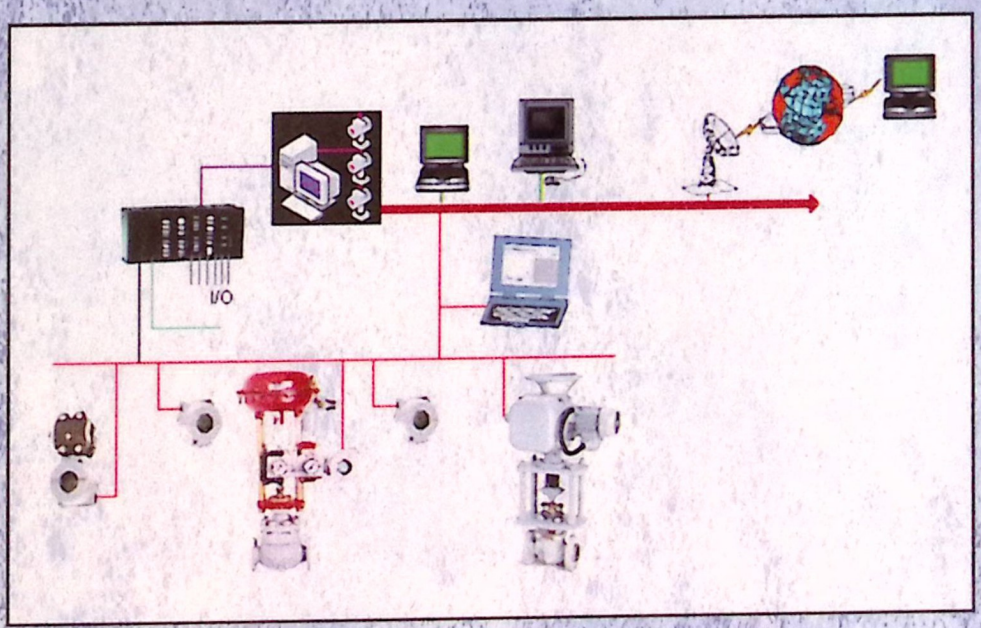
RESEARCH IN COMPUTING SCIENCE

ISSN: 1870-4069

Special issue: Industrial Informatics

Luis Pastor Sánchez Fernández
Oleksiy Pogrebnyak
Elsa Rubio Espino
(Eds.)



Vol. 31

RCS
Research in Computing Science

Special Issue: Industrial Informatics

Research in Computing Science

Series Editorial Board

Comite 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 (Frane)
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:
Coordinacion Editorial

Blanca Miranda Valencia

Formating:
Formacion

Luis Pastor Sánchez Fernández
Oleksiy Pogrebnyak
Elsa Rubio Espino

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 31,** Noviembre, 2007. Tiraje: 500 ejemplares. *Certificado de Reserva de Derechos al Uso Exclusivo del Título* No. 04-2004-062613250000-102, 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 Responsable: *Juan Humberto Sossa Azuela, SOAJ 560723*

***Research in Computing Science** is published by the Center for Computing Research of IPN. **Volume 31,** November, 2007. 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, October, 2007, in the IPN Graphic Workshop – Publication Office.*

Volume 31

Volumen 31

Special Issue: Industrial Informatics

Volume Editors:
Editores del Volumen

Luis Pastor Sánchez Fernández
Oleksiy Pogrebnyak
Elsa Rubio Espino

Instituto Politécnico Nacional
Centro de Investigación en Computación
México 2007

ISSN: 1870-4069

Copyright © Instituto Politécnico Nacional 2007
Copyright © by Instituto Politécnico Nacional 2007

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

The editors and the Publisher of this journal have made their best effort in preparing this special issue, but make no warranty of any kind, expressed or implied, with regard to the information contained in this volume.

All rights reserved. No part of this publication may be reproduced, stored on a retrieval system or transmitted, in any form or by any means, including electronic, mechanical, photocopying, recording, or otherwise, without prior permission of the Instituto Politécnico Nacional, except for personal or classroom use provided that copies bear the full citation notice provided on the first page of each paper.

Indexed in LATINDEX and PERIODICA
Indexada en LATINDEX y PERIODICA

Printing: 500
Tiraje: 500

Printed in Mexico
Impreso en México

Preface

Computer sciences have experimented enormous advances in recent years, that is the reason to include in this volume of *Research in Computing Science* several contributions obtained by specialists in areas such as: Automation, Virtual Instrumentation and Signal Processing, Real Time Systems, Infrastructure, Technology and Data Industrial Network, Enterprise Management and Applications of Industrial Informatics, specialized areas that are having a high impact on contemporary societies.

We received several papers from researches of many parts of the world. We accepted only the best papers. Every paper was evaluated by three reviewers with the common objective of select the papers with the highest quality.

We want to express sincerely acknowledge to all the people that contribute with this magazine. We are specially indebted with authors, editors, reviewers and program committee members for their outstanding work and advice. Finally We would also like to thank to the National Polytechnic Institute for facilities given in order to achieve the success of this special issues, which we believe, is so much important for the academic and scientific life in this country.

November, 2007

Luis Pastor Sánchez Fernández
Oleksiy Pogrebnyak
Elsa Rubio Espino

Scientific Committee

Abelardo del Pozo	Institute of Cybernetics, Mathematics & Physics (ICIMAF). Cuba
Alberto Aguado	Institute of Cybernetics, Mathematics & Physics (ICIMAF). Cuba
Alfonso Gutiérrez Aldana	Center for Computing Research, National Polytechnic Institute (IPN). Mexico
Amaury Caballero	Florida International University. USA
Ana Isabel González Santos	Higher Polytechnical Institute "José Antonio Echeverría". Cuba
Anatoliy Popov	Aerospace National University " Kharkov Aviation Institute " of Jarkov, the Ukraine.
Ángel Costa Montiel	Higher Polytechnical Institute "José Antonio Echeverría". Cuba
Carlos Aguilar	Center for Computing Research, National Polytechnic Institute. Mexico
Carlos Canudas de Wit	Institut National Polytechnique de Grenoble (INPG). France
Cesar de Prada Moraga	University of Valladolid, Spain
Cornelio Yañez	Center for Computing Research, National Polytechnic Institute. Mexico
Domingo Cortés Rodríguez	Center for Computing Research, National Polytechnic Institute. Mexico
Eduardo B. Fernandez	Florida Atlantic University. USA
Eduardo Quiles Cucarella	Polytechnic University of Valencia. Spain.
Elsa Rubio Espino	Center for Computing Research, National Polytechnic Institute. Mexico
Emilio García Moreno	Polytechnic University of Valencia. Spain.
Eugenio César Sánchez	Higher Polytechnical Institute "José Antonio Echeverría". Cuba
Felipe Espinosa Zapata	University of Alcalá de Henares, Spain
Humberto Ferasoli Filho	São Paulo State University - UNESP.

	Brazil
Jeff Meisel	Academic Accounts. National Instruments Mexico
João E. M. Perea Martins	São Paulo State University - UNESP. Brazil
José de Jesús Medel	Center for Computing Research, National Polytechnic Institute. Mexico
José G. Rodríguez	CINVESTAV. Mexico
José Luis Verdegay	Granada University. Spain
Kang Yen	Florida International University. USA
Lázaro Gorostiaga Canepa	Center of Automatization, Robotics and Technologies of the Information and the Manufacture. Spain
Luis Enrique Ramos Velasco	Universidad Autónoma del Estado de Hidalgo
Luis Miguel Fernández	Institute of Cybernetics, Mathematics & Physics (ICIMAF). Cuba
Marcelo Nicoletti Franchin	São Paulo State University - UNESP. Brazil
Marco A. Moreno Armendariz	Center for Computing Research, National Polytechnic Institute. Mexico
Miguel García Silvente.	Granada University. Spain
Norberto Flores Guzmán	Mexican Institute of Petroleum (IMP). Mexico
Oleksiy Pogrebnyak	Center for Computing Research, National Polytechnic Institute. Mexico
Orestes Llanes Santiago	Higher Polytechnical Institute "José Antonio Echeverría". Cuba
Oscar Camacho Nieto	Center for Computing Research, National Polytechnic Institute. Mexico
Pedro Castillo García	Centre de Recherches de Royallieu 60205 Compiègne. France
Pedro Guevara	Research Center for Applied Science and Advanced Technology (CICATA). National Polytechnic Institute. Mexico
Raynitchka Tzoneva	Cape Peninsula University of

	Technology. Republic of South Africa
Roger Misa Llorca	Higher Polytechnical Institute "José Antonio Echeverría". Cuba
Romeo Ortega.	Ecole Supérieur d'Electricité. France
Sergio Suárez Guerra	Center for Computing Research, National Polytechnic Institute. Mexico
Valery Moreno Vega	Higher Polytechnical Institute "José Antonio Echeverría". Cuba
Virgilio López Morales	Autonomous University of Hidalgo State. Mexico
Volodymir Lukin	Aerospace National University "Kharkov Aviation Institute" of Jarkov, the Ukraine.
Volodymir Ponomaryov	Aerospace National University "Kharkov Aviation Institute" of Jarkov, the Ukraine.
	IPN, ESIME Culhuacan, México

Table of Contents

Virtual Sensors. An overview *(invited paper)* 1
Pedro Albertos, I. Peñarrocha, R. Sanchis

Modeling and Manipulating 3D Datasets through the Extreme Vertices Model in the n-Dimensional Space (nD-EVM) 15
Ricardo Pérez -Aguila

Design A Perceptron Neural Network Block With Complementary Training Function For Industrial PLCs By Ladder Language 25
Hamid Abdi, Abolfazl Salami, Abolfazl Ahmadi, Majid Abdi

Effective Threshold Control in a Coded Communication System with Information Feedback Driven by Gaussian and Impulsive Noises 33
Francisco Tejeda, Vladimir Kazakov

Fuzzy Directional Adaptive Recursive Temporal Filter for Noise Suppression of Video Color Sequences 43
Alberto Rosales-Silva, Volodymyr Ponomaryov, Francisco Javier Gallegos-Funes, Oleksiy Pogrebnyak

Adaptability of the Information Systems to the changes in the Business Processes 53
Sandra Dinora Orantes Jiménez, Maximo Lopez, Agustin Francisco Gutierrez Tornes

PID control under sampling period constrains 63
Alfredo Cuesta, Luis Grau, Ignacio López

Speed control of a sugar cane belt conveyor with a fuzzy controller 73
Mercedes Ramírez, Modesto Angulo, Ariel Domínguez, Pedro Albertos

Prefiltering and Robustness of GPC with Structured Perturbations 85
Carolina Mañoso, Angel P. de Madrid, Roberto Hernandez, Miguel Romero

Industrial Temperature Control using an AFLC 95
Nitin Patil, Rajan Chile, L. M. Waghmare

Preliminary Processing and Lossy Compression of Multichannel Information Data 105
Vladimir Lukin, Mikhail Zriakhov, Anatoliy Popov, Oleksiy Pogrebnyak

Unsupervised Remote Sensing Data Classification Using Multimodal Statistical Model 115
Anatoliy Popov, Oleksiy Pogrebnyak, Alexandra N. Brashevan

IIR lifting DWT for Lossless Image Compression 125
Oleksiy Pogrebnyak, Pablo Manrique Ramírez, Luis Pastor Sánchez Fernández

Attitude Subsystem Development for an Educative Satellite Based on Reaction/Momentum Wheel and Magnetic Torquing Coils 133
Esaú Vicente Vivas, Emilio Jiménez Madrigal, Rodrigo Alva, Rodrigo Córdova

Diagonal Recurrent Neural Networks for Speed Control of a DC motor .. 143
Alfonso Noriega, Carlos Silva, Jesús Pichardo

A Modified Rete Match Algorithm for Predicate Management in Real Time Planning and Execution Systems 153
Marcelo Franchin

Flight Computer for a Human Resources Training System in Small Satellite Technology 163
Esaú Vicente Vivas, Emilio Jiménez Madrigal, Francisco Gómez Islas, Zaira Carrizales Muñoz

Aircraft Noise Evaluation: First Stage 173
Arturo Rojo, Luis Pastor Sánchez Fernández, Luis Alejandro Sanchez

Color Image Segmentation Using Relation of Equivalence 183
Pablo Manrique Ramírez, Oleksiy Pogrebnyak

Software Architecture Implementation of an e-HUB to offer e-Services for SMEs 193
Javier Espadas, David Concha, Teresa Najera, Nathalie Galeano, David Romero, Arturo Molina

FPGA Implementation of Turbo Codification Technique for Error Control in Communication Channels 203
Pablo Manrique Ramírez, Rafael Antonio Márquez Ramírez, Oleksiy Pogrebnyak

Experiences concerning the application of electric machines steady state models for diagnostic purposes 213
Angel Costa Montiel, Miriam Vilaragut Llanes, Rodrigo Lopez Cardenas

Development of Mobile Robots - A Proposal of an Open and Flexible Control Architecture 223
Wilson Yonezawa, Humberto Ferasoli, Renê Pegoraro, Marco Caldeira, João Rosário

Nonlinear Modelling of Water Quality in Shrimp Ponds based on Artificial Neural Networks 233
José Carbajal Hernández, Luis Sánchez Fernández and Jorge Hernández

Mechanical Faults Diagnosis in Induction Motors based on Neural Networks and Vibration Patterns 243
Beatriz Jaime, Luis Pastor Sánchez Fernández, Rodrigo López Cárdenas

Two Reconstruction Algorithms of the Process on the Output of the Exponential Converter 253
Vladimir Kazakov, Monica Sedeno

Robust Telecontrol with Output Feedback 261
Luis Gonzalez, Jose Rodriguez

Author Index 271

Virtual Sensors. An overview

P. Albertos[1], I. Peñarrocha[2], R. Sanchis[2]

[1] Departamento de Ingeniería de Sistemas y Automática
Polytechnic University of Valencia, Spain.
pedro@aii.upv.es
[2] Departament d'Enginyeria de Sistemes Industrials i Disseny
Universitat Jaume I de Castelló, Spain.
{ipenarro,rsanchis}@esid.uji.es

Abstract. A virtual sensor is an information-processing based device designed to get information about an internal process variable which is not directly accessible. It may receive information from several sensors as well as the sequence of process inputs, updated following a given pattern. By using a model of the process, the desired variables are estimated at the desired rate. Virtual sensors can be used to monitor the evolution of some internal variables, some quality indices or some parameters in the process. Moreover, the estimated variables can be also used to feed a control system or to provide some sort of redundancy in the measurements. In this way, richer feedback control loops can be implemented or more reliable redundant control systems can be designed. In this plenary talk, some of the main issues in the design of virtual sensors are discussed and their main features, such as robustness or convergence, are analyzed. Some illustrative applications illustrate the main concepts.

1 Introduction

The main goal of a sensing device is to provide reliable, precise, up-to-date and concise information about a process variable, parameter or index, to be used for different monitoring or control purposes. The basic *transducer* device converts a variable into another one which is easier to handle. It is also typical to get an improvement in the sensed variable by correcting this basic measurement by means of related information based on the characteristics of the transducer (non-linear compensation) or the environment (temperature compensation).

Modern computer-based systems allow for more versatile sensors providing such information without the need for extra hardware. Additional features such as device identification, change in the range of operation, availability, storage of previous values and so on are also typical in modern sensors. This leads to the concept of *smart sensors*, that is, sensors devices providing not only the value of the sensed variable (one of the possible optional measurements) but also data about the operating conditions, location of the device and quality of the measurement. This data allow for a better management of the full data acquisition system.

© L. Sánchez, O. Pogrebnyak and E. Rubio (Eds.)
Industrial Informatics.
Research in Computing Science 31, 2007, pp. 1-14

The use of AI techniques also allows for some sort of intelligence in performing these tasks. As a result, these devices are sometimes known as *intelligent sensors*, being also able to validate, communicate and report the sensed variable and reconfigure its operation based on additional knowledge about environmental conditions.

The common features of advanced sensors include: to convert the sensed variable into a treatable signal (in digital form), to cover a wide range of interest, to provide the best measurements in terms of accuracy, timing (conversion time, delays), disturbances (noise and drifts) or linearity, to prevent maintenance and detect malfunctioning, and to communicate with the user as well as the rest of components.

Most of these devices, originally working in continuous time, are conceived to operate under regular sampling/updating patterns.

The concept of a *virtual sensor* is rather different. The estimated value of the variable is computed as a result of the action of the process inputs on a model of the process, combined with additional information taken from the process. In this way, the sensed variable is not directly measured, it may be even not accessible or physically measurable, but there exists a model relating its time evolution with that of the available variables. Virtual sensors may also incorporate some of the features already mentioned for smart or intelligent sensors. Virtual sensors can be used to monitor the evolution of some internal variables, some quality indices or some parameters in the process. Moreover, the estimated variables can be also used to detect faulty conditions, to feed a control system or to provide some sort of redundancy in the measurements. In this way, richer feedback control loops can be implemented or more reliable redundant control systems can be designed.

Moreover, the constraint of regular sampling/updating pattern may be removed in such a way that the measurements are gathered at different time instants. Figure 1 represents a continuous-time process manipulated by several continuous-time inputs ($u(\tau), \tau \in \mathbb{R}$, measurable, and $v_c(\tau)$, unmeasurable) and several output signals ($y(\tau)$) that can be measured with different rates and at different time instants. Some of them could be measured with redundant sensors of different precision, different sampling rates and different associated delay, leading to the discrete-time measurement values $m_{i,k}$, where i refers to the sensor number.

Assume that the values of some internal variables $z(\tau)$, which are not directly measurable but are a function of the measurable ones, are needed for monitoring or control purposes. Moreover, assume they are required at a fixed frequency on predefined time instants. In this scenario, a virtual sensor is defined as the information-processing device that based on the knowledge of the applied inputs and the measurement of the available outputs at arbitrary instants, predicts the values of the desired signals at the required time instants [1].

Among the virtual sensors applications the following cases can be considered: a) the process variable cannot be directly measured, i.e. measuring some performance; b) the physical sensor is too expensive, not enough accurate or too slow, like a chromatograph; c) the sensor placement is not accessible, like the

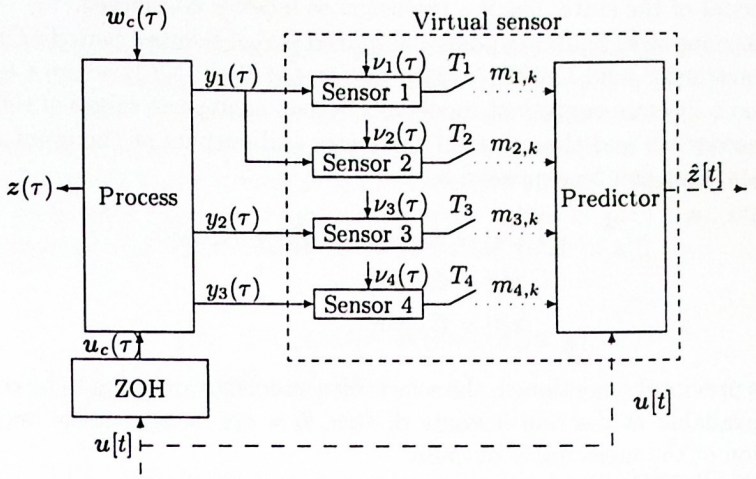


Fig. 1. General setting for internal knowledge estimation in MIMO process with several sensors.

burning temperature in a cement kiln; or d) it should be too far away of the place of interest, like the moisture of the paper web at the headbox exit. Other situations may also require a virtual sensor if, for instance, there is no room for a physical device or it is very expensive to maintain. Some applications of virtual sensors are reported, i.e. in [2] and [3].

Several approaches can be used for this prediction. The simplest one is to interpolate or extrapolate the measurements with any interpolation method. This idea could result in a computationally simple predictor that can be applied if the targeted variables are measurable. However, if the frequency of the measurements is not high enough the predictions could have a large error. In these situations the predictions can be significantly improved if a model of the process is used and the inputs as well as the output measurements are taken into account in the predictor algorithm.

The model of the process can be identified off line or can be estimated online by an adaptive algorithm running on the virtual sensor. In this work the general case of unmeasured output prediction is studied, where several signals that are related to the output are assumed to be measured at different instants (the desired output may be one of them, but not necessarily).

The process is assumed to be a continuous linear time-invariant system modeled by the equations

$$\dot{\boldsymbol{x}}(\tau) = \boldsymbol{A}_c\,\boldsymbol{x}(\tau) + \boldsymbol{B}_c\,\boldsymbol{u}_c(\tau) + \boldsymbol{B}_{wc}\,\boldsymbol{w}_c(\tau), \tag{1a}$$

$$\boldsymbol{y}(\tau) = \boldsymbol{C}_y\,\boldsymbol{x}(\tau), \tag{1b}$$

where $\boldsymbol{x} \in \mathbb{R}^n$ is the state, $\boldsymbol{u} \in \mathbb{R}^{n_u}$ is the input vector, $\boldsymbol{w}_c(\tau) \in \mathbb{R}^{n_w}$ is the continuous time disturbance vector and $\boldsymbol{y} \in \mathbb{R}^{n_y}$ is the measured output vector. The vector of targeted variables can be written as

$$\boldsymbol{z}(\tau) = \boldsymbol{C}_z\,\boldsymbol{x}(\tau), \tag{1c}$$

where $z(\tau) \in \mathrm{R}^{n_z}$. In the general case, $z(\tau)$ can be a nonlinear function or a functional of the state, like if a performance index is considered.

If the input signals are updated at a fixed period (control period) T by means of a zero order hold (ZOH) ($u(\tau) = u[t], \tau \in [tT, tT + T)$) where $t \in \mathbb{N}$, then there is a discrete equivalent model of (1) that relates the values of the discrete input sequence and the values of the states and outputs at the input updating instants, that can be expressed as

$$x[t+1] = A\,x[t] + B\,u[t] + B_w\,w[t], \tag{2a}$$

$$y[t] = C_y\,x[t], \tag{2b}$$

$$z[t] = C_z\,x[t]. \tag{2c}$$

As previously mentioned, the sensor measurements are going to be considered only available at discrete instants of time $\tau = \tau_k$, being a noisy and delayed function of the measurable outputs:

$$m_{i,k} = c_i\,x(\tau_k - \delta_{i,k}) + \nu_{i,k}, \quad i = 1, \ldots, n_m \tag{3}$$

where $m_{i,k}$ is the available measurement of sensor i at the k-th sampling (not all sensor measurements are available at each sampling time), with delay $\delta_{i,k}$ and measurement noise $\nu_{i,k}$, and n_m is the number of sensors. The availability factor $\alpha_{i,k}$ defined as

$$\alpha_{i,k} = \begin{cases} 1, \text{ if } m_{i,k} \text{ is available,} \\ 0, \text{ if } m_{i,k} \text{ is not available,} \end{cases} \tag{4}$$

indicates which are the available measurements at the k-th sampling instant. If some of the sensors have their own non negligible dynamics, the last representation is also valid just by using some of the process states as sensor states (see [4,5]).

If the sensor measurements are assumed to be only available at instants $t = t_k$ (synchronously with the input update) and are affected by a variable time delay, the measurement equation (3) can be written

$$m_{i,k} = c_i\,x[t_k - d_{i,k}] + \nu_{i,k}, \quad i = 1, \ldots, n_m \tag{5}$$

where $d_{i,k}$ is the discrete delay (measured in number of control periods) assigned to sensor i in the k-th sampling.

The rest of this paper is structured as follows. In the next section the basic structure of a virtual sensor is described. The simplest and more useful situations are analyzed and the sensor properties are summarized. For the sake of clarity, all the treatment is done in discrete time. Continuous time and random occurrence of measurements and input updating considerations introduce additional complexity in the treatment [5]. The general case of arbitrary sampling/updating pattern is considered in Section 3. A Kalman filter approach, which is also valid in the previous cases, is generalized in Section 4. Then, a typical application is presented and some conclusions are drafted in the final Section.

2 Virtual sensor structure

In the following, only discrete time systems are considered. Assume that some variable values are required at each input updating, $\boldsymbol{z}[t]$ ($\boldsymbol{z}[t] = \boldsymbol{z}(tT)$), but they are not available. Instead, the values of some of the elements of the vector $\boldsymbol{y}(\tau)$ are measured at different sampling instants. The virtual sensor is a dynamic system similar to (2) whose inputs are the input action at rate T and an *innovation* term $\mathbb{J}$ elaborated from the irregularly sampled values of $\boldsymbol{y}(\tau)$ (i.e., $m_{i,k}$) and their estimates, and whose output is the predicted value of $\boldsymbol{z}[t]$.

$$\hat{\boldsymbol{x}}[t+1] = \boldsymbol{A}\,\hat{\boldsymbol{x}}[t] + \boldsymbol{B}\,\boldsymbol{u}[t] + \mathbb{J}(\boldsymbol{m}, \hat{\boldsymbol{y}}), \tag{6a}$$

$$\hat{\boldsymbol{y}}[t] = \boldsymbol{C}_y\,\hat{\boldsymbol{x}}[t], \tag{6b}$$

$$\hat{\boldsymbol{z}}[t] = \boldsymbol{C}_z\,\hat{\boldsymbol{x}}[t]. \tag{6c}$$

The innovation term should be designed to assure some properties of the predictor. In particular, the sensing error

$$\tilde{\boldsymbol{z}}[t] = \boldsymbol{z}[t] - \hat{\boldsymbol{z}}[t] \tag{7}$$

should be small and converge to zero under nominal and constant operating conditions, regardless the stability condition of the plant (the plant matrix A not necessarily being Hurwitz). Also, the sensitivity of this error with respect to measurement noises and plant disturbances should be bounded.

Depending on the sampling scenario, different prediction algorithms can be used to estimate the state and the targeted variables at the input updating instants.

2.1 Conventional sampling

The simplest case corresponds to a regular sampling/updating of all the signals, with a period T and without any delay in any of the measurement signals. The innovation term is made proportional to the output estimation error, i.e.

$$\mathbb{J}(\boldsymbol{m}, \hat{\boldsymbol{y}}) = \boldsymbol{L}(\boldsymbol{m} - \boldsymbol{C}\hat{\boldsymbol{x}}), \tag{8}$$

where $\boldsymbol{C}$ is the matrix formed with the rows $\boldsymbol{c}_i$ that define each sensor equation (5). As it is well known [6], the error dynamics is characterized by the eigenvalues of $\boldsymbol{A} - \boldsymbol{L}\,\boldsymbol{C}$.

In this regular case, this innovation sequence can be filtered by a discrete transfer matrix $\boldsymbol{L}(z)$. The virtual sensor can be represented in block diagram as depicted in Figure 2, where it has been assumed that $n_m = n_y$.

As already mentioned, by $\boldsymbol{L}(z) = \boldsymbol{L}$ a constant matrix, the stability of the sensor (like in any Luenberger-type observer) can be ensured, if the pair (A, C) is observable. By the appropriate selection of the *filter* $\boldsymbol{L}(z)$, the sensitivity functions relating the estimated output with respect to noise and plant disturbance can be fine tuned (see, i.e., [4] and the references therein).

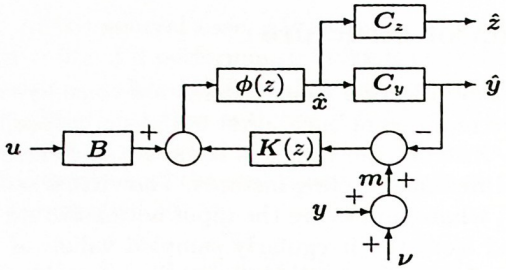


Fig. 2. Innovation in virtual sensor.

Denoting by $\phi(z) = (zI - A)^{-1}$, the noise sensitivity function is

$$T_{\hat{z}\nu} = C_z \left(\phi(z)^{-1} + L(z)C\right)^{-1} L(z). \tag{9}$$

An analysis of the shape of the frequency response of this function will enlighten the possible options in selecting L.

2.2 Nonconventional regular sampling

Assume now that all the measurements are available without delay periodically at the same instants of time, being N the number of input updates between two consecutive measurement instants ($N = t_k - t_{k-1}$). In this case, the innovation term (8) is only applied at sampling instants, leading to the prediction error dynamics (in the absence of disturbances):

$$\tilde{x}_{k+1} = \mathcal{A}\tilde{x}_k = (I - LC)\, A^N \tilde{x}_k, \tag{10}$$

$$\tilde{z}_k = C_z \tilde{x}_k, \tag{11}$$

The state estimation error is defined as $\tilde{x}_k = x[t_k] - \hat{x}[t_k]$. The necessary and sufficient condition to stabilize the predictor is to take a matrix L such that all eigenvalues of $\mathcal{A}$ are inside the unit circle ([5]).

In this case, the analysis by using matrix transfer functions is rather complex involving $\mathcal{Z}$-modified transform representations ([7]).

3 Nonconventional general sampling

Assume now that measurements are taken synchronously with the input updating instants, but not all sensors are available at every sampling instant. Assume also that the available sensors have a different time-varying delay. Then, a different prediction algorithm must be developed in order to make use of all the available measurements in spite of their availability and delays. The state is first periodically updated running the model (2) in open loop from the last sampling

instant t_{k-1}

$$\hat{\boldsymbol{x}}[t|t_{k-1}] = \boldsymbol{A}^{t-t_{k-1}}\,\hat{\boldsymbol{x}}[t_{k-1}] + \sum_{i=1}^{t-t_{k-1}} \boldsymbol{A}^{i-1}\boldsymbol{B}\,\boldsymbol{u}[t-i]. \tag{12a}$$

where $\hat{\boldsymbol{x}}[t|t_{k-1}]$ represents the state estimation at time t with the available measurements until t_{k-1}, and $\hat{\boldsymbol{x}}[t_{k-1}]$ represents the state estimation at t_{k-1} with the available measurements until t_{k-1}, i.e., the best estimation of $\boldsymbol{x}(t_{k-1})$. If a new measurement is available at $t = t_k$ the state is updated as

$$\hat{\boldsymbol{x}}[t_k] = \hat{\boldsymbol{x}}[t_k|t_k - 1] + \sum_{i=1}^{n_m} \boldsymbol{\ell}_{i,k}\,(m_{i,k} - \boldsymbol{c}_i\,\hat{\boldsymbol{x}}[t_k - d_{i,k}|t_k - 1])\,\alpha_{i,k}. \tag{12b}$$

where $\boldsymbol{\ell}_{i,k}$ is the innovation gain vector used to update the state estimation with the measurement $m_{i,k}$. Equation (12b) uses the delayed state estimation $\hat{\boldsymbol{x}}[t_k - d_{i,k}|t_k - 1]$, that is calculated as

$$\hat{\boldsymbol{x}}[t_k - d_{i,k}|t_{k-1}] = \boldsymbol{A}^{-d_{i,k}}\,\hat{\boldsymbol{x}}[t_k|t_{k-1}] + \sum_{j=1}^{d_{i,k}} \boldsymbol{A}^{j-1-d_{i,k}}\,\boldsymbol{B}\,\boldsymbol{u}[t_k - j]. \tag{12c}$$

To reduce the computational cost on the delayed state estimation, an extended state vector including all the necessary past estate estimations can be used (see [8]). The desired output prediction is obtained by means of the output equation $\hat{\boldsymbol{z}}[t] = \boldsymbol{C}_z\,\hat{\boldsymbol{x}}[t]$.

The features of the predictor are determined by the gains $\boldsymbol{\ell}_{i,k}$ used to update the state estimation upon the sampling of each measurement, that is, the innovation matrix gain

$$\boldsymbol{L}_k = \left[\boldsymbol{\ell}_{1,k}\;\boldsymbol{\ell}_{2,k}\;\cdots\;\boldsymbol{\ell}_{n_m,k}\right] \tag{13}$$

The predictor matrix gain must be designed to assure the predictor stability, robustness to the sporadic data availability, and disturbance and noise attenuation. The predictor gain is, in general, time-varying, but the case of a constant gain can be also feasible. In order to design a predictor, i.e., to determine the predictor matrix gain (13), with those properties several techniques can be used, depending on the nature of the sampling scenario and disturbances. The different design techniques considered are pole placement, LMI based disturbance attenuation, and Kalman filtering. The situations in which each of the techniques are more suitable will be discussed in the following.

The case developed in subsection 2.2 is a particular case in which the measurements were taken regular and periodically and, therefore, the pole placement technic is applicable. If the measurements are taken synchronously with the input update but they have irregular delay and availability, the pole placement technique cannot be applied. In order to derive an LMI based design technique some definitions must be made first.

The measurements $m_{i,k}$ are assumed to be available irregularly at discrete instants $t = t_k$, $t \in \mathbb{N}$, $k \in \mathbb{N}$, being $N_k = t_k - t_{k-1}$ the number of input updates from t_{k-1} to t_k, and, therefore $t_k = \sum_{i=1}^{k} N_i$ represents the instant in which the t-th input update occurs and the k-th sample is available (formed with the values of some of the sensors). The availability matrix at instant t_k is defined as $\boldsymbol{\alpha}_k = \text{diag}\{\alpha_{1,k}, \ldots, \alpha_{n_m,k}\}$. The sampling scenario parameter s_k is defined as the combination of sensor availability, time between samples and sensor delay $(\boldsymbol{\alpha}_k, N_k, d_{i,k})$ that defines a sample, and enumerates all the possible sampling situations as

$$s_k \in \mathcal{S} = \{1, 2, \ldots, n_{\mathcal{S}}\}, \tag{14}$$

where $n_{\mathcal{S}}$ is the number of possible combinations. All the variables that define the sampling scenario can be expressed as a function of this parameter, i.e., $N_k = N(s_k)$, $d_{i,k} = d_i(s_k)$, and $\boldsymbol{\alpha}_k = \boldsymbol{\alpha}(s_k)$. If the updating gain matrix is also defined to have a different value for each possible value of the sampling parameter, that is

$$\boldsymbol{L}_k = \boldsymbol{L}(s_k) \in \boldsymbol{\mathcal{L}} = \{\boldsymbol{L}(1), \boldsymbol{L}(2), \ldots, \boldsymbol{L}(n_{\mathcal{S}})\}, \tag{15}$$

then, the prediction error dynamics is demonstrated to be of the parametric form (see [8])

$$\tilde{\boldsymbol{x}}_k = \boldsymbol{\mathcal{A}}(s_k)\, \tilde{\boldsymbol{x}}_{k-1} + \boldsymbol{\mathcal{B}}(s_k)\, \mathbb{V}_k \tag{16a}$$

$$\tilde{\boldsymbol{z}}_k = \boldsymbol{C}_z\, \tilde{\boldsymbol{x}}_k, \tag{16b}$$

with $\mathbb{V}_k = \begin{bmatrix} \boldsymbol{w}[t_k - 1] \\ \boldsymbol{w}[t_k - 2] \\ \vdots \\ \boldsymbol{w}[t_k - \beta_k] \\ \boldsymbol{\nu}_k \end{bmatrix}$, $\beta = \max\{d_1(s_k), \ldots, d_{n_m}(s_k), N(s_k)\}, \forall s_k \in \mathcal{S}$,

and where

$$\boldsymbol{\mathcal{A}}(s_k) = (\boldsymbol{I} - \boldsymbol{L}(s_k)\, \boldsymbol{\alpha}(s_k)\, \boldsymbol{C}_d(s_k))\, \boldsymbol{A}^{N(s_k)}, \qquad \boldsymbol{C}_d(s_k) = \begin{bmatrix} \boldsymbol{c}_1\, \boldsymbol{A}^{-d_1(s_k)} \\ \vdots \\ \boldsymbol{c}_{n_m}\, \boldsymbol{A}^{-d_{n_m}(s_k)} \end{bmatrix}_{n_m \times n},$$

$$\boldsymbol{\mathcal{B}}(s_k) = \begin{bmatrix} \boldsymbol{\Lambda}(N(s_k)) - \boldsymbol{L}(s_k)\, \boldsymbol{\alpha}(s_k)\, \boldsymbol{\mathcal{C}}_d(s_k) & -\boldsymbol{L}(s_k)\, \boldsymbol{\alpha}(s_k) \end{bmatrix}$$

$$\boldsymbol{\mathcal{C}}_d(s_k) = \begin{bmatrix} \boldsymbol{c}_1\, \boldsymbol{A}^{-d_1(s_k)}\, (\boldsymbol{\Lambda}(N(s_k)) - \boldsymbol{\Lambda}(d_1(s_k))) \\ \vdots \\ \boldsymbol{c}_{n_m}\, \boldsymbol{A}^{-d_{n_m}(s_k)}\, (\boldsymbol{\Lambda}(N(s_k)) - \boldsymbol{\Lambda}(d_{n_m}(s_k))) \end{bmatrix}_{n_m \times \beta n}.$$

$$\boldsymbol{\Lambda}(N(s_k)) = [\overbrace{\boldsymbol{B}_w \;\; \boldsymbol{A}\boldsymbol{B}_w \;\; \boldsymbol{A}^2\boldsymbol{B}_w \;\; \cdots \;\; \boldsymbol{A}^{N(s_k)-1}\boldsymbol{B}_w}^{N(s_k)} \;\; \underbrace{\boldsymbol{0} \;\; \cdots \;\; \boldsymbol{0}}_{\beta}]_{n \times \beta n_w}. \tag{17}$$

With the introduction of parameter s_k the error dynamics is a parametrically time-varying linear system, specifically a jump linear system, where the disturbances and noise vectors ($\mathbb{V}_k$) are the inputs, the state estimation error ($\tilde{\boldsymbol{x}}_k$) is the state, and the desired output prediction error ($\tilde{\boldsymbol{z}}_k$) is the output.

The predictor design objective is to find a procedure to calculate the matrix $\boldsymbol{L}(s_k)$ (time varying with s_k, or constant) that leads to an adequate performance in terms of disturbance attenuation and robustness to the time varying delay and availability. The main idea consists of minimizing the norm of the transfer function from input $\mathbb{V}_k$ to output $\tilde{\boldsymbol{z}}_k$. Different norms can be used depending on the known norms that characterize the disturbances. If the ℓ_2 norm of the disturbances is known, the $\mathcal{H}_\infty$ norm can be minimized in order to minimize the ℓ_2 norm of the prediction error. If the ℓ_∞ or RMS norm of the disturbances is known, minimization of $\mathcal{H}_\infty$ implies the minimization of the RMS value of the prediction error. If the disturbances are white noises of known variance, minimization of $\mathcal{H}_2$ norm will imply the minimization of RMS norm of the prediction error.

