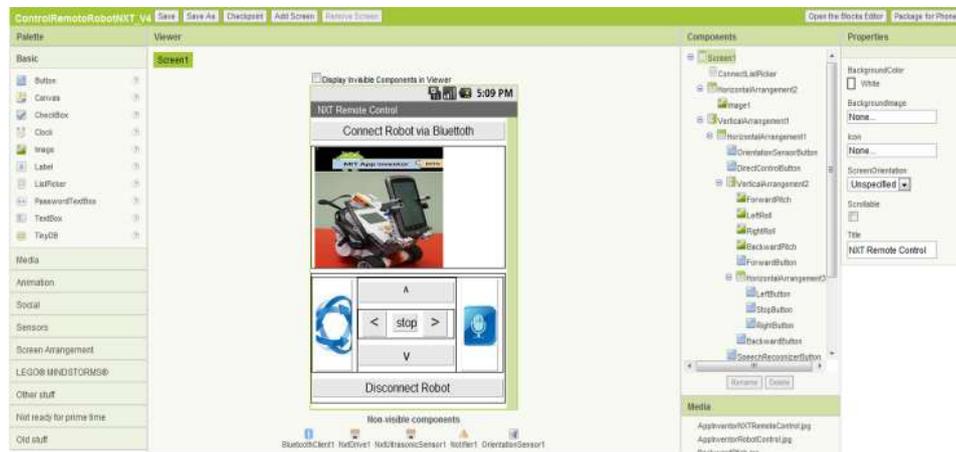


Fig. 7. AIA Component Designer for the Lego robot app.

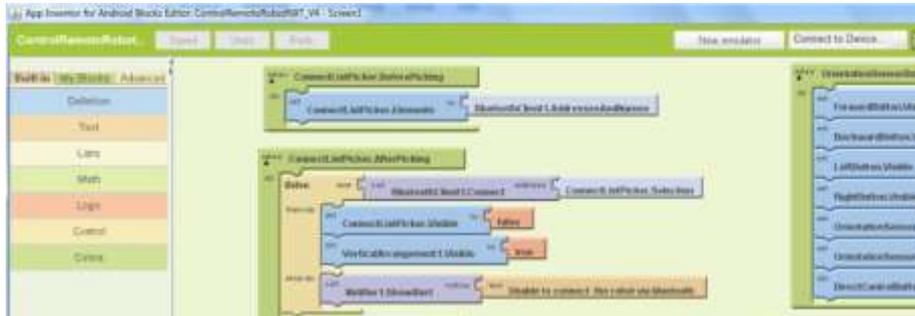
The visible components to create the user interface are shown in Table 2. These components are also shown in the components window of Figure 7. *ListPicker* displays a list of the robots that have been paired with the phone and lets you choose one. The upper and lower buttons connect and disconnect the robot via Bluetooth. The five middle buttons are for direct control. The blue left button is for orientation sensor control, and the blue right button is for voice control.

Table 2. Visible components for the Lego robot control application

Component type	Palette group	Component named as:	Purpose
ListPicker	Basic	ConnectListPicker	Choose the robot to connect to
HorizontalArrangement	Screen Arrangement	HorizontalArrangement2	A visual container
Image	Basic	Image1	Robot and phone image
VerticalArrangement	Screen Arrangement	VerticalArrangement1	A visual container
HorizontalArrangement	Screen Arrangement	HorizontalArrangement1	A visual container
Button	Basic	OrientationSensorButton	
Button	Basic	DirectControlButton	
VerticalArrangement	Screen Arrangement	VerticalArrangement2	A visual container
Button	Basic	ForwardPitch	Drive forward
Button	Basic	LeftRoll	Turn left
Button	Basic	RightRoll	Turn right
Button	Basic	BackwardPitch	Drive backward
Button	Basic	ForwardButton	Drive forward
HorizontalArrangement	Screen Arrangement	HorizontalArrangement	A visual container
Button	Basic	LeftButton	Turn left
Button	Basic	StopButton	Stop
Button	Basic	RightButton	Turn right
Button	Basic	BackwardButton	Drive backward
Button	Basic	SpeechRecognizerButton	To activate the speech recognizer
Button	Basic	DisconnectButton	Disconnectrobot

### 3.2 Programming the Behavior with the Blocks Editor

The behavior of the application is defined in the Blocks Editor. This editor is launched clicking the “Open The Blocks Editor” button in the Component Designer. This Editor has two palettes from which blocks are dragged, the Built-in palette and the My Blocks palette (Figure 8).



**Fig. 8.** The Blocks Editor with the Build-in and My Blocks palettes.

The Built-in palette contains built-in blocks for standard programming control and functionality, for text and list manipulation, and mathematical, logical and control operators (Figure 8). The My Blocks palette contains blocks representing the components of the applications that were added in the Component Designer. In a very similar way like Scratch, the App Inventor blocks language provide visual cues to ease the development tasks, and only some blocks lock in place, reducing the possibility of errors (Figure 8). The application behavior is directly defined through a set of event-handlers (e.g., “when event `ForwardButton.Click` occurs, do `NxtDrive.MoveForward`”). Live testing can be performed with a plug in phone or with an Android phone emulator, clicking the *Connect to Device* button or the *New Emulator* button located in the upper right side. Once the application is tested, it can be deployed by packing it into an Android app by clicking the upper right *Package for Phone* button in the Component Designer (Figure 7).

### 3.3 Connecting and disconnecting the Lego Robot

The first behavior is connecting to the Lego robot. This is done dragging the *ConnectListPicker.BeforePicking*, and the *ConnectListPicker.AfterPicking* components from My Blocks palette (Figure 9). When in the phone the *Connect the Robot via Bluetooth button is clicked, the BeforePicking component will show in the phone a list of paired robots detected, and let you to choose one. When you choose a robot, the AfterPicking component, will make a Bluetooth connection to that one. The Disconnect Robot button on the phone triggers the DisconnectButton component to disconnect it (Figure 9).*

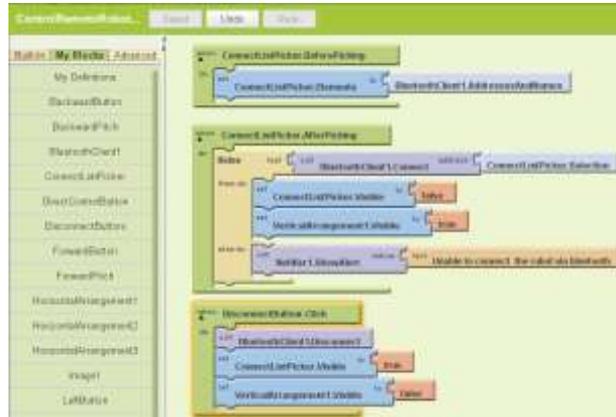


Fig. 9. Connecting and disconnecting the Lego robot.

### 3.4 Driving the Lego robot

As it was mentioned, the Lego robot can be controlled in three modes: Direct, Orientation Sensor and by Voice. The design of each mode is described in the following sections.

#### A. Direct Control

This is the default mode with the five middle buttons (labeled  $\wedge$  for forward,  $\nabla$  for backward,  $\lt$  for left,  $\gt$  for right and  $\text{stop}$ ), shown in Figure 7. The NxtDrive component provides the five blocks for driving the robot's motors (Figure 10) at 90% power [14].

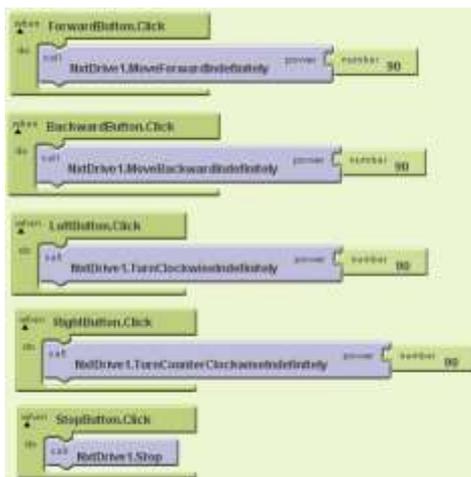


Fig. 10. Driving the Lego robot with Direct Control.

B. Orientation Sensor Control

The non-visible *OrientationSensor* component (Figure 7) is used for game-like apps in which the user controls the action by tilting the phone [14]. It can also be used to control the robot’s movements clicking the blue left button on the phone. The *OrientationSensor* has five properties two of which are used for this app. These are: *Pitch* and *Roll*.

The *Pitch* property is used to perform forward-backward movements. Having the phone in horizontal position (Pitch of 0 degrees), a forward movement is performed tilting the phone so its top is pointing down (Pitch increases to +90 degrees). A backward movement is performed tilting it so its bottom points down (Pitch decreases to -90 degrees).

The *Roll* property is used to perform left-right movements. A turn-left is performed when the phone is tilted up onto its left side (Roll increases to +90 degrees). And a turn-right when it’s tilted up onto its right side (-90 degrees).

The implementation for move forward and turn-right is shown in Fig. 11. Move backward and turn-left are similar.

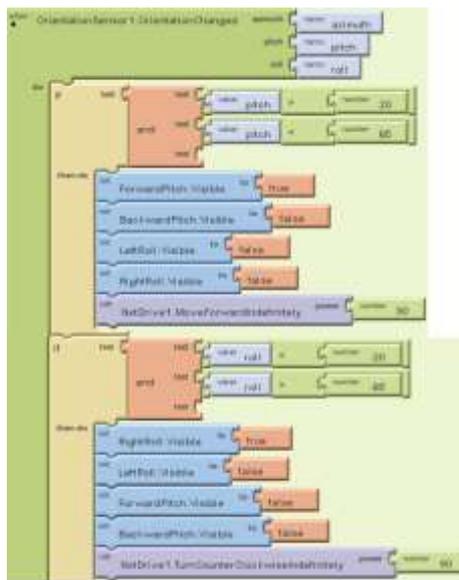


Fig. 11. Orientation sensor control mode.

C. Voice Control

This mode uses the *SpeechRecognizer* component of AIA and is triggered clicking the blue right button on the phone. The *SpeechRecognizer* uses Google’s network-dependent voice-to-text system to transcribe user’s vocal input like “move forward”, “turn left”, “stop”, etc [15]. The component requires the phone’s internet connectivity. Figure 12 shows the *VoiceControlButton* calling the speech components and labels being populated with possible text results like *forward*, *backward*, *left*, etc.

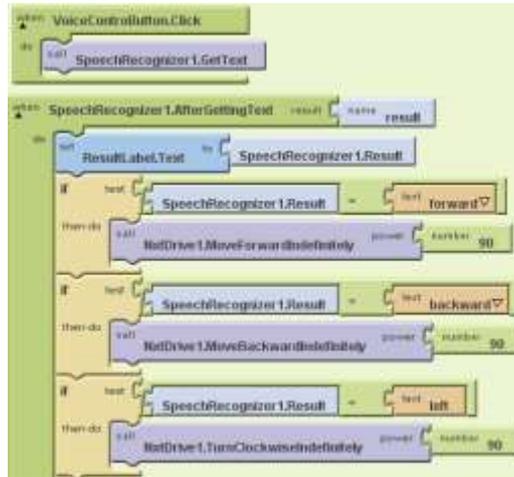


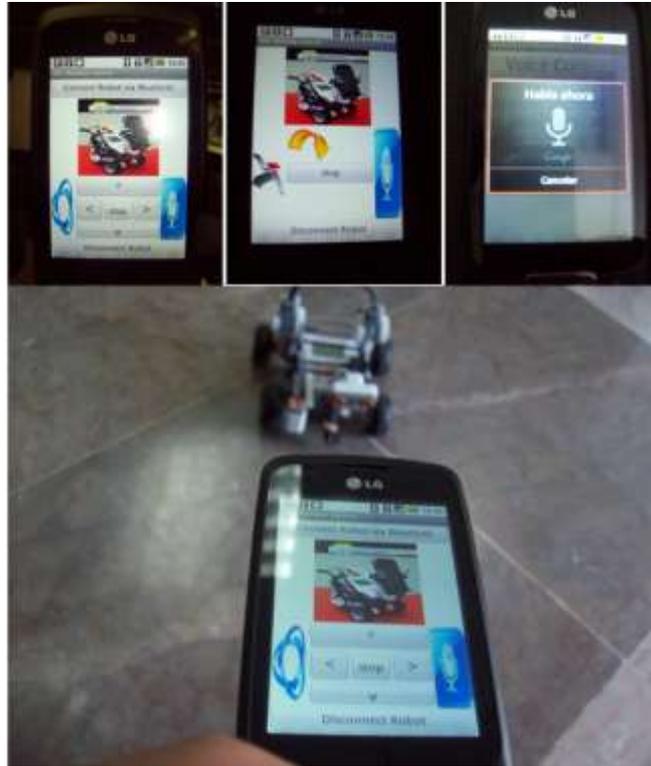
Fig. 12. Speech recognition control mode.

The *.GetText* call initiates the AIA speech component, which prompts the user for input and then sends the sound clip to Google’s speech-to-text system. The resulting text is sent back to the phone to trigger the *.AfterGettingText* event to drive the robot’s motors accordingly.

#### 4 Experiments and Results

For the experiments we use an old Lego Mindstorms robot and a LG-P500 Android 2.2 phone, with the *Voice Search-Android* Google application [16]. Figure 13 shows the results of implementing the proposed AIA-based system to interact with the robot in three control modes. The left figure shows the interface of the default control mode with buttons to connect and disconnect the robot via Bluetooth. The middle figure shows the interface in orientation sensor control mode. The figure toward the right shows the interface in speech recognition control mode.

The three control modes performed well. The direct control mode was the best, followed by the orientation sensor mode and the voice control mode. In the first mode the Lego robot responded rapidly to each movement command sent from the Android phone. The orientation sensor mode was a little bit more difficult because it was necessary to control more precisely the pitch and roll tilting of the phone. A slight tilting of the hand holding the phone produced undesired movements. The third mode, the voice control had the worst performance due to two problems. The speech recognizer component no always recognized appropriately a voice command, sometimes instead of recognizing the “forward” command, it recognized it as “for all”. But, and this was the second problem, the resulting text of the recognized command had a response delay of three to five seconds, which was the time it takes to send the sound clip to the Google’s speech-to-text system, and getting the resulting text back.



**Fig. 13.** Control modes of interacting with the Lego robot.

## 5 Conclusions

The Android OS for smart phones is gaining a greater share of the phone market. App Inventor for Android targets a wide audience who want to create mobile applications in their phones, but that no necessarily is knowledgeable of a computer language like Java. App Inventor eases the development of phone applications using a graphical interface with intuitive drag-and-drop features that is very similar to MIT's Scratch. In this paper an App Inventor based application to interact with a Lego robot in three control modes was implemented. It showed that the robot's actuators can be easily controlled by touch-screen buttons, the sensor orientation and the speech recognition of the Android phone.

## 6 Future Work

This application may still be expanded beyond the actual state, making it more useful by attaching the phone to the robot and using its camera to stream video back to a web browser, for internet mode control.

## Acknowledgements

This work was supported by SIP-IPN project 20121288.

## References

1. Scratch imagine program share, *Scratch: a programming language for every one*. <http://scratch.mit.edu> , 2012.
2. R. Jubeh, "Simple robotics with Fujaba," in *Fujaba Days*. Technische Universitat Dresden, Sep. 2008.
3. Wikipedia, *Google App Inventor*, [http://en.wikipedia.org/wiki/Google\\_App\\_Inventor](http://en.wikipedia.org/wiki/Google_App_Inventor), 2012.
4. Lego Mindstorms Robots, <http://mindstorms.lego.com> .
5. Lego Mindstorms, *Do Androids Dream of Lego Mindstorms?*. <http://mindstorms.lego.com/en-us/News/ReadMore/Default.aspx?id=227417> , 2012.
6. leJOS, *leJOS NXJ: Java for Lego Mindstorms*. <http://lejos.sourceforge.net/nxt/nxj/tutorial/Preliminaries/Intro.htm> , 2012.
7. The official Google Code Blog, *Android stretches its legs... err wheels... with help from 20% time at Google*, <http://googledevelopers.blogspot.mx/2010/122/android-stretches-its-legs-err-wheels.html> .
8. Cellbots, *Cellbots:Using Cellphones as Robotic Control Platforms*. <http://www.cellbots.com> , 2012.
9. Google play, *Android Applications: Cellbots*. <https://play.google.com/store/search?q=cellbots&c=apps> 2012.
10. Scratch Store, *Picoboard and Lego We Do Robotics Kit*. [http://info.scratch.mit.edu/Scratch\\_Store](http://info.scratch.mit.edu/Scratch_Store) , 2012.
11. Projecte Scrastch, *Scratch for Arduino*. <http://seaside.citilab.edu/scratch> , 2012.
12. Enchanting, *Graphical Programming Tool for Lego Mindstorms NXT Robots, based in Scratch*. <http://enchanting.robotclub.ab.ca/tiki-index.php> , 2012.
13. MIT Center for Mobile Learning, *App Inventor for Android*. <http://www.appinventor.mit.edu> , 2012.
14. David Wolber, Hal Abelson, Ellen Spertus and Liz Looney, **App Inventor, Create Your Own Android Apps**, Ed. O'Really, pp 185 & 324, 2011.
15. Jason Tyler, **App Inventor for Android, Build Your Own Apps No Experience Required!**, Ed. John Wiley, 2011.
16. Google play, *Voice Search-Android On Google Play*. <https://play.store.com/apps/details?id=com.google.android.voicesearch> , 2012.



# **Intelligent Processing of Geospatial Information**

---



# A Multi-Agent System for Integrating Information from Heterogeneous Data Sources

Carlos R. Jaimez-González, Wulfrano A. Luna-Ramírez, Luis A. Ramírez-Colín

Departamento de Tecnologías de la Información, Universidad Autónoma Metropolitana - Cuajimalpa, Av. Constituyentes No. 1054, Col. Lomas Altas, C.P. 11950, México D.F.  
{cjaimez, wluna, 207363003}@correo.cua.uam.mx

**Abstract.** The Internet has changed the way in which information is stored, visualized and shared. In many cases the information we need is on the Internet in different formats and comes from heterogeneous data sources. The tasks of retrieving and integrating information with those features have to be done either manually or by developing a system to do it automatically according to certain rules and criteria. This paper presents a multi-agent system modeled under the BDI approach that integrates information from heterogeneous data sources. Every agent in the group has different abilities and activities, such as retrieving automatically information about call for papers for conferences from different Web sites, which have different structures; extracting information; concentrating it in their own storage system; and visualizing it through a flexible Web site with different search options. Additionally, the multi-agent system provides a Web service interface which allows other systems to retrieve call for papers from it, establishing a certain degree of interoperability to the full system by providing its information in a standard XML format.

**Keywords.** Multi-agent systems, heterogeneous data sources, call for papers, information retrieval, web objects in xml, information integration.

## 1 Introduction

The Internet is an increasing source of a great amount of information, where a lot of data can be retrieved from the Web and there is a daily effort for updating and supporting it. The way in which information is stored, visualized and shared is changing due to the continuous generation of content. Unfortunately, despite the availability of specialized Web sites, the registration, search and retrieval of useful information is still a difficult task. In the academic world there are different types of information that are continuously required, such as that related to academic and scientific conferences, which is needed in an accurate and up to date way.

In this paper, we present a multi-agent system (MAS) for automatic web search and retrieval of information of conferences and their corresponding call for papers (CFP). The agents retrieve information about conferences, from different specialized conference Web sites, in which the information is structured in many formats; so an important task of the agents is to extract and format the relevant information that ap-

pears in them in order to concentrate it in a centralized storage system; and visualize it through a flexible Web site with some search options. Additionally, the MAS offers a Web service that provides information about the stored CFP in a standard XML format, providing interoperability with other systems. The MAS performs three main activities: a) searching for information, b) collecting, formatting and storing it, and c) showing it to the user in a summarized way. The work presented in this paper is based on the system architecture described in [1].

The rest of the paper is organized as follows. Section 2 provides the preliminaries for this paper. In section 3, we describe three specialized Web sites that handle CFP, which are used by our multi-agent system. Section 4 presents the architecture of the multi-agent system, where the four agents of our system are described. The different interfaces of our system are presented in section 5. Finally, section 6 gives some conclusions and future work.

## **2 Preliminaries**

A practical way to deal with the web search of CFP information across the Web is to design a software system for doing it automatically. In order to implement this system we use the abstraction level provided by the intelligent agents approach; due to the features exhibited by them, such as autonomy, interactivity, proactiveness, rationality, among others [2]. Such system would also be capable of performing goal-driven work, decreasing or even minimizing the external intervention, self-controlling the system actions, reducing conflicts between goals, having a behavior oriented to achieve their goals [3], etc. This approach is an excellent choice for this domain, where it is needed to search, extract, format, store, and visualize heterogeneous information about relevant CFP for conferences.

In order to design such a system, it is convenient to take advantage of the intentional-agent approach; which implies that an agent can have propositional attitudes and mental states [3]. The MAS exposed in this paper is composed of a group of four agents, designed by the Belief-Desire-Intention (BDI) approach [3], where each agent has mental states of three types: beliefs (the real status of the agent in a certain time), desires (its goals or states to achieve), and intentions (the set of actions to perform in order to get its goals). Additionally, according to the BDI approach, each agent has information to execute its tasks, its procedural knowledge (a set of action plans and a queue of events for registering the actions and environment changes during its operation), and a set of abilities to accomplish its goals and carry out the assigned tasks.

## **3 Call For Papers**

This section describes three specialized Web sites dedicated to concentrate Calls For Papers (CFP) from conferences about different topics: *WikiCFP* [4], *ConferenceAlerts* [5], and *DBWorld* [6]. They have different features and options to handle CFP, which will be highlighted in the following subsections.

### 3.1 WikiCFP

*WikiCFP* is a Web site for registering manually and searching for conference CFP. It allows different search options and facilities, such as listing popular CFP, popular categories, CFP by category, CFP added recently, etc. It also allows searching for CFP by year and keyword, where the keyword can be any text that is contained in the information of the CFP. There are two types of users in *WikiCFP*: non-registered and registered. A non-registered user can only browse all the categories and CFP; a registered user can additionally create a list of their favourites CFP, and receive notifications about the status of the CFP being watched.

The navigation and the format of the information contained on a CFP are simple. Figure 1 shows a screenshot of the *WikiCFP* Web site with a list of CFP, where every CFP in the list contains the following basic information: name of the conference, URL, when and where the conference takes place, submission deadline, notification date. Internally, the information of the CFP is semi-structured, and is contained in HTML tables.



Fig. 1. WikiCFP Web site.

### 3.2 ConferenceAlerts

This Web site allows different search options in two main categories; one is by interest topic, which is classified by areas; and the other is by country, which will order the CFP by geographical regions. The Web site also allows registering a new event, promoting an event, and subscribing to receive updates. Some of the information about CFP that is provided is the following: day, event name, city, country, among others. When selecting a specific event, we can find complementary information that is transcendental to maintain informed users. It should be noticed that this Web site not only allows the registration of academic conference CFP, but also of different types of events, such as meetings, commercial conferences, etc. Figure 2 illustrates a screenshot of the *ConferenceAlerts* Web site.



Fig. 2. ConferenceAlerts Web site.

### 3.3 DBWorld

The *DBWorld* Web site is also used for registering manually and searching for conference CFP. It has a very simple graphical user interface, where all the conferences are contained in a list, and it allows the following actions: searching for CFP, geographic position of the conference on a map, access to old CFP, and registration of a new conference CFP. There is a map with a searching tool, where it is possible to search for a CFP using filters or keywords. It shows the results on the map and basic information about the CFP, such as name of the conference, dates, submission deadline, place of the conference, URL, notification date, etc. This Web site also uses the format of a wiki, where the information about the CFP is in a list with a specific pattern. Although the information is semi-structured, it has a common pattern to represent the information of a CFP in tables. A screenshot of the *DBWorld* Web Site is shown in Figure 3.

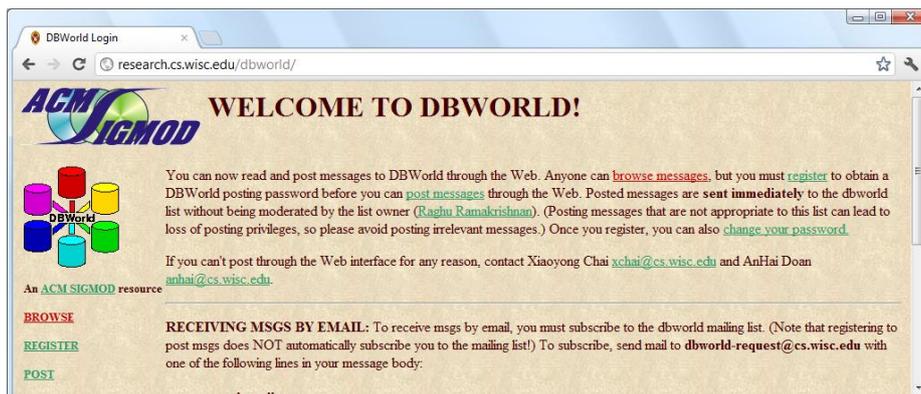


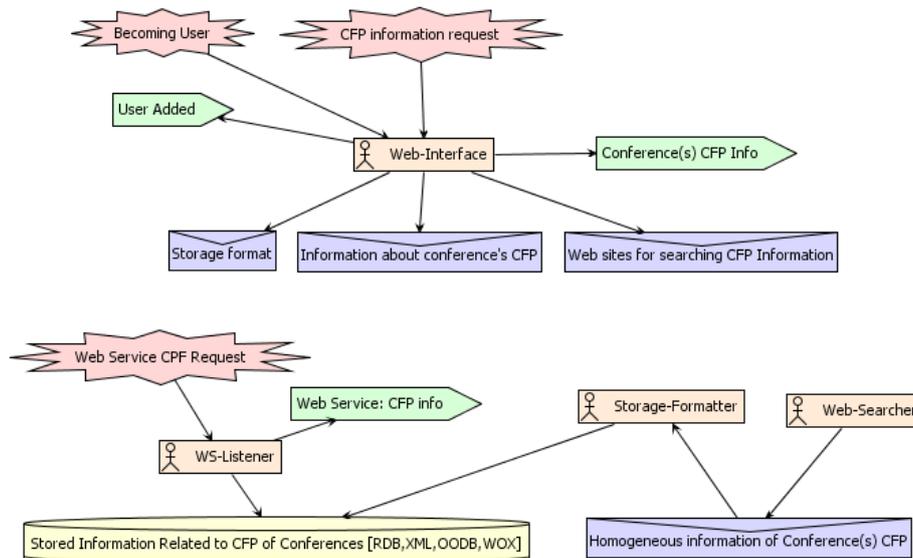
Fig. 3. DBWorld Web site.

### 4 Multi-Agent System Architecture

The architecture of our multi-agent system is composed of four intentional agents, designed using the BDI approach [3], [7]; and the Prometheus methodology, using the Prometheus Design Tool (PDT) [8], a graphical editor for specifying, designing and implementing intelligent agents. The development process following this methodology has three stages:

1. **System specification:** oriented to identify the human-agent interaction described by perceptions and actions, defining the goals, and some steps of functionality based in the defined roles.
2. **Architectural design:** where the agents of the system are defined, the MAS entire structure is determined, and the system dynamics is captured by means of interaction protocols.
3. **Detailed design:** where the plans, data, capabilities and actions of the agents are defined. PDT was used for designing this detailed design of agents.

Figure 4 illustrates the System Overview Diagram of PDT, with the architecture of the MAS, where it can be seen the agents and the three main processes mentioned above, which are associated with them.



**Fig. 4.** The MAS architecture, with the Web-Interface, Web-Searcher, Storage-Formatter and the WS-Listener agents. The CPF information is stored in different formats: relational, object oriented, XML and Web Objects in XML.

In the first place it is depicted the *Web -Interface* agent, which has the following activities: it provides the communication with the users of the system; it gives to the

other three agents the instructions for searching information on the Internet, it specifies the format for storing the heterogeneous information collected from the Internet; and it shows to the users the results of their requests for CFP information. The *Web-Searcher* agent is in charge of retrieving CFP from different Web sites, and downloading this information to the system. The *Storage-Formatter* agent takes the downloaded CFP and cleans them up, by converting them into XHTML files and finally takes the clean CFP, reads them, extracts their information, and stores them in the storage system. Additionally, the *WS-Listener* is the agent responsible for providing the responses of Web Services requests of external systems; it executes the queries provided in the Web Service and sends the result to the requesting system.

The group of BDI agents were implemented using *Jason* [9], which is an interpreter of the agent-oriented language *AgentSpeak(L)* [3], wrote in the Java programming language. *Jason* is a graphical environment for agent development, which allows the edition and execution of agent programs, and includes the opportunity to extend the abilities of each agent through Java programs (called internal actions) that implement a variety of tasks in order to achieve certain goals.

The following three subsections explain in detail the activities carried out by the two main agents of our system: *Web-Searcher* and *Storage-Formatter*.

#### 4.1 Retrieving a CFP

The *Web-Searcher* agent retrieves a set of CFP from the Web sites described in previous sections (*WikiCFP*, *ConferenceAlerts* and *DBWorld*). In order to carry out this task, the agent uses the `wget` command, which is launched automatically once a day, according to the agent goals. The following code fragment shows a sample of the *Web-Searcher's* execution of the `wget` command to retrieve only a specific CFP from the *WikiCFP* Web site.

```
wget -N --html-extension
http://www.wikicfp.com/cfp/servlet/event.showcfp?eventid=$a
```

It can be seen that the `wget` command has some parameters to specify how to download a specific CFP, which is contained in a Web page. The `-N` parameter is used to specify how deep the command goes and retrieves the contents of the Web page; the `--html` and the `-extension` parameters specify to store the Web page retrieved as an HTML file; and finally the rest of the string represents the URL of the Web page (CFP) that will be downloaded. The command uses the `$a` variable to represent the `id` of the CFP from the *WikiCFP* Web site (this `id` goes from 1 to 19,000 approximately, and this number is being increased daily as new CFP are added to the Web site). In order to illustrate a real example of how this agent retrieves a Web page with a CFP, we will use the CFP of the *12th Congress on Computing (CORE 2012)*, which is already registered in the Web site. The agent retrieves the Web page representing this CFP from the following URL:

```
http://www.wikicfp.com/cfp/servlet/event.showcfp?eventid=25031
```

Figure 5 shows a screenshot of the Web page that shows the *CORE 2012* CFP with *id=25031*. This is retrieved from the *WikiCFP* Web site.



**Fig. 5.** CFP for CORE 2012, retrieved from the *WikiCFP* Web site.

Once the *Web-Searcher* has finished retrieving the CFP from the Web site, its next goal is to generate an HTML file and store it in the MAS data storage system, as described previously. The file name for this specific CFP is the following: `event.showcfp?eventid=25031.html` (the *Web-Searcher* agent takes the last part of the URL to name the HTML file). The following code fragment, written in Jason [9], shows the plan followed by the *Web-Searcher* agent to accomplish its goal of downloading a Web page, before starting to process it.

```
// Web-Searcher Agent in project WebCFPAgent.mas2j
/* Initial beliefs and rules */
...
/* Initial goals */
...
/* Plans */
...
+!searchCFP(CFPID): true <-
    cfpActions.searchCFP(CFPID).
...
```

The code starts by defining the *Web-Searcher* beliefs and rules, its goals, and plans; then it can be seen a small fragment of the plan to satisfy the `searchCFP` goal, which takes the `CFPID` as parameter representing the Web page to be downloaded. The plan executes an internal action, implemented in the Java class `cfpActions.searchCFP`,

and it is the responsible to execute the `wget` command as previously described. It should be noticed that this action is executed for the three CFP Web sites currently considered for this system. All actions carried out by the agents are also implemented through Java classes. The next step in the process is to clean up the resulting file in order to get it ready to be read, and extract its conference information.

## 4.2 Cleaning Up a CFP

In this task, the *Storage-Formatter* agent takes every downloaded HTML file (CFP), and cleans it from any unmatched start and end tags, converting the stored HTML file into a well-formed XHTML file. Having this file is essential in the MAS, in order to use the *Storage-Formatter* agent's ability of XML parsing and extracting its conference information. The following code fragment shows part of the structure of the XHTML file generated by this agent. It can be observed information about the conference, such as its full name, URL, conference dates, place of the conference, abstract registration deadline, submission deadline, notification date, among other information. We have removed some markup from the generated XHTML file for display purposes, and to show only the important information that the agent needs to retrieve, which is highlighted in bold and red colour.

```
<span property="v:description">CORE 2012 : 12th Congress on
Computing</span>

<tr><td align="center">Link: <a href=
"http://www.core2012.cic.ipn.mx/" target="_newtab">
http://www.core2012.cic.ipn.mx/</a></td></tr>

<table cellpadding="3" cellspacing="1" align="center">
  <tr bgcolor="#e6e6e6">
    <th>When</th>
    <td align="center">Nov 28, 2012 - Nov 30, 2012</td>
  </tr>
  <tr bgcolor="#f6f6f6">
    <th>Where</th>
    <td align="center">Distrito Federal, México</td>
  </tr>
  <tr bgcolor="#f6f6f6">
    <th>Submission Deadline</th>
    <td align="center">
      <span property="v:startDate" content="2012-08-
19T00:00:00">Aug 19, 2012</span>
    </td>
  </tr>
  <tr bgcolor="#e6e6e6">
    <th>Notification Due</th>
```

```

        <td align="center">
            <span property="v:startDate" content="2012-10-
21T00:00:00">Oct 21, 2012</span>
        </td>
    </tr>
</table>

```

This XHTML file can now be taken by this agent to be read and extract its information for storing it in the MAS storage system. Next subsection explains this process, which is also carried out by the *Storage-Formatter* agent.

### 4.3 Parsing and Storing a CFP

The final step in the process is carried out by the *Storage-Formatter* agent, which stores the CFP information permanently in the system. The agent takes the XHTML file; reads it using its XML parsing ability; extracts all the CFP information according to the structure of the XHTML file; and stores it in the desired storage medium. It should be noticed that the agent carries out an analysis of the internal structure of the CFP, in order to extract its information. With this analysis the *Storage-Formatter* agent determines exactly where every piece of information is located, such as the name of the conference, submission deadline, notification date, conference dates, etc.

Concerning the storage medium for the CFP, the MAS has implemented five different storage mechanisms, which can be selected through the *Web-Interface* agent. These mechanisms are the following: *MySQL* as a relational database [10]; *DB4O* as an object-oriented database [11]; *eXist* as an XML database [12]; XML files using the *Web Objects in XML* serializer [13], [14]; and XML files using custom XML serialization. The MAS is under development, testing and evaluation of the performance of the five different storage mechanisms, in order to determine the most efficient implementation for its final version.

This final process aims to leave all the important information about CFP in only one homogeneous place, from which it can be accessed later through different interfaces. The basic structure we use to store a CFP depends on the specific implementation, but the information is always the same. The following code fragment illustrates all the fields used to store the CFP information internally, represented as Java attributes of a particular class, which is part of the internal actions for this agent.

```

private String abreviatura;
private String nombre;
private String fechaInicio;
private String fechaFin;
private String fechaDeadline;
private String ciudad;
private String pais;
private String url;
private ArrayList listaCategorias;

```

The outcome of this step is to have the information of every CFP stored in our system permanently. An example of the information that the agent extracts from the CFP for the *CORE 2012* is shown in Figure 6. It should be noticed that this process is executed by the agents for every CFP in the Web sites mentioned previously. Additionally, the following code fragment illustrates part of the XML document generated for the *CORE 2012* CFP with all its information.

```
<conferencia>
  <abreviatura>CORE 2012</abreviatura>
  <nombre>12th Congress on Computing</nombre>
  <fechaInicio>Nov 28, 2012</fechaInicio>
  <fechaFin>Nov 30, 2012</fechaFin>
  <fechaDeadline>Aug 19, 2012</fechaDeadline>
  <ciudad>Distrito Federal</ciudad>
  <pais>Mexico</pais>
  <url>http://www.core2012.cic.ipn.mx</url>
  <categoria>computer science</categoria>
</conferencia>
```

Nombre	Tipo	Valor
this	ReadWcfp	#406
almacen	XMLcfp	#407
builder	SAXBuilder	#408
CFP	UAMcfp	#409
catago	ArrayList	"size = 1"
[0]	String	"computer science"
FechaInicio	String	"Nov 28, 2012 "
FechaFinal	String	" Nov 30, 2012"
ciudad	String	"Distrito Federal"
pais	String	" México"
abreviatura	String	"CORE 2012 "
nombre	String	" 12th Congress on Computing"
url	String	"http://www.core2012.cic.ipn.mx/"
deadline	String	"Aug 19, 2012"

Fig. 6. Information extracted from the CORE 2012 CFP.

## 5 Interfacing with the Multi-Agent System

This section briefly describes the two types of interfaces of the MAS. The first interface is using the *Web-Interface* agent, which is the responsible for the interchange of information between a normal user and the MAS. The other interface with the MAS is through the *WS-Listener* agent, which executes a Web service that allows the interaction of the user with the system by retrieving specific information using CFP queries.

### 5.1 Web-Interface Agent

The *Web-Interface* agent exhibits as part of its abilities a Web interface, which is divided in categories and countries, but it is also possible to search CFP with specific criteria, such as their name, dates, deadline for submitting a paper, place where the conference will take place, or a combination of them. The *Web-Interface* agent can communicate the parameters established by the user to execute searching of conference CFP, so that the *Web-Searcher* and the *Web-Interface* agents can start working together. The *Web-Interface* agent receives from the Web interface a set of parameters representing the search options specified by the user, and builds a set of queries used by the *Web-Searcher* agent to search for the CFP that comply with the desired criteria; then the CFP selected are shown in an ordered list.

Additionally, the *Web-Interface* agent has the ability to allow searching and downloading individual CFP. The following Jason code fragment shows the plan followed by the *Web-Interface* agent to accomplish its goal of searching for an individual CFP and downloading the corresponding Web page. After downloading the CFP, it is passed to the *Web-Searcher* agent to be processed as described in section 4.

```
// Web-Interface Agent in project WebCFPAgent.mas2j
/* Initial beliefs and rules */
...
/* Initial goals */
...
!startSearch.
/* Plans */
...
/* Searching Plan*/
+!startSearch : cfpid(CFPID) <-
    .send(web_Searcher, achieve, searchCFP(CFPID)).
...
```

The code starts by defining the *Web-Interface* beliefs and rules, its goals, and plans as in the rest of the agents. It can be seen a short fragment of the plan, where the agent tries to satisfy a requirement from the user by starting the retrieval of a CFP. The user's action fires the *startSearch* goal of this agent; and sends a message to the *Web-Searcher* agent in order to begin the task of retrieving the CFP information as the user requested; just as described in section 4, where the *Web-Searcher* agent takes the control, downloads the CFP, and process it.

### 5.2 WS-Listener Agent

The *WS-Listener* is the agent that provides the communication of the MAS with external systems through a Web service interface. This agent is the responsible for listening requests from systems, executing the Web services indicated by the parameters of each request, and sending to the requesters the information about CFP that is stored

in the MAS. It should be noticed that the *WS-Listener* agent has a set of Web services, depending on the type of request. At the moment this agent is under development, but it is designed to return the same information as the *Web-Interface* agent; the only difference is that the *WS-Listener* returns the result in XML standard messages, and the *Web-Interface* agent gives the result through a Web interface.

## 6 Conclusions and Future Work

This paper presented an intentional agent-based system that automates and facilitates the retrieval of CFP information from heterogeneous sources, specifically from three different specialized Web sites dedicated to concentrate information about scientific conferences and CFP. This task is performed by a group of four agents designed under the BDI approach: *Web-Interface* agent, *Web-Searcher* agent, *Storage-Formatter* agent and *WS-Listener* agent. These agents retrieve the relevant information of CFP for conferences, store it in the storage system, and show it to users by means of a flexible Web site, which is part of the *Web-Interface* agent duties. Some interoperability with other external software systems is incorporated in our MAS through the *WS-Listener* agent, which listens for Web service requests, and returns the results in the form of standard XML messages.

The MAS presented has five different implementations for storing the heterogeneous information retrieved: MySQL as a relational database, DB4o as an object-oriented database, eXist as an XML database, Web Objects in XML as an XML serializer, and XML files with custom serialization. A sample Jason code of some of the agents is shown, in order to demonstrate the types of plans that they follow. It is being carried out an investigation and evaluation of the suitability and performance of every storage approach, in order to determine the best way for storing the information retrieved by the three specialized Web Sites. As part of the future work, we also consider the inclusion of two more intentional agents: the *CFP-Recommender*, which would be in charge of the generation of recommendations about conference CFP, based on the user profile and similar user profiles; and the *CFP-Ranking* agent which would be responsible of ranking the CFP according to specialized ranking Web sites for conference CFP.