As an example, the design procedure for one of the transfer function norms is shown. If the ℓ_2 norm of the disturbance and noise signals are known, it is possible to minimize the norm $\|\tilde{\boldsymbol{z}}_k\|_2$ by means of minimizing the sum

$$\sum_{i=1}^{n_w} \gamma_{w_i}^2 \|w_{i,k}\|_2^2 + \sum_{i=1}^{n_m} \gamma_{\nu_i}^2 \|\nu_{i,k}\|_2^2$$

along all variables γ_{w_i}, γ_{ν_i}, $\boldsymbol{P}(s_k) = \boldsymbol{P}(s_k)^\top \in \mathbb{R}^{n\times n}$, $\boldsymbol{Q}(s_k) \in \mathbb{R}^{n\times n}$, $\boldsymbol{X}(s_k) \in \mathbb{R}^{n\times n_m}$, $s_k \in \mathcal{S}$ that satisfy the LMI equation

$$\begin{bmatrix} \boldsymbol{Q}(s_k) + \boldsymbol{Q}(s_k)^\top - \boldsymbol{P}(s_k) & \boldsymbol{M}_A(s_k) & \boldsymbol{M}_B(s_k) \\ \boldsymbol{M}_A(s_k)^\top & \boldsymbol{P}(s_{k-1}) - \boldsymbol{C}_y^\top \boldsymbol{C}_y & \boldsymbol{0} \\ \boldsymbol{M}_B(s_k)^\top & \boldsymbol{0} & \boldsymbol{\Gamma}^2 \end{bmatrix} \succ 0 \tag{18}$$

with

$$\boldsymbol{M}_A(s_k) = (\boldsymbol{Q}(s_k) - \boldsymbol{X}(s_k)\boldsymbol{\alpha}(s_k)\,\boldsymbol{C}_d(s_k))\;\boldsymbol{A}^{N(s_k)}, \tag{19}$$

$$\boldsymbol{M}_B(s_k) = \left[\boldsymbol{Q}(s_k)\boldsymbol{\Lambda}(N(s_k)) - \boldsymbol{X}(s_k)\boldsymbol{\alpha}(s_k)\,\boldsymbol{C}_d(s_k) \;-\boldsymbol{X}(s_k)\boldsymbol{\alpha}(s_k)\right] \tag{20}$$

$$\boldsymbol{\Gamma} = \mathrm{diag}\{\boldsymbol{\Gamma}_w', \boldsymbol{\Gamma}_\nu\}, \qquad \boldsymbol{\Gamma}_w' = \frac{1}{\beta}\mathrm{diag}\{\boldsymbol{\Gamma}_w, \ldots, \boldsymbol{\Gamma}_w\}_{\beta n\times \beta n},$$

$$\boldsymbol{\Gamma}_w = \mathrm{diag}\{\gamma_{w_1}, \ldots, \gamma_{w_n}\}, \qquad \boldsymbol{\Gamma}_\nu = \mathrm{diag}\{\gamma_{\nu_1}, \ldots, \gamma_{\nu_{n_m}}\}.$$

with matrix $\boldsymbol{\Lambda}(N(s_k))$ defined as (17). This problem can be solved with LMI standard solvers. Finally, the matrix gain that minimizes the norm error is $\boldsymbol{L}(s_k) = \boldsymbol{Q}(s_k)^{-1}\boldsymbol{X}(s_k)$. This solution gives a different gain for each sampling scenario, that leads to a scheduled-gain predictor implementation. If some restrictions are made over matrices $\boldsymbol{Q}(s_k)$ and $\boldsymbol{X}(s_k)$, it is possible to obtain different implementations. For example, if $\boldsymbol{Q}(s_k)$ and $\boldsymbol{X}(s_k)$ are chosen to be constant (i.e., $\boldsymbol{Q}(s_k) = \boldsymbol{Q}$ and $\boldsymbol{X}(s_k) = \boldsymbol{X}$) a constant gain $\boldsymbol{L}$ is obtained leading to the predictor with the lowest computational cost (see [9]).

Application to asynchronous sampling. When dealing with asynchronous measurements (taken at arbitrary time instants), the use of a continuous time model is necessary, implying matrix exponentials computation and, therefore, a high computational cost. In order to avoid this, an interpolation is proposed to first estimate the value of the output measurement at synchronous instants with the input update (when $\tau = tT$). If τ_k ($tT < \tau_k < (t+1)T$) is the instant in which an asynchronous measurement $y(\tau_k)$ takes place, the idea is to use an interpolation technique that gives a synchronous measurement m_k corresponding to the instant t or $t+1$, i.e., $m_k = f(y(\tau_k), \tau_k)$. When this technique is applied, a new measurement noise is introduced, whose value depends on the interpolation technique. As the interpolation technique is known, the noise norm due to interpolation errors can be bounded and, therefore, the previous technique is also applicable (see [5]).

4 Kalman filter techniques

Assume that the disturbances $\boldsymbol{w}[t]$ and $\boldsymbol{\nu}_k$ are white noise with variance-covariance matrices $\boldsymbol{W}$ and $\boldsymbol{V}$ respectivelly. ($\mathcal{E}\{\boldsymbol{w}[t]^\top \boldsymbol{w}[\tau]\} = \boldsymbol{W}\delta(t-\tau)$, $\mathcal{E}\{\boldsymbol{\nu}_k^\top \boldsymbol{\nu}_j\} = \boldsymbol{V}\delta(k-j)$). Assume that some of the elements of vector $\boldsymbol{y}(\tau)$ are measured at arbitrary discrete instants t_k, $k \in \mathbb{N}$ (at least one element $m_{i,k}$ is available at instant t_k). Let us define the $\mu_k \times n_m$ matrix $\boldsymbol{M}_k$ formed with the rows of an identity matrix that correspond to the position of the elements $m_{i,k}$ available at instant t_k, thus μ_k is the number of variables measured at instant t_k. N_k is the elapsed time between the last two consecutive sampling instants ($N_k = t_k - t_{k-1}$) and $d_{i,k}$ is the delay of an individual measurement $m_{i,k}$. Under these suppositions, the gain matrix $\boldsymbol{L}_k$ that minimizes the error variance (Kalman filter) of the algorithm (12) is calculated on-line by ([1,5])

$$\hat{\boldsymbol{Q}}_{k|k-1} = \boldsymbol{A}^{N_k}\hat{\boldsymbol{Q}}_{k-1}(\boldsymbol{A}^\top)^{N_k} + \sum_{i=0}^{N_k-1} \boldsymbol{A}^i \boldsymbol{B}_w \boldsymbol{W} \boldsymbol{B}_w^\top (\boldsymbol{A}^i)^\top,$$

$$\boldsymbol{L}_k = \hat{\boldsymbol{Q}}_{k|k-1}\left(\boldsymbol{M}_k\boldsymbol{C}_{d,k}\right)^\top \Big(\boldsymbol{M}_k\left(\boldsymbol{C}_{d,k}\hat{\boldsymbol{Q}}_{k|k-1}\boldsymbol{C}_{d,k}^\top + \boldsymbol{V} + \boldsymbol{W}_d\right)\boldsymbol{M}_k^\top\Big)^{-1}\boldsymbol{M}_k$$

$$\hat{\boldsymbol{Q}}_k = \left(\boldsymbol{I} - \boldsymbol{L}_k\boldsymbol{C}_{d,k}\right)\hat{\boldsymbol{Q}}_{k|k-1}$$

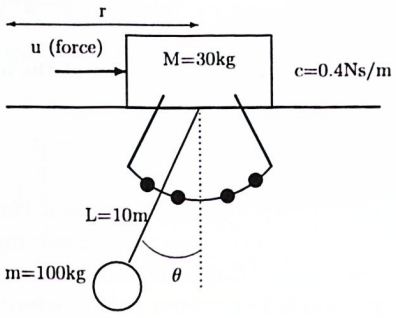
being

$$\boldsymbol{C}_{d,k} = \begin{bmatrix} \boldsymbol{c}_1 \boldsymbol{A}^{-d_{1,k}} \\ \vdots \\ \boldsymbol{c}_{n_m} \boldsymbol{A}^{-d_{n_m,k}} \end{bmatrix}_{n_m \times n}, \quad \boldsymbol{W}_d = \underbrace{\begin{bmatrix} \boldsymbol{c}_1 \boldsymbol{A}^{-d_{1,k}} \boldsymbol{\Lambda}(d_{1,k}) \\ \vdots \\ \boldsymbol{c}_{n_m} \boldsymbol{A}^{-d_{n_m,k}} \boldsymbol{\Lambda}(d_{n_m,k}) \end{bmatrix}}_{\star} \begin{bmatrix} \boldsymbol{W} & & \boldsymbol{0} \\ & \ddots & \\ \boldsymbol{0} & & \boldsymbol{W} \end{bmatrix} [\star]^\top$$

If the measurements are available periodically and the delay in each sensor is constant, an stationary solution of the above equations can be obtained and then, the predictor gains can be calculated off-line and applied as an scheduled gain predictor (as in the solution in the LMI based approach).

5 Application example

Consider the system shown in the figure 3. The input of the system is the force that is assumed to be updated by a controller at a constant rate of $T = 0.5s$. The measured signals are θ and r, but they are assumed to be measured by binary sensors that produce a digital pulse when the signal reaches some predefined values. A low cost encoder produces a pulse every 1 m for the r signal. The θ signal is assumed to produce a pulse at values $\{-0.07 \ -0.02 \ 0.02 \ 0.07\}$. As a consequence there are measurements of both θ and r at random instants that are asynchronous with the input update. The linearized model of this system (see [10]) is defined by the shown matrices, where $\boldsymbol{B_w} = \boldsymbol{B_c}$. The output variable



$$\boldsymbol{A_c} = \begin{bmatrix} 0 & 1 & 0 & 0 \\ 0 & -\frac{c}{M} & -\frac{mg}{M} & 0 \\ 0 & 0 & 0 & 1 \\ 0 & -\frac{c}{ML} & -\frac{(M+m)g}{ML} & 0 \end{bmatrix}, \quad \boldsymbol{B_c} = \begin{bmatrix} 0 \\ \frac{1}{M} \\ 0 \\ \frac{1}{ML} \end{bmatrix},$$

$$C_y = \begin{bmatrix} 1 & 0 & 0 & 0 \\ 0 & 0 & 1 & 0 \end{bmatrix}^T$$

Fig. 3. Crane with binary sensors.

of interest is assumed to be θ, and hence $C_z = [0\ 0\ 1\ 0]$. The objective of the predictor is then to estimate the value of θ at the synchronous instants (i.e. to obtain $\hat{\theta}[t]$). For the simulation, the input is generated as a band limited white noise of period 1 sec. and power 100, that is contaminated by another band limited white noise of power 2 and period 0.2 sec. The measurements are contaminated by additive random normal noise of variance $0.01^2 rad^2$ and $0.1^2 m^2$ respectively. A Kalman filter is designed taking first the matrices $\boldsymbol{V} = [0.1^2\ 0; 0\ 0.01^2]$ and $\boldsymbol{W} = 4$. In the figure 5 the true and the estimated states at the measurement instants are shown, as well as the true output ($\theta[t]$) and the predicted one ($\hat{\theta}[t]$) at the input updating instants. The average quadratic prediction error is $1.357 \cdot 10^{-5}$.

If the $\boldsymbol{V}$ and $\boldsymbol{W}$ matrices are not exactly known, the behavior is not optimal, but the predictor still works well. As an example, if the matrices used in the predictor equation are $\boldsymbol{V} = [0.1^2\ 0; 0\ 0.02^2]$ and $\boldsymbol{W} = 2$, the quadratic prediction error average is $1.94 \cdot 10^{-5}$, and with $\boldsymbol{V} = [0.1^2\ 0; 0\ 0.005^2]$ and $\boldsymbol{W} = 8$, the quadratic prediction error average is $2.52 \cdot 10^{-5}$.

Consider now the case when the output of the system is measured by 2 sensors with different precision and measurement availability instants. This idea is covered by the general model taking $C_y = [C_z^\top\ C_z^\top]^\top$ and $\boldsymbol{V} = [v_1\ v_{12}\ ;\ v_{12}\ v_2]$.

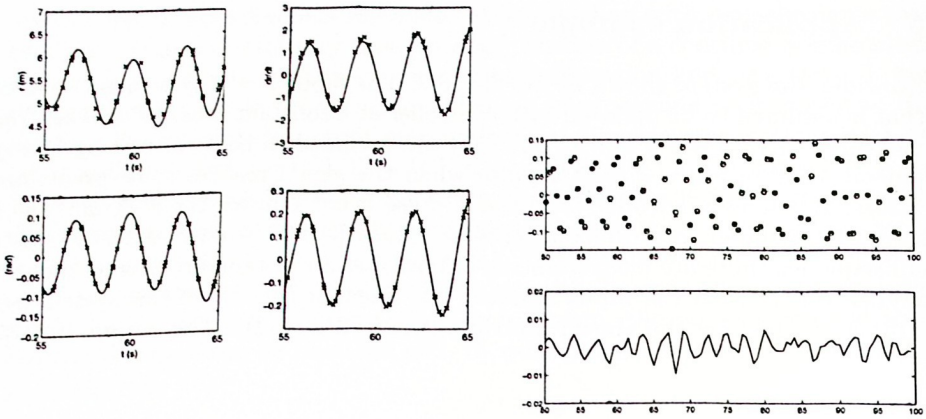


Fig. 4. True (- left, o right) and estimated (x) output, and prediction error at the input updating instants (tT).

Assume now that only θ is measured with 2 sensors. The first one is a binary sensor fixed on position $\theta = 0$ that gives scarce but very precise measurements (null noise variance assumed). The second sensor is a continuous sensor that gives a measurement every $N = 2$ input updating periods but with a noise variance of 0.006^2 (hence $\boldsymbol{V} = [0.006^2\ 0; 0\ 0]$). With the same input conditions as before the average quadratic prediction error is $4.35 \cdot 10^{-6}$. In order to compare the sensor fusion effect of the predictor, the same simulation is carried out assuming that only the second sensor is available, obtaining an error of $10.5 \cdot 10^{-6}$. If only the first sensor is used the resulting error is $10.7 \cdot 10^{-6}$. In the figure 5 the prediction errors are shown.

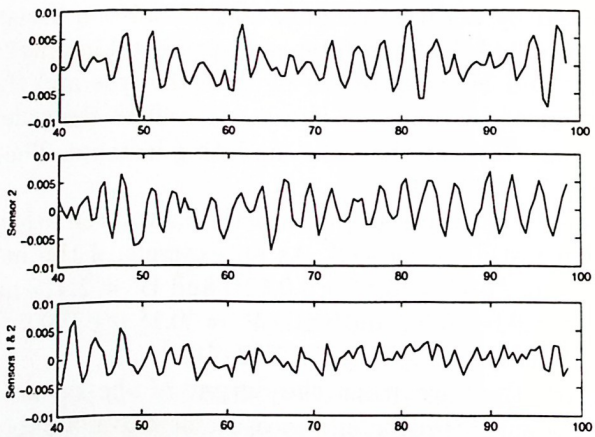


Fig. 5. Error with different sensor measurements.

Consider now a control period of $T = 0.2$ s. Assume that the angle and horizontal position are measured synchronously with the input update every 2 or 4 seconds ($N = \{10, 20\}$), and assume also that the signal transmission takes 0.4 or 0.8 seconds ($d_i = \{2, 4\}$) because they are accessed through a shared network. It is assumed that r and θ are measured with zero mean white noise sensors with typical deviation $\sigma_r = 0.01$ m and $\sigma_\theta = 0.001$ rad. In this scenario, both $\mathcal{H}_\infty$ and *Kalman* predictor are applied in order to implement a closed loop control using the state estimation. A disturbance $v(\tau) = 0.2$, $\tau > 200$s has been applied. Figure 6 show the results with $\mathcal{H}_\infty$ and Kalman filter approaches. Before the constant disturbance, the performance with $\mathcal{H}_\infty$ and Kalman filter approaches is similar. With the disturbance, the $\mathcal{H}_\infty$ approach leads to a prediction error characterized with $\sigma_r = 0.048$, while the Kalman filter leads to $\sigma_r = 0.078$. An explanation to this fact is that the Kalman filter is optimum only if the disturbances are white noise with zero mean. On the other hand, the computational cost of the Kalman filter is 8 times the $\mathcal{H}_\infty$ predictor.

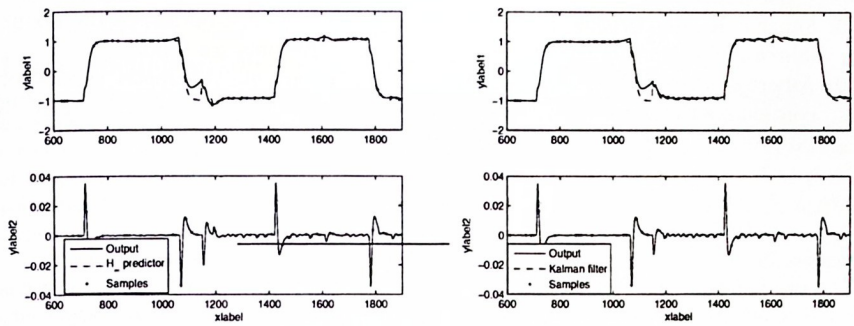


Fig. 6. Crane Control with different virtual sensors.

6 Conclusions

A virtual sensor has been defined as the intelligent device that estimates the desired outputs of a process using the information of the model, the known control inputs, and the accessible outputs measured with different sensors with different availability, noise and delays. Two different estimation algorithms that address sensor availability, scarce measurements and delays as a whole have been proposed (continuous and discrete one). Different techniques have been proposed for different sampling scenarios: pole placement (for periodic sampling), LMI based approach (when the different sampling scenarios are known and finite) and Kalman filter based techniques (for the general case with white noise disturbances). The Kalman filter needs an on-line calculation of the predictor gain, while the other techniques calculate the gain off-line, leading to a lower computational cost predictor.

References

1. Sanchis, R., Albertos, P.: Virtual sensors: model based output prediction. First EurAsia Conference on Advances in Information and Communication Technology (2002)
2. Gustafsson, F., Drevo, M., Forssell, U., Lofgren, M., Persson, N., Quicklund, H.: Virtual sensors of tire pressure and roadfriction. Society of Automotive Engineers **2001-01-0796** (2001)
3. Long, T., Hanzevack, E., Bynum, W.: Sensor fusion and failure detection using virtual sensors. Proceedings of the Arnencan Control Conference (1999)
4. Albertos, P., Goodwin, G.: Virtual sensors for control applications. Annual Reviews in Control **26** (2002) 101–112
5. Peñarrocha, I.: Virtual Sensors for processes with scarce and delayed measurements. PhD Thesis, Polythecnic University of Valencia (2006)
6. Astrom, K., Wittenmark, B.: Computer controlled systems. Prentice Hall (1997)
7. Cuenca, A., Salt, J., Albertos, P.: Implementation of algebraic controllers for non-conventional sampled-data systems. Real-Time Systems **35**(1) (2007) 59–89
8. Peñarrocha, I., Sanchis, R., Albertos, P.: Design of low cost virtual sensors. IEEE - Instrumentation and Measurement Technology Conference (IMTC) (2006)
9. Sanchis, R., Peñarrocha, I., Albertos, P.: Design of robust output predictors under scarce measurements with time-varying delays. Automatica (2007)
10. Albertos, P., Sanchis, R., Sala, A.: Fault detection via parameter estimation in continuous-time systems with random sampling (scarce measurements). IFAC Symposium on Low Cost Automation. Buenos Aires (1995)

Modeling and Manipulating 3D Datasets through the Extreme Vertices Model in the n-Dimensional Space (nD-EVM)

Ricardo Pérez-Aguila

Universidad Tecnológica de la Mixteca
Carretera Huajuapan-Acatlima Km. 2.5.
Huajuapan de León, Oaxaca 69000, México
ricardo.perez.aguila@gmail.com

Abstract. The objective behind this work is to describe the Extreme Vertices Model in the n-Dimensional Space (nD-EVM) and the way it represents n-Dimensional Orthogonal Pseudo-Polytopes (nD-OPP's) by considering only a subset of their vertices: the Extreme Vertices. Once we have the elements for managing nD-OPP's through the nD-EVM, we will exemplify the way 3D datasets can be represented through the model. These examples are analyzed in order to compare storage requirements according to their respective representations through voxelizations and through the nD-EVM. The basic algorithms on the nD-EVM provide us useful information about the polytopes being modeled through our scheme.

Keywords: n-Dimensional Orthogonal Polytopes Modeling, Geometrical and Topological Interrogations, 3D Voxelizations Compression, Computational Geometry.

1 Introduction and Problem Statement

Coxeter defines an **n-Dimensional Euclidean Polytope** Π_n as a finite region of n-dimensional Euclidean space enclosed by a finite number of (n-1)-dimensional hyperplanes [3]. The finiteness of the region implies that the number N_{n-1} of bounding hyperplanes satisfies the inequality $N_{n-1}>n$. The part of the polytope that lies on one of these hyperplanes is called a cell. Each cell of a Π_n is an (n-1)D polytope, Π_{n-1}. The cells of a Π_{n-1} are Π_{n-2}'s, and so on; we thus obtain a descending sequence of elements Π_{n-3}, Π_{n-4}, ... , Π_3 (a volume), Π_2 (a polygon), Π_1 (an edge), Π_0 (a vertex) [3].

The representation of a polytope through a scheme of Hyperspatial Occupancy Enumeration is essentially a list of identical hyperspatial cells occupied by the polytope. Specific types of cells, called hypervoxels [4] are hyper-boxes (hypercubes, for example) of a fixed size that lie in a fixed grid in the n-dimensional space. Jonas defines two kinds of hypervoxels [4]:

- Centered Hypervoxel: an n-dimensional hyper-box whose dimensions are given by x_1Side, x_2Side, ..., x_nSide and it is represented by the coordinates of its centroid.
- Shifted Hypervoxel: whose characteristics are same that those for the centered hypervoxel, except that its representation is given by some of its 2^n vertices.

By instantiation, we know that a 2D hypervoxel is a pixel while a 3D hypervoxel is a voxel; the term *rexel* is suggested for referencing a 4D hypervoxel [4].

The collection of hyperboxes can be codified as an n-dimensional array $C_{x_1,x_2,...,x_n}$ of binary data.

The array will represent the coloration of each hypervoxel:

- If $C_{x_1,x_2,...,x_n}=1$ the **black** hypervoxel $C_{x_1,x_2,...,x_n}$ represents an occupied region from the nD space.
- If $C_{x_1,x_2,...,x_n}=0$ the **white** hypervoxel $C_{x_1,x_2,...,x_n}$ represents an empty region from the nD space.

In fact, the set of black cells represents an nD-OPP p whose vertices coincide with some of the black cells' vertices.

By using the representation through a binary array, the computation of some operations just control the operations between bits for all the elements. It is well known, for example, that Boolean set operations are trivial under this scheme, however, the spatial complexity of an hypervoxelization is $\prod m_i$ where m_i, $1 \leq i \leq n$, is the length of the grid along the X_i-axis. For example, a three-dimensional grid with $m_1 = m_2 = m_3 = 1000$ requires to store 1 billion (1×10^9) voxels.

It is well known that hypervoxelizations are the native way in which some datasets are represented and stored. Moreover, common 3D datasets such as the found in medical applications, for example, are required to have a high degree of precision because of the importance of the

© L. Sánchez, O. Pogrebnyak and E. Rubio (Eds.)
Industrial Informatics.
Research in Computing Science 31, 2007, pp. 15-24

information obtained from them. However, to more precision usually a cost in spatial complexity must be paid. In this last sense, efficient procedures for compression are required such that the precision in the datasets is not compromised or at least it is affected as minimum as possible. It is clear that the union of the black cells in an nD hypervoxelization defines an nD polytope, in fact, an nD Orthogonal Pseudo-Polytope (nD-OPP). The objective behind this work is to describe the Extreme Vertices Model in the n-Dimensional Space (nD-EVM) and the way it represents nD-OPP's by considering only a subset of their vertices: the Extreme Vertices (**Section 2**). The Extreme Vertices Model (3D-EVM) was originally presented, and widely described, in [1] for representing 2-manifold Orthogonal Polyhedra and later considering both Orthogonal Polyhedra (3D-OP's) and Pseudo-Polyhedra (3D-OPP's) [2]. This model has enabled the development of simple and robust algorithms for performing the most usual and demanding tasks on solid modeling, such as closed and regularized Boolean operations, solid splitting, set membership classification operations and measure operations on 3D-OPP's. It is natural to ask if the EVM can be extended for modeling n-Dimensional Orthogonal Pseudo-Polytopes (nD-OPPs). In this sense, some experiments were made, in [6], where the validity of the model was assumed true in order to represent 4D and 5D-OPPs. Finally, in [7] was formally proved that the nD-EVM is a complete scheme for the representation of nD-OPPs. The meaning of complete scheme was based in Requicha's set of formal criterions that every scheme must have rigorously defined: Domain, Completeness, Uniqueness and Validity. Although the EVM of an nD-OPP has been defined as a subset of the nD-OPP's vertices, there is much more information about the polytope hidden within this subset of vertices. We will describe basic procedures and algorithms in order to obtain this information (**Sections 2.6, 2.7** and **2.8**). Once we have the elements for managing nD-OPP's through the nD-EVM, we will describe the way 3D datasets can be represented through the model (**Section 3**). We will show some examples and we will compare storage requirements according to their respective representations through voxelizations and through the nD-EVM. We will indicate the way the basic algorithms on the nD-EVM provide useful information about the polytopes being modeled under our scheme.

2 The Extreme Vertices Model in the n-Dimensional Space (nD-EVM)

2.1 Preliminary Background: n-Dimensional Orthogonal Pseudo-Polytopes

Definition 2.1: *A Singular n-Dimensional Hyper-Box in* $\mathbb{R}^n$ *is the continuous function*

$$I^n: [0,1]^n \rightarrow [0,1]^n$$
$$x \sim I^n(x) = x$$

For a general singular kD hyper-box c we will define the boundary of c.

Definition 2.2: *For all i,* $1 \le i \le n$, *the two singular (n-1)D hyper-boxes* $I^n_{(i,0)}$ *and* $I^n_{(i,1)}$ *are defined as follows: If* $x \in [0,1]^{n-1}$ *then*

- $I^n_{(i,0)}(x) = I^n(x_1,...,x_{i-1},0,x_i,...,x_{n-1}) = (x_1,...,x_{i-1},0,x_i,...,x_{n-1})$
- $I^n_{(i,1)}(x) = I^n(x_1,...,x_{i-1},1,x_i,...,x_{n-1}) = (x_1,...,x_{i-1},1,x_i,...,x_{n-1})$

Definition 2.3: *In a general singular nD hyper-box c we define the (i,α)-cell as* $c_{(i,\alpha)} = c \circ I^n_{(i,\alpha)}$

The next definitions indicate what we consider as the orientation of an (n-1)D cell.

Definition 2.4: *The orientation of an (n-1)D cell* $c \circ I^n_{(i,\alpha)}$ *is given by* $(-1)^{\alpha+i}$.

Definition 2.5: *An (n-1)D oriented cell is given by the scalar-function product* $(-1)^{i+\alpha} \cdot c \circ I^n_{(i,\alpha)}$

Definition 2.6: *A formal linear combination of singular general kD hyper-boxes,* $1 \le k \le n$, *for a closed set A is called a k-chain.*

Definition 2.7 [8]: *Given a singular nD hyper-box* I^n *we define the (n-1)-chain, called the boundary of* I^n, *by* $\partial(I^n) = \sum_{i=1}^{n} \left(\sum_{\alpha=0,1} (-1)^{i+\alpha} \cdot I^n_{(i,\alpha)} \right)$

Definition 2.8 [8]: *Given a singular general nD hyper-box c we define the (n-1)-chain, called the boundary of c, by* $\partial(c) = \sum_{i=1}^{n} \left(\sum_{\alpha=0,1} (-1)^{i+\alpha} \cdot c \circ I^n_{(i,\alpha)} \right)$

Definition 2.9 [8]: *The boundary of an n-chain* $\sum c_i$, *where each* c_i *is a singular general nD hyper-box, is given by* $\partial\left(\sum c_i\right) = \sum \partial(c_i)$

Definition 2.10: *A collection* $c_1, c_2, \ldots, c_k$, $1 \leq k \leq 2^n$, *of general singular nD hyper-boxes is a combination of nD hyper-boxes if and only if*

$$\left[\bigcap_{\alpha=1}^{k} c_\alpha([0,1]^n) = (\underbrace{0,\ldots,0}_{n})\right] \wedge \left[(\forall i, j,\ i \neq j,\ 1 \leq i, j \leq k)\left(c_i([0,1]^n) \neq c_j([0,1]^n)\right)\right]$$

In the above definition the first part of the conjunction establishes that the intersection between all the nD general singular hyper-boxes is the origin, while the second part establishes that there are not overlapping nD hyper-boxes.

Definition 2.11: *We say that an n-Dimensional Orthogonal Pseudo-Polytope p, or just an nD-OPP p, will be an n-chain composed by nD hyper-boxes arranged in such way that by selecting a vertex, in any of these hyper-boxes, we have that such vertex describes a combination of nD hyper-boxes (**Definition 2.10**) composed up to* 2^n *hyper-boxes.*

Describing nD-OPP's as union of disjoint nD hyper-boxes in such way that by selecting a vertex, in any of these hyper-boxes, we have that such vertex is surrounded up to 2^n hyper-boxes, will be very useful because in the following propositions we consider geometrical and/or topological local analysis over such vertices and their respective incident hyper-boxes.

2.2 The nD-EVM: Foundations

Definition 2.12: *Let c be a combination of hyper-boxes in the n-Dimensional space. An Odd Edge will be an edge with an odd number of incident hyper-boxes of c.*

Definition 2.13: *A brink or extended edge is the maximal uninterrupted segment, built out of a sequence of collinear and contiguous **odd edges** of an nD-OPP.*

Definition 2.14: *We will call Extreme Vertices of an nD-OPP p to the ending vertices of all the brinks in p. EV(p) will denote to the set of Extreme Vertices of p.*

The brinks in an nD-OPP p can be classified according to the main axis to which they are parallel. Since the extreme vertices mark the end of brinks in the n orthogonal directions, is that any of the n possible sets of brinks parallel to X_i-axis, $1 \leq i \leq n$, produce to the same set EV(p).

Definition 2.15: *Let p be an nD-OPP. $EV_i(p)$ will denote to the set of ending or extreme vertices of the brinks of p which are parallel to x_i-axis, $1 \leq i \leq n$.*

Theorem 2.1 [7]: *A vertex of an nD-OPP p, $n \geq 1$, when is locally described by a set of surrounding nD hyper-boxes, is an extreme vertex if and only if it is surrounded by an odd number of such nD hyper-boxes.*

Theorem 2.2 [7]: *Any extreme vertex of an nD-OPP, $n \geq 1$, when is locally described by a set of surrounding nD hyper-boxes, has exactly n incident linearly independent odd edges.*

Definition 2.16: *Let p be an nD-OPP. A kD couplet of p, $1<k<n$, is the maximal set of kD cells of p that lies in a kD space, such that a kD cell e_0 belongs to a kD extended hypervolume if and only if e_0 belongs to an (n-1)D cell present in $\partial(p)$.*

Theorem 2.3 [7]: *Let p be an nD-OPP with its associated sets $EV_1(p)$, $EV_2(p)$, ..., $EV_{n-1}(p)$, $EV_n(p)$. Then $EV_1(p) = EV_2(p) = \ldots = EV_{n-1}(p) = EV_n(p)$.*

Let Q be a finite set of points in $\mathbb{R}^3$. In [2] was defined the ABC-sorted set of Q as the set resulting from sorting Q according to coordinate A, then to coordinate B, and then to coordinate C. For instance, a set Q can be ABC-sorted is six different ways: $X_1X_2X_3$, $X_1X_3X_2$, $X_2X_1X_3$, $X_2X_3X_1$, $X_3X_1X_2$ and $X_3X_2X_1$. Now, let p be a 3D-OPP. According to [2] the Extreme Vertices Model of p, EVM(p), denotes to the ABC-sorted set of the extreme vertices of p. Then EVM(p) = EV(p) except by the fact that EV(p) is not necessarily sorted. In this work we will assume that the coordinates of extreme vertices in the Extreme Vertices Model of an nD-OPP p, $EVM_n(p)$ are sorted according to coordinate X_1, then to coordinate X_2, and so on until coordinate X_n. That is, we are considering the only ordering $X_1 \ldots X_i \ldots X_n$ such that $i-1 < i$, $1 < i \leq n$.

Definition 2.17: *Let p be an nD-OPP. We will define the Extreme Vertices Model of p, denoted by $EVM_n(p)$, as the model as only stores to all the extreme vertices of p.*

Theorem 2.4 [7]: *Card(EV$_i$(p))=Card(EV(p))=Card(EVM$_n$(p)) is an even number, $1 \le i \le n$.*

2.3 Sections and Slices of nD-OPP's

Definition 2.18: *We define the Projection Operator for (n-1)D cells, points, and set of points respectively as follows:*

- *Let $c(I^n_{(i,a)}(x)) = (x_1,..,x_n)$ be an (n-1)D cell embedded in the nD space. $\pi_j\left(c(I^n_{(i,a)}(x))\right)$ will denote the projection of the cell $c(I^n_{(i,a)}(x))$ onto an (n-1)D space embedded in nD space whose supporting hyperplane is perpendicular to X_j-axis: $\pi_j\left(c(I^n_{(i,a)}(x))\right) = (x_1,...,\hat{x}_j,...,x_n)$*
- *Let $v=(x_1,...,x_n)$ a point in $\mathbb{R}^n$. The projection of that point in the (n-1)D space, denoted by $\pi_j(v)$, is given by: $\pi_j(v) = (x_1,...,\hat{x}_j,...,x_n)$*
- *Let Q be a set of points in $\mathbb{R}^n$. We define the projection of the points in Q, denoted by $\pi_j(Q)$, as the set of points in $\mathbb{R}^{n-1}$ such that $\pi_j(Q) = \left\{p \in \mathbb{R}^{n-1} : p = \pi_j(x),\ x \in Q \subset \mathbb{R}^n\right\}$*

In all the cases $\hat{x}_j$ is the coordinate corresponding to X_j-axis to be suppressed.

Definition 2.19: *Consider an nD-OPP p:*

- *Let np_i be the number of distinct coordinates present in the vertices of p along X_i-axis, $1 \le i \le n$.*
- *Let $\Phi^i_k(p)$ be the k-th (n-1)D couplet of p which is perpendicular to X_i-axis, $1 \le k \le np_i$.*

Definition 2.20: *A Slice is the region contained in an nD-OPP p between two consecutive couplets of p. $Slice^i_k(p)$ will denote to the k-th slice of p which is bounded by $\Phi^i_k(p)$ and $\Phi^i_{k+1}(p)$, $1 \le k < np_i$.*

Definition 2.21: *A Section is the (n-1)D-OPP, n > 1, resulting from the intersection between an nD-OPP p and a (n-1)D hyperplane perpendicular to the coordinate axis X_i, $1 \le i \le n$, which not coincide with any (n-1)D-couplet of p. A section will be called external or internal section of p if it is empty or not, respectively. $S^i_k(p)$ will refer to the k-th section of p between $\Phi^i_k(p)$ and $\Phi^i_{k+1}(p)$, $1 \le k < np_i$.*

2.4 Computing Couplets and Sections

Theorem 2.5 [7]: *The projection of the set of (n-1)D-couplets, $\pi_i\left(\Phi^i_k(P)\right)$, of an nD-OPP P, can be obtained by computing the regularized XOR (⊗) between the projections of its previous $\pi_i\left(S^i_{k-1}(P)\right)$ and next $\pi_i\left(S^i_k(P)\right)$ sections, i.e., $\pi_i\left(\Phi^i_k(P)\right) = \pi_i\left(S^i_{k-1}(P)\right) \otimes^* \pi_i\left(S^i_k(P)\right),\ \forall k \in [1, np_i]$*

Theorem 2.6 [7]: *The projection of any section, $\pi_i\left(S^i_k(p)\right)$, of an nD-OPP p, can be obtained by computing the regularized XOR between the projection of its previous section, $\pi_i\left(S^i_{k-1}(p)\right)$, and the projection of its previous couplet $\pi_i\left(\Phi^i_k(p)\right)$.*

2.5 The Regularized XOR operation on the nD-EVM

Theorem 2.7 [2]: *Let p and q be two nD-OPP's having $EVM_n(p)$ and $EVM_n(q)$ as their respective EVM's in nD space, then $EVM_n(p \otimes^* q) = EVM_n(p) \otimes EVM_n(q)$.*

This result allows expressing a formula for computing nD-OPP's sections from couplets and vice-versa, by means of their corresponding Extreme Vertices Models. These formulae are obtained by combining **Theorem 2.7** with **Theorem 2.5**; and **Theorem 2.7** with **Theorem 2.6**, respectively:

Corollary 2.1 [2]: $EVM_{n-1}\left(\pi_i(\Phi^i_k(p))\right) = EVM_{n-1}\left(\pi_i(S^i_{k-1}(p))\right) \otimes EVM_{n-1}\left(\pi_i(S^i_k(p))\right)$

Corollary 2.2 [2]: $EVM_{n-1}\left(\pi_i(S^i_k(p))\right) = EVM_{n-1}\left(\pi_i(S^i_{k-1}(p))\right) \otimes EVM_{n-1}\left(\pi_i(\Phi^i_k(p))\right)$

2.6 Basic Algorithms for the nD-EVM

According to **Sections 2.2** to **2.4** we can define the following primitive operations which are based in the functions originally presented in [2]:

```
Output: An empty nD-EVM.
Procedure InitEVM( )
{ Returns the empty set.                    }
```

```
Input:  An nD-EVM p
Output:
An (n-1)D-EVM embedded in (n-1)D space.
Procedure ReadHvl(EVM p)
{ Extracts next (n-1)D couplet
  perpendicular to X1-axis from p.          }
```

```
Input/Output: An (n-1)D-EVM p embedded
in (n-1)D space.
Input: A coordinate coord of type
CoordType (CoordType is the chosen type
for the vertex coordinates: Integer or
Real)
Procedure SetCoord(EVM p, CoordType
coord)
{ Sets the X1-coordinate to coord on
  every vertex of the (n-1)D couplet
  p. For coord = 0 it performs the
  projection π1(p).                         }
```

```
Input:  An nD-EVM p
Output: A Boolean.
Procedure EndEVM(EVM p)
{ Returns true if the end of p
  along X1-axis has been reached.           }
```

```
Input:
An (n-1)D-EVM hvl embedded in
nD space.
Input/Output: An nD-EVM p
Procedure PutHvl(EVM hvl, EVM p)
{ Appends an (n-1)D couplet hvl,
  which is perpendicular to
  X1-axis, to p.                            }
```

```
Input:  Two nD-EVM's p and q.
Output: An nD-EVM
Procedure MergeXor(EVM p, EVM q)
{ Applies the Exclusive OR
  operation to the vertices of
  p and q and returns the
  resulting set.                            }
```

Function MergeXor performs an XOR between two nD-EVM's, that is, it keeps all vertices belonging to either $EVM_n(p)$ or $EVM_n(q)$ and discards any vertex that belongs to both $EVM_n(p)$ and $EVM_n(q)$. Since the model is sorted, this function consists on a simple merging-like algorithm, and therefore, it runs on linear time [2]. Its complexity is given by $O(Card(EVM_n(p)) + Card(EVM_n(q))$ since each vertex from $EVM_n(p)$ and $EVM_n(q)$ needs to be processed just once. Moreover, according to **Theorem 2.7**, the resulting set corresponds to the regularized XOR operation between p and q.

From the above primitive operations, the **Algorithms 2.1** and **2.2** may be easily derived. The **Algorithm 2.3** computes the sequence of sections of an nD-OPP p from its nD-EVM using the previous functions [2]. It sequentially reads the projections of the (n-1)D couplets *hvl* of the polytope p. Then it computes the sequence of sections using function *GetSection*. Each pair of sections S_i and S_j (the previous and next sections about the current *hvl*) is processed by a generic processing procedure (called *Process*), which performs the desired actions upon S_i and S_j.

```
Input: An (n-1)D-EVM corresponding to
section S. An (n-1)D-EVM corresponding
to couplet hvl.
Output: An (n-1)D-EVM.
Procedure GetSection(EVM S, EVM hvl)
    // Returns the projection of the
    // next section of an nD-OPP
    // whose previous section is S.
    return MergeXor(S, hvl)
end-of-procedure
```

Algorithm 2.1. Computing $EVM_{n-1}(\pi_1(S_k^i(p)))$ as

$$EVM_{n-1}(\pi_1(S_{k-1}^i(p))) \otimes EVM_{n-1}(\pi_1(\Phi_k^i(p)))$$

```
Input: An (n-1)D-EVM corresponding to
section Si. An (n-1)D-EVM corresponding
to section Sj.
Output: An (n-1)D-EVM.
Procedure GetHvl(EVM Si, EVM Sj)
    // Returns the projection of the
    // couplet between consecutive
    // sections Si and Sj.
    return MergeXor(Si, Sj)
end-of-procedure
```

Algorithm 2.2. Computing $EVM_{n-1}(\pi_1(\Phi_k^i(p)))$ as

$$EVM_{n-1}(\pi_1(S_{k-1}^i(p))) \otimes EVM_{n-1}(\pi_1(S_k^i(p)))$$

```
Input:  An nD-EVM p.
Procedure EVM_to_SectionSequence(EVM p)
        EVM hvl          // Current couplet.
        EVM Si, Sj       // Previous and next sections about hvl.
        hvl = InitEVM( )
        Si = InitEVM( )
        Sj = InitEVM( )
        hvl = ReadHvl(p)
        while(Not(EndEVM(p)))
                Sj = GetSection(Si, hvl)
                Process(Si, Sj)
                Si = Sj
                hvl = ReadHvl(p) // Read next couplet.
        end-of-while
end-of-procedure
```

Algorithm 2.3. Computing the sequence of sections from an nD-OPP p represented through the nD-EVM.

```
Input: An nD-EVM p and the number n of
dimensions.
Output: The content of nD space
enclosed by p.
Procedure Content(EVM p, int n)
  // Variable cont stores the content
  // of nD space enclosed by p.
  real cont = 0.0
  // Couplets between a slice of p.
  EVM hvl1, hvl2
  // Current section of p.
  EVM s
  if(n=1) then
    // Compute the length of the input
    // 1D-OPP expressed under the EVM.
    return Length(p)
  else
    n = n - 1
    hvl1 = InitEVM( )
    hvl2 = InitEVM( )
    s = InitEVM( )
    hvl1 = ReadHvl(p)
    while(Not(EndEVM(p)))
        hvl2 = ReadHvl(p)
        s = GetSection(s, hvl1)
        // Recursive Call.
        cont = cont + Content(s, n) *
                      dist(hvl1, hvl2)
        hvl1 = hvl2
    end-of-while
    return cont
  end-of-else
end-of-procedure
```

Algorithm 2.4. Computing the content of nD space enclosed by p.

```
Input: An nD-EVM p and the number n of
dimensions.
Output: The content of (n-1)D space
enclosed by the boundary of p.
Procedure BoundaryContent(EVM p, int n)
   // cont stores the content of
   // (n-1)D space enclosed by the
   // boundary of p.
   real cont = 0.0
   // Couplets between a slice of p.
   EVM hvl1, hvl2
   EVM s     // Current section of p.
   if(n=2) then
     // Compute the perimeter of
     // the input 2D-OPP expressed
     // under the 2D-EVM.
     return x1Sum(p) + x2Sum(p)
   else
     n = n - 1
     hvl1 = InitEVM( )
     hvl2 = InitEVM( )
     s = InitEVM( )
     hvl1 = ReadHvl(p)
     while(Not(EndEVM(p)))
         hvl2 = ReadHvl(p)
         s = GetSection(s, hvl1)
         // Call to algorithm Content
         // and recursive call.
         cont=cont + Content(hvl1, n) +
               BoundaryContent(s, n) *
                     dist(hvl1, hvl2)
         hvl1 = hvl2
     end-of-while
     cont = cont + Content(hvl1, n)
     // hvl1 contains the last
     // couplet of p.
     return cont
   end-of-else
end-of-procedure
```

Algorithm 2.5. Computing the content of (n-1)D space enclosed by the boundary of p.

2.7 Computing the Content of an nD-OPP

Now, we will describe a procedure in order to compute the content of nD space enclosed by an nD-OPP (length of a 1D-OPP, area of a 2D-OPP, volume of a 3D-OPP, and so on). In this case we will consider the partition induced by its Slices. Let p be an nD-OPP. The nD space enclosed by p, denoted by $Content_{(n)}(p)$, can be computed as the sum of the contents of its nD slices (**Equation 2.1**):

$$Content_{(n)}(p) = \begin{cases} Length(p) & n = 1 \\ \sum_{k=1}^{np_i - 1} Content_{(n-1)}\left(S_k^i(p)\right) \cdot dist\left(\Phi_k^i(p), \Phi_{k+1}^i(p)\right) & n > 1 \end{cases}$$

Where np_i is the number of couplets of p perpendicular to X_i-axis; $S_k^i(p)$ is the k-th section of the nD-OPP p which is perpendicular to X_i-axis and it is between couplets $\Phi_k^i(p), \Phi_{k+1}^i(p)$ whose distance is given by $dist\left(\Phi_k^i(p), \Phi_{k+1}^i(p)\right)$. The **Algorithm 2.4** implements **Equation 2.1** in order to compute the content of nD space enclosed by an nD-OPP p expressed through the nD-EVM.

2.8 Computing the Content of the Boundary of an nD-OPP

Now, we will describe the way to compute the content of (n-1)D space enclosed by the boundary of an nD-OPP (perimeter of a 2D-OPP, area of the boundary of a 3D-OPP, volume of the boundary of a 4D-OPP, and so on). Let p be an nD-OPP. The (n-1)D space enclosed by p, denoted by $BoundaryContent_{(n-1)}(p)$, can be computed as follows (**Equation 2.2**):

$$BoundaryContent_{(n-1)}(p) =$$

$$\begin{cases} x_1 Sum(p) + x_2 Sum(p) & n = 2 \\ \sum_{k=1}^{np_i} Content_{(n-1)}\left(\Phi_k^i(p)\right) + \sum_{k=1}^{np_i-1} BoundaryContent_{(n-1)}\left(S_k^i(p)\right) \cdot dist\left(\Phi_k^i(p), \Phi_{k+1}^i(p)\right) & n > 2 \end{cases}$$

According to previous equation we reach the basic case when n = 2. In this situation, the perimeter of a 2D-OPP p is computed as $x_1 Sum(p) + x_2 Sum(p)$ where $x_i Sum(p)$ is the sum of the lengths of all brinks parallel to X_i-axis. The **Algorithm 2.5** implements **Equation 2.2** in order to compute the content of (n-1)D space enclosed by the boundary of p expressed through the nD-EVM.

3 Manipulating "Real World" 3D Datasets with the nD-EVM

In this section we will describe some results related to the conversion from voxelizations to our specific implementation of nD-EVM when n = 3. Such voxelizations correspond to "real world" datasets taken from the MoViBio Research Group [5], and the University of Tübingen's Project VolRen [9]. As commented in the **Section 1**, a 3D voxelization is a set of black and white cells where each cell is a convex orthogonal polyhedron. The set of black cells represents an nD-OPP p whose vertices coincide with some of the black cells' vertices. Each of these vertices may be common to (surrounded by) up to 8 black cells. So, according to **Theorem 2.1**, if a vertex is surrounded by an odd number of black cells then it is an Extreme Vertex. Thus, a 3D voxelization to the 3D-EVM conversion algorithm is as simple as collecting every vertex that belongs to and odd number of cells, and discarding the remaining vertices. The **Tables 1** and **2** show the measures we obtained when we converted 3D voxelizations, taken from the mentioned research groups, to our implementation in the Java Language, of the nD-EVM when n = 3. We report the following execution times:

- Time for computing the conversion from 3D voxelization to 3D-EVM.
- Time for computing the content of the 3D-OPP (**Section 2.7**) expressed under the 3D-EVM.
- Time for computing the boundary content of the 3D-OPP (**Section 2.8**) represented by the EVM.
- Time for computing the 2D sections, starting from the 2D couplets (**Section 2.6**) perpendicular to X_1-axis, of the resulting 3D-OPP.

The descriptions corresponding to the set of objects being modeled in each voxelization are given in **Tables 1** and **2**, as well as the total number of voxels in each representation. The conversions and tasks were performed in a computer with Intel Celeron Processor at 900 Mhz and 256 megabytes in RAM memory.

From **Tables 1** and **2** can be observed, in first place, that the processing time for the conversion from the 3D voxelizations to the 3D-EVM was the largest from all the considered tasks. In this situation model *Skull* (**Table 2**) required 2,626.757 seconds (almost 45 minutes) for its conversion while, on the other hand, the shortest time for conversion corresponds to the model *Marschner/Lobb* (**Table 1**) with 1.332 seconds. It is interesting to observe that the model *Skull* does not contain the maximum number of extreme vertices in its corresponding EVM. The 3D model *Leg of Statue* (**Table 1**) was represented with 1,604,538 extreme vertices while *Skull* was represented with 1,302,134 extreme vertices, however, *Leg of Statue* required 480.902 seconds for its conversion to the 3D-EVM (approximately 8 minutes). The model *Marschner/Lobb* has both the lowest number of extreme vertices (872) and the lowest processing time for its conversion (1.332 seconds).

Now, let p be a 3D-OPP expressed under a voxelization with size ($x_1Size \times x_2Size \times x_3Size$) and with $EVM_3(p)$ as its corresponding EVM. Consider the ratio

$$\frac{x_1Size \cdot x_2Size \cdot x_3Size}{Card(EVM_3(p))}$$

As can be seen, the idea behind such ratio is to express the number of times the quantity of voxels in the original representation of the object p is greater than the number of extreme vertices in its corresponding representation through the 3D-EVM. For example, consider model *CSF* (**Table 2**). Its source voxelization has size (256 × 256 × 124) which implies that we require to store 8,126,464 voxels. The 3D-EVM associated to *CSF* has 86,570 extreme vertices (see **Table 2**). Hence, our proposed ratio gives us the value 93.87 which implies that the number of stored voxels that belong to the original representation of the object is precisely 93.87 times greater than the number of obtained

extreme vertices. The **Table 3** shows the ratio Number-of-voxels/Number-of-Extreme-Vertices for the models described in **Tables 1** and **2**. According to the results we obtained, the number of voxels in the model *Lobster* is 131.38 times greater than the cardinality of its corresponding 3D-EVM. In fact, model *Lobster* has the largest ratio from our seven tested models. On the other hand, the model *Leg of Statue* (**Table 1**) has a number of voxels which is 6.73 times greater than the number of extreme vertices required for representing it. The value shared by our ratio depends on the topology and geometry of the objects being modeled, but it shows to us the conciseness, related to storing requirements, when we represent such objects through the EVM.

Table 1. Results from the conversion of 3D voxelizations (*Bonsai*, *Leg of statue*, *Lobster* and *Marschner/Lobb*) to the 3D-EVM.

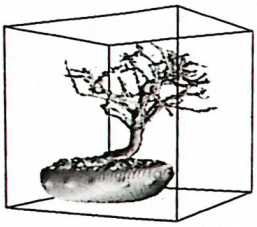
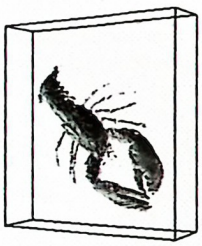
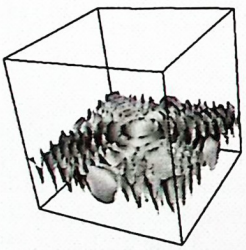
	Voxelization size:	$(256 \times 256 \times 256) \equiv 16{,}777{,}216$ voxels
	Name and Description:	**Bonsai.** Computed tomography of a bonsai tree.
	EVM size:	**641,462**
	Time for conversion:	**858.905 s**
	Content:	3,412,818 u^3
	Time for computing 3D content:	**33.709 s**
	Boundary content:	2,098,246 u^2
	Time for computing boundary content:	**71.322 s**
	Time for computing 2D sections:	**36.282 s**
	Voxelization size:	$(341 \times 341 \times 93) \equiv 10{,}814{,}133$ voxels
	Name and Description:	**Leg of statue.** Computed tomography of a leg of a bronze statue.
	EVM size:	**1,604,538**
	Time for conversion:	**480.902 s**
	Content:	2,499,595 u^3
	Time for computing 3D content:	**136.847 s**
	Boundary content:	3,244,280 u^2
	Time for computing boundary content:	**367.579 s**
	Time for computing 2D sections:	**130.257 s**
	Voxelization size:	$(301 \times 324 \times 56) \equiv 5{,}461{,}344$ voxels
	Name and Description:	**Lobster.** Computed tomography of a lobster contained in a block of resin.
	EVM size:	**41,566**
	Time for conversion:	**335.722 s**
	Content:	3,831,352 u^3
	Time for computing 3D content:	**0.921 s**
	Boundary content:	263,910 u^2
	Time for computing boundary content:	**1.582 s**
	Time for computing 2D sections:	**0.751 s**
	Voxelization size:	$(41 \times 41 \times 41) \equiv 68{,}921$ voxels
	Name and Description:	**Marschner/Lobb.** Simulation of high frequencies where 99% of the sinusoids are right below the Nyquist frequency.
	EVM size:	**872**
	Time for conversion:	**1.332 s**
	Content:	68,637 u^3
	Time for computing 3D content:	**0.010 s**
	Boundary content:	11,070 u^2
	Time for computing boundary content:	**0.020 s**
	Time for computing 2D sections:	**0.001 s**

Table 2. Results from the conversion of 3D voxelizations (*Silicium, Skull* and *CSF*) to the 3D-EVM.

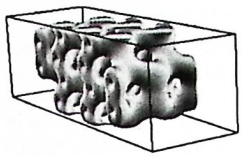
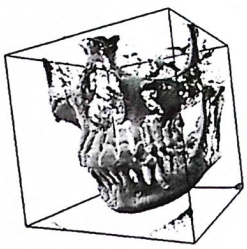
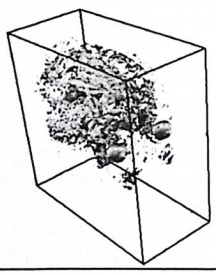
	Voxelization size:	$(98 \times 34 \times 34) \equiv 113{,}288$ voxels
	Name and Description:	**Silicium.** Simulation of a silicium grid.
	EVM size:	**1,164**
	Time for conversion:	**1.412 s**
	Content:	66,163 u^3
	Time for computing 3D content:	**0.010 s**
	Boundary content:	11,274 u^2
	Time for computing boundary content:	**0.020 s**
	Time for computing 2D sections:	**0.001 s**
	Voxelization size:	$(256 \times 256 \times 256) \equiv 16{,}777{,}216$ voxels
	Name and Description:	**Skull.** Rotational computer arm x-ray scan of a human skull.
	EVM size:	**1,302,134**
	Time for conversion:	**2,626.757 s**
	Content:	14,834,427 u^3
	Time for computing 3D content:	**87.907 s**
	Boundary content:	2,656,000 u^2
	Time for computing boundary content:	**245.873 s**
	Time for computing 2D sections:	**78.723 s**
	Voxelization size:	$(256 \times 256 \times 124) \equiv 8{,}126{,}464$ voxels
	Name and Description:	**CSF.** Dataset corresponding to the Cerebro-Spinal-Fluid in a human head.
	EVM size:	**86,570**
	Time for conversion:	**798.458 s**
	Content:	8,115,182 u^3
	Time for computing 3D content:	**2.293 s**
	Boundary content:	323,810 u^2
	Time for computing boundary content:	**3.395 s**
	Time for computing 2D sections:	**2.283 s**

Table 3. The ratio Number-of-voxels/Number-of-Extreme-Vertices for the 3D voxelizations shown in **Tables 1** and **2**.

Object p	Voxelization Size (Number of voxels)	Card(EVM_3(p))	$\frac{x_1Size \cdot x_2Size \cdot x_3Size}{Card(EVM_3(p))}$
Bonsai	16,777,216	641,462	26.15
Leg of Statue	10,814,133	1,604,538	6.73
Lobster	5,461,344	41,566	131.38
Marschner/Lobb	68,921	872	79.03
Silicium	113,288	1,164	97.32
Skull	16,777,216	1,302,134	12.88
CSF	8,126,464	86,570	93.87

The importance behind a "real world" 3D dataset is the information can be obtained about it. As commented in the introduction of this section, we computed for each model shown in **Tables 1** and **2**, through its corresponding 3D-EVM, the volume (u^3), the area of its boundary (u^2), and its sections. The last task refers to the case if our 3D models are represented through the 3D-EVM then, according to the procedures mentioned in **Section 2.6**, their sections will describe to us the interior of the modeled objects with the objective to perform the appropriate analyses according to the application. The 3D model *Leg of Statue* (that with the highest number of extreme vertices) required the maximum processing times for computing both its volume as the area of its boundary: 136.847 seconds and 367.579 seconds respectively. Conversely, the 3D model *Marschner/Lobb* (that with the lowest number of extreme vertices) required the minimum processing times for computing both its volume as the area of its boundary: 0.01 seconds and 0.02 seconds respectively. In the case related to

the computing of sections, we found again that *Leg of Statue* required the maximum processing time: 130.257 seconds. A "tie" was found in the minimum processing time for sections corresponding to 3D models *Marschner/Lobb* and *Silicium* (**Tables 1** and **2** respectively): their sections were computed in only one millisecond.

4 Conclusions and Future Work

In this work we have described the **Extreme Vertices Model in the n-Dimensional Space (nD-EVM)**. The Extreme Vertices Model allows representing nD-OPP's by means of a single subset of their vertices: the *Extreme Vertices*. **Section 2** is in fact a very brief description of the capabilities of the model because we have developed simple and robust algorithms, besides the ones presented in this work, for performing the most usual and demanding tasks on polytopes modeling such as closed and regularized Boolean operations, boundary extraction, and set membership classification operations (see [2] and [7] for more details). These procedures and algorithms provide us a way to obtain much more information than the extracted from the examples presented in **Section 3**. Moreover, **Table 3** shows the conciseness of the nD-EVM respect to voxelizations because we have obtained in all our described examples that the ratio Number-of-voxels/Number-of-Extreme-Vertices is between 6.73 and 131.38.

In this work we have concentrated about the representation of binary voxelizations through the EVM. However, it is well known that some devices express their outputs through 3D datasets whose voxels contain more information besides the empty/occupied property. In this sense, we propose as future work the representation of these datasets through nD-OPP's by considering their voxels' additional properties as independent spatial dimensions. Our proposal implies that these datasets can be expressed and manipulated as polytopes whose number of dimensions in greater than three. Furthermore, we will consider the representation of time varying 3D and 4D datasets through the nD-EVM. Because of the intensive use of the XOR operation in the nD-EVM (**Sections 2.4**, **2.5** and **2.6**), we exploit spatial and temporal redundancies between 2D and 3D frames, in these 3D and 4D datasets respectively, without compromising the quality of the information for the required purposes. We can guarantee, according to previous experiences, the model will represent these datasets in a very concise way.

Finally, we mention the development of other "real world" practical applications under the context of the nD-EVM, which are widely discussed and modeled in [7]. These practical applications, through we have showed the versatility of application of the nD-EVM, consider: (1) the representation and manipulation of 2D and 3D color animations; (2) a method for comparing images oriented to the evaluation of volcanoes' activity; (3) the way the nD-EVM enhances Image Based Reasoning; and finally, (4) an application to collision detection between 3D objects through the nD-EVM. As previously commented, details and results about these four practical applications can be found in [7].

References

1. Aguilera, A. & Ayala, D. Orthogonal Polyhedra as Geometric Bounds in Constructive Solid Geometry. Fourth ACM Siggraph Symposium on Solid Modeling and Applications SM'97, pp. 56-67. Atlanta, USA, 1997.
2. Aguilera, A. Orthogonal Polyhedra: Study and Application. PhD Thesis. Universitat Politècnica de Catalunya, 1998.
3. Coxeter, H.S.M. Regular Polytopes. Dover Publications, Inc., New York, 1963.
4. Jonas, A. & Kiryati, N. Digital Representation Schemes for 3-D Curves. Technical Report CC PUB #114, The Technion - Israel Institute of Technology, Haifa, Israel, July 1995.
5. MoViBio Research Group (Modeling and Visualization of Biomedical data). Web Site: http://truja.lsi.upc.edu/movibio (Site visited in May, 2007).
6. Pérez-Aguila, R. The Extreme Vertices Model in the 4D space and its Applications in the Visualization and Analysis of Multidimensional Data Under the Context of a Geographical Information System. MSc Thesis. Universidad de las Américas, Puebla. Puebla, México, May 2003.
7. Pérez-Aguila, R. Orthogonal Polytopes: Study and Application. PhD Thesis. Universidad de las Américas - Puebla. Cholula, Puebla, México, November 13th, 2006.
8. Spivak, M. Calculus on Manifolds: A Modern Approach to Classical Theorems of Advanced Calculus. HarperCollins Publishers, 1965.
9. VolRen: Project Volume Rendering with Graphics HW. Web Site: http://www.gris.uni-tuebingen.de/areas/volren (Site visited in May, 2007).

Design A Perceptron Neural Network Block With Complementary Training Function For Industrial PLCs By Ladder Language

Hamid Abdi
Faculty member of Arak faculty of science and technology
Iran University of science and technology
Email: HamidAbdi@iust.ac.ir

Abolfazl Salami, Abolfazl Ahmadi
Faculties member of Arak faculty of science and technology
Iran University of science and technology

Majid Abdi
Research Manager of Electronic development Co Of researchers
Email: Majidabdi@aryarobot.com

Abstract: Programmable logic controllers are the main controllers in the today's industries; they are used for several applications in industrial control systems and there are lots of examples from the PLC applications in industries especially in big companies and plants such as refineries, power plants, petrochemical companies, steel companies, and food companies. In PLCs there are some functions in the function library in the programming and configuration software that can be used in PLC programs as basic program elements. The aim of this project are introducing and implementing a new function block of a neural network to the function library of PLC. This block can be applied for some control applications or nonlinear functions calculations if it had been trained for desired applications. The implemented neural network is a Perceptron neural network with three layers, three input nodes and one output node. The block can be used in manual or automatic mode.
In this paper the structure of the implemented function block, the parameters and the training method of the network are presented. All of these are showed by considering especial subjects in PLC programming and its complexities. Finally the application of the new block is compared with a classic simulated block and the results are presented.

Keywords: Programmable Logic Controller, PLC Programming, Neural Networks, Perception Network, Intelligent Control

1. Introduction

Developing the new control strategies and implementation and extending new calculation tools in PLC systems are interesting subjects for PLC manufacturers

© L. Sánchez, O. Pogrebnyak and E. Rubio (Eds.)
Industrial Informatics
Research in Computing Science 31, 2007, pp. 25-32

and special advanced users. Manufacturers presented new functions by introducing their new PLCs. They try to improve their last functions, simple to understand or applying by PLC users [6], [7]. So declaration of the new function blocks is an applicable research field in industrial control, however industrial applications of the neural network are limited by needs of the high level engineering knowledge in control, intelligence or level of automation and the applications. Moreover it needs to increase the reliability of neural network function block especially for industrial applications.

In PLCs, there are several mathematical and operational function blocks which can be used easily but neural network function block has not be defined in PLC which we know them [1],[6],[7]. Our study on most known PLC manufacturer products like siemens, Omron, Allen Bradley ,Yokogawa , Hitachi and ... show that NN is nut implemented with PLCs yet. In this paper in the first section the importance of the subject and prior art are discussed.

Then in part II artificial neural networks are reviewed and then in Part III PLCs and PLC programming are studied and in part IV and V structure of the implemented neural network for PLC application and it's operation are introduced then in part VI the function block operation is analyzed in comparison with a simulated system and finally some conclusion have been made and the results are presented.

2. NEURAL NETWORKS

In a neuron, inputs information and parameters excite the input channel of a real neuron with an electric signal. Then these input signals are manipulated by the synapse (Weight) and the effect of them with impact of the other parameter (Bias) is used to excite the output of the neuron. In artificial NN synapses weight the input signals and they are added with a bias value finally decision function produce the output of the artificial NN. So neurons it produces a constant current in the output paths of the neuron toward other neurons that have link of them [3], [6].

An Artificial NN is a network of neurons model. NN has some distinct mathematical and graphical models. These models are used to show the structure of suggested artificial neural network.

To use a NN the parameters of the NN should be trained at first and then it can be used. There are two main topologies in neural networks. The first one is feed forward and the second one is feedback topology. Most of the recent applications of neural network use the feed forward topology. The most conventional method for feed forward NN training is the back error propagation method [4], [5].

Perceptron is one of the feed forward neural networks. Figure-1 Multilayer Perceptron networks are a example of feed forward neural network which have several applications in complex mathematical problem solving, classification and nonlinear equation solving. Multi layer Perceptron neural network can be used by proper selection of the network parameters and decision functions [3], [2] for a lot of problems.

In these networks there are an input layer as an input part of the network, an output layer as the output part of the network and some hidden layers which are used between input and output layers (Figure-1). In these networks the information passes straightforward the network to produce the output. There are considerable contributions on the use of perceptron neural networks for industrial control or calculation problems. [3], [4]

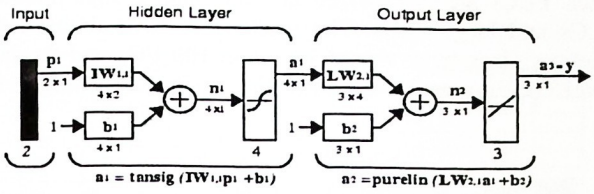


Fig-1 Multi layer Perceptron neural network

Table-1 Training routine of the Perceptron neural networks

$E(w)=\frac{1}{2}\sum_{j=1}^{p}(y_j-d_j)^2$	Y_j: network j-th output value, d_j : target value of j-th output, p see [1]: number of neurons in the output layer	1
$\Delta w_{ij}=-\eta\cdot\frac{\partial E}{\partial w_{ij}}$ $\Delta w_{ij}=-\eta\cdot\frac{\partial E}{\partial w_{ij}}$ $\frac{\partial E}{\partial w_{ij}}=\frac{\partial E}{\partial y_j}\cdot\frac{dy_j}{dS_j}\cdot\frac{\partial S_j}{\partial w_{ij}}$	These are the weight changes due to the j-th output y_j – j-th neuron output value, S_j – weighted total of output signals determined by the formula (1). See[1]	2
$\frac{\partial S_j}{\partial w_{ij}}\equiv x_i$	x_i is the neuron i-th input value See[1]	3
$\frac{\partial E}{\partial y_j}=\sum_k\frac{\partial E}{\partial y_k}\cdot\frac{dy_k}{dS_k}\cdot\frac{\partial S_k}{\partial y_j}=\sum_k\frac{\partial E}{\partial y_k}\cdot\frac{dy_k}{dS_k}\cdot w_{jk}^{(n+1)}$	K is the number of neurons in layer n+1 See[1]	4
$\delta_j^{(n)}=\frac{\partial E}{\partial y_j}\cdot\frac{dy_j}{dS_j}$	$\delta_j^{(n)}$ n-th layer if we know $\delta_k^{(n+1)}$ of the next (n+1) layer. See[1]	5
$\delta_j^{(n)}=\left[\sum_k\delta_k^{(n+1)}\cdot w_{jk}^{(n+1)}\right]\cdot\frac{dy_j}{dS_j}$	$\delta_j^{(n)}$ for the last layer of the neuron See[1]	6
$\delta_j^{(N)}=\left(y_j^{N}-d_j\right)\cdot\frac{dy_j}{dS_j}$	Network the vector of those values that the neural network must generate with the given set of input values. See [1]	7
$\Delta w_{ij}^{(n)}=-\eta\cdot\delta_j^{(n)}\cdot x_i^n$	And finally let's write down the formula (6) in expand form Weigh function changes	8

3. Programmable logic controllers

PLCs are the most applied controllers in industrial application especially for interlocking and supervising control to make sequences of operation or safety logics. However the today's PLCs have control abilities for applying in continuous systems (like PID facility). It should be noticed that there are more than 100 manufacturers in the PLC forum. They produce various types of PLCs in different sizes. PLCs are categorized in three main groups including small PLCs, medium PLCs and large PLCs. The small sizes have almost less than 40 inputs/outputs, medium sizes have almost less than 100 I/Os and the large sizes have more than 500 I/Os. Figures 2 and 3 show the application strategy of a large PLC.

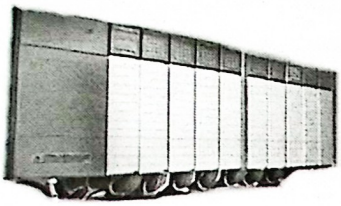


Fig-2 a large PLC

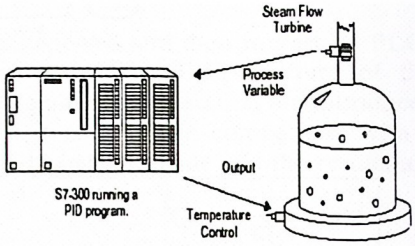


Fig-3 PLC application terminology

It should be noted that for comparing PLCs there are several other items. Some characteristics such as CPU, Cycle time, simplicity, reliability, expansion capability, peripherals, networking features are some considerable items.

The PLC systems can be programmed with three languages. These are Ladder programming (LAD), function block diagram programming (FBD) and statement list programming (STL). These methods of programming vary in different PLCs but they have a little difference between most of the PLCs. To write a structural PLC program, we have to use some basic instructions or some programmed blocks. It also needs to write some new blocks which should be programmed for the application.

The organization blocks (OB), program blocks (PB) and some function blocks (FB,FC) are some of these blocks in Siemens PLCs. For example OB blocks are function blocks manage the PLCs and they act like an operating system of a computer. They control PLC operation sequences and hardware parts of the PLC.

The PB blocks are program blocks; they can be used as the basic applied control program or the subroutine of the user programs. The FC blocks are basic applied functions in the PLC systems. They perform a simple software or hardware task which generally can be used by the user program. Some of these blocks have been designed or reserved by PLC manufacturers and some others can be programmed by the user of the PLC due to the applications. Usually FC blocks are used in PB Blocks; PB and FC blocks are used in OB blocks. It should be noted that the main

program which calls the other PB blocks or FB Blocks or even FC Blocks should be in OB1 block. One of the function blocks which have been used as a classic controller is shown in figure 4.
This structure is used in Siemens S7 PLCs generation; in the same manner other PLCs use the like structures for programming. The differences between these PLCs and other manufacturer's products are less than almost 10%. The differences are only in the names of the blocks and some indicative symbols or methods of combination.

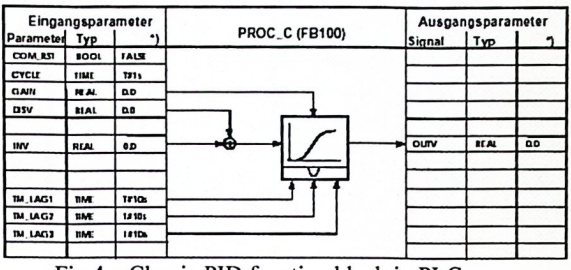


Fig 4 – Classic PID function block in PLC

4. Implementation of neural network function block

The implemented block is a three layers Perceptron neural network. The block includes three input nodes and seven nodes in second layer and one node in output layer. The block in PLC S7300 has two automatic and manual modes. In the manual mode the parameters of the network are declared manually in the program (W01...W21, B01...B21). They are set by the user when the network is used. However if they are not defined some uniform random number are selected. In the automatic mode parameters of the network are calculated automatically by the back error propagation training method.
It is the most conventional method for training of these networks. The network is completely trained after almost 10 minute of the time which the PLC is placed in RUN mode (for 500 data for training). The training data should be sorted in some parts of the PLC memory and it makes a matrix with four vectors including three column vectors as input data and one other vector as the output date of the network. The inputs and output data are used for training the network parameters. Every training sequence is calculated in one machine cycle of the PLC system. (The machine cycle is almost 1 microsecond for our test program)

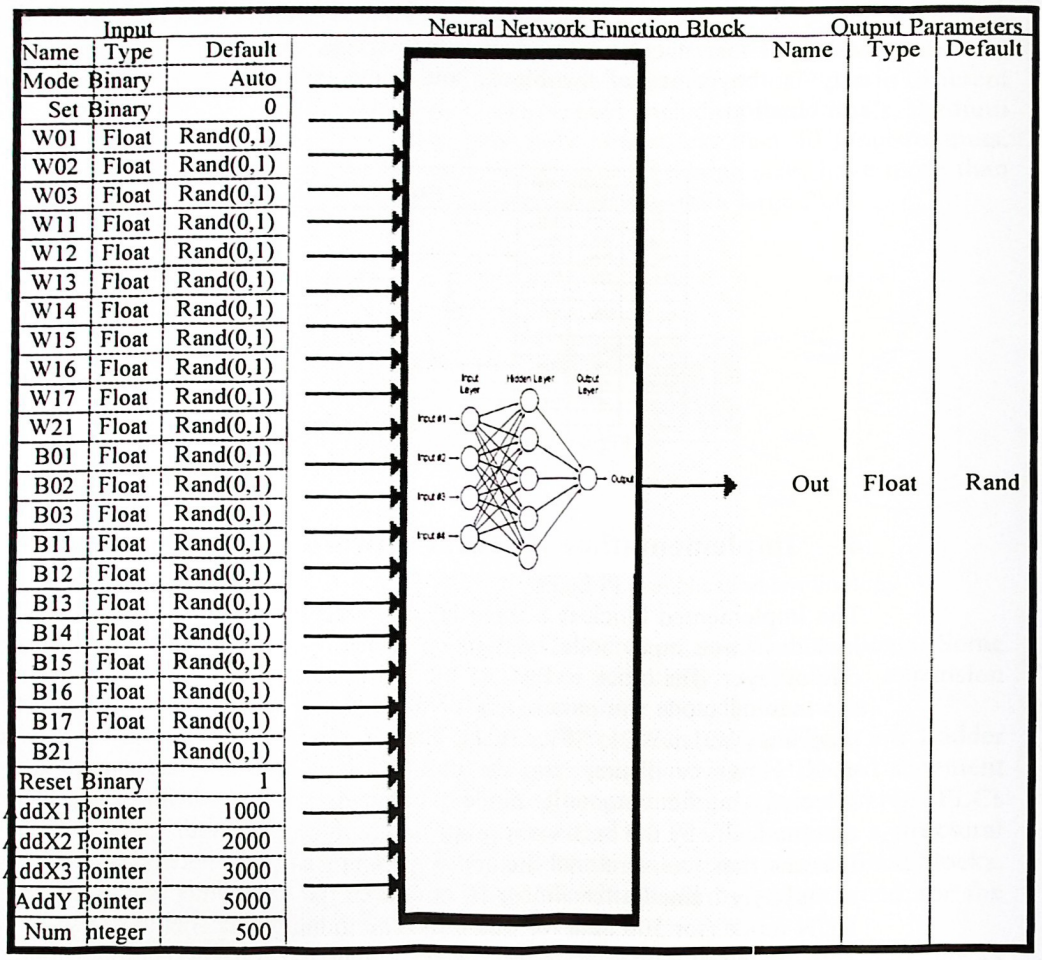


Fig-5 Implemented function block

5. Operation

To test the designed network in a siemens PLC, the behavior of a PID controller for a first order simple model is studied in Matlab software. The model is used to extract data to form four vectors, the samples are gathered from almost 8 minutes analyzing of the PID controller on the model with 1000 sampling points. Table 2 shows the sampled data variables and related their considerations. The information of the PID controlled systems transferred to the desired places in PLC memory to make the training matrix for the neural network. Then the block is trained by the matrix. After that the behavior of the network is compared with the PID controlled in the model to analyze the operation of the implemented neural network function block. Figure 6 compares the Function block in PLC with the PID controller due to applying a five level set point both in model and PLC. Some other inputs are used in manual mode for manual parameter defining of the NN. Others are for control the operation and stages.

Table 2- Training matrix variables

Type	Name	Address	NN Name
Input	Set Point	MW1000-1999	X0
	Error	MW2000-2999	X1
	Control Output	MW3000-3999	X2
Output	Process Variable	MW5000-5999	Y

6. Conclusion

In this paper the structure of neural networks by aiming of the implementation in the industrial PLCs are studied. The new function block is introduced for the PLCs applications and the operation of the block is studied too. Then the accuracy of the implemented block is investigated regarding to the calculation and programming limits of the PLCs. In addition the back error propagation training method is implemented by PLC software as a complementary PB. Then the implemented function is analyzed by using a computer controlled model. The presented results show that the function performs the desired operation. It should be noted that this application is gathered based on the fully industrial controller

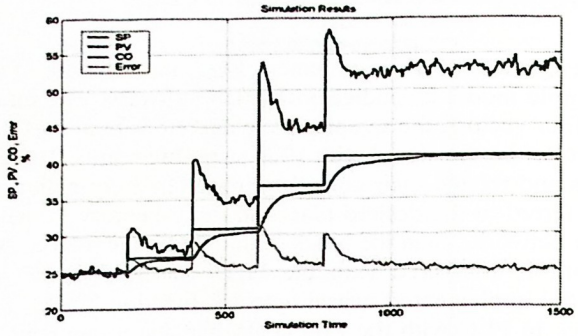


Fig 6 – Neural network block output compare to PID output

7. Acknowledgments

The authors would like to appreciate NIOC and Rahavaran's automation laboratories and IUST University, Arak faculty of engineering.

8. References

[1] VOICU, "Advances in automatic control", Publication date : 10-2003
[2] Goldman, RM. "Mathematical Methods for Neural Network Analysis and Design" MIT Press, 1996.
[3] Hinton, G.E, "How neural networks learn from experience", Sci. American, 1992
[4] MATH WORKS Inc, "Neural Network Toolbox User's Guide" , "MATLAB USER MANUAL " , 1992-2002
[5] Beale & Jackson, "Neural Computing: An Introduction." Inst. Phys. 1996
[6] Technical Manual References of Siemens PLC- S7300
[7] Technical Manual References of Siemens PLC- S7400
[8] [7] Technical Manual References of Omron PLC- CVM
[9] Hamid Abdi, "Control in petroleum industries review and technology trend", national conference on control and communication in petroleum industry, Sep 2003, Sharif University of technology, Tehran, Iran

Effective Threshold Control in a Coded Communication System with Information Feedback Driven by Gaussian and Impulsive Noises

Vladimir Kazakov, Francisco Tejeda

Department of Telecommunications Engineering. SEPI-ESIME-IPN
Av. Instituto Politécnico Nacional s/n. 07738 México D.F., México
Phone (+5255) 57296000 ext. 54757
ftejeda@ipn.mx, vkazakov41@hotmail.com

Abstract. The expressions for the error probability and the average consumption of energy for the coded communication system with information feedback are obtained. A simple effective method of the threshold control at the feedback channel receiver is suggested when the Gaussian and Impulsive noises are present in the system. This method makes possible to avoid the useless consumption of energy in the coded system when the signal to noise ratio on the feedback channel is small and to improve the efficiency in other cases, taking into account the sum of both noises yields a quite destructive random process.

Keywords: Gaussian white noise, impulsive noise, coded communication system with information feedback.

1 Introduction

The Communication Systems with Information Feedback (CSIF) are within the communication systems with feedback channel [1-10]. The efficiency of the coded CSIF is characterized by the average consumption of energy and the error probability. We obtain the general expressions for these main characteristics of the CSIF. Furthermore, when the signal to noise ratio in the feedback channel is small, there is a useless consumption of energy, in other cases there is a chance to improve the characteristics of the CSIF. We suggest a simple and effective threshold control in the feedback channel receiver.

Let us describe shortly the logic operation of the coded CSIF [11]. At the transmitter part the message $a_k(\overline{1,N})$ is stores in memory, it is coded by the codeword A_k, and after that the word A_k is sent through direct channel. In the input of the direct channel receiver there is the codeword $\tilde{A} = A_k + S_d$ (here S_d is a disturbed combination occurred owing to the presence of noise in the direct channel). The space of the signals in the output of the direct channel decoder is $G = \{G_1, G_2, \dots, G_N\}$, where $G_i(i = \overline{1,N})$ is the subspace of the codeword A_i and correspondingly of the message a_i. The received message a_i is stored in memory and is temporally not delivered to the recipient. On the basis of a_i the feedback channel coder generates the codeword B_i (generally from another code) and sends it through the feedback channel.

© L. Sánchez, O. Pogrebnyak and E. Rubio (Eds.)
Industrial Informatics
Research in Computing Science 31, 2007, pp. 33-42

In the input of the feedback channel decoder there is a disturbed codeword $\tilde{B} = B_k + S_f$ (here S_f is a disturbed combination occurred owing to the presence of noise in the feedback channel). The space of the signals in the output of the feedback channel decoder is $D = \{D_1, D_2, \ldots, D_N\}$ where $D_j (j = \overline{1,N})$ is the subspace of the codeword B_j and correspondingly of the message a_j. In the transmitter side there is a comparison scheme. At the first input of this comparison scheme the message a_k is delivered from the memory block, at the second input the result of the decision of the feedback channel decoder is passed. When $\tilde{B} \in D_k$ occurs, there is a coincidence between the message sent through the direct channel and the message received in the feedback channel receiver. In this case, the comparison scheme produces an acknowledgement control signal in order to send the next message through the direct channel. The acknowledgement control signal is received by the direct channel receiver which sets free the message stored in its memory and it is delivered to the recipient. When $\tilde{B} \notin D_k$ there is no coincidence between the index of the sent and received messages in the input of the comparison scheme. In this case, a rejection control signal is generated by the comparison scheme. The rejection control signal and the same message a_k are sending from the transmitter part. The reception of the rejection control signal blocks out the delivery of the received message from the memory of the direct channel receiver to the recipient. This process goes on until the indexes in both inputs of the comparison scheme are the same and the system passes to the transmission of the next message.