## References

1. Luna-Ramírez, W.A., Jaimez-González, C.R., Ramírez-Colín, L.A.: An Agent-Based System Architecture for Retrieving Call for Papers. In: IADIS International Conference on Applied Computing 2012, pp. 434-436, ISBN: 978-989-8533-14-2, Madrid, Spain (2012)
2. Wooldridge, M., Jennings, N.R.: Intelligent Agents: Theory and Practice. Knowledge Engineering Review 10(2), (1995)
3. Rao, A.S.: AgentSpeak(L): BDI Agents Speak out in a Logical Computable Language. In: Van de Velde, W. and Perram, J. (eds), Proceedings of the Seventh Workshop on Modelling Autonomous Agents in a Multi-Agent World (MAAMAW'96), 22-25 January. Eindhoven, Netherlands. LNAI 1038:42-55, Springer-Verlag, Heidelberg, Berlin, (1996)

4. A Wiki for Calls For Papers. Available at: <http://www.wikicfp.com/>. Last accessed in October 2012
5. Conference Alerts: Academic Conferences Worldwide. Available at: <http://www.conferencealerts.com/>. Last accessed in October 2012
6. DBWorld. Available at: <http://research.cs.wisc.edu/dbworld/>. Last accessed in October 2012
7. Rao, A.S., George, M.P.: BDI-Agents: from Theory to Practice. In: The First International Conference on Multiagent Systems, San Francisco, (1995)
8. Padgham, L., Winikoff, M.: Developing Intelligent Agent Systems: A Practical Guide. Available at: <http://www.cs.rmit.edu.au/agents/prometheus/>, John Wiley and Sons, (2004). Last accessed in October 2012
9. Bordini, R.H., Hubfiner, F.J.: Jason. A Java-based AgentSpeak Interpreter Used with Saci for Multi-agent Distribution over the Net, (2006). Available at: <http://jason.sourceforge.net/wp/>. Last accessed in October 2012
10. MySQL DB: The world's most popular open source database. Available at: <http://www.mysql.com/>. Last accessed in October 2012
11. DB4Objects DB: by Versant. Available at: <http://www.db4o.com/>. Last accessed in October 2012
12. eXist: An open source database management system built using XML technology. Available at: <http://exist-db.org/exist/index.xml>. Last accessed in October 2012
13. Web Objects in XML: an XML serializer for Java and C# objects. Available at: <http://woxserializer.sourceforge.net/>. Last accessed in October 2012
14. Jaimez-González, C.R., Lucas, S.M.: Easy XML Serialization of C# and Java Objects. In: Balisage: The Markup Conference 2011, vol. 7, Balisage Series on Markup Technologies, ISSN: 1947-2609, doi:10.4242/BalisageVol7.Jaimez01, Montréal, Canada (2011)



# **Micro technologies & Embedded Systems**

---



# A Hardware/Software Co-Execution Model Using Hardware Libraries for a SoPC Running Linux

Alejandro Gómez-Conde, José de Jesús Mata-Villanueva, Marco A.  
Ramírez-Salinas, and Luis A. Villa-Vargas

Center for Research in Computing,  
National Polytechnic Institute, D.F., Mexico.  
alegzc@yahoo.com.mx, gsusmata@gmail.com, mars@cic.ipn.mx,  
lvilla@cic.ipn.mx

**Abstract.** A SoPC (System on a Programmable Chip) is the combination of computer architecture, IP designs, and an embedded operating system implemented on a FPGA. A SoPC developer needs to possess skills in advanced hardware design and efficient software programming to make high performance solutions on this platforms type. These requirements have resulted in a complicated and error-prone development process, therefore the dissemination of this promising technology has been confined to a small group of developers. To address this problem, our proposal exemplifies a Hardware/Software co-execution model where software tasks are assisted by hardware libraries in a SoPC.

**Keywords:** System on a Programmable Chip, Linux FPGA system, Mixed execution mode, Function interception.

## 1 Introduction

A SoPC system requires a FPGA with a processor core that can be either *hard* or *soft*. This designation refers to the flexibility or configurability of the core. *Hard cores* have a custom VLSI layout that is added to the FPGA in the manufacturing process. They are less configurable and tend to have higher performance characteristics than *soft cores*. Although *soft cores* are highly configurable they use logic elements as well as routing resources within a FPGA that increases the complexity of hardware design and reconfigurable stages. The overall implementation of a SoPC system needs to meet one or more of these constraints:

- Performance
- Execution time
- Small footprint area
- Communication and connectivity
- Implementation cost
- Upgradability
- Low power

The outline of this paper is as follows: Section 2 describes our proposal. Section 3 presents the experiments and results for the implementation process of a SoPC that uses a Hardware/Software co-execution model that includes the creation of a hardware platform, the integration of the Operating System and the implementation of the bitswapper hardware function. Section 4 discusses related work. Finally a conclusion is given.

## 2 Hardware/Software Co-Execution Model

In a traditional execution model, a user program often includes references to software libraries. Once the user-program starts execution it requests software functions; those functions are grouped together in software libraries that may be statically or dynamically attached to the user program.

A Purely software-based execution model does not always take advantage of the hardware within the platform nor is it able to delegate software tasks to hardware modules. As a direct consequence the execution model does not always meet the performance or execution time constraint for a specific task in a SoPC system. Our approach was to develop a Hardware/Software co-execution model for critical applications in a SoPC. Our proposal uses a library-based access model and extends its functionality to support the execution of hardware functions. Figure 1 shows the proposed execution model. The idea behind interception is to assist a software task by hardware modular functions in a seamlessly manner. If the hardware function is not available when the request is issued then the process that requests such functionality should either wait for the hardware function to be programmed into the FPGA or follow its execution path until the requested function is provided. Our proposal waits for the hardware function to be configured and then continues its execution assisted by a hardware function. This capability adds upgradability to a SoPC system because there is no longer the need to modify the application source code to embedded hard-coded routines that interact with hardware functions. If a new SoPC system is needed, our project will provide a Hardware/Software partitioning model to get the best performance at the lowest implementation cost using Hardware/Software partitioning, co-design and co-synthesis techniques but even if there is not access to the source code, an application can be assisted to reduce its execution time.

Our proposal uses the Dynamic Linking Loader service but instead of *calling* the original software library function, it *calls* a hardware modular function that might be programmed at run-time in the reconfigurable area of the SoPC. The communication protocol between the user program and the hardware function is implemented by a dynamic library and a character device driver. The main reason for using a character device driver is that it can be dynamically loaded at the system startup and all the functionality is in the kernel module; the dynamic library could be set up to capture function calls for specific programs while the rest of the systems work as usual. This behavior could be activated or deactivated at any time while the system is still running as long as the hardware function is not in use. Once the hardware function has been configured and the

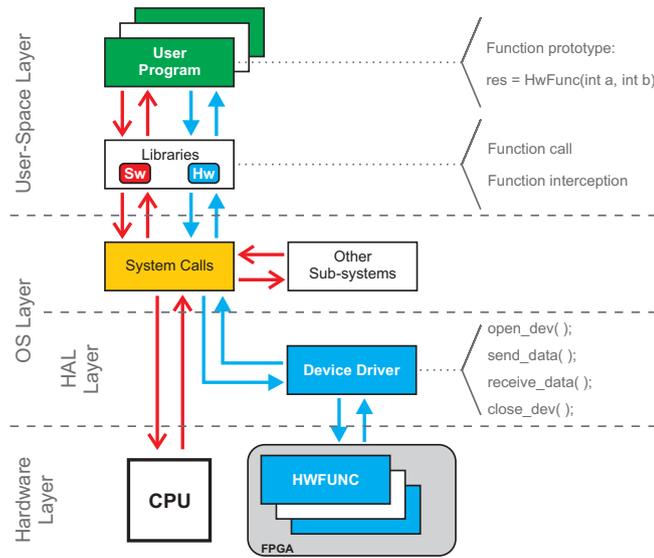


Fig. 1. Proposed execution model.

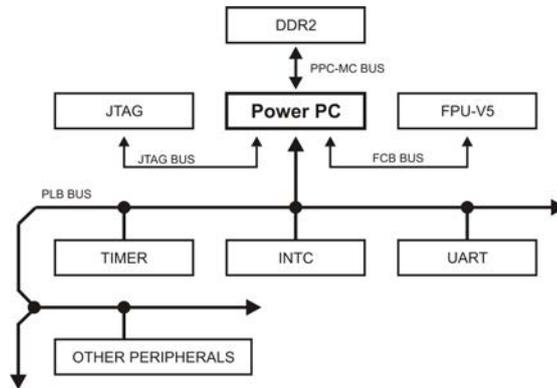
software access layer has been properly set up, the program can take advantage of the functionality provided by the hardware module. A performance penalty is expected to exist as a direct consequence of the reconfiguration process at runtime when a hardware function is called for the first time. When a large amount of data is processed by the function, the impact of the reconfiguration process becomes smaller.

### 3 Experiments and results

#### 3.1 Hardware Platform

A Linux-capable SoPC hardware platform was built up using Xilinx’s EDK tools. Figure 2 shows a brief description of the architecture **S1** implemented in the FPGA of the ML507 development board. It integrates the PowerPC 440 Processor, a DDR2 memory controller, interrupt controller, serial port, JTAG debug port, float point unit, System ACE, a timer, and some other peripherals. Although Xilinx’s tools helped in the integration and configuration processes, it was necessary to incorporate a description of the elements within the hardware platform besides of the `xparameters.h` description file.

D. Gibson and B. Herrenschmidt in [1] propose a *device tree model* to generate a hierarchical description of a hardware platform. The most important properties of the *flattened tree* obtained from the *device tree model* are: relocatable properties that allow the bootloader or the kernel to move around the blob



**Fig. 2.** Architecture of the hardware platform S1.

tree as a whole, without need to parse or adjust its internals. It is also possible to *insert* or *delete* a node or a whole *subtree*; this property will help us integrate hardware functions without having to recompile the kernel.

Xilinx in collaboration with the Open Source Community developed a BSP package that was integrated into the SDK to generate a *device tree* description for the S1 hardware platform.[2]

The hardware platform S1 was tested using `Push_Buttons_5Bit_GPIO` and `LEDs_8Bit_GPIO` hardware modules in stand alone mode using *Xilinx's SDK*.

### 3.2 Integration of the Operating System

An embedded operating system was generated for the ML507 Board using Poky framework tools[3] and `meta-xilinx`[4] specification layer. This framework provides a structured scheme to integrate architecture specifications in the kernel compilation processes and provides a cross-compile platform to create, test and debug user-space as well as kernel-space applications. Figure 3 shows the workflow used in Poky to generate the Operating System.

U-boot was chosen as the bootloader because Xilinx's modified version of U-Boot has been successfully adapted and tested to work with Virtex5 FPGAs.[5] The `meta-xilinx` configuration and specification layer was modified to successfully generate the kernel image and the bootloader.

The most important files generated using Poky framework are: The file `u-boot-ml507.ace` contains the FPGA configuration *bitstream* and the executable file for U-Boot in a System ACE file format. It will be loaded at startup to configure the FPGA, and start the system's boot process. The file `uImage-vertex5.bin` is a binary file of the kernel image. It is adapted to be *launched* by U-Boot. The file `uImage-vertex5.dtb` holds a *flattened tree blob* that represents the internal hardware architecture implemented within the FPGA. The file `poky-image-minimal-vertex5.tar.gz` is a compressed image of the SoPC file system (including

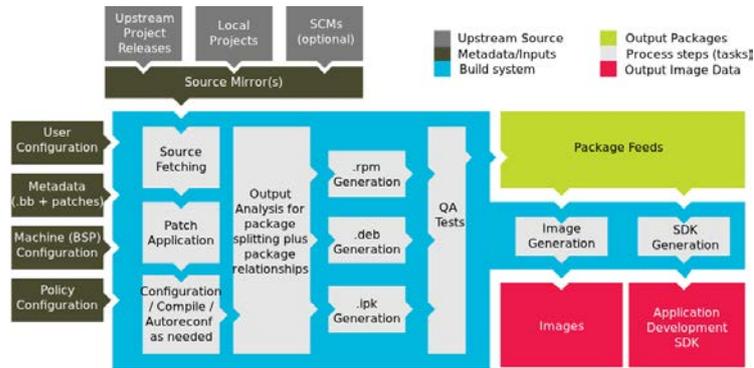


Fig. 3. Poky Architecture Workflow.

/boot, /dev, /etc, /home, /lib and others). The file *modules-2.6.37+-r12-virtex5.tgz* is another compressed file that contains kernel modules that are dynamically loaded at *run time*.

In order to run the Operating System on the ML507 board the user must set the configuration switch SW3 so that the FPGA is configured by the SysACE as described in [6]. The Compact Flash must be partitioned in three sections; the first one must be formatted as *FAT16*, the second partition should be *ext3* and the third partition should be used as the swap partition. The first partition must have the *u-boot-ml507.ace*, *uImage-virtex5.bin* and *uImage-virtex5.dtb* files in it. The second partition must be filled with the content of *poky-image--minimal-virtex5.tar.gz* and the *modules-2.6.37+-r12-virtex5.tgz*.

The serial communication program *kermit* was used to explore the boot process and the kernel startup. The last part of the output text shown in the console for the boot process is presented below. The last four lines of text are associated to the *device tree blob* and the *kernel* load respectively. The device tree specifies architecture details that the kernel will use to properly setup the system. The kernel image was generated using the *uimage* format that contains the compressed *vmlinux* plus a few extra bytes of *metadata* that describe the kernel load address and the image name.

```
CPU:   Xilinx PowerPC 440 UNKNOWN (PVR=7ff21912) at 400 MHz
DRAM:  256 MB
FLASH: 32 MB
In:    serial
Out:   serial
Err:   serial

reading uimage.dtb
25059 bytes read

reading uimage.bin
```

2096237 bytes read

At the end we obtained a fully functional Linux-based Operating System that runs on the ML507 board; the kernel was generated using the 2.6.37 Linux kernel source code.

### 3.3 The Hardware Functional Unit

The hardware functional unit *bitswapper* changes the endianness of a data word from big endian to little endian and vice versa. This function uses Xilinx's IPIF module as an interface between the functional unit and the system. This functional unit has been chosen for its simplicity and ease of implementation; it is not yet a real functional unit, but it lets us debug integration procedures and provided valuable information to select the best communication model to be implemented between a hardware function and the software layer. Our first approach was to attach it to the PLB bus. It uses two registers and a basic communication protocol to send and receive data to and from the system through the IPIF interface.

Xilinx's EDK tools were used to attach the *bitswapper* hardware module to the existing platform; the core functionality of the proposed function is presented below.

```
process(datain)
begin
  for i in datain'low to datain'high loop
    my_swap_sig(i) <= datain((C_SLV_DWIDTH-1)-i);
  end loop;
end process;
dataout <= my_swap_sig;
```

Figure 4 shows the synthesized modular hardware function and fig. 5 shows the simulation results for the core functionality of the proposed hardware modular function.

Our implementation used a character device driver to provide communication services between the hardware function and the user program. Therefore a special file entry in the `/dev` folder was created; it was named `hwfn1`. The kernel module implements `device_init`, `device_exit`, `send_data`, `receive_data`, `device_open`, `device_read`, `device_write`, `device_release`, and other functions that are commonly related to character device drivers. Our proposal generates a dynamic library as described in [7]; to prove the function interception process the software library function `bitswap` was created; in our embedded development environment this software function is provided by the dynamic library `libhw.so`. The Dynamic Linking Loader uses a special feature that is controlled by the system variable `LD_PRELOAD`; in order to intercept the `bitswap` call another dynamic library was created using the same function prototype but implementing the communication protocol through the file descriptor associated to `hwfn1`. The kernel uses the services provided by the character device driver to send data to and receive data from the hardware function.

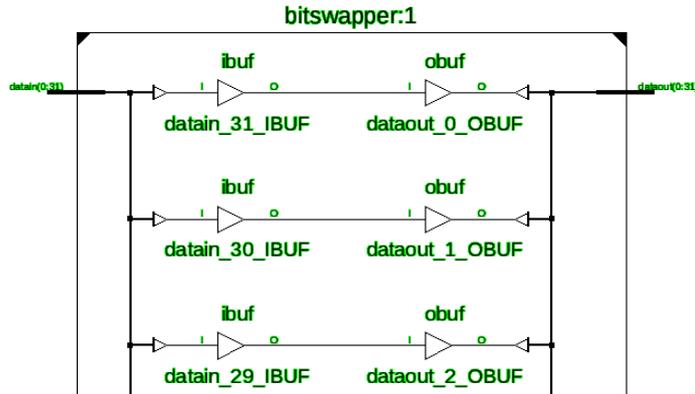


Fig. 4. RTL Schematic representation for Hardware Functional Unit.

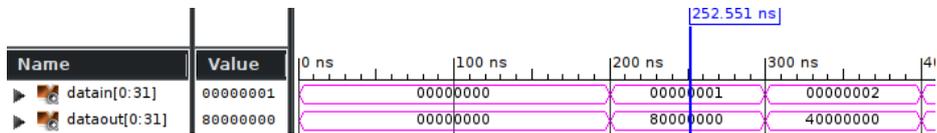


Fig. 5. Simulation results for the core functionality of the *bitswapper* function.

Table 1 presents a summary of the FPGA resources used in the *bitswapper* hardware modular function and for the SoPC system implementation.

Table 1. FPGA resource utilization.

Resource type	Bitswapper	SoPC System	Available
Slices	88	5668	11200
Slice Registers	109	8939	44800
Slice LUTs	86	9926	44800
IOBs	0	210	640
BlockRAMs	0	40	148
Memory used (kib)	0	1404	5328
ICAPs	0	1	2
DCM	0	1	6

#### 4 Related work

Hayden Kwok-Hay So *et al.*[8] presented a very interesting Hardware/Software co-design methodology based on BORPH (*Berkeley Operating system for Re-*

*Programmable Hardware*). It included an Operating System designed for reconfigurable computers that implements hardware modules designed by the user; it also provides a native kernel support to implement processes in FPGAs using a homogeneous Unix interface for hardware and software processes. An interesting issue of this project was that a hardware process implemented into an FPGA inherits the same level of service from the kernel, as a software process does. BORPH includes file system support, interrupt support, and scheduling capabilities for hardware processes; it also provides common Unix APIs that interact with hardware processes in a seamlessly manner. The unified file interface allows hardware and software processes to communicate via standard Unix file pipes. BORPH was implemented on a development board with five FPGAs; one was used as a master unit while the others each held an implementation of a hardware process. They also developed a Hardware/Software interface that allows application development, migration of software procedures to hardware modules and a comparison among Hardware/Software implementations of the same application. Although they provided a revolutionary implementation of hardware processes, they only implement user hardware design into FPGAs and did not consider partial reconfiguration.

Qingxu Deng, *et al.*[9] propose a Hardware/Software unified architectural model for an FPGA implementing partially runtime reconfigurable services. Their proposal implements scheduling services and on-line placement of hardware modules within the same FPGA. Hardware and software tasks were implemented as processes but each hardware task uses a software communication layer to interact with the Operating System. They also implement a modified version of the system call *exec* to recognize HELF files (*Hardware Executable and Linkable Format*) and extract the information required to program a hardware IP-Core and its dedicated communication software layer from the HELF file; this new type of executable file format is an extension to ELF. A key feature of this proposal is that it implements a routing and placement algorithm to reduce the required time to schedule and execute hardware tasks in the reconfigurable area of the FPGA. Even though this proposal presented a new approach to create an abstract model of a hardware process that can be launched from user-space it is only fully functional for that specific platform since other platforms may not recognize the HELF executable file format; besides that, the functional elements of this proposal, at kernel level, cannot be easily migrated into other architectures due to the fact that they are deeply merged into a specific version of the Linux kernel.

Vaibhawa Mishra *et al.*[10] propose a dynamically reconfigurable SoPC that uses an Operating System based on the 2.6.34 Linux kernel. They implement a minimal hardware platform in the ML507 development board using the FPGA, the PowerPC 440 *hard core* processor and other peripherals. This proposal uses floating-point hardware modular functions to prove the reconfigurability of the system. They propose a functional model that is similar to a peripheral device

model with the only difference of implementing one *virtual device* that can be re-programmed at runtime to implement any of the four floating-point functional units developed. Although this proposal implements reconfigurability using the *hwicap* IP-Core on a Virtex5 FPGA, it did not provide a major advantage over other proposals made in the past.

## 5 Conclusion and future work

SoPC systems are about to be a major improvement in the mobile devices field. Therefore our proposal is building up the required infrastructure to be able to generate, test, validate, and improve our knowledge of this type of systems.

Task execution will no longer use a purely software execution model; nevertheless, there cannot be a solid conclusion about which methodology shall prevail.

This work aims to provide an overview of the challenges that are faced while designing and implementing SoPC systems.

The *products* generated in this work are: a reconfigurable hardware platform for SoPCs and Poky-Linux framework tools and a configuration layer to build and integrate an Operating System to the ML507 board. The on going work implements another communication scheme between the hardware function and the software layer to compare and select a lower latency communication scheme for the SoPC system. Our team is developing a set of hardware modules to complement the actual hardware library.

Hardware functional unit *bitswapper* was developed and used to test the SoPC system for both correctness and performance. Although it was designed for test purposes, this modular hardware unit lets us explore the overall development process over the little-known path of mixed execution model.

Although the operating system is still in the development phase and, even when the hardware platform has not been tuned up to the maximum allowable performance, early results are promising. The Hardware/Software execution model reduces the CPU processing load when using hardware functional units.

This work is part of the MASA Project, the current development stage focuses on implementation specifically on upgradability.[11] Another on-going work of this project has developed a framework for custom routing and partitioning, that will enable effectively implementing relocatable hardware functions of variable size (in terms of reconfigurable partitions) and number of input/output ports.

Future work will design, implement and compare our model with a system-call based access model and a memory mapped access model that attaches a hardware function to the user-process virtual memory.

## References

1. D. Gibson and B. Herrenschmidt. Device trees everywhere. OzLabs, IBM Linux Technology Center. February 13, 2006.
2. Xilinx Inc. Device Tree Generation. *Online resources*. [http://elinux.org/Device\\_Trees](http://elinux.org/Device_Trees) and <http://wiki.xilinx.com/device-tree-generator> Visited on March 2011.
3. Poky Platform Builder. *Online resource*. [www.pokylinux.org/](http://www.pokylinux.org/) Visited on January 2011.
4. Xilinx Inc. BSP Specification Layer in Poky to support ML50x development platforms. *Online resource*. <http://git.yoctoproject.org/cgit/cgit.cgi/meta-xilinx/> Visited on March 2011.
5. Xilinx Inc. Xilinx's modified versión of U-Boot. *Online resource*. <http://git.xilinx.com/?p=u-boot-xlnx.git;a=summary> Visited on April 2011.
6. Xilinx Inc. ML507 Evaluation Platform. *Online resource*. [http://www.xilinx.com/support/documentation/boards\\_and\\_kits/ug347.pdf](http://www.xilinx.com/support/documentation/boards_and_kits/ug347.pdf) Visited on Nov 2010.
7. U. Drepper. How To Write Shared Libraries. *Online resource*. <http://people.redhat.com/drepper/dsohowto.pdf> , 2011.
8. H. Kwok-Hay, A. Tkachenko and R. Brodersen. A Unified Hardware/Software Runtime Environment for FPGA-Based Reconfigurable Computers using BORPH. Department of Electrical Engineering and Computer Science University of California, Berkeley. CODES+ISSS'06, October, 2006.
9. Q. Deng, Y. Zhang, N. Guan and Z. Gu. A Unified HW/SW Operating System for Partially Runtime Reconfigurable FPGA based Computer Systems. 2008.
10. V. Mishra, K. Solomon Raju and P. Tanwar. Implantation of Dynamically Reconfigurable Systems on Chip with OS Support. International Journal of Computer Applications. 2012.
11. L. Villa , M. Ramírez, O. Espinosa, and C. Peredo. Modular Architecture for Synthesized Applications. Center for Research in Computing, IPN. Mexico, DF. <http://www.microse.cic.ipn.mx/masa-es> , 2010.

# Real Time Stereo Vision with a modified Census transform and fast tracking in FPGA

Juan Manuel Xicoténcatl-Pérez, Arturo Lezama-León, José Miguel Liceaga-Ortiz-De-La-Peña, Rubén O. Hernández-Terrazas

Polytechnique University of Pachuca  
Km. 20 Carretera Pachuca Sahagún Ex Hacienda de Sta. Bárbara , Zempoala, Hidalgo, México  
jmxico@upp.edu.mx

**Abstract.** In this work, architecture for real stereo processing is implemented using a modified census transform based in a non central pixel census technique. The non central census technique allows a compact architecture. Proposed architecture is segmented in image rectification module to avoid lens distortion and to align epipolar lines, stereo processing module with modified census transform and finally, a post processing module with a propagation algorithm to correct false disparity values. Proposed architecture uses low hardware resources and memory requirement in a way that Spartan Low Cost FPGA can be used for implementation. Additionally, a correlation tracking module is incorporated to one camera with neglected hardware cost in comparison with other architectures for security applications.

**Keywords:** real time, stereo vision, fast tracking

## 1 Introduction

Stereo vision, compared with other range sensors such as laser scanner or time-of-flight, is a technology that can deliver the sufficient description of the surrounding environment. It is purely passive technology and thus offers low cost and has potential uses in many visual domains such as autonomous navigation in which accurate 3D information about the road is crucial, object 3D reconstruction, object segmentation, and surveillance systems [10]. However, calculation of three dimensional depth maps on signal processors that meets these requirements is very time consuming. In this way, real-time dense stereo is difficult to be achieved with general purpose processors even CUDA. For real-time requirements of most applications, the specific algorithms were often implemented using dedicated hardware and it is possible because of many stereo vision algorithms do not enforce a purely sequential implementation and are therefore apply to parallelized solutions. Additionally, stereo applications are used for security applications or video games in parallel with a tracking algorithm or system. For this reason, a tracking architecture is designed using the correlation structure from the stereo module.

In the last few years, the GPUs have become more and more popular. Using GPUs for stereo acceleration can directly be a solution for PC-oriented applications. But, the high power consumption limits their applications. FPGAs have already shown their high performance capacity for image processing tasks in parallel especially for embedded systems. In this paper, it is simulated and synthesized a stereo vision core algorithm implemented in VHDL for the Spartan XC3S1000 from XILINX, an FPGA that is suitable for this kind of application. The algorithm is based on a Census transform modified algorithm with produces a small hardware implementation. It is small enough to enable the pre- and post-processing of the images on the same FPGA, but with a maximum disparity of 50 pixels all in real time for a compact system.

## 2 Background

The task of a stereo vision algorithm is to analyze the images taken by a pair of cameras and to extract the displacement of the objects in both images. This displacement is counted in pixels and called disparity. All these disparities form the disparity map, which is the output of a stereo vision algorithm and enables the calculation of distances to objects using triangulation.

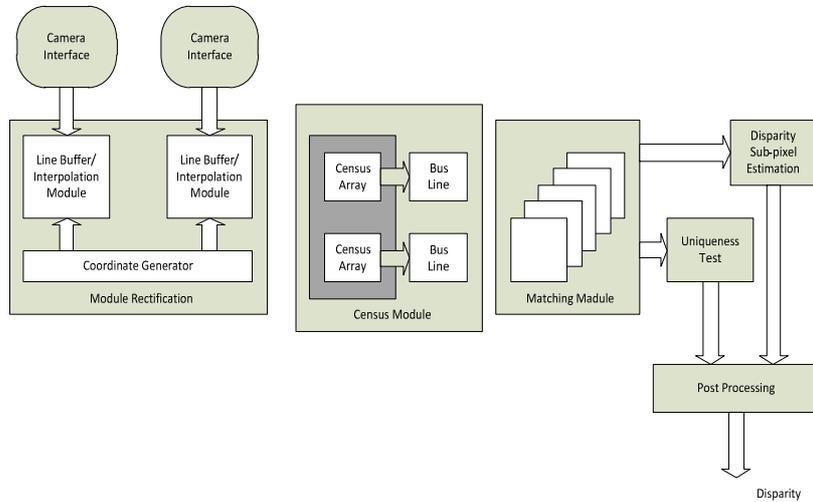
Detecting conjugate objects in stereo images to obtain dense disparity maps is a challenging research problem known as the correspondence problem, i.e. to find for each point in the left image, the corresponding point in the right one. To determine a conjugate pair, it is necessary to measure the similarity of the points. The point to be matched should be distinctly different from its surrounding pixels. In order to minimize the number of false correspondences and try to diminish time processing in the image pair, several constraints have been imposed. The uniqueness constraint requires that a given pixel from one image cannot correspond to more than one pixel on the other image. In the presence of occluded regions within the scene, it may be impossible at all to find a corresponding point. The ordering constraint [6] requires that if a pixel is located to the left of another pixel in image, i.e. left image, the corresponding pixels in right image must be ordered in the same manner, and vice versa, i.e. ordering of pixels is preserved across the images. The ordering constraint may be violated if an object in the scene is located much closer to the camera than the background, and one pixel corresponds to a point on the object while the other pixel corresponds to a point in the background. Finally, the continuity constraint [4], which is valid only for scenarios in which smooth surfaces are reconstructed, requires that the disparity map should vary smoothly almost everywhere in the image. This constraint may be violated at depth discontinuities in the scene.

According to a recent taxonomy [17], stereo algorithms that generate dense depth measurements can be divided into two classes, global and local algorithms. Global algorithms, e.g. [14], rely on iterative schemes that carry out disparity assignments on the basis of the minimization of a global cost function. These algorithms yield accurate and dense disparity measurements but exhibit a very high computational cost that renders them unsuited to real-time applications. Local algorithms [6, 8, 13] can be based in different concepts to establish a correspondence between images, so it is

possible to find area-based and feature based. Area-based stereo algorithms are approaches that propose a dense solution for calculating high-density disparity maps. Additionally, these approaches have a regular algorithmic structure which is suitable for convenient hardware architecture. A simple method would be to calculate the absolute difference between two pixels; this method is extremely cheap but not robust. The second possible method seeks to improve upon the previous one by considering a small window around the pixels in the left and right image and then using the sum of the differences. Small windows give support to central pixel to avoid false matching during the stereo process. In this paper, it is used a modified Census transform to achieve a stereo correlation. Modified census transform uses a non central pixel reference, i.e., it uses the most right pixel in the central row from a given window. This change just affects the generated census vector but results and advantages are same like other census implementations. Census transform is a non-parametric measure used during the matching process for measuring similarities and obtaining the correspondence between the points into the left and right images. Advantages with census transform are: a windows support, tolerance to illumination changes and possibility to hardware parallel implementation. On the other hand, feature-based algorithms rely on certain points of interest. These points are selected according to appropriate feature detectors. The major limitation of all feature-based techniques is that they cannot generate dense disparity maps, and hence they often need to be used in conjunction with other techniques. Because of the sparse and irregularly distributed nature of the features, the matching results should be augmented by an interpolation step if a dense disparity map of the scene is desired. Additionally, an extra stage for extensive feature detection in the two images is needed, which will increase the computational cost. Thus feature-based methods are not suitable for real-time applications.

### **3 Hardware implementation**

Stereo hardware implementation can be segmented as next figure indicates



**Fig. 1.** Stereo diagram block. Camera module is an interface with stereo camera board. Module rectification uses a predefined algorithm for image coordinate transformation. After that, census module obtain a census vector for census matching and finally, postprocessing and subpixel calculus is achieved.

Hardware modules in figure 1 show stages for rectification, modified census transform and census matching. Additional hardware is used to check the uniqueness restriction and sub pixel estimation.

### 3.1 Rectification process

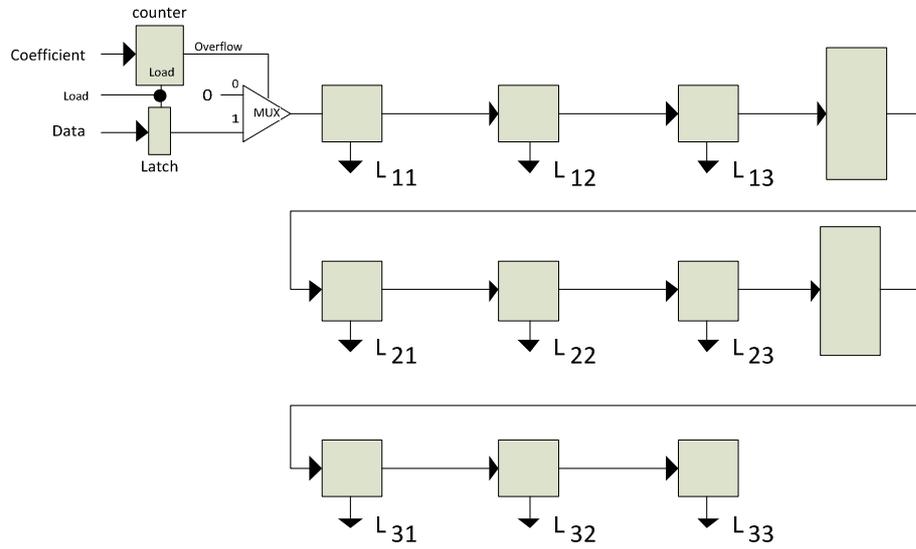
Rectification involves lens distortion correction, alignment rows in stereo images, back or reverse projection, and bilinear interpolation. In first place, due to manufacturing errors there will always appear distortions caused by the lens and a misalignment between lens and camera chip. These distortions happen before the light ray hits the image chip and consequently affect coordinates. In order to obtain a pair of rectified images from the original images after processing lens distortion and to align scanlines a homography is applied. In the equation (1)  $(x, y)$  and  $(x', y')$  are coordinates of a pixel in the original images and the rectified images, respectively.

$$\begin{bmatrix} x'_{lr} \\ y'_{lr} \\ z'_{lr} \end{bmatrix} = H_{LR}^{-1} \begin{bmatrix} x''_{lr} \\ y''_{lr} \\ 1 \end{bmatrix} \quad \begin{bmatrix} x_{lr} \\ y_{lr} \end{bmatrix} = \begin{bmatrix} x'_{lr} \\ z'_{lr} \\ y'_{lr} \\ z'_{lr} \end{bmatrix} \quad (1)$$

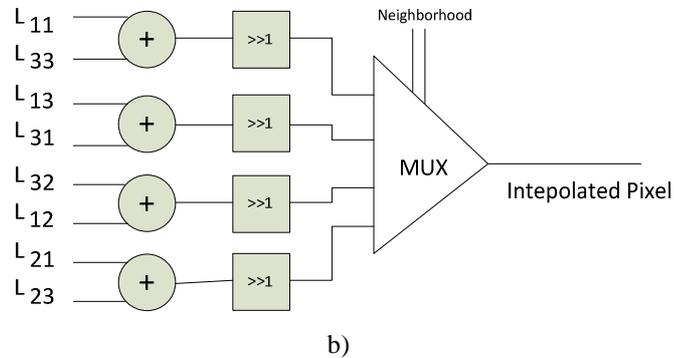
To avoid problems such as reference or coordinate duplication, reverse mapping is used with interpolation. Once the image pair is rectified, 1-D searching with the corresponding line is sufficient to evaluate the disparity.

A common hardware implementation of the rectification process is through Look-Up Tables. Only at the start, an offline calibration using MatLab is done and from calibration stage two LUTs -one for each camera – results [1]. Generated data includes: new pixel positions and best pixels for interpolation (Figure 2). Although, using LUTs directly in FPGA is possible, it is more suitable to have the LUT in external memory to extent capacity for future architectural expansions.

In hardware, rectification is applied as in [1][2] with some important differences: in proposed architecture new pixels positions are previously computed and a defined neighborhood (from calibration stage) is used to interpolate absent pixels. In this way, it is no necessary to process coordinates using a homography in the architecture. Finally, rectification module is shown in figure 2.



a)



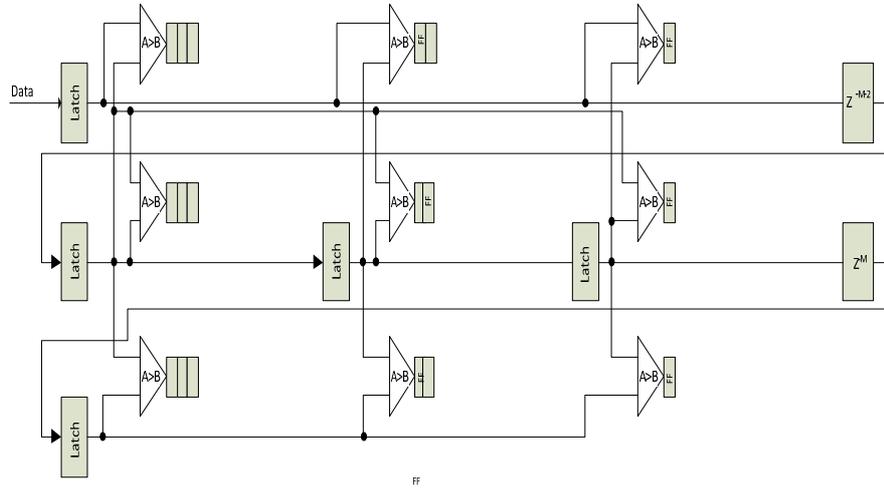
**Fig. 2.** Rectification module, a)Hw for the interpolation table, it detects absent pixels with coordinates from LUTs , b)Bilinear Interpolation module uses neighborhood for establish pixels values in real time.

In figure 2a, new pixel coordinates generates black pixels in the rectified image, so these black pixels are filled with interpolated pixels from specified neighborhood obtained in the calibration process. Architecture is based in row buffers to avoid store images. In figure 2b, a module for bilinear interpolation is shown. In both cases  $3 \times 3$  windows are used for explication purpose, real architecture uses  $11 \times 11$  windows.

### 3.2 Stereo Processing

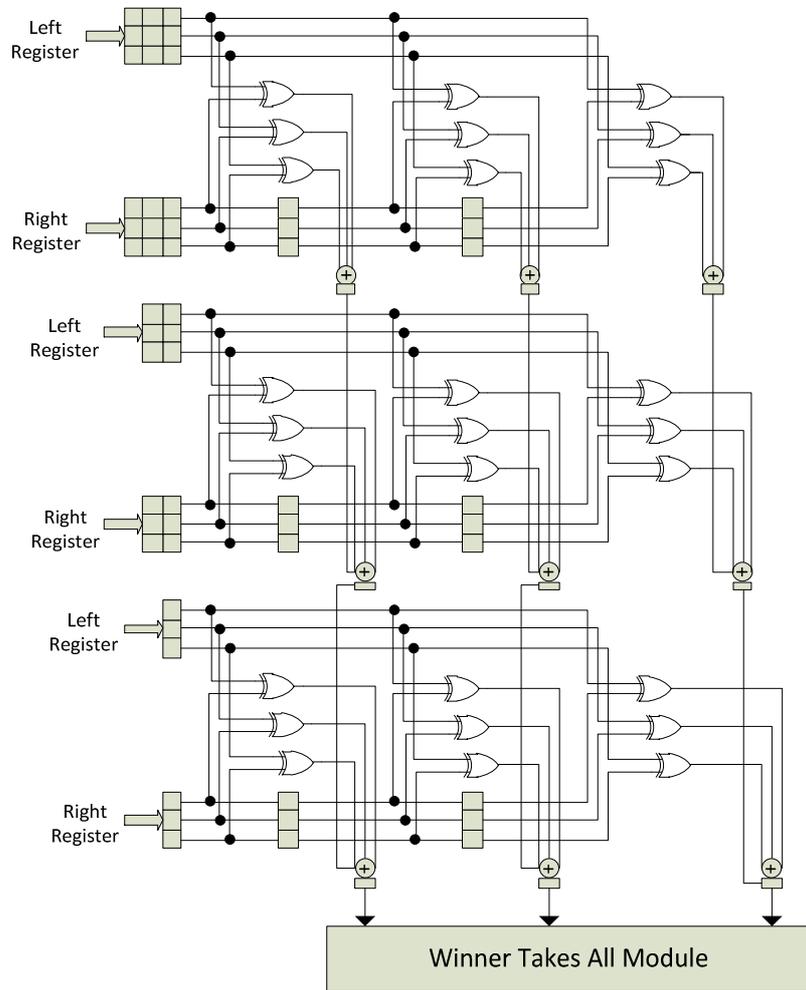
Stereo matching is divided into two stages, the census transform stage and the correlation stage. In the census transform stage, the left and right images are transformed into images with census vector pixel values instead gray-level intensity. This transform is a non-parametric measure for window based processing used during the matching process for measuring similarities and obtaining the correspondence between the points into the left and right images. In a classic hardware implementation for census transform, pixel neighborhood associated with central pixel needed to be accessed simultaneously to calculate in a single instant census vector. Last technique increases HW requirement with consequent frequency decrement and complexity [2]. However, in the presented architecture, a modified census transform is used. Proposed modification just takes, instead central pixel for census, first right pixel in central row. In test changes with proposed modification were not appreciable. With this in mind, necessary hardware to calculate census transform decrease significantly. But pixels in the same column need to be present at same time. To achieve this, architecture uses a scan-line buffer and windows buffer. Scan-line buffer is memory which is able to contain a row from input image in order to synchronize data and avoid additional access to external memory and a window buffer, which is a set of shift registers with the pixels belong to the window. Such windows buffers in other architectures [5][15] consists of 8 bit registers, but in this case registers contains only one bit. The scan-line buffer used in the proposed system consists of 10 dual-port memories, and each memory can store one scan-line of an input image. Assuming that the coordinates of

the current input pixel are  $(x, y)$  and the intensity value of the pixel is  $I(x, y)$ , the connections between the memory are shown in Fig. 3.



**Fig. 3.** 3x3 Architecture example. Implemented architecture is 11x11. Blocks in dark gray are 1 bit registers. In this case, after 3 latency clocks, registers contains a valid census transform. There are in the architecture two modules one for each channel.

Figure 3 shows a scan-line buffer converting a single row pixel input into a column pixel vector output. A window buffer is a set of 1 bit shift registers, but central line can store one 8 bit pixel from input image. One bit registers with comparison modules store values and delay results until all columns have been processing. Intensity values in central row register are shifted from left to right at each pulse clock to build the census transform. In comparison, with the classical technique where all  $11 \times 11$  pixel registers are implemented, it is obtained a windows buffer with  $121 \times 8 = 968$  bits, but in proposed architecture just 21 registers (11 left column register + 10 central row registers)  $\times 8 + 10 \times 10 = 261$  bits are necessary to implement census transform.



**Fig. 4.** Correlation module uses Hamming distance to establish similarity. Here is shown a  $3 \times 3$  windows module with a 3 level disparity as example. Proposed architecture uses 50 disparity levels.

Correlation stage evaluates the correlation between the census vectors generated by the left and right census transform. Stereo procedure in the correlation stage is:  $N$  hamming distances are evaluated using a template window for a pixel in the left image and the corresponding  $N$  correlation windows for pixels in the right image. After the comparison, the two pairs with the shortest hamming distances are used to define the resulting disparity. Since the windows being compared can be regarded as bit vectors, it is possible to obtain the hamming distance by counting '1' in the vector after applying an XOR operation [5]. Here, proposed architecture introduces another hardware artifice to decrease hardware: proposed architecture uses pipeline from fig-

ure 3 to calculate Hamming distance from left and right images without using a sum combinatorial tree (figure 4).

In order to decide upon the disparity result, the template window in the left image is compared with all  $N$  candidate windows from the right image. First, the census vector from the census transform module is delayed for  $N$  pixel clocks. Next, the distance between any two census vectors is calculated. Tournament selection method is used to find the shortest distance among these  $N$  hamming distances and winner takes all. The candidate window, which has the shortest distance from the template window, is selected as the closest match, and the coordinate difference of the selected windows along with the x-axis is extracted as the disparity result. In proposed system there are not sum combinatorial trees but RTL structures which cut combinational paths and increase frequency. Most left blocks are one bit registers en figure 3.