2 New random process

Let us consider *RC* filters are in the receivers of both direct and feedback channels. Let us assume that in the input of the *RC* filter the following process is present on both channels

$$\xi_i(t) = \theta s_{in_i}(t) + n_i(t) + m_i(t), \tag{1}$$

where θ is a parameter of presence of the signal (i.e., θ=0 the signal is no present; θ=1 the signal is present), the index i indicates the channel (it could be direct or feedback), $s_{in}(t)$ is the input rectangular signal with amplitude A_i, $n_i(t)$ is the Gaussian process and $m_i(t)$ is the impulsive process. The resulting output process is expressed

$$x_i(t) = s_{out_i}(t) + x_{1_i}(t) + x_{2_i}(t), \tag{2}$$

where

$$s_{out_i}(t) = A_i(1 - e^{-\alpha_i t}), \tag{3}$$

$$\alpha_i = \frac{1}{R_i C_i}, \tag{4}$$

α_i is the constant of the *RC* filter, $s_{out}(t)$ is the output signal, $x_1(t)$ is a Gaussian random variable and $x_2(t)$ is a random variable described by the Beta distribution [12]. The Probability Density Function (PDF) of the Gaussian random variable is well known

$$w_1(x_1) = \frac{1}{\sigma_1\sqrt{2\pi}} e^{-\frac{(x_1-\theta s_{out}(T))^2}{2\sigma_1^2}}, \tag{5}$$

where $T=T_d=T_f$ is the duration of the signals in both channels, σ_1^2 is the variance in stationary state and is known; its mathematical expectation depends on the parameter θ, it means, if $\theta=0$, $s_{out}(T) = 0$; if $\theta=1$, $s_{out}(T) = A_i(1 - e^{-\alpha_i T})$. The Beta distribution is defined by

$$w_2(x_2) = \frac{x_2^{\lambda_1-1}(1-x_2)^{\lambda_2-1}}{B(\lambda_1,\lambda_2)}, 0 \le x_2 \le 1, \tag{6}$$

where λ_1, λ_2 are the Poisson's flow intensities, $B(\cdot)$ is the Beta function. The mathematical expectation is

$$\langle x_2 \rangle = \frac{\lambda_1}{\lambda_1 + \lambda_2}, \tag{7}$$

and its variance

$$\sigma_2^2 = \frac{\lambda_1\lambda_2}{(\lambda_1 + \lambda_2)^2(\lambda_1+\lambda_2 + 1)}. \tag{8}$$

In order to get the PDF of the sum of the Gaussian and impulsive noises the convolution theorem is applied

$$w(x) = \int_{-\infty}^{\infty} w_1(x_1 - x_2)\, w(x_2) dx_2 . \tag{9}$$

The resulting PDF depends on many parameters inherit from the two random processes. However, the Poisson's flow intensities have a strong influence in the shape of the new PDF, as it could be symmetrical ($\lambda_1 = \lambda_2$) or asymmetrical ($\lambda_1 \neq \lambda_2$). The mathematical expectation of this new PDF is determined by

$$\langle x \rangle = \theta s_{out}(T) + \langle x_2 \rangle , \tag{10}$$

and the variance is

$$\sigma^2 = \sigma_1^2 + \sigma_2^2 . \tag{11}$$

3 General expressions

We analyze the coded CSIF with the following assumptions: 1) the number of possible repetitions of the message is equal to infinity; 2) the transmission of the acknowledgement and rejection control signals is ideal; 3) the coding schemes in both channels are not changed in the repetition process.

We note the principal difference between the problems solved by both decoders: the direct channel decoder must solve the problem of determining one among many possible words; therefore, in its output any word $A_l(\overline{1,N})$ can appear; the feedback channel decoder must solve the problem of the detection of one known word B_k because the index k is a priori known in the transmitter part of the system. In other words, the feedback channel decoder solves the following binary problem: $\tilde{B} \in D_k$ or $\tilde{B} \notin D_k$ (Fig. 1). Taking into account this difference, one can characterize any cycle of the message a_k transmitted by the following three probabilities:

1) the probability of correct reception $\hat{p}$

$$\hat{p} = P_k[\tilde{A} \in G_k]P_k[\tilde{B} \in D_k], \tag{12}$$

where $P_k[\cdot]$ means the probability of the event $[\cdot]$ if the messages a_k and b_k were transmitted in the direct and feedback channels correspondingly.

2) The error probability $\hat{q}$

$$\hat{q} = \sum_{i \neq k} P_k[\tilde{A} \in G_i]P_i[\tilde{B} \in D_k], \tag{13}$$

3) The probability of error detection $\hat{u}$

$$\hat{u} = P_k[\tilde{A} \in G_k]\sum_{i \neq k} P_k[\tilde{B} \in D_{\sim}] + \sum_{i \neq k} P_k[\tilde{A} \in G_i]P_i[\tilde{B} \in D_i]. \tag{14}$$

It is clear that

$$\hat{p} + \hat{q} + \hat{u} = 1. \tag{15}$$

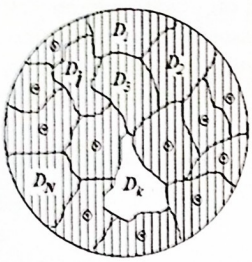


Fig. 1. The space D and the subspace of reception D_k of the codeword B_k.

However, we must specify these general expressions. We extend the coding formulas from a simple direct channel to the coded CSIF [13]. Therefore, the probability of correct reception $\hat{p}$ for any lineal and systematic code is

$$\hat{p} = \sum_{m=0}^{t_d} \binom{m}{n_{低}} q_d^m (1-q_d)^{n_d-m} \sum_{m=0}^{t_f} \binom{m}{n_f} q_{f_1}^m (1-q_{f_0})^{n_f-m}, \tag{16}$$

$$q_d = 0.5 \int_h^{+\infty} w_d(x)dx + 0.5 \int_{-\infty}^{h} w_d(x-\langle x \rangle)dx, \tag{17}$$

$$q_{f_0} = \int_{\beta_0}^{+\infty} w_f(x)\,dx, \tag{18}$$

$$q_{f_1} = \int_{-\infty}^{\beta_0} w_f(x-\langle x \rangle)\,dx, \tag{19}$$

where q_d is the probability of the symbol error in the direct channel; q_{f_0} and q_{f_1} are the error probabilities for the zero and one symbols in the feedback channel; where h is the threshold decisions in the direct receiver; β_0 is the "initial" threshold decision in the feedback channel receiver; $w_d(\cdot)$ and $w_f(\cdot)$ are the PDF's in the output of the RC filters in the direct and feedback receivers; t_d and t_f are the maximum capacity that the code schemes have to correct errors in the direct and feedback channels; m is the amount of errors that the codeword has; n_d and n_f are the length of the codeword in the direct and feedback channels. The expression for the error probability per codeword is

$$\hat{q} \le (N_d - 1) \sum_{m=0}^{t_d} \binom{m}{n_d} q_d^{n_d-m} (1-q_d)^{n_d-d_d+m} \sum_{m=0}^{t_f} \binom{m}{n_f} q_{f_1}^{n_f-m} (1-q_{f_0})^{n_f-d_f+m}, \tag{20}$$

where d_d and d_f are the minimum code distances in the direct and feedback channels; $N_d = 2^{k_d}$ is the number of codewords and k_d is the number of informative symbols in the codeword in the direct channel. The error probability per codeword of the CSIF with an infinity number of repetitions is determined by the formula:

$$Q = \hat{q} + \hat{u}\hat{q} + \hat{u}^2\hat{q} + \cdots = \frac{\hat{q}}{1-\hat{u}} = \frac{\hat{q}}{\hat{p}+\hat{q}}. \tag{21}$$

In the same manner, one can obtain the formula for the average consumption of energy <*E*> of one codeword transmitted

$$\langle E \rangle = \frac{n_d E_d + n_f E_f}{\hat{p}+\hat{q}}, \tag{22}$$

where E_d and E_f are the energy of any elementary signal in the direct and feedback channels. We divide (22) by the white noise spectral density $N_0/2$, thus

$$\langle Z^2 \rangle = \frac{n_d z_d^2 + n_f z_f^2}{\hat{p} + \hat{q}}, \tag{23}$$

where $z_d^2 = 2E_d/N_0$ and $z_f^2 = 2E_f/N_0$ are the square of the signal to noise ratio (SNR) in both direct and feedback channels.

The expressions (21) and (23) describe the principal characteristics of the coded CSIF regards to a codeword. But as it is known, different codes have a different efficiency with the same SNR. It means that in order to make a correct comparison among the efficiency of different codes in the coded CSIF, it is necessary to recalculate the expressions (21) and (23) per one symbol. Using the methodology described in [14], we can define the error probability equivalent per one informative symbol

$$q = 1 - (1 - Q)^{\frac{1}{k_d}}. \tag{24}$$

In order to calculate the average consumption of energy per one informative symbol, one can write down the following formula

$$\langle z^2 \rangle = \frac{\langle Z^2 \rangle}{k_d} = \frac{n_d z_d^2 + n_f z_f^2}{k_d(\hat{p} + \hat{q})}. \tag{25}$$

The formulas (24) and (25) are used for all calculations, since they provide a possibility to characterize the efficiency of any CSIF by the function $q(\langle z \rangle)$. It is very useful to introduce another function

$$q_0(z_0) = q_d(z_d), \tag{26}$$

where q_0 is the error probability in a communication system with one channel which uses amplitude modulation and the a priori probabilities are equal.

4 Example

For our goal it is enough to choose a simple variant of the coded CSIF. Namely, we suppose that the code (7, 4, 1) is used in both channels (here n_d=n_f=7, k_d=4, t_d=t_f=1). Besides this, we must take into account others parameters: the *RC* factor α = 1, the duration of the signal T = 1, the impulsive noise amplitude is 1. We choose the Poisson's flow intensities $\lambda_1 = \lambda_2 = 1$ since the simplest case for illustration is a symmetrical one. Nonetheless, these parameters could take others values. The results of calculations following the formulas (24) and (25) are presented in Fig. 2. There are two curves: 1) the curve *a* represents the function $q_0(z_0)$; 2) the curve *b* consists of two parts b_1 and b_2; it characterizes the efficiency of the coded CSIF $q(\langle z \rangle)$ when $z_d = 2$

and the value of z_f is widely changed ($0 \le z_f \le 3.5$), these values are marked by circles.

The first part b_1 is related to the case $z_f \le 1.1$ (or $z_f < z_d$). As one can see, when the parameter z_f increases from 0 to 1.1, the error probability q decreases a little ($-0.98 \le \log q \le -0.66$). Furthermore, we note that the average consumption of energy $\langle z \rangle$ is quite big in comparison with the values z_d and z_f. The reason of this effect is connected with a fairly error probabilities $q_{\dagger_0}$ and q_{f_1} which leads to the fairly probability of repetitions $\hat{u}$ as well as the increasing average number of repetitions. It is clear that this regime is very bad and undesirable owing to the huge consumption of energy and the enormous amount of errors in the auxiliary feedback channel. The efficiency of the coded CSIF in this regime under consideration is so much worse than the efficiency of the system with one channel which uses amplitude modulation and has the same parameter values because the location of the curve b_1 is higher than the location of the curve a.

The second part b_2 is related with the case $z_f > 1.1$. Here the situation is changed mainly: when z_f increases then the error probability q decreases quickly. The coded CSIF efficiency in this regime is higher than the system efficiency with one direct channel due to the fact that the location of the curve b_2 is lower than the location of the curve a. It is important to note that the calculations of the coded CSIF efficiency with others codes leads to the same shape of the function $q(\langle z \rangle)$.

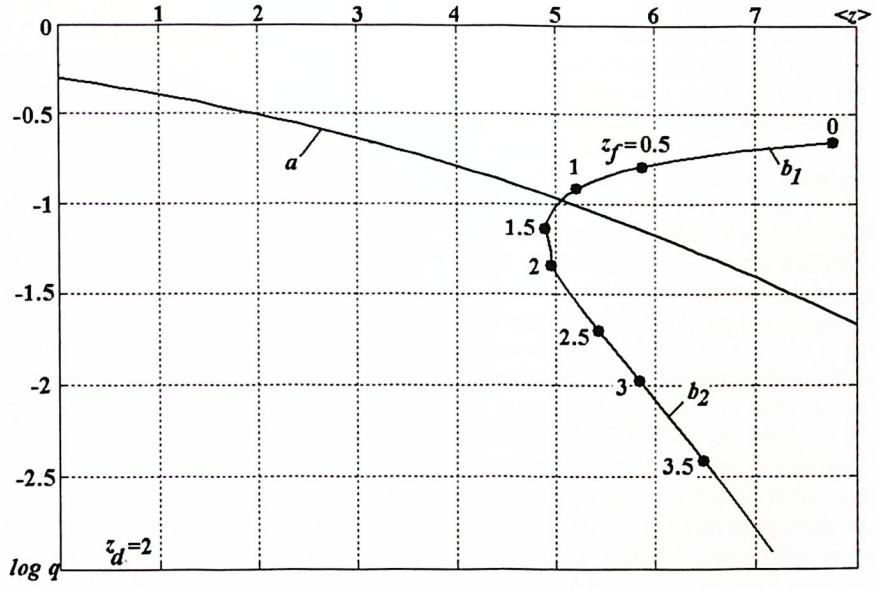


Fig. 2. The coded CSIF efficiency (curves b_1 and b_2) with the elementary signals reception in the feedback channel with initial threshold.

5 The proposed method

In order to avoid the regime describe by the curve b_1, it is theoretically necessary to increase the subspace D_k (see Fig. 1). In this case, the probability of the correct reception of the codeword B_k (correspondingly of the message a_k) must be increased also. This effect leads to decrease the number of repetitions and the consumption of energy. Hence, we need to change the feedback channel threshold β in a wide range so that we could obtain an optimal regime of system functioning. Now let us rewrite the formulas for the probabilities $\hat{p}$ and $\hat{q}$ taking into account an arbitrary threshold β in the feedback channel receiver. Then instead of (18) and (19), we have

$$q_{f_0} = \int_{\beta_0-\beta}^{+\infty} w_f(x)dx\,, \tag{27}$$

$$q_{f_1} = \int_{-\infty}^{\beta_0+\beta} w_f(x-\langle x\rangle)dx\,. \tag{28}$$

Putting (27) and (28) into (16) and (20), let us apply these formulas for the efficiency calculation of the CSIF with codes (7, 4, 1) on both channels. Like in Fig. 2, let us choose the same parameter values for the system, $z_d = 2$ and $z_f = 0.5, 1.0, 1.5, 2.0, 2.5, 3$. Changing the threshold β ($-\infty \le \beta \le +\infty$) and applying the recalculation procedure (24) and (25) per one informative symbol, we obtain the family of curves $q(<z>)$ depicted in Fig. 3.

The curve c is the envelope of this family and characterizes the optimal regime of functioning for the chosen parameter $z_d = 2$. The crossing points of the family curves $q(<z>)$ with curve b are marked by circles. As one can see, the initial thresholds are optimal near the value $z_f = 2.5$. In other cases, it is necessary to choose different values for the thresholds. If $z_f < 1.1$, the receiver must be slightly open for an expected codeword. The concrete values of the threshold can be determined as points of contact of the family curves with the envelope curve c as in table 1 is illustrated.

Table 1. The concrete values of the threshold regarding z_f in the envelope curve c.

z_f	$\mp\beta$
0.5	0.6
1.0	0.4
1.5	0.2
2	0.1
2.5	0

In table 1, the double signs mean the negative value for the symbol 1 and the positive value for the symbol 0. Such choice of thresholds avoids the issue discussed in Fig. 2: the bad regime with huge consumption of energy (compare the envelope c higher the point $z_f = 1.1$ and the curve b_1). The family curves $q(<z>)$ with $z_f \ge 1.1$ demonstrate

that it is necessary to decrease the subspace D_k. It means that the signs of the thresholds must be invert in the points of contact of the family curves with the envelope c. In this case we have: $z_f = 3, \pm\beta = 0.1$

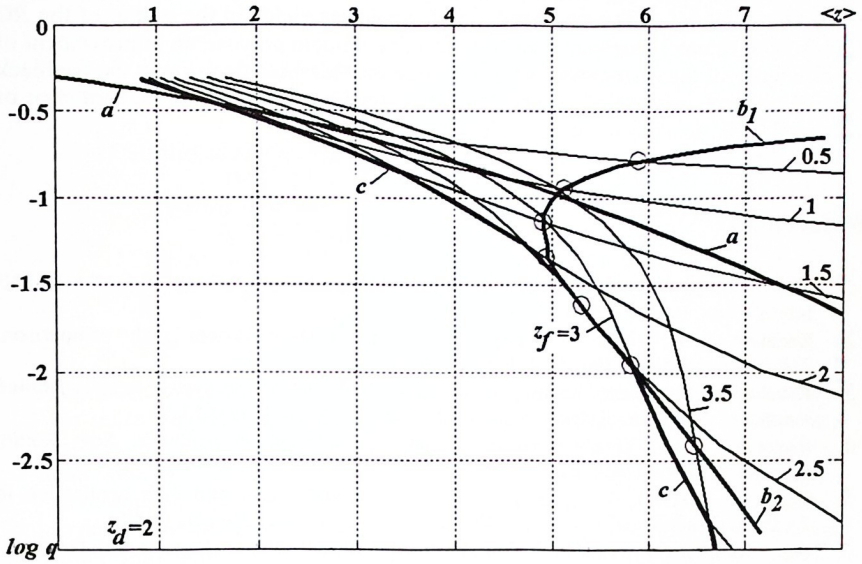


Fig. 3. The coded CSIF with several values of the parameter z_f when the thresholds in the feedback channel receiver are varied according to the proposed method.

The shape of the envelope c demonstrates: 1) the application of the proposed method provides a possibility of improving the CSIF efficiency in comparison with the variant of initial threshold. Actually, if $z_f < z_d$ (to be exact here $z_f < 1.1$) the undesirable regime with huge consumption of energy is avoided; if $z_f \geq z_d$, the error probability q decreases with the same consumption of energy. 2) if $z_f < z_d$ (to be exact here $z_f < 1.1$) the CSIF efficiency is worse than the efficiency of a communication system with one direct channel since the curve c is higher than curve a or curve $q_0(z_0)$.

It is necessary to emphasize that all features above mentioned are valid for any type of systematic code used in the CSIF. Nevertheless, the optimal thresholds β values and the corresponding values of q and $\langle z \rangle$ must be others. Although there exist others parameters which determine the performance of the system also, such as: Poisson's flow intensities, RC factor α, duration of the signal T, impulsive noise amplitude, etc.

6 Conclusion

A simple practical method of the threshold control in a feedback channel receiver is suggested when the coded CSIF is driven by Gaussian and impulsive noises. Despite of the fact that sum of Gaussian and impulsive noises yields in the output of the *RC* filter in both channel receivers a quite destructive random process, an improvement of the efficiency of the system was achieved when the threshold decision in the feedback channel receiver was varied in a wide range. In spite of the optimal performance of the coded CSIF depends on many parameters, it was found.

References

1. Chang, S.S.L.: Improvement of two-way communication by means of feedback. IRE Intern. Conv. Record, (1961) Pt. 4
2. Kazakov, V.A.: Noise immunity of a communication system with comparison. Telecommunications, Vol. 22, No. 2, (1968) 18-23
3. Kazakov, V.A.: Noise immunity of a memoryless comparison system with a limited number of repetitions. Telecommunications, Vol. 22, No. 8, (1968) 13-19
4. Kanevskij, Z.M.: Transmission of message by information feedback. Sov. Radio, Moscow (1968) – in Russian
5. Klove, T., Korzhik, V.: Error detected code. General theory and their application in feedback communication systems. Kluver academic publisher, Boston, (1995)
6. Ooi, J.M.: A framework for low-complexity communication over channels with feedback. Ph.D. dissertation, Massachusetts Institute of Technology, (1998)
7. Ooi, J.M.: Coding for channels with feedback. Kluver academic publisher, Massachusetts (1998)
8. Wicker, S.: Error control systems for digital communication and storage. Prentice Hall, New Jersey (2002)
9. Tatikonda, S. C.: Control under communication constraints. Ph.D. dissertation, Massachusetts Institute of Technology, (2000)
10. Klein, T. E.: Capacity of Gaussian noise channels with side information and feedback. Ph.D. dissertation, Massachusetts Institute of Technology, (2001)
11. Bargh, M. S.: Coding strategies for channels with feedback. Ph.D. dissertation, Eindhoven University of Technology, Netherlands, (1999)
12. Stuart, A., Ord, K.: Kendall's advanced theory of statistics. Vol. 1, Arnold, sixth edition, London, (1994) p. 200
13. Lin, S., Costello, D.: Error control coding. Prentice Hall, New Jersey (1983)
14. Blokh, E.L.: Noise immunity of a decision feedback communication system. Science academy URSS, (1963) – in Russian

Fuzzy Directional Adaptive Recursive Temporal Filter for Noise Suppression of Video Color Sequences

Alberto Rosales-Silva[1], Volodymyr Ponomaryov[1], Francisco Gallegos-Funes[2]
Oleksiy Pogrebnyak[3]

National Polytechnic Institute of Mexico
Mechanical and Electrical Engineering Higher School[1,2]
Center for Computing Research[3]
[1] ESIME-Culhuacan; Av. Santa Ana 1000, Col. San Francisco Culhuacan,
04430, Mexico D.F., Mexico, vponomar@ipn.mx
[2] ESIME-Zacatenco; Av. IPN s/n, U.P.A.L.M. Col. Lindavista,
07738, Mexico D.F., Mexico, fgallegosf@ipn.mx
[3] CIC; Av. Juan de Dios Batiz S/N, esq. Miguel Othon de Mendizábal,
07738, Mexico, D.F., Mexico, olek@pollux.cic.ipn.mx

Abstract. We present the fuzzy directional adaptive recursive temporal filter for noise suppression of video color sequences. Considering spatial-temporal information allows improving the efficiency of filtering in case of fast motion and noise presence We use the differences between images, such as, the angle deviations to obtain several parameters to apply them in the proposed algorithm to detect and differentiate movement in background of noise. Extensive simulation results have demonstrated that the proposed filter can consistently outperforms other filters by balancing the tradeoff between noise suppression and detail preservation.

1 Introduction

The problem of motion detection is very complex due it is not always easy to distinguish illumination changes from real motion and because of the problem of aperture. In several applications it is sufficient to detect changes in the scene rather than actual motion, and even to detect only some of the changes [1].

In this paper, a method is proposed to distinguish noise and changes of motion camera and movement present in video color sequences. Different operations depending of parameters obtained by fuzzy logic membership functions permit to consume less time and to realize robust noise suppression and movement detection. The goal of proposed method is to use adaptive threshold from local pixel statistics and spatial pixel context, and it is adapted with temporal and spatial information [1-3].

The proposed method is robust to noise. In the case of motion detection during the noise suppression process, where the detection is used for temporal filtering, undetected changes in an object can lead to motion blur, but in the same time if some noise is labelled as motion it is no so critical. Using fuzzy logic techniques we define a confidence measure with respect to the existence of motion [2].

© L. Sánchez, O. Pogrebnyak and E. Rubio (Eds.)
Industrial Informatics
Research in Computing Science 31, 2007, pp. 43-52

2 Proposed Method

The proposed Fuzzy Directional Adaptive Recursive Temporal (FDARTF_G) filter for simultaneous noise suppression and motion detection in video color sequences is depicted in Figure 1.

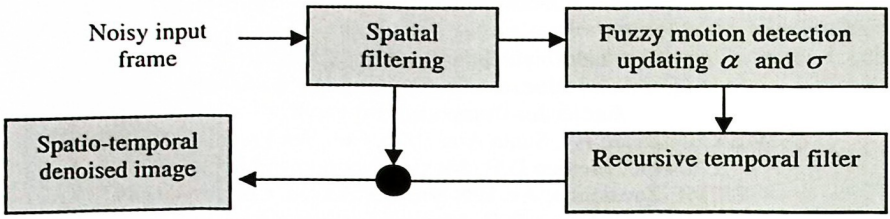


Fig. 1. Block diagram of proposed algorithm.

2.1 Spatial Filtering Stage

To suppress noise in the first stage we use the Gaussian estimation algorithm:
Step 1) IF $\theta_c \le F/255$ THEN Histogram is increased in "1", else is "0".
Step 2) Compute probabilities for each one of these samples:

$$p_i = Histogram_i / S \text{ , } i=0,\ldots,255. \tag{1}$$

Step 3) Compute standard deviation σ'_T :

$$\mu = \sum_{i=0}^{255} i \cdot p_i ; \sigma^2 = \sum_{i=0}^{255} (i-\mu)^2 \cdot (p_i) \text{ , } \sigma'_T = \sqrt{\sigma_T^2} \tag{2}$$

where $S = \sum_i Histogram_i$, $\theta_c = A(\bar{y}, y_c)$ is the angle deviation, $\bar{y} = 1/N \sum_{i=1}^{N} y_i$ is the mean value with i=1,…,N, N=9, and y_c is the central pixel.

2.2 Fuzzy Motion Detection Stage

Each plane of video sequence is processed and the parameters $\sigma'_T = \sigma'_{red} = \sigma'_{green} = \sigma'_{blue}$ are adapted along of video sequence. The angle deviations $\theta_i = A(x_i, x_c)$ are calculated [4], where i=1,…,N-1, N=9, and $i \neq central\ pixel$.

The uniform regions are detected using the mean weighted filtering algorithm [5]:
Step 1) IF θ_2 *AND* θ_4 *AND* θ_7 *AND* $\theta_5 \ge \tau_1$ THEN,

$$y_{out} = \sum_{\substack{i=1 \\ i \neq c}}^{N-1} x_i \left[\frac{2}{\left(1+e^{\theta_i}\right)^r} \right] + x_c \Bigg/ \sum_{i=1}^{N-1} x_i \left[\frac{2}{\left(1+e^{\theta_i}\right)^r} \right] + 1 \tag{3}$$

Step 2) IF θ_6 *AND* θ_3 *AND* θ_1 *AND* $\theta_8 \geq \tau_1$ THEN, Use eq. (3).

The central pixel has the highest weight with "1" to preserve some features in uniform regions. A 5x5 window is used to estimate the standard deviation in same way as Gaussian estimation algorithm to obtain the values σ_T, and then we compare with the values σ'_T to have a similarity value for each sample and to have a criterion in perform more or less filtering charge. We obtained that if $\sigma'_T < \sigma_T$, then $\sigma_T = \sigma'_T$, otherwise $\sigma'_T = \sigma_T$, where T can be the component red, green, or blue, that permits to improve temporal filtering algorithm. By optimum PSNR and MAE we defined a threshold $Th_T = 2\sigma_T$ to preserve some features in a spatially filtered frame that will be used in temporal algorithm.

For each pixel *(i,j)* of the any component image, we use a 3x3 neighbourhood window as illustrates Figure 2. If A_T denotes one component input image, the gradient can be defined as,

$$\nabla_{(k,l)} A_T(i,j) = \left| A_T(i+k, j+l) - A_T(i,j) \right| \text{ with } k,l \in \{-1,0,1\} \tag{4}$$

where the pair *(k,l)* corresponds to one of eight directions that are called *the basic gradient values* [2], and *(i,j)* is called the centre of the gradient.

To avoid blurred in presence of an edge, we use one basic gradient for each direction and two related gradient values. Three gradient values for a certain direction are finally connected together into one single value called fuzzy gradient value. We take pixels as vectors to have directional process, taking the same procedure as in gradient values. By this way we obtain Fuzzy vector gradient values that are defined by fuzzy rule 1. Two related gradient values in the same direction and the basic gradients are determined by the centres making a right-angle with the direction of the corresponding basic gradient [3].

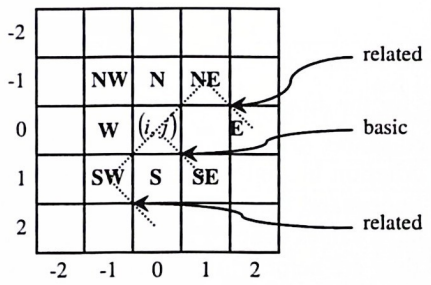


Fig. 2. Basic and related gradients and vector gradients

The basic vector gradient values are determined by the following algorithm,

Step 1) IF $\nabla_{\gamma\beta} < T_{s\beta}$ THEN compute angle deviation in the direction $\alpha_{\gamma\beta}$ and obtain weight value,

$$\alpha'_{\gamma\beta} = 2 \Big/ \left(1 + e^{\alpha_{\gamma\beta}}\right)^{r} \tag{5}$$

Step 2) Compute basic vector gradient using membership function.

Step 3) IF $\nabla_{\gamma\beta} > T_{s\beta}$ THEN $\mu_{BIG} = 0$.

where $T_{s\beta} = Th_T$, γ=*NW,N,NE,E,SE,S,SW,W* [6], $\beta = red, gree, blue$, $r = 1$ channels in video sequence, the membership function is $\mu_{BIG} = \max(x, y)$ with $x = \alpha'_{\gamma\beta}$ and $y = \left(1 - \nabla_{\gamma\beta} / T_{s\beta}\right)$.

To obtain the angle deviation in each plane of the frame we select to work in the angle formed by vectors in only one coordinate [4]:

$$\alpha = \cos^{-1}\left(\left(r_1 r_2 + g_1 g_2 + b_1 b_2\right) \Big/ \sqrt{\left(r_1^2 + g_1^2 + b_1^2\right)\left(r_2^2 + g_2^2 + b_2^2\right)}\right) \tag{6}$$

where, (r_1, g_1, b_1) and (r_2, g_2, b_2) are coordinates of two pixels.

To determine related vector gradients the procedure is the following:

Step 1) IF $\nabla_{\gamma\beta(R1,R2)} < T_{s\beta}$ THEN compute angle deviation in direction $\alpha_{\gamma\beta(R1,R2)}$ and obtain weight value [5],

$$\alpha'_{\gamma\beta(R1,R2)} = 2 \Big/ \left(1 + e^{\alpha_{\gamma\beta(R1,R2)}}\right) \tag{7}$$

Step 2) Obtain related vector gradient using membership function.

Step 3) IF $\nabla_{\gamma\beta(R1,R2)} > T_{s\beta}$ THEN $\mu_{BIG} = 0$.

where $(R1, R2)$ are the related vector gradients and the membership function is $\mu_{BIG} = \max(x, y)$ with $x = \alpha'_{\gamma\beta(R1,R2)}$ and $y = \left(1 - \nabla_{\gamma\beta(R1,R2)} / T_{s\beta}\right)$.

The fuzzy rule 1 is defined as,

Fuzzy Rule 1: defining the fuzzy vector gradient value $\nabla^F_{\gamma\beta} A_\beta(i, j)$,
IF $\nabla_{\gamma\beta}$ is BIG AND $\nabla_{\gamma\beta R1}$ is BIG, OR $\nabla_{\gamma\beta}$ is BIG AND $\nabla_{\gamma\beta R2}$ is BIG, THEN $\nabla^F_{\gamma\beta} A_\beta(i, j)$ is BIG,

where $\nabla_{\gamma\beta}$ is the basic vector gradient value, and $\nabla_{\gamma\beta R1}$ and $\nabla_{\gamma\beta R2}$ are two related vector gradient values for the direction γ in the channel β.

If basic and related vector gradients are close enough, in absolute difference or norm, or in a vector criterion in angle distances, the proposal is developed to obtain robust parameters to understand the nature of pixels in a window processing. Under this criterion we will have values denoted as fuzzy vector gradients that means nearby in pixels related, and they are helpful to suppress Gaussian noise. The noise suppression is done by a weighted mean procedure where nearby close to 1 have bigger weights in the algorithm due to proposed procedure used in membership function. It suppresses noise more efficiently but smoothes details and edges, in our complete algorithm the temporal filtering are designed. The reference values where found modifying their parameters according to optimum PSNR and MAE values.

The weighted mean algorithm is implemented by:

$$y_{out} = \sum_{\substack{i=0 \\ i \neq c}}^{N-1} y_\gamma \cdot x_{\gamma i} \Big/ \sum_{i=0}^{N-1} x_{\gamma i} \tag{8}$$

2.3 Recursive Temporal Filtering Stage

The reference values of spatial filter presented in sec. 2.2 are used in the final stage in the proposed filter. The past and present frames are used to avoid dramatic charge in memory requirements and processing time. The fuzzy logic rules are used in each plane of two frames. We found angle deviations and gradient values by the central pixel in present frame respect to his neighbours in past frame in each plane,

$$\theta_i^1 = A\left(x_i^A, x_c^B\right) \; ; \; \nabla_i^1 = \left|x_i^A - x_c^B\right|; i = 1,3 \; , N; \; N = 9 \tag{9}$$

where x_c is central pixel in present frame, and A and B are past and present frames by planes, respectively.

The membership functions are defined to obtain a value that indicates the degree, in which a certain gradient or vector value matches the predicate. If a gradient or vector value has membership degree one for the fuzzy set SMALL, it means that it is SMALL for sure in this fuzzy set. Selection of membership functions is given from nature of pixels, where a movement is not a linear response, and a pixel has different meanings in each frame of video sequence.

Membership functions SMALL and BIG for angles and gradients are given by [7]:

$$\mu_{SMALL}(M) = \begin{cases} 1 & M < med \\ \exp\left(-\dfrac{(M - med)^2}{2\sigma_S^2}\right) & \text{otherwise} \end{cases} \tag{10}$$

$$\mu_{BIG}(M) = \begin{cases} 1 & M < med \\ \exp\left(-\dfrac{(M - med)^2}{2\sigma_B^2}\right) & \text{otherwise} \end{cases} \tag{11}$$

where $\sigma_S^2 = 0.1$, $\sigma_B^2 = 1000$, and M can be the angle θ or gradient ∇, for angles *med*=0.2 and *med*=0.9 for membership functions SMALL and BIG, respectively, and for gradients *med*=60 and *med*=140 for SMALL and BIG functions, respectively.

We use the fuzzy rules 2, 3, 4, and 5 to acquire the corresponding values:

Fuzzy Rules
Fuzzy Rule 2: Defining the fuzzy gradient-vector value $SBB(x,y,t)$. IF $\theta^1(x,y,t)$ is SMALL AND $\theta^2(x,y,t)$ is BIG AND $\theta^3(x,y,t)$ is BIG AND $\nabla^1(x,y,t)$ is SMALL AND $\nabla^2(x,y,t)$ is BIG AND $\nabla^3(x,y,t)$ is BIG THEN $SBB(x,y,t)$ is true.
Fuzzy Rule 3: Defining the fuzzy gradient-vector value $SSS(x,y,t)$. IF $\theta^1(x,y,t)$ is SMALL AND $\theta^2(x,y,t)$ is SMALL AND $\theta^3(x,y,t)$ is SMALL AND $\nabla^1(x,y,t)$ is SMALL AND $\nabla^2(x,y,t)$ is SMALL AND $\nabla^3(x,y,t)$ is SMALL THEN $SSS(x,y,t)$ is true.
Fuzzy Rule 4: Defining the fuzzy gradient-vector value $BBB(x,y,t)$. IF $\theta^1(x,y,t)$ is BIG AND $\theta^2(x,y,t)$ is BIG AND $\theta^3(x,y,t)$ is BIG AND $\nabla^1(x,y,t)$ is BIG AND $\nabla^2(x,y,t)$ is BIG AND $\nabla^3(x,y,t)$ is BIG THEN $BBB(x,y,t)$ is true.

Fuzzy Rule 5: Defining the fuzzy gradient-vector value $BBS(x,y,t)$.
IF $\theta^1(x,y,t)$ is BIG AND $\theta^2(x,y,t)$ is BIG AND $\theta^3(x,y,t)$ is SMALL AND $\nabla^1(x,y,t)$ is BIG AND $\nabla^2(x,y,t)$ is BIG AND $\nabla^3(x,y,t)$ is SMALL THEN $BBS(x,y,t)$ is true.

where $\theta^r(x,y,t)$ are angles values, $\nabla^r(x,y,t)$ are gradient values, and $r = 1,2,3$.

and the following rules:

Fuzzy Rule *SBB(x,y,t)*
If *SBB(x,y,t)* is the biggest value found from the others:
Step 1) IF {(*SBB(x,y,t)* > *SSS(x,y,t)*) AND (*SBB(x,y,t)* > *BBB(x,y,t)*) AND (*SBB(x,y,t)* > *BBS(x,y,t)*)} THEN Weighted mean using *SBB(x,y,t)*,

$$y_{out} = \frac{\sum p^A(x,y,t)\cdot SBB(x,y,t)}{\sum SBB(x,y,t)}$$

Step 2) Update standard deviation for next frames to divide details from uniform regions,

$$\sigma_T{}' = (\alpha\cdot\sigma_{TOTAL}) + (1-\alpha)\cdot(\sigma_T{}')$$

where $T = red, green, blue$, $\sigma_{TOTAL} = (\sigma_{red} + \sigma_{green} + \sigma_{blue})/3$, and $\alpha = \alpha_{SBB} = 0.875$.

Fuzzy Rule *SSS(x,y,t)*
If *SSS(x,y,t)* is the biggest value found from the others:
Step 1) IF {(*SSS(x,y,t)* > *SBB(x,y,t)*) AND (*SSS(x,y,t)* > *BBB(x,y,t)*) AND (*SSS(x,y,t)* > *BBS(x,y,t)*)} THEN Weighted mean using *SSS(x,y,t)*,

$$y_{out} = \frac{\sum (p^A(x,y,t)\cdot 0.5 + p^B(x,y,t)\cdot 0.5)\cdot SSS(x,y,t)}{\sum SSS(x,y,t)}$$

Step 2) Update standard deviation for next frames to divide details from uniform regions.

Fuzzy Rule *BBB(x,y,t)*
If *BBB(x,y,t)* is the biggest value found from the others:
Step 1) IF {(*BBB(x,y,t)* > *SBB(x,y,t)*) AND (*BBB(x,y,t)* > *SSS(x,y,t)*) AND (*BBB(x,y,t)* > *BBS(x,y,t)*)} THEN motion-noise = true.

$$y_{out} = \frac{\sum p^A(x,y,t)\cdot SBB(x,y,t)}{\sum SBB(x,y,t)}$$

Step 2) If $\sqrt{motion-noise\ confidence} = 1$. then $\alpha = 0.875$,
else if $\sqrt{motion-noise\ confidence} = 0$ then $\alpha = 0.125$, else $\alpha = 0.5$.
Step 3) $y_{out} = (1-\alpha)\cdot(pres\_fr_{central\_pixel}) + \alpha\cdot(past\_fr_{central\_pixel})$

Fuzzy Rule *BBS(x,y,t)*
If *BBS(x,y,t)* is the biggest value found from the others:
Step 1) IF {(*BBS(x,y,t)* > *SBB(x,y,t)*) AND (*BBS(x,y,t)* > *SSS(x,y,t)*) AND (*BBS(x,y,t)* > *BBB(x,y,t)*)} THEN Weighted mean using *BBS(x,y,t)*,

$$y_{out} = \frac{\sum p^B(x,y,t)\cdot(1-BBS(x,y,t))}{\sum(1-BBS(x,y,t))}$$

Step 2) Update standard deviation for next frames to divide details from uniform regions.

where *SBB(x,y,t)* value says that central pixel is in movement because of big differences in local and gradient values, *SSS(x,y,t)* shows that a central pixel is not in movement because of small differences in all directions, *BBB(x,y,t)* value shows that a central pixel and its neighbours have not relation among the others and it is highly probably that this pixel is in motion or is a noisy pixel, $p^A(x,y,t)$ and $p^B(x,y,t)$ are

the pixels in last and present frames that fulfil with the IF condition, y_{out} is the spatial and temporal filtered output, and $\alpha = \alpha_{SSS} = 0.1255$

3 Experimental Results

From the simulation experiments we obtained the properties of proposed Fuzzy Directional Adaptive Recursive Temporal (FDARTF_G) filter in terms of PSNR and MAE criteria. We compared our approach with the FMRSTF (Fuzzy Motion Recursive Spatial-Temporal) and FVMRSTF (Fuzzy Vectorial Motion Recursive Spatial-Temporal) filters, these filters are designed to work with gradients and angle deviations, respectively [1-3].

We use the 176x144 QCIF video sequences "Flowers" and "Miss America" to qualify effectiveness of the proposed filter. The frames of video sequences are treated in an RGB color space with 24 bits and 8 bits for each channel.

Figure 3 presents the performance results in terms of PSNR for the frame #100 of video sequence "Miss America" corrupted with Gaussian noise from 0.00 to 0.05 in variance with zero mean by use different filters. In Figure 4 we show the MAE performance for the same frame of video sequence "Miss America". From the Figures 3 and 4, one can see that the best results in PSNR and MAE criteria are given by the proposed FDARTF_G filter.

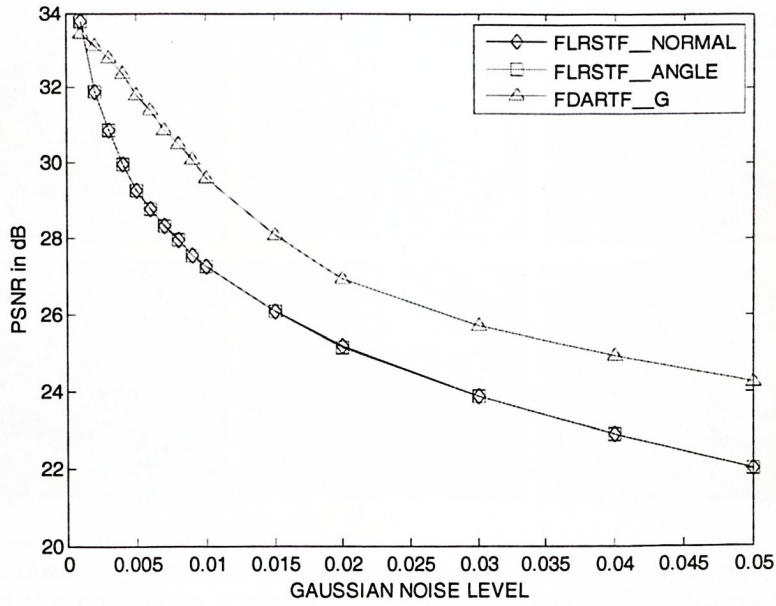


Fig. 3. PSNR for the frame #100 of video sequence "Miss America" by use different filters.

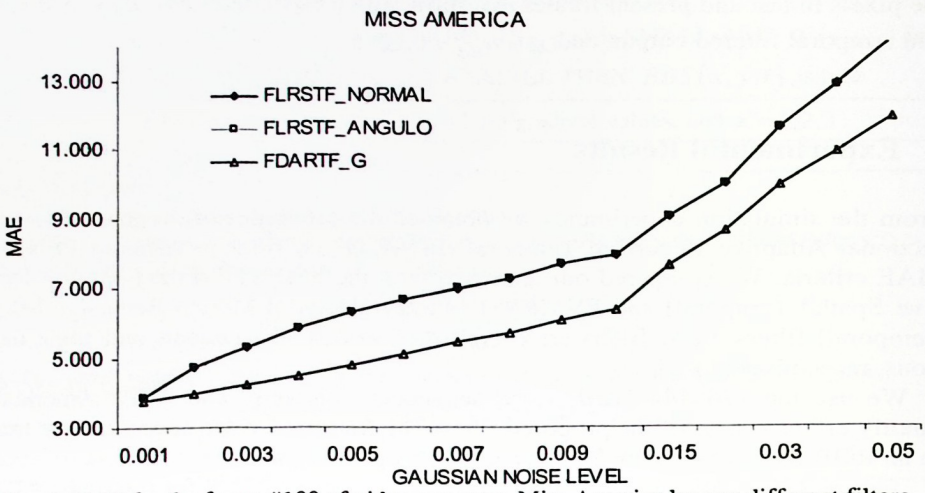
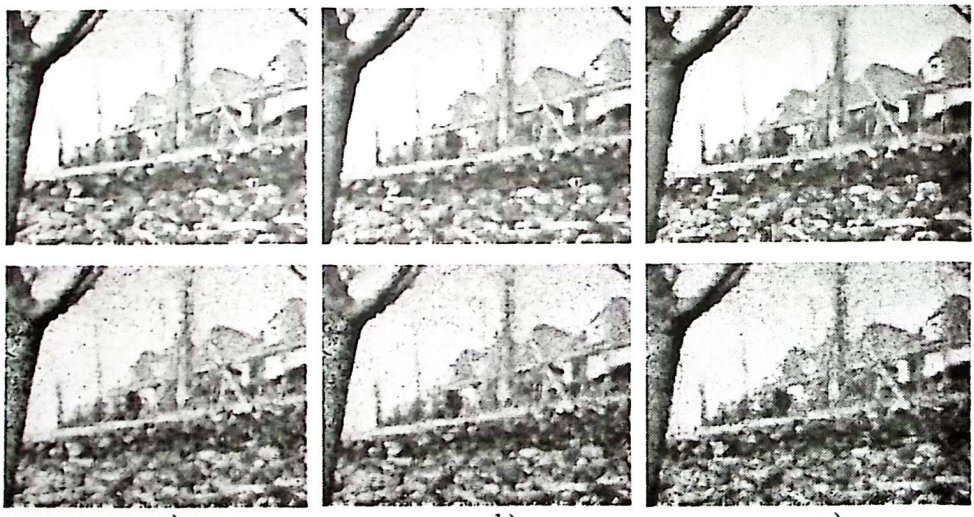


Fig. 4. MAE for the frame #100 of video sequence Miss America by use different filters.

Figure 5 depicts the visual results in a frame of video sequence "Flowers" degraded with 0.01 and 0.03 of Gaussian noise by means of use of the FLRSTF_NORMAL, FLRSTF_ANGLE, and the proposed FDARTF_G filter. From this Figure, one can see that the restored frames by means of use the proposed filters appear to have a better subjective quality.



a) b) c)

Fig. 5. Visual results in a frame of video sequence "Flowers" degraded with variance of 0.01 and 0.03 from top to bottom, a) Column of restored images by FLRSTF_NORMAL, b) Column of restored images by FLRSTF_ANGLE, and c) Column of restored images by proposed FDARTF_G.

Figure 6 presents the visual result in the color image "Mandrill" in the case of 30% of impulsive noise degradation by means of use the proposed FDARTF_G filter. From this Figure, we observe that the restored image appears to have a good subjective quality.

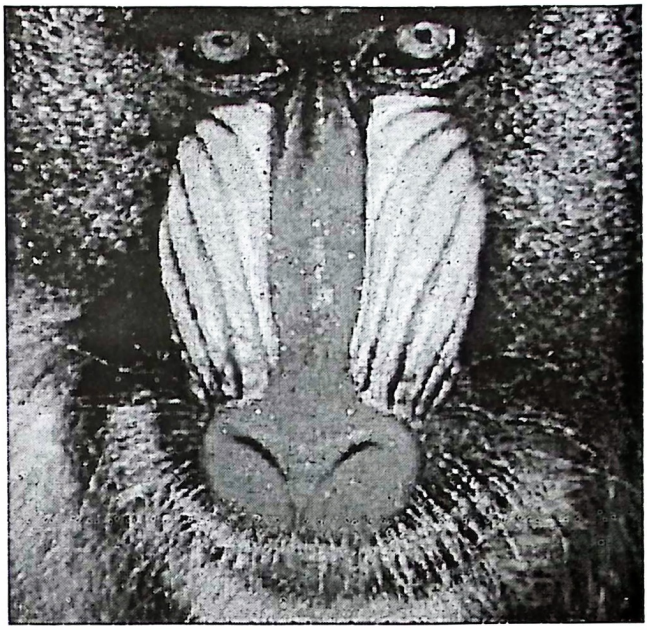


Fig. 6. Visual results in the color image "Mandrill" restored from 30% of impulsive noise by proposed FDARTF_G filter.

4 Conclusions

We present an adaptive recursive scheme for fuzzy logic based motion detection. The proposed algorithm realizes the spatial and temporal filtering to improve the noise suppression and detail preservation. We demonstrate that taking into account the gradients and vectors, and connecting them together, we can realize a robust algorithm improving the techniques that use such features in a separate form.

5 Acknowledgments

This work is supported by National Polytechnic Institute of Mexico and CONACyT.

References

1. Zlokolica V., De Geyter M., Schulte S., Pizurica A., Philips W., Kerre E.: Fuzzy Logic Recursive Motion Detection for Tracking and Denoising of Video Sequences. IS&T/SPIE Symposium on Electronic Imaging, San Jose, California, USA, January (2005)
2. Zlokolica. V.: Advanced Non-Linear Methods for Video Denoising. PhD Thesis, Gent University, (2006)
3. Zlokolica V., Schulte S., Pizurica A., Philips W., Kerre E.: Fuzzy Logic Recursive Motion Detection and Denoising of Video Sequences. Journal of Electronic Imaging, Vol.15, No.2, (2006)
4. Trahanias P. E., Venetsanopoulos A. N.: Vector Directional Filters-A New Class of Multichannel Image Processing Filters. IEEE Transactions on Image Processing, Vol.2, No.4, (1993)
5. Plataniotis K. N., Venetsanopoulos A. N.: Color Image Processing and Applications. Springer Verlag, Berlin, (2000)
6. Schulte S., De Witte V., Nachtegael M., Van der Weken D., Kerre E.: Fuzzy Two-Step Filter for Impulse Noise Reduction from Color Images. IEEE Transactions on Image Processing, Vol.15, No.11, (2006)
7. Fuzzy Logic Fundamentals, Chapter 3, pp. 61-103, 2001, www.informit.com/content/images/0135705991/samplechapter/0135705991.pdf

Adaptability of the Information Systems to the changes in the Business Processes

Sandra Dinora Orantes Jiménez[1], Máximo López Sánchez[2], Agustín F Gutiérrez Tornés[3]

[1]Center for Computing Research-National Polytechnic Institute, Av. Juan de Dios Bátiz s/n, Mexico, D.F., CP 07738, Mexico. Telephone number: (55) 57296000 Ext. 56523, Fax 56607
dinora@cic.ipn.mx
[2]National Center of Research and Technological Development, Interior Internado Palmira s/n, Colony Palmira. Cuernavaca, Morelos, Mexico, C.P. 62490. Telephone number: (777) 318 - 7741, 01 (777) 312-2314 Ext: 145
maximo@cenidet.edu.mx
[3]Coordinator of Systems, BANAMEX; Professor of Subject, ITESM-CCM, Mexico, DF, Mexico. Telephone number: (55) 22629184
afgutitor@banamex.com

Abstract. A company doesn't exist if it doesn't have clients and definite well business processes. A company that doesn't have clients doesn't have business processes. Actually is most important the automatization of this processes trough information systems that generates and maintain the control of the activities. This process is obtaining a product that generates value for the company and for the clients. Achieve that the organizations have a high yield; it depends on the strategic decisions and the efficiency in the business processes and of the rapidity with which it is obtained. The present investigation, arises with the interest to analyze the problems that make consider to the "Adaptability" like an attribute of Quality in Information systems, with the objective to turn it a characteristic, that includes all the stages of the development of software oriented to the businesses, that is to say, observing not only the technical problems, but the problematic one, from the point of view of as software adapts to changes in the business.

Key words: Business process, Software Quality, Information Systems, business engineering, management.

1. Introduction

" A business process is a concrete way of arranging a set of activities, not annotated by organizational barriers, with a beginning and an end, coordinated and orientated to the attainment of a product that generates value for a final or intermediate client ". [1] The fundamental element in the previous definition is the search of the satisfaction of the final or intermediate client. If there is no client then there is not any business

© L. Sánchez, O. Pogrebnyak and E. Rubio (Eds.)
Industrial Informatics
Research in Computing Science 31, 2007, pp. 53-62

process. Both elements are so close that he is the client the one that gives content to concept the business process. In fact the company is defined as the sum of business process and the client sees her when it uses someone of them.

Multiple business studies have shown evidence that organization do often is inconsistent with the environment conditions and generally not the right thing at the right time. Most of the time they follow a given strategy direction. Changes may seem continuous, but occurs in the context of that orientation. Organizations usually "do more of the same", some times improved. Most organizations like periods of stability because they achieve success not by changing strategies, but by practicing the ones they know and have successful proved before. This can be happen because the same types of conditions are cyclic and tend to occur again. In a rapidly changing world, across every industry, businesses environment is becoming more complex, more competitive, fast-paced and unpredictable. To compete and succeed, businesses have to become an adaptive and agile enterprise that can accommodate change and adapt to the change quickly. It is there where this research arises in the search of concepts, definitions, formalizations that lead to turning the attribute Adaptability, in one Characteristic to demand in information systems orientated to the business, across all your life's cycle, guaranteeing Adaptable Information Systems to the changes in the business processes.

For understand because it becomes necessary to propose the creation of a model that allows the evaluation of the Adaptability as characteristic of the Systems of Information the changes of the business processes, the present article is divided of the following form: First there are mentioned brief big paradigms that they precede to the emergence of the need of development of Adaptable Information Systems. Then, it talk the new paradigm Adaptability in Information Systems proposing a model based in patterns [7] that it gives place to the offer of a group of subcharacteristics with the objectives to turn the Adaptability into a characteristic that the enclose. Finally, an example, conclusions and work future relating to the investigation are provided.

2. Problem with the Traditional Information Systems

Generally, the development of the Information Systems takes a lot of time and money and even, on having used them they become fragile, particularly, in situations for which explicitly it wasn't designed. One can speak about paradigms in the development of the information systems:

1. Structured programming. Which arose in the 70s and which was trying to make the system construction feasible on a large scale, having to rely on a specification of the results wished from the beginning of the project and bearing in mind that the specification rare times changed.
2. Programming OO. It was the following great paradigm that appears in the 80s, trying to do "easily" to be reorganized when the specification of the system was changing, dividing from that the functionality is separated in classes that are designed to have a minimal interaction between them. Nevertheless, every change of specification or of the environment of work, there needs the intervention of analysts, designers and programmers, which, carries to expenses in re-designing,

re-helping, reconstructing and passing again for the whole life's cycle of software development.

Today, one presents another type of problem that with the methodologies constructed and orientated to objects, is not so easy to face, to programmer adaptable information systems to the change.

3. Adaptability in Information Systems

For a company being adaptable is to be capable of operating productivity in a competitive environment of continually and unpredictable changes [8].

For an individual being adaptable is to be capable of contributing to the bottom line of a company that is constantly reorganizing its human and technological resources in response to unpredictable market change. With the Internet emergence (Information Systems Web, eBusiness, eCommerce, Services Web) or companies that handle your information across Intranet, it is allowed that there should appear applications that typically they need to adapt to the rapid and constant changes of the way.

Certainly, it is here where there are needed other methodologies that improve the software, which is orientated to the problem of handling the change in the business processes, adaptability in information systems that enclosed are anticipated to the change and almost automatically identify in a program traversing that is what one will affect, bearing in mind what says the definition of adaptable software: "Adaptive software uses available information about changes in its environment to improve its behavior."[2]

Adaptability enables information systems to follow changes that occur along the lifetime of a software system. In this spirit, adaptability is a visionary concept. A vision that assumes that business support is given almost autonomously during build and runtime of the software system. Terminology on adaptability is not consistent. The most essential terms comprise for instance flexibility, adaptation and agility. From a research viewpoint adaptable systems represent the ideal final condition reflecting unlimited problem space. Therefore it is necessary to extract the qualities that are responsible for adaptable behaviour and to transfer these qualities on information systems.

4. Proposal use Patterns to structure model for Adaptability in Information Systems

Patterns [7] provide a way to capture and document expert knowledge to be used by non-experts. The pattern approach provides the foundation for various pattern styles such as design and analysis pattern for instance. The assumption is that changes in the organization can see like patterns, is possible to observe a number of patterns that exist specific to adaptability requirements in enterprise systems. The detected patterns are based on iterative analyses of business process requirements and technical features with regard to adaptability and its attributes such as efficiency and flexibility for instance. The Adaptability patterns are found in the characteristic of quality that is

necessary an information system for that considered like adaptable. First is necessary identifies the structures that exist in the information systems themselves the second recognizes patterns in the process of intervention which bring about changes in the enterprise system (as a consequence of business process changes). The classification is presented like subcharacteristics in the sections 4.1.1, 4.1.2 and 4.1.3. It necessary to heel that the subcharacteristics that moment it identified has the objective in the Information System the managing Structural Analogy, Knowledge, Redundancy and Customizing of business. The use patterns or the business dimension characterize the circumstances of the usage of an information system. They reflect that adaptability-enabling factors are also related to decisions referring to the deployment of the system. In this area patterns are for instance the capabilities of personnel (person-bound) knowledge; existing guidelines to proper deploy a software system.

4.1 Propose of Subcharacteristics

It necessary explain that the model doesn't formally established, that is based into all the stages of a software development cycle and taking as a base the structure proposed in the Mexican standard NMX-I-O55-NYCE [4], which in turn bases on the standard international ISO/IEC 9126 [5], they appear and they propose some Subcharacteristics that will have to be contained inside the characteristic Adaptability and for which will have to establish metric concrete. System patterns describe the immanent qualities of the information system itself. The system is considered a closed system to measure its capacities to manage change. The subcharacteristics and metric offers must include three big stages of the evolution of development of software: correct Analysis of the problem, Design of the problem of correct form and that the Implementation uses as base for another development. Can see that subcharacteristics are identified like patterns. (See Fig. 1)

System-based (to guarantee a correct Analysis of the problem and that an implementation uses as base for another development)		Business Use
Validation	Design (to assure a Design of the problem of correct form)	Structural Analogy
Utility	Monitoring	Knowledge
Usability	Reliability	Redundancy
Adaptability Feasibility	Correctness	Customizing
Extensibility	Faults rate	
Flexibility	self-similarity	
Tuning	Availability	
Fixability of the changes		
Reusability		
Scalability		
Modularity		
Mobility		
Interoperability		
Self-organization		

Fig. 1. Identified Patterns of Adaptability

4.1.1 Subcharacteristic to guarantee a correct Analysis of the problem

Validation. The metric ones elaborated for this subcharacteristic will have to guarantee that the requirements were understood of precise form and besides, bear in mind the time that went to understand the requirements.

Metrics some of them who consider this objective can be: *How much effort took the comprehension and precise understanding of the requirements? What so much will they change the requirements from the moment that they were raised by the client until the system is dedicated? There has been born in mind that the requirements of the business are a mobile objective? Have measurements been taken in order that the requirements are satisfied at all time? Has it achieved that the system is adaptable to the desires of the client?*

The important of the subcharacteristic validation is to assure that the information system is adaptable to any reasonable change and that it will be able to be constructed in the time that the business like that it needs.

Also is necessary considered the subcharacteristics that repeat like patterns: *Scalability, Modularity, Mobility, Interoperability* and *Self-organization.*

4.1.2 Subcharacteristics to assure a Design of the problem of correct form

Monitoring. That the requirements have been translated from faithful form into the requirements of the solution.

Reliability. That doesn't exist distrust any respect that the resultant design will give an application of agreement to the requirements of the business. There must be born in mind that the companies not always are ready to pay money extra to guarantee reliability.

Correctness. That the design allows any adequacy changes, without of an effect wave. There must be born in mind all the requirements of the business that with the time can change.

Faults Rate. The faults and mistakes must be corrected from the first stages, since no company expects to invert money extra (though the tariffs are low) to correct faults. If some it is not possible to avoid them, there must exist a plan of contingency that allows remedying immediately the problems that arise. When changes arise in the business processes, there must be guaranteed that they must not introduce new faults. The business have changed your attitude with regard to the Information Systems that nowadays not only are instruments of finance, but rather functions control essentials of the business, which if it is definitive is that the clients demand simple interfaces and reliable results. The standards at present, they are those who dictate the rules of the game relating to the quality of the information systems, which they allow to make competitive to the business and to achieve that the clients of the same one go away with the business that thanks to the support of the information systems seem to them to be more reliable.

Self-similarity. Also has an important consequence for the design of computer networks, as typical network traffic has self-similar properties for drive business process.

Availability. The availability of a repairable system as "the probability that the system is operating at a specified time t".

4.1.3 Subcharacteristics to guarantee that an implementation uses as base for another development

There exist ways of guaranteeing that the developments of information system at present, they could provide a good base for the construction of the systems of tomorrow, for it one of the important characteristics in this stage is the Reusability.

Reusability. The construction of reusable parts likes a great game of LEGO where an assembly can be disarmed in pieces to build something different. For the case of the software, there is born in mind the reusability of analysis, designs and certainly the software itself (class libraries) that include from processes up to interfaces graphs and frames of work for data bases. Nevertheless, it is important bear in mind that the Reusability stretches to be less successful in the highest levels of applications of the business; motives exist for this, but the most common is the general tension in the company between the Reutilization and the Utility.

Utility. It hopes that between more Reutilization it is, this does not influence that the resultant product is "less easy" to use.

Facility of Use (Usability). On the other hand, doing that the system is " easy to use " there diminishes your aptitude to be re-used in a different authority or objectives (tools with specific intention are in general " easy to use " compared with multipurpose tools). Because of this and other factors, the components of high level in a product, generally they are not reusable to construct the different one and the manufacture of reusable components of high level across authorities of the problem, typically they increase the cost of your utilization.

Adaptability Feasibility. This subcharacteristic that will have to be born in mind in this phase, already is own of the model who is in structure, besides it is the one that guarantees the feasibility of adapting the existing system to the needs of the tomorrow of the business. This approach is lightly different from that of reusable parts, which are those that in group they constitute to the system and which depend on the experience of the programmers to construct the information systems of the tomorrow with the parts of the current one; with the approach of obtaining of a Adaptable Information System it hopes that the system could be personalized by the final users or for a person who did not have to see with the development. Four factors exist in the Information Systems, which need to be adaptable and which bear the consideration of 4 more subcharacteristics in this stage; 2 imply changes to very high level (subcharacteristics Extensibility and Flexibility) and two changes to low level (subcharacteristics Tuning and Fixability of the changes).

Extensibility. This subcharacteristic capture in counts the facility with which the system is capable of growing with the needed changes. For example, it is said that a Information System is extensible if it is relatively "easy" to add the managing of graphical devices or other formats or any another managing of statistical graphs that the current system does not possess.

Flexibility. It bears in mind that so easy is to change the capacities of the system into styles, categories, certain processes, etc. Without one affects your integrity.

Tuning. This subcharacteristic also can be thought as an activity of change. To tune to the development with the possible changes in the business processes. A system that it is considered to be "easily tuning" can be changed of forms that affect your performance for good. An example of tuning agreement to [6] would be: For example, CORBA objects are location-independent, which allows objects to be physically

moved around on the network (e.g., to reduce network traffic) without impacting very much, if any, code. Java is even more extreme in this aspect, since the objects can be moved to any machine on the network on the fly. Naturally none of this negates the need to do up-front performance engineering work.

Fixability of the changes. This subcharacteristic is defined as the skill of fixing a change without affecting the rest of the system; this can be difficult in big systems, since it is needed bear in mind many aspects, to separate the interfaces of the implementation and to approach the specification of the behavior of every component. The first attempt of this subcharacteristic can think in the Design Patterns [7] that they provide the solution of a definite well, adaptable problem to new needs.

4.1.4 Evaluation of the Adaptability of Information Systems

Table 1 shows propose of the evaluation subcharacteristics.

Subcharacteristic: "x"			
Atribute	Weight	Effect	Points
.....			
Subcharacteristic index		⇒	

Table 1. Determination of the Subcharacteristic Index for evaluation characteristic Adaptability in Information Systems

The business dimension characterizes the circumstances of usage for an information system. It is outlined that the degree of adaptability, which can be reached by a specific information system, is influenced by decisions referring to the deployment of the system within a business organization. The calculation of the index is not detailed in this paper.

5. Business Strategy Generation: Intuitive Example

Let us consider an insurance broker agency formed by ten employees. As a broker, the agency sells policies for different companies. The main products are life and automobile policies. For selling and advertising the insurance company obtains detailed information from potential customers (C), and from private and governmental agencies (A). This information is distributed between the company's agents (AG) that contact potential clients via phone and try to set up a conference call; however, they also have their own sources of information. At the interview, the agent examines the client's current insurance coverage and tries to find an opportunity for a policy that will best fit the customer's needs. Before obtaining an insurance policy, the new client suffers an identity investigation. In the case of a life insurance, the client has, in addition, to approve a physical examination test in an accredited hospital (H). In the case that the investigation is positive both parts sign a policy and keep a copy of the contract. If during the investigation irregularities are found, the agent is informed who meets with the client in order to find new options. The insurance policy is in effect when the client makes the first insurance premium payment. Every policy carries with a schedule of premiums, which varies with the type and coverage. Each policy provides a commission for the agency. The commission varies with the insurance company, policy type and coverage. The insurance company management (M) defines

the commissions politic, which varies from agency to agency. The agency splits the commission received for each policy with the agent who sold it; the rate depends on the seniority of the agent. Once a policy has been sold, the agency submits premium bills to the client, collects payment and sends the payment, minus it commission, to the insurance company. If a client fails to pay premiums, the agent who sold the policy is informed, so that he can contact the client. Claims can be made on insurance policies as specified in the policy itself. Clients or beneficiaries (B) contact the agent to file such claims. For an automobile insurance policy, claims are made when the car is involved in an accident, damaged or stolen. The beneficiaries may make life insurance claims on the death of the insured. In both cases, the insurance company sends an adjuster (AD) to legitimate the claim and arrange the final insurance details. Life insurance company's objectives, goals, strategies and Critical Success Factors (CSFs) are defined and the cross-reference identifies the relationship established between the elements. For instance, an external organizational entity identified in the life insurance company context diagram is Management (M). The management has the objective to achieve market leadership position (O1). To accomplish that objective, management's establish to reduce operation cost 5% (G1), achieve 30% of market participation (G2) and improve stockholders gains 15% (G3). The organizational strategies involve penetration into new markets (S2). CSFs like growth through acquisition (C7) will be taken out to each objective and goals proposed.

Facts and beliefs from the confidential knowledge base:

⇑) Leadership. - It is the second seller of life assurance reaching $50 millions this year, with selling projected in $75 millions next year. In car assurance hold the sixth national place.

⇑) Acknowledgement. - Recognized prestige in products and services quality, supported by ISO-9000 certification and the norms of the American Institute of Banking.

⇑) Growing. - Growing potential is attractive in the life assurance segment by virtue of it dominant position in the market.

⇑) Supporting. - The Company is part of one of the must important financial groups of the country, having a significant financial, technology and holding support.

⇑) Strategic alliance with other life assurance companies. - This produced lower operative costs, complement of products and better distribution infrastructure.

⇓) Competence. - In the life assurance segment the company has a dominant role. However, in the car segment the competence has been growing and the company fall from the third to the sixth place.

⇓) Financing. - Nevertheless the company has 1,200,000 clients in the car segment only the 5% of the population pay in advance. In the life assurance segment the 30% of the clients pay in advance.

⇓) Needing of distribution channels. - This had affected the selling and the collecting of car insurance.

⇓) Investment. - The company has to assign a very important part of the investment in new information systems and building restoring.

⇓) Demanding. - Depressing on car demand will affect the selling volume of car assurances.

Facts and beliefs from the popular knowledge base:

⇑) Improving. - Recovering of the internal life assurance the last two years.
⇑) Earning. - Width range of suppliers and clients that allow to increase operation profit.
⇑) Capital outlay. - The lowing of interest rates allows offering high price articles with credit from six to twelve month with out interest.
⇑) Innovation. - Constantly innovation of new programs and business schemes.
⇑) Projecting. - The government agency estimates that only ten percent of the population has a life insurance.
⇑) Demographic growing. - The regular demographic growing guarantees a constant increasing of the sector.
⇓) Investment. - The sector needs high investment to sustain its increasing.
⇓) Competence. - There will be a competence increment in the market.
⇓) Promotion. - There are indispensable intense promotion campaigns for market developing.
⇓) Price changes. - This area of services is sensible to the price alteration and signified changes are difficult to implement.
⇓) Earning level. - Company performance is closely related to the earning level of the population.
⇓) Taxes regulation. - Government controls on selling life and car insurance.

First is necessary establishing Business Strategy but that changes in the organization are limited by the core competencies, i.e. an enterprise that sells computational equipment can be transformed in to an enterprise that sells telecommunication equipment, but would be very difficult to transform it into a financial institution. Then analyze that subcharacteristic will be demanded and establishing weight in agreement to the effects that they cause to the moment to develop the information system. To show the feasibility of the pattern approach, open-source system has been tested on the basis of weighted patterns. The system-based and the business use dimension has analyzed separately, showing that both dimensions are minor developed for analyzed system.

6. Conclusions and Future Works

The main line of this paper contributes to the debate on the adaptability of information systems within a changing organizational environment. To overcome the lack of methodical support, patterns embedded in procedure models to assess and evaluate the adaptability of information systems were outlined. For this has treated to the identified of certain subcharacteristics, thanks pattern in the business ambient for establishing a model who allows to turn the attribute Adaptability into a characteristic to demand into all the stages of the life's cycle Information System development, guaranteeing that will be possible to modify it agree also they change the business processes, bearing in mind that increasingly the systems are in use not for taking only the finance in a company, but for competing and field for being opened in bought increasingly competitive. One of the identified subcharacteristics that it are considered to be they must form a part of the model it is the verification, which, it happens after guaranteeing that the requirements of the information system orientated

to the business went due validation. Of the subcharacteristics that are suggested for a future model of evaluation of the Adaptability in Information Systems to the changes in business processes, the Flexibility is often more difficult to achieve that the Extensibility specially when during the development they change and new needs and requirements are specified. The complexity of the current systems does more improbable than some offers of Software Engineering for the development of information system be perfect, if for no other reason, more that for the fact that the system has changed, has to modify the developed work though it has been elaborated well, this bears to that the subcharacteristic Tuning should be demanded nowadays in the software. Only if the Information Systems are adaptable, it will be reasonable to introduce changes and/or to arrange faults, with a low probability that the new changes introduce new faults. Adaptability in Information Systems to the changes in the business processes, do that the company in question manages to identify like reliability for your clients and eligible instead of other options.

Our proposed future direction would be to continue to working with model, formalizer, but to deepen the analysis from the accumulation of comparative observation and from application of theoretical perspectives. In addition it propose for other authors it concerned with the development of design of the agent-based information system for administrative documentation management in this type the project. We would hope eventually to move in the direction of active models, but would need to add a political analysis before we took that step.

References

[1] Miguel Arjona Torres. Dirección estratégica: Principios y aplicaciones de la gestión del rendimiento. Published 1999. Ediciones Díaz de Santos. 236 pages. ISBN 8479783869

[2] Peter Norvig and David Cohn. Harlequin Incorporated. http://www.harlequin.com

[3] Mohamed Fayad & Marshall P. Cline. Aspects of Software Adaptability. Communications of the ACM archive, Volume 39, Issue 10 (October 1996). Pages: 58 – 59.1996. ISSN:0001-0782. ACM Press New York, NY, USA

[4] NMX-I-055-NYCE: Tecnología de la información-Ingeniería de software-calidad de producto- Parte 01, Parte 02, Parte 03, Parte 04, 2005.

[5] ISO/IEC 9126, Software Product Evaluation, Quality, Characteristics and Guidelines for their Use. Information Technology, 1991.

[6] Berg, B., Cline, M., Girou, M. Lessons learned from the OS/400 OO project. *Commun. ACM 38*, 10 (1995), 54-64.

[7] Erich Gamma, Richard Helm, Ralph Johnson y John Vlissides. Design Patterns Elements of Reusable Object-Oriented Software. Addison Wesley Publishing Company. 1996.

[8] Balve, P., Wiendahl, H, Westkämper, E.: Order management in transformable business structures - basics and concepts. In: Robotics and Computer Integrated Manufacturing, Volume 17, Issue 6, 2001; 461-468.

[9] Gutiérrez A., Clempner J. Information Technology Strategic Planning: Business Process Modeling. 6th World Multiconference on Systemics, Cybernetics and Informatics (SCI 2002) Orlando, USA, in July 14-18. 2002.

[10] ISO/IEC JTC1/SC7 3790. NWI Proposal - Software Engineering. Software Quality Requirements and Evaluation (SQuaRE) . Evaluation Module for Recoverability. 2007-06-28.

[11] Katja Andresen, Norbert Gronau. Managing change –determining the adaptability of informaiton systems. European and Mediterranean Conference on Information Systems (EMCIS) 2006, July 6-7 2006, Costa Blanca, Alicante, Spain.

PID control under sampling period constrains

*Alfredo Cuesta, **Luis Grau, and **Ignacio López

*Felipe II College (Complutense Univ.)
**Communications and Control Systems Dept. (UNED)
acuesta@cesfelipesegundo.com
{lgrau,ilopez}@scc.uned.es
http://www.scc.uned.es/protedico/

Abstract. Digital redesign by emulation of the continuous is a quite frequent method for building digital controllers. However, as the sampling period increases the performance is degraded. One solution is to design the controller directly in the discrete domain, with the advantage of having the sampling period as an extra parameter. But when there are sampling constrains either in sensors or actuators it is necessary to choose as sampling period the slowest one. Moreover, such a period becomes a lower bound. Multi-rate controllers arise due to these sampling constrains releasing from the hypothesis of having a single period. This is why it is interesting to consider a so widespread controller such as the PID in multi-rate situations. In this paper a direct design method, based on Modified Ziegler Nichols, for PID controllers is presented. The novelty of this paper is to consider the sampling period as a degree of freedom and, when it comes too large, to improve the performance by turning it into a multi-rate PID. With the method proposed both gain and phase margin requirements can be obtained.

Key words: Multi-rate, PID, Ziegler-Nichols, Tuning methods, Kranc operators

1 Introduction

It is remarkable that the design of controllers is still carried out in the continuous domain even when these are finally built in digital devices. Just considering PID controllers, there are about one thousand tuning rules compiled in [1], all of them for continuous systems. Afterwards, taking a sampling period fast enough and then applying some approximation of the derivative such as Backward differences or Tustin, a discrete PID is obtained. However this method is no longer right if the sampling period is above a threshold. Designing techniques directly in the discrete domain considering the sampling period as an extra degree of freedom have been proposed [2],[3] . Besides sampling constrains either on sensors or actuators can prevent from using a single sampling rate or even from choosing it within a given range, let alone small enough to emulate the continuous. A system under this constrains is said to be Multi-rate (MR). There are two possible control strategies: Multi-Rate Input Controller (MRIC) in which the manipulated

© L. Sánchez, O. Pogrebnyak and E. Rubio (Eds.)
Industrial Informatics
Research in Computing Science 31, 2007, pp. 63-72

variable is updated faster than the measured variable, and Multi-Rate Output Controller (MROC) which is the complementary case. Chemical processes [4], polymer reactors [5] or sector servo of hard disk drives [6] are MRIC examples where the control action could be updated N_u times for each measure of the output. On the other hand, artificial vision systems are a MROC example in which the large amount of incoming data is faster than the algorithm working on it.

Multi-rate PID (MRPID) controllers have been built for both strategies with different techniques [7]-[12]. All prove very good performance but the controller obtained is not expressed in term of the classical gains K_P, K_I and K_D. Using Kranc operators technique [13] the authors presented a model for a MRIC PID controller that preserves the gains [14] but has the drawback of dealing with MIMO systems.

This work considers the general case for MRPID in which the measured variable is updated every T_e seconds and the manipulated variable every T_u seconds. As a consequence both MRIC and MROC strategies are covered. The contribution of this paper is to present the advantages of such a MRPID for obtaining a range of useful phase margins and a given gain margin, bringing it face to face with the conventional single-rate PID (SRPID) in two different scenarios: first taking the sampling period as a parameter design; and then assuming there is a lower bound for the sampling period which is the slowest one found in the control system. The disadvantages remarked above are coped so that finally a simple tuning method in the frequency domain is obtained. The rest of the paper is organised as follows. Section II presents the MRPID controller and its modelling. Section III is devoted to the tuning method in the frequency domain. Section IV shows an study case and finally the conclusions are given.

Notation

$G(s)$:	Continuous transfer function
$G(z), G^{T_s}(z)$:	Discrete transfer function (from an algorithm with period T_s)
$\underline{G}(z)$:	Kranc Operator of G
l.c.m.$(a, b, c, ...)$:	Least common multiple of a, b, c, ...
$\underline{R}^T, \underline{E}^T, \underline{U}^T, \underline{Y}^T$:	Discrete transfer function of vectorised signals sampled at metaperiod T.

N_e : Error rate	N_u : Control rate
N_P : Proportional rate	N_I : Integral rate
N_D : Derivative rate	T : Metaperiod
m: $l.c.m.(N_e, N_P, N_I, N_D)$	n: $l.c.m.(N_u, N_P, N_I, N_D)$

2 Model

Let us consider a parallel digital PID controller in which the backward differences approximation of the derivative, $s = (z-1)/zT$, is used. Let us assume that the

error is measured with period T_e and the control action is updated with period T_u. Hence, there are two intuitive approaches to the way the controller internally works. The simplest one is to compute the control action value in terms of the last error sample and send it out with period T_u until a new error sample is received. By doing so a SRPID working at the larger period is finally obtained. A more elaborate and natural decision is to consider the error as a sequence of steps and the control action as the sum of every basic control component (i.e. proportional, integral and derivative) response to such steps. Still a more general case can be considered with this approach if the proportional, integral and derivative components work at periods T_{prp}, T_{int} and T_{der} respectively. This scheme is depicted in Fig. 1(a), where the superscript gives extra information about the sampling time to the Z-transform.

Any given basic action will seldom take place at the same instant than any other one or even the final control action or the error sampling. Therefore, following the second approach, the simplest solution to this problem is to hold the error sample for T_e seconds so that it is always available for every component. Every basic action is also hold during its own period making feasible to add the three of them at any time. This output is finally sampled every T_u seconds. Kranc operators methodology [13] has been used throughout this paper. Kranc operators transform a MR system into a Single Rate MIMO system working at metaperiod T, defined as the l.c.m. of all the periods found in the system. The two basic ideas behind Kranc operators are the vector switch decomposition (VSD) [15] and the approximation of continuous signals by fictitious samplers at high rates [16]. Thus a sampler element with period T/N is modelled by an expand operator $[N^+]$ which vectorises the input signal in N channels so that every channel is advanced with respect to the next one T/N seconds; then all the channels are sampled at metaperiod T so that in T seconds N samples of the signal are taken and finally the signal is reconstructed with a reduce operator $[N^-]$ as defined in [13]. So, first of all, the periods appearing in Fig. 1(a) must be rewritten in terms of the metaperiod T: $T_e = T/N_e$, $Tu = T/N_u$, $T_{prp}=T/N_P$, $T_{int}=T/N_I$, $T_{der}=T/N_D$. Due to space limitations and since of the paper is focused on the tuning method in the frequency domain we encourage to find an exhaustive description of the model in [17]

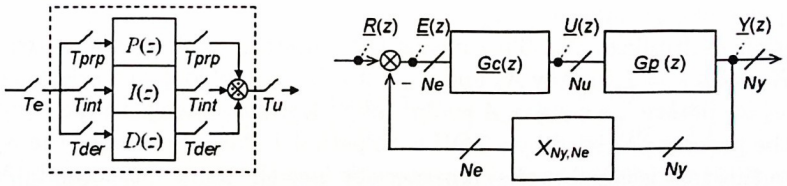


Fig. 1. MRPID scheme (a) and block diagram using Kranc Operators (b).

When the MRPID controller is inserted in a control loop, the process and the feedback must be also modelled using Kranc operators [13]. The control loop is finally represented by the block diagram shown in Fig. 1(b) Here $Gc(z)$ is the model of the MRPID controller obtained above and $Gp(z)$ is the Kranc operator of the continuous process $Gp(s)$ preceded by a ZOH with period T/N_u. The continuous output of the process is approximated with a fictitious sampler that measures N_y times in a metaperiod. Since just N_e out of these N_y samples must be fed back, the block $X_{Ny,Ne}$ is necessary. The figure also represents the vectorisation by a crossing line that indicates the number of channels it has in.