### 3.3 Tracking module

Tracking objects by correlation is a basic technique in tracking systems. In this case, add a tracking module in a real time architecture which does correlations with windows it is not a problem, although this tracking module just is added to one camera because of hardware circuitry complexity. Figure 5 shows the tracking module

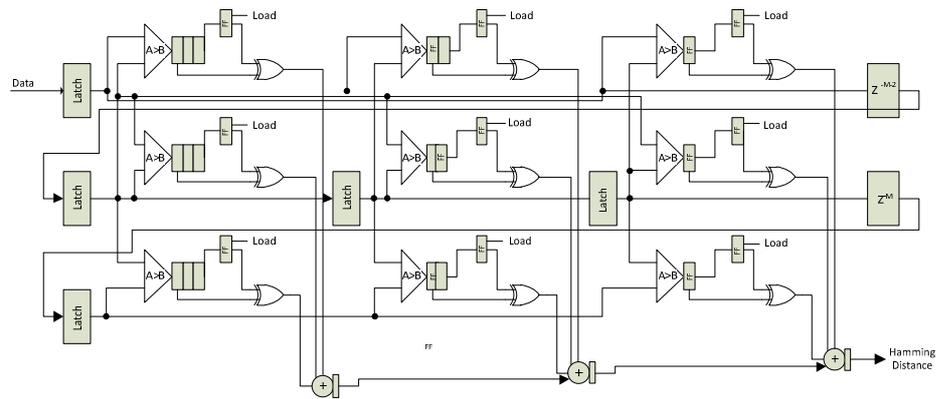


Fig. 5. 3x3 module for object tracking.

From figure 5, module is exactly same that census module except for some details: additional flip flop in the output from shift registers, XOR gates to evaluate hamming distance and sum sequential tree. Output from sequential tree is input to a coordinate module that gives to the architecture a coordinate range where objective tracking is calculated.

### 3.4 Post Processing

Post processing uses online subpixel disparity, i.e., no winning disparity is stored with near pixels to after calculate a subpixel estimation. Additionally, in order to avoid false correspondences, classical left-right check is substituted by uniqueness restriction. Briefly, assume that the left image is chosen as reference and the disparity candidates range build a disparity array  $[0, \dots, d_{max}]$ .  $L(x,y)$  is one point of left image, the algorithm searches for the best candidate by minimizing, in this case, a matching cost  $C$ . Suppose now the best match found for  $L(x-a+d_{max}, y)$  is  $R(x, y)$ , with matching cost  $C(x-a+d_{max}, x, y)$ . And another point of the left image  $L(x-b+d_{max}, y)$  has previously been matched with  $R(x, y)$  with cost  $C(x-b+d_{max}, x, y)$ . And another point of the left image  $L(x-b+d_{max}, y)$  has previously been matched with  $R(x, y)$  with cost  $C(x-b+d_{max}, x, y)$ . Based on the uniqueness constraint, we conclude that at least one of the two matches is incorrect and only the match with minimum cost is retained. This implies that the proposed approach allows for recovering from previous errors as long as better matches are found during the search. During the implementation, it only needs to set up  $d_{max}$  registers to keep track of the best match and corresponding matching cost for right image points in the range of interval. The match newly created for  $R(x,y)$  is compared with previous match, and the one be replaced will be labeled “incorrect”.

## 4 Results

The proposed real-time stereo vision system is designed and coded using VHDL and implemented using a Spartan XC3S1000 FPGA from Xilinx.. The implemented system interfaces two MT9M112 CMOS sensors from Micron as a stereo camera pair. Table I summarizes the device utilization reports from the Xilinx synthesis tool in ISE release 13i, Used FPGA resources for architecture are indicated in table 1:

**Table 1.** Device utilization summary

	Used	Available	Utilization
Occupied Slices	7554	7680	98%
Rectification	2311		30%
Census transform	1034		13%
Hamming Distance	1566		20%
Correlation	1867		24%
Post processing	776		10%

Since the hardware was built for real-time processing of an incoming image, the disparity results of the proposed design were generated through HDL functional simulation, i. e., a test bench was generated to probe architecture.

Table 2 compares different architectures with the proposed in this paper.

**Table 2.** Real time performance of reported stereo vision systems based on FPGA

Implemented system	Image Size	Matching method	Disparity	Fps
MSVM-III	640x480	Census	64	30
Kunh et al.	256x192	Ssd/census	25	50
Proposed Architecture	640x480	Census	50	52

Table 2 shows some systems found in literature the proposed architecture is based in a cheap FPGA and it is comparable with bigger and elaborate systems. Proposed architecture uses less hardware and is more suitable for mobile applications like robotic platforms. DeepSea founded in the literature is a multi board platform so it can be used for comparison.

Finally, Fig. 4 is the resultant disparity image from test images captured in different environments. The images were processed and obtained from the implemented system at different post-processing levels.



**Fig. 6.** Test stereo images and disparity maps from the proposed architecture

## 5 Conclusions

In this article, we have proposed a high performance FPGA-based stereo vision system with minimum cost using modified census transform, which can provide dense disparity information with additional sub-pixel accuracy in real time. The proposed system was implemented within a single FPGA including all the pre and post-processing functions such as rectification and uniqueness test. To achieve the targeted performance and flexibility, architecture was focused on the intensive use of pipelining and modularization. The proposed system can be used for higher level vision applications such as intelligent robots, surveillances, automotives, and navigation. Additional application areas in which the proposed stereo vision system can be used will continue to be evaluated and explored.

## 6 References

1. Vancea, C. and S. Nedevschi: LUT-based Image Rectification Module Implemented in FPGA. *Intelligent Computer Communication and Processing, 2007 IEEE International Conference on*, pages 147–154, Sept. 2007.
2. S. Birchfield and C. Tomasi. Depth discontinuities by pixel-to-pixel stereo. *Int. J. Comput. Vision*, 35(3):269–293, 1999.
3. M. Bleyer, M. Gelautz, C. Rother, and C. Rhemann. A stereo approach that handles the matting problem via image warping. In *CVPR09*, pages 501–508, 2009.
4. D. Comaniciu and P. Meer. Mean shift: A robust approach toward feature space analysis. *IEEE Trans. PAMI*, 24:603–619, 2002.
5. Y. Deng and X. Lin. A fast line segment based dense stereo algorithm using tree dynamic programming. In *Proc. European Conf. on Computer Vision (ECCV 2006)*, volume 3, pages 201–212, 2006.
6. P. F. Felzenszwalb and D. P. Huttenlocher. Efficient belief propagation for early vision. *Int. J. Comput. Vision*, 70(1):41–54, 2006.
7. S. Gehrig, F. Eberli, and T. Meyer. A real-time low-power stereo vision engine using semi-global matching. In *CVS09*, pages 134–143, 2009.
8. M. Gong, R. Yang, W. Liang, and M. Gong. A performance study on different cost aggregation approaches used in real-time stereo matching. *Int. Journal Computer Vision*, 75(2):283–296, 2007.
9. H. Hirschmuller. Accurate and efficient stereo processing by semi-global matching and mutual information. In *Proc. Conf. on Computer Vision and Pattern recognition (CVPR 2005)*, volume 2, pages 807–814, 2005.
10. H. Hirschmuller. Stereo processing by semi-global matching and mutual information. *IEEE Trans. on PAMI*, 2(30):328–341, 2008.
11. H. Hirschmuller and D. Scharstein. Evaluation of stereo matching costs on images with radiometric differences. *IEEE Trans. Pattern Anal. Mach. Intell.*, 31(9):1582–1599, 2009.
12. A. Hosni, M. Bleyer, M. Gelautz, and C. Rhemann. Local stereo matching using geodesic support weights. In *ICIP*, 2009.
13. A. Klaus, M. Sormann, and K. Karner. Segment-based stereo matching using belief propagation and a self-adapting dissimilarity measure. In *ICPR '06*, pages 15–18, 2006.

14. C. Lei, J. Selzer, and Y. Yang. Region-tree based stereo using dynamic programming optimization. In CVPR06, pages II: 2378–2385, 2006.
15. S. Mattoccia. A locally global approach to stereo correspondence. In 3DIM2009, pages 1763–1770, Kyoto, Japan, 2009.
16. S. Mattoccia, S. Giardino, and A. Gambini. Accurate and efficient cost aggregation strategy for stereo correspondence based on approximated joint bilateral filtering. In Proc. Of ACCV2009, 2009.
17. D. Scharstein and R. Szeliski. A taxonomy and evaluation of dense two-frame stereo correspondence algorithms. *Int. Jour. Computer Vision*, 47(1/2/3):7–42, 2002. 1, 2, 6



# Towards the improvement of the linearity-efficiency trade-off in TXs for MedRadio

Luis A. Sánchez-Gaspariano<sup>1</sup>, Carlos Muñoz-Montero<sup>1</sup>, Salvador A. Arroyo-Díaz<sup>1</sup>,  
José M. Trevera-Rodríguez<sup>1</sup>, Víctor H. Ponce-Ponce<sup>2</sup>

<sup>1</sup>Universidad Politécnica de Puebla, Tercer Carril del Ejido “Serrano” S/N, San Mateo Cuanalá,  
Juan C. Bonilla, Puebla, México.

<sup>2</sup>Centro de Investigación en Computación del Instituto Politécnico Nacional, Av. Juan de Dios  
Bátiz, Esq. Miguel Othón de Mendizábal, Col. Nueva Industrial Vallejo, Delegación Gustavo  
A. Madero, México, D.F.  
luisabraham.sg@gmail.com

**Abstract.** The design trade-offs of a transmitter (TX) for the Medical Device Radiocommunications Service (MedRadio), which encompasses low power consumption in a bandwidth from the 401MHz to the 406MHz, small form factor and robustness, are discussed in this work. The aim is to determine the key factors for linearity and efficiency in a direct-up TX architecture suitable for MedRadio. A solution based on the use of a predriver amplifier preceding the power amplification stage with output power modulation is sketched. Experimental results of the predriver prototype in a double poly, three metal layers, 0.5 $\mu$ m CMOS technology show the effectiveness of the proposed circuit for high efficiency frequency translation.

**Keywords:** MedRadio, Wireless Direct-Up TX, predriver amplifier.

## 1 INTRODUCTION

The use of wireless technology in medical devices is nowadays a reality which is turning into a common practice for both, preventive and corrective health care. Some examples of such wireless devices include endoscopes [1], pacemakers [2], fingertip pulse oximeters [3], and diverse sensors for m-Health systems [4]. In order to make easier the employment of wireless in implanted medical devices, the Federal Communications Commission (FCC) has established the MedRadio, whose frequency band lies in the 401 to 406MHz frequency range [5]. In addition, the effective isotropic radiated power (EIRP) is limited to  $-16$ dBm and the channel bandwidth restricted to 300kHz maximum. Depending on the application, a MedRadio transceiver has to meet several criteria. Nevertheless, most of the wireless medical applications demand the following [6]: very low power consumption, small form factor, robust link with low Bit Error Rate (BER), and channel selectivity within the MedRadio band.

In general terms, to achieve a low cost and a small form factor in the implementation of CMOS transceivers, commercial vendors as well as researchers have focused

adfa, p. 1, 2011.

© Springer-Verlag Berlin Heidelberg 2011

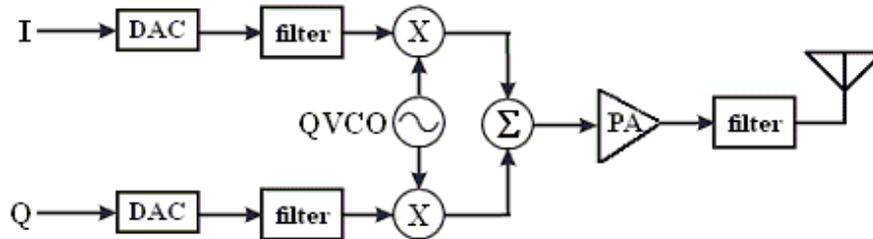


Fig. 1. Zero-IF TX architecture.

on direct conversion architectures [7]. On the other hand, to reach a high spectral efficiency in the TX, digital modulation techniques such as different forms of Quadrature Amplitude Modulation (QAM) are adopted, resulting in a transmit waveform with a high Peak-to-Average Power Ratio (PAPR). This in turn requires a linear Power Amplifier (PA) to drive the antenna and typically, linear PAs exhibit a low-Power Added Efficiency (PAE), resulting in high-power consumption. Thus, the design of an adequate TX architecture for MedRadio still is an open problem which must be assessed.

This paper is organized as follows: in section two, a quick review of the direct-up TX is presented focusing on its most important characteristics as well as pointing out its advantages as well as its drawbacks with respect to the MedRadio system; in section three, the study of the key factors for low power consumption and channel selectivity in the TX is addressed; later in section four, a possible solution is sketched; section five describes a circuit implementation example, with its experimental results, of the proposed solution in a double poly, three metal layers,  $0.5\mu\text{m}$  CMOS technology; finally, the conclusions of the work are drawn in section six.

## 2 THE ZERO-IF TX ARCHITECTURE

Figure 1 shows the block diagram of a zero-IF TX. As can be seen, the quadrature signals pass through a digital-to-analog converter (DAC) and once filtered those are up converted to RF by the quadrature VCO. Finally, the PA reinforces the output signal, which is filtered again before to reach the antenna in order to suppress harmonic components and intermodulation (IM) products produced by the non-linearities of the PA.

According with the results reported in diverse works [7], the direct-up TX architecture is attractive since cost and form factor are good enough and its integrability is high. Moreover, it presents no image problems. Unfortunately, complexity rises by the LO-leakage. The use of techniques such as offset VCO and LO-leakage calibration are somehow necessary to alleviate this problem. It is important to remark that most of the wireless medical applications demand, among others, very low power consumption and channel selectivity within the MedRadio band. Therefore, if direct-up architecture is intended to be employed in a MedRadio TX, it has to exhibit linearity and low power consumption.

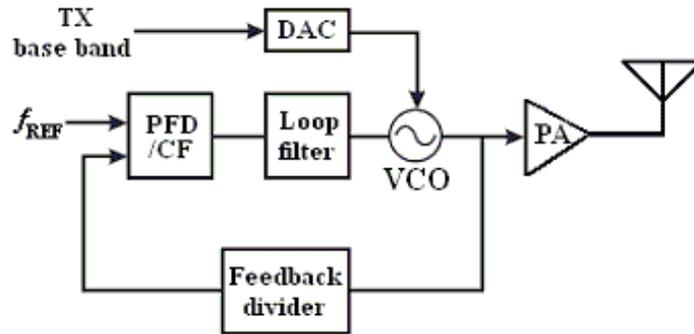


Fig. 2. Zero-IF TX architecture.

One possibility for having both, linearity and low power consumption is to use a zero-IF constant envelope TX [6]. Since it can be driven close to compression, this results in a high PAE of the PA; which means lower power consumption. Fig. 2 depicts the block diagram of the TX proposed in [6] for the Medical Implant Communication Service. It is a constant envelope architecture which utilizes Frequency Shift Keying (FSK) modulation. As can be appreciated, the modulation control of the transmitter is embedded within a Phase Locked Loop (PLL). This yields several benefits. First, it is another step toward lowering power consumption, since this scheme removes the two up converters in Fig. 1. Second, the PLL provides inherent filtering for the modulated signal. Third, the use of direct VCO modulation modifies the control voltage via a DAC and since the transfer function from the VCO-to-PLL is of the high-pass form, the modulation must concentrate most of its power spectral density outside of the PLL bandwidth.

Another possibility for high linearity is the power up converter (PU) proposed in [8] based on the polyphase multipath technique and shown herein in Fig. 3. The differential pairs are driven as transconductors by the corresponding phase shifted base band signal (BB) meanwhile the transistor below acts like a switch driven by the corresponding phase shifted square wave from the local oscillator (LO). The elements labeled as  $L_{RF}$  are choke inductors and  $Z_L$  is the load of the matching network between the antenna and the driver. A good linearity is achieved with the employment of this circuit, -40dBc worst case harmonic rejection with a six phases system @ 2.4GHz [8]. However, the drain efficiency of the PU is rather low (11% for 9dBm output power) and this implies a high power consumption. On the other hand, the PU does not necessarily require an output filter since unwanted spectral components can be significantly reduced in a flexible manner with the use of multiple paths and multiple phases. This is an attractive characteristic considering that it reduces costs and improves the form factor of the TX since dedicated filters out of chip can be eliminated. Moreover, direct conversion TX with a polyphase multipath scheme provides a flexible architecture for multiple standards [8]. This is a crucial step to TXs for MICS, MedRadio and Wireless Medical Telemetric Service (WMTS) [6].

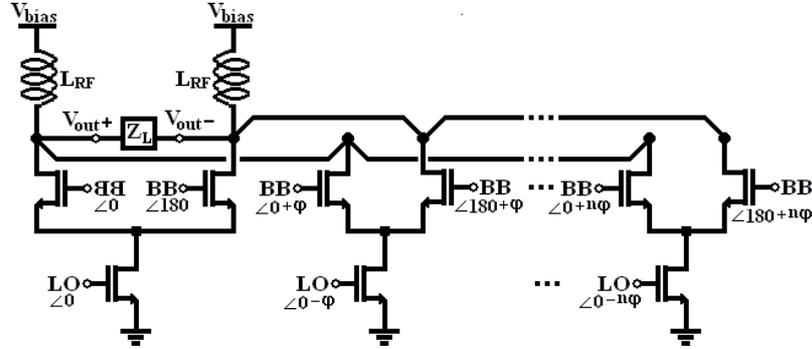


Fig. 3. Power up converter based on the polyphase multipath technique.

### 3 POWER SAVING AND LINEARITY

In the previous section we saw that a polyphase multipath direct conversion TX is a good choice for linearity, form factor, and flexibility for multi-standard wireless technologies. Nevertheless, the PU reported in [8] exhibits poor efficiency and therefore its power consumption is rather high. That is an unwanted feature in medical applications. On the other hand, in order to save power the solution proposed in [6] utilizes a constant envelope nonlinear PA. Thus, by using both techniques, an efficient nonlinear PA within a polyphase multipath architecture, a trade-off between power consumption, channel selectivity and form factor can be reached.

There are three parameters which play an important role in the efficiency of any PA: the input power ( $P_{IN}$ ); the output power ( $P_{OUT}$ ); and the DC power consumption ( $P_{DC}$ ). The ratio between  $P_{OUT}$  and  $P_{IN}$  defines the drain efficiency,  $\eta_D$ , whereas the ratio between the difference of  $P_{OUT}$  minus  $P_{IN}$  and  $P_{DC}$  defines the power added efficiency,  $PAE$ , which is the most popular measure in industry. An alternative manner to express the  $PAE$  is as a function directly proportional to  $\eta_D$  and inversely proportional to the power gain ( $A_p$ ), i.e.

$$PAE = \frac{P_{OUT} - P_{IN}}{P_{DC}} = \frac{P_{OUT}}{P_{DC}} \left( 1 - \frac{P_{IN}}{P_{OUT}} \right) = \eta_D \left( 1 - \frac{1}{A_p} \right) \quad (1)$$

Hence,  $\eta_D$  and  $A_p$  are relevant for a high efficiency performance. In terms of the conduction angle, the  $\eta_D$  for switched (nonlinear) PAs can be expressed as [9]

$$\eta_D = \frac{2 \sin^2(\phi)}{\phi(\pi - \phi)} \quad (2)$$

where  $\phi$  is half the total conduction angle,  $2\pi$ , of the driving signal.

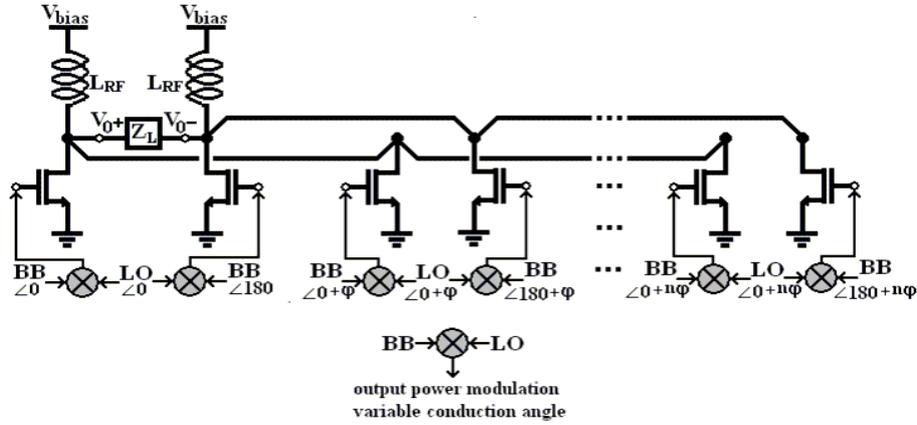


Fig. 4. Direct conversion TX based on the polyphase multipath architecture.

By inspection of (1) and (2) it can be seen that, if sufficient  $A_P$  can be attained and a good  $\eta_D$  is exhibited by the driver, then the PAE of the power amplification stage can be high and therefore the power consumption of the TX adequate for medical applications. Probably, one of the simplest manners in which the  $A_P$  can be enlarged is to reduce the  $P_{IN}$ . Fig. 4 shows the circuit diagram of the proposed direct-up TX within a polyphase multipath architecture. It consists of  $n$  pseudo differential pairs of switched PAs driven by predriver stages compound by mixers. In this case, the mixer is at the back driving the power amplification stage. In such a way, the power transference of this stage to the load can be improved by modulating the power delivered by the mixer, i.e. the input power of the PA. Besides, if the mixed signal is fed into the PA such that it is capable of modifying the conduction angle of the amplifier, then also the  $\eta_D$  can be controlled. Moreover, the power demands on the BB can be relaxed. In sum, the allocation of the mixer out of the power amplification stage is more convenient for power efficiency and consequently lower power consumption.

#### 4 LINEAR AND EFFICIENT TX SCHEME FOR MeDRadio

Typically, the usual metrics to evaluate the performance of a mixer are: conversion gain/loss, noise figure, port isolations, linearity, and power consumption [10]. From those, noise figure and port isolations are of major concern in case that the mixer is intended to be used in a receiver circuit. In addition, with the use of the multipath polyphase technique, linearity is, in principle, not a major issue. Instead, the demands on the performance of the mixer used as a predriver stage just like depicted in Fig. 7 in multipath polyphase architecture with improved efficiency can focus in the following aspects:

- Output power modulation.
- Capability of modifying the conduction angle of its output signal, i.e. the driving signal of the PA.

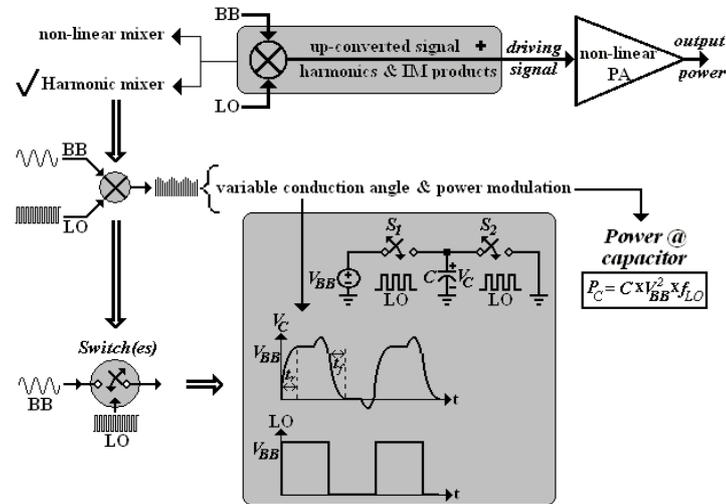


Fig. 5. Switched circuit used as a mixer.

From the many mixer architectures available, it is difficult to find one which exhibits these features. There are basically two ways to accomplish the frequency translation of two signals. The first method is by means of a nonlinear circuit whose nonlinearities generate harmonics and IM products, and the second possibility is with the use of a time variant circuit. Switching circuits known as harmonic mixers are used in zero-IF TXs to up convert the BB to RF. Since harmonic mixers possess a low self-mixing DC offset, they result attractive for zero-IF TXs, with the drawback of having a conversion gain usually small [7].

Fig. 5 depicts the action of a simple switching circuit with two switches,  $S_1$  and  $S_2$ , and one capacitor,  $C$ . The LO signal controls the action of  $S_1$  and  $S_2$  such that they operate in turn. When LO is at a high value  $S_2$  is open and  $S_1$  is closed leading to the charge of capacitor  $C$  at voltage  $V_{BB}$  which is provided by the voltage source on the left. Assuming there is some resistance value between the terminals of  $S_1$  when it is closed then there is an RC circuit given by the series connection of  $C$  and the resistive value of the switch. The time constant of the RC circuit establishes a rise time,  $t_r$ , in which  $C$  is charged at approximately  $V_{BB}$ . On the other hand, when LO is at a low value  $S_1$  is open meanwhile  $S_2$  is closed; again, assuming there is an series RC circuit given by the resistive value between the terminals of  $S_2$  when it is closed and the capacitor  $C$ , the time constant given by this RC circuit establishes a fall time,  $t_f$ , in which  $C$  is discharged completely. Note that if there are different rise and fall times in the circuit it is possible to alter the duty cycle of the resulting signal. In other words, we can modify the conduction angle of the signal at the output by changing the time constants in the circuit, which can be done easily by changing the resistive value of the switches. Moreover, the power at the output of circuit such as the one depicted in Fig 5 is given by [11]

$$P_C = C \times V_{BB}^2 \times f_{LO} \tag{3}$$

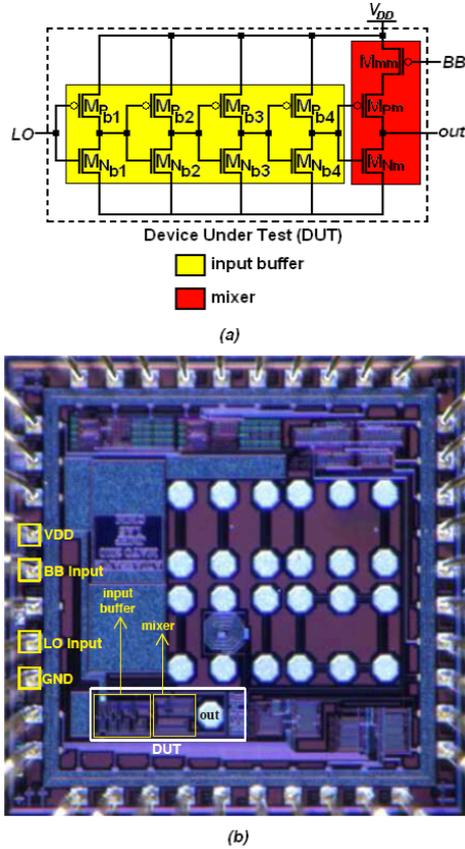


Fig. 6. Proposed predriver CMOS circuit: (a) circuit diagram and (b) prototype die photo.

where  $f_{LO}$  is the frequency of the LO. According with this expression, the output power is proportional to the quadratic value of the  $V_{BB}$  voltage. Thus, the output power can be controlled by changing the value of  $V_{BB}$ . Therefore, it is convenient to use a switched RC network as a pre-driver mixer for the PA stage of the TX since it allows to modulate the power at its output and to change the duty cycle of the output waveform.

## 5 EXPERIMENTAL RESULTS

A prototype of the predriver stage was designed and fabricated in a double poly three metal layers  $0.5\mu\text{m}$  CMOS technology from MOSIS foundry. Fig. 6 (a) shows the circuit diagram of the fabricated circuit. As can be seen, it consists of an input buffer which turns the input sinusoidal signal from the LO into a square wave signal with 50% duty cycle followed by the mixer. A total of 11 transistors are used. From those, only three belong to the mixer. The  $(W/L)$  ratios used for the devices in the

**Table 1.** Experimental results of the predriver stage.

<b>LO Frequency</b>	393.5MHz
<b>BB Frequency</b>	10MHz
<b>Power Supply, <math>V_{dd}</math></b>	3.3V
<b>Tuning control</b>	1.0V – 1.8V
<b>Maximum Power Consumption</b>	88.2 $\mu$ W
<b>Minimum Power Consumption</b>	68.2 $\mu$ W
<b>Maximum Output Power</b>	-13dBm
<b>Minimum Output Power</b>	$\approx$ -19dBm
<b>Maximum Duty Cycle</b>	50%
<b>Minimum Duty Cycle</b>	$\approx$ 30%
<b>Maximum Pulse Amplitude</b>	2.5V
<b>Minimum Pulse Amplitude</b>	0.4V

buffer were (19.8 $\mu$ m /0.6 $\mu$ m), (66 $\mu$ m /0.6 $\mu$ m), (115.5 $\mu$ m /0.6 $\mu$ m) and (264 $\mu$ m /0.6 $\mu$ m) for the transistors  $M_{pb1}$ ,  $M_{pb2}$ ,  $M_{pb3}$  and  $M_{pb4}$ , respectively; whereas the ratios for the transistors  $M_{Nb1}$ ,  $M_{Nb2}$ ,  $M_{Nb3}$  and  $M_{Nb4}$  were (15 $\mu$ m /0.6 $\mu$ m), (30 $\mu$ m /0.6 $\mu$ m), (45 $\mu$ m /0.6 $\mu$ m) and (120/0.6), respectively. For the case of the switches in the mixer, the ratios occupied were (180 $\mu$ m /0.6 $\mu$ m) for transistors  $M_{pm}$  and (540 $\mu$ m /0.6 $\mu$ m) for transistors  $M_{Nm}$ . Finally, the ratio of transistors  $M_{mm}$  was (480 $\mu$ m /0.6 $\mu$ m). The prototype area is 0.472 $\mu$ m X 0.148 $\mu$ m including the output pad. The prototype die photo is depicted in Fig. 6(b). As can be appreciated, the bias input,  $V_{DD}$ , the BB input, the LO input and the ground are fed to the bond pads of the chip meanwhile the output node must be taken from an inner-pad labeled as out.

Table 1 summarizes the most important characteristics of the prototype. The LO frequency was 393.5MHz meanwhile the BB frequency was 10MHz. The power supply for the circuit was of 3.3V. The tuning control was of 800mV ranging from 1.0V to 1.8V. The maximum power consumption was 88.2 $\mu$ W and the minimum power consumption was 68.2 $\mu$ W. The maximum output power was -13dBm meanwhile the minimum output power was -19dBm approximately. The duty cycle of the prototype was modifiable in a range of 20% from 50% to 30%. Finally, the maximum pulse amplitude of the output signal was 2.5V whereas the minimum was 400mV.

Fig. 7 (a) shows the waveform obtained at the output of the prototype for a voltage value of the tuning control of 1.2V. On the other hand, Fig. 7 (b) depicts the spectrum at the output of the prototype within the MedRadio band. It can be seen that the frequency components  $\omega_{LO} \pm 4\omega_{BB}$ ,  $\omega_{LO} \pm 3\omega_{BB}$ ,  $\omega_{LO} \pm 2\omega_{BB}$ ,  $\omega_{LO} \pm \omega_{BB}$ , are present with enough power to be taken into account ( $\omega_{LO} = 2\pi f_{LO}$ ). Nevertheless, further cancellation of the intermodulation distortion in the transmitter chain can be performed by using a band pass filter to select the frequency of interest,  $\omega_{LO} + \omega_{BB}$ .

## 6 CONCLUSIONS

The design issues of TXs for MedRadio circuits have been discussed. Due to the requirements of medical applications such as low power consumption, small form

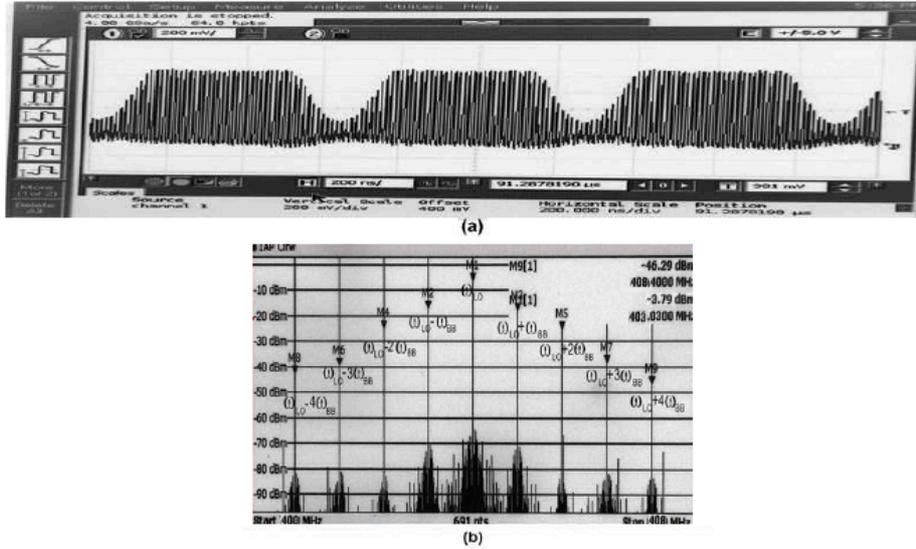


Fig. 7. Output signal of the prototype: (a) waveform in time domain and (b) spectrum within the MedRadio band.

factor and channel selectivity, a direct-up approach intended to exhibit a good efficiency within a polyphase multipath architecture have been sketched as an alternative to satisfy those demands. The key aspects for power saving lies in the pre-driver stage which must perform two actions: mixing the LO and BB and driving the PA with variable conduction angle and modulated output power.

The experimental results were: a tuning control of the predriver of 800mV ranging from 1.0V to 1.8V; a maximum power consumption of  $88.2\mu\text{W}$  and a minimum power consumption of  $68.2\mu\text{W}$ ; a maximum output power of -13dBm meanwhile the minimum output power was -19dBm; a duty cycle modifiable in a range of 20% from 50% to 30%; a maximum pulse amplitude of the output signal of 2.5V whereas the minimum was 400mV. In addition, the frequency components  $\omega_{LO} \pm 4\omega_{BB}$ ,  $\omega_{LO} \pm 3\omega_{BB}$ ,  $\omega_{LO} \pm 2\omega_{BB}$ ,  $\omega_{LO} \pm \omega_{BB}$ , are present with enough power to be taken into account. However, further cancellation of the intermodulation distortion in the transmitter chain can be performed by using a band pass filter to select the frequency of interest,  $\omega_{LO} + \omega_{BB}$ .

According to the results obtained in the characterization of the prototype, we conclude that the behavior of the predriver stage follows the course anticipated in the synthesis of the mixer performed in section 4.

## 7 ACKNOWLEDGEMENTS

Authors would like to thank the Programa de Mejoramiento del Profesorado (PROMEP) of México for the financial support given through the project number UPPUE-PTC-038

## 8 REFERENCES

1. K. Chul and S. Nooshabadi, "A DTR UWB transmitter/receiver pair for wireless endoscope", in *Proc. IEEE Asian Solid-State Circuits Conference (ASSCC)*, Taipei, 2009, pp. 357-360.
2. D. Halperin et al. "Pacemakers and Implantable Cardiac Defibrillators: Software Radio Attacks and Zero-Power Defenses", in *Proc. IEEE Symposium on Security and Privacy (SP)*, San Francisco California, 2008, pp. 129-142.
3. N. Watthanawisuth et al. "Wireless wearable pulse oximeter for health monitoring using ZigBee wireless sensor network", in *Proc. IEEE Electrical Engineering/Electronics Computer Communications and Information Technology (ECTI-CON)*, Chiang Mai, 2010, pp. 575-579.
4. R.S.H. Istepanian et al. "Guest Editorial: Introduction to the Special Section on M-Health: Beyond Seamless Mobility and Global Wireless Health-Care Connectivity", *IEEE Transactions on Information Technology in Biomedicine*, vol. 8, issue. 4, pp. 405-414, Dec. 2004.
5. Federal Communications Commission, "*Wireless Networks of Implanted Medical Devices*" [Online]. Available: <http://www.fcc.gov/document/wireless-networks-implanted-medical-devices>.
6. C. Siu and K. Iniewski, "CMOS Circuits for Wireless Medical Applications", in *VLSI circuits for Biomedical Applications*, K. Iniewski, Ed., London: Artech House, 2008, pp. 111-131.
7. Pui-in Mak et al. "Transceiver architecture selection: review, state-of-the-art survey and case study", *IEEE circuits and systems magazine*, vol. 7, Issue 2, pp. 6-25, Jun. 2007.
8. Sherestha, R. et al. "A Polyphase Multipath Technique for Software-Defined Radio Transmitters", *IEEE J. Solid-State Circuits*, vol. 41, No. 12, pp. 2681 – 2692, Dec. 2006.
9. S.C. Cripps, "RF Power Amplifiers for Wireless Communications", 2<sup>nd</sup> ed. Artech House, 2006, ch. 7, pp 179-218.
10. B. Razavi, "RF Microelectronics", 2<sup>nd</sup> ed. P. Hall, 2012, ch. 6, sec. 6.5, pp. 408-424.
11. J. M. Rabaey et al. "Digital Integrated Circuits: A Design Perspective", 2<sup>nd</sup> ed. Pearson Education, Inc., 2003, ch. 5, sec 5.5, pp. 214-227.

# **Natural Language Processing**

---



# A Proposal of Automatic Error Correction in Text

Wulfrano A. Luna-Ramírez, Carlos R. Jaimez-González

Departamento de Tecnologías de la Información, Universidad Autónoma Metropolitana –  
Cuajimalpa, Av. Constituyentes No. 1054, Col. Lomas Altas, C.P. 11950, México, D.F.  
{wluna,cjaimez}@correo.cua.uam.mx

**Abstract.** The great amount of information that can be stored in electronic media is growing up daily. Many of them is got mainly by typing, such as the huge of information obtained from web 2.0 sites; or scanned and processing by an Optical Character Recognition software, like the texts of libraries and government offices. Both processes introduce error in texts, so it is difficult to use the data for other purposes than just to read it, i.e. the processing of those texts by other applications like e-learning, learning of languages, electronic tutorials, data minning, information retrieval and even more specialized systems such as tiflogologic software, specifically blinded people-oriented applications like automatic reading, where the text would be error free as possible in order to make easier the text to speech task, and so on. In this paper it is showed an application of automatic recognition and correction of ortographic errors in electronic texts. This task is composed of three stages: a) error detection; b) candidate corrections generation; and c) correction -selection of the best candidate. The proposal is based in part of speech text categorization, word similarity, word dictionaries, statistical measures, morphologic analisys and n-grams based language model of Spanish.

**Keywords.** Automatic error correction, language models, n-grams, word eskeleton, part of speech, text processing, string comparision, edit distance.

## 1 Introduction

The great amount of electronic texts comes mainly from the internet, where the irruption of social networks and the Web 2.0 has provide us a huge quantity of data in just a little period of time. Another great source of information is the data obtained from the scanning of printed documents by means of an Optical Character Recognition (OCR), used by the libraries and government offices in order to preserve files of historic information. Given this enormous quantity of data, there is a need to apply automatic tools to process, transform and extract this information. Even more, a lot of such text contains typing errors and this makes difficult to keep the processing of such texts by other applications; of course, OCR is another source of errors in texts [1],[2],[3],[6],[7]. So, to retrieve text without orthographic errors is a difficult task.

Additionally, when a software tool is used, the main problem of having text with errors is the difficult to use this information as the input of more specialized and even

useful applications, like the wide set of Natural Language Processing (NLP) tools [2],[11]. Some examples of such applications are: data mining (and of course web mining), information searching and retrieval, document categorization, and those systems related with education (learning of foreign languages), phonetic translation, code and text computer aided edition, and aided to people with disabilities: text to speech conversion, blind people software assistance, accessibility applications, etc. [9],[10].

Fortunately, the same knowledge of the NLP field can help to design some methods to cope with errors in texts, like the linguistic knowledge. In this way, according to the linguistic point of view, there are different levels of text treatment or processing, some of them are: a) orthographic and morphologic level: describes the structure and external features of words, i.e. the way the letters are articulated in certain language; b) syntactic level: describes the paragraphs and the phases, i.e. the word organization within a text, and the grammatical categories of words, (commonly called Part of Speech -POS); c) semantic level: fully related with the meaning of text, takes care of the context where phrases and words appear; d) contextual or pragmatic level: describe the specific use of words, locutions, phrases, etc., in a certain situation (the discursive and temporal use) within a domain. On this way, the use of certain linguistic knowledge is an adequate approach to design automatic applications of text like the error correction task [1],[2].

In this work, it is used knowledge of the morphologic and contextual levels, specifically a POS tagger, and different techniques of word matching, additionally a language model based in n-grams is used [10],[11].

Briefly, the method is based on speech text categorization, word similarity, word dictionaries, statistical measures, and n-grams model of language.

In the next sections it is exposed the proposal. The section 2 describes the process of error correction and a brief classification of text errors is given; the section 3 describes the method of error correction; after that, some implementation details, experiments, results and discussion are exposed in section 4; section 5 includes conclusion and future work.

## 2 The Process of Error Correction in Electronic Text

The Error Correction in Electronic Text is performed in three stages [1],[6],[7]:

1. **Error detection:** It is oriented to text revision in order to identify the chain of characters from the text (words and other marks) as part of a given language, typically by means of a dictionary comparison, or a grammatical structure (morphological, syntactical or semantic).
2. **Candidate correction generation:** detects some possibilities of correction to a given error.
3. **Correction:** it is the selection of a specific candidate correction and the substitution of it in the text.

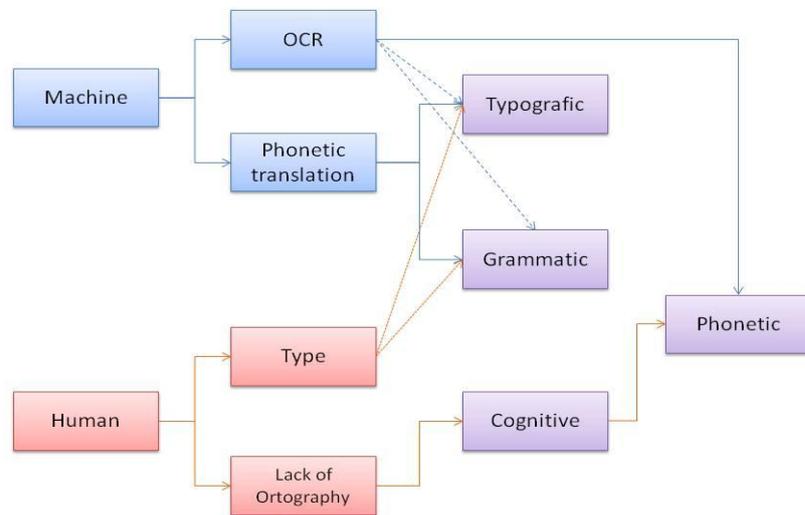
In **interactive error correction** the system corrector performs the first and the second stages in an autonomous way, and left to the user the final stage, where the

decision of which is the real correction is taken. In the other hand, the **automatic error correction**, the system does the three stages without the need of the user final decision [1].