3 Tuning Method

The open loop transfer matrix of the MR system shown in Fig. 1 is given by:

$$M(z) = X_{Ny,Ne} \cdot \underline{Gp}(z)\underline{Gc}(z), \tag{1}$$

Thus $M(z)$ is a square transfer matrix with N_e rows and columns and therefore the Nyquist curve must be generalised to the characteristic loci [18], which is the eigenvalue set of $M(e^{j\omega T})$ for $0 \leq \omega \leq \pi/T$. In order to generalise the Modified ZN method to the MR case, first note that

$$\underline{XG}(z) = X_{Ny,Ne} \cdot \underline{Gp}(z) = [g_{ij}(z)], \tag{2}$$

with $1 \leq i \leq Ne$ and $1 \leq j \leq Nu$, is the open loop transfer matrix without the controller. Let us consider that a MR Proportional controller $\underline{C_P}(z)$ with $K_P = 1$ and $N_P = 1$ is connected. Such a controller is represented by a transfer matrix in which every element is equal to zero but the first column is equal to one. Then the element in the position (i, j) of $M_P(z) = \underline{XG}(z)\underline{C_P}(z)$ is given by:

$$m_{ij} = \begin{cases} \sum_{q=1}^{Nu} g_{i,q}(z) & \text{if } j = 1 \\ 0 & \text{if } 1 < j \leq Ne \end{cases} \tag{3}$$

The eigenvalue set of the matrix $M_P(e^{j\omega T})$ is

$$\lambda_{1,2,\ldots,Ne} = \sum_{q=1}^{Nu} g_{1,q}(e^{j\omega T}) = Gp^T(e^{j\omega T}) \tag{4}$$

with multiplicity N_e. This means that inserting a Proportional controller with $K_P = N_P = 1$, for any N_e and N_u, into the control loop and selecting a frequency ω_0 we determine a point $A = Gp^T(e^{j\omega_0 T})$, which belongs to the Nyquist curve of the process preceded by a ZOH with period T. Moreover, since the Nyquist curve in this case is equal to the characteristic loci, the point A is actually N_e coincident points $A = A_{P1} = \ldots = A_{PN_e}$ where the subscript P indicates they were obtained just with a proportional controller. These A_{Pi} are complex numbers which have a well-known correspondence in $\mathbb{R}^2$, so from now on complex points will be referred

to as elements (or points) of $\mathbb{R}^2$. Increasing the proportional internal rate N_P splits point A so that in the most general case $A \neq A_{P1} \neq ... \neq A_{PN_e}$. Then, modifying K_P moves every A_{Pi} over the line that goes through the origin and A_{Pi} so that $|A_{Pi}|$ is a function of K_P. Let $\underline{C_I}(z)$ be a MR Integral controller with $N_I = 1$ and $K_I = 0$, which is now connected instead of $\underline{C_P}(z)$ in (1). Following the same reasoning as above, point A is now split, by means of the controller, for the frequency ω_0, into N_e points $A_{I1} \neq ... \neq A_{IN_e} \neq A$; and again for each A_{Ii} its module $|A_{Ii}|$ is a function of K_I. The same can be said for a MR Derivative controller $\underline{C_D}(z)$ with $N_D = 1$ and $K_D = 0$ so that in the end each $|A_{Di}|$ is a function of K_D.

We are now in a position of considering the complete MRPID $\underline{Gc}(z) = \underline{C_P}(z) + \underline{C_I}(z) + \underline{C_D}(z)$, for any set of internal rates and gains. Once it is connected into the open loop, (1) becomes

$$M(z) = X_{Ny,Ne} \cdot \underline{Gp}(z)\,(\underline{C_P}(z) + \underline{C_I}(z) + \underline{C_D}(z)) \tag{5}$$

The eigenvalue set of (5) for $z = e^{j\omega_0 T}$ is therefore given by

$$\{B_i = A_{Pi} + A_{Ii} + A_{Di}\}, \text{ with } i = 1, ..., Ne \tag{6}$$

Consequently once the internal rates N_P, N_I and N_D as well as the frequency ω_0 have been fixed, the eigenvalue set of (5) depends only on K_P, K_I and K_D. In this paper we propose to choose gains K_P, K_I and K_D to assure a closed loop gain margin of γ dB, which means that the characteristic loci has satisfy two conditions: it must cut the negative x-axis in $B = -10^{-\gamma/20}$ and nowhere else to the left. This is the task the next two subsections are devoted to.

3.1 Computing the Proportional, Integral and Derivative main directions

Since N_e points B_i are obtained with (6) it is necessary to select just one. Let $B_1 = A_{P1} + A_{I1} + A_{D1}$ be the eigenvalue that will match the required point B, then we define the unitary vectors

$$\mathbf{V_P} = \frac{A_{P1}}{|A_{P1}|},\ \mathbf{V_I} = \frac{A_{I1}}{|A_{I1}|},\ \mathbf{V_D} = \frac{A_{D1}}{|A_{D1}|} \tag{7}$$

and the proportional, integral and derivative main directions as the directions established by $\mathbf{V_P}$, $\mathbf{V_I}$ and $\mathbf{V_D}$ respectively. It is clear that points A_{P1}, A_{I1} and A_{D1} must be properly chosen out of their respective sets $\{A_{Pi}\}$, $\{A_{Ii}\}$ and $\{A_{Di}\}$ for satisfying the conditions of the given gain margin. As a first approach we recommend to choose A_{P1} as

$$A_{P1} = \min_{A_{Pi}} \text{abs}\left(\text{angle}\left(\frac{A_{Pi}}{|A_{Pi}|}; Ne\right) - \text{angle}\left(\frac{A_P}{|A_P|}; 1\right)\right) \tag{8}$$

where the first term is the angle of the unitary vector established by A_{Pi} using the actual controller (i.e. with the given $N_e \geq 1$) and the second term is the angle of the unitary vector established by the unique A_P if N_e was 1. A similar criterion should be used for A_{I1} and A_{D1}.

3.2 Computing K_P, K_I and K_D

Now it is possible to generalise Modified ZN method to the MR situation shown in Fig. 1 with the following procedure:

1. Connect a MRPID controller with $N_P = N_I = N_D = K_P = 1$ and $K_I = K_D = 0$.
2. Select a frequency $\omega_0 \in [0, \omega_0/T)$. This frequency determines a point $A = Gp^T(e^{j\omega_0 T})$ in the Nyquist plot of the process preceded by a ZOH with period T .
3. Make N_P, N_I and N_D equal to the greater of N_e and N_u.
4. Obtain $\{A_{Pi}\}$, the eigenvalue set of the open loop transfer matrix considering only proportional action with $K_P = 1$. Similarly, obtain $\{A_{Ii}\}$ and $\{A_{Di}\}$ considering only integral ($K_I = 1$) and derivative action ($K_D = 1$) respectively.
5. Select $\{A_{P1}\}$, $\{A_{I1}\}$ and $\{A_{D1}\}$ according to (8) and obtain $\mathbf{V}_\mathrm{P}$, $\mathbf{V}_\mathrm{I}$ and $\mathbf{V}_\mathrm{D}$ with (7).
6. Select a gain margin of γ dB. This gain margin determines a point $B = -10^{-\gamma/20}$.
7. Point B can be rewritten as $B = \rho_P \mathbf{V}_\mathrm{P} + \rho_I \mathbf{V}_\mathrm{I} + \rho_D \mathbf{V}_\mathrm{D}$, where ρ_P, ρ_I and ρ_D depend on K_P, K_I and K_D respectively. Although the relation between K and ρ is hard (or impossible) to determine, an iterative and simple Newton algorithm rapidly converges on a solution. The main steps are:
   - Consider only proportional action and give an initial value to K_P.
   - Obtain $\{A_{Pi}\}$ for that K_P and select A_{P1} as the one in the direction of $\mathbf{V}_\mathrm{P}$.
   - If $|A_{P1}| = \rho_P$ then K_P is the solution. Otherwise repeat the process with another value of K_P.
   - Repeat the steps for integral and derivative actions, obtaining K_I and K_D.

It is important to remark the following issues:

1. Unlike for the continuous controller, $\mathbf{V}_\mathrm{P}$, $\mathbf{V}_\mathrm{I}$ and $\mathbf{V}_\mathrm{D}$ are generally not orthogonal.
2. The choice of $\mathbf{V}_\mathrm{P}$, $\mathbf{V}_\mathrm{I}$ and $\mathbf{V}_\mathrm{D}$ according to criterion (8) is not always right so, once the controller has been tuned, it is necessary to verify that point B is actually in the characteristic loci. If not then other main directions must be chosen.
3. The gain margin must be also checked out, for the method only assures one value of the characteristic loci which is point B.

3.3 Derivative action effect on both gain and phase margins

When using MRPI or MRPD controllers there is a unique couple (ρ_P, ρ_I) or (ρ_P, ρ_D) but for MRPID infinite combinations of ρ_P, ρ_I and ρ_D are possible so an extra constrain is needed for fixing a unique combination (ρ_P, ρ_I, ρ_D). When

$N_e = 1$, it has been proposed to try different K_P (i.e. ρ_P) for it has the effect of modifying the phase margin keeping the gain margin [14]. The same can be applied now for $N_e > 1$. This paper gets deeper in the study of when both gain and phase margins requirements can be satisfied. So far a frequency ω_0 must be previously fixed so that point A is moved to position B to satisfy the gain margin requirement; but another frequency could also accomplish the same goal, perhaps with a wider or more appropriate phase margin range. Thus, if ω_0 is a degree of freedom the following algorithm is proposed:

1. Carry out the procedure of computing K_P, K_I and K_D selecting the ultimate frequency as the initial value for ω_0 and fixing $K_D = 0$. Consequently the proportional gain gets its maximum value K_{Pmax}.
2. Fix $K_P = 0.95K_{Pmax}$ and compute K_I and K_D again.
3. Repeat step 2 reducing another 5% each time until $K_P = 0$ or the gain margin is not satisfied because of the characteristic loci has been excessively bent. Thus a phase margin range has been obtained for the same gain margin.
4. Set another frequency ω_0, smaller than the previous one since it was the upper bound and repeat steps 1, 2 and 3. Thus a new phase margin range is obtained, probably intersecting with others so eventually a gain margin range for a given phase margin is also obtained

4 Example

Let us consider the continuous process given by the transfer function $Gp(s) = 25(s+1)^{-1}(s^2+s+25)^{-1}$ and a the following robustness requirements: a gain margin of $\gamma = 26$ dB and a phase margin $\phi = 55°$. Two controllers were tested under three error sampling periods: $T_e = \{50 \text{ ms}, 125 \text{ ms}, 250 \text{ ms}\}$; the first one was a conventional discrete PID (SRPID) that updates the control action at $T_u = T_e$, and the second one a MRPID decreasing T_u to 40% of T_e; i.e. $T_u = \{20 \text{ ms}, 50 \text{ ms}, 100 \text{ ms}\}$. Notice that both cases can be modelled by Kranc Operators technique, taking a metaperiod $T = \{100 \text{ ms}, 250 \text{ ms}, 500 \text{ ms}\}$, with $N_e = N_u = N_P = N_I = N_D = 2$ for the SRPID and $Ne = 2; N_u = N_P = N_I = N_D = 5$ for the MRPID. Let $Gp^T(z)$ denote the discrete model of the process with a given period T.

As a first approach, the frequency $\omega_0 = 4.4$ rad/s was taken. We choice such a ω_0 because it is associated to a point A in the Nyquist plot of $Gp^T(z)$. Then both SRPID and MRPID were tuned for different K_P, always satisfying the gain margin requirement. The results are depicted in Fig. 2(a), where the greater and lower phase margin obtained with the MRPID ($\times$) and with the SRPID ($\cdot$) are shown. Thus an upper and lower bound for phase margins are obtained. It can also be seen that just for $T_e = 20$ ms the MRPID is able to fulfil the robustness while the MRPID allows to have 20 ms $\leq T_e \leq$ 100 ms.

In a second approach all the points in the third quadrant, instead of a single point A, belonging to the Nyquist plot of $Gp^T(z)$ were scanned. Again both SRPID and MRPID where tuned with several K_P, always satisfying the gain margin requirement. The results have been depicted in Fig. 2(b), where the

symbols × and · have the same meaning as above. Again an upper and lower bound is found for both controllers. Thus, any point within the solid lines means a phase margin that is reachable for a certain error period Te using the MRPID controller. The same can be said for the SRPID and the dotted lines. Notice that the figure proves that the requirements can be satisfy but does not include the information about neither K_P nor ω_0. However since Fig. 2 was obtained with an iterative method such information can be saved and retrieved whenever.

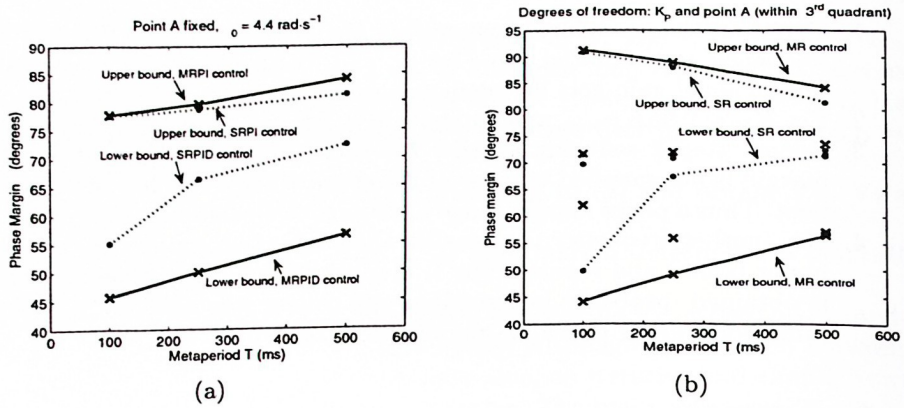


Fig. 2. Phase margin range when point A is fixed (a) and when is free within the 3rd quadrant (b)

This example shows how MRPID controllers can improve the performance of a system in which there are severe sampling constrains in the sensors while the actuators can work much faster. For instance, from Fig. 2(b) it can be seen that if $T_e = 125$ ms and $T_u = 50$ ms then it is impossible to satisfy both gain and phase margin requirements using a conventional SRPID but it is using a MRPID instead. Thus, tuning the both SRPID and MRPID to obtain the phase margin closest to 55° satisfying the gain margin we obtain the characteristic loci, the step response and control action depicted in Fig. 3.

5 Conclusions

This paper introduces an indirect tuning method for the derivative gain for the more general multi-rate situation. It is indirect because decisions are taken over K_P in terms of robustness requirements and K_D is recomputed in terms of K_P. This is why MRPID controllers are an interesting control strategy, not only for coping with sampling constrains either in the error or in the control action, but also for taking advantage of the sampling period considering it as an extra parameter since SRPID can be considered as a particular case. Due to it

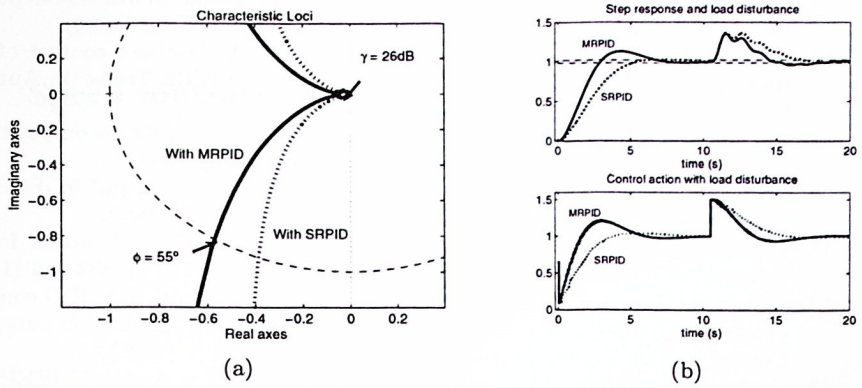


Fig. 3. Characteristic loci (a), step response, load rejection and control action (b) with both MRPID and SRPID. The sampling constrains are $T_e = 125$ ms and $T_u = 50$ ms and 55° are required. The SRPID is not able to satisfy it.

is necessary to solve an eigenvalue equation it not possible to give an analytical solution and even for those cases in which it is, the computational solution is more straightforward.

Acknowledgments. This work was supported by Ministerio de Educación y Ciencia de España. (Proj. TEDICO. Ref. DPI2004-05903).

References

1. O'Dwyer, A. "Handbook of PI and PID" Ed. Imperial College Press (2003).
2. López, I., Dormido, S., and Morilla, F. "The sampling period as a control parameter" in 2nd Symp. on Intelligent Components and Instruments for Control Applications, Budapest (1994).
3. González, R. López, I., Morilla, F and Pastor, R. "SintoLab: The REPSOL-YPF PID tuning tool" in Control Engineering Practice, Pergamon. pp. 1469-1480 (2003).
4. Morant, F., Albertos, P., and Crespo, A. "Oxide composition control in a raw material mill" in IFAC Symp. on Microprocessor in Control. Istambul (1986).
5. Ohshima, M., Hashimoto I., Takeda, M., Yoneyama, T. and Goto, F. "Multirate multivariable model predictive control and its application to a semi-commercial polymerization reactor" in Proc. of the ACC, 2, pp.1576-1581 (1992).
6. Nemani, M., Tsao, T.C., and Hutchinson, S. "Multirate analysis and design of visual feedback servo control systems" in ASME Journal of Dynamic Systems, Measurement and Control. Vol 116, no. 1. pp.45-55 (1994).
7. Araki, M., Yamamoto, K. "Multivariable multirate sampled-data systems: State-space description, transfer characteristics, and Nyquist criterion" in IEEE Trans. on Automatic Control, Vol. AC-31, No. 2, pp. 145-154. (1986).
8. Araki, M., Hagiwara, T. "Pole assignment by multirate-sampled data output feedback" in Int. journal of control. Vo. 44, pp.1661-1673 (1986).

9. Hagiwara, T., Fujimura, T.F., Araki, M. "Generalizad multirate-output controllers" in Int. Journal of Control. Vol 52, No. 3, pp. 597-612 (1990).
10. Khargonekar, P.P., Poolla, K., and Tannenbaum, A. "Robust control of linear time invatiant plants using periodic compensation" in IEEE Trans. on Automatic Control, Vol. AC-30. No. 11, pp.1088-1096.
11. Salt, J. and Albertos, P. "Multirate controllers design by rate decomposition" in Proc. of 30th conference on decision and control, Sydney (2000).
12. Salt, J. and Albertos, P. "Model-based multirate controllers design" in IEEE trans. On control systems technology. Vol. 13, No.6, pp. 988-997 (2005).
13. Thompson, P.M. "Gain and phase margins of multirate sampled-data feedback systems" in International Journal of Control. Vol. 44, No. 3, pp.833-846 (1986).
14. Cuesta, A. Grau, L., López, I. "CACSD tools for tuning Multi-Rate PID controllers in time and frequency domain" in Proc. of the IEEE Int. Conf. on Computer Aided Control System Design. pp. 3036-3041. Munich, Germany (2006).
15. Kranc, G.M. "Input-output analysis of multirate feedback systems". IEEE Trans. on Automatic Control. Vol. AC-3, pp.21-28 (1957).
16. Anderson, B.D.O., Keller, J.P. "Discretization techniques in control systems". Control and Dynamic systems. Vol 66. pp. 47-93 (1994).
17. Cuesta, A. Grau, L., López, I. "Generalised tuning method for Multi-Rate PID controllers" in Proc. of the CSC'07, May 16-18, 2007, Marrakech, Morocco (2007).
18. Maceijowski, J.M. "Multivariable Feedback Design" Addison-Wesley. Cornwall, England (1989) Cuesta, A. Grau, L., López, I. "CACSD Tools for Tuning Multi-Rate PID Controllers in Time and Frequency Domains" in Proc. of the IEEE Symposium on CACSD, Munich (2006).

Speed control of a sugar cane belt conveyor with a fuzzy controller

Mercedes Ramírez[1], Modesto Angulo[1], Ariel Domínguez[1],
Pedro Albertos[2]

[1] Department of Automatic Control, Faculty of Electrical Engineering,
Universidad de Oriente. Santiago de Cuba, Cuba.
{mramirez, angulo, arield}@fie.uo.edu.cu
[2] Department of Systems Engineering and Control,
Universidad Politécnica de Valencia, España.
pedro@aii.upv.es

Abstract. This paper describes the experience attained by the authors in the design of a fuzzy controller for speed control of an induction motor in field oriented, that drives a sugar cane belt conveyor, without making the compensations in the d-q axes. The Mamdani type fuzzy controller was implemented which utilizes speed error and the rate of change of the speed error as input variables and the current component in the q-axis as output variable. The simulation results reached show that the behavior of the control system under sudden changes of load and reference are better than those results obtained using PID classic.

Keywords: Fuzzy control, Speed control, Induction motor, field oriented control Simulation, PI controller.

1 Introduction

The function of a raw sugar factory is to produce sugar from the juice of sugar canes delivered to the factory. On arrival at the plant, the sugar cane is transported on belt conveyors to the shredder where it is prepared for the removal of juice by the extraction station. The shredder prepares the cane by smashing it up into small pieces. The extraction process is done by crushing mills. The prepared cane from the shredder is passed through a series of mills called the milling train as a whole. The mills squeeze the cane to separate the juice which contains the sugar from its fibrous part.

The feeding conveyor is moved by an electric motor with winding rotor and double or triple stages of speed reduction. In this process the control of speed takes place by insertion of external resistances in the rotor circuit, it is a discrete control and introduces strong electric losses and heating. On the other hand the presence of gearboxes introduces mechanical losses. Both situations contribute to reduce the efficiency of these systems, being one of the fields that can be improved.

© L. Sánchez, O. Pogrebnyak and E. Rubio (Eds.)
Industrial Informatics
Research in Computing Science 31, 2007, pp. 73-84

To achieve a high performance of the AC motor drives the vector control is used [1]. The decoupling characteristics of vector controlled induction motor are adversely affected by the parameter changes in the motor [2].

Actually in a wide range of power the induction machine is fed with an inverter with semiconductor devices acting as voltage source using Pulse Width Modulation techniques (PWM) whose commutation frequency is usually in the range 3KHz -12 KHz. [3].

Since this converters are used with powers up to 1000 kW (and higher) and one of the main applications is the position servomechanism of high dynamic behavior, nowadays it is possible to reach a quick control of the stator currents with sinusoidal references of stator current in closed loop. For it, the stator currents can be controlled by quick regulation loops, and the stator voltage or currents equations can be used, depending of motor capacity.

The motor control issues have been traditionally handled by proportional integral derivative controllers (PID). However, the fixed gain controllers are very sensitive to parameter variations, load disturbances, etc. So, the controller parameters have to be continuously adapted. The problem can be solved by several adaptive control techniques such as model reference adaptive control (MRAC) [4], sliding mode control (SMC) [5], variable structure control (VSC) and self tuning PI controllers [6], etc.

The design of all of the above controllers depends on the exactitude of the mathematical model of the system. However, it is often difficult to develop an accurate mathematical model of the system due to unknown load variation, unknown and unavoidable parameter variations due to saturation, temperature variations and system disturbances.

The necessity to obtain more accurate systems, facing with success their unavoidable disturbances and nonlinearities and the loads associated to them has propitiated the use of intelligent techniques: fuzzy control, neural nets, genetic algorithms and others [7], [8], [9], [10]. To solve the problem outlined in this work a fuzzy controller (FLC), is proposed, which is based on the Fuzzy Logic introduced at the first time by Zadeh [11].

The FLC has some advantages such as: (1) it does not need any exact system mathematical model; (2) it can handle nonlinearity of arbitrary complexity and (3) it is based on the linguistic rules with IF-THEN general structure which are deduced from the knowledge of an expert [12].

In the following section the mathematical models of the components of the system are described in terms of the controller's inputs and outputs. In the section 3 the developed fuzzy controller is analyzed in detail, explaining their structure and operation.

In the section 4 the simulation scheme is outlined and in the section 5 the results are examined in comparison with those obtained with the use of classic controllers PID. Finally, in section 6 the conclusions of this work are exposed.

2 Mathematical Model of System Components

2.1 Induction motor model.

The model that will be used in this work is determined by the equations which define the dynamic behavior of the machine in oriented field, in this case in the reference system of the axes d (direct) and q (transversal) [2].

$$\sigma\tau_s \frac{di_{sd}}{dt} + i_{sd} = \frac{u_{sd}}{R_s} + \sigma\tau_s\omega_{mr}i_{sq} - (1-\sigma)\tau_s \frac{di_{mr}}{dt} \tag{1}$$

$$\sigma\tau_s \frac{di_{sq}}{dt} + i_{sq} = \frac{u_{sq}}{R_s} - \sigma\tau_s\omega_{mr}i_{sd} - (1-\sigma)\tau_s\omega_{mr}i_{mr} \tag{2}$$

$$\tau_r \cdot \frac{di_{mr}}{dt} + i_{mr} = i_{ds} \tag{3}$$

$$0 = i_{qs} - \left(\omega_{mr} - \omega_r\right)\tau_r \cdot i_{mr} \tag{4}$$

Where:

σ : Bondel's dispersion coefficient

τ_s : Stator's time constant (s)

τ_r : Rotor's time constant (s)

ω_{mr} : Speed of the reference axis (1/s)

i_{sd} : Current component in d axis (A)

i_{sq} : Current component in q axis (A)

u_{sd} : Voltage component in d axis (V)

u_{sq} : Voltage component in q axis (V)

L_m : Per phase magnetizing inductance (H)

i_{mr} : Magnetizing current (A)

The previous equations represent the conversion of u_{sd} and u_{sq} in i_{sd} and i_{sq} and in this case they are the consequence of feeding the motor with a voltage source. That is to say that they are the equations of the induction motor in a reference system

that rotates at the speed of the rotor magnetic flux (dynamics of the induction motor in field oriented) [13].

Equations of electromagnetic torque

$$T_e = \frac{3}{2}\frac{P}{2}\frac{L_m}{L_r}\cdot\lambda_r\cdot i_{qs} = \frac{3}{2}\frac{P}{2}\frac{L_m^2}{L_r}\cdot i_{mr}\cdot i_{qs} \tag{5}$$

Where:

λ_r: component of rotor magnetic flux
P: number of poles of the rotor

Equation of mechanical part

$$J\frac{d\omega}{dt} + B\omega = T_e - T_l \tag{6}$$

Where J is the total inertial torque, ω is the axis rotation speed, B is the friction, T_e is the machine torque and T_l is the load torque. The equation of mechanical part as a function of electric speeds is:

$$\frac{1}{P}J\frac{d\omega}{dt} + B\omega = Pki_{mr}i_{sq} - T_l \tag{7}$$

The blocks diagram of the motor fed with a voltage source (See Fig. 1) can be obtained from the previous model and the interrelation among its equations.

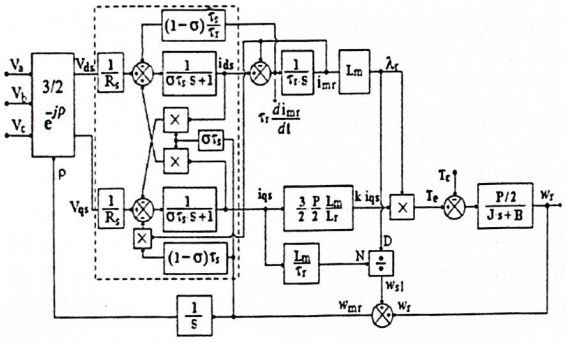


Fig. 1. Induction motor in field oriented fed with voltage source.

2.2 Mathematical model of cane belt conveyor

The cane belt conveyor consists of a long horizontal tract located below the floor level which transports the cane to the feeding conveyor and is built with steel splints. A scheme is shown in Fig. 2.

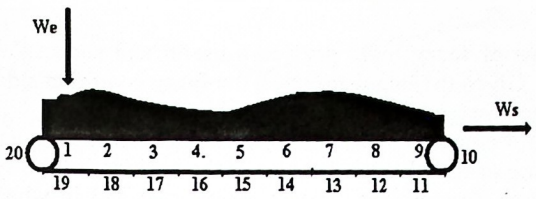


Fig. 2. Schematic representation of a belt conveyor.

When there is variation of speed due to fluctuations in the load or in the reference of the motor, it will present a dynamic that should be kept in mind in the model of the conveyor.

The static and dynamic behavior of the drive is determined by the mechanical equation of the system. When all the components that contribute both to the generation of the moment of the motor and to the resistant moment are known, the behavior of the group can be expressed simply by means of the equations of the Newton classic mechanics.

Thus, the equation (8) that appears developed in [14] is taken as dynamic model of the sugar cane conveyor. The model is represented in schematic form in Fig. 3.

$$T_l(s) = \frac{W_o r}{\eta i}\frac{1}{s} + \frac{m_{eq} r^2}{n i^2} s\omega(s) \tag{8}$$

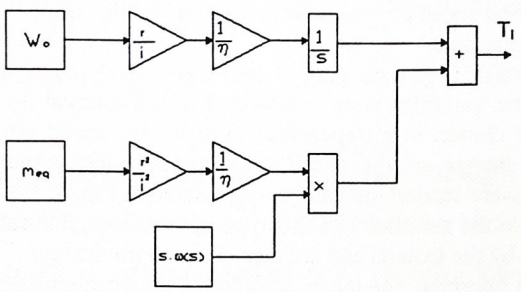


Fig. 3. Model equivalent of the sugar cane conveyor.

Where:

Wo: Traction force in stationary state

η: Efficiency of the system transmission

m_{eq}: Equivalent mass

r: ratio of toothed wheel

i: transmission relationship among the axis that moves the toothed wheel and motor axis

T_l: Load torque

3 Design of FLC

The mechanisms of fuzzy logic provide a useful and successful tool to implement linguistic rules. Given the linguistic rules, the fuzzy controller's development does not demand too much time and its cost is not high.

The fact that the controller is carried out by means of linguistic rules makes easy the understanding of its behavior. In short, if the behavior of the system in closed loop is not satisfactory, determining the responsible rule and in what address it must be modified is usually possible [15].

In this work, a fuzzy controller of Mamdani type was designed to control the speed of the group motor-conveyor and the "Fuzzy Logic" Toolbox was used as tool. This controller should provide the control of the current i_q that is required in order that the motor reaches the reference speed.

Two input linguistic variables were considered: The speed error ($e\omega(k)$) and the rate of change of the speed error ($ce\omega(k)$) and an output variable: the current of the axis q (iq). These variables were defined respectively as:

$$e\omega(k) = \omega_r(k) - \omega(k) \tag{9}$$

$$ce\omega(k) = e\omega(k) - e\omega(k-1) \tag{10}$$

where:
ω_r: reference speed
ω: the rotation speed of the motor in each moment
k: the discrete time.

The linguistic variables were normalized in the interval [-1 1]. The membership functions were chosen in a trapezoidal form for the speed error and in a triangular form for the change of the speed error and for the output. The corresponding partitions that were carried out can be appreciated in Fig. 4, 5 and 6. Such an election was done due to the simplicity of this type of functions, that facilitates the validation and correction by the experts and are also easier to implement.

As it can be observed, the labels assigned to the different sets have been created in an abbreviated form, using the initials, with the purpose of facilitating their notation and understanding .

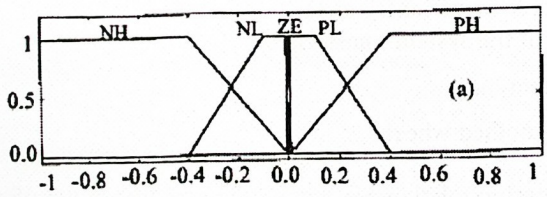


Fig. 4. Speed error

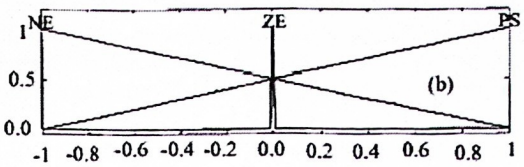


Fig. 5. Rate of change of the speed error

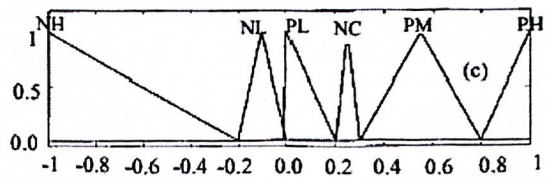


Fig. 6. The current of the axis q i_q

The fuzzy controller includes a certain number of rules that can usually be expressed by linguistic statements like:

$$\textit{If } x_1 \textit{ is } A_1^i \textit{ and } x_2 \textit{ is } A_2^i \textit{ and} \ldots \textit{and } x_{n_x} \textit{ is } A_{n_x}^i \textit{ then } z_1 \textit{is } Z_1^i, z_2 \textit{ is } Z_2^i, \ldots, z_{n_z} \textit{ is } Z_{n_z}^i \tag{11}$$

where A_j^i is the fuzzy set of the input j (n_x is the number of inputs) and Z_j^i is the fuzzy set of the output j (n_y is the number of outputs) both relative ones to the rule i.

For the group motor-conveyor the rules base is integrated by 7 rules that are shown next:

1) *If* $e\omega(k)$ is PH (positive high) *then* i_q is PH (positive high)
2) *If* $e\omega(k)$ is PL (positive low) *then* i_q is PM (positive medium)
3) *If* $e\omega(k)$ is ZE (zero) *and* $ce\omega(k)$ is PS (positive low) *then* i_q is PL (positive low)
4) *If* $e\omega(k)$ is ZE (zero) *and* $ce\omega(k)$ is NE (negative) *then* i_q is NC (no change)
5) *If* $e\omega(k)$ is ZE (zero) *and* $ce\omega(k)$ is ZE (zero) *then* i_q is NC (no change)
6) *If* $e\omega(k)$ is NL (negative low) *then* i_q is NL (negative low)
7) *If* $e\omega(k)$ is NH (negative high) *then* i_q is NH (negative high)

The membership functions, fuzzy sets for the input/output variables and the rules used in this work are selected by trial and error to obtain the optimum drive performance. Zadeh's compositional operator "sup-min" was used for the fuzzy reasoning and the

centroid method was used for the defuzzification. The output is obtained by means of the following expression:

$$Output = \frac{\sum_{i=1}^{n} x_1 \mu(x_i)}{\sum_{i=1}^{n} \mu(x_i)} \tag{12}$$

Where n is the total number of rules and $\mu(x_i)$ denotes the degree of membership for the rule *i*. This method is the most frequently used one and it also guarantees a soft behavior of the control.

4 Control Scheme. Proposed Configuration

The basic configuration of the drive consists of an induction motor fed by a current controlled voltage source inverter. The speed error and the rate of change of the speed error are processed by the fuzzy controller to generate the appropriate electromagnetic torque in the motor [16].

In order to obtain the system model for simulation, the configuration selected was the Current Controlled Voltage Inverter Source. In this case, the inverter feeds the motor with voltage and allows using the non compensated motor's equation. At the motor control schemes with compensation to variables decoupling, compensation's blocks are placed before coordinates transformations and the inverter is fed by voltage without motors current's feedback. These are feed backing to flux model.

Fig. 7. shows the simulation scheme developed for the validation experiments.

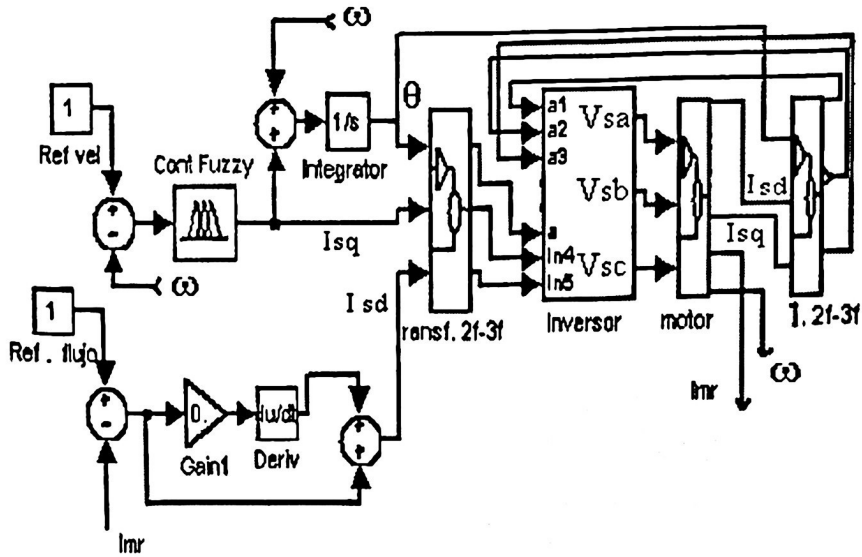


Fig. 7. Simulation scheme of the fuzzy control of an induction motor.

The reference currents of each phase are compared with the current values to carry out a control Bang-Bang for drive the power semiconductors of each branch of the inverter bridge to generate the voltage to feed the motor. The corresponding blocks are used to carry out the transformations of coordinates.

5 Simulation Experiments

To carry out the simulation experiments, the package SIMULINK®-MATLAB® was used as tool and the models that appear in epigraph 2 were taken to simulate the process. The fuzzy inference system was implemented by means of the fuzzy logic toolbox based in the structure shown in epigraph 3. The simulation scheme used can be seen in Fig. 7.

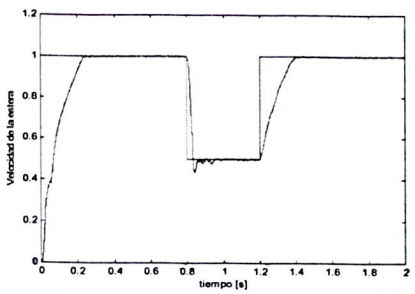


Fig. 8. Response to changes in the reference with fuzzy controller.

Motor power and belt conveyor parameters have been determinates for real conditions at the feeder belt conveyor drive in a Cuban sugar factory [14]. Simulations variable's results matches with experimental tests at referred drive conditions.

The following pictures show the response of the system to changes in the reference signal and in the load with PID controller and with fuzzy controller for comparison.

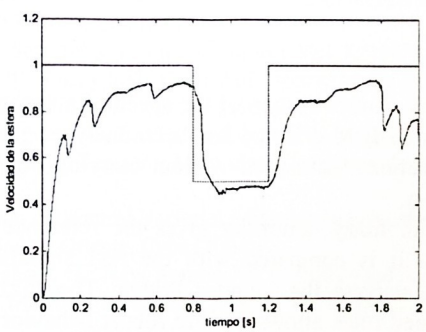


Fig. 9. Response to changes in the reference with PID controller.

The responses of the system to changes in the reference signal with fuzzy controller and with PID controller respectively are shown in Fig. 8 and Fig. 9. There you can appreciate how the response of the control system with fuzzy control is faster without presenting stationary state error. The response of the system with both types of regulators to disturbance in the load can be seen in Fig. 10 and Fig. 11. It can be appreciated that the fuzzy controller's response is better because the response is faster and more accurate. In all the cases the PI controller is tuned to give optimum response.

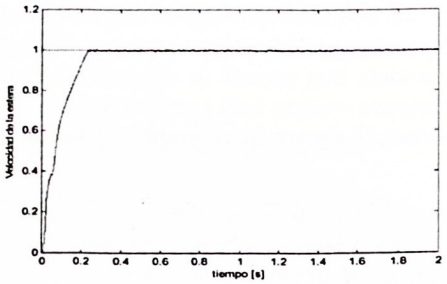


Fig. 10. Response to step in the reference signal and in the load disturbance with fuzzy controller.

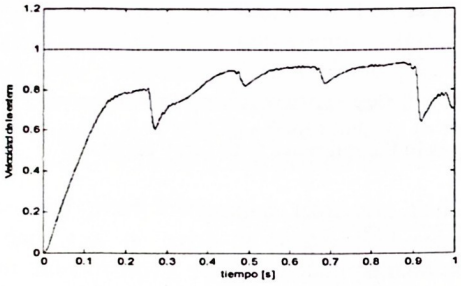


Fig. 11. Response to step in the reference signal and in the load disturbance with PID controller.

6 Conclusions

The behavior of a fuzzy control system of the speed of a sugar cane conveyor driven by an induction motor in field oriented has been analyzed. In the simulation of the motor the voltage equations were used without carrying out the decoupling of the variables in the axes d-q.

The speed loop with fuzzy controller gives the reference current signal to the bridge inverter where it is compared with the real signals of the motor. Their difference is used to activate the inverter bridge. The fuzzy controller designed according to the outlined rules allows a more robust behavior of the control system when facing variations of parameters and external disturbances that the classic PID controller.