As can be inferred, the automatic error correction is necessary to PLN applications where the full processing is performed with no user intervention.

Even more, the correction can be divided in two categories [1]: a) **isolated word error** correction: where the three stages described before are performed just word by word; b) **the contextual error correction**: where can be observed the context of the word, i.e. the phrase where it is used.

On the other hand, there is a brief classification of errors [1,7] that can be founded in text (obtained from humans –typing- or by machines –mainly OCR methods). Some error can derivate in linguistic mistakes by the accidental generation of invalid words within the language, or can be match with another valid word but not the correct one. The error classes could be part each other, but in order to understand the nature of error it is useful to try one classification, which is depicted in **Figure 1**: a) typographical: contains wrong characters in the string or word; b) grammatical: violate the articulation rules of a certain language; c) cognitive: the origin of the errors is the lack of knowledge of orthographic rules; d) phonetical: they are wrong representations of a given linguistic utterance (a phonetic chunk of language), those are the worst case of error, because generate a great word malformation and the semantic knowledge implied is high.



**Figure 1.** A classification of error in text according to their origin: human being or computer systems. One error can originate different cases of linguistic mistake, as is depicted,

for example an OCR error can be derivate in a phonetic error.

From the linguistic knowledge, those different types of errors can be organized accordingly with the language level related: a) morphologic; b) syntactic; c) semantic; d) speech structure; and e) pragmatic.

The types of errors are used to design techniques which can be applied to the stages of error correction in texts. One case of error that has been mainly treated by PLN applications are word error, which are described in the following subsection given the importance for the automatic or semiautomatic error correction methods that had been tried [1].

## 2.1 Word level errors

The level word errors can be treated relatively easy by a computer program, but the propagation of the mistake accordingly with the linguistic levels needs to be coped with techniques which imply knowledge about the other levels of the language. The most successful application use morphological and syntactic knowledge, but the superior levels when are incorporated in them originate domain dependent applications.

The two main sources of electronic texts (human or OCR) introduce certain classes of errors that had been afforded by different algorithms, like the edit distance (called Levenshtein distance) for the morphological errors; and n-gram based comparisons, for the syntactic and (in a very restricted sense) semantic errors [1]:

- **Insertion:** from the addition of one character in the string.
- **Duplication:** the doubling of one character in the string.
- **Deletion:** to omit one character in the string.
- **Substitution:** to replace one character by other.
- **Transposition:** where two characters are swapped.
- **Segmentation:** when two strings are formed from just one word.
- **Union:** when two words originate one string.

The segmentation errors are the most frequent in texts from OCR, on the other hand, the human being commonly produce union errors when typing.

## 2.2 N-grams

In order to verify if a sequence of linguistics elements (words, letters, phrases, etc.) within a text is valid for a given language, it is common use models of languages which represent the constraints of a certain combination of linguistics elements in order to build right structures in language [1],[11],[12].

One model of languages is n-grams, a statistical model of sequences of linguistic elements: typically n word or letters where  $n \geq 2$ . It is based in the estimation of sequences probability calculated from a text corpus called the **transition probability**, it can give a prediction of the n next element from the n previous according to the frequency with they appear in corpus. N-grams model uses the history based Markov supposition: a linguistic element is affected by the local context, so the previous text

affects the future text [1],[11]. Some examples of n-grams or letters and words are showed in **Figure 2**.

Letters N-gram order	N-gram example
monogram	{i,n,f,o,r,m,a,t,i,c,s}
bigram	{in,nf,fo,or,rm,ma,at,ti,ic,cs}
trigram	{inf,nfo,for,orm,rma,mat,ati,tic,ics}

Words N-gram order	N-gram example
monogram	{computing,is,not,easy}
bigram	{computing,is} {is,not} {not,easy}
trigram	{computing,is,not} {is,not,easy}

**Figure 2.** The first table shows n-gram examples of the word *informatics*, and the second contains n-gram examples of the phrase: *computing is not easy*. Where  $n = \{1,2,3\}$

N-grams had been used in error detection and the candidate generation stages of the error correction task [1]. In the proposal presented in this paper lexicons of bigram and trigram were used.

In the next sections the proposal of automatic error correction and the implementation are exposed.

### 3 The Automatic Error Correction Proposal

The error correction can be focused to the isolated words or taking account the local context of them. The proposal presented in this paper (POS-Tagged Automatic Error Correction -PAEC) is oriented to automatic error correction of the chain of words in a paragraph, and eventually in the full text [14]. It is based on POS text categorization, word similarity comparison, word dictionaries, statistical measures, and n-grams language models.

Some techniques used in automatic detection, candidate generation and selection of the correction use probability and morphology of words. The proposal adds to those techniques a POS-tagging process for the sake to augmenting the linguistic knowledge in the full process, i.e. uses the morphological and syntactical information present in the text under revision.

The proposal uses the following resources (texts written in Spanish): 1) a plain text corpora (CT), 2) a POS-tagged corpora (CA) with ten grammatical categories defined: verbs, nouns, conjunctions, idioms, articles, adjectives, adverbs, pronouns, interjections and miscellaneous; 3) a text corpora with inserted errors (CE), from a OCR and random deletions, insertions or substitutions; 4) a set of word lexicons: nine of each POS defined, and one of words from all grammatical categories; 5) a n-gram model of language (bigrams and trigrams); and 6) a n-gram model of language of POS-tags, which identify the n-gram occurrence of certain POS-tags (bigrams and trigrams).

The PAEC proposal is showed in **Figure 3**. It is composed of the following modules for to analyze a text:

1. **Preprocessing:** the abbreviation words are expanded, upper case and punctuation symbols are extracted, and the resulting text is separated in sentences and words.
2. **POS-tagging.** The text is POS-tagged in order to identify the grammatical category they belong. The POS-tagger used is an implementation of TBL POS-tagger [4], trained which semiautomatic annotated Spanish text.
3. **Word extraction:** it separates the words and the POS-tagged related, in order to recover the original text, but identifying the grammatical category of words. Produce a list of word for being analyzed.
4. **Contextual error detection:** this task makes a list of possible wrong words in text. It is carried out in two phases:
  - a. Seeks each word from the text in their corresponding POS-lexicon.
  - b. An analysis of bigrams that can be formed with the words from the text for identifying the abnormal combinations, because it can discover the presence of an error in text.

The process is showed in the **Algorithm 1**, as follows:

**Input:**

P: List of words to be corrected  
 LN: Lexicon of language model (n-grams)

**Output:**

PE: List of words identified as errors

**Var:**

k: It represents the index of P word it is being reviewed in each iteration.  
 P\_k: It is the k member of P  
 |P|: Total number of words in the list P  
 LP: Word lexicon  
 pa: It is the word is being reviewed in each iteration (P\_k)  
 ant: previous word of pa  
 post: next word of pa  
 flag: It indicates if a bigram of pa has been found  
 <I>: It is an inserted pseudo-word at begin of each line

**Functions:**

add(l,e): add the e element to the list l

**Begin**

1. k = 0
2. While k < |P| do
3.   ant = P\_k-1
4.   pa = P\_k
5.   post = P\_k+1
6.   If pa != <I>
7.       If pa not in LP then
8.           add(PE,pa)

```

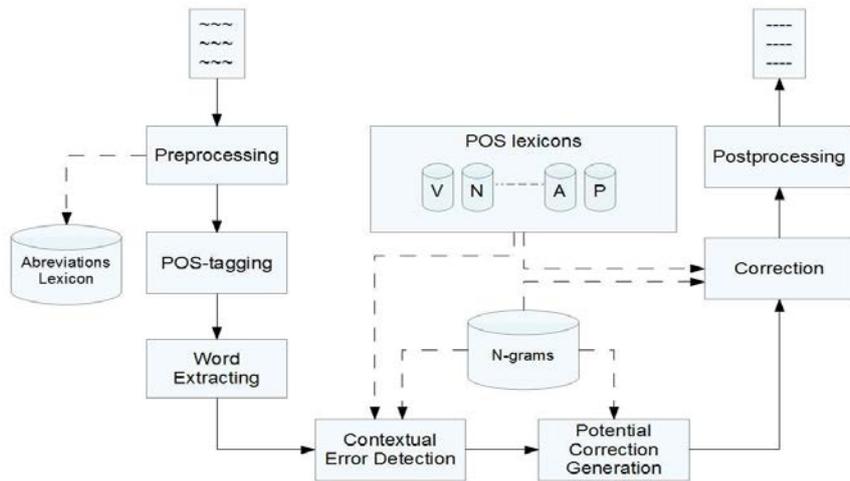
9.     else
10.    If [ant,pa] not in LN then
11.      If ant not in LP then
12.        add(PE,ant)
13.      else
14.        flag = 1
15.      If [pa,post] not in LN then
16.        If post not in LP then
17.          add(PE,post)
18.        else
19.          If post not in PE
20.            flag = 2
21.      If flag = 2 then
22.        add(PE,pa)
23.      If flag = 1 then
24.        If ant not in PE then
25.          add(PE,pa)
26.    k = k + 1
27.    Return PE

```

**End**

**Algorithm 1.** Contextual Error Detection. This algorithm receives the list of words to be corrected and the lexicons where the words are searching. It generate the words identified as errors, this is the input for the next stages of error correction.

5. **Potential corrections generation:** taking in count the POS-tag of the potential errors detected in the previous modules, it makes a candidate generation of corrections based in n-grams and morphological comparisons (skeleton and edit distance), seeking in the right lexicon according to the grammatical category of words.
6. **Correction:** it selects the best potential correction from the candidate words, this is the responsible for taking the right one. This selection is based in: a) n-gram analysis for identify the most probable combination of those candidates and the local context of word; b) minimum edit distance of words and candidates; c) skeleton comparisons; and d) size of words and word skeletons.
7. **Postprocessing:** once the correction has been done in text, the upper case, abbreviations and punctuation symbols are reintroduced and the final text is saved.



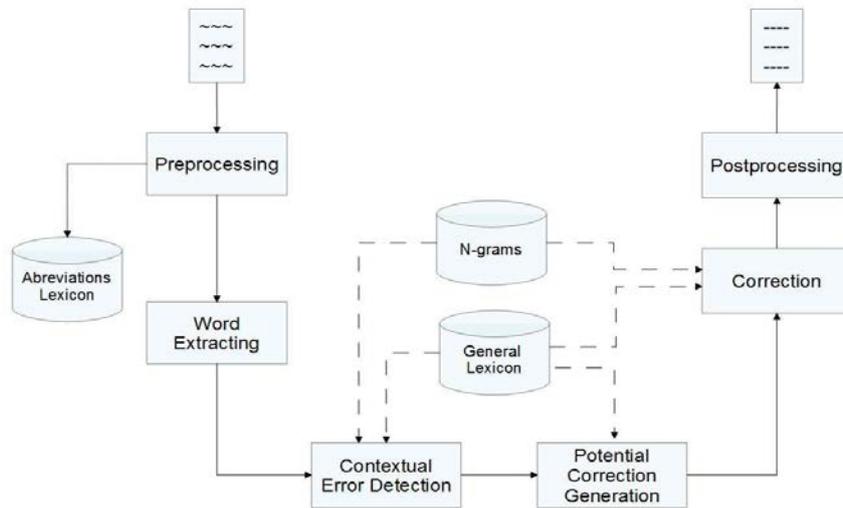
**Figure 3.** The PAEC proposal. The automatic error correction method is based in the identification of POS of the text and the model of language based in n-grams.

In the following section are discussed the implementation matters, and the experiments carried out in order to verify de efficacy of PAEC proposal.

#### 4 Implementation, experiments and results

The PAEC proposal was implemented using the PERL programming language, because it is relatively easy to construct text analyzer functions, string matching, and lexicon searching [14].

In order to test the PAEC proposal of error correction an alternative method of correction was implemented Morphological Automatic Error Correction (MAEC). This method does not include a phase of POS-tagging for the sake of represent a more traditional way to do the error correction in text, i.e. it contains less linguistic knowledge, specifically morphologic and syntactic information, so a framework of comparison can be established. The MAEC method is depicted in **Figure 4**.



**Figure 4.** The Morphologic Automatic Error Correction (MAEC) method, it uses a n-grams model of language and a General Lexicon for making the three stages of text error correction. It was used for testing the PAEC proposal of error correction.

#### 4.1 Data description

As it was mentioned in the previous section, many linguistic resources were used to implement the PAEC and MAEC methods of error correction in texts. They are described in the **Tables 1, 2, 3, 4, 5 and 6.**

**Table 1.** Plain Text Corpus characteristics. The texts were extracted from different Web sites written in Spanish from the following domains: Essay, Politics, Literature and Philosophy.

Language	Spanish
Format	ASCII
Files	497
Text lines	137,103
Words	2,877,834
Characters	18,075,061

**Table 2.** The POS-tagging manually annotated text corpus characteristics. This corpus was used for the initial training of the POS-tagger which is an iterative and supervised process, where results in a bigger (automatically) annotated corpus for being used in the error correction method. In the first stage it was used 0.1509% of the plain text corpus.

Language	Spanish
----------	---------

Format	ASCII
Files	1
Text lines	243
Words	4,346

In the **Table 3** is described the set of testing files, which contains a lot of errors, one set is obtained from an OCR (Hewlett Packard, ScanJet 3200c, their features can be seen in the **Table 4**) process and the other one was generated according to the following process (for simulating typing errors):

1. The file is formatted in order to keep 10 words per line.
  2. One word of each line is randomly selected
  3. From each word it is randomly selected one character
  4. In the position selected is randomly inserted one error (insertion, deletion or substitution)
  5. The modified word is reinserted in its corresponding position in the file.
- So, it is gotten one random error each ten lines.

**Table 3.** Text Corpus with errors used for testing the error correction methods. The *Errors 1* column shows the automatically inserted errors. The *Errors 2* refers to the error from the OCR process.

Texts	Lines	Strings	Characters	Errors 1	Errors 2
1	15	79	587	7	5
2	17	92	634	7	8
3	16	126	831	13	2
4	11	147	1068	14	9
5	14	150	985	15	4
6	19	260	1770	25	28
7	17	310	2061	30	10
8	20	358	2045	55	52
9	17	591	3800	116	29
10	19	1153	6774	41	15
<b>Total</b>	<b>165</b>	<b>3,266</b>	<b>20,555</b>	<b>323</b>	<b>162</b>

**Table 4.** Features of the scanner and the OCR software used to generate the *Error 2* set of files.

Scanner	HP ScanJet 3200c
ORC software	HP Precisionscan LT
Resolution	600 x 1200 dpi
Format of output	Plain text (.txt) or riched text (.rtf)

The set of lexicons are showed in the **Table 5**. The General Lexicon is used by the MAEC method, while the other lexicons (POS-based) are used by the proposal PAEC. The contents of each lexicon were taken from the Dictionary of Usual Spanish [12], and the structure of them is given by the orthographic form or each word and

their n-grams of letters. Word between one to two letters has just their monogram, but words of four letter or more has either bigram and trigram.

**Table 5.** The features of the lexicons used by the error correction methods. A = words total number. B = larger size of word. C = smaller size of word. D = media of size of words.

Lexicon	A	B	C	D
General	131,648	42	1	17
Verbs	10,206	30	2	11
Prepositions	32	13	1	17
Pronouns	83	8	2	5
Interjections	26	7	2	5
Conjunctions	38	20	1	8
Articles	9	3	2	3
Adjectives	5,508	26	2	14
Adverbs	1,516	23	2	12
Nouns	14,517	25	1	12

**Table 6.** Features of the N-grams based model of language: larger, smaller and media size of just bigram are showed.

Bigrams	566,521
Larger frequency	25,542
smaller frequency	1
Media frequency	4

## 4.2 Experiment Description

The experiment carried out for testing the MAEC and PAEC methods of error correction in text is composed of two evaluations:

- **Simulation of typing errors** (errors automatically inserted): the purpose was to identify the performance of insertion, deletion and substitution errors.
- **OCR errors in text:** this is a life real set of evaluation, the files comes from an OCR method after being scanned.

Additionally, for the PAEC proposal, one more test was made: manually POS-tagged text was tested in order to obtain an ideal tagging, and verify the efficacy of the method by means to avoid the possibility of getting some deviation because the precision of the automatic POS-tagger.

The following accounts were made in the texts from both methods:

C: corrected words. Real errors detected and corrected properly.

E: detected errors. Total of errors identifies by the methods.

e: non detected errors. Total of errors not identified by the methods.

I: introduced errors. Wrong substituted words due to the methods identify some false errors.

F: false error correction. The candidate correction selected for correcting the error detected was wrong, so the error persist.

o: original errors. They are the original errors presents in text.

On the other hand, the rate of errors was calculated using the following formula [66]:

$$c = p - i / p$$

Where c: rate of error correction; p: words of text; and i: errors of text (wrong words).

### 4.3 Results of experiment

The results of the testing of MAEC method of error correction are showed in the **Table 7** and the results of PAEC method are showed in **Table 8**. The results of the third experiment (manually POS-tagged files) are showed in **Table 9**.

**Table 7.** The results of the automatic error correction based in n-grams and morphologic analysis MAEC. The first row is the result of the typing simulation set of testing, while the second row is the result of the texts from the OCR.

Set of testing	p	o	i	C	E	e	I	F	c
Errors 1	3266	323	140	202	311	31	19	90	0.957
Errors 2	3185	162	104	86	180	10	38	56	0.967

**Table 8.** The results of the automatic error correction based in n-grams and morphologic analysis PAEC. The first row is the result of the typing simulation set of testing, while the second row is the result of the texts from the OCR.

Set of testing	p	o	i	C	E	e	I	F	C
Errors 1	3266	323	370	112	474	28	167	180	0.886
Errors 2	3185	162	229	84	304	8	152	70	0.928

**Table 9.** The results of the automatic error correction based in n-grams and morphologic analysis PAEC. *Errors 3* row is the test performed with manual POS-tagged files that come from the simulated typing errors, while the *Errors 4* are files that come from an OCR process.

Set of testing	p	o	i	C	E	e	I	F	c
Errors 3	3266	323	133	217	238	26	27	80	<b>0.959</b>
Errors 4	3266	162	74	104	186	8	33	45	<b>0.977</b>

### 4.4 Discussion

As can be seen in the results tables, in the experiments on the sets of files *Errors 1* and *Errors 2* the MAEC method is better than PAEC proposal. These results are discussed as follows.

The MAEC method has a high rate of no detected errors (*e*), this affects mainly to the selection of errors because it introduces new errors (*I*) originated by errors that originate valid Spanish words, as a consequence the wrong word is not identified as error but belongs to a n-gram with low frequency; so, another word would be treated as an error.

The most mismatched words are words of one or two letters: y, a, e, la, el, lo, no; because when a deletion occurs they disappears or become in valid words difficult to identify and the introduction or error is possible. On this way, the number  $F$  is high too, because the n-gram model can't offers information when an error appears in a low frequency n-gram, so when it is weighted by the module of selection of candidate for making the correction there is a low possibility to select the right one.

On the other hand, the PAEC method was affected by the POS-tagging in two aspects:

- In error detection: when a word is taken as an error but the POS-tag is not so precise, then it is searched in the wrong dictionaries; so,  $I$  rate is increased.
- In potential correction generation: as in error detection, the candidates are generated from a possible wrong dictionary. So, the candidate can be different from the right one.

Finally, in **Table 6** can be seen the result from the POS-tagged files. As can be appreciated, the results are improved with respect to both previous tests. On this way, the PAEC proposal improves its performance given these facts:

- In error detection: now it is possible to identify wrong words that originates no valid words in Spanish, because can be matched in the right dictionary according to their POS-tag. Additionally, the errors that originate valid words are detected because the POS-tag indicate a low frequency n-gram: a wrong word belongs to a different grammatical category due to the mistake, and it is identified.
- In potential correction generation: the candidate generation is performed by its morphologic features, the n-gram probability and its grammatical category adding by this means the morphologic and syntactic knowledge.
- In correction: due to the correct detection the right candidate is selected in most of the times, so the correction is right.

## 5 Conclusions and Future Work

According to the results presented in the previous section the PAEC method is better in error correction when a precise POS-tagging is performed on the text because the morphologic and syntactic knowledge is properly added to the process, so the rationale is the right one but there are some improvement to do; if it is not the case, there are a lot of mistakes in the entire correction process.

Additionally, in order to improve the performance of the PAEC method it is required to do:

- To explore the use of syntactic n-grams (sn-grams), which are an improved form of n-grams, constructed using paths in syntactic trees in order to reduce the noise given to presence of many arbitrary terms in the n-gram. The sn-grams can be applied to the three stages of the error correction process to calculate POS sn-grams and word sn-grams [13].

- To improve the precision of the POS-tagger:
  - To train the algorithm in a more intensive way, adding more annotated text to start the learning of tagging rules.
  - To develop an alternative way to tagging words not identified and increase the lexical files of the tagger.
  - To use more than one tagger in order to improve the rate of correct tagging by the weighting the results of them.
  - To assign more than one POS-tag to words, accordingly to the probability of tags.
- To improve the detection of errors for searching the potential errors in more than one grammatical category lexicon given the assignation of more than one POS-tag.
- To refine the lexicons in two ways:
  - Increase the words they contain, in order to extend its lexical coverage.
  - By means of a lemmatization process, so the index of word would be more flexible because it can contain not only the complete orthographic form.

As a final commentary is the method can be tested on real world errors and a larger amount of texts. They can be taken from:

- Contents of Web 2.0 pages.
- Government pages of complains, administrative or legal processes and opinion pages.
- A bank of OCR-processed documents from libraries and government offices.

## References

1. Kukich K. (1992). Techniques for automatically correcting words in text. *ACM Computing Survey*. Vol. 24, pp. 377--439.
2. Jurafsky D. & Martin J. H. (2000). *Speech and language processing. An introduction to natural language processing, computational linguistics and speech recognition*. USA: Prentice-Hall.
3. Jones M. P. & Martin J. (1997). Contextual spelling correction using latent semantic analysis. In *Proceedings of the Fifth Conference on Applied Natural Language Processing*. (pp. 166--173). USA: Washington DC.
4. Brill E. (1995). Unsupervised learning of disambiguation rules for part of speech tagging. In D. Yarowsky & K. Church (Eds.), *Proceedings of the third ACL Workshop on Very Large Corpora (WVLC-95)*. (pp. 1--13). Somerville, New Jersey: Association for Computational Linguistics.
5. Jin R., Hauptmann A. G., & Xiang Z. C. (2003, January). A content-based probabilistic correction model for ocr document retrieval. In T. Kanungo, E. H. Barney Smith, H. Jianying, P. B. Kantor (Eds.) *Document Recognition and Retrieval X (Proceedings of SPIE/IS&T)*. (pp. 128--135). Sta. Clara, California: The Society for Imaging Science and Technology.
6. Klein, S. T., & Kopel, M. (2002, august). A voting system for automatic OCR correction. In K. Järvelin (General Chair), *Information Retrieval and OCR: From Converting Con-*

- tent to Grasping Meaning. Workshop conducted at the 25th ACM SIGIR Conference, 2002, Tampere, Finland.
7. Taghva, K., & Stofsky, E. (2001). OCRSpell: An interactive spelling correction system for OCR errors in text. *International Journal on Document Analysis and Recognition*, 3(3), 125--137.
  8. Sproat, R., & Olive, J. (1998). Text-to-speech synthesis. In V. K. Madisetti, & D. B. Williams (Eds.), *The digital signal processing handbook* (pp. 46-1--46-11). Boca Raton, Florida: CRC Press.
  9. Alvarez Cabán, J. M. (2000). Alternatives for access and use of Spanish language assistive technology equipment by individuals with visual disabilities. In California State University Northridge (Comp.), *Proceedings of the Technology and Persons with Disabilities Conference, 2000*. Los Angeles: California State University Northridge. Retrieved from [http://www.csun.edu/cod/conf/2000/proceedings/0187Alvarez\\_Caban.html](http://www.csun.edu/cod/conf/2000/proceedings/0187Alvarez_Caban.html)
  10. Moreno Sandoval, A. (1998). *Lingüística Computacional*. Madrid: Síntesis.
  11. Manning C. D. & Schütze H. (2001). *Foundations of statistical natural language processing*. Cambridge, Massachusetts: The MIT Press.
  12. Lara Luis Fernando (s.f). *Diccionario del Español Usual en México*. Biblioteca Virtual Miguel de Cervantes. Available in : <http://bib.cervantesvirtual.com/servlet/SirveObras/06811650999196173088968/index.htm?na=26041>
  13. Grigori Sidorov, Francisco Velasquez, Efstathios Stamatatos, Alexander Gelbukh, and Liliana Chanona-Hernández. Syntactic Dependency-based N-grams as Classification Features. *LNAI 7629*, 2012, pp. 1-11.
  14. Luna Ramírez Wulfrano Arturo & Barrón Machado Jorge. (2004). Mejoras al reconocimiento óptico de caracteres y a la correccion de textos electronicos en los sistemas de lectura automatica de texto. Bsc. Thesis. Supervisor: Esmeralda Uraga Serratos. Classification: 001 00623 L1 2004. 366 pps. <http://bcct.unam.mx/web/tesiunam.htm> & [http://oreon.dgbiblio.unam.mx:8991/F/YMJ36QMYP4YQCSJFJHG7QS5S22MP3K9VREJM6RAP6QSSUYANAG-37691?func=service&doc\\_number=000334454&line\\_number=0007&service\\_type=TAG](http://oreon.dgbiblio.unam.mx:8991/F/YMJ36QMYP4YQCSJFJHG7QS5S22MP3K9VREJM6RAP6QSSUYANAG-37691?func=service&doc_number=000334454&line_number=0007&service_type=TAG)



# Recognizing Textual Entailment with Similarity Metrics

Miguel Rios<sup>1</sup> and Alexander Gelbukh<sup>2</sup>

<sup>1</sup> University of Wolverhampton,  
Research Group in Computational Linguistics,  
Stafford Street, Wolverhampton, WV1 1SB, UK  
M.Rios@wlv.ac.uk

<sup>2</sup> Center for Computing Research,  
National Polytechnic Institute,  
Mexico City, Mexico  
www.gelbukh.com

**Abstract.** We present a system for the Recognizing Textual Entailment task, based on various similarity metrics, namely (i) string-based metrics, (ii) chunk-based metric, (iii) named entities-based metric, and (iv) shallow semantic metric. We propose the chunk-based and named entities-based metrics to address limitations of previous syntactic and semantic-based metrics. We add the scores of the metrics as features for a machine learning algorithm. Then, we compare our results with related work. The performance of our system is comparable with the average performance of the Recognizing Textual Entailment challenges systems, though lower than that of the best existing methods. However, unlike more sophisticated methods, our method uses only a small number of simple features.

## 1 Introduction

The Recognizing Textual Entailment (RTE) task consists in deciding, given two textual expressions, whether the meaning of one of them, called Hypothesis (H), is entailed by the meaning of the other one, called Text (T) [5]. The RTE Challenge is a generic task which addresses common semantic inference needs across Natural Language Processing (NLP) applications.

In order to address the task of RTE, different methods have been proposed. Most of these methods rely on machine learning (ML) algorithms. For example, a baseline method proposed by Mehdad and Magnini [9] consists in measuring the word overlap between the Text and Hypothesis; the word overlap is the number of words shared between the two textual expressions. Their method is organized into three main steps: (i) pre-processing; all T–H pairs are tokenized and lemmatized; (ii) computing of the word overlap; (iii) building a binary classifier. An overlap threshold is computed over the training data, and the test data is classified based on the learned threshold. If the word overlap score is greater than the threshold, then the entailment decision is TRUE (there is entailment), otherwise

it is FALSE (there is no entailment). The motivation behind this paradigm is that a T–H pair with a strong similarity score has good chances to represent an entailment relation. Different types of similarity metrics are applied over the T–H pair in order to extract features and to train a classifier.

Similarity metrics that deal with semantics usually use information from ontologies or semantic representations given by parsers [2]. However, the comparison between texts is done by matching the semantic labels, and not by matching the content of those units.

In this work we describe an RTE system based on various similarity metrics. In addition, we propose new similarity metrics based on different representations of text for RTE that are: (i) chunks and (ii) Named Entities. The goal of the introduction of these new features is to address limitations of previous syntactic- and semantic-based metrics. We add the scores of the new metrics along with simple string-based similarity metrics and a shallow-semantic-based metric [11] as features for a machine learning method for RTE. Then, we compare our results with related work on RTE. The performance of our system is comparable with the average performance of the RTE challenges, though it is lower than that of the best known methods.

In the remainder of this paper we discuss the related work (Section 2), describe our RTE system (Section 3) and compare its performance with previous work (Section 4). Finally, we give conclusions and suggest some future work (Section 5).

## 2 Related Work

Burchardt et al. [2] introduced new features for RTE that involve deep linguistic analysis and shallow word overlap measure. Their method consists of three steps: first, they represent the T–H pair with the Frame Semantics (FS) and Lexical Functional Grammars (LFG) formalisms; this representation is similar to the Semantic Role Labeling. Then, they calculate a similarity score based on matching the LFG graphs, and finally make a statistical entailment decision. They used the RTE-2 and RTE-3 datasets as training data, and extracted 47 features from the deep and the shallow overlap. These features consist of combinations of predicates overlaps, grammatical functions match, and lexical overlaps.

The methods that use Semantic Role Labeling (SRL) for RTE use the annotation provided by a semantic parser to measure the similarity between texts. However, they only measure the similarity in terms of how many labels the two texts share (overlaps) and not in terms of the content marked with those labels.

Delmonte et al. [8] introduced semantic-mismatch features, such as locations, discourse markers, quantifiers, and antonyms. Their entailment decisions are based on applying rewards and penalties over the semantic similarity and shallow similarity scores. Later, Delmonte et al. [6] participated in the RTE-2 challenge with an enhanced version of their previous system. Their new system uses new features based on heuristics, such as Augmented Head Dependency Structures, grammatical relations, negations, and modal verbs.

Roth and Sammons [12] used semantic logical inferences for RTE, where the representation method is a Bag-of-Lexical-Items (BoLI). The BoLI relies in word overlap. It states that the entailment relation holds if the overlap score is above a certain threshold. An extended set of stopwords is used to select the most important concepts for the BoLI, such as auxiliary verbs, articles, exclamations, discourse markers, and words in WordNet. Also, in order to recognize relations in the T–H pairs, the system checks matchings between SRLs, and then applies a series of transformations over the semantic representations to make easier to determine the entailment. Their system uses the following transformation operations:

- *annotate*, which make some implicit property of the meaning of the sentence explicit;
- *simplify* and *transform*, which remove or alter some section of the text T in order to improve annotation accuracy or make it more similar to H;
- *compare*, which compares some elements of the two members of the entailment pair and assigns a score that correlates to how successfully those elements of the H can be subsumed by the T.

### 3 Experimental Design

The RTE task can be seen as a binary classification task, where the entailment relations are the classes. Then the RTE benchmark datasets can be used to train a classifier [4].

Our RTE system is based on a supervised machine learning algorithm. We train the machine learning algorithm with similarity scores computed over the T–H pairs extracted from different classes of metrics described below.

With these metrics we build a vector of similarity scores used as features to train a machine learning algorithm. We use the development datasets from the RTE 1 to 3 benchmark to train different ML algorithms, using their implementations from the WEKA toolset<sup>3</sup> without any parameter optimization. Then, we test the models with a tenfold cross-validation over the development datasets to decide which algorithm to use for the comparison against related work over the test datasets.

The metrics we used as follows.

#### 3.1 Lexical Metrics

We use the following string-based similarity metrics: precision, recall, and  $F_1$ :

$$\text{precision}(T, H) = \frac{|T \cap H|}{|H|} \quad (1)$$

$$\text{recall}(T, H) = \frac{|T \cap H|}{|T|} \quad (2)$$

<sup>3</sup> <http://www.cs.waikato.ac.nz/ml/weka/>

$$F_1(T, H) = 2 \times \frac{\text{precision}(T, H) \times \text{recall}(T, H)}{\text{precision}(T, H) + \text{recall}(T, H)} \quad (3)$$

As input for the metrics we use a bag-of-words (BoW) representation of the T–H pairs. However, we only use content words to compute the similarity score in the T–H pairs.

### 3.2 Chunking

Shallow parsing (or chunking) consists in tagging a text with syntactically correlated parts. This alternative to full parsing is more efficient and more robust. Chunks are non-overlapping regions of text that are sequences of constituents that form a group with a grammatical role (e.g. noun group). The motivation for introducing a chunking similarity metric consists in that a T–H pair with a similar syntax structure can hold an entailment relation. The chunking feature is defined as the average of the number of similar chunks (in the same order) in the T–H pair:

$$\text{chunking}(T, H) = \frac{1}{m} \sum_{n=1}^m \text{simChunk}(t_n, h_n), \quad (4)$$

where  $m$  is the number of chunks in  $T$ ,  $t_n$  is the  $n$ -th chunk tag and content in the same order, and  $\text{simChunk}(t_n, h_n) = 1$  if the content and annotation of the chunk are the same, and 0.5 if the content of the chunk is different but the chunk tag is still the same.

The following example shows how the chunking metric works. Consider:

T: *Along with chipmaker Intel , the companies include Sony Corp. , Microsoft Corp. , NNP Co. , IBM Corp. , Gateway Inc. and Nokia Corp.*

H: *Along with chip maker Intel , the companies include Sony , Microsoft , NNP , International Business Machines , Gateway , Nokia and others.*

First, for each chunk, this metric compares and scores the content of the tag: whether it is the same chunk group and whether it is the same order of chunks. Table 1 shows how this metric scores each chunk for the previous example.

Finally, the chunking metric (4) computes the individual scores and gives a final score of  $\text{chunking}(T, H) = 0.64$  for this example.

### 3.3 Named Entities

Named Entity Recognition (NER) is a task that identifies and classifies parts of a text into predefined classes such as *names of persons, organizations, locations, expressions of times, quantities, monetary values, percentages*, etc. For example, from the text: “Acme Corp bought a new...”, *Acme Corp* is identified as a named entity and classified as an *organization*.

**Table 1.** Example of partial scores given by the chunking metric

Tag	Content	Tag	Content	Score
PP	<i>Along</i>	PP	<i>Along</i>	1
PP	<i>with</i>	PP	<i>with</i>	1
NP	<i>chipmaker Intel</i>	NP	<i>chip maker Intel</i>	0.5
NP	<i>the companies</i>	NP	<i>the companies</i>	1
VP	<i>include</i>	VP	<i>include</i>	1
NP	<i>Sony Corp.</i>	NP	<i>Sony</i>	0.5
NP	<i>Microsoft Corp.</i>	NP	<i>Microsoft</i>	0.5
NP	<i>IBM Corp.</i>	NP	<i>International Business Machines</i>	0.5
NP	<i>Gateway Inc.</i>	NP	<i>Gateway</i>	0.5
NP	<i>Nokia Corp.</i>	NP	<i>Nokia and others.</i>	0.5

The motivation of a similarity measure based on NER is that the participants in H should be the same as those in T, and H should not include more participants in order to hold an entailment relation. The goal of the measure is to deal with synonymous entities.

Our approach for the NER similarity measure consists in the following: first, the named entities are grouped by type; then, the content of the same type of groups (e.g. *Scripps Hospital* is an organization) is compared using the cosine similarity equation. However, if the surface realizations of the same named entity in T and H are different, we retrieve words that share the same context as the named entity in question; the words are retrieved from Dekang Lin's thesaurus. Therefore, the cosine similarity equation will have more information than just the named entity.

For instance, consider the T–H pair from the previous example. The entity from T: *IBM Corp.* and the entity from H: *International Business Machines* have the same tag *organization*. Our metric groups them and adds words from the similarity thesaurus resulting in the following bag-of-words (BoW) representation:

T's entities: {*IBM Corp.*, ..., *Microsoft*, *Intel*, *Sun Microsystems*, *Motorola* / *Motorola*, *Hewlett-Packard* / *Hewlett-Packard*, *Novell*, *Apple Computer*, ...}

H's entities: {*International Business Machines*, ..., *Apple Computer*, *Yahoo*, *Microsoft*, *Alcoa*, ...}.

Finally, the metric computes the cosine similarity between these BoWs.

### 3.4 TINE

TINE [11] is an automatic metric based on the use of shallow semantics to align predicates and their respective arguments between a pair of sentences. The metric combines a lexical matching with a shallow semantic component to address adequacy for machine translation evaluation. The goal of this metric is to provide a flexible way of align shallow semantic representations (semantic role

labels) by using both the semantic structure of the sentence and the content of the semantic components.

A verb in the hypothesis H is aligned to a verb in the text T if they are related according to the following heuristics: (i) the two verbs share at least one class in VerbNet, or (ii) the pair of verbs holds a relation in VerbOcean.

For example, in VerbNet the verbs *spook* and *terrify* share the same class, namely, *amuse-31.1*, and in VerbOcean the verb *dress* is related to the verb *wear*.

The following example shows how the alignment of verbs and predicates is performed. Consider:

H: *The lack of snow discourages people from ordering ski stays in hotels and boarding houses.*

T: *The lack of snow is putting people off booking ski holidays in hotels and guest houses.*

Then, the algorithm proceeds with the following steps:

1. Extract verbs from H:  $V_H = \{\textit{discourages}, \textit{ordering}\}$
2. Extract verbs from T:  $V_T = \{\textit{putting}, \textit{booking}\}$
3. Similar verbs aligned with VerbNet (shared class get-13.5.1):  
 $V = \{(v_H = \textit{order}, v_T = \textit{book})\}$
4. Compare arguments of  $(v_H = \textit{order}, v_T = \textit{book})$ :  
 $A_H = \{A0, A1, AM-LOC\}$   
 $A_T = \{A0, A1, AM-LOC\}$
5.  $A_H \cap A_T = \{A0, A1, AM-LOC\}$
6. Exact matches:  
 $H_{A0} = \{\textit{people}\}$  and  $T_{A0} = \{\textit{people}\}$
7. Different word forms:  
 expand the representation:  
 $H_{A1} = \{\textit{ski}, \textit{stays}\}$  and  $T_{A1} = \{\textit{ski}, \textit{holidays}\}$   
 to:  
 $H_{A1} = \{\{\textit{ski}\}, \{\textit{stays}, \textit{remain}, \dots, \textit{journey}, \dots\}\}$   
 $T_{A1} = \{\{\textit{ski}\}, \{\textit{holidays}, \textit{vacations}, \textit{trips}, \dots, \textit{journey}, \dots\}\}$
8. Similarly with  $H_{AM-LOC}$  and  $T_{AM-LOC}$

Here,  $V_H$  is the set of verbs in the hypothesis H,  $V_T$  is the set of verbs in the text T,  $A_H$  is the set of arguments of the hypothesis H, and  $A_T$  is the set of arguments in the text T.

The metric aligns similar verbs with the ontology and similar arguments with a distributional thesaurus. Then, the metric computes a similarity score given the previous alignment points.

## 4 Experimental Results

We compared our method with other machine learning-based methods and with methods that use a SRL representation as one of its features.

**Table 2.** The 10-fold cross-validation accuracy results over the RTE development datasets

Algorithm	RTE-1	RTE-2	RTE-3
SVM	<b>64.90%</b>	<b>59.00%</b>	<b>66.62%</b>
NaïveBayes	62.25%	58.25%	64.50%
AdaBoost	<b>64.90%</b>	57.75%	62.75%
BayesNet	64.19%	<b>59.00%</b>	65.25%
LogitBoost	62.25%	52.50%	61.00%
MultiBoostAB	64.55%	<b>60.50%</b>	64.00%
RBFNetwork	61.90%	54.25%	64.80%
VotedPerceptron	63.31%	57.75%	65.80%

We used the RTE-1, RTE-2, and RTE-3 development datasets to train the classifiers. Table 2 shows the tenfold cross-validation results.

The SVM achieved the best results in the experiments during the training phase. We use this algorithm to perform the classification over the RTE test datasets. The data used for classification are the test datasets of the RTE challenge. The experimental results are summarized in Table 3.

**Table 3.** Comparison with previous accuracy results over the RTE test datasets

Method	RTE-1	RTE-2	RTE-3
Roth and Sammons [12]	–	–	65.56%
Burchardt and Frank [1], Burchardt et al. [2]	54.6%	59.8%	62.62%
Delmonte et al. [8], [6], [7]	59.25%	54.75%	58.75%
Our method with SVM	53.87%	55.37%	61.75%

Table 4 shows the overall accuracy results of the RTE systems on the RTE test datasets against our method. Our method is close to the average performance but below the best method.

However, the systems that showed the best results in the RTE challenge are complex and sophisticated systems. In contrast, our method relies on a small number of simple features. Our main semantic feature is focused in predicate-argument information, while other methods tackle several semantic phenomena such as negation and discourse information [12] or rely on a large number of features [2].

**Table 4.** Comparison with overall accuracy results over the RTE test datasets

Challenge	Our method	Average	Best
RTE-1	53.87%	55.12%	70.00%
RTE-2	55.37%	58.62%	75.38%
RTE-3	61.75%	61.14%	80.00%

Error analysis shows that the most common source of errors for our method is the TINE similarity metric. The following categories of errors made by this metric are the most common ones:

1. Lack of coverage in the ontologies, for example:

T: *This year, women were awarded the Nobel Prize in all fields except physics.*

H: *This year the women received the Nobel prizes in all categories less physical.*

The lack of coverage in the VerbNet ontology prevented the detection of the similarity between *receive* and *award*.

2. Matching of unrelated verbs, for example:

T: *If snow falls on the slopes this week, Christmas will sell out too, says Schiefert.*

H: *If the roads remain snowfall during the week, the dates of Christmas will dry up, said Schiefert.*

In VerbOcean, *remain* and *say* are incorrectly indicated to be related. The VerbOcean dictionary was created by a semi-automatic extraction algorithm [3] with an average accuracy of 65.5% and thus contain a considerable number of errors.

3. Incorrect tagging of the semantic roles by the semantic parser SENNA<sup>4</sup>, for example:

T: *Colder weather is forecast for Thursday, so if anything falls, it should be snow.*

H: *On Thursday, must fall temperatures and, if there is rain, in the mountains should.*

The position of the predicates affects the SRL tagging. The predicate *fall* has the roles (A1, V, and S-A1) in the reference, and the roles (AM-ADV, A0, AM-MOD, and AM-DIS) in the hypothesis H. As a consequence, the metric cannot match the fillers. Also, SRL systems do not detect phrasal verbs: e.g., the action *putting people off* is similar to *discourages* but current SRL systems do not detect this.

As we see, the quality of the semantic parser and the coverage of the ontologies are significant causes that affect the performance of our method.

In addition, on the RTE-1 test dataset with 800 T-H pairs, the coverage of the semantic metric is 491 pairs. This means that the system only predicts a certain amount of pairs. On the RTE-3 dataset, on which we obtain the best result, also has 800 T-H pairs, but the coverage on this dataset is much better: 556 pairs. Accordingly, our method has a smaller amount of errors due to a greater number of semantic-scored pairs.

<sup>4</sup> SENNA, <http://ml.nec-labs.com/senna/>

## 5 Conclusions and Future Work

We have presented a machine learning-based system for Recognizing Textual Entailment (RTE) task, based on new similarity metrics as well as simple string-based metrics and a shallow-semantic metric. The new similarity measures are based on: (i) chunking, (ii) named entities.

Our method has performance comparable with the average performance of methods in the RTE challenges. However, its performance is below that of the best know methods. On the other hand, our method relies on a small number of simple features, and our system only tackles one semantic phenomenon: predicate-argument information.

A preliminary error analysis shows that a main source of errors is the alignment of predicates by the TINE measure. However, if the system has more pairs tagged with predicate-argument information, then its performance improves.

In order to improve the performance of our current machine learning-based system, in our future work we will attempt to resolve the errors caused by the TINE metric based on the error analysis, or will use a different semantic approach to RTE [10].

Our semantic metric uses a distributional thesaurus to measure the similarity between arguments, so that, for example, *cat* and *dog* will be aligned because they share the same context. A possible direction to improve the semantic metric is to add hard constraints over the core arguments. These constrains can be defined as thresholds learned over the training dataset.

## Acknowledgments

This work was partially supported by the Mexican National Council for Science and Technology (CONACYT), scholarship reference 309261, and SIP-IPN grant 20121823.

## References

- [1] Burchardt, A., Frank, A.: Approaching textual entailment with LFG and FrameNet frames. In: Proceedings of the Second PASCAL Challenges Workshop on Recognising Textual Entailment. Venice, Italy (2006)
- [2] Burchardt, A., Reiter, N., Thater, S., Frank, A.: A semantic approach to textual entailment: System evaluation and task analysis. In: Proceedings of the ACL-PASCAL Workshop on Textual Entailment and Paraphrasing. pp. 10–15. Association for Computational Linguistics, Prague (June 2007)
- [3] Chklovski, T., Pantel, P.: VerbOcean: Mining the Web for Fine-Grained Semantic Verb Relations. In: Lin, D., Wu, D. (eds.) Proceedings of EMNLP 2004. pp. 33–40. Barcelona, Spain (Jul 2004)
- [4] Dagan, I., Dolan, B., Magnini, B., Roth, D.: Recognizing textual entailment: Rational, evaluation and approaches – erratum. *Natural Language Engineering* 16(1), 105 (2010)

- [5] Dagan, I., Glickman, O.: The PASCAL Recognising Textual Entailment challenge. In: In Proceedings of the PASCAL Challenges Workshop on Recognising Textual Entailment (2005)
- [6] Delmonte, R., Bristot, A., Boniforti, M.A.P., Tonelli, S.: Coping with semantic uncertainty with VENSES. In: Proceedings of the Second PASCAL Challenges Workshop on Recognising Textual Entailment. Venice, Italy (2006)
- [7] Delmonte, R., Bristot, A., Piccolino Boniforti, M.A., Tonelli, S.: Entailment and anaphora resolution in RTE 3. In: Proceedings of the ACL-PASCAL Workshop on Textual Entailment and Paraphrasing, pp. 48–53. Association for Computational Linguistics, Prague (June 2007)
- [8] Delmonte, R., Tonelli, S., Piccolino Boniforti, M.A., Bristot, A., Pianta, E.: VENSES – a linguistically-based system for semantic evaluation. In: In Proceedings of the PASCAL Challenges Workshop on Recognising Textual Entailment (2005)
- [9] Mehdad, Y., Magnini, B.: A word overlap baseline for the Recognizing Textual Entailment task (2009)
- [10] Pakray, P., Barman, U., Bandyopadhyay, S., Gelbukh, A.: A statistics-based semantic textual entailment system. *Lecture Notes in Artificial Intelligence* 7094, 267–276 (2011)
- [11] Rios, M., Aziz, W., Specia, L.: TINE: A metric to assess MT adequacy. In: Proceedings of the Sixth Workshop on Statistical Machine Translation, pp. 116–122. Association for Computational Linguistics, Edinburgh, Scotland (July 2011)
- [12] Roth, D., Sammons, M.: Semantic and logical inference model for textual entailment. In: Proceedings of the ACL-PASCAL Workshop on Textual Entailment and Paraphrasing, pp. 107–112. Association for Computational Linguistics, Prague (June 2007)

# **Neural Networks & Unconventional Computation**

---



# Image transform based on an alpha-beta convolution model

Antonio Alarcón-Paredes, Elías Ventura-Molina, Oleksiy Pogrebnyak, and Amadeo Argüelles-Cruz

Center for Computing Research, National Polytechnic Institute. México, D.F.  
aparedesb07@sagitario.cic.ipn.mx, eventura\_a12@sagitario.cic.ipn.mx,  
olek@cic.ipn.mx, jamadeo@cic.ipn.mx

**Abstract.** In this paper, it is presented a novel method based on an alpha-beta convolution model, to be used at transformation stage in an image compression system. This method takes the alpha-beta's associative memories theory and is applied to a set of images in grayscale. Since these associative memories are used for data with binary inputs and outputs, it is also presented a modification to the original alpha and beta operators in order to be applied directly to the the pixel values of an image. The proposed method is applied as a traditional convolution, with the difference that instead of making sums of products, there are performed maximum or minimum operations of the alphas and betas. The Shannon entropy is used to measure the amount of bits of information contained in the images. The traditional images transform usually do not provide any kind of compression and they also use complex operations. Therefore, this new method represents an advantage by offering a lower amount of entropy in the transformed image that in the original image by making use of simple operations such as addition, subtraction, minimums, and maximums.

## 1 Introduction

The imminent growth in the amount of the existing information has given the guideline to think on mathematical methods that help us to represent the information in a compact manner, reducing the number of bits used for its representation. For this reason, data compression systems and, in a particular case, image compression systems have been created. There exists lossless image compression systems, whose stages are: transformation and coding; as well as image compression systems, whose stages are: transformation, quantization and coding [5], [13]. Although there are a large number of image transforms, the most common one is the DCT (Discrete Cosine Transform). The DCT was proposed by Ahmed, Nataraj and Rao in 1974 [1]. It is a transform that is applied to blocks of pixels of an image, each block is usually constituted by 8 x 8 or 16 x 16 pixels, and it consists in a bijective function that maps one to one the image values allowing DCT to be reversible, thus there is no loss of information, but it does not compress the image neither [10], [17]. From the early 80s, the CCITT (Consultative

Committee for International Telegraphy and Telephony) and ISO (International Organization for Standardization) began to work together in order to develop an international standard for image compression, which was achieved in 1992 with the acronym JPEG (Joint Photographic Experts Group) [9], [20] that uses the DCT in its transformation stage. From this date, diverse modifications have been developed to the DCT for its fast implementation, such as Chen [4], who takes advantage of the symmetry of the cosine function to reduce the needed operations to implement the transform. Subsequently, Arai [3] develops the DCT only taking into consideration the real part of a DFT (Discrete Fourier transform) and using the FFT (Fast Fourier Transform) algorithm that was proposed by Winogard in [22]. In addition there have been new developments for the DCT which can be consulted in [12], [15], [16] and [24]. In Shannon information theory [18], the quantity of information is defined as a probabilistic process, taking the image as the information source; denoted by  $S$  with  $n$  elements  $s_1, s_2, \dots, s_n$  and  $i = 1, 2, \dots, n$ , its entropy is defined as:

$$H(S) = - \sum_i P(s_i) \log_2(P(s_i)) \quad (1)$$

for this reason, it is used the Shannon entropy as a measure to know the amount of bits which represent each image in this paper.

This paper is organized as follows: in section 2 it is shown the theoretical framework for the development of the new method, section 3 describes the proposed method. The results and conclusions are shown in section 4 and section 5, respectively.

## 2 Theoretical Support

This paper presents a new method based on a modification of the original algorithm of the alpha-beta associative memories [23], and focuses in the image transformation stage. This section shows the theory which is the base for the proposed method.

### 2.1 Associative memories

The AM (Associative memories) [8] are pattern recognition's algorithms, whose purpose is to recover full patterns from input patterns that could be altered. Input patterns are represented by column-vectors  $\mathbf{x}$  and output patterns by column-vector  $\mathbf{y}$ . For each input pattern  $\mathbf{x}$ , there is one and only one output pattern  $\mathbf{y}$ , forming an association as an ordered pair:  $(\mathbf{x}, \mathbf{y})$ . The set of the  $p$  pattern associations is named fundamental association set or simply fundamental set, with  $\mu = 1, 2, \dots, p$ , and it is represented as:

$$\{(\mathbf{x}^\mu, \mathbf{y}^\mu) \mid \mu = 1, 2, \dots, p\} \quad (2)$$

The operation of an AM is divided into two phases: the learning phase, where input patterns  $\mathbf{x}$  are associated with their corresponding output patterns  $\mathbf{y}$  to

generate the associative memory; and the recovery phase, in which we introduce a pattern  $\mathbf{x}$  as the input to the memory, and as result we expect to receive a corresponding pattern  $\mathbf{y}$  in the outcome. There are two types of AM which are classified according to the nature of their pattern: auto-associative memories and hetero-associative memories. The memory is auto-associative if it fulfills that  $\forall \mu, \mathbf{x}^\mu = \mathbf{y}^\mu$ , *i.e.* each input pattern is equal to its corresponding output pattern; on the other hand, the memory will be hetero-associative if it is true that  $\exists \mu \in \{1, 2, \dots, p\} | \mathbf{x}^\mu \neq \mathbf{y}^\mu$ , *i.e.* if there exists at least an input pattern different to its corresponding output pattern.

Traditional models of AM, such as Lernmatrix [19], Correlograph [21], as well as Linear Associator [2], [11] operate within the theory of artificial neuron model of McCulloch and Pitts [14]. That is, its operation is based on sums of products. In addition there are the morphological AM [6] and the alpha-beta AM [23] which are the only ones that are handled outside of this theory and instead of sums of products, they use maximums (or minimums) values of the sums, in the case of the morphological; and maximum (or minimum) of alphas and betas, in the case of the alpha-beta.

## 2.2 Convolution

On the other hand, in the theory of image processing, convolution is widely used. The convolution of an image is an operation in the spatial domain, *i.e.*, a method which operates directly on the value of the pixels in the image. To perform a convolution is required to define a mask, also called window, whose choice of values should be carefully taken; the mask is usually  $3px \times 3px$ . Convolution can be expressed as follows:

$$g(x, y) = T[f(x, y)] \quad (3)$$

where  $f(x, y)$  is the input image,  $g(x, y)$  is the output image and  $T$  is an operator in  $f$  that is defined on specific neighbors points to  $(x, y)$ ; we will place the mask on each of these points to make the necessary operations. These operations consist of multiplying the pixels in the neighborhood  $(x, y)$  with the coefficients of the mask, adding the results to obtain the response in the pixel  $(x, y)$  of the resulting image [7].

Note that in both the classical AM models, and the convolution of images there are used sums of products; in the case of the alpha-beta AM, to its operation the sum of products is changed to maximums (or minimums) of alpha and betas. For the development of this work, it was thought to exchange the traditional function of convolution for a new method that is implemented as a convolution based on maximum or minimum of alphas and betas.

## 2.3 Alpha and beta operations

The alpha-beta AM [23] uses maximums and minimums and two binary operations proposed specifically: alpha ( $\alpha$ ) and beta ( $\beta$ ), that establish the name

of alpha-beta AM. In order to make the definition of the binary operations alpha and beta, there should be specified the set  $A$  and the set  $B$  beforehand, as following:

$$A = \{0, 1\} \text{ and } B = \{0, 1, 2\} \tag{4}$$

Binary operation  $\alpha : A \times A \rightarrow B$  is defined as:

Table 1  $\alpha : A \times A \rightarrow B$

$x$	$y$	$\alpha(x, y)$
0	0	1
0	1	0
1	0	2
1	1	1

Binary operation  $\beta : B \times A \rightarrow A$  is defined as:

Table 2  $\beta : B \times A \rightarrow A$

$x$	$y$	$\beta(x, y)$
0	0	0
0	1	0
1	0	0
1	1	1
2	0	1
2	1	1

### 3 Proposed method

As can be shown, in *Table 1* and *Table 2*, the alpha and beta operators could only handle binary inputs and outputs, hence it was required to extend them in order to handle real-valued numbers and thus operate directly over the pixel value of an image.

The new operation  $\alpha_{\mathbb{R}} : \mathbb{R} \times \mathbb{R} \rightarrow \mathbb{R}$  is defined as

$$\alpha_{\mathbb{R}}(x, y) = x - y + 1 \tag{5}$$

The alpha-beta AM can be max type and min type; besides, the method proposed in this paper can operate both types: max or min. Thus, the operation  $\beta_{\mathbb{R}} : \mathbb{R} \times \mathbb{R} \rightarrow \mathbb{R}$  has an operation for each type of recovery, max ( $\beta_{\mathbb{R}}^{\vee}$ ) or min ( $\beta_{\mathbb{R}}^{\wedge}$ ), as follows:

$$\text{If } x = y \rightarrow \beta_{\mathbb{R}}(x, y) = 1 \tag{6}$$

$$\beta_{\mathbb{R}}^{\vee}(x, y) = y - |x| - 1 \tag{7}$$

$$\beta_{\mathbb{R}}^{\wedge}(x, y) = x - |y| - 1 \tag{8}$$

**Definition 1.** Let  $\mathbf{A} = [a_{ij}]$  be a matrix of size  $m \times n$  representing an image, and let  $\mathbf{sb} = [sb_{ij}]$  be a matrix  $d \times d$ -dimensional and represents an **image sub-block** of  $\mathbf{A}$ , such as:

$$sb_{ij} = a_{r_it_j} \quad (9)$$

where  $i, j = 1, 2, \dots, d$ ,  $r = 1, 2, \dots, m$ ,  $t = 1, 2, \dots, n$ , and  $sb_{ij} = a_{r_it_j}$  represents the pixel value given by coordinates  $(r + i, t + j)$ , where  $(r, t)$  and  $(r + d, t + d)$  are the beginning and the end of the sub-block, respectively.

**Definition 2.** Let  $\mathbf{A} = [a_{ij}]$  be a matrix of size  $m \times n$  that represents an image. The value denoted by  $\varepsilon$  is a value greater than the maximum value that could assume a pixel of an image, thus:

$$\varepsilon > \bigvee_{\forall i,j} a_{ij} \quad (10)$$

**Definition 3.** Let  $\mathbf{sb} = [sb_{ij}]$  be a matrix  $d \times d$ -dimensional and represents an image sub-block, and let  $\varepsilon > \bigvee_{\forall i,j} sb_{ij}$  be a value greater than the maximum value that could assume a pixel of the sub-block  $\mathbf{sb}$ . The **transformation mask** of max type, denoted as  $\mathbf{mt}^\vee = [mt_{ij}^\vee]_{h \times h}$ , is initialized to 0, except for its central pixel that assumes the value  $\varepsilon$ , when the max type ( $\beta_{\mathbb{R}}^\vee$ ) is used:

$$mt_{ij}^\vee = \begin{cases} \varepsilon & \text{central pixel} \\ 0 & \text{other case} \end{cases} \quad (11)$$

**Definition 4.** Let  $\mathbf{sb} = [sb_{ij}]$  be a matrix  $d \times d$ -dimensional and represents an image sub-block, and let  $\varepsilon > \bigvee_{\forall i,j} sb_{ij}$  be a value greater than the maximum value that could assume a pixel of the sub-block  $\mathbf{sb}$ . The **transformation mask** of min type, denoted as  $\mathbf{mt}^\wedge = [mt_{ij}^\wedge]_{h \times h}$ , is initialized to 0, except for its central pixel that assumes the value  $-\varepsilon$ , when the min type ( $\beta_{\mathbb{R}}^\wedge$ ) is used:

$$mt_{ij}^\wedge = \begin{cases} -\varepsilon & \text{central pixel} \\ 0 & \text{other case} \end{cases} \quad (12)$$

By means of simplicity, in the following definitions, the notation of the sub-block  $\mathbf{sb} = [sb_{ij}]$  and the transformation mask (either if is max or min)  $\mathbf{mt} = [mt_{ij}]$  as coordinates, thus,  $[sb_{ij}] = \mathbf{sb}(i, j)$  y  $[mt_{ij}] = \mathbf{mt}(i, j)$ .

**Definition 5.** Let  $\mathbf{sb} = [sb_{ij}]$  be a matrix of size  $d \times d$  representing an image sub-block, let  $\mathbf{mt}^\vee = [mt_{rt}^\vee]_{h \times h}$  be a transformation mask max type, and let  $\mathbf{t} = [t_{ij}]$  be the transformed sub-block; the **alpha max convolution** operation ( $*\alpha_{\max}(\mathbf{sb}, \mathbf{mt}^\vee)$ ), is expressed as:

$$\mathbf{t}(i, j) = *\alpha_{\max}(\mathbf{sb}, \mathbf{mt}^\vee) = \bigvee_{i=-a}^a \bigvee_{j=-b}^b \alpha_{\mathbb{R}}(\mathbf{sb}(i+r, j+t), \mathbf{mt}^\vee(r, t)) \quad (13)$$

where  $a = \frac{h-1}{2}$  and  $b = \frac{h-1}{2}$ .

**Definition 6.** Let  $\mathbf{sb} = [sb_{ij}]$  be a matrix of size  $d \times d$  representing an image sub-block, let  $\mathbf{mt}^\vee = [mt_{rt}^\vee]_{h \times h}$  be a transformation mask max type, and let  $\mathbf{t} = [t_{ij}]$  be the transformed sub-block; the **beta** min **convolution** operation  $(*\beta_{\min}(\mathbf{sb}, \mathbf{mt}^\vee))$ , is expressed as:

$$\mathbf{t}(i, j) = *\beta_{\min}(\mathbf{sb}, \mathbf{mt}^\vee) = \bigwedge_{i=-a}^a \bigwedge_{j=-b}^b \beta_{\mathbb{R}}^\wedge(\mathbf{sb}(i+r, j+t), \mathbf{mt}^\vee(r, t)) \quad (14)$$

where  $a = \frac{h-1}{2}$  and  $b = \frac{h-1}{2}$ .

**Definition 7.** Let  $\mathbf{sb} = [sb_{ij}]$  be a matrix of size  $d \times d$  representing an image sub-block, let  $\mathbf{mt}^\vee = [mt_{rt}^\vee]_{h \times h}$  be a transformation mask max type, and let  $\mathbf{t} = [t_{ij}]$  be the transformed sub-block; the **alpha** min **convolution** operation  $(*\alpha_{\min}(\mathbf{sb}, \mathbf{mt}^\wedge))$ , is expressed as:

$$\mathbf{t}(i, j) = *\alpha_{\min}(\mathbf{sb}, \mathbf{mt}^\wedge) = \bigwedge_{i=-a}^a \bigwedge_{j=-b}^b \alpha_{\mathbb{R}}(\mathbf{sb}(i+r, j+t), \mathbf{mt}^\wedge(r, t)) \quad (15)$$

where  $a = \frac{h-1}{2}$  and  $b = \frac{h-1}{2}$ .

**Definition 8.** Let  $\mathbf{sb} = [sb_{ij}]$  be a matrix of size  $d \times d$  representing an image sub-block, let  $\mathbf{mt}^\vee = [mt_{rt}^\vee]_{h \times h}$  be a transformation mask max type, and let  $\mathbf{t} = [t_{ij}]$  be the transformed sub-block; the **beta** max **convolution** operation  $(*\beta_{\max}(\mathbf{sb}, \mathbf{mt}^\wedge))$ , is expressed as:

$$\mathbf{t}(i, j) = *\beta_{\max}(\mathbf{sb}, \mathbf{mt}^\wedge) = \bigvee_{i=-a}^a \bigvee_{j=-b}^b \beta_{\mathbb{R}}^\vee(\mathbf{sb}(i+r, j+t), \mathbf{mt}^\wedge(r, t)) \quad (16)$$

where  $a = \frac{h-1}{2}$  and  $b = \frac{h-1}{2}$ .

### Alpha-beta convolution transform algorithm

1. As usual on traditional image transforming methods, the alpha-beta convolution method proposed in here, is applied individually to an image sub-blocks of size  $d \times d$ . The image denoted as  $\mathbf{A} = [a_{ij}]_{m \times n}$  is divided into  $\kappa = (m/d) \cdot (n/d)$  sub-blocks  $\mathbf{sb}^\omega | \omega = 1, 2, \dots, \kappa$ , where  $m$  and  $n$  are the image height and width respectively,  $a_{ij}$  is the  $ij$ -th pixel of image with  $a \in \{0, 1, 2, \dots, L-1\}$  being  $L$  the number of bits that represents the value of a pixel.
2. Initialize to 0 all values in the resulting image  $\mathbf{T} = [t_{ij}]_{m \times n}$ .
3. Create the transforming mask  $\mathbf{mt}$  depending on the desired usage: max or min, according to the *Definition 2*, *Definition 3* and *Definition 4*.
4. Apply the *alpha* max (or min) *convolution* to each sub-block according to the *Definition 5* and *Definition 7* and place the outcome on resulting image  $\mathbf{T}$ .  $\square$

### Inverse alpha-beta convolution transform algorithm

1. The inverse alpha-beta convolution method, is applied individually to the sub-blocks of transformed image  $\mathbf{T}$ .
2. Initialize to 0 all values in the resulting recovered image  $\mathbf{A}' = [a'_{ij}]_{m \times n}$ .
3. Locate each sub-block of the transformed image  $\mathbf{T}$  and apply the *beta* min (or max) *convolution* according to the *Definition 8* and *Definition 6*, using **the same transforming mask** used in the alpha-beta convolution transform algorithm and place the outcome on recovered image  $\mathbf{A}'$ .  $\square$

Since each time the **mt** is created, the maximum value of the current sub-block is taken, it is required to store every maximum in a vector.

## 4 Results

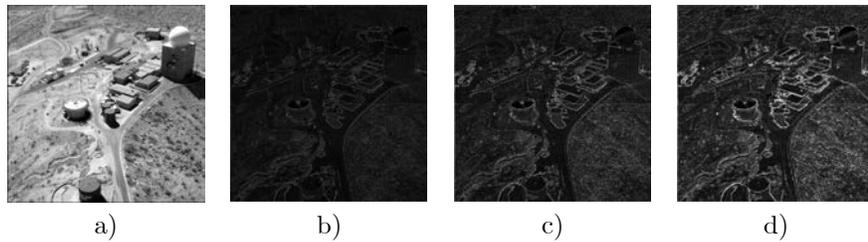


Fig. 1 Aerial image (a), with sub-block different sizes:  $4 \times 4$  (b),  $8 \times 8$  (c),  $16 \times 16$  (d).

In order to measure the proposed transform, a comparison between the original and the transformed image was performed using the Shannon entropy [18]. The method presented in this paper was applied to a set of 20 images widely used in many other papers, and was applied varying the size of **sb** ( $4 \times 4$ ,  $8 \times 8$ ,  $16 \times 16$ ) for each image, obtaining 60 transformed images. The set of testing images are grayscale, and 8 bits/pixel, and have different sizes.

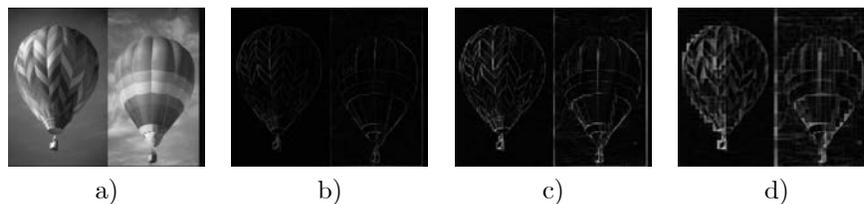


Fig. 2 Balloon image (a), with sub-block different sizes:  $4 \times 4$  (b),  $8 \times 8$  (c),  $16 \times 16$  (d).



Fig. 3 *Girl* image (a), with sub-block different sizes:  $4 \times 4$  (b),  $8 \times 8$  (c),  $16 \times 16$  (d).

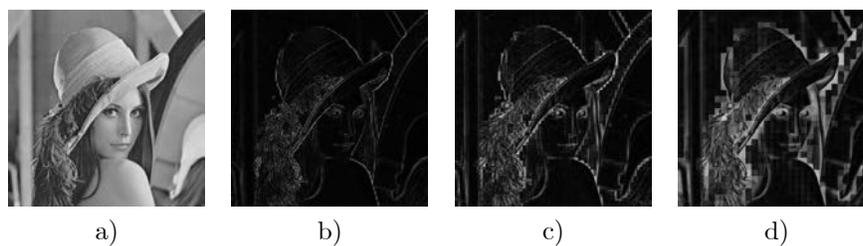


Fig. 4 *Lenna* image (a), with sub-block different sizes:  $4 \times 4$  (b),  $8 \times 8$  (c),  $16 \times 16$  (d).

Figures 1 to 6 show the original image (a) and some transformed images using  $4 \times 4$  sub-block size at (b),  $8 \times 8$  sub-block size at (c), and  $16 \times 16$  sub-block size at (d). The results that compare the entropy between the original image versus the transformed images are shown at *Table 3*.

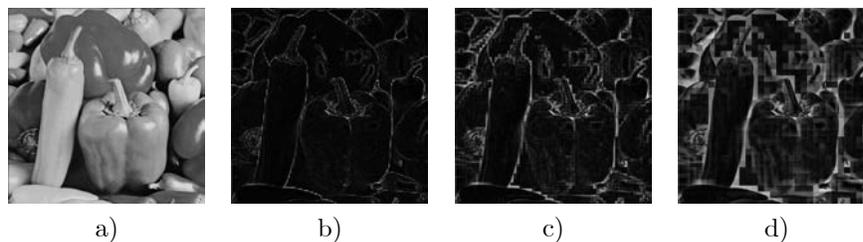


Fig. 5 *Peppers* image (a), with sub-block different sizes:  $4 \times 4$  (b),  $8 \times 8$  (c),  $16 \times 16$  (d).



Fig. 6 Zelda image (a), with sub-block different sizes:  $4 \times 4$  (b),  $8 \times 8$  (c),  $16 \times 16$  (d).

Table 3. Entropy results with different **sb** sizes.

Image	Size	Entropy	Entropy $4 \times 4$	Entropy $8 \times 8$	Entropy $16 \times 16$
Aerial	$2048 \times 2048$	7.1947	5.7846	6.4473	6.8813
Baboon	$512 \times 512$	7.3577	6.4610	6.9275	7.1501
Baloon	$720 \times 576$	7.3459	3.8442	4.7928	5.6843
Barbara	$720 \times 576$	7.4838	5.7309	6.3724	6.8072
Bike	$2048 \times 2560$	7.0219	5.3234	6.0152	6.5303
Board	$720 \times 576$	6.8280	4.5892	5.3763	6.2891
Boats	$512 \times 512$	7.1914	5.5520	6.2688	6.8104
Couple	$512 \times 512$	7.2010	5.3092	6.0276	6.5776
Elaine	$512 \times 512$	7.5060	5.1970	5.8841	6.5101
F-16	$512 \times 512$	6.7043	4.8730	5.5588	6.0325
Girl	$720 \times 576$	7.2878	5.0300	5.8262	6.4912
Goldhill	$720 \times 576$	7.5300	5.3532	6.1223	6.6593
Hotel	$720 \times 576$	7.5461	5.4036	6.2258	6.9116
Lenna	$512 \times 512$	7.4474	5.1027	5.8879	6.5444
Man	$512 \times 512$	7.1926	5.5016	6.3273	6.9066
Peppers	$512 \times 512$	7.5943	5.1029	5.9029	6.662
Sailboat	$512 \times 512$	7.4847	5.7920	6.5942	7.1399
Tiffany	$512 \times 512$	6.6002	4.7339	5.3748	5.8841
Woman	$2048 \times 2560$	7.2515	5.4033	6.0797	6.5413
Zelda	$720 \times 576$	7.3335	4.5692	5.3577	6.1714

The data contained in the table refers to the image dimensions in pixels, entropy of the original image (Entropy), the entropy of the transformed image using a **mt** of  $4 \times 4$  pixels (Entropy  $4 \times 4$ ), the entropy of the transformed image using a **mt** of  $8 \times 8$  pixels (Entropy  $8 \times 8$ ), and the entropy of the transformed image using a **mt** of  $16 \times 16$  pixels (Entropy  $16 \times 16$ ). The transformation stage of an image compressor does not provide any information reduction, its main purpose is to make easier the compression at following steps (quantization and coding). Since the alpha-beta convolution transform provides information reduction, it is clear that represents an advantage over the traditional transforming methods, besides the method proposed in this paper uses very simple operations such as addition, subtraction, minimums and maximums (comparisons).

## 5 Conclusions

The modified alpha and beta operators are capable to handle real valued inputs and may be used as well in image processing, and it is clear the next step could be the use of modified alpha-beta associative memories to perform the image transform.

By replacing the sums of products in traditional convolution with the maximums or minimums of alpha and beta, was possible to create an alternative for image transform in an image compression system, offering low computational cost by means of using simple operations.

The experimental results show that the alpha-beta convolution transform is reversible, that is, this new method has no information loss. Although 8 definitions were presented, it is needed the proposition and demonstration of some lemmas or theorems is required in order to formally prove this method is reversible.

As mentioned, the transformation stage in an image compression system does not provide any compression to images, so since the method proposed in this paper offers a smaller entropy on the transformed images, it also represents an advantage for the quantization and coding stages in an image compression system.

## References

1. Ahmed, N., Natarajan, T., Rao, K. R.: Discrete cosine transform. *IEEE Transactions on computers*. **23** (1974) 90-93
2. Anderson, J. A.: A simple neural network generating an interactive memory, *Mathematical Biosciences*. **14** (1972) 197-220
3. Arai, Y., Agui, T., Nakaajima, M.: A fast DCT-SQ scheme for image. *Transactions of the IEICE*. **71** (1988) 1095-1097
4. Chen, W., H., Smith, C., H., Fralick, S. C.: A fast computational algorithm for the discrete cosine transform. *IEEE Transactions on Communications*. **25** (1977) 1004-1009
5. Cosman, P., Gray, R., Olshen, R.: *Handbook of medical imaging processing and analysis*. Academic Press. USA. (2000) ISBN 0-12-077790-8
6. Díaz de León Santiago, J.L. & Yáñez Márquez, C.: *Memorias Morfológicas Heteroasociativas*. IT-57, Serie Verde, ISBN 970-18-6697-5, CIC-IPN, México (2001)
7. González, R. C., Woods, R. E. & Eddins, S. L.: *Digital Image Processing using MATLAB*. Prentice Hall. (2003). ISBN 0130085197
8. Hassoun, M. H.: *Associative Neural Memories*. Oxford University Press, New York. (1993)
9. ISO/IEC IS 10918-1 | CCITT T.81: Digital Compression and Coding of Continuous-Tone Still Image. (1992) ISO/IEC
10. Jain, A. K.: A sinusoidal family of unitary transforms. *IEEE Transaction on Pattern Analysis and Machine Intelligence*. **4** (1979) 356-365
11. Kohonen, T.: Correlation matrix memories, *IEEE Transactions on Computers*, C-21, **4** (1972) 353-359

12. Kok, C. W.: Fast algorithm for computing discrete cosine transform. *IEEE Transactions on Signal Processing*. **45** (1997) 757-760
13. Lakac, R., Plataniotis, K. N.: *Color image processing methods and applications*. CRC Press, Taylor & Francis. USA. (2007) ISBN 978-0-8493-9774-5
14. McCulloch, W. & Pitts, W.: A logical calculus of the ideas immanent in nervous activity. *Bulletin of Mathematical Biophysics*. **5** (1943) 115-133
15. Pan, W.: A fast 2-D DCT algorithm via distributed arithmetic optimization. *IEEE Proceeding Image Processing*. **3** (2000) 114-117
16. Ramírez, J., García, A., Frenadse, P. G., Parrilla, L., Lloris, A.: A new architecture to compute the discrete cosine transform using the quadratic residue number system. *IEEE Int. Symposium on Circuits and Systems*. (2000)
17. Rao, K. R., Yip, P., C.: *The transform and data compression handbook*. The Electrical Engineering and Signal Processing Series. (2001) ISBN 0-8493-3692-9
18. Shannon, C. E.: A mathematical theory of communication. *Bell system technical journal*. **27** (1948) 379-423 y 623-656
19. Steinbuch, K. V.: Die Lernmatrix. *Kybernetik*, **1**, **1** (1961) 36-45
20. Wallace, G. K.: The JPEG still picture compression standard. *Proceedings of Communications of the ACM*. **34** (1991) 30-44
21. Willshaw, D., Buneman, O. & Longuet-Higgins, H.: Non-holographic associative memory, *Nature*. **222** (1969) 960-962
22. Winograd, S.: On computing the discrete Fourier transform. *Proceedings of the National Academy of Sciences of the United States of America*. **73** (1976) 1005-1006
23. Yáñez, C.: *Associative memories based on order relations and binary operators* (in Spanish). Doctoral Theses. Center for Computing Research. (2002)
24. Yu, S., Swartzlander, E. E.: DCT implementation with distributed arithmetic. *IEEE Transactions on Computers*. **50** (2001) 985-991



# Security Token for Web Bank Applications Using a Linear and Congruential Random Number Generator

Luis Orantes<sup>1</sup>, Marco Ramírez<sup>2</sup>, Pablo Manrique<sup>2</sup>, Victor Ponce<sup>2</sup>, Aniceto Orantes<sup>3</sup>, Victor Salazar<sup>3</sup>, Antonio Montes<sup>3</sup>, Carlos Hernández<sup>4</sup>, Eric Gómez<sup>5</sup>

<sup>1</sup>Center for Research in Computing, Av. Juan de Dios Bátiz, Mexico City 07738, Mexico  
lorantesg1101@alumno.ipn.mx

<sup>2</sup>Center for Research in Computing, Av. Juan de Dios Bátiz, Mexico City 07738, Mexico  
{mars, pmanriq, vponce}@cic.ipn.mx

<sup>3</sup>HighBits, Av. Central Poniente #847 Int. 3, Tuxtla Gutiérrez, Chiapas 29000, Mexico  
{aorantes, vsalazar, antonio}@highbits.com

<sup>4</sup>Unidad Profesional Interdisciplinaria en Ingeniería y Tecnologías Avanzadas, Av. Instituto Politécnico Nacional 2580, Mexico City 07340, Mexico  
carlos@highsecret.com

<sup>5</sup>Escuela Superior de Ingeniería Mecánica y Eléctrica Unidad Zacatenco, Av. Instituto Politécnico Nacional S/N. Unidad Profesional Adolfo López Mateos. Mexico City 07738, Mexico  
ergomez@ipn.mx

**Abstract.** This paper presents a new algorithm for using with an one-time password security token; the objective is to provide security for the authentication of customers using bank websites even in the cases when the user has been the victim of a phishing or spyware attack and their bank account secret password has been stolen. For the token's performance, the algorithm make use of a Linear and Congruential Random Number Generator (LCG) (for a better understanding of the presented algorithm a short introduction to this arena is given), and an exhaustive algorithm for the validation of the one-time password keys is presented. This paper shows that the present algorithm is easy to implement and safer than a competing algorithm widely used in today's security tokens.

**Keywords:** Security token, cryptography, random number generators.

## 1 Introduction

Bank institutions have modernized their operations by allowing their customers to perform almost any account transaction using the Internet. One of these operations has been to transfer funds electronically to other bank accounts, bringing with this a profound danger. The user may be a victim of phishing or may have, without knowing it, installed spyware on their computer. Thieves use the user's stolen passwords to empty the user's bank account and transferring the funds to a ghost bank account for later withdrawal.

One solution that has been given for this problem is the use of one-time password generators (security tokens), which create a password that is valid for bank access just once. With this technique, if the password is stolen it is rendered useless for accessing the bank website. In addition to the conventional password the bank website will request the one-time password which if stolen it's useless because it will be already expired just right after the legitimate user introduced it to the bank website. In other words, the token system proves that the user is who claims to be and acts as an electronic key to access the bank website services, offering security to the user even in cases where their password has been stolen.

A security token is a hardware device usually with a LCD screen (this kind of security token doesn't need to be connected to the computer when used) or provided with a USB plug (this kind of security token needs to be connected to the computer when used). Regarding the kind of approach they are used, security tokens can be classified into several categories with some of the most common approaches: 1) one-time passwords, 2) time-synchronized passwords and 3) challenge/response passwords. In this paper, the algorithm presented belongs to the one-time password category.

It has been done plenty of research about token security but in this paper just one reference will be done, just to the most common security token used nowadays. This security token is the secureID token developed by RSA Security which uses a 64 bit secret key for a hash function called Alleged SecureID Hash Function (ASHF). In [14] it has been shown that the core of this security token can be broken in a few milliseconds and they conclude that it doesn't provide the security demanded by institutions nowadays including banks. In contrast, the algorithm presented in this paper offer much higher security just as it is (with a 64 bits secret key) but it may be virtually unbreakable resizing the presented algorithm to a larger key size as explained later.

The algorithm presented in this paper uses a Pseudo Random Number Generator (RNG) as the encryption mechanism which is necessary for the generation of the one-time passwords. The following section has a brief explanation of what a RNG is because randomness is the core of the presented algorithm and a basic background on this topic is required for a better understanding of it.

## **2 Random Number Generators (RNG)**

Computers can't be random. What computers can do is to simulate a random process by using a RNG. A RNG is an equation, which can generate a sequence of pseudorandom numbers. This sequence of pseudorandom numbers is finite and, after a certain quantity of pseudorandom numbers created, the sequence is repeated again in a cyclic way. The length of the cycle is called the RNG period and this is given, in the best cases, by the number of bits of the RNG and it's limited by the bit number of

the mathematical operations that can be computed. In ordinary PCs it is only 32 bits, nevertheless it is possible to simulate 64 bits in C++ simply by defining a long long variable type or even larger number of bits using a long int library.

There are several random number generators; the generator is selected depending on the type of a specific. There are several kinds of applications for a RNG; the most typical applications are: simulations (e.g., of physical systems to be simulated with the Monte Carlo method), cryptography and procedural generation. For the selection of the appropriate random number generator it should be taken on account not just the quality of the random numbers the generator creates but also the complexity of the generator.

For instance, for the particular application presented, it should be taken on account that this algorithm may be implemented in hardware devices (e.g. a microcontroller). These kind of devices may have limited resources such as computing capabilities and battery life. Another factor to take on account in the generator selection is energy consumption because we want the token to last as much as possible (at least a couple of years). If the generator is too complex (i.e. computationally speaking too expensive) the token life would last just a few months. A token with a short life span would be too impractical for being used in real life applications. Also it is necessary the token to generate the one-time passwords instantaneously and for achieving this goal a random number generator quick to compute is mandatory.

On one hand there are very quick generators such as xorshift [1] [2] [3] which in some cases may generate a full period but they generate very low quality random sequences. The Linear feedback shift register RNG (LFSR) [4] is a popular generator which in the past has been implemented in hardware [5] but it doesn't generate a full period. This is a very popular generator which has been implemented in several applications; important LFSR-based stream ciphers include A5/1 and A5/2, used in GSM cell phones, E0, used in Bluetooth, and the shrinking generator. Nevertheless its drawbacks were revealed when the A5/2 cipher has been broken and both A5/1 and E0 have serious weaknesses [6] [7].

On the other hand there are very high quality random number generators such as Blum Blum Shub [8] [9], Yarrow algorithm [10] (incorporated in Mac OS X and FreeBSD), Fortuna [11] [12] and CryptGenRandom [13] (incorporated in Windows) that have the inconvenient of being computationally speaking too expensive for the purpose of this algorithm.

## **2.1 The selected random number generator**

The selected random number generator for being incorporated in the algorithm presented in this paper is the Linear and Congruential Generator". This RNG was selected by its satisfactory quality; also it has the advantage that it computes the random

sequences very quickly. Because of its low complexity this RNG is suitable for being implemented in hardware applications with limited resources (e.g. a microcontroller).

A sequence of random numbers is obtained by evaluating the following equation:

$$Z_i = (Z_{i-1} + a) \cdot c \pmod{m} \quad (1)$$

This RNG requires a seed  $Z_0$  which is the initial state of the RNG; this can be seen as the index or initial point of the random table and it will be the first value of the random table. The next value of the random table is calculated by replacing  $Z_{i-1}$  with the new obtained value and this process is repeated again. In other words,  $Z_{i-1}$  is the value of the previously computed random number.

$a$ ,  $c$  and  $m$  are constants, nevertheless these values of these constants need to be chosen carefully because the quality of the random table depends on them. There are values for these constants that generate a poor quality random table and others that don't generate any random table at all. The value of  $m$  gives the cycle size of the random table, that is to say, from which number the random table would repeat again. It is desirable to have as many random numbers as possible; therefore the value of  $m$  usually is as big as the maximum value it can be computed.

Multiple researches have been conducted to find out what the values of these constants are the best. There are several approaches for determining the quality of the pseudorandom number sequence generated by a given constant values, for instance in [15] this quality is examined by scatter plots and spectral test but there are many other techniques for determining the quality of a pseudorandom sequence and if this quality is acceptable enough. The generator, "Linear and Congruential Generator" is especially very sensitive to the choice of these constants values and in the past have been poor choices for the values of these constant with not good results [16].

These constants need to meet some conditions; for instance to guarantee a generation of a full period for any seed values when having a non-zero value for  $c$ ; they need to meet the following conditions [15]:

1.  $c$  and  $m$  must be relative primes,
2.  $a-1$  must be divisible by all prime factors of  $m$ ,
3.  $a-1$  must be a multiple of 4 if  $m$  is a multiple of 4.