The use of a classical PI algorithm for the control of the AC motor drives made evident the necessity of using the decoupling of the variables in the axes d-q, which is completely unnecessary when using the fuzzy controller. It is also necessary to keep in mind that there are nonlinearities in the model such as the hysteresis for drive the power semiconductors.

Due to the improvement of speed control this result makes it possible to minimize time waste in the process and to eliminate or minimize the continuous and much stressed starting process of the motor and then the process efficiency must be higher.

References

1. Blaschke, F.: The Principle of Field Orientation as Applied to The New Transvector Closed-Loop Control for Rotating Field Machines. Siemens Review, Vol. 34, No.3 (1972) 217-220
2. Bose, B. K.: High Performance Control of Induction Motor. Department of Electrical Engineering of University of Tennessee, Knoxvill. On line. (1998)
3. Vass, P.: Vector Control of AC Machines, Oxford University Press, Oxford (1993)
4. Sugimoto, H., Tamai, S.: Secondary Resistance Identification of an Induction Motor Applied Model Reference Adaptive System and its Characteristics. IEEE Trans. and Ind. Appl, Vol. 2, (1987) 296-303
5. Arellano, J., Asher, G., Sumner, M.: Robust Fuzzy-Sliding Mode Control for motor Drives Operating with Variable Loads and Pre-defined System Noise Limits. School of Electrical and Electronic Engineering. University of Nottingham, Nottingham, England (2001)
6. Uddin, M.N., Radwan, T. S., Radhman, M.A.: Performances of Novel Fuzzy Logic Based Indirect Vector Control for Induction Motor Drive. Faculty of Enginneering and Applied Science, Memorial University of Newfoundland, St, John's, NF, Canada. On line. (2003)
7. Spiegel, M., Turner, M., McCormick, V.: Fuzzy logic based controllers for efficiency optimization of inverter fed induction motor drives. IEEE Transaction on Power Electronic, Vol. 137 No. 3. (2003) 387-401
8. Lee J.: On methods for improving performance of PI- type fuzzy logic controllers. IEEE Transactions of Fuzzy Systems, Vol. 1. No 4. (2000) 298-301
9. Lei S., Langari R.: Synthesis and approximation of fuzzy logic controllers for nonlinear system. International journal of Fuzzy Logic, Vol. 5. No 2 (2003) 98-104
10. Grabouwski, P.Z., Kazmierkouski, M.P., Bose, B.K., Blaabjeng, F.: A Simple Direct-Torque Nreuro-fuzzy Control of PWM-Inverter-Fed Induction Motor Drive. IEEE Trans. Ind. Electron. Vol. 47, No. 4 (2000) 863-870
11. Zadeh, L. A.: Fuzzy Sets, Information and Control, Vol. 8 (1965) 338-353
12. Cleland, J., Tumer, W.: Fuzzy logic Control of Electric Motor and Motor Drives Feasibility Study. EPA/600/SR-95/175 April. On-line (1996)
13. Angulo, M., Domínguez, A., Ferrán, J.: Control Vectorial de un Motor de Asincrónico Alimentado con Fuente de Tensión. Memorias II congreso de Automática. Bucaramanga Colombia (1997)
14. Cobas, F.: Simulación del control Vectorial y Escalar en una estera transportadora. Tesis de Maestría. Universidad de Oriente, Cuba (2004)
15. Margaliot, M., Langholz, G.: New Approaches to fuzzy modeling and control: design and analysis. Series in Machine Perception Articial Intelligence, World Scientic, Singapore (2000)
16. Eid B., Ehab H.: Integrated IGBT PWM converter/inverter system feeding three phase induction motors. Doctoral thesis. Cairo University. (2001)

Prefiltering and Robustness of GPC with Structured Perturbations

C. Mañoso, A. P. de Madrid, R. Hernández, M. Romero

Dpto. de Sistemas de Comunicación y Control - UNED
Juan del Rosal, 16. 28040 – Madrid (Spain)
carolina@scc.uned.es

Abstract. This paper deals with prefiltering for improving the robust stability of Generalized Predictive Control (GPC) against structured perturbations. Some guidelines for choosing an optimal value of the prefilter T have been widely accepted in the literature. However, this work shows how these guidelines are not always adequate and can even worsen robustness. Thus, once a given value of the prefilter T has been chosen a deep robustness analysis is needed.

Keywords: Model Based Predictive Control, GPC, Prefitering, Structured Perturbations, Stability Analysis, Robustness Analysis.

1 Introduction

All predictive controllers share a common methodology: at each "present" instant t, future process outputs $y(t+k \mid t)$ are predicted for a certain time window, $k = 1, 2, \ldots, N$, using a model of the process. The optimal control law is obtained by minimizing a cost function (only the unconstrained case is considered):

$$J(\Delta u,t) = E\left\{\sum_{j=N_1}^{N_2}\gamma(j)\left[r(t+j\mid t)-y(t+j\mid t)\right]^2+\sum_{j=1}^{N_u}\lambda(j)\left[\Delta u(t+j-1\mid t)\right]^2\right\} \quad (1)$$

where $E\{\cdot\}$ is the expectation operator, N_1 is the minimum costing horizon, N_2 is the maximum costing horizon, N_u is the control horizon, γ is a future errors weighting sequence, and λ is a control weighting sequence.

Generalized Predictive Control (GPC) [1, 2] is one of the most representative predictive controllers. GPC assumes a CARIMA model (a transfer function plus a colored and integrated white noise) to describe the system dynamics:

$$A(z^{-1})y(t) = B(z^{-1})u(t)+\frac{T(z^{-1})}{\Delta}\xi(t) \quad (2)$$

where Δ is the increment operator and $\xi(t)$ represents uncorrelated zero-mean white noise. $T(z^{-1})$ is a polynomial which implements a prefilter. In practice, T is not considered a model parameter but a controller parameter, as its value is chosen to

© L. Sánchez, O. Pogrebnyak and E. Rubio (Eds.)
Industrial Informatics
Research in Computing Science 31, 2007, pp. 85-94

improve the closed loop robustness. In fact, prefiltering is the most popular approach to robust design in GPC.

The effects of prefiltering on robustness were initially considered by Robinson and Clarke [3], who gave some guidelines for selecting T. However, these results are applied only for mean-level or dead-beat GPC. Soeterboek [4] pointed out that robustness would be enhanced by choosing $T = A(1-\alpha z^{-1})$, as α increases. Megías *et al.* [5] showed that α cannot be increased unlimitedly to improve robustness because performance would deteriorate. Yoon and Clarke [6] extended these works, suggesting that a more general guideline, $T = A(1-\alpha z^{-1})^{N_1}$, improves the robust stability because it ensures the presence of a low-pass filter in the control loop that rejects the high frequency unmodeled responses. This recommendation is accepted in the literature widely and deserves some further explanations.

GPC control structure can be expressed in a classical LTI form (Fig. 1). From it, it is easy to derive the expression of the closed loop characteristic equation:

$$\delta_0(z^{-1}) = RA_0\Delta + BSz^{-1} \tag{3}$$

where A_0 and B_0 represent the transfer function of the actual plant (generally different to A and B, the model transfer function used to design the GPC controller), R and S are the following polynomials

$$R = \frac{T + \sum_{i=N_1}^{N_2} k_{1i} H_i}{\sum_{i=N_1}^{N_2} k_{1i} q^{-N_2+k}}, \qquad S = \frac{\sum_{i=N_1}^{N_2} k_{1i} F_i}{\sum_{i=N_1}^{N_2} k_{1i} q^{N_1+k}} \tag{4}$$

and H_i and F_i are polynomials that can be derived from some diofantine equations [1].

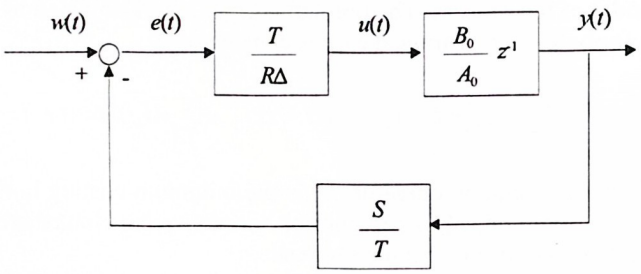


Fig. 1. GPC structure.

Assuming unstructured uncertainties and using the Small Gain Theorem, Yoon and Clarke [6] concluded that it is required that ST be a low-pass filter and therefore $T \neq 1$ should be used, otherwise ST $(= S)$ would be a high-pass filter. According to this, they proposed a "natural" choice for T:

$$T = A(1-\alpha z^{-1})^{N_1} \tag{5}$$

where α lies in the neighborhood of the dominant root of A.

The main handicap of this approach is that it has been shown that (5) is fairly simplistic. For this reason, deep analysis for each given case is needed. Sensitivity analysis (to multiplicative uncertainty, to disturbances, and to noise) has been proposed by Rossiter [7]. In practice, this analysis shows that the effect of T is a trade-off between disturbances or noise rejection and robustness to model parameter uncertainty at different frequency bands, in a way that is not always beneficial. If the chosen T-filter does not achieve the desired sensitivities, it may not be obvious how to redesign it to improve matters.

Following this line of argument, in this paper we will study the influence of T on robustness when structured perturbations are considered using tools applicable to polytopes of polynomials. It will be shown how (5) can lead to poorer robustness when these perturbations are present and how this fact can be depicted in geometrical terms.

This paper has been structured as follows: In Section 2 the GPC control of plants with structured uncertainties is introduced. Section 3 shows by examples how the general guideline for choosing T is not always adequate. Finally, section 4 draws the main conclusions of this paper.

2 GPC with Structured Uncertainties

Structured perturbations mean that the uncertainties are in the coefficients, *i.e.* the numerator and the denominator of the actual plant are given by uncertain polynomials. Affine linear uncertainty structures will be considered. Thus, given a set of real parameters $q_i, i = 0,\ldots,l$, which can vary between a maximum and a minimum value, $q_i^- \le q_i \le q_i^+, i = 0,\ldots,l$, the coefficients of the numerator and denominator polynomials are affine linear functions of the uncertainty parameter vector $\mathbf{q} = (q_0,\ldots,q_l) \in \Re^l$, that is $a_i(\mathbf{q}) = \alpha_i^T \mathbf{q} + \beta_i, i = 0,\ldots,m$, $b_i(\mathbf{q}) = \gamma_i^T \mathbf{q} + \rho_i, i = 0,\ldots,n$ where α_i and γ_i are $1 \times l$ vectors and β_i and ρ_i are scalars. Then, the actual plant is defined by the family of plants:

$$G_0(\mathbf{q}, z^{-1}) = \frac{B_0(\mathbf{q}, z^{-1})}{A_0(\mathbf{q}, z^{-1})} = \frac{\sum_{i=0}^{n} b_i^0(\mathbf{q}) z^{-i}}{1 + \sum_{i=1}^{m} a_i^0(\mathbf{q}) z^{-i}} \tag{6}$$

where ($m \ge n \ge 0$). With this structure of uncertainties, $A_0(\mathbf{q}, z^{-1})$ and $B_0(\mathbf{q}, z^{-1})$ *are polytopes of polynomials in* z^{-1}, and $G_0(\mathbf{q}, z^{-1})$ *is a polytope of plants in* z^{-1}.

It has been shown that the family of characteristic polynomials $\delta_0(z^{-1})$ of the closed loop system constituted by a GPC controller (4) and the family of plants (6) is a polytope of polynomials [8, 9]:

$$\delta_0(z^{-1}) = R(z^{-1}) A_0(\mathbf{q}, z^{-1}) \Delta + S(z^{-1}) B_0(\mathbf{q}, z^{-1}) z^{-1} \tag{7}$$

This fact allows the use of a very mature theory from the point of view of the robust stability analysis. Nowadays it can be said that there are powerful results to analyze the stability and the robust performance of families of polynomials formed by interval polynomials or by polytopes of polynomials. The main tools for the analysis of polynomial families are Kharitonov Theorem [10] for interval polynomials and the Edge Theorem [11] and Rantzer Theorem [12] for polytopes of polynomials.

In order to determine which T leads to the best robustness when structured uncertainties are present, we propose the analysis of the stability region in the parameter space derived from the closed loop characteristic equation. The method of Ackermann [13] will be used to draw this region because it has low computational cost and no conservatism. The stability hypersphere around the nominal process will be also analyzed.

In the following two examples will illustrate how guideline (5) influences robustness in the presence of structured uncertainties.

3 Examples

3.1 Example 1

Let us revisit an example proposed by Yoon and Clarke [6],

$$G = \frac{B}{A} = \frac{0.2}{1 - 0.8z^{-1}} \tag{8}$$

and let us assume that the plant is actually represented by the following family (interval) of plants:

$$G_0 = \frac{0.2}{1 + (-0.8 + \alpha_1)z^{-1} + \alpha_2 z^{-2}} \tag{9}$$

with α_1 and α_2 the uncertainty parameters. The GPC controller is tuned with the following predictive control settings: $N_1 = 1$, $N_u = 2$, $N_2 = 5$ and $\lambda = 0.01$.

Case $T = 1$

In the absence of prefiltering (*i.e.* $T = 1$) the polynomials R and S (4) are the following:

$$\begin{aligned} R(z^{-1}) &= 0.2624 \\ S(z^{-1}) &= 1.9231 - 0.9231z^{-1} \end{aligned} \tag{10}$$

and the closed loop characteristic equation is:

$$\begin{aligned}\delta_0(z,\alpha_1,\alpha_2) &= 0.2624z^3 + (-0.0877 + 0.2624\alpha_1)z^2 + \\ &+(0.0253 - 0.2624\alpha_1 + 0.2624\alpha_2)z + \\ &-0.2624\alpha_2\end{aligned} \tag{11}$$

Case T=A(1− 0.8z^{-1})

Now we will follow the standard guideline (5): $T = A(1 - \alpha z^{-1})^{N_1}$, with $\alpha \in (0,1)$. It is extensively accepted that the robustness will be better when α lies in the neighborhood of the dominant root of A. Therefore, we will take $\alpha = 0.8$ as it was proposed in [6]. Thus R and S are

$$\begin{aligned}R(z^{-1}) &= 0.2624 - 0.0752z^{-1} + 0.0202z^{-2} \\ S(z^{-1}) &= 0.2 - 0.16z^{-1}\end{aligned} \tag{12}$$

and

$$\begin{aligned}\delta_0(z,\alpha_1,\alpha_2) &= 0.2624z^5 + (-0.5075 + 0.2624\alpha_1)z^4 + \\ &+(0.3335 - 0.3376\alpha_1 + 0.2624\alpha_2)z^3 + \\ &+(-0.09652 + 0.0954\alpha_1 - 0.3376\alpha_2)z^2 + \\ &+(0.01616 - 0.0202\alpha_1 + 0.0954\alpha_2)z - 0.0202\alpha_2\end{aligned} \tag{13}$$

The families of polynomials (11) and (13) are polytopes as it was stated above (7). Fig. 2 shows their stability regions.

The area of the stability region for $T = 1$ is smaller than the one that follows the recommendation (5). Therefore, it could be concluded that the robust stability, in terms of the associated stability areas, has been improved with prefiltering.

However, the radius of the stability hypersphere (in this case just a circle) around the nominal process is smaller with prefiltering. Even though the stability region is bigger, from this point of view prefiltering following (5) deteriorates the robustness of the closed loop system.

3.2 Example 2

Now one of the examples proposed by Rossiter [7] is considered,

$$G = \frac{1}{1 - 1.4z^{-1} + 0.45z^{-2}} \tag{14}$$

assuming that the actual perturbations are structured and therefore the actual plant is given by a family (interval) of plants

$$G_0 = \frac{1}{1 + (-1.40 + \alpha_1)z^{-1} + (0.45 + \alpha_2)z^{-2}} \tag{15}$$

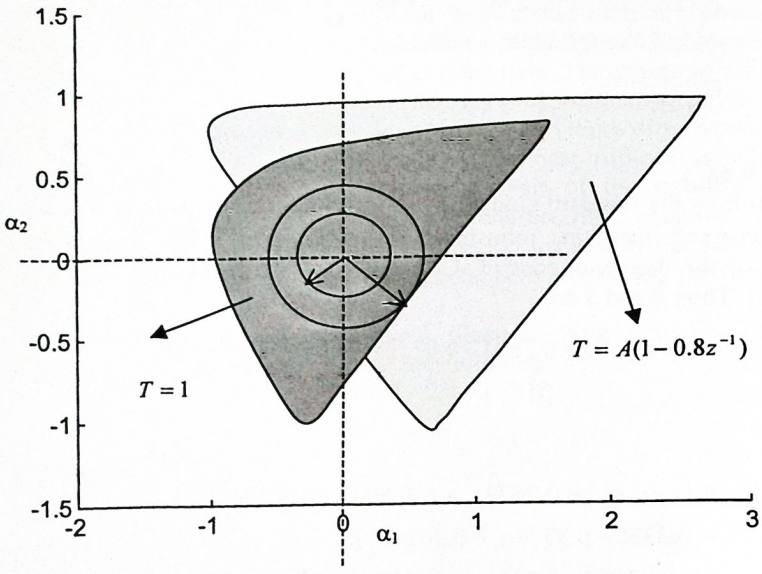


Fig. 2. Example 1. Stability regions for $T = 1$ and $T \neq A\ (1 - 0.8z^{-1})$.

with α_1 and α_2 the uncertainty parameters. The GPC controller has been tuned with $N_1 = 1, N_u = 1, N_2 = 10$ and $\lambda = 0.01$.

This example illustrates the situations considered in [7]: No prefiltering (*i.e.* $T = 1$, the situation rejected by Yoon and Clarke [6]) and T following the general recommendation $(1-0.8z^{-1})^n$, $n = 1$ or 2. (This choice of α is intuitive in that if sampling at about 1/10 of the rise time, a common guideline for predictive control, then a typical dominant process pole would be around 0.8. Hence this is a sensible pole for a low-pass filter on output measurements.)

For the nominal system, the sensitivity analysis performed in [7] concludes that the inclusion of a T-filter has given good reductions in every sensitive function over the high frequency: the sensitivity function to multiplicative uncertainty is actually better over the whole frequency range, the output sensitivity is worse at mid and low frequencies and better at high frequencies, and the input sensitivity is better over the whole frequency range.

Now let us consider the effect of T-filter on structured uncertainties.

Case $T = 1$

The controller is given by the following expressions

$$R(z^{-1}) = 8.9245$$
$$S(z^{-1}) = 9.9977 - 13.0137z^{-1} + 4.0159z^{-2} \tag{16}$$

and the closed loop characteristic equation is

$$\begin{aligned}\delta_0(z,\alpha_1,\alpha_2) &= 8.9245z^3 + \\ &+(-11.421+8.9245\alpha_1)z^2 + \\ &+(3.4963-8.9245\alpha_1+8.9245\alpha_2)z + \\ &-8.9246\alpha_2 - 0.000125\end{aligned} \tag{17}$$

Case $T = (1 - 0.8z^{-1})$
Now we obtain

$$\begin{aligned}R(z^{-1}) &= 8.9245 - 0.0001z^{-1} \\ S(z^{-1}) &= 2.8583 - 3.8772z^{-1} + 1.2189z^{-2}\end{aligned} \tag{18}$$

and

$$\begin{aligned}\delta_0(z,\alpha_1,\alpha_2) &= 8.9245z^4 + \\ &+(-18.561+8.9245\alpha_1)z^3 + \\ &+(12.633-8.9246\alpha_1+8.9245\alpha_2)z^2 + \\ &+(-2.7973-0.0001\alpha_1-8.9246\alpha_2)z + \\ &-0.0001\alpha_2 + 0.000045\end{aligned} \tag{19}$$

Case $T = (1- 0.8z^{-1})^2$

$$\begin{aligned}R(z^{-1}) &= 8.9245 - 4.9728z^{-1} + 0.0001z^{-2} \\ S(z^{-1}) &= 0.6914 - 0.9632z^{-1} + 0.3118z^{-2}\end{aligned} \tag{20}$$

$$\begin{aligned}\delta_0(z,\alpha_1,\alpha_2) &= 8.9245z^5 + (-25.7+8.9245\alpha_1)z^4 + \\ &+(27.482+13.897\alpha_1+8.9245\alpha_2)z^3 + \\ &+(-12.904+4.9729\alpha_1-0.9632\alpha_2)z^2 + \\ &+(2.2379-0.0001\alpha_1+4.9729\alpha_2)z + \\ &-0.0001\alpha_2 - 0.000045\end{aligned} \tag{21}$$

Fig. 3 shows the stability regions of the polytopes (17), (19), and (21). The stability region for T=1 is smaller than the ones that follow the recommendation (5). However, as in the previous example the stability hypersphere around (0, 0) –absence of uncertainties– is bigger when there is no prefiltering ($T = 1$). Therefore, in this sense robustness is not improved by the guideline.

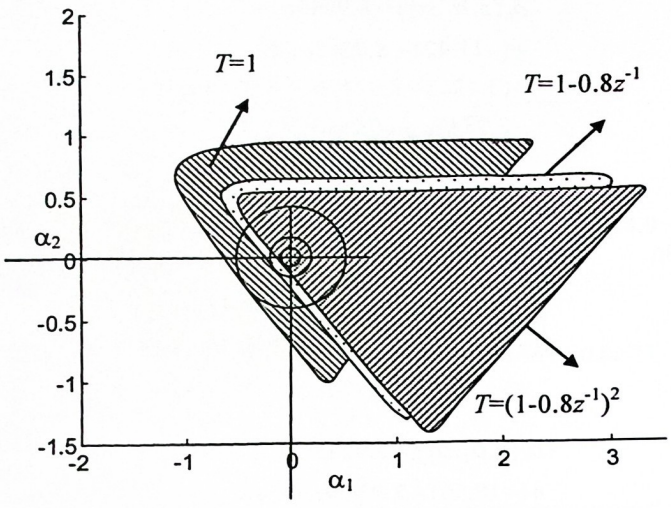


Fig. 3. Example 2. Stability regions for $T = 1$, $T = (1 - 0.8z^{-1})$,and $T = (1 - 0.8z^{-1})^2$.

4 Conclusions

This paper has focused on the study of the influence of prefilter T on the robustness of GPC against structured uncertainties. Given that the closed loop is a polytope of polynomials, it is possible to analyze the robust stability with tools based on polytopes with ease and no conservatism.

Some examples have shown that the widely accepted guideline for choosing T does not guarantee better robustness. In fact, this has been shown with simple geometrical measurements, such as the area of the stability region or the radius of the stability hypersphere in the uncertainties space.

Examples have shown that there exist "directions" in the uncertainties space where the robust stability margins are better and "directions" where they are worse. This situation is somehow similar to the sensitivity analysis when unstructured disturbances are considered, where there could be frequency bands where sensitivity is improved and bands where it is worsened.

For these reasons, no matter the type of disturbances that are present, once a given value of the prefilter T has been chosen a deep analysis of robustness (sensitivity functions, stability regions, etc.) is needed.

Acknowledgments. This work was supported by the Spanish *Ministerio de Educación y Ciencia*, under project ref. DPI2004-05903.

References

1. Clarke D. W., Mohtadi, C., Tuffs, P. S.: "Generalized Predictive Control - Part I. The Basic Algorithm, Part II. Extensions and Interpretations", *Automatica*, **23** (1987), 137-160.
2. Bitmead R. R., Gevers, M., Wertz, V.: *Adaptive Optimal Control: The Thinking Man's GPC.* Prentice Hall International (1990).
3. Robinson, B. D., Clarke, D. W.: "Robustness Effects of a Prefilter in Generalised Predictive Control", *IEE Proc. D*, **138** (1991), 2-8.
4. Soeterboek, R.: *Predictive Control. A unified Approach*, Prentice Hall (1992).
5. Megías, D., Serrano J., de Prada, C.: "Uncertainty Treatment in GPC: Design of T polynomial", *European Control Conference* (1997).
6. Yoon, T. W, Clarke, D.W.: "Observer Design in Receding-Horizon Predictive Control", *Int. J. Control*, **61** (1995), 171-191.
7. Rossiter, J. A.: *Model Based Predictive Control, A practical Approach*. CRC Press (2003).
8. Mañoso, C., Hernández, R., de Madrid, A. P., Dormido, S.: "Robust stability analysis of predictive controllers using extreme point results", *Proc CESA '96 IMACS Multiconference*, **1** (1996), 483-488.
9. Mañoso, C., Hernández, R., de Madrid, A. P., Dormido, S.: "Robust Stability analysis of GPC: an Application to Dead-Beat and Mean-Level Predictive Controllers", *Proc. Mediterranean IEEE Conference*. CD-ROM Proceedings (1997) , paper 78.
10. Kharitonov, V. L.: "Asymptotic stability of an equilibrium position of a family of systems of linear differential equations", *Differential'nye Uraveniya*, **14** (1978), 1483-1485.
11. Bartlett, A. C., Hollot, C.V., Lin, H.: "Root locations of an entire polytope of polynomials: it suffices to check the edges", *Mathematics of Control, Signals and Systems*, **1** (1988), 61-71.
12. Rantzer, A.: "Stability Conditions for Polytopes of Polynomials", *Transactions on Automatic Control*, **3**(7) (1992), 79-89.
13. Ackermann, J. E.: "Parameter Space Design of Robust Control Systems", *IEEE Transactions on Automatic Control*, **25**(6) (1980), 1058-1072.

Industrial Temperature Control using an AFLC

Nitin Jagannath Patil[I], Dr. R. H. Chile[II], Dr. L.M. Wagmare[III]

[I]D.N. Patel College of Engineering, Shahada, India,
nitinpatil_2002@yahoo.com,
[II]S.G.G.S.Institute of Engineering & Technology, Nanded, India,
rhchile@yahoo.com,
[III]S.G.G.S.Institute of Engineering & Technology, Nanded, India,
lmwaghmare@yahoo.com.

Abstract. A closed loop control system incorporating adaptive fuzzy logic has been developed for a class of industrial temperature control problems. A unique fuzzy logic controller (FLC) structure with an efficient realization and a small rule base that can be easily implemented in existing industrial controllers was proposed. It was demonstrated in both software simulation and hardware test in an industrial setting that the fuzzy logic control is much more capable than the current temperature controllers. This includes compensating for thermo mass changes in the system, dealing with unknown and variable delays, operating at very different temperature set points without retuning, etc. Also for auto-tuning the FLC adaptation mechanism is included in one of the input to fuzzy logic controller. It is achieved by implementing, in FLC, a classical control strategy and an adaptation mechanism to compensate for the dynamic changes in the system. The proposed FLC was applied to a temperature process and a significant improvement in the system performance is observed.

Keywords: Adaptive, Fuzzy Control, Temperature Regulation.

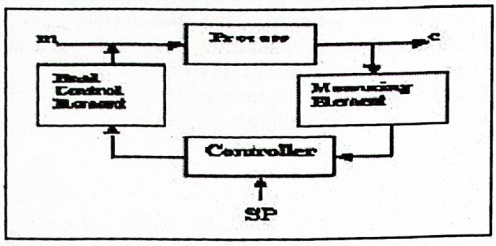
1 Introduction

While modern control theory has made modest inroad into practice, fuzzy logic control has been rapidly gaining popularity among practicing engineers. This increased popularity can be attributed to the fact that fuzzy logic provides a powerful vehicle that allows engineers to incorporate human reasoning in the control algorithm. As opposed to the modern control theory, fuzzy logic design is not based on the mathematical model of the process. The controller designed using fuzzy logic implements human reasoning that has been programmed into fuzzy logic language (membership functions, rules and the rule interpretation).

It is interesting to note that the success of fuzzy logic control is largely due to the awareness to its many industrial applications. An industrial interest in fuzzy logic control as evidenced by the many publications on the subject in the control literature has created an awareness of its increasing importance by the academic community. Starting in the early 90s, the Applied Control Research Lab. at Cleveland State University, supported by industry partners, initiated a research program investigating the role of fuzzy logic in industrial control. The primary question at the time was:

© L. Sánchez, O. Pogrebnyak and E. Rubio (Eds.)
Industrial Informatics
Research in Computing Science 31, 2007, pp. 95-104

"What the fuzzy logic control does that the conventional control can not do?" The research results over the last few years have been reported in [1-5]. In this paper, we concentrate on fuzzy logic control as an alternative control strategy to the current PID method used widely in industry. Consider a generic temperature control application shown in Fig. 1:



m- Manipulated Variable, c- Controlled Variable, SP – Set point

Fig. 1: A typical Industrial Temperature Control problem

The temperature is measured by a suitable sensor such as Thermocouples, RTD, Thermistor, etc. and converted to a signal acceptable to the controller. The controller compares the temperature signal to the desired set point temperature and actuates the control element. The control element alters the manipulated variable to change the quantity of heat being added to or taken from the process. The objective of the controller is to regulate the temperature as close as possible to the set point. To test the new fuzzy logic control algorithms, two temperature regulation processes were used in this research.

One uses hot and cold water as manipulated variable and a valve as the controller element, the other uses electricity as a power source to a heater, actuated by a Solid State Relay (SSR). The new algorithms were tested extensively in both simulation and the hardware tests.

1.1 Motivation

Currently, the classical PID (Proportional, Integral and Derivative) control is widely used with its gains manually tuned based on the thermal mass and the temperature set point. Equipment with large thermal capacities requires different PID gains than equipment with small thermal capacities. In addition, equipment operation over wide ranges of temperatures (140° to 500°), for example, requires different gains at the lower and higher end of the temperature range to avoid overshoots and oscillation. This is necessary since even brief temperature overshoots, for example, can initiate nuisance alarms and costly shut downs to the process being controlled. Generally, tuning the Proportional, Integral, and Derivative constants for a large temperature control process is costly and time consuming. The task is further complicated when incorrect PID constants are sometimes entered due to the lack of understanding of the temperature control process. The difficulty in dealing with such problems is compounded with variable time delays existed in many such systems. Variations in manufacturing, new product development and physical constraints place the RTD

temperature sensor at different locations, inducing variable time delays (dead time) in the system.

It is also well known that PID controllers exhibit poor performance when applied to systems containing unknown non-linearity such as dead zones saturation and hysteresis. It is further understood that many temperature control processes are nonlinear. Equal increments of heat input, for example, do not necessarily produce equal increments in temperature rise in many processes, a typical phenomenon of nonlinear systems. The complexity of these problems and the difficulties in implementing conventional controllers to eliminate variations in PID tuning motivate us to investigate intelligent control techniques such as fuzzy logic as a solution to controlling systems in which time delays, non-linearities, and manual tuning procedures need to be addressed.

1.2 The Time Delay Problem and Existing Solutions

To study the temperature control problem using classical control techniques, a simplified block diagram, in Fig. 2, is used, where C(s) represents the controller and $G(s)e^{-s\tau}$ the plant with a pure time delay of τ. It is well known that the time delay makes the temperature loops hard to tune. The time delay problem may be characterized by *large* and *small* delays. A linear time invariant system with finite delay τ can be modeled as $G(s)\,e^{-s\tau}$ where $G(s)$ is a rational transfer function of s. Note that the delay corresponds to a phase shift of $-w\tau$, where ω denotes the frequency. Small phase shifts at frequencies of interest may be viewed as perturbations and incorporated into a delay free design with sufficient phase margin. A large delay is classified as a delay that significantly affects the stability and phase margins to the point that delay free design methods will not be sufficient.

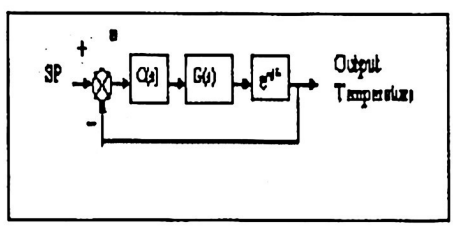


Fig. 2: A Closed Loop Temperature Control System

A number of time delay compensation and prediction schemes have been developed and/or improved with modifications as shown in [7-12]. The performance of Smith Predictor Control (SPC) was studied experimentally in [8]. It shows that the system performs well if the process model is accurate, but that performance degrades rapidly with inaccuracy in the process parameters and time delay. Clearly for an unknown or variable time delay, Smith predictive compensation is no longer a viable technique. Several control design methods for systems with varying time delays have appeared in recent literature including an estimation and self-tuning method proposed by Brone and Harris [10], a variable structure controller by Shu and Yan [11], and a model reference adaptive approach by Liu and Wang [6], to name a few. For systems

with large time delays, most design approaches use a prediction mechanism as part of the controller to simulate the process for given system parameters and time delay. In the well known Smith predictor [7], the controller output is fed through models of the process with delay, and the process without delay, respectively. The difference of the output signals is added to the actual plant output and then fed back to the controller, thus allowing the controller to act on the prediction of the plant output.

1.3 Fuzzy Logic Control

Fuzzy control is an appealing alternative to conventional control methods when systems follow some general operating characteristics and a detailed process understanding is unknown or traditional system models become overly complex [6]. The capability to qualitatively capture the attributes of a control system based on observable phenomena is a main feature of fuzzy control. These aspects of fuzzy control have been demonstrated in various research literature, see [13-15, 16,17] and commercial products from vendors like Reliance Electric and Omron. The ability of fuzzy logic to capture system dynamics qualitatively, and execute this qualitative idea in a real time situation is an attractive feature for temperature control systems. The analytical study of fuzzy logic is still trailing its implementation and much work is still ahead, particularly in the area of stability and performance analysis. Furthermore, as solutions to practical problems, fuzzy logic control design is problem dependent and the adaptation of an exiting fuzzy logic controller to a different control problem is not straightforward. The available design tools, such as the Fuzzy Toolbox provided by Mathworks Inc., generally require further improvements before they become acceptable to control engineers. In this paper, the validity of fuzzy logic control as an alternative approach in temperature control applications is investigated.

2 Adaptive Fuzzy Logic Control Design

The FLC developed here is a two-input single-output controller. The two inputs are the deviation from set point *error, e(k)*, and *error rate, $\Delta e(k)$*. The FLC is implemented in a discrete-time form using a zero-order-hold as shown in Fig. 3a. The operational structure of the Fuzzy controller is shown in Fig. 3b.

2.1 Fuzzification / Defuzzification

Fuzzification and defuzzification involve mapping the fuzzy variables of interest to "crisp" numbers used by the control system. Fuzzification translates a numeric value for the error, *e(k)*, or error rate, *De(k)*, into a linguistic value such as positive large with a membership grade. Defuzzification takes the fuzzy output of the rules and generates a "crisp" numeric value used as the control input to the plant.
The FLC membership functions are defined over the range of input and output variable values and linguistically describes the variable's universe of discourse as shown in Figure 4. The triangular input membership functions for the linguistic labels

zero, small, medium, and large, had their membership tuning center values at 0, 0.2, 0.35, and 0.6, respectively. The universe of discourse for both *e* and *De* is normalized from -1 to 1. The left and right half of the triangle membership functions for each linguistic label was chosen to provide membership overlap with adjacent membership functions. The straight line output membership functions for the labels zero, small, medium, and large are defined as shown in Fig. 4 with end points corresponding to 10, 30, 70, and 100% of the maximum output, respectively. Both the input and output variables membership functions are symmetric with respect to the origin.

Selection of the number of membership functions and their initial values is based on process knowledge and intuition. The main idea is to define partitions over the plant operating regions that will adequately represent the process variables.

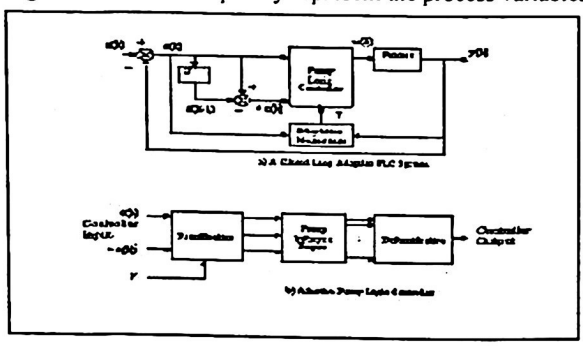


Fig. 3: Adaptive Fuzzy Control System

2.2 Rule Development

Our rule development strategy for systems with time delay is to regulate the overall loop gain to achieve a desired step response. The output of the FLC is based on the current input, *e(k)* and *De(k)*, and adaptation mechanism output γ without any knowledge of the previous input and output data or any form of model predictor. The main idea is that if the FLC is not designed with specific knowledge of mathematical model of the plant, it will not be dependent on it. The rules developed in this paper are able to compensate for varying time delays on-line by tuning the FLC output membership functions based on system performance. The FLC's rules are developed based on the understanding of how a conventional controller works for a system with a fixed time delay. The rules are separated into two layers: the first layer of FLC rules mimics what a simple PID controller would do when the time delay is fixed and known; the second rule layer deals with the problem when the time delay is unknown and varying.

In developing the first layer rules, consider the first order plant, $G(s)e^{-s\tau}$ where $G(s)\dfrac{a}{s+a}$. In the PID design, the following assumptions are made:

- The time delay τ is known

- The rise time, t_r, or equivalently, the location of the pole is known.
- t_r is significantly smaller than τ
- The sampling interval is T_s

The conventional PI-type controller in incremental form is given by:

$$u(k) = u(k-1) + f(e, \Delta e(k)) \qquad (1)$$

where *f(e,Δe)* is computed by a discrete-time PI algorithm. This control algorithm was applied to a first order plant with delay. Using the Ziegler-Nichols method carried out initial tuning of PI parameters. The step response obtained has about a 20% overshoot for a fixed time delay. Next a fuzzy logic control law was set up where *F(e,Δe)*, the output of the FLC for the k_{th} sampling interval, replaces *f(e,Δe)* in the incremental controller described in (1). The rules and membership functions of the FLC were developed using an intuitive understanding of what a PI controller does for a fixed delay on a first order system. They generalized what a PI controller does for each combination of *e* and Δe in 12 rules as shown in Table 1.

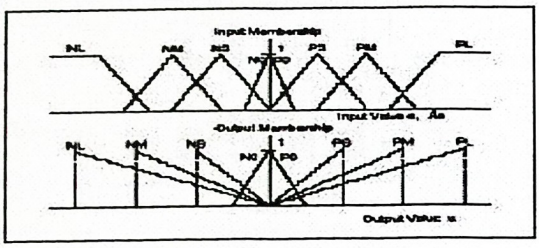


Fig. 4: Fuzzy Membership Functions

Table 1: FLC Control Rules

		Δe							
		NL	NM	NS	NO	PO	PS	PM	PL
e	NL	NL							
	NM	NM					PS		
	NS	X	X	X	NM	PS	X	X	X
	NO	NO							
	PO	PO							
	PS	X	X	X	NS	PM	X	X	X
	PM	NS			PM				
	PL	PL							

X – No Control Action

The output from each rule can be treated as a fuzzy singleton. The FLC control action is the combination of the output of each rule using the weighted average defuzzification method and can be viewed as the center of gravity of the fuzzy set of output singletons.

2.3 Tuning Membership Functions in Design Stage

Since there is little established theoretical guidance, the tuning of rules and membership functions in the design stage is largely an iterative process based on

intuition. The number of membership functions can vary to provide the resolution needed. Note that the number of rules can grow exponentially as the number of input membership functions increases. The input membership functions for e and Δe generate 64 combinations, which can be grouped into twelve regions corresponding to each rule in Table 1. The center and slopes of the input membership functions in each region is adjusted so that the corresponding rule provides an appropriate control action. In case when two or more rules are fired at the same time, the dominant rule, that is the rule corresponding to the high membership grade, is tuned first. Modifying the output membership function adjusts the rules contribution relative to the output universe of discourse. Once input membership rule tuning is completed, fine-tuning of the output membership functions is performed to achieve the desired performance. Although this FLC is constructed based on the assumption that the time delay is fixed and known, the only element of the controller that is a function of the delay is the universe of discourse for the output. It is shown below that with some adjustment and extra rules, the FLC can be made to adapt to an unknown nature or change in delay.