It is recommended for a 64-bit variable such as an ordinary PC can simulate by defining a long long variable type in C++ to be the next:

$$\begin{aligned} a &= 6364136223846793005 \\ c &= 1442695040888963407 \\ m &= 2^{64} \end{aligned}$$

### 3 Algorithm for the generation of one-time passwords on the token side

Equation 1 is capable of generating a pseudo-random table; nevertheless it is necessary to have a unique random table for every token that is constructed. It is desirable to have a different sequence of numbers for every token. It is possible to achieve this by encrypting equation 1 by performing a XOR operation of the generated random number with a secret key  $K$ . After this, it is necessary to resize the encrypted sequence to values that can fit within the eight digits of a LCD display (i.e. values that range from  $0..99,999,999 = 9^8$ ) This is done by performing a modulus operation of the calculated  $Z_i$  value with  $9^8+1$ . With this the one-time passwords will range from 0 to  $9^8$ . The equation then becomes:

$$Z_i = (KZ_{i-1} + c) \text{ mod } m \otimes \quad (2)$$

$$\text{OneTimePassword} = Z_i \text{ mod } (9^8 + 1) \quad (3)$$

The security token needs to be initialized, it is necessary to have a seed value for this and the same value of the secret key  $K$  as the seed for the token which will be the  $Z_i$  value. This value will be stored in the token’s memory and it will be used for computing the next one-time password by the security token and also at the server side to validate a one-time password. The first 1,000 one-time passwords are generated at the security token side for being wasted (this is done for allowing detection of 1,000 expired one-time password as explained later).

On the token side the value of  $K$  is obtained and is used as seed  $Z_0$  for evaluating Equation 2 and 3 and generating a one-time password. This equation is evaluated again to obtain another one-time password. In this way, every time the “generate a one-time password” button is pushed a pseudorandom encrypted sequence is obtained. See Table 1 for an example of this.

**Table 1.** Example of a sequence of numbers generated by the token

Counter	One-time password
1	82475249
2	82040631
3	72383201
4	68714439
5	32340945
6	88383319
7	94725313
8	10436071

#### 4 Algorithm for the one-time password validation at the server side

In order to validate a one-time password on the server side, it is necessary to decrypt the one-time password introduced by the user. However, a modulus operation can't be reversed (the resulting double modulus operation used in Equation 3 is even more irreversible). Therefore it is not possible to decrypt the one-time password introduced by the user as usual (i.e. performing the inverse operations); for this reason it is decrypted using an exhaustive algorithm.

At the server side, the one-time password is validated by computing  $N_{exp}$  (it's computed for 1,000) expired one-time passwords generated from the actual value of  $Z_i$  that is stored in the server using the algorithm previously explained (refer to section 3). These are previous valid one-time passwords that were introduced by the user but they already expired; this is done to give feedback to the user and the user may distinguish between an invalid or an expired one-time password (valid but expired; it was already introduced by the user to the bank website before). If this is found within the  $N_{exp}$  expired it comes to inform the user that introduced one-time password was valid in the past but it already expired (i.e. it won't be accepted as valid anymore).

After this,  $N_{val}$  (10,000) valid one-time passwords are generated using the same previous approach. These are the possible valid keys for the actual state of the token. If the one-time password introduced by the user is within this range the key is accepted as valid. The value of  $Z_i$  for the key that matched minus  $N_{exp}$  expired is stored in the server. This value will be necessary for computing the  $N_{exp}$  and  $N_{val}$  one-time passwords for a future validation. If no match is found it comes to reject the introduced one-time password by the user and the  $Z_i$  value stored at the server side remains unchanged.

One advantage of the algorithm presented in this paper is that a range of only  $N_{val}$  valid keys are accepted. One-time passwords that are not valid in one moment become valid in another. This is according to the actual value of  $Z_i$  stored at the present moment at the server side. This allows us to have a very low probability that a one-time password is accepted by the server as valid, which redounds to a highly secure system.

It is taken into account that the user may waste one-time passwords by generating and not using them (i.e. the user didn't introduced to the bank website a one-time password generated by the token); for this reason there is a range  $N$  valid. This range gives us a tolerance in the number of acceptable one-time passwords; it is necessary this value to be as small as possible because this will give us a lower probability that an attacker, in a random way, may guess a one-time password. This further contri-

butes to a higher level of security. Given that these one-time passwords are of eight numeric digits and have a range  $N_{val}$  (of 10,000 valid one-time passwords). This results in a probability of almost one in 100 million that a given one-time password will be accepted as valid. On the other hand, if the user wastes more than 10,000 keys by generating one-time passwords and not introducing it at the server, a desynchronization occurs and the token becomes useless; therefore the value for the range of  $N_{val}$  should be carefully considered.

Noteworthy this maximum value for wasted one-time passwords is reset between validations, which means that if the user has a considerable amount of one-time passwords wasted the range  $N_{val}$  is reset to 10,000 at the server side once a one-time password is accepted, giving the user the maximum number of one-time passwords that can be wasted again.

One advantage of this algorithm is that it isn't necessary to store the generated one-time passwords in the server. only the value of  $Z_i$  is stored in the server. For this reason, this algorithm doesn't waste space in the server for storing expired one-time passwords nor resources to determine if the one-time password introduced by the user is part of the expired ones. Perhaps this is not significant with one user but it's taken into account that a bank institution may have millions of clients then the saved space and resources becomes significant. The algorithm also has the advantage that once a one-time password has been accepted and validated, all the previous ones are automatically expired, even in cases where they haven't been introduced to the server.

## 5 Number of maximum generated one-time passwords

In spite of the fact, that this algorithm is able to compute eight digits range, it is not recommended to use the full range. It is possible that an attacker may be storing the one-time passwords that have been generated by the token as they are introduced to the website by the user. This leads to the hypothetical possibility that the probability of guessing a one-time password in an arbitrary way increases. This is because the attacker knows which one-time passwords already were used in the past and he wouldn't try them again. The recommendation is to cancel the token after a million one-time passwords are remaining within it. Thereby, the probability of guessing a one-time password goes from approximately 1-in-100 million to 1-in-one million. This calculation was done by assuming that there was an attacker storing the entire history of the generated one-time passwords by the token for years. This scenario is very impractical. This is a very theoretical scenario but this is done as an extra security measure.

## 6 Attacking the algorithm

The way to attack the presented algorithm is to have an attacker storing the entire one-time password introduced by the user in the bank web site. Let's say the attacker has two one-time passwords collected; to attack the algorithm a full search through the 64 bits generated random sequence is performed. This is done for every possible key  $K$  to find which key produce a sequence that matches those the attacker has. It may occur that there is more than one single  $K$  than match the sequence; if this is the case the attacker needs to wait for another one-time password. This will help to reduce the number of candidate secret keys  $K$  which match the stored sequence and could be the secret key. This process needs to be repeated again and again until it's found that only one key  $K$  match the sequence of one-time passwords the attacker has collected. If this is the case the secret key  $K$  has been for the token under attack. The following one-time passwords can be predicted for sure, therefore breaking the token security.

Performing this attack as the algorithm was presented in this paper (i.e. having a 64 bits Linear and Congruential Random Number Generator) is already computationally speaking a hard code to break with a PC because the attacker needs to try with  $2^{64}$  one-time passwords times the  $2^{64}$  possible keys resulting in  $2^{128}$  one-time passwords in total. Nevertheless it is possible to strengthen the algorithm and protect it from this kind of attack by using a larger number of bits for the RNG; (let's say 1024 bits for instance). The algorithm presented in this paper was of 64 bits; this is because 64 bits operations can be computed with ordinary PC instructions. As explained previously it is possible to simulate operations of larger number of bits using a long integer library (this library is commonly used for cryptographic applications). With this it is possible to assure that the algorithm is totally unbreakable.

## 7 Open Research Issues

Equation 2 is capable to generate multiple derived random sequences for a given values for the constants  $a$ ,  $c$  and  $m$ . A research needs to be conducted for determining if the derived random sequences are as random as the original one, or at least random enough. Another issue that is left for future research is related with the constants of the RNG. In this paper the algorithm was presented assuming a 64 bits RNG and it was also proposed to enlarge the bit number of the RNG to make the algorithm stronger. Nevertheless, a research needs to be conducted to find out the best constants values for  $a$  and  $c$  that generate the best random sequence for a 1024 bit  $m$  size.

To keep in touch with the research advances of this project please visit <http://www.highsecret.com> where related information will be posted continuously.

**Acknowledgments.** The first author acknowledges support from the Mexican Council of Science and Technology (CONACYT) to pursue MSc studies at CIC-IPN. The second author acknowledges National Polytechnic Institute of Mexico.

## Bibliography

1. Marsaglia, George (July 2003). "Xorshift RNGs". *Journal of Statistical Software*
2. Brent, Richard P. (August 2004). "Note on Marsaglia's Xorshift Random Number Generators". *Journal of Statistical Software*
3. Panneton, François (October 2005). "On the xorshift random number generators". *ACM Transactions on Modeling and Computer Simulation (TOMACS)*
4. M. Goresky and A. Klapper, Algebraic Shift Register Sequences, Cambridge University Press, 2012
5. Linear Feedback Shift Registers in Virtex Devices, Maria George and Peter Alfke, *Xilinx press*
6. Barkam, Elad; Biham, Eli; Keller, Nathan (2008), *Journal of Cryptology*
7. Lu, Yi; Willi Meier; Serge Vaudenay (2005). "The Conditional Correlation Attack: A Practical Attack on Bluetooth Encryption"
8. Lenore Blum, Manuel Blum, and Michael Shub. "A Simple Unpredictable Pseudo-Random Number Generator", *SIAM Journal on Computing*
9. Lenore Blum, Manuel Blum, and Michael Shub. "Comparison of two pseudo-random number generators", *Advances in Cryptology: Proceedings of Crypto*
10. Notes on the Design and Analysis of the Yarrow Cryptographic Pseudorandom Number Generator, J. Kelsey, B. Schneier, and N. Ferguson, *Sixth Annual Workshop on Selected Areas in Cryptography, Springer Verlag*
11. Niels Ferguson and Bruce Schneier, *Practical Cryptography*, published by Wiley
12. John Viega, "Practical Random Number Generation in Software," *acsac*, pp. 129, 19th Annual Computer Security Applications Conference
13. Dorrendorf, Leo; Zvi Gutterman, Benny Pinkas. "Cryptanalysis of the Random Number Generator of the Windows Operating System"
14. Cryptanalysis of the Alleged SecurID Hash Function, Biryukov Alex
15. A Collection of Selected Pseudorandom Number Generators, Kart Entacher
16. Press, William H., et al. (1992). *Numerical Recipes in FORTRAN 77: The Art of Scientific Computing* (2nd ed.).



# **Simulation & Modeling**

---



# Classical Realization of Grover's Quantum Search Algorithm using Toffoli gates

Manuel-Iván Casillas-del-Llano<sup>1</sup> and Álvaro-Lorenzo Salas-Brito<sup>2</sup>

<sup>1</sup>Universidad Autónoma Metropolitana. Unidad Azcapotzalco. D.F., México  
al210180113@alumnos.azc.uam.mx

<sup>2</sup>Universidad Autónoma Metropolitana. Unidad Azcapotzalco. D.F., México  
asb@correo.azc.uam.mx

**Abstract.** Grover's algorithm is used to search for quantum data. However, this algorithm procedure is described by means of concepts and operators from quantum theory; concepts hardly known by computer scientists. In this work we propose an alternative classical computing model of Grover's algorithm, using Toffoli gates connected with elementary gates. Our model has been programmed on a high-level programming language and tested using arbitrary elements on a data set. Our results are concordant with those presented on the reference section.

**Keywords:** Grover algorithm, Toffoli gates, Quantum computing.

## 1 Introduction

A computer is a physical device that aids us to process information while running some algorithms. An algorithm is well defined procedure, with finite description, that executes some information processing task. A task of this kind can be done by means of physical processes.

At the design level of complex algorithms, it is useful and essential to work with some idealized computational model. However, while analyzing the true limitations of a computer device, especially for practical reasons, it is important not to forget the link between computing and physics. Idealized models can not fully represent all the details of these computational devices.

Classical computing has several limitations. There are problems that cannot be deal with actual computing, such as the impossibility to run on polynomial time the travelling agent problem algorithm or integer factorization.

However, it has been shown that those kinds of problems can be handled by **quantum computing**. Quantum computing uses the phenomena described by quantum theory in order to process information and execute tasks faster than classical computing. Devices that process quantum information are named **quantum computers**.

## 2 Grover's Algorithm

Suppose there is a non sorted data base of size  $N$  consisting (without loss of generality) of numbers from 0 to  $N - 1$ . Using traditional algorithms, we must look up for every element on the data base in order to find the desired item. The average number of steps needed is  $N/2$ , and  $N$  on the worst case scenario; therefore, searching for an element has order of  $O(N)$  time complexity. However, using quantum mechanics procedures, Grover's algorithm only requires  $O(\sqrt{N})$  steps. [1]

Initially, an  $n$  qubit system, with  $N = 2^n$  elements, is set in an equal superposition of all basis states, expressed as

$$|\Psi_0\rangle = \frac{1}{\sqrt{N}} \sum_{i=0}^{N-1} |i\rangle \quad (1)$$

It is possible to search for a specific element inside this system. This particular element is defined as the **marked state**, while the remaining elements of the set are defined as the **collective state** [1]. In order to perform the searching of the marked state, two special operators  $C$  and  $D$  are used, defined as inversion and diffusion operator, respectively. Operator  $C$  has the effect to invert or change the sign of the amplitude in the marked state, and ignores the rest of the elements belonging to the collective state. When operator  $D$  is applied to the superposition of states, it increases the amplitude of the marked state, decreasing the amplitude of the collective state. If we define the compound operator as  $U \equiv DC$ , then each operation of  $U$  is called an iteration. It has been shown that after  $U$  is repeated  $O(\sqrt{N})$  times, the probability of getting the marked state when a measurement is made approaches 1 [1].

### 2.1 Representation of inversion and diffusion operators.

Before reading the following sections, we encourage you to read [4], where the most important operations on quantum computing are explained in great detail. Also, for a wide explanation of Grover's Algorithm insights, we recommend reading [1].

Given the superposition of states  $|\Psi_0\rangle = \left(\frac{1}{\sqrt{N}}\right) \sum_{i=0}^{N-1} |i\rangle$ , we will denote the marked state as  $|M\rangle$ .

Inversion operator  $C$  is defined as  $C \equiv I - 2|M\rangle\langle M|$ , where  $I$  is the identity matrix[5]. Similarly, diffusion operator  $D$  is defined as  $D \equiv 2|\Psi\rangle\langle\Psi| - I$ .

To be able to represent these operators by means of Toffoli gates and elementary operations, it is necessary to know the effect produced by them on the superposition of states. Let's split  $|\Psi_0\rangle$  into two parts: the marked state and the collective state, that is, a linear combination of  $|\Psi_0\rangle$  defined as:

$$|\Psi_0\rangle = \alpha|\Psi\rangle + \beta|M\rangle \quad (2)$$

where  $\alpha$  and  $\beta$  are their respective amplitudes of the collective state and the marked state.

Next, we will apply the inversion operator to this linear combination of states:

$$\begin{aligned} C(\alpha|\Psi\rangle + \beta|M\rangle) &= (I - 2|M\rangle\langle M|)(\alpha|\Psi\rangle + \beta|M\rangle) \\ C(\alpha|\Psi\rangle + \beta|M\rangle) &= \alpha|\Psi\rangle - \left(\frac{2\alpha}{\sqrt{N}} + \beta\right)|M\rangle \end{aligned} \quad (3)$$

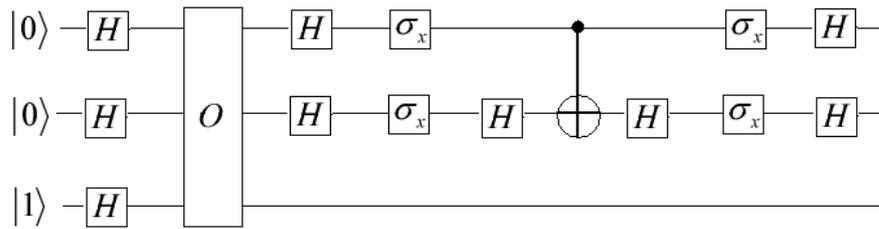
At this moment,  $C$  operator has changed the sign of the marked state. Next step is to apply the diffusion operator on equation (2). That is:

$$\begin{aligned} D\left[\alpha|\Psi\rangle - \left(\frac{2\alpha}{\sqrt{N}} + \beta\right)|M\rangle\right] &= (2|\Psi\rangle\langle\Psi| - I)\left[\alpha|\Psi\rangle - \left(\frac{2\alpha}{\sqrt{N}} + \beta\right)|M\rangle\right] \\ D\left[\alpha|\Psi\rangle - \left(\frac{2\alpha}{\sqrt{N}} + \beta\right)|M\rangle\right] &= \left(\alpha - \frac{4\alpha}{N} - \frac{2\beta}{\sqrt{N}}\right)|\Psi\rangle + \left(\frac{2\alpha}{\sqrt{N}} + \beta\right)|M\rangle \end{aligned} \quad (4)$$

We have applied once the compound operator  $U \equiv DC$ , and so, an iteration of Grover's algorithm have been made. According to [7], the number of iterations needed to approach the amplitude of the marked state to 1 is  $\lfloor \pi\sqrt{N}/4 \rfloor$ .

**2.2 Alternative representation of Grover’s algorithm using different approaches.**

Grover algorithm can be also represented by circuits using interconnected quantum gates and the Toffoli gate [8]. A circuit for 2 qubit system is shown in Figure 1.



**Fig. 1.** A quantum circuit that implements Grover’s algorithm for N = 4 elements. Pauli matrix  $\sigma_x$  behaves as a NOT gate. Block with letter O denotes a query to the oracle.

Another representation of Grover’s algorithm can be implemented using optical approaches, as described in [16] for a system of 2 qubit elements. An implementation using two trapped atomic ion qubits for a system of 2 qubit elements is also proposed [17]. Other representations such as nuclear magnetic resonance are described in [5].

As it will be shown on the next sections, we will build a model that will also represent Grover’s algorithm, but using only classical (non-quantum) elementary gates (with its limitations, see Conclusions section).

**3 Toffoli gates**

Toffoli gates were invented by Tommaso Toffoli [15]. Its main characteristic is that it is a universal reversible logic gate. It is a universal gate because any logic gate can be constructed by means of several Toffoli gates interconnected. It is a reversible logic gate because, given a certain output, we can obtain its corresponding input. Toffoli gates can be modeled using the billiard ball model [2]. This gate is also known as a controlled-controlled-NOT gate, because it flips the third bit on a 3-bit gate if and only if the first two bits are 1. Fig. 2a) shows Toffoli gate truth table when applied to three bits, and Fig. 2b) shows its circuit representation.

Toffoli gates are crucial for our proposed model, since it will aid us in the construction of the inversion operator C because it can detect if the marked state was found or not (see Section 4.2). The output of the inversion operator C will be zero if the element is not the marked state, and it will be 1 if the marked state was found. These outputs will be used then for further calculations.

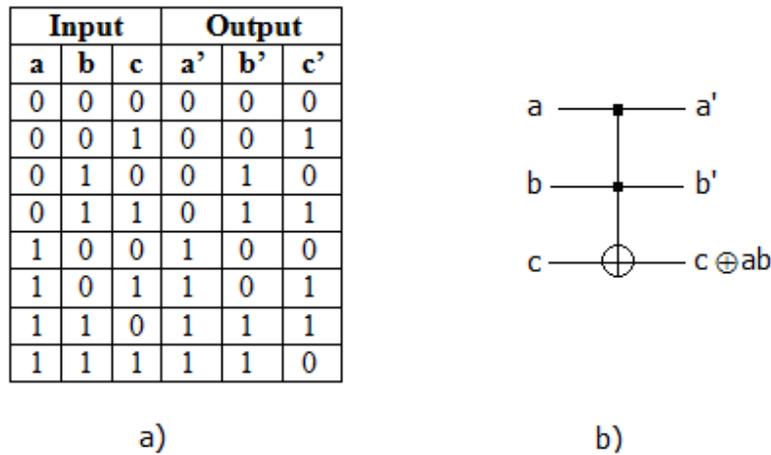


Fig. 2. a) Toffoli gate truth table when applied to three bits. b) Circuit implementation.

#### 4 Realization of Grover's Algorithm using Toffoli gates

Now that we have obtained the necessary equations from the preceding section, we now are able to model Grover's algorithm. If we look closely to Eq. 3 and 4, we can see that the operations involved are elementary additions, subtractions, multiplications and divisions, such as  $\alpha - 4\alpha/N - 2\beta/\sqrt{N}$ . Thus, we need basic gates that perform these operations, such that this model will be constructed interconnecting classical logic gates. The elements involved in these basic operations are needed for the model, so we must supply them at the beginning of the execution. We call these elements as the **control data** of the model. Also, the **input data** will consist of the initial element list, which contains the superposition of states. Fig. 3 shows a general diagram for the proposed model.

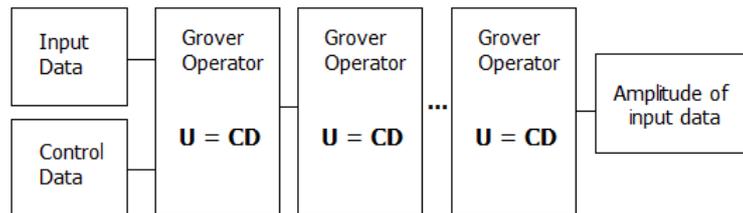


Fig. 3. General schematic of Grover's algorithm procedure. Grover operator is applied the optimum number of iterations in order to increase the amplitude of the marked state, and thus increasing its probability.

Elements from Figure 3 are described below.

**Input data:** It consists of the element list, which stores the marked state and the collective state.

**Control data:** These are fixed data that must be supplied to the algorithm before its execution. Control data consists of  $\alpha$ , which represents the amplitude of the marked state;  $\beta$ , which stores the amplitude of all the collective set; and finally  $1/N$  and  $1/\sqrt{N}$ , where  $N$  represents the total number of states on the database.

**Grover operator:** This operator was defined in section 2 as  $U \equiv DC$ .

**Input data amplitudes.** This is the set of the final amplitudes of the marked state and the collective state. After applying Grover operator a total of  $\lfloor \pi\sqrt{N}/4 \rfloor$  iterations to the superposition of states, we expect that the amplitude of the marked state is almost 1.

#### 4.1 Elementary gates

In order to implement Grover’s algorithm using non-quantum operations through classical gates, we need to define them first. Such gates constitute the set of basic operators of the model.

##### $\Pi$ GATE

This gate requires two input elements. It returns the product of both elements. (See Fig. 4)

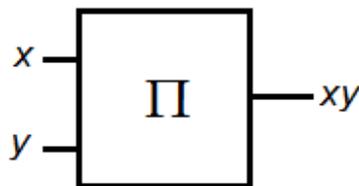


Fig. 4.  $\pi$ -gate: it returns the product of  $x$  times  $y$

$\Sigma$  GATE

This gate also requires two input elements. It returns the sum of both elements. (See Fig. 5)

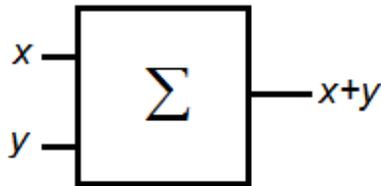


Fig. 5:  $\Sigma$ -gate: it returns the sum of x and y.

$\sigma$  GATE

For this gate, two input data are needed. If the second element is set to 1, the first element will suffer a change of its algebraic sign, otherwise, it will remain unaltered. (See Fig. 6)

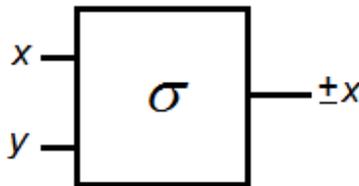
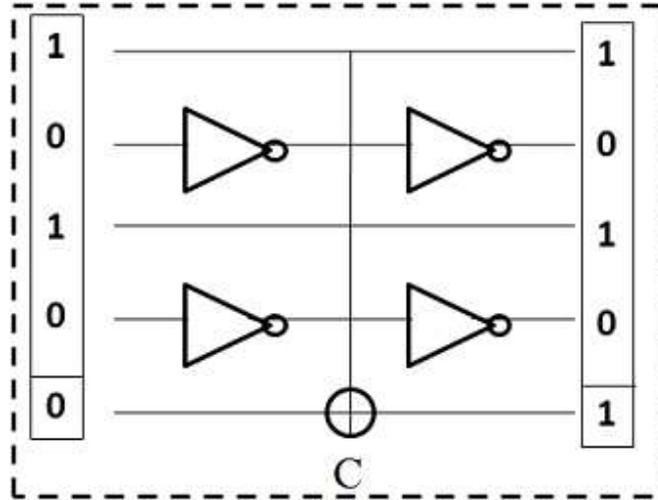


Fig. 6:  $\sigma$ -gate: if y is set to 1, this gate will change x's sign. Oterwhise, x will keep unchanged.

4.2 Modeling of Grover operator using Toffoli and elementary gates.

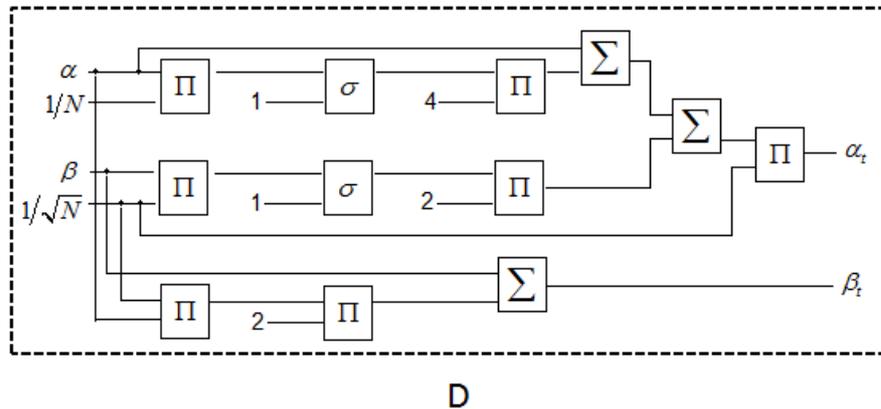
As stated before, Grover operator consists of the application of C operator, followed by D operator, that is,  $U \equiv DC$ .

Inversion operator C is constructed according to the element we are searching for, that is, the marked state. We constructed inversion operator using the binary representation of the element as follows: if the marked state contains zeroes, two NOT gates are put sequentially, otherwise, no gate is needed. These gates are connected by means of a **Toffoli gate**; this gate acts as follows: if every bit is set to 1, it means that the marked state was found, and it will return a 1 as an output, otherwise, it will return a zero, (that is, the marked state was not found). Suppose we want to search for the element 10 on the superposition of states (whose binary representation is 1010 for a 4-bit system). Construction for the inversion operator C using elementary and Toffoli gates is shown in Fig 7.



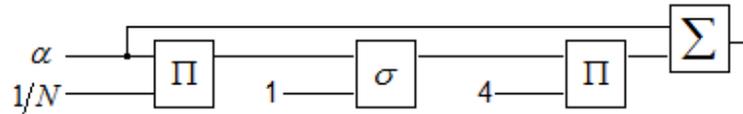
**Fig. 7:** Construction of inversion operator C for element 10 (which binary representation is 1010). Fifth bit is set to 0, so if every remaining bit at the end was set to 1, this bit will be also switched to 1, meaning that the marked element was found.

Diffusion operator D is needed to increase the amplitude of the marked state. According to Equation 3, this operator can be constructed using elementary gates as shown in Fig 8.



**Fig. 8:** Diffusion operator. Interconnection of classical gates allows to increase the amplitude of the marked state. This operator along with the inversion operator successfully simulates Grover operator  $U = CD$ .  $\alpha_t$  and  $\beta_t$  are variables that will be used on further calculations.

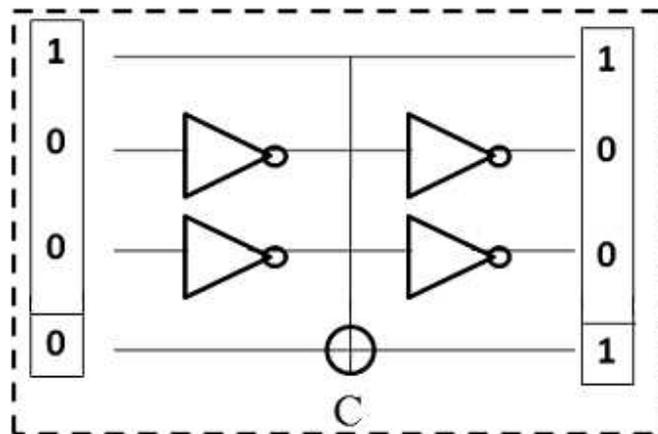
For illustrative purposes, let's take a closer look at the first part of Fig. 8. Eq. 4 shows that the partial chain of operations  $\alpha - 4\alpha/N$  is performed. This is done by the section described in Fig. 9, taken from the diagram in Fig. 8



**Fig. 9:** Carefully following each element on the diagram, we see that the operation  $\alpha - 4\alpha/N$  is successfully made.

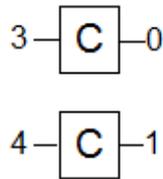
### 4.3 Searching for a marked state using the model.

In order to show how the model works, let's look for the element labeled "4" stored in a list of 3 q-bit elements. First, we construct the Inversion operator C for element 4 using a Toffoli gate and NOT gates (see Fig. 10)



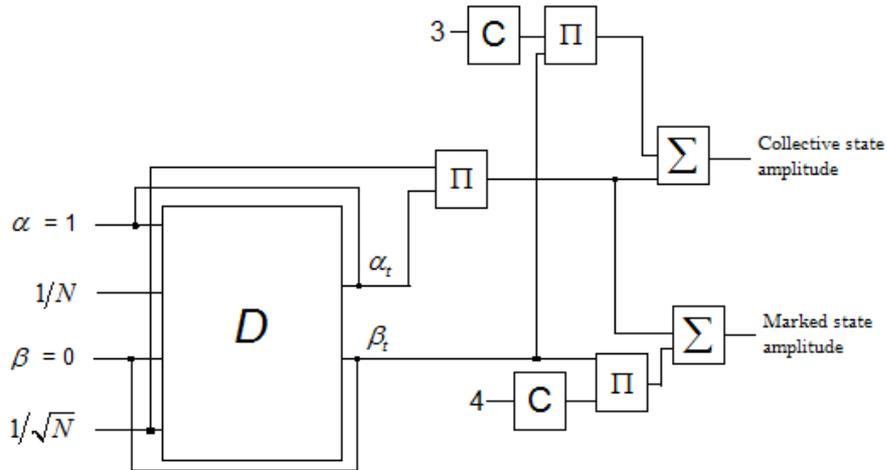
**Fig. 10:** Construction of inversion operator C for element 4 (which binary representation is 100). Fourth bit is set to 0, so if every remaining bit at the end was set to 1, this bit will be also switched to 1, meaning that the marked element was found.

Toffoli gates are useful on the task of finding the marked state, since it detects wherever the marked state is there or not. Since the entire list consists of 8 elements (from 0 to 7), we must apply the inversion operator for every single element belonging to the list. Fig. 11 shows the results of applying the inversion operator on element 3 and 4.



**Fig. 11:** When inversion operator is applied to element “3”, it returns a 0 since it is not the marked state. On the other hand, C operator returns a 1 when applied to the element 4, that is, the marked state was found. This outputs are used for further calculations.

Next step is to increase the amplitude of the marked state using the diffusion operator. Fig. 12 shows how while combining the outputs from both in andversion diffusion operators we increase the amplitude of the marked state, while the collective state remains without change. Notice that only the element 3 and 4 are shown for simplicity, but this has to be done for every element on the list of elements.



**Fig. 12:** With the output of inversion operator, it is now possible to increment the amplitude of the marked state. To accomplish this, we apply diffusion operator D and combine it with the output of inversion operator C, just like Eqs. 3 and 4 describe. Notice how the outputs from the inversion operator are multiplied by the auxiliary variable  $\beta_t$  and then added to the final result. Since the output from the inversion operator is always zero for elements of the collective state, the product is also zero and nothing is added, except from the marked state.

## 5 Simulation of the model using a high-level programming language.

At this stage, we can program an algorithm that simulates the processes of Grover's algorithm for quantum search. This algorithm will be programmed on a high-level programming language to run some tests, in order to validate our model.

### 5.1 General Algorithm

We present the general steps needed to simulate correctly Grover's Algorithm. This algorithm is based on Equations 3 and 4. Its advantage consists of the few steps needed to simulate Grover's algorithm.

Classic Grover Algorithm

Input

$N$ : the total number of elements on the system.

$\bar{\beta} = (\beta_1, \beta_2, \dots, \beta_{N-1})$ : the amplitude coefficient vector of the collective state.

$\alpha_M$ : the amplitude of the marked state.

Output

$\bar{\beta}' = (\beta'_1, \beta'_2, \dots, \beta'_{N-1})$ : the changed amplitude coefficient vector of collective state.

$\alpha'_M$ : the new amplitude of the marked state.

Variables

$\beta$ : the value of the amplitude of the collective state. We can store it on a single variable since all the collective state will have the same amplitude through the entire algorithm.

*coef*: an auxiliary variable used to store intermediate values.

Begin

//Initialize every element of the vector  $\beta_i \in \bar{\beta}$  and  
//the amplitude of the marked state to  $\alpha_M$  an  
//equally superposition of states

$\beta \leftarrow 1/\sqrt{N}$

$\alpha_M \leftarrow 1/\sqrt{N}$

```

Repeat from  $i = 1$  to  $\lfloor \pi\sqrt{N}/4 \rfloor$ 
  //Apply the operations from Equation 3 to the
  //amplitude of the marked state and the
  //collective state.
   $\alpha'_M \leftarrow \alpha_M - 4\alpha_M/N - 2\beta/\sqrt{N}$ 
   $\beta \leftarrow 2\alpha/\sqrt{N} + \beta$ 

  //Store the new normalized marked state to the
  //auxiliary variable.
   $coef \leftarrow \alpha'_M/\sqrt{N}$ 

  //Assign each element  $\beta_i \in \bar{B}$  its new amplitude.
   $\beta_i \leftarrow coef$ 
  //Assign the new amplitude of the marked state.
   $\alpha'_M \leftarrow coef + \beta$ 

End Repeat

```

In order to validate our proposed model, we implemented it using the programming language C++. Data input consists of the number of qubits  $n$  and the number of desired iterations. It let you choose the optimum number of iterations  $\lfloor \pi\sqrt{N}/4 \rfloor$ , or any other number of iterations.

## 5.2. Tests using the optimum number of iterations

Table 1 shows the results for databases built from 1 to 10 qubits, using the optimum number of iterations. The **probability of the marked state** is the probability of obtaining the marked state when a measure of the superposition of states is made. The **probability of the collective state** (or probability of failure) is the probability of obtaining one of the elements of the collective state. Since the probability of the marked state is always same, despite the element searched for, it is unnecessary to list the marked state. This means that if we are looking for the element numbered as 3 on a list of 16 elements, the probability of obtaining it after applying Grover's algorithm is 96.1319%, and that probability will not change if we are looking for the element numbered as 5 on that same list.

**Table 1.** Probabilities of obtaining the marked state and the collective state embedded into an  $n$ -qubit system, using the optimum number of iterations.

No. of qubits $n$	Number of elements N	Number of iterations $\lfloor \pi\sqrt{N}/4 \rfloor$	Probability of the marked state	Probability of the collective state
1	2	1	50%	50%
2	4	1	100%	0%
3	8	2	94.5313%	5.4691%
4	16	3	96.1319%	3.8685%
5	32	4	99.9182%	0.0806%
6	64	6	99.6586%	0.3402%
7	128	8	99.5620%	0.4318%
8	256	12	99.9947%	0.0052%
9	512	17	99.9448%	0.0511%
10	1024	25	99.9461%	0.1023%

### 5.3 Tests using an arbitrary number of iterations

We made tests using an arbitrary number of iterations, instead of the optimum number. Table 2 shows the results for states from 1 to 10 qubit.

**Table 2.** Probabilities of obtaining the marked state and the collective state embedded into an  $n$ -qubit system, using an arbitrary number of iterations.

No. of qubits $n$	Number of elements N	Number of iterations	Probability of the marked state	Probability of the collective state
1	2	3	50%	50%
2	4	3	25%	75%
3	8	6	99.9786%	0.0217%
4	16	7	36.4913%	63.5085%
5	32	7	20.9918%	79.0097%
6	64	12	0.0071%	99.9936%
7	128	10	91.9442%	8.0518%
8	256	18	54.1236%	45.7980%
9	512	20	94.2684%	5.7232%
10	1024	27	97.8187%	2.1483%

## 6. Analysis of results.

Table 1 shows that if we use the optimum number of iterations, probability of obtaining the marked state is almost 100% (except for the 2-qubit case, which is an expected outcome [14]). Particularlry, Table 1 shows that for a 3-qubit system, a probability of 94.5313%, which is also a result obtained at [1]. On Table 2 it is shown that unexpected results are obtained if a different number of iterations than the optimal is used. There is no way to know if there is an improvement of getting a better probability for the marked state. For example, we again can see on Table 1 that for a 3-qubit system, we get a probability of 94.5313% using the optimum number of iterations (in this case, 2); however, Table 2 shows that we get a better probability (about 99.9786%) if we use 6 iterations. This does not necessarily mean that if we use a bigger number of iterations, the probability of obtaining the marked state will be better. For example, calculating the probability of a 6-qubit system yields a 99.6586% for the marked state, using the optimum number of iterations (6). But if we use twice that number of iterations, Table 2 shows that we get a quite bad result (0.0071%), meaning that the probability of failure (that is, the probability of obtaining an element of the collective state when making a measure on the system) is about 99.9936%. Similar results can be seen when we are measuring probabilities on a database with 16 and 32 elements, or 4 and 5-qubits systems, respectively.

## 7. Conclusions

As powerful as classical computing is, it has several limitations. Research on unconventional computing, such as biological inspired computing and quantum computing is made in order to deal with these limitations. Despite that a lot of research work has been made on quantum computing, there is still so much work to do. There are a lot of obstacles that must be solved before a complete physical implementation of a quantum computer can be made. To be able to analyze and foresight the inner concepts and advantages of quantum computing, it is necessary to have a good understanding about theoretical quantum physics. Such knowledge is often lacked by classical computer scientists. For the aforementioned reasons, we need an alternative model capable of explaining the quantum processes (in this case, quantum data search) on terms that can be understood by classical computing developers and researches. In this work, we have presented an alternative theoretical implementation of Grover's quantum search algorithm. With this model, we are able to analyze the behavior of Grover's algorithm on each iteration, without knowing quantum physics or owning an actual quantum computer. With the analysis of the main quantum operators of the algorithm, an equivalent model using only elementary operators by means of interconnecting Toffoli gates and elementary gates was built. This model was implemented on a high level programming language using equations 2 and 3. Many simulations were made using different databases with several elements. Results obtained coincide with those obtained on [1] and [13].

It is important to mention that our proposed model shows the behavior of the Grover's algorithm for quantum data. However, since we are implementing a quan-

tum algorithm on a classical computer, advantages acquired from quantum theory are lost, such as superposition of states, that is, the process that affects the elements on the database is made one by one, and not at the same time like Grover's algorithm quantum implementation. Our model only shows how the probability of obtaining the marked states increases with the algorithm, but it behaves just as a classical algorithm.

## References

1. C. Lavor, L.R.U. Mansur, R. Portugal. "Grover's Algorithm: Quantum Database Search", arXiv:quant-ph/0301079v1.
2. E. Fredkin and T. Toffoli. "Conservative logic", *International Journal of Theoretical Physics*, 21, pp. 219–253, 1982.
3. A. Steane and E. Rieffel. "Beyond bits: the future of Quantum Information Processing", *IEEE Computer Society*, Vol. 33, pp.38–45, January 2000.
4. N. Mermin. "Lecture Notes on Quantum Computation", Cornell University, Physics, pp-481-681, 2006.
5. M. Nielsen and I. Chuang. "Quantum Computation and Quantum Information", Cambridge University Press, 2000.
6. A. Childs and W. Dam. "Quantum Algorithms for Algebraic Problems", *Reviews of Modern Physics*, Vol. 82, No. 1, pp. 1-52, January-March 2010.
7. P. Kaye, R. Laflamme, M. Mosca. "An Introduction to Quantum Computing", Oxford University Press, 2007.
8. G. Benenti, G. Casati, G. Strini. "Principles of Quantum Computation and Information", World Scientific, 2004.
9. S. Lomanco. "A lecture on Shor's Quantum Factoring Algorithm", arXiv:quant-ph/0010034v1.
10. I. L. Chuang, N. Gershenfeld, and M. Kubinec, "Experimental implementation of fast quantum searching", *Physical Review Letters*, 80, pp. 1050–1053, 1999.
11. Z. Diao, "Exactness of the Original Grover Search Algorithm", arXiv:quant-ph/1010.3652v1.
12. L. Grover, "A fast quantum mechanical algorithm for database search", arXiv:quant-ph/9605043
13. S. Lomanco. "A lecture on Grover's Quantum Search Algorithm", arXiv:quant-ph/0010034v1
14. K.A. Brickman, P.C. Haljan, P.J. Lee, M. Acton, L. Deslauriers, C. Monroe, "Implementation of Grover's Quantum Search Algorithm in a Scalable System", arXiv:quant-ph/0510066v2
15. T. Toffoli, "Reversible computing," in *Proceedings of the 7th Colloquium on Automata, Languages and Programming*. Springer-Verlag London, UK, 1980, pp. 632–644.
16. P.G. Kwiat, J.R. Mitchell, P. D. D. Schwindt, A. G. White, "Grover's search algorithm: An optical approach", arXiv:quant-ph/9905086v1
17. K.A. Brickman, P.C. Haljan, P.J. Lee, M. Acton, L. Deslauriers, C. Monroe, "Implementation of Grover's Quantum Search Algorithm in a Scalable System", arXiv:quant-ph/0510066v2



# Modelado y simulación de una barra actuada por dos rotores con movilidad espacial restringida

O. Sánchez-Rodríguez<sup>1\*</sup>, J.C. Chimal-Eguia<sup>1</sup>, and R. Barrón-Fernández<sup>1</sup>

Centro de Investigación en Computación, Instituto Politecnico Nacional,  
Ciudad de México 07738, México  
{osanchezrod}@gmail.com

**Resumen** Este trabajo se dirige hacia la estabilización de una barra actuada por dos rotores con movilidad espacial restringida. El problema de controlar el sistema, se resuelve empleando un novedoso método propuesto por Astolfi y Ortega, llamado estabilización por Inmersión e Invariancia en el cual se robustece un controlador, por inmersión en un sistema con mejores características de desempeño y su invariancia nos asegura que se converge a un punto de estabilidad. Se valida el algoritmo de control obtenido mediante software de simulación numérica e implementación en un ambiente de realidad virtual, el modelo del ambiente de realidad virtual se diseño en software de CAD.

**Abstract** This work is directed toward stabilizing rod actuated by two rotors with restricted spatial mobility. The problem of controlling the system is solved using a new method proposed by Astolfi and Ortega, named Immersion and Invariance stabilization. The method consists of a controller which is strengthened by immersion in a system with better performance characteristics and its invariance ensures that will converge to a point of stability. This is validated by numerical software simulation and implemented on a virtual reality environment which was designed in CAD software.

**Keywords:** Estabilización, Controlador, Inmersión e Invarianza, Simulación

## 1. Introducción

Durante los últimos años se han generado diversos trabajos referentes al desarrollo de vehículos aéreos no tripulados más eficientes, con mejores prestaciones y altos rendimientos, debido a la necesidad del sector aeroespacial de contar con vehículos de tamaño pequeño y su amplia gama de aplicaciones que van desde el monitoreo de tráfico vehicular, inspección de zonas de riesgo, vigilancia de fronteras, medición de condiciones meteorológicas, búsqueda y rescate en desastres naturales y en aplicaciones militares como vigilancia de zonas en conflictos bélicos

---

\* Agradecemos el apoyo del IPN a través del proyecto SIP-20121154 así como las valiosas opiniones y comentarios de las personas que revisaron este trabajo.

donde es necesario hacer un reconocimiento de la zona sin arriesgar las vidas de los pilotos. Uno de los principales objetivos para los investigadores actualmente es la de lograr automatizar la dinámica de vuelo de tales aeronaves de forma que sean estables y se desempeñen en la vecindad de una trayectoria deseada.

El problema que se aborda en este trabajo es el de estabilizar el desacople entre la dinámica longitudinal y lateral, el estudio entonces se hace sobre un PVTOL con dinámica cartesiana restringida, el sistema se describe como un sistema con dos propulsores dispuestos a una distancia sobre el plano longitudinal ( $x$ ) y dinámica de traslación restringida.

Para lo cual obtenemos el modelo matemático que describe la dinámica de la barra actuada por dos rotores para la que se diseña un algoritmo de control que estabilice al sistema, tales algoritmos se validan mediante simulaciones en software y con su implementación en un Ambiente de Realidad Virtual. El artículo está organizado de la siguiente manera, en la sección 1 presentamos una breve introducción, la sección 2 aborda los antecedentes, en la sección 2 se obtiene el modelo matemático a estudiar y se analizan sus propiedades, mientras que en la sección 4 se diseña el algoritmo de control, en las secciones 5 y 6 se muestran las simulaciones y en la sección 7 se hacen las conclusiones del artículo.

## 2. Antecedentes.

Durante la última década se han venido dando esfuerzos por parte de la comunidad científica y tecnológica orientados a la estabilización y seguimiento de trayectorias de las aeronaves de ala rotativa. Altug et al. [1] proponen un algoritmo de control para estabilizar el cuatrirotor usando visión artificial y una cámara como el sensor principal. Ellos estudian dos métodos, en el primero usan un algoritmo de control por linealización y en el otro la técnica de backstepping. Sus resultados fueron probados exitosamente en simulaciones.

En [2] Heredia et al. Tratan el problema de controlar un helicóptero autónomo y lo hacen mediante simulaciones computacionales de estrategias de control basadas en lógica difusa y control tracking no-lineal en dos escenarios posibles ascenso vertical y un movimiento simultáneo longitudinal y lateral. El controlador consiste en un MIMO (múltiples inputs múltiples outputs) de lazo interno para la estabilización y cuatro SISO en lazos guiados para velocidad y posición. Pounds et al. [3] Conciben y desarrollan un algoritmo de control para un prototipo de cuatro rotores ellos usan una unidad de medición inercial (IMU por sus siglas en inglés) para la medición de la velocidad y aceleración angular, usando una linealización del modelo dinámico conciben el algoritmo de control. Los resultados fueron probados por simulación.

Vissière David and Petit Nicolas [4] consideran el problema de desarrollar un sistema modular embebido en tiempo real para aplicaciones de control en UAVs, sus esfuerzos los guían en dirección de la programación y la propuesta de estrategias de control. Para probar sus resultados reportan la implementación de un filtro extendido de Kalman a  $75Hz$  usado para la estimación de los estados de un pequeño helicóptero. Su trabajo es más bien un estudio detallado de los

pasos y elementos necesarios para la implementación de un sistema embebido. Chowdhary Girish and Lorenz Sven [5] conciben el problema de estabilización de un UAV VTOL mediante estados retroalimentados en línea como una técnica de control lineal sencilla con el único problema de la envolvente de vuelo, el problema es a menudo acentuado debido a un inadecuado modelo lineal, al ruido de medición en los sensores y a las perturbaciones externas. Ellos presentan una arquitectura de control basada en una extensión válida de la ley de control óptima lineal por estados completos retroalimentados. Una extensión del filtro de Kalman es usada en el problema de los estados y la estimación de los parámetros. Basados en la estimación de los parámetros la ganancia de estado retroalimentado es calculada resolviendo la ecuación de Ricatti para la optimización cuadrática del control en línea.

Arda Özgür Kivrak [6] en su tesis doctoral aborda el problema del diseño de un sistema embebido para un vehículo aéreo cuatrorotor equipado con sensores inerciales, el sistema de control lo desarrolla en Matlab/Simulink y su implementación en tiempo real usando el módulo de Simulink Real time Windows target, ahí diseña un regulador cuadrático lineal para la estabilización de la actitud de vuelo. El hardware lo integra una tarjeta de adquisición de datos, los controladores de los motores de CD, un conjunto de sensores, motores de CD y la plataforma DraganFlyer V Ti.

Salazar-Cruz et al. [7] describen el diseño de un sistema de control embebido para un vehículo aéreo no tripulado (UAV) del tipo cuatrorotor con capacidades para realizar vuelos estacionarios. El modelo dinámico del vehículo se presenta mediante ecuaciones de Euler-Lagrange y proponen una estrategia de control basada en la saturación de integradores. Un sistema de control embebido describe la arquitectura de vuelo autónomo estacionario. Los principales componentes del sistema son un microcontrolador, una unidad de medición inercial (IMU), un sistema de posicionamiento global (GPS) y sensores infrarrojos. Los ángulos de Euler se calculan utilizando un algoritmo de fusión de datos. Los resultados experimentales demuestran que el sistema de control funciona de manera satisfactoria para vehículos autónomos volando en interiores.

Adigbli Patrick [8] en su trabajo presenta lo referente al desarrollo de un control remoto fiable para asistir en el control de un mini robot aéreo con cuatro rotores y que garantice las capacidades de un vuelo estable. Como primera fase obtiene el modelo dinámico mediante ecuaciones de Euler-Lagrange y prueba tres diferentes tipos de leyes de control, por estados retroalimentados, backstepping y modos deslizantes todos para estabilización y posición del UAV, el autor menciona que todos ellos fueron comparados en simulaciones aunque no menciona las ventajas de cada uno de ellos.

Mian Y Wang [9] proponen un controlador no lineal para la estabilización de un helicóptero, la estrategia se basa en la saturación de integradores. Debido a los logros positivos que se tienen en este tipo de estrategia es que les permitió tomar en cuenta las condiciones de acoplamiento. El controlador mostró buen resultado en simulaciones respecto de otros controladores, y gracias a los sensores embebidos y de control es que es capaz de realizar vuelos autónomos en tiempo

real. Sus resultados muestran que la estrategia de control es capaz de realizar tareas de forma autónoma como el despegue, aterrizaje y el vuelo estacionario. Finalmente Ollero y Merino [10] examinan los métodos y tecnologías que se han aplicado en la robótica aérea, presentan varios vehículos aéreos no tripulados, resumen las distintas técnicas de control incluidas las arquitecturas de control y métodos de control. Además examinan brevemente los equipos, las técnicas de visión robótica aérea.

### 3. Modelo matemático

#### 3.1. Modelo dinámico de la barra

Para obtener el modelo dinámico de la barra actuada se considera que los dos rotores producen una fuerza normal al plano horizontal de la barra. Debido

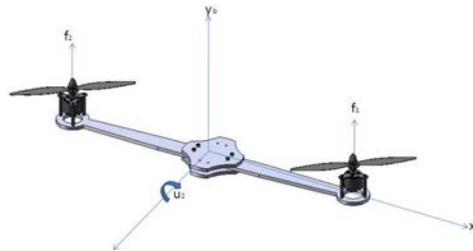


Figura 1: Modelo del sistema

a que la barra tiene 2 rotores que le proporcionan la fuerza de empuje, el empuje total está dado por,  $T_t = \sum_{i=1}^2 T_i$  las fuerzas provocadas por la acción de los motores  $T_1$  y  $T_2$  producen un par respecto del centro de gravedad. El par rotacional total se da por la siguiente expresión  $u_\theta = (T_2 - T_1)l$ , donde  $l$  es la distancia del centro de gravedad de la barra al eje de los motores.

Realizando el análisis de fuerzas, aplicando la segunda ley de Newton y considerando como variables de estado a los elementos que acumulan energía que en este caso es una masa que gira a cierta velocidad respecto de su centro de gravedad, la ecuación que modela el comportamiento dinámico del sistema es la siguiente ecuación diferencial

$$J\ddot{\theta} + B\dot{\theta} = u \quad (1)$$

donde  $\phi(u)$  es una función de control desconocida que depende del tiempo.

$$J\ddot{\theta} + B\dot{\theta} = \phi(u) \quad (2)$$

Pasamos a una representación en ecuaciones de primer orden

$$\dot{\theta}_1 = \theta_2 \tag{3}$$

$$\dot{\theta}_2 = -\frac{B}{J}(\theta_2) + \frac{1}{J}\phi(u) \tag{4}$$

El modelo se representa en variables de estado, para lo cual se definen los siguientes estados  $x_1 = \theta_1, x_2 = \theta_2$  y se obtiene la siguiente representación matricial

$$\dot{x} = f(x, u) = \begin{bmatrix} 0 & 1 \\ 0 & -\frac{B}{J} \end{bmatrix} x + \begin{bmatrix} 0 \\ \frac{1}{J} \end{bmatrix} [u] \tag{5}$$

El valor de  $J$  fue obtenido de un análisis del modelo hecho en SolidWorks y la masa pesando cada uno de los elementos que la conforman,  $B$  es el amortiguamiento del sistema modelado por la siguiente ecuación

$$B = -F | \dot{\theta} | \dot{\theta}$$

### 3.2. Estabilidad

Se analiza la estabilidad del sistema calculando el polinomio característico del sistema en lazo abierto el cual es,

$$\lambda^2 + \lambda \frac{B}{J} \tag{6}$$

al igualar 6 a cero se obtienen los valores propios del sistema  $\lambda_1 = 0$  y  $\lambda_2 = -\frac{B}{J}$  Por lo tanto para cualquier valor del sistema este tendrá un valor propio en cero y un valor propio negativo, solo se infiere estabilidad interna, ya que la matriz  $A$  no es Hurwitz.

## 4. Diseño del algoritmo de control

El algoritmo de control usado para la estabilización del sistema de la barra se basa en el método propuesto por Alessandro Astolfi y Romeo Ortega [11], utilizado para la estabilización asintótica y el diseño de leyes de control adaptables para sistemas no lineales.

Consideremos un sistema de la forma

$$\dot{x} = f(x, u)$$

y el problema básico de estabilización de encontrar una ley de control  $u = u(x)$  (esto cuando sea posible) de forma tal que el sistema en lazo cerrado sea localmente (globalmente) asintóticamente estable. El procedimiento propuesto para resolver este problema consiste de dos pasos. Primero encontrar un sistema dinámico objetivo  $\dot{\xi} = \alpha(\xi)$  que sea localmente (globalmente) asintóticamente

estable y de dimensión estrictamente menor que  $x$ , un mapeo  $x = \pi(\xi)$ , y una función  $c(x)$ , tal que

$$f(\pi(x), c(\pi(\xi))) = \frac{\partial \pi}{\partial \xi}(\xi)\alpha(\xi)$$

esto es que cualquier trayectoria  $x(t)$  del sistema  $\dot{x} = f(x, c(x))$  es la imagen a través del mapeo  $\pi(\cdot)$  de una trayectoria del sistema objetivo. Nótese que el mapeo  $\pi : \xi \rightarrow x$  es una inmersión, es decir, el rango de  $\pi$  es igual a la dimensión de  $\xi$ . En segundo lugar, aplicar una ley de control que contribuye a la variedad  $x = \pi(\xi)$  atrae y mantiene las trayectorias en lazo cerrado acotadas. De esta manera se tiene que el sistema en lazo cerrado se comportará asintóticamente como el sistema objetivo deseado y la estabilidad estará garantizada.

Retomamos las ecuaciones 3 y 4 y se reescriben como

$$\dot{\theta}_1 = \theta_2 \tag{7}$$

$$\dot{\theta}_2 = \xi(t) + \frac{1}{J}\phi(u) \tag{8}$$

donde  $\xi(t)$  es una función desconocida que depende del tiempo, y se obtiene el siguiente sistema

$$\dot{\theta}_1 = \theta_2 \tag{9}$$

$$\dot{\theta}_2 = \xi(t) + \frac{1}{J}u \tag{10}$$

Considérese el siguiente sistema objetivo de orden completo, se diseñará un nuevo controlador para dicho sistema, donde  $u = u(t)$  es cualquier ley de control estabilizante por retroalimentación para dicho sistema

$$z_1 = \theta_2 - \hat{\theta}_2 + \beta_1(\theta_1) \tag{11}$$

$$z_2 = \xi(t) - \rho_1 + \beta_2(\theta_1) \tag{12}$$

$$z_3 = \dot{\xi}(t) - \rho_2 + \beta_3(\theta_1) \tag{13}$$

se despejan las variables de estado de interés para la inmersión en el sistema de orden superior, se debe encontrar una función  $\psi(x, z)$  que preserve acotadas a las trayectorias y establezca asintóticamente a cero al dinámica

$$\theta_2 = z_1 + \hat{\theta}_2 - \beta_1(\theta_1) \tag{14}$$

$$\xi(t) = z_2 + \rho_1 - \beta_2(\theta_1) \tag{15}$$

$$\dot{\xi}(t) = z_3 + \rho_2 - \beta_3(\theta_1) \tag{16}$$

supóngase  $\ddot{\xi} \approx 0$ , y aplíquese la siguiente entrada de control

$$u = J[(\rho_1 - \beta_2) + k_p\theta_1 + k_d\theta_2]$$

y se sustituyen las ecuaciones 9 y 10 con lo cual es obtenido el siguiente sistema

$$\dot{\theta}_1 = \theta_2 \tag{17}$$

$$\dot{\theta}_2 = -\xi + \rho_1 - \beta_2 + k_p\theta_1 + k_d\theta_2 \tag{18}$$

considérese la ecuación 12 se reescribe el sistema de la siguiente forma

$$\dot{\theta}_1 = \theta_2 \quad (19)$$

$$\dot{\theta}_2 = -z_2 + k_p\theta_1 + k_d\theta_2 \quad (20)$$

es derivado el sistema objetivo para buscar la función que nos preserve dicha condición de estabilidad

$$\dot{z}_1 = \xi + \frac{1}{J}u - \dot{\theta}_2 + \frac{\partial\beta_1}{\partial\theta_1}\theta_2 \quad (21)$$

$$\dot{z}_2 = \dot{\xi} - \dot{\rho}_1 + \frac{\partial\beta_2}{\partial\theta_1}\theta_2 \quad (22)$$

$$\dot{z}_3 = -\dot{\rho}_2 + \frac{\partial\beta_3}{\partial\theta_1}\theta_2 \quad (23)$$

se hace la sustitución de  $\xi$ ,  $\dot{\xi}$  y  $\theta_2$  en las ecuaciones anteriores, obteniéndose el siguiente sistema

$$\dot{z}_1 = z_2 + \rho_1 - \beta_2 + \frac{1}{J}u - \dot{\theta}_2 + \frac{\partial\beta_1}{\partial\theta_1}(z_1 + \hat{\theta}_2 - \beta_1) \quad (24)$$

$$\dot{z}_2 = z_3 + \rho_2 - \beta_3 - \dot{\rho}_1 + \frac{\partial\beta_2}{\partial\theta_1}(z_1 + \hat{\theta}_2 - \beta_1) \quad (25)$$

$$\dot{z}_3 = -\dot{\rho}_2 + \frac{\partial\beta_3}{\partial\theta_1}(z_1 + \hat{\theta}_2 - \beta_1) \quad (26)$$

son despejadas nuevamente las variables de interés, quedando el siguiente sistema

$$\dot{\hat{\theta}}_2 = \rho_1 - \beta_2 + \frac{1}{J}u + \frac{\partial\beta_1}{\partial\theta_1}(\hat{\theta}_2 - \beta_1) \quad (27)$$

$$\dot{\rho}_1 = \rho_2 - \beta_3 + \frac{\partial\beta_2}{\partial\theta_1}(\hat{\theta}_2 - \beta_1) \quad (28)$$

$$\dot{\rho}_2 = \frac{\partial\beta_3}{\partial\theta_1}(\hat{\theta}_2 - \beta_1) \quad (29)$$

Este sistema cumple con la condición de la función  $\psi(\theta, z)$  de preservar acotadas las trayectorias y estabilizarlas asintóticamente en cero. Ahora se resolverá el problema de encontrar una función y un control  $u$  tal que describa a la variedad invariante. Esto requiere de la solución de una ecuación diferencial parcial, esta se comienza a construir considerando el siguiente sistema de ecuaciones, tomado 21, 22 y 23 y sustituyéndoles  $z_1$ ,  $z_2$  y  $z_3$ , el sistema es

$$\dot{z}_1 = \frac{\partial\beta_1}{\partial\theta_1}z_1 + z_2 \quad (30)$$

$$\dot{z}_2 = \frac{\partial\beta_2}{\partial\theta_1}z_1 + z_3 \quad (31)$$

$$\dot{z}_3 = \frac{\partial\beta_3}{\partial\theta_1}z_1 \quad (32)$$

se deriva 30 y se sustituye 31

$$\dot{z}_1 = \frac{\partial \beta_1}{\partial \theta_1} z_1 + \dot{z}_2 = \frac{\partial \beta_1}{\partial \theta_1} z_1 + \frac{\partial \beta_2}{\partial \theta_1} z_1 + z_3 \quad (33)$$

se deriva 33 y se sustituye 32 y se obtiene

$$\ddot{z}_1 = \frac{\partial \beta_1}{\partial \theta_1} \dot{z}_1 + \frac{\partial \beta_2}{\partial \theta_1} \dot{z}_1 + \dot{z}_3 = \frac{\partial \beta_1}{\partial \theta_1} \dot{z}_1 + \frac{\partial \beta_2}{\partial \theta_1} \dot{z}_1 + \frac{\partial \beta_3}{\partial \theta_1} z_1 \quad (34)$$

se iguala a cero 33 y finalmente se obtiene un polinomio con términos que son ecuaciones diferenciales parciales

$$z_1^{(3)} + \beta_1 \ddot{z}_1 + \beta_2 \dot{z}_1 + \beta_3 z_1 = 0 \quad (35)$$

El resultado final es el sistema y el controlador, donde las variables del controlador son obtenidas de las ecuaciones 27, 28 y 29. El sistema estable asintóticamente es descrito por las siguientes ecuaciones

$$\dot{\theta}_1 = \theta_2 \quad (36)$$

$$\dot{\theta}_2 = -\frac{F}{J} + \frac{1}{J}u \quad (37)$$

$$u = J[(\rho_1 - \beta_2) + k_p \theta_1 + k_d(\hat{\theta}_2 - \beta_1)] \quad (38)$$

## 5. Resultados

### 5.1. Sintonización de las ganancias del controlador y del observador

En esta sección se valida el algoritmo de control mediante los resultados obtenidos por simulación numérica de la implementación hecha en *Matlab* y *Simulink*. Se hace un primer conjunto de simulaciones con la finalidad de observar el comportamiento de las ganancias tanto del observador como las ganancias del controlador PD. La figura 2 hace una primera visualización del comportamiento con diferentes ganancias para el controlador mientras que las del observador permanecen constantes en 1,0, es claro que la ganancia optima se encuentra por arriba de 100 y debajo de 1000. Por lo tanto se grafican valores en ese intervalo, la figura 3 muestra el comportamiento a diferentes valores, es fácil concluir que a valores mayores el sobretiro negativo que muestra el sistema es menor, pero de la figura 2 se observa que si es muy grande el sistema no se estabiliza en cero. Por lo tanto se utilizara una ganancia para el controlador de 400.

A continuación como se muestra en la figura 4 se mantienen constantes las ganancias del controlador en 1,0 y se hacen variar las ganancias del observador. Es fácil notar que la ganancia óptima es de 1,35.

### 5.2. Primera simulación del sistema en lazo cerrado

Se hace una primera simulación con los parámetros del controlador y el observador de la tabla 1 La figura 5 y 6 muestran la posición y la velocidad rotacional

Polos del controlador	Polos del observador
1.0	1.0
1.0	1.0
1.0	

Cuadro 1: Parámetros del sistema para la primera simulación

del sistema en lazo cerrado, es apreciable que ambas se estabilizan después de una serie de oscilaciones. Si se observa la figura 7 se notara que la variable estimada  $\hat{\theta}_2$ , muestra un comportamiento similar a la variable  $\theta_2$ , la figura 8 muestra una comparación entre ambas variables, se observa que el observador estima a la variable  $\theta_2$  en 4.5 segundos. La figura 9 muestra el error de estimación del observador y la figura 10 muestra la señal de control que estabiliza al sistema en cero grados.

### 5.3. Segunda simulación del sistema en lazo cerrado

En esta parte se hace una segunda simulación tomando en consideración los polos obtenidos de la sintonización de las ganancias. Los parámetros utilizados para esta simulación son las de la tabla 2 La figura 11 y 12 muestran la posi-

Polos del controlador	Polos del observador
400.0	1.35
400.0	1.35
400.0	

Cuadro 2: Parámetros del sistema para la segunda simulación

ción y la velocidad rotacional del sistema en lazo cerrado, se puede notar que la estabilización es inmediata. Si observamos la figura 13 se nota que la variable estimada  $\hat{\theta}_2$ , muestra un comportamiento igual a la variable  $\theta_2$ , la figura 14 muestra una comparación entre ambas variables, se observa que el observador estima a la variable  $\theta_2$  inmediatamente. La figura 15 muestra el error de estimación del observador y la figura 16 muestra la señal de control que estabiliza al sistema en cero grados. Podemos concluir que al sintonizar adecuadamente el observador este muestra un desempeño excelente, ya que estabiliza al sistema inmediatamente, también debe observarse que la dinámica más rápida de controlador es la que tiene un impacto más notable en el desempeño del lazo de control y que la dinámica del observador debe ser solo un poco más rápida ya que de lo contrario se desestabiliza el sistema, esto debido a una saturación en el lazo de control.

## 6. Ambiente de realidad virtual

Los resultados obtenidos de la simulación en ambiente de realidad virtual son satisfactorios y el sistema describe las trayectorias de estabilización mostradas en las gráficas 5 y 11 a diferencia que se muestra con un modelo de en 3D del sistema. La figura 17 muestra una etapa de la estabilización del sistema y la figura 18 muestra al sistema estabilizado.

## 7. Conclusiones

En el presente trabajo se desarrolla un algoritmo de control para un sistema formado por una barra actuada por dos rotores, el algoritmo propuesto está basado en la teoría propuesta por Astolfi y Ortega [11]. Este controlador mostró un excelente desempeño en las simulaciones cuando las ganancias fueron las adecuadas. En las simulaciones presentadas se observa que la ganancia del controlador es la que tiene mayor presencia en la dinámica del sistema sin embargo una buena elección de las ganancias del observador nos asegura una dinámica lo suficientemente rápida para la correcta estimación del estado  $\theta_2$ , pero debe tenerse cuidado ya que una dinámica demasiado rápida satura el lazo y nos lleva a inestabilidades. Las condiciones iniciales no tienen una consecuencia importante en la estabilización del sistema siempre que se mantengan por debajo de 45 grados sobre la horizontal.

Las gráficas 9 y 15 muestran el error de estimación, el cual como puede verificarse tiende a hacerse cero, nuevamente una correcta sintonización de las ganancias hace que la dinámica del error sea más rápida. El modelo de realidad virtual ayudó a visualizar con anticipación el comportamiento del sistema físico. Esto facilita la interpretación del algoritmo de control y su desempeño.

## Referencias

1. Altug, E., Ostrowski, J.P., Taylor, C.J.: Control of a quadrotor helicopter using dual camera visual feedback. *The International Journal of Robotics Research* 24 **24/5** (2005) 329–341
2. Sensor and actuator fault detection in small autonomous helicopters. In: *A Journal of IFAC. Volume 18 of Mechatronics The Science of Intelligent Machines.*, Elsevier (2008)
3. Pounds, et al: Nonlinear control for systems with bounded inputs: Real-time embedded control applied to uavs. *Proceedings of the 45th IEEE Conference on Decision and Control, San Diego California E.U.A* (2006) 5888–5893
4. Vissière, D., Petit, N.: An embedded system for small-scaled autonomous vehicles. *International Conference on Informatics in Control, Automation and Robotics* (2008) 158–164
5. Chowdhary, G., Lorenz, S.: Control of a vtol uav via online parameter estimation. *Guidance, Navigation, and Control Conference and Exhibit*
6. Arda-Özgür, K.: Design of control systems for a quadrotor flight vehicle equipped with inertial sensors. Tesis doctoral de la Universidad de Atilim, Turquía (Diciembre de 2006)

7. Salazar-Cruz, S., Escareño, J., Lara, D., Lozano, R.: Embedded control system for a four-rotor uav. *International Journal of Adaptive Control and Signal Processing*
8. Adigbil, P.: Nonlinear attitude and position control of a micro quadrotor using sliding mode and backstepping techniques. 3rd US-European Competition and Workshop on Micro Air Vehicle Systems (MAV07) and European Micro Air Vehicle Conference and Flight Competition (EMAV2007) (2007)
9. Mian, A.A., bo Wang, D.: Dynamic modelling and nonlinear control strategy for an underactuated quad rotor rotorcraft. *Journal of Zhejiang University* (2008) 539–545
10. Ollero, A., Merino, L.: Control and perception techniques for aerial robotics. *Annual Reviews in Control* **28** (2004) 167–178
11. Astolfi, A., Ortega, R.: Immersion and invariance: A new tool for stabilization and adaptive control of nonlinear systems. *IEEE Transactions on Automatic Control* **48** (April 2003) 590–606
12. Bishop, R.H., Dorf, R.C.: *Sistemas de Control Moderno*. Décima edn. Pearson, USA (2005)
13. Hamacher, C., et al: *Organización de computadores*. Quinta edn. Mc Graw Hill, USA (2003)
14. Chi-Tsong, C.: *Linear System Theory and design*. third edn. Oxford University Press, New York USA (1999)
15. Chi-Tsong, C.: *Introduction to linear systems*. first edn. Rinehart and Winston line, New York USA (1970)
16. Domínguez, S., et all: *Control en el espacio de estados*. Primera edn. Prentice Hall, Madrid España (2002)
17. Assad, G.D.: *Instrumentación, Acondicionamiento electrónico y adquisición de datos*. Pre edn. Trillas, México (2000)
18. García, E.: *Compilador C, CSS y simulador PROTEUS para microcontroladores dsPIC*. Primera edn. Alfa-Omega, México (2008)
19. Hinrichsen, D., Pritchard, A.: *Mathematical Systems Theory Modeling State Space Analysis Stability and Robustness*. First edn. Springer, Germany (2005)
20. Close, C., Frederick, D.: *Modeling and Analysis of Dynamic Systems*. Third edn. John Wiley Sons Inc., USA (2000)
21. Khalil, H.: *Non linear systems*. Second edn. Prentice Hall, USA (1996)
22. Lozano, R., et all: Real-time stabilization and tracking of a four mini rotorcraft. *IEEE Transactions on Control Systems Technology* **12** (2004) 510–516
23. : *Virtual Reality Toolbox 3.0 Help*, Matlab 2010
24. : *V Real Builder. User's Guide and Reference*

### Apéndice: Gráficas

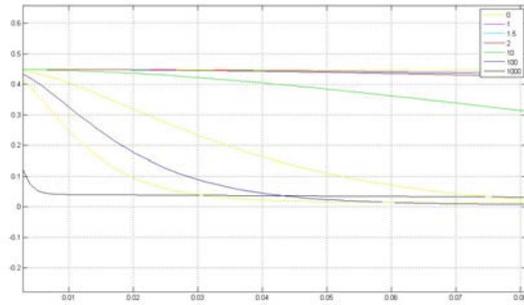


Figura 2: Ganancias del observador

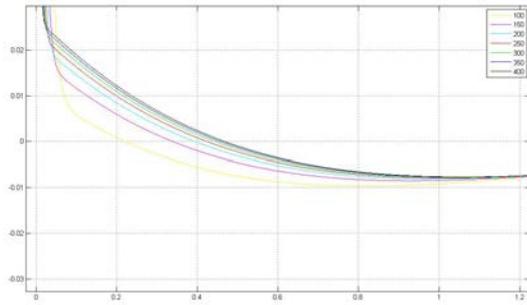


Figura 3: Ganancias del controlador en el intervalo de 100 a 400

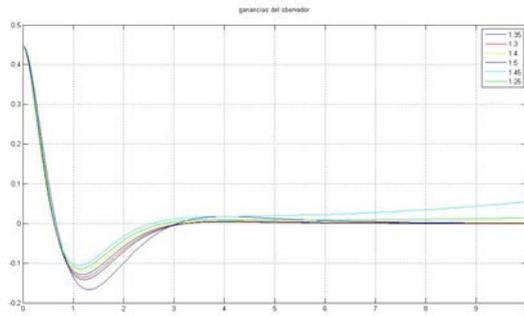


Figura 4: Ganancias del observador en el intervalo 1.0 a 1.5

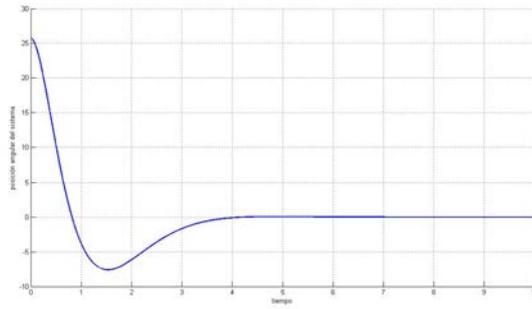


Figura 5: Posición angular  $\theta_1$  del sistema en lazo cerrado

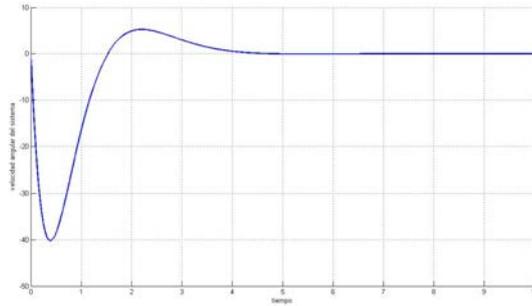


Figura 6: Velocidad angular  $\theta_2$  del sistema en lazo cerrado

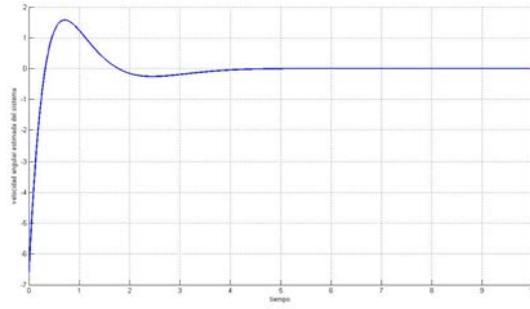


Figura 7: Velocidad angular estimada  $\hat{\theta}_2$  del sistema en lazo cerrado

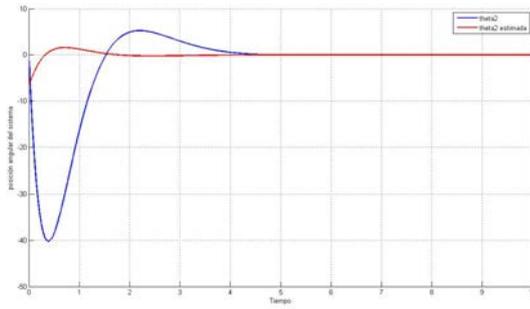


Figura 8:  $\theta_2$  y  $\hat{\theta}_2$

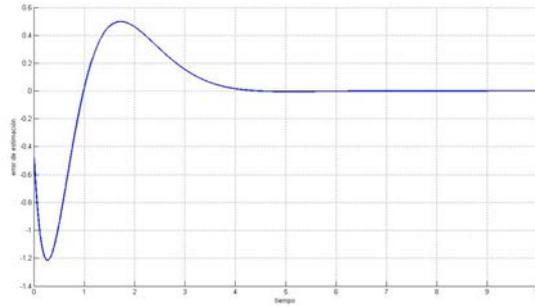


Figura 9: Error de sistema en lazo cerrado

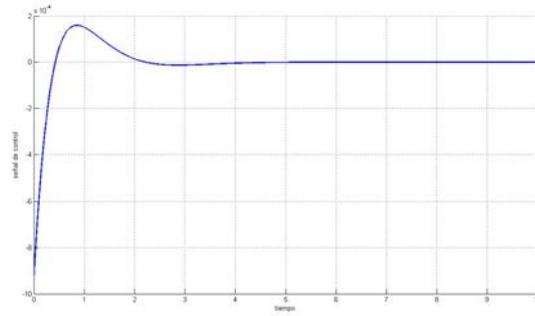


Figura 10: Señal de control para el sistema en lazo cerrado

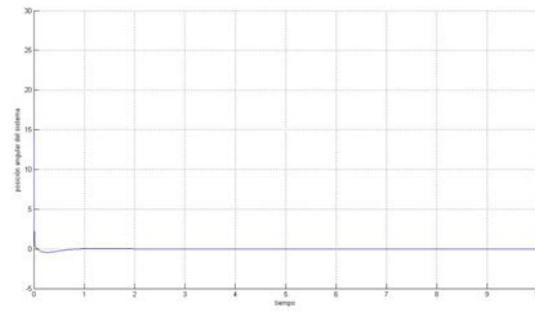


Figura 11: Posición angular  $\theta_1$  del sistema en lazo cerrado

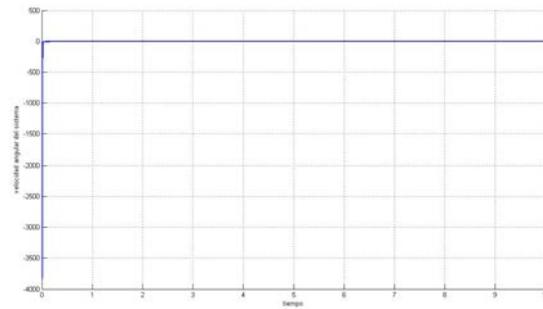


Figura 12: Velocidad angular  $\theta_2$  del sistema en lazo cerrado

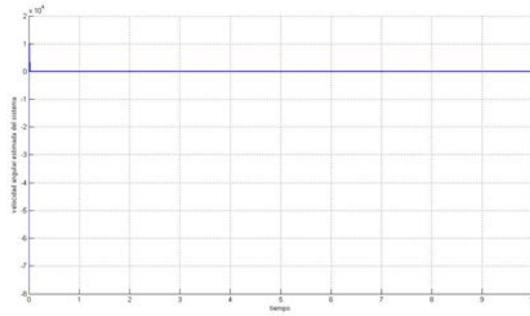


Figura 13: Velocidad angular estimada  $\hat{\theta}_2$  del sistema en lazo cerrado

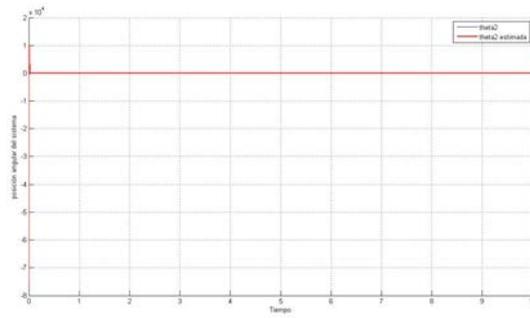


Figura 14:  $\theta_2$  y  $\hat{\theta}_2$

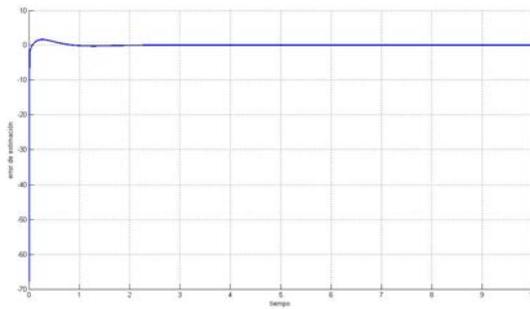


Figura 15: Error de sistema en lazo cerrado

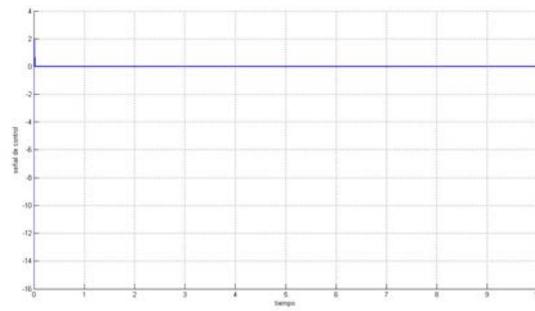


Figura 16: Señal de control para el sistema en lazo cerrado

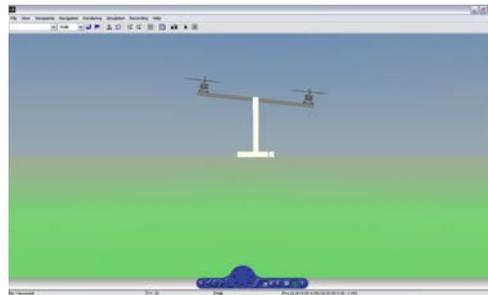


Figura 17: Transitorio de la estabilización del sistema

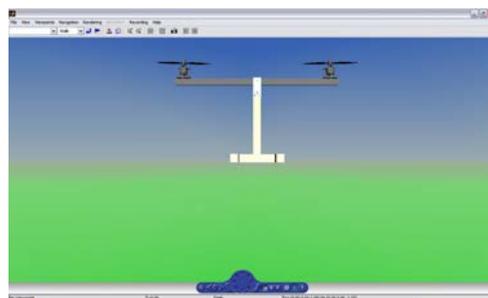


Figura 18: Sistema estabilizado



# Modelo de Autómata Celular 1-dimensional para una EDP hiperbólica

I. Huerta-Trujillo<sup>1</sup>, J.C. Chimal-Eguía<sup>1</sup>, S.V. Flores-Cortes<sup>1</sup>, and N. Sanchez-Salas<sup>2</sup>

*ihuerta@ipn.mx, chimal@cic.ipn.mx, sfloresco@ipn.mx,  
norma@esfm.ipn.mx*

<sup>1</sup> Centro de Investigación en Computación  
Av. Juan de Dios Bátiz s/n casi esq. Miguel Othón de Mendizábal, U. P. Adolfo  
López Mateos, Del. Gustavo A. Madero, C.P. 07738, México D.F.

<sup>2</sup> Escuela Superior de Física y Matemáticas  
Av. Instituto Politécnico Nacional Edificio 9, U. P. Adolfo López Mateos, Del.  
Gustavo A. Madero, C.P. 07738, México D.F.

**Abstract.** In this paper, we briefly describe the classic study of a vibrating string by partial differential equations whose evolution in time and under certain boundary conditions can have a very complicated analytical solution. However, we can see that by using a discrete model as a cellular automaton with simple transition rules we can study the same system without need to incorporate the intrinsic complexity of partial differential equations. Thus, we observed how the dynamics of the whole phenomenon is reflected in the simulation of collective effects of individual elements.

En este artículo, se presenta de manera breve el estudio clásico de una cuerda vibrante mediante ecuaciones diferenciales parciales hiperbólicas cuya evolución en el tiempo y bajo ciertas condiciones de frontera puede tener una solución analítica complicada. Sin embargo, podemos observar que mediante el uso de un modelo discreto como un autómata celular con reglas de transición simples se puede estudiar el mismo sistema sin necesidad de incorporar la complejidad intrínseca de las ecuaciones diferenciales parciales. Es decir, observamos como la dinámica de todo el fenómeno se traduce en la simulación de efectos colectivos de elementos discretos.

**Keywords:** simulación; cuerda vibrante; autómatas celulares; EDP hiperbólica; simulating; vibrant string; cellular automata; PDE hyperbolic.

## 1 Introducción

La simulación por computadora es una de las estrategias de las que dispone la ciencia actual para predecir sucesos en sistemas con un alto grado de complejidad. Esta complejidad es tal que la matemática clásica no resulta suficiente

para expresar todos los fenómenos que ocurrirán en el sistema en este sentido la simulación nos auxilia para ver los comportamientos de los sistemas deseados.

Shannon define el proceso de simulación como: *el proceso de diseñar un modelo de un sistema real y llevar a término experiencias con él, con la finalidad de comprender el comportamiento del sistema o evaluar nuevas estrategias -dentro de los límites impuestos por un cierto criterio o un conjunto de ellos - para el funcionamiento del sistema [7].*

La simulación consiste en la experimentación aplicada a los modelos del sistema o fenómeno que se quiere estudiar con el objeto de obtener información que ayude a entender este sistema. La simulación está fuertemente ligada a un proceso iterativo compuesto por el diseño del modelo, ejecución del modelo, análisis de la ejecución [5].

Frecuentemente los modelos matemáticos empleados para describir las vibraciones de un sistema físico incluyen ecuaciones diferenciales parciales no lineales, tradicionalmente se ha usado el método de elemento finito para resolver este tipo de problemas, la principal dificultad con dicho método se presenta al hacer el análisis para plantear la ecuación diferencial a resolver, la cual depende tanto de la geometría del sistema como de las condiciones de frontera y de las condiciones iniciales, [3].

Las ecuaciones diferenciales están intrínsecamente relacionadas con modelos continuos, al resolverlas numéricamente solo se resuelven sobre un punto o grupo de puntos de un conjunto compuesto por un continuo infinito[11], por lo que las soluciones resultan ser aproximadas.

Los modelos de escala celular o modelos discretos, consideran que el elemento (no el fenómeno en sí) juega un papel fundamental, dado que la dinámica de todo el fenómeno se observa en la simulación de efectos colectivos [9].

Curiosamente, los autómatas celulares tienen la capacidad de representar comportamientos complejos a partir de una dinámica sencilla. Debido a esto, desde su origen se les ha utilizado como elementos de la computación para la modelación de fenómenos biológicos y físicos. Además, los autómatas celulares son estudiados como objetos matemáticos debido al interés intrínseco relativo a los aspectos formales de su comportamiento

Los modelos basados en autómatas celulares han resultado apropiados en sistemas físicos con un régimen altamente no lineal, también en sistemas químicos y biológicos donde tienen lugar umbrales discretos [13].

Un autómata celular (AC), es un sistema dinámico discreto, que consiste en un arreglo de células (nodos) en alguna dimensión  $d$  [4]. Wolfram [12] los define como idealizaciones matemáticas de sistemas físicos, cuyo espacio y tiempo son discretos, en donde las cantidades físicas se pueden circunscribir a un conjunto finito de valores.

El concepto de AC lleva implícitamente asociado otros conceptos, como espacio y localidad de influencia. Se asume que el sistema representado está distribuido en el espacio y que regiones cercanas tienen mayor influencia entre sí, que otras que se encuentren apartadas dentro del sistema[1].

Para definir un AC es necesario definir otros elementos que lo integran, tales como:

- Una lattice (retícula) regular de  $N$  máquinas de estado finito idénticas, llamadas células [8], que cubre el espacio  $d$ -dimensional; cada una de las células cuenta con patrones idénticos y conexiones locales con otras células.
- Un conjunto de estados que pueden ser asignados a cada célula.
- Una regla de transición, que especifica la evolución en el tiempo de los estados.
- Una vecindad; las interacciones locales que toda célula tiene es con células que pertenecen solo a su vecindad (área de influencia).

Estas características son las que en conjunto dan forma a un AC.

### 1.1 Definición Formal de Autómata Celular

**Definición 11** Una “lattice” o “retícula” es un arreglo uniforme, generalmente infinito [12], formado por objetos idénticos llamados “células”. Este arreglo puede ser  $n$ -dimensional, pero para efectos de simulación de sistemas naturales se implementa de 1, 2 ó 3 dimensiones, de tamaño finito.

**Definición 12** Una vecindad para una célula  $c \in L$  es  $V(c) = \{k_0, k_1, \dots, k_n \mid k_j \in L, j = 0, 1, \dots, n\}$ ; es decir, una vecindad es un conjunto de células para las cuales la célula  $c$  es el punto de referencia para el área de influencia.

En el caso de AC 1-dimensionales, la vecindad para la  $i$ -ésima célula (la vecindad incluye a la célula  $i$ , sin embargo, la actualización por la función no depende del estado de  $i$  en el tiempo  $t$ ), puede estar compuesta por las células adyacentes a ésta en el lado derecho e izquierdo (se busca que esta vecindad respete la simetría respecto a la célula central, aunque no es obligatorio), por lo que se dice que esta vecindad es de radio  $r$ , es decir:

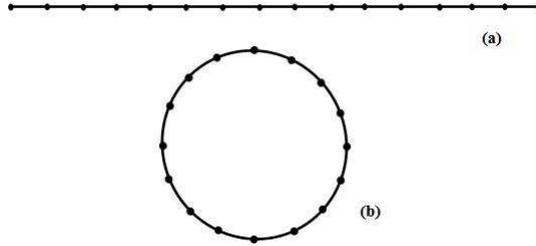
$$V(c) = \{c_{i-r}, \dots, c_{i-1}, c_i, c_{i+1}, \dots, c_{i+r}\}$$

La definición de lattice por si misma nos permite considerar lattices de tamaño infinito, pero en la práctica esta implementación resulta imposible, es por eso que los AC son representados como sistemas en espacios finitos, a estas condiciones que nos permiten limitar el espacio de operación del AC las llamamos *condiciones de frontera*. Para el caso de nuestro modelo se define una frontera cerrada (véase [3]).

**Definición 13** La función de transición en un AC 1-dimensional es local y se aplica simultáneamente de acuerdo a la siguiente ecuación:

$$c_j^{t+1} = \Phi(c_{j+\delta_1}^t + c_{j+\delta_2}^t + \dots + c_{j+\delta_q}^t) \quad (1)$$

donde  $j + \delta_k$  representa las células que conforman la vecindad de  $c_j$ .



**Fig. 1.** Representación de un AC 1-dimensional (a) con condiciones de frontera periódica (b).

**Definición 14** Un AC es una 4-tupla  $AC = (L, S, V, \Phi)$  donde:

- L*: Es una retícula regular y  $L = \{c \in \mathbb{C}^d\}$  para una lattice  $d$ -dimensional.
- S*: Es el conjunto finito de todos los posibles estados de las células,  $c \in L$ .
- V*: Es el conjunto finito de células que definen la vecindad para una célula.
- $\Phi$ :  $S^d \rightarrow S$ , es una función de transición aplicada simultaneamente a las células que conforman la lattice.

La actualización en el estado de las células requiere que se conozca el estado de las células vecinas, a esta región del espacio que la célula necesita conocer se le conoce como *vecindad* [3].

En este artículo, se presenta de manera breve el estudio clásico de una cuerda vibrante mediante ecuaciones diferenciales parciales cuya evolución en el tiempo y bajo ciertas condiciones de frontera puede tener una solución analítica muy complicada. Sin embargo, podemos observar que mediante el uso de un modelo discreto como un autómata celular con reglas de transición relativamente simples se puede estudiar el mismo sistema sin necesidad de incorporar la complejidad intrínseca de las ecuaciones diferenciales parciales. Es decir, observamos como la dinámica de todo el fenómeno se traduce en la simulación de efectos colectivos de elementos discretos. Este modelo de autómata celular para estudiar un modelo dinámico no tan simple como el de una cuerda vibrante, puede conducirnos a una moraleja muy interesante, a saber: ¿Podemos estudiar sistemas complejos, dinámicamente representados por ecuaciones diferenciales parciales, mediante sistemas discretos como los autómatas celulares?, si la respuesta es afirmativa, entonces nos encontramos frente a una nueva manera alternativa de estudiar sistemas físicos complejos (como las ecuaciones de Navier-Stokes) mediante sistemas discretos como los autómatas celulares, pero con el poder computacional actual como plataforma de trabajo.

El artículo está organizado de la siguiente manera: En la sección 1 se presenta una breve introducción de los autómatas celulares y del problema que se pretende resolver. En la sección 2, presentamos de forma breve el modelo analítico de la cuerda vibrante. En la sección 3, presentamos el modelo de un autómata celular

que representa a una cuerda vibrante, la sección 4 se presenta la definición del modelo de AC propuesto, la sección 5 . Finalmente en la sección 6, se presentan algunas conclusiones de este trabajo.

## 2 Modelo Tradicional

El modelo del sistema físico para una cuerda vibrante se representa a través de una ecuación diferencial parcial hiperbólica. Algunas formas de esta ecuación o una generalización de ella se obtienen siempre inevitablemente en cualquier análisis matemático de fenómenos que involucren propagación de ondas en un medio continuo [2].

Considérese una cuerda elástica perfectamente flexible, extendida firmemente entre soportes fijos al mismo nivel (Figura 2). El eje  $x$  está localizado a lo largo de la cuerda, con los extremos de esta sujetos, si la cuerda es puesta en movimiento en un tiempo inicial, vibrará libremente en un plano vertical libremente si la resistencia del aire es despreciada.

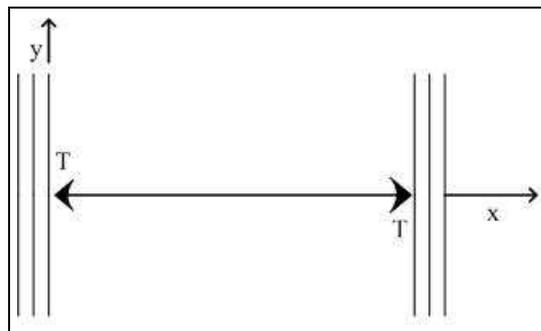


Fig. 2. Cuerda sujeta en ambos extremos.

El modelo matemático que rige el movimiento de esta cuerda está dado por la ecuación (2) y se conoce como ecuación de onda en una dimensión [6].

$$\frac{\partial^2 u(x, t)}{\partial x^2} - \frac{1}{c^2} \frac{\partial^2 u(x, t)}{\partial t^2} = 0 \quad (2)$$

Donde  $u(x, t)$  es el desplazamiento respecto a la posición de equilibrio y  $c^2 = T/\rho$ , con  $\rho$  = masa de la cuerda por unidad de longitud y  $T$  la tensión de la cuerda. Las condiciones de frontera del sistema son:

$$u(0, t) = 0 \quad \text{y} \quad u(l, t) = 0 \quad \forall t \geq 0 \quad (3)$$

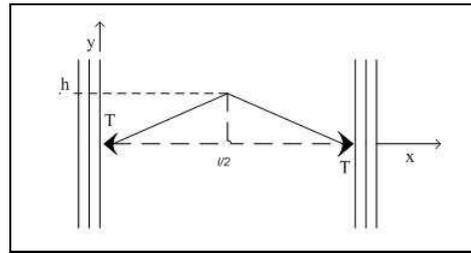
Es decir, los extremos se encuentran fijos en todo momento.

Las condiciones iniciales para el sistema son:

$$u(x, 0) = f(x) \quad \text{y} \quad \left. \frac{\partial u(x, t)}{\partial t} \right|_{t=0} = g(x) \tag{4}$$

Donde  $f(x)$  es la ecuación que marca la posición inicial de todos los puntos al inicio del análisis del sistema y  $g(x)$  es la velocidad inicial del sistema.

El caso que nos interesa en este momento es el caso para una cuerda de longitud  $l$  con una deflexión inicial triangular y velocidad inicial igual a cero (Figura 3).



**Fig. 3.** Cuerda con deflexión inicial triangular.

Las condiciones iniciales para el sistema son:

$$u(x, 0) = f(x) = \begin{cases} \frac{2h}{l}x & \text{para } 0 < x \leq \frac{1}{2}l \\ \frac{2h}{l}(l-x) & \text{para } \frac{1}{2}l < x < l \end{cases} \tag{5}$$

$$\left. \frac{\partial u(x, t)}{\partial t} \right|_{t=0} = g(x) = 0 \tag{6}$$

la solución que satisface a la ecuación (2) es:

$$u(x, t) = \frac{8h}{\pi^2} \sum_{n=1}^{\infty} \frac{1}{n^2} \sin\left(\frac{n\pi}{l}x\right) \sin\left(\frac{n}{2}\pi\right) \cos\left(\frac{n\pi}{l}ct\right) \tag{7}$$

Cabe resaltar que esta ecuación es una serie infinita y el resultado se aproxima si tendemos a tomar un número lo suficientemente grande de términos en la serie, y evaluar las derivadas parciales para cada punto del cual deseemos conocer su posición en cualquier instante de tiempo.

### 3 Modelo de Autómata celular para una cuerda vibrante

#### 3.1 Obtención de la regla de evolución

Supongamos que una cuerda es una sucesión de puntos con masa específica unidos por resortes (sistema masa-resorte), en el cual la masa de la cuerda se

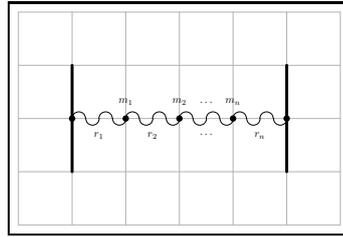


Fig. 4. Representación de una cuerda como un sistema masa-resorte.

encuentra distribuida en los puntos de unión y no en los resortes y sus extremos se encuentran fijos a una superficie (Figura 4).

Llamaremos *deq* o distancia de equilibrio, a la distancia de separación entre las masas, o longitud del resorte, que conforman la cuerda, cuando estas se encuentran en reposo.

Para el sistema de cuerda vibrante con deflexión inicial triangular (Figura 3) que se quiere estudiar, tomaremos las condiciones iniciales de posición descritas en la ecuación (5) y de velocidad descrita en la ecuación (6), de este modo tenemos que cada partícula interna de la cuerda, está sometida a dos fuerzas que actúan en dirección de los vectores  $\vec{\Delta r}_1$  y  $\vec{\Delta r}_2$  (Figura 5).

Es necesario conocer las fuerzas que se ejercen sobre la partícula  $m_i$  con el objeto de encontrar la velocidad con la que se va a mover y la posición final que tendrá dicha partícula transcurrido un tiempo  $t$ .

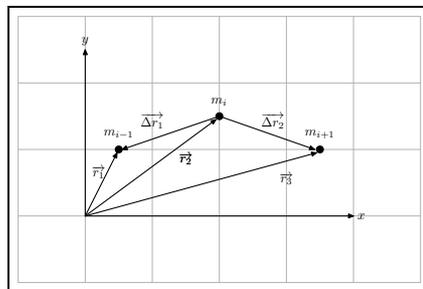


Fig. 5. Fuerzas ejercidas sobre la partícula  $m_i$  por sus vecinos.

Procedemos a encontrar  $\vec{\Delta r}_1$  y se procede de forma análoga para  $\vec{\Delta r}_2$ . Tomando en cuenta la Figura 5, vemos que:

$$\vec{r}_2 + \vec{\Delta r}_1 = \vec{r}_1 \Rightarrow \vec{\Delta r}_1 = \vec{r}_1 - \vec{r}_2 \tag{8}$$

El vector  $\overrightarrow{\Delta r_1}$  puede representarse como el producto de un vector unitario  $\widehat{\Delta r_1}$  de la misma dirección y sentido que el vector, multiplicado por el módulo de  $\overrightarrow{\Delta r_1}$  [10] entonces:

$$\overrightarrow{\Delta r_1} = |\overrightarrow{\Delta r_1}| \widehat{\Delta r_1}$$

Tomando en cuenta que el módulo de  $\overrightarrow{\Delta r_1}$ , representa la distancia de separación de las masas, se puede escribir este módulo como la suma de la distancia de equilibrio o longitud del resorte más la deformación del resorte debida al cambio de posición de la masa central, tenemos entonces que;

$$|\overrightarrow{\Delta r_1}| \widehat{\Delta r_1} = (deq + \Delta deq) \widehat{\Delta r_1}$$

Donde  $\Delta deq$  es el incremento presentado por el resorte de unión entre las masas  $m_i$  y  $m_{i-1}$  en dirección del vector unitario, es necesario encontrar este incremento a fin de conocer en incremento de la fuerza desde el punto de equilibrio al punto de análisis, por lo que tenemos,

$$|\overrightarrow{\Delta r_1}| \widehat{\Delta r_1} \Rightarrow \Delta deq \widehat{\Delta r_1} = (|\overrightarrow{\Delta r_1}| - deq) \widehat{\Delta r_1} \quad (9)$$

Ambos lados de la ecuación (9) son vectores, su representación en componentes son:

$$\Delta deq \widehat{\Delta r_1} = (\Delta x_1, \Delta y_1) \quad (10)$$

$$\left( |\overrightarrow{\Delta r_1}| - deq \right) \widehat{\Delta r_1} = \left( |\overrightarrow{\Delta r_1}| - deq \right) \left( \frac{x_{i-1} - x_i}{|\overrightarrow{\Delta r_1}|}, \frac{y_{i-1} - y_i}{|\overrightarrow{\Delta r_1}|} \right) \quad (11)$$

Dado que dos vectores son iguales si y solo si, sus componentes respectivas son iguales, entonces podemos igualar componente a componente del vector (10) con el vector (11), por lo que tenemos:

$$\Delta x_1 = \left( |\overrightarrow{\Delta r_1}| - deq \right) \frac{x_{i-1} - x_i}{|\overrightarrow{\Delta r_1}|} \quad (12)$$

$$\Delta y_1 = \left( |\overrightarrow{\Delta r_1}| - deq \right) \frac{y_{i-1} - y_i}{|\overrightarrow{\Delta r_1}|} \quad (13)$$

De ésta manera obtenemos las componentes  $\Delta x_1$  y  $\Delta y_1$  que son los incrementos de desplazamiento en los ejes  $X$  y  $Y$  de la partícula  $m_i$  para  $\overrightarrow{\Delta r_1}$ . Procedemos de la misma forma para encontrar  $\Delta x_2$  y  $\Delta y_2$  para  $\overrightarrow{\Delta r_2}$ .

Por *ley de Hooke* para un sistema masa resorte en una dimensión, tenemos que:

$$F = -k \Delta x \quad (14)$$

En general, para la partícula  $m_i$ , existen dos fuerzas ejercidas por  $m_{i-1}$  en dirección de  $\overrightarrow{\Delta r_1}$ , debido a las componentes  $x$  y  $y$  del vector, y dos fuerzas ejercidas por  $m_{i+1}$  en dirección de  $\overrightarrow{\Delta r_2}$ , entonces sustituyendo los valores encontrados

en la ecuación (12) y la ecuación (13) y sus análogos para  $\Delta x_2$  y  $\Delta y_2$  en la ecuación (14) obtenemos:

$$F_x = -k_1 \Delta x_1 - k_2 \Delta x_2$$

Suponiendo que los resortes que unen a las masas de la cuerda son exactamente iguales, entonces  $-k_1 = -k_2 = -k$  :

$$F_x = -k(\Delta x_1 + \Delta x_2) \quad (15)$$

Análogamente:

$$F_y = -k(\Delta y_1 + \Delta y_2) \quad (16)$$

Estas fuerzas son las que actúan sobre la partícula  $m_i$ , contribuyendo a su aceleración en el momento en que la partícula se encuentra oscilando y permiten acelerar a la misma. Utilizando la *segunda ley de Newton*  $\vec{F} = m\vec{a}$ , y la ecuación de velocidad para un movimiento uniformemente acelerado, tenemos que la velocidad final para una partícula está dada en función de su velocidad inicial y de la aceleración que siente la partícula en un instante de tiempo  $t$ , en consecuencia:

$$\begin{aligned} \vec{v}_f &= \vec{v}_i + \vec{a}t \\ &= \vec{v}_i + \frac{\vec{F}}{m}t \end{aligned}$$

Haciendo una separación de componentes y empleando los valores de fuerza encontrados en las ecuaciones (15) y (16), obtenemos las velocidades por componente para  $m_i$  que son;

$$v_{fx} = v_{ix} + \frac{F_x}{m}t \quad (17)$$

$$v_{fy} = v_{iy} + \frac{F_y}{m}t \quad (18)$$

En este punto hemos encontrado la velocidad que tendrá la partícula  $m_i$  transcurrido un tiempo  $t$ , esto nos proporciona la información necesaria para calcular la nueva posición de la partícula para el mismo instante de tiempo, para esto empleando los resultados de las ecuaciones (17) y (18) y la ecuación de desplazamiento uniformemente acelerado:

$$x_{fi} = x_i + v_{ix}t + \frac{1}{2} \frac{F_x}{m}t^2 \quad (19)$$

$$y_{fi} = y_i + v_{iy}t + \frac{1}{2} \frac{F_y}{m}t^2 \quad (20)$$

Las ecuaciones (17), (18), (19) y (20), son las que se emplean en la definición de la función de evolución para el AC propuesto en la siguiente sección.

### 4 Definición del Autómata Celular para una Cuerda Vibrante

Basado en la definición 14, el modelo de AC para un sistema de cuerda vibrante fijo en los extremos de una longitud  $l$  con una deflexión inicial triangular de altura  $h$  con  $0 < h \ll l$  es una 4-tupla  $AC = (L, S, V, \Phi)$  donde cada célula  $c \in L$  está definida por su masa, posición inicial y su velocidad inicial, cuando la cuerda está en reposo, siendo:

$L$ : Es una retícula regular y  $L = \{c_i \in \mathbb{C}^1 : i = 1, 2, \dots, n\}$

$S = \left\{ \begin{array}{l} \overrightarrow{P_{c_i}^t} : \text{vector de posición en } t \\ \overrightarrow{V_{c_i}^t} : \text{velocidad de } c \text{ en } t. \end{array} \right\} \forall c_i \in \mathbb{C}^i$

$V: V = \{(c_{i-1}, c_i, c_{i+1})\}$

$\Phi: \mathbb{R}^2 \rightarrow \mathbb{R}^2,$

$$\Phi = \left\{ \begin{array}{l} \text{a) } \overrightarrow{P_{c_i}^{t+1}} = \overrightarrow{P_{c_i}^t} + \overrightarrow{V_{c_i}^t}t + \frac{\overrightarrow{F_{i-1}^t} + \overrightarrow{F_{i+1}^t}}{2m}t^2 \\ \text{b) } \overrightarrow{V_{c_i}^{t+1}} = \overrightarrow{V_{c_i}^t} + \frac{\overrightarrow{F_{i-1}^t} + \overrightarrow{F_{i+1}^t}}{m}t \end{array} \right\}$$

donde:  $\overrightarrow{F_{i-1}^t} + \overrightarrow{F_{i+1}^t}$  es la fuerza que las células  $c_{i+1}$  y  $c_{i-1}$  ejercen sobre  $c_i$  en el tiempo  $t$ ;  $\overrightarrow{P_{c_i}^{t+1}}$ , es la posición final de la célula en el espacio. y  $\overrightarrow{V_{c_i}^{t+1}}$ , es la velocidad final de la célula en el tiempo  $t + 1$ .

La función de transición  $\Phi$ , está compuesta por dos reglas fundamentales, ambas reglas se aplican simultáneamente a todas las células que conforman la lattice.

La regla *a)* define la posición de la célula en el tiempo  $t + 1$ , tomando la velocidad en el tiempo  $t$ , esta posición se actualiza, siendo la nueva posición inicial para  $t + 2$  y así sucesivamente. Similarmente, para *b)* la velocidad final para el tiempo  $t + 1$  se actualiza, siendo la velocidad inicial para el tiempo  $t + 2$ .

#### 4.1 Obtención experimental de la constante $k_l$

Se toma como base para obtener la constante  $k_l$ , una cuerda de longitud  $l = 0.90m$  sin ninguna tensión aplicada y una masa  $m_c = 0.00507kg$ . De forma experimental, se sujeta la cuerda  $l$  en sus extremos, y aplicando una fuerza  $f = mg$ , en  $\frac{l}{2}$ . Esta fuerza cambiará al variar la masa  $m$ , como se muestra en la Figura 6.

donde:

- $l_e$ , es la longitud alcanzada por la cuerda al aplicarle la fuerza del peso.
- $h$ , la longitud que se desplaza el punto  $\frac{l}{2}$ , desde su posición de origen.
- $m$ , la masa variable aplicada en el experimento.

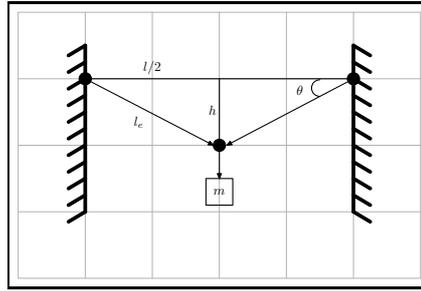


Fig. 6. Esquema experimental para obtener la constante  $k_l$  de la cuerda.

Realizado el ajuste por mínimos cuadrados obtenemos que  $k_l = 3963.63$ , que es la constante de restitución de la cuerda. Con este resultado, es posible calcular la constante  $k_c$  que se utiliza en el AC propuesto. Se tiene que:

$$k_l = \frac{1}{\frac{1}{k_{c1}} + \frac{1}{k_{c2}} + \dots + \frac{1}{k_{cn}}} = \frac{k_c}{n}$$

Lo que implica que:

$$k_c = n \cdot k_l \tag{21}$$

Es decir, la constante de restitución de un resorte (ecuación 21), para el modelo de AC propuesto, es directamente proporcional al número de resortes dentro del sistema por la constante de restitución de la cuerda.

## 5 Comparación ente el modelo tradicional y el modelo AC propuesto

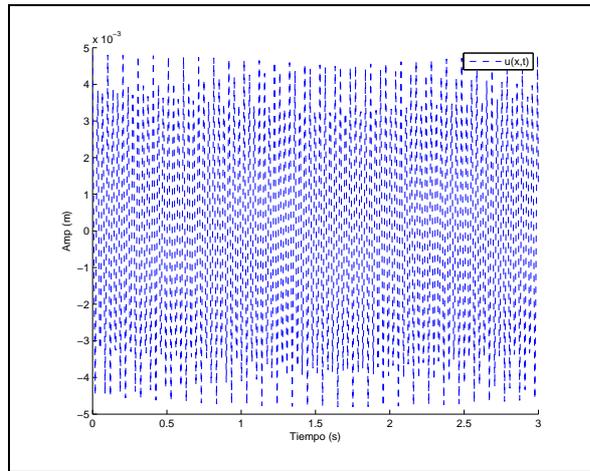
Para llevar a cabo la comparación entre el AC y la ecuación de movimiento encontrada (vea pag. 6) que satisface la ec. diferencial (2), es necesario definir las condiciones iniciales con las que se iniciará la ecuacion; pensando en una cuerda de guitarra, se tomarán las condiciones en las que vibra una cuerda sexta:

- $l$ : longitud de la cuerda  $0.90m$  la cual se estira  $4mm$ .
- $h$ : desplazamiento desde el origen del punto medio  $h = 5mm$ .
- $x$ : punto que se verifica su oscilación  $x = \frac{l}{2}$ .
- $T$ : tensión de la cuerda una vez estirada  $T = 15.84N$ .
- $m$ : masa de la cuerda  $masa = 0.00507kg$ .
- $\rho$ : densidad lineal de la cuerda  $\rho = \frac{m}{l}$ .
- $c$ : constante definida en la ecuación que rige el movimiento con  $c = \sqrt{\frac{T}{\rho}} = 53.144$ .

Sustituyendo estos valores en la ecuacion (7), tenemos:

$$u\left(\frac{l}{2}, t\right) = \frac{8 \cdot 0.005}{\pi^2} \sum_{n=1}^{\infty} \frac{1}{n^2} \sin\left(\frac{n\pi}{0.904} \frac{0.904}{2}\right) \sin\left(\frac{n}{2}\pi\right) \cos\left(\frac{n\pi}{0.904} 53.144 \cdot t\right)$$

Simulando para  $t = 0, \dots, 3$ , la ecuación  $u\left(\frac{l}{2}, t\right)$ , se genera la gráfica de la Figura 7, en la cual se puede ver un movimiento oscilatorio del punto  $x = \frac{l}{2}$ . En la gráfica de la Figura 8 se presenta el movimiento de la célula  $c_{\frac{l}{2}}$  del modelo del AC desarrollado.



**Fig. 7.** Gráfica generada por la ecuación de movimiento  $u\left(\frac{l}{2}, t\right) : t = 0, \dots, 3$ .

Haciendo una comparación de los modelos y sobreponiendo las gráficas sinusoidales se obtiene la gráfica de la Figura 9, puede observarse que tanto la trayectoria que sigue el punto  $\frac{l}{2}$ , correspondiente al modelo matemático, como la trayectoria de la célula  $c_{\frac{l}{2}}$ , tienen una correspondencia de fase.

Obteniendo los espectros de frecuencia para ambas señales y graficándolos, se aprecia que la gráfica de la Figura 10, los espectros de frecuencia de las dos señales se encuentran prácticamente empalmadas, observando la gráfica de la Figura 11, que presenta las frecuencias de la frecuencia fundamental y de las primeras dos armónicas es visible que hasta la segunda armónica se aprecia un ligero defasamiento de las gráficas.

En la gráfica de la Figura 12, puede observarse que existe una perfecta superposición de las frecuencias fundamentales de ambas señales la cual se encuentra al rededor de los  $30Hz$ .

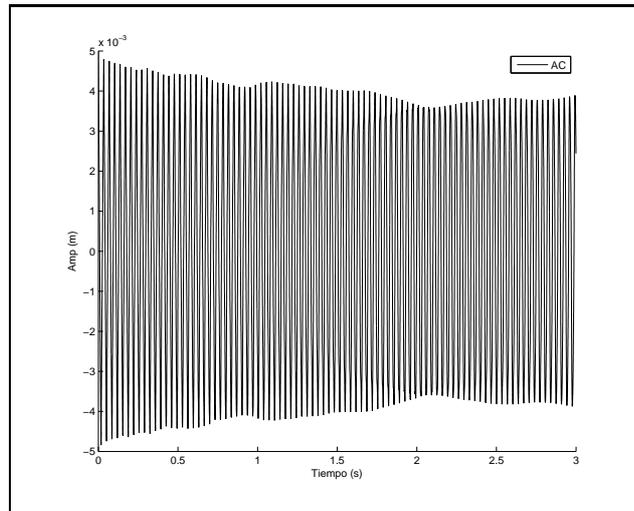


Fig. 8. Gráfica generada por el modelo de AC propuesto durante un tiempo  $t = 0, \dots, 3$ .

Matemáticamente, el primer término de la sumatoria de la ecuación (7),  $u(x, t)$ , cuando  $n = 1$ , define la frecuencia angular o fundamental de la señal por lo que  $\cos(\omega t) = \cos\left(\frac{\pi}{0.904} 53.144 \cdot t\right)$  siendo  $\omega$  la frecuencia fundamental; dado que  $\omega = \frac{2\pi}{T} = 2\pi f \Rightarrow f = \frac{\omega}{2\pi}$  donde  $T$  es el periodo y  $f$  la frecuencia. Sustituyendo datos tenemos:

$$f = \frac{\frac{\pi}{0.904} 53.144}{2\pi} = \frac{53.144\pi}{2\pi \cdot 0.904} = 29.4 Hz \quad (22)$$

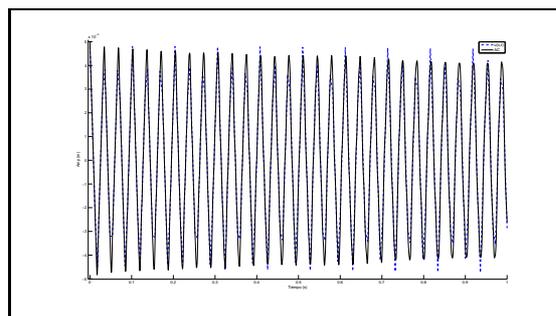
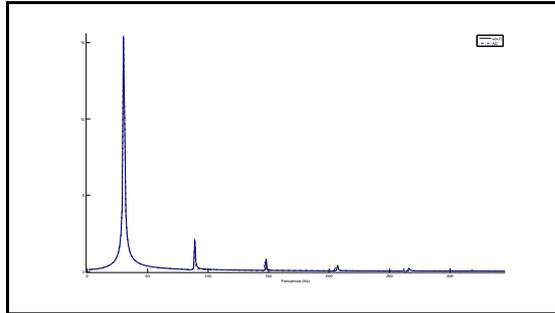
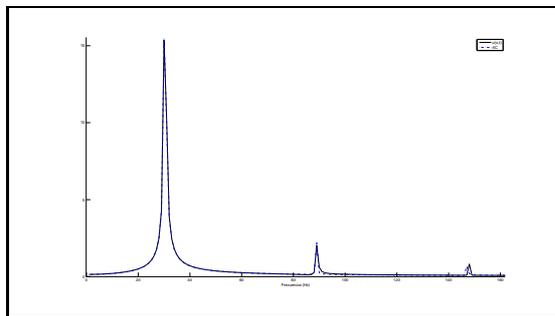


Fig. 9. Gráfica generada por la la superposición de las gráficas sinusoidales ampliada.



**Fig. 10.** Gráfica generada por la superposición de los espectros de frecuencia obtenidos de las señales del modelo matemático y el modelo AC.

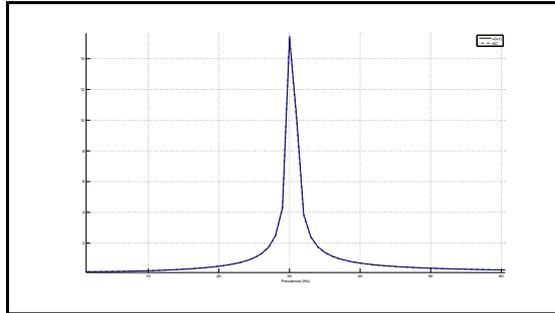


**Fig. 11.** Gráfica de frecuencia fundamental y dos primeras armónicas.

lo cual es congruente con la gráfica de espectros de frecuencia con la frecuencia fundamental.

## 6 Conclusiones

Cuando se estudian sistemas físicos a través de ecuaciones diferenciales parciales, usualmente debido a las no linealidades intrínsecas al modelo continuo, la solución de estos sistemas puede tornarse en extremo complicada. En este trabajo hemos mostrado como, modelando un sistema masa-resorte en el cual la masa de la cuerda se encuentra distribuida en los puntos de unión y no en los resortes, mediante un autómata celular con reglas de transición simples se producen los mismos resultados que el modelo analítico. Es importante notar que cuando tomamos parámetros de una cuerda real de una guitarra y los sustituimos en la ecuación (7) al graficar esta función que representa el desplazamiento lateral  $u(x, t)$  obtenida de manera analítica y la obtenida por el autómata celular (figura 9) los resultados son congruentes. Una manera de verificar la exactitud de ambos modelos fue encontrar el espectro de frecuencia de ambos modelos y obtener la frecuencia fundamental además de 2 armónicas principales. En la



**Fig. 12.** Gráfica de frecuencias fundamentales.

gráfica de la figura 12 se puede observar que la frecuencia fundamental se encuentra a los  $30\text{Hz}$ , lo cual corresponde de manera casi perfecta con el cálculo analítico de esta frecuencia fundamental. Así mismo, se puede ver en esta misma gráfica la concordancia entre las otras dos armónicas principales.

El modelo de AC para una cuerda vibrante resulta útil una vez que se cuentan con los datos necesarios para llevar a cabo una simulación ya que no está sujeto a las condiciones iniciales de la cuerda, como sucede en el caso de la ecuación diferencial que representa el movimiento, el AC solo es desarrollado una vez y puede asignarsele cualquier condición inicial.

Para el caso de la ecuación diferencial, es necesario el conocimiento de las condiciones iniciales de la cuerda que se pretende simular ya que el resultado de la solución depende de estos datos, por lo que el AC es una herramienta que puede ser igualmente efectiva que el modelo matemático o una alternativa a las ecuaciones diferenciales. A lo largo de las pruebas, el modelo resultó ser efectivo con respecto al modelo analítico por lo que la simulación de movimiento amortiguado es factible tomando en cuenta el factor de amortiguamiento según el medio en donde oscile la cuerda y la fricción que el medio ejerce sobre las células que conformen la cuerda. Cuerdas como la de los instrumentos musicales pueden ser simuladas con el AC, si se aplica define un coeficiente de fricción que amortigüe la amplitud de la oscilación, para esto es necesario tomar en cuenta el diámetro de la cuerda y la forma de las células para definir el coeficiente de amortiguamiento del aire.

No cabe ninguna duda que el éxito obtenido a lo largo de la historia del modelado de fenómenos físicos ha sido a través de ecuaciones diferenciales. Sin embargo en este trabajo se muestra que la dinámica de un fenómeno simulado computacionalmente como la suma de efectos colectivos de elementos discretos representados por un autómata celular, es parecida a la que puede generar el modelo analítico. Consideramos que esto es de mucha importancia ya que, en la actualidad existen muchos fenómenos físicos representados por ecuaciones diferenciales no parciales que podrían estudiarse utilizando modelos discretos, como los autómatas celulares, que si bien tienen reglas de transición simples utilizan el gran poder de computo que dan los procesadores actuales.

Es posible ampliar el modelo de AC propuesto para simular membranas en 3-dimensiones implementando una red de 2-dimensiones en el AC, e incluyendo un componente más en el vector de posición y de fuerza, empleando una vecindad de Moore para simular la cohesión de las células en la membrana.

Agradecimientos: Queremos agradecer al CONACyT su valioso apoyo para la realización de este trabajo. Así mismo los autores agradecen al EDI, y COFAA por su apoyo económico sin el cual este trabajo no se hubiera podido realizar.

## Referencias

1. Y. Bar-Yam. *"Dynamics of Complex Systems"*. Addison-Wesley, USA, 1997.
2. W.E. Boyce. *"Ecuaciones Diferenciales y Problemas con Valores en la Frontera"*. Ed. Limusa, México, 1990.
3. B. Chopard and et al. *"Cellular Automata Modeling of Physical Systems"*. Cambridge University Press, USA, 1998.
4. R. Espericueta. *"Cellular Automata Dynamics Explorations in Parallel Processing"*. Bakersfield College, USA, 1997.
5. P. Fishwick. "Simulation Model Design". *Proceeding of the Winter Simulation Conference*, 1995.
6. H. P. Hsu. *"Análisis de Fourier"*. Prentice Hall, México, 1998.
7. J. Johannes and R. Shannon. "Systems simulation: the art and science". *IEEE Transactions on Systems, Man and Cybernetics*, vol. 6(10):pp. 723–724., 1976.
8. M. Mitchell. "Computation in cellular automata: A selected review". *Nonstandard Computation*, pages 95–140, 1998. Weinheim: VCH Verlagsgesellschaft.
9. S. Motta and V. Brusici. *"Mathematical Modelling of the Immune System"*. Springer, USA, 2004.
10. M. R. Spiegel. *"Análisis Vectorial"*. Schaum's. McGraw Hill, México, 1991.
11. T. Toffoli. "Occam, Turing, von Neumann, Jaynes: How much can you get for how little? (A conceptual introduction to cellular automata)". *InterJournal, MIT*, December 1994. Cambridge, MA.
12. S. Wolfram. "statistical mechanics of cellular automata". *Reviews of Modern Physics*, (55):601–644, 1983.
13. S. Wolfram. "cellular automata as models complexity". *Nature*, (311):419–424, 1984.

## Author Index

### Índice de autores

Alarcón-Paredes Antonio	349		
Alvarado-Cervantes Rodolfo	79	Manrique Pablo	361
Álvarez Ballesteros Salvador	163	Mata-Villanueva José de Jesús	285
Arellano-Verdejo Javier	33	Montes Antonio	361
Argüelles-Cruz Amadeo	349	Moreno Escobar Jesús Jaime	17
Arroyo-Díaz Salvador	309	Muñiz-Montero Carlos	309
Barrón-Fernández Ricardo	33,389	Olivares Ceja Jesús Manuel	57
Calzada-Navarrete Viridiana Rubí	3	Olmedo-Aguirre José Oscar	195
Carreto Arellano Chadwick	163	Olvera Y.	107
Casillas-del-Llano Manuel-Iván	373	Orantes Aniceto	361
Chávez Pacheco Juan Ernesto	163	Orantes Luis	361
Chimal-Eguía J. C.	389,407	Pogrebnyak Oleksiy	237,349
Cortés Domingo	221	Ponce-Ponce Victor H.	309,361
De la Fuente-Arriaga José Abel	117	Priego Pérez Fausto Pavel	57
Duarte Calderon Alizari	149	Ramírez-Salinas Marco A.	285,361
Estrada Chávez Carlos D.	131	Ramírez-Colín Luis A.	269
Felipe-Riverón Edgardo M.	79,91	Rangel Tirado Pamela Araceli	237
	117,131	Ríos Miguel	337
Flores-Cortes S. V.	407	Rivera Aguilar Diana Naim	57
Galicia Galicia Pedro	181	Rivera Meraz Mariano	131
Gallegos Francisco	45	Rodríguez-Aguilar Rosa María	69
García-Martínez Edgar A.	33	Rosales Alberto	45
Garduño-Calderón Eduardo	117	Salas-Brito Álvaro-Lorenzo	373
Gelbukh Alexander	337	Salazar Victor	361
Godoy-Calderon Salvador	33	Sánchez de Antuñano Jorge Manuel	
Gómez Eric	361	Alejandro	69
Gómez-Conde Alejandro	285	Sánchez Fernández Luis Pastor	237
González García Luis G.	221	Sánchez-Gaspariano Luis A.	309
Gutierrez-Aldana Alfonso	251	Sanchez Meraz Miguel	149
Hernández Carlos	361	Sánchez-Rodríguez O.	389
Hernández-Terrazas Rubén O.	295	Sanchez-Salas N.	407
Huerta-Trujillo I.	407	Sandoval-Reyes Sergio	251
Jaime Fonseca Beatriz Adriana	17	Serrano Talamantes José Félix	57
Jaimez-González Carlos R.	205	Sosa Paz Carlos	149
	269,321	Torres-Huitzil César	3
Kazakov V.	107	Tortolero Agustín	45
Kinani Jean Marie	45	Trevera-Rodríguez José M.	309
Laureano-Cruces Ana Lilia	69	Van Horebeek Johan	131
Lezama-León Arturo	295	Ventura-Molina Elías	349
Liceaga-Ortiz-De-La-Peña		Villa-Vargas Luis A.	285
José Miguel	295	Villalobos-Castaldi Fabiola M.	91
Luna-Ramírez Wulfrano A.	205	Xicoténcatl-Pérez Juan Manuel	295



## **Editorial Board of the Volume**

### **Comité editorial de volumen**

Cornelio Yáñez Márquez  
Rolando Menchaca Méndez  
Jesús Guillermo Figueroa Nazuno  
Alexander Gelbukh  
Grigori Olegovico Sidorov  
Sergio Suárez Guerra  
Edgardo Manuel Felipe Riverón  
Osvaldo Espinosa Sosa  
Juan Carlos Chimal Eguía  
Marco Antonio Ramírez Salinas  
César Torres Huitzil  
Mario Arcadio Hernández Zarate  
Juan Manuel Xicoténcatl Pérez  
Sandra Dinora Orantes Jiménez  
Chadwick Carreto Arellano

Maricela Quintana  
Saturnino Job Morales Escobar  
Herón Molina Lozano  
Nareli Cruz Cortés  
Jorge Cortés Galicia  
Yulia Nikolaevna Ledeneva  
Elsa Rubio Espino  
Germán Tellez Castillo  
Hiram Francisco Calvo Castro  
Iliac Huerta Trujillo  
Abril Uriarte Arcia  
Erandi Castillo Montiel  
Amadeo José Argüelles Cruz  
Jorge Cortés Galicia

This volume contains 30 papers, which are grouped into nine thematic areas representative of the main current areas of interest within the computing science, such as:

- Artificial Intelligence
- Communications & Computer Networks
- Databases & Software Technology
- Digital Signal Processing
- Intelligent Processing of Geospatial Information
- Micro technologies & Embedded Systems
- Natural Language Processing
- Neural Networks & Unconventional Computation
- Simulation & Modeling

The volume will be useful for researches and students working in the respective areas of the computing science, as well as, for all readers interested in the most recent advances of this field of knowledge.

[www.cic.ipn.mx](http://www.cic.ipn.mx)

ISSN: 1870-4069



INSTITUTO POLITÉCNICO NACIONAL  
LA TÉCNICA AL SERVICIO DE LA PATRIA