2.4 Adaptation Mechanism

The FLC structure presented above can be directly modified to compensate for changes in the plant dynamics and variable time delays by adding a second layer of adaptation mechanism as one of the input to fuzzy inference engine of FLC. More details on adaptive methods can be found in [1,2].
In the case of varying time delay, the FLC gain must be adjusted to offset the effects of the changes in delay. It will be shown in Section IV that the maximum gain or control action is inversely proportional to the time delay. Therefore, if the delay increases, we should decrease the FLC gain to reduce the control action, and vice versa. Based on this relationship, the system performance can be monitored by a second layer of rules that adapts the output membership functions of the first layer of rules to improve the performance of the fuzzy controller.

The design strategy for the second layer of rules is based on two different aspects of tracking performance, i.e., rise time and overshoot calculated from $(e,\Delta e)$. The second layer rules are listed in Table 2. They monitor the plant response and reduce or increase the FLC controller output universe of discourse. The fuzzy membership functions are defined using a membership configuration similar to the control strategy in Figure 3. The adjustment rules perform two actions; they reduce the FLC gain when the plant is significantly overshooting the desired response, and increase the gain when rise time performance is slow.

Fuzzy logic control temperature control scheme is further tested in an industrial application where several components in a machine have to be temperature regulated. These components are of different thermo mass and may be regulated at different temperatures. Currently, a separate PID controller is tuned for each component at each temperature set point, which is quite labor intensive. Furthermore, the PID parameters need frequent adjustments due to the changes in operating conditions. The goal of fuzzy control is to replace this set of PID controllers with one self-tuning fuzzy controller and to eliminate the needs for further tuning, once the machine is in operation.

Table 2 : FLC Output Adaptation

Rise Time Rules If Tracking is L_1 then adjustment is L_2		Overshoot Rules If overshoot is L_3 then adjustment is L_4	
L_1	L_2	L_3	L_4
SS	PS	L	NL
MS	PM	M	NM
VS	PL	S	NS

3 AFLC APPLIED TO INDUSTRIAL PROCESS

3.1 Hardware Setup

A generic diagram of the process that applies to all components of the machine is shown below:

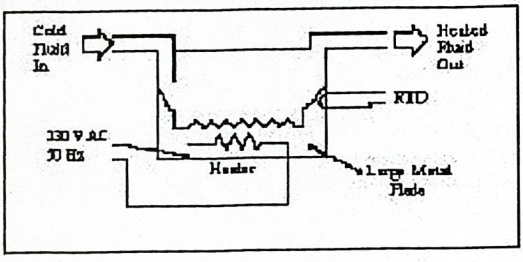


Fig. 5: An Industrial Temperature Control Application

This heating equipment of high temperature liquids has a large thick metal plate on the underside of the tank between the bottom and the inside of the tank, as shown in Fig. 5. It can be shown that this is a second order system with two thermal time constants. The first one correspond the thermal resistance from the heater to the plate and the plate heat capacity [1]. The second one comes from the thermal resistance of the plate to the material and the heat capacity of the material. There are many variations in the dynamics of the system. The thermo capacity is proportional to the size of the tank, which is quite different from one component to another. The time delay in the system is quite sensitive to the placement of the RTD. The heater can be found to be undersized or oversized. The heater on and off is controlled by a 24V pulse width modulated (PWM) signal applied to the SSR, as show in Fig. 6.

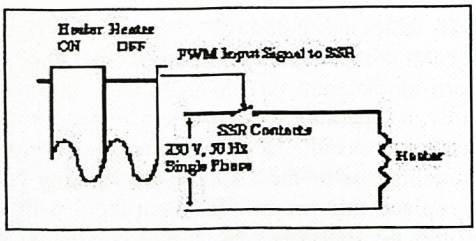


Fig. 6: Heater Excitation

3.2 Hardware Test Results

The proposed Adaptive fuzzy control algorithm was compared experimentally with the existing PID control used in industry. In this application, it is important to prevent over shoots, which seriously affect the quality of the product. It is also desirable to have a smooth control signal that does not require excessive on and off actions in the heater. The results are shown in Figure 7(a)(b)(c)(d). The top portion of each figure is a comparison of the PID vs. Adaptive Fuzzy temperature response, while the bottom portion is their respective heater *on times.* The Temperature Control Node was used to control the process for both controllers under the same conditions (i.e. same ambient temperature, delays, etc.). The results were obtained by actually controlling the process in its industrial setting.

The comparison of the performance of the FLC and PID controllers was performed under different set points, different thermal mass and different time delays. In each case, the FLC was able to successfully meet all design specifications without operator's tuning. On the other hand, it is a standard practice that for each of these different testing conditions, the PID controller needs to be manually tuned. Otherwise, the resulting response produced by PID controller would usually be unsatisfactory, as can be seen in Fig. 7(a)(b)(c)(d).

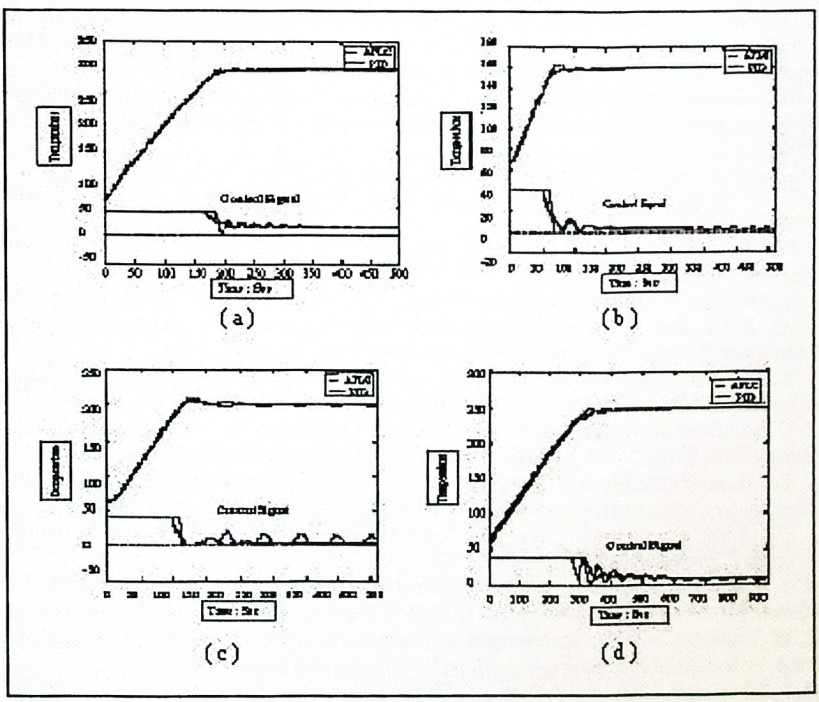


Fig. 7 : PID and Adaptive Fuzzy Control a) at High Temperature b) at Low Temperature c) with Time Delay d) with large Thermal Mass

4 Conclusions

Unlike some fuzzy controllers with hundreds, or even thousands, of rules running on dedicated computer systems, a unique AFLC using a small number of rules and simple scaling adaptation mechanism and straightforward implementation is proposed to solve a class of temperature control problems with unknown dynamics or variable time delays commonly found in industry. Additionally, the FLC can be easily programmed into many currently available industrial process controllers. The FLC was implemented for a tank temperature control problem. The results show significant improvement in maintaining performance. The AFLC also exhibits robust performance for plants with significant variation in dynamics.

References

1. James Dawson, "Fuzzy Logic Control of Linear Systems with Variable Time Delay," M.S. Thesis, Cleveland State University, June 1994.
2. Thomas A. Trauztch, "Self-Tuning Temperature Control Using Fuzzy Logic", M.S. Thesis, Department of Electrical Engineering, Cleveland State University, June 1996
3. David J. Elliott, "Fuzzy Logic Positional Servo Motor Control Development Platform", ", M.S. Thesis, Department of Electrical Engineering, Cleveland State University, June 1997.
4. Wilfred Nonnenmacher, "Fuzzy Logic Position Control of a Servo Motor", M.S. Thesis, Department of Electrical Engineering, Cleveland State University, April 1997.
5. Nitin Dhayagude, Zhiqiang Gao and Fouad Mrad, "Fuzzy Logic Control of Automated Screw Fastening", Journal Robotics and Computer Aided Manufacturing, Vol. 12, No.3 pp. 235-242, 1996
6. K. Passino and S. Yurkovich, Fuzzy Control, , Addison- Wesley, 1998.
7. O.J.M. Smith, "A Controller to Overcome Dead Time " ISA Journal, No. 2,28, February 1959.
8. P.S. Buckly, "Automatic Control of Processes with Dead Time," Proc. IFAC World Congress, Moscow, 1960, p33-40.
9. J.E. Marshall, "Control of Time Delay Systems," Stevenage, UK; NY P. Peregrinus, c1979.
10. Q. Brone, and S. Harris, "Varying Time Delay Estimation and Self-Tuning Control," Proceedings form the 1991 American Controls Conference, v2, p1740- 1741.
11. K. Shu, and J. Yan, "Robust Stability of Uncertain Time Delay Systems and its Stabilization by Variable Structure Control," Int. Journal of Control, 1993, v57 n1, p237-246.
12. G.P. Lui, and H. Wang, "Adaptive Controller for Continuous-Time Systems with Unknown Varying Time Delay," 1991 Int. Conf. of Cont. IEE Conf. Pub. v2, n332, p1084-1088.
13. C.C. Lee, "Fuzzy Logic in Control Systems: Fuzzy Logic Controller -- Parts 1 & 2 " IEEE Trans. on Sys. Man, and Cybernetics, Vol 20, No.2, pp404-435 March/April 1990.
14. P.J. King, and E.H. Mamdani, "The Application of Fuzzy Control Systems to Industrial Processes," Automatica, v11, p235-242, 1977.
15. S. Chiu, S. Chand, D. Moore, and A. Chaudhary, "Fuzzy Logic for Control of Roll and Moment for a Flexible Wing Aircraft", IEEE Cont, Sys. Mag. Vol. 11,No. 4, '91, pp42-48.
16. S. Tzafestas, and N. Papanikolopoulos, "Incremental Fuzzy Expert PID Control," IEEE Trans. on Industrial Electronics, Vol. 37, pp365-371, October 1990.
17. P. Oliveria, P. Lima, and J. Sentierio, "Fuzzy Supervision of Direct Controllers," Proc. 5th IEEE International Symposium on Intelligent Control, pp638- 643, 1990.

Preliminary Processing and Lossy Compression of Multichannel Information Data

Vladimir Lukin[1], Mikhail Zriakhov[1], Anatoliy Popov[1],
Oleksiy Pogrebnyak[2]

[1] National Aerospace University, Dept of Transmitters, Receivers and Signal Processing, 17 Chkalova St, 61070 Kharkov, Ukraine
lukin@xai.kharkov.ua
[2] Instituto Politecnico Nacional, Centro de Investigacion en Computacion, Ave. Juan de Dios Batiz S/N, C.P. 07738, Mexico, D.F., Mexico
olek@pollux.cic.ipn.mx

Abstract. A task of compressing multichannel information data (MID), i.e., signals from multiple sensors, is considered. Our approach is based on the following distinctive peculiarities: a) representation of multichannel information data as a specific image (2-D data array) and applying lossy DCT-based compression in rectangular shape blocks; b) use of preliminary statistical and correlation analysis for exploiting inter-channel correlation for performing specific pre-processing operations, namely, re-ordering, normalization, cyclic shifting; c) allocation of service data that describe parameters of pre-processing operations into a separate bit-stream and its lossless coding. It is shown that these steps carried out together, can lead to considerable increase of compression ratio for a given fixed measure of distortions (PRD) introduced by lossy compression. The proposed method is tested for simulated data for quantitative evaluation and then verified for real life multichannel signals.

Keywords: multichannel information data, lossy compression.

1 Introduction

Modern industrial control and diagnostic systems are often equipped by a set of sensors that produce simultaneous measurements of certain parameters at given instants of time. Let us give some examples of such systems. One important class is a number of spatially distributed sensors that measure radiation in certain locations around nuclear power stations or other dangerous objects [1]. Another example is a set of sensors that control flight parameters of aircrafts or unmanned aerial vehicles [2]. Modern medical diagnostic systems that register electrocardiograms (ECG) or electroencephalograms (EEG) are also usually multichannel [3,4].

Aforementioned systems and applications might seem having nothing in common since they differ from each other by basic properties of registered signals, sampling rate, number of bits used for each sample representation, etc. However, there are two similar aspects. First, as already said, the obtained (recorded) data in all considered

© L. Sánchez, O. Pogrebnyak and E. Rubio (Eds.)
Industrial Informatics
Research in Computing Science 31, 2007, pp. 105-114

systems are multichannel and for some subsets of channels these data can exhibit some degree of correlation. Another unifying aspect is that it is desirable to compress the obtained multichannel data either for transferring them via communication lines or for archiving [5,6].

There exist many different approaches to compression of multichannel information data (signals). First, data can be coded in a lossless or lossy manner [7]. Lossless compression has been a subject of intensive research for several decades but a provided compression ratio (CR) in many practical situations is unable to satisfy user requirements since CR is too small (slightly larger than unity). Due to this, lossy compression techniques have gained popularity in such application areas as audio, telecommunications, telemedicine [6,7], etc. Therefore, below we pay attention to lossy compression of MID.

Second, MID can be compressed component-wise, i.e., separately for each channel. This way is simple but it does not allow exploiting inter-channel correlation that might exist for all or some signals in different channels. Particular solutions already proposed for multichannel ECG and other medical signals have demonstrated that it is worth to exploit inter-channel correlation in one or another manner to improve compression performance [3,4]. Note that inter-channel correlation exists not only for medical MID, but for other types of sensors that register data from a common information source that can be, for example, placed at different distances from particular sensors.

People who dealt with lossy image compression know that such conventional techniques as JPEG and JPEG2000 allow taking advantage of spatial redundancy of data in vertical, horizontal and other directions simultaneously by using data decorrelation [7,8]. Our idea is to exploit this useful property of image coders by representation of MID as a specific image that has much larger amount of "pixels" in rows that correspond to temporal axis of registered (sampled and quantized) data than in columns that correspond to channel indices. MID can be represented as such an image as they are, i.e. column index corresponds to a channel index and no pre-processing is applied. We show that this is not the best way. Channel reordering, signal power normalization in each channel and, possibly, signal cyclic shifting could be operations the use of which results in considerable improvement of data compression performance. The parameters that describe these operations should be either known at decoder side or have to be compressed together with other type of coded data, quantized discrete cosine transform (DCT) coefficients for our variant of MID compression scheme. On one hand, service data on these parameters that form a separate bit-stream occupy some (additional) space. On the other hand, it is shown that in majority of practical situations, this space is very small in comparison to the main bit-stream, and the positive effect of the proposed pre-processing is sufficient.

The influence of the proposed operations on performance of the proposed coder is analyzed by computer simulations for specially synthesized test signals. These signals possess only the most general properties of MID. Particular items like block size and lossless coder selection are discussed as well. At the end, we give some details concerning particular applications of the designed lossy compression scheme and present performance gain.

2 Test Multichannel Signal

In order to simulate basic properties of real life MID, we have created a test multichannel signal. It has been simulated in the following manner. First, a set of mutually independent zero mean i.i.d. random signal $\{A_{kl}\}$ has been generated where l denotes a temporal (row) index and k is a channel (column) index, l=1,...,L, k=1,...,K. For each k, random values of $\{A_{kl}\}$ had uniform distribution within the limits [-127; 127] and they have been rounded-off to the nearest integer. For the original $\{A_{kl}\}$ the parameters L and K were larger than a required temporal sample size L_s and a number of channels Λ, respectively.

Then, the obtained array $\{A_{kl}\}$ has been subject to a sliding mean filter with a scanning window size NxN with obtaining a new array $\{A^f_{kl}\}$ as $A^f_{kl} = [\sum_{p=l}^{l+N-1} \sum_{q=k}^{k+N-1} A_{pq} / N^2]_{r-o}$ where l=1,...,L_s and $L \geq L_s$+N-1; k=1,...,K and $K \geq \Lambda$+N-1, $[\cdot]_{r-o}$ means rounding-off to the nearest integer. This way one obtains the following properties of the multichannel test signal $\{A^f_{kl}\}$. First, it is an array of 8-bit integers where in a random process in each channel is zero mean and approximately Gaussian (according to the central limit theorem that holds if NxN is large enough). These random processes are characterized by the same (temporal) autocorrelation defined by N and they have approximately equal variances σ^2_k. Inter-channel correlation is also determined by N and the signals in channels with smaller channel indices' difference $|k_1 - k_2|$ have larger correlation factor $R_{k1k2} = \sum_{p=1}^{L} A^f_{k1l} A^f_{k2l} / (L\, \sigma_{k1}\, \sigma_{k2})$. If $|k_1 - k_2| \geq N$, then $R_{k1k2} \approx 0$. Also note that for $|k_1 - k_2| < N$ the maximal value of cross-correlation function $R_{k1k2}(\Delta l)$ is commonly observed for $\Delta l = 0$. This can be treated as zero mutual shift of channel signals that have some degree of similarity.

All aforementioned properties remain after the next operation defined as $A^*_{kl} = [127 A^f_{kl} / A_{\max}]_{r-o}, A_{\max} = \max_{kl}\{A^f_{kl}\}$ which is intended on "stretching" the simulated data array to the range from –127 to 127. Then the simulated multichannel signal looks like it is shown in Fig. 1 (N=9, Λ=8).

Having the mentioned properties of $\{A^*_{kl}\}$ (in fact, in our simulations $\Lambda = 8$, $L = 4096$), it is easy to simulate different properties of real life multichannel signals. In particular, it is possible to simulate any arbitrary order of one-channel signals in multichannel array and, respectively, an arbitrary set of $\{R_{k\,k+1}, k = 1,...,\Lambda - 1\}$. It is also possible to simulate different powers of signals in different channels by multiplying $\{A^*_{kl}, l = 1,...,L\}$ by the corresponding factors F_k for each k-th channel. By cyclic shifting the signals in different channels, one can simulate mutual shifts $\Delta l_{k1k2} \neq 0$, i.e. "delayed" versions of similar signals.

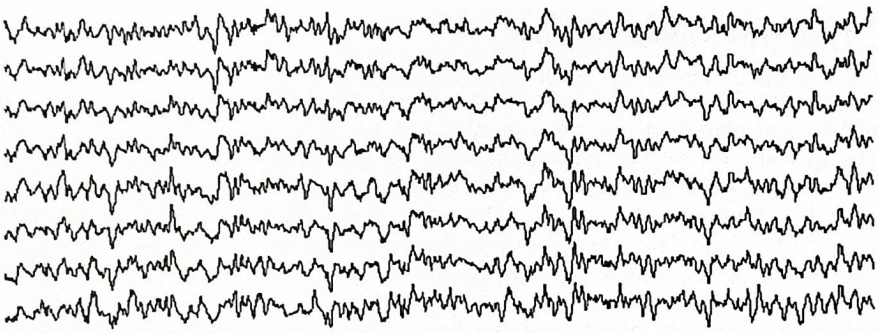


Fig. 1. An example of one realization of the simulated multichannel test signal.

All these effects happen in practice. For example, for standard (Franck) 8-lead system of multichannel ECG the factors R_{k1k2} are commonly of the order from 0.4 to 0.95, powers of channel (lead) signals can differ by several times, and the absolute values Δl_{k1k2} for which maxima of $R_{k1k2}(\Delta l)$ are observed are of the order of tens for sampling rate of hundreds Hz.

For MID that are observed in UAV control, the channel signal properties are essentially different. Considerable (of the order 0.3...0.5) values of R_{k1k2} are observed for few pairs of channels (speed projections, acceleration projections), for other pairs of channel signals that correspond to data of that or of different origin (e.g., altitude and pressure or temperature) the values R_{k1k2}, as can be expected, are practically equal to zero. The values Δl_{k1k2} are equal to zeroes for the channels for which the values R_{k1k2} are large enough. Signal powers in different channels can differ by tens of times since they might correspond to different physical processes and be expressed using particular calibration.

In opposite, radiation MID correspond to signals (sequences of measurements) of the same origin. The values R_{k1k2} for them are commonly larger, Δl_{k1k2} for different pairs of channels can differ from zero, channel signal powers differ a lot.

As seen, the observed phenomena have some degree of similarity (existence of correlation in all or in some pairs of channels), but there are also quite many peculiarities for each particular application. Intuitively, cross-channel correlation should be taken into account in design of MID lossy compression techniques.

3 Proposed Approach to Lossy Compression of MID

As mentioned in Abstract and Introduction, the proposed approach to lossy compression of MID has several distinctive features. The first feature is the use of preliminary statistical and correlation analysis for exploiting inter-channel correlation for performing specific pre-processing operations, namely, re-ordering, normalization and cyclic shifting. The second feature is representation of MID as a specific image

and applying lossy DCT-based compression in rectangular shape blocks. The third feature is allocation of service data that describe parameters of pre-processing operations into a separate bit-stream and its lossless coding. Note that in some cases not all operations of pre-processing are to be performed.

Let us consider the proposed approach more in detail. Suppose that for all channels a sampling rate is the same and one has an original MID $\{P_{kl}\}$, k=1,...,Λ, l=1,...,L. A way in which rows of $\{P_{kl}\}$ are assigned to particular signals can be different. For example, sensors can be given by a set of indices and then the k-th row of $\{P_{kl}\}$ corresponds to data from sensor with an index k. If such correspondence is established, let us determine cross-correlation functions $R_{k1k2}(\Delta l), k1 = 1,..., L-1, k2 = k1+1,..., L$ and powers $\sigma_k^2 = \sum_{l=1}^{L}(P_{kl} - \overline{P}_k)^2 /(L-1), \overline{P}_k = \sum_{l=1}^{L} P_{kl} / L$ for all k of $\{P_{kl}\}$. Then find the maximal value $R_{k1k2}^{\max}$ of $R_{k1k2}(\Delta l)$ and the corresponding $\Delta l_{k1k2}^{\max}$, also calculate $R_{k1k2} \approx R_{k1k2}^{\max} / \sigma_{k1}\sigma_{k2}$.

Suppose now that there is a correspondence between the set $\{k, k$=1, ...,$\Lambda\}$ and some set $\{m\}$ (this correspondence also defines mappings $\{\sigma_k, k = 1,..., \Lambda\} \rightarrow \{\sigma_m\}$ and $\{R_{k1k2}\} \rightarrow \{R_{m1m2}\}$). The total number of such correspondences is $\Lambda!$. Our proposition is to find such reordering of channels that a product $\prod_{m=1}^{\Lambda-1} R_{mm+1}$ attains maximal value. In other words, signals in channels of MID are to be placed in such a manner that inter-channel correlation factors for neighbor channels are large.

It might seem that the preliminary operation of channel re-ordering requires a lot of computation. However, in practice it is not always necessary. For example, the values of cross-correlation factors in multichannel ECG do not change considerably depending upon a patient and for several recordings of the same patient. Similarly, for UAV multichannel control data it is possible to establish the "optimal" order of channels in advance. This means the following. After carrying out the corresponding experiments it is possible to recommend some optimal (or, at least, quasi-optimal) order of channel representation of MID in 2-D array (let us denote it as $\{P_{kl}^{opt}\}$. For 8-channel ECG, the optimal order of leads is I-V6-V5-II-V4-V3-V2-V1.

For UAV multichannel data, one has to "group together" the channels that have the same origin and large values of cross-correlation factors, i.e., to place in neighbor channels speed projection data, acceleration data, etc. If the optimal order of channels is fixed and known in advance, there is no need to perform cross-correlation analysis of MID. Moreover, it is supposed known to coder and, respectively, to decoder. Then, no service information on the used order of channels is required. In the opposite case, i.e., if the preliminary operation of channel re-ordering is done, service information dealing with it is the accepted order of channels. It occupies $(\Lambda-1)[\log_2 \Lambda]_{\uparrow}$ bits where $[\square]_{\uparrow}$ denotes rounding-off to the nearest larger integer.

Recall that we consider the compression framework that includes Preliminary processing of MID → Performing 2-D DCT in rectangular shape blocks → Quantization of DCT coefficients → Lossless coding of service data and quantized DCT coefficients represented as 1-D array. Within this framework (details will be given below), the use of channel data re-ordering produces CR reduction by 1.03...1.2 times in comparison to the case when optimal ordering is not carried out.

Another operation of preliminary processing is elimination of non-zero Δl_{k1k2}. More in detail, one needs to know $\Delta l_{1k}, k = 2,...,\Lambda$ for the selected (used) order of channels in $\{P_{kl}^{opt}\}$. Then, the mutual shift (with respect to the first channel) are to be eliminated by cyclic shifting of data by $\Delta l_{1k}, k = 2,...,\Lambda$. The shift values form another subset of service information. Its volume expressed in bits is $(\Lambda - 1)[\log_2 \max(\Delta l_{1k}, k = 2,...,\Lambda)]_{\uparrow}$. If a channel order is fixed and the values $\Delta l_{1k}, k = 2,...,\Lambda$ are stable and known in advance for a given type of MID, the considered operation does not require to perform the cross-correlation analysis of channel data. Also, in this case $\Delta l_{1k}, k = 2,...,\Lambda$ can be known at decoder side and there is no need in coded bit-stream of service data.

According to simulations performed for the test signals described in Section 2, the operation of elimination of non-zero Δl_{k1k2} can result in CR reduction (for a given rate of distortions introduced) by 1.02...1.1 times. The benefit is commonly larger when channel signals have a higher degree of cross-correlation and wider main lobes of auto-correlation functions.

The third operation of MID pre-processing is signal power normalization in channels. Here the term "normalization" means that we would like to have (approximately) equal variances $\{\sigma_k^2\}_{norm}$ (or standard deviations) for $\{P_{kl}^{opt}\}$. This can be done in different ways. For example, it is possible to determine $\{\sigma_k, k = 1,...,\Lambda\}$ for $\{P_{kl}^{opt}\}$, to find the maximal value $\sigma_{\max}$ among $\{\sigma_k, k = 1,...,\Lambda\}$, and to obtain the normalized version of $\{P_{kl}^{opt}\}$ as

$$P_{kl}^{on} = P_{kl}^{opt} \phi_k, l = 1,...L;\ \phi_k = \sigma_{\max} / \sigma_k, k = 1,...,\Lambda. \quad (1)$$

The set of normalizing factors $\{\phi_k, k = 1,...,\Lambda\}$ is one more kind of service data that describe MID pre-processing. This set is subject to lossless compression (coding). Since the values $\{\phi_k, k = 1,...,\Lambda\}$ are not integers, their coding might require, at most, 32Λ bits. Note that normalization is a rather simple and fast operation. But it is very efficient in the sense of CR reduction (see next Section).

On the contrary to JPEG, we propose to apply rectangular shape blocks an array $\{P_{kl}^{opt}\}$ is divided to. The block size is ΛL_{bl} where L_{bl} is supposed to be larger than Λ and equal to power of 2 providing faster calculation of DCT in blocks. Rectangular shape of blocks opens several opportunities to scan quantized DCT coefficients while forming a 1-D array subject to lossless coding. We have analyzed the proposed coder efficiency for several variants of scanning and finally decided to apply the modified

zigzag scanning illustrated in Fig. 2. We have also tested several different algorithms of lossless coding. The best results have been provided by bzip2 [9].

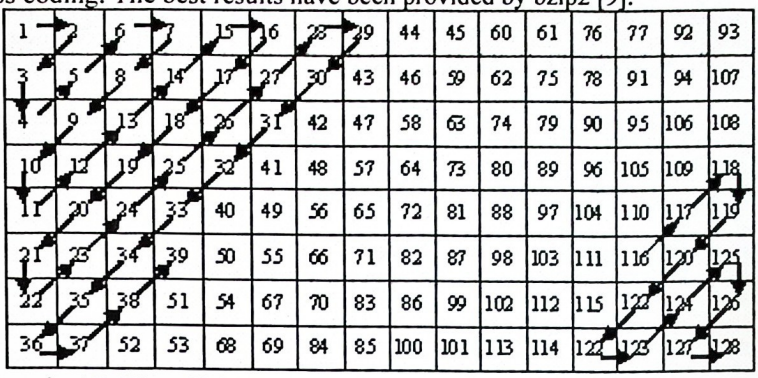


Fig. 2. Quantized DCT coefficients scanning in rectangular block (ΛL_{bl}=8x16)

4 Experimental Result Analysis

Consider now the test signal $\{A^*_{kl}\}$ and its "derivatives". To characterize its compression, let us use a criterion PRD [3, 4] that has been widely exploited in analysis of medical signal lossy coding. For MID, PRD can be expressed as

$$PRD\left(S,S^{dec}\right)=\sqrt{\frac{\sum_{k=1}^{\Lambda}\sum_{l=1}^{L}(S_{kl}-S_{kl}^{dec})^2}{\sum_{k=1}^{\Lambda}\sum_{l=1}^{L}(S_{kl}-\langle S_k\rangle)^2}}\times 100\,. \qquad (2)$$

where $\{S_{kl}\}$ and $\{S_{kl}^{dec}\}$ denote the original and decoded MID, respectively; $\langle S_k\rangle$ is the mean in the k-th channel.

Using the test signal $\{A^*_{kl}\}$, we have obtained different signals $\{S_{kl}\}$ and analyzed the dependences $PRD(bpp)$ (since $\{A^*_{kl}\}$ is an 8-bit data array, CR is strictly connected with bpp=8/CR). The procedure of modeling MID with different powers in channels was the following: Case 1 - $S_{kl}=A^*_{kl}, l=1,...,L, k=1,...,\Lambda$ (this case simulates the use of preliminary normalization), Case 2 - $S_{1l}=1.5A^*_{1l}, l=1,...,L$; $S_{kl}=A^*_{kl}, l=1,...,L, k=2,...,\Lambda$; $S_{8l}=2.5A^*_{8l}, l=1,...,L$ (i.e., only in the first and the eighth channels the signal power is larger than in other channels); Case 3 - $S_{1l}=1.5A^*_{1l}$; $S_{2l}=1.7A^*_{2l}$; $S_{3l}=A^*_{3l}$; $S_{4l}=0.5A^*_{4l}$; $S_{5l}=2A^*_{5l}$; $S_{6l}=2.5A^*_{6l}$; $S_{7l}=2.2A^*_{1l}$; $S_{8l}=0.75A^*_{8l}, l=1,...,L$, i.e. in all channels the signal powers are not

equal to each other and they differ by several times. The obtained dependences *PDR*(*bpp*) for all three considered Cases are presented in Fig. 3. As seen, the CRs for considerably different signal powers (Case 3) are essentially larger than for the Case of normalized signals (Case 1). Note that the benefit of normalization (degree of CR increasing) is larger than for other operations of pre-processing considered above.

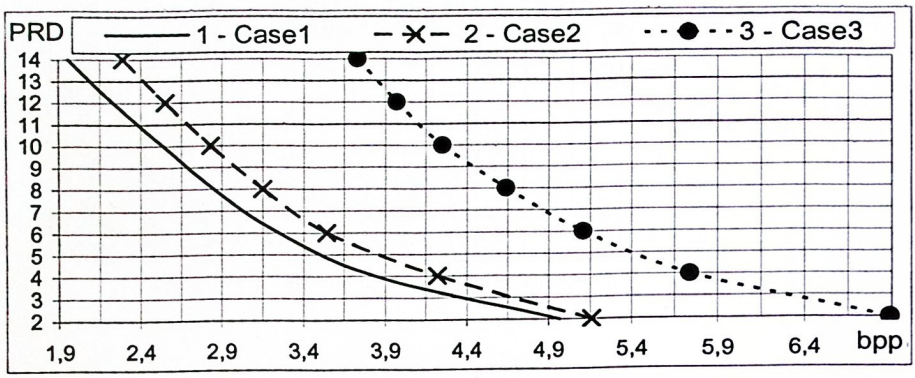


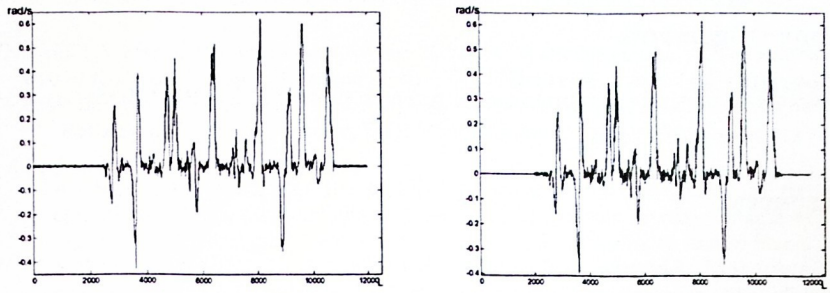
Fig 3. Dependences *PDR*(*bpp*) for MID with different powers in channels.

The CR increasing due to normalization is not the only one benefit of the operation of MID normalization. Our studies have also shown that for the case of normalized MID the PRD values in all channels occur to be approximately equal to each other and equal to the PRD (2). If normalization is not used, PRD values in channels with smaller power occur to be larger than PRD (2), this means that larger distortions are introduced in those channels, and this is undesirable.

We have also tested the influence of block size L_{bl} on CR. It occurred that, in general, L_{bl} increasing leads to larger CR. However, starting from some $L_{bl}^{\min}$ its further increasing does not result in noticeable reduction of bpp. Our recommendation is to select $L_{bl}^{\min}$ by 20...50 times larger than the signal auto-correlation main lobe width expressed in number of samples. For most practical situations $L_{bl}^{\min} = 256$ is enough.

Let us also demonstrate experimental results obtained for real life data. Fig. 4,a presents an original one-channel telemetric signal for testing UAV control system designed by Professor V.I. Kortunov, National Aerospace University, Kharkov, Ukraine (http://k504.xai.edu.ua/eng/nauka_eng.php?link=autocontrol_eng). In aggregate, this system records data for 11 channels, but the plot in Fig. 4,a shows only angular speed data (sampling rate is 50 Hz). The same channel signal after compression (together with other channel UAV data using the proposed method) with PRD=5% and decompression is represented in Fig. 4,b, CR=12. Note that in case of separate compression of data in each channel with the same PRD the obtained CR is 1.3 times smaller. The improvement is provided due to accounting inter-channel correlation.

As seen, no considerable distortions of information are observed; at the same time, noise filtering effect is visible. This is one more positive feature of lossy compression in case of its applying to signals corrupted by noise. Earlier similar effects have been observed in lossy compression of noisy images [10-12].



a b

Fig. 4. Original one-channel signal of UAV control data (a) and the result of its compression and decompression (b)

We have also applied the proposed approach to lossy compression of 8-channel ECG kindly passed to us by the Center XAI-Medica (http://www.xai-medica.com/), National Aerospace University, Kharkov, Ukraine. The reached average (for 9 patients) CR in case of PRD=5% is about 26. This is approximately 1.7 times better than for channel-by-channel lossy compression of the considered ECGs with the same PRD. These results are slightly better than those ones reported in [3,4]. Also note that in case of 12-channel ECG records where there are 4 sum-difference channels it is possible to compress 8 basic channels and to recover sum-difference channel ECGs from the corresponding decompressed channel ECGs. PRD values in sum-difference channels are then about 1.4 times larger than in basic channel ECGs and this is appropriate for practical applications.

5 Conclusions

A new method of lossy compression of multichannel information data is proposed. Peculiar features of the proposed approach consist in applying the operations of preliminary analysis and processing of MID like channel signal re-ordering, normalization and cyclic shifting for improvement of inter-channel correlation of data and, thus, increasing the CR. Another specific feature is the use of MID representation as an image next applying DCT in rectangular shape blocks.

The proposed method has been tested for specially simulated random MID with certain intra and inter-channel correlation properties. It has been shown that the use of all aforementioned operations leads to CR increasing, and the main improvement is contributed by normalization.

Verification for real life MID (UAV control signals and multichannel ECGs) has been performed as well. The obtained results demonstrate that due to incorporating

inter-channel correlation of data it is possible to increase CR by 30...70% in comparison to channel-by-channel compression under the same level of introduced distortions. Moreover, these "distortions" may relate to noise removal if original signals subject to compression are noisy.

Acknowledgements

This work was partially supported by the research project SEMAR #11055.

References

1. Brennan, S.M., Maccabe, A.B., Mielke, A.M., Torney, D.C.: Radiation Detection with Distributed Sensor Networks. IEEE Computer, V. 37, 8 (2004) 1-10
2. Herwitz, S.R., Leung, J.G., Higgins, R.G., Dunagan, S.E., Arvesen, J.C.: Remote Command-and-Control of Imaging Payloads Using Commercial Off-the-Shelf Technology. Proceedings of IGARSS (2002) 2726-2728
3. Cohen, A., Zigel, Y.: Compression of Multichannel ECG Through Multichannel Long-Term Prediction. IEEE Engineering in Medicine and Biology (1998) 109-115
4. Zigel, Y., Cohen, A., Katz, A.: ECG Signal Compression Using Analysis by Synthesis Coding. IEEE Transactions on Biomedical Engineering. V. 47, 10 (2000) 1038-1316
5. Keller, P.E., Kouzes, R.T., Kangas, L.J.: Applications of Neural Networks to Real-Time Data Processing at the Environmental and Molecular Sciences Laboratory (EMSL). Proceedings of 8th IEEE Conference on Real-Time Computer Applications in Nuclear, Particle and Plasma Physics (1993) 4 p.
6. Rosen, E.: Mobile telemedicine arrives. Telemedicine Today, V. 40, 10 (1997) 14–42
7. Salomon, D.: Data Compression. The Complete Reference. 3-rd edn. Springer-Verlag, Berlin Heidelberg New York (2004)
8. Taubman, D., Marcellin, M.: JPEG 2000: Image Compression Fundamentals. Standards and Practice. Kluwer, Boston (2002)
9. Burrows, M., Wheeler, D.J.: A Block-sorting Lossless Data Compression Algorithm. SRC Research Report 124, Digital Systems Research Center, Palo Alto (1994)
10. O.K. Al-Shaykh, R.M. Merserau: Lossy Compression of Noisy Images. IEEE Trans. On Image Processing, V. 7 12 (1998) 1641-1652
11 Ponomarenko, N., Lukin, V., Zriakhov, V., Egiazarian, K., Astola, J.: Lossy compression of images with additive noise. Proceedings of International Conference on Advanced Concepts for Intelligent Vision Systems, Antwerpen, Belgium (2005) 381-386
12. Ponomarenko N., Lukin V., Zriakhov M., Pogrebnyak O.: Methods for Lossy Compression of Images Corrupted by Multiplicative Noise. Proceedings of the International Conference "Modern Problems of Radioengineering, Telecommunications and Computer Science" (TCSET), (2006) 278-281

Unsupervised Remote Sensing Data Classification Using Multimodal Statistical Model

Anatoliy V. Popov[a], Oleksiy Pogrebnyak[b], Alexandra N. Brashevan[a]

[a]National Aerospace University of Ukraine "Kharkov Aviation Institute", Chkalova Str. 17, 61070 Kharkov, Ukraine, Telephone/Fax: +38(0572) 441186, E-mail: off@xai.edu.ua
[b]Centro de Investigacion en Computacion, Instituto Politecnico Nacional, Av. Juan de Dios Batiz S/N, C.P.07738, Mexico, D.F., Mexico, E-mail: olek@pollux.cic.ipn.mx

Abstract. A solution for the problem of unsupervised recognition in the conditions of a priori indefinite number of object classes in radar images is presented. The designed algorithm performs image clustering to divide image objects into classes. The region of interest is can be chosen by user and then probabilistic filtering is applied to recognize the objects of the predetermined class on the entire image. The algorithm is operated on the multichannel data and shows stable recognition results.

Keywords: image processing, unsupervised classification, remote sensing

1 Introduction

The airborne and spaceborne remote sensing (RS) systems are widely used for ecological monitoring of environment, mapping and extraordinary situation prevention. The analysis of aerospace images allow to detect the waters pollutions such as oil spills, monitor the borders of rivers and lakes, detect ice obstruction, evaluate water resources, soil state and erosion and so on [1]. Aerospace remote sensing of the Earth surface is performed as by radar systems (including polarimetric radars) as in optical an IR bands [2]. Remote sensing data often are represented by multichannel images where the image brightness corresponds to the amplitude of the signal reflected from the sensed surface [3].

For example, the radar image shown Fig. 1 contains many spatially distributed objects whose number a priori is unknown. Object separation and contouring, detection of the similar objects and their classification (object clustering) using the level of reflected signal intensity is one of the goals of the processing of such images. The noise contained in the images is caused by equipment and reflected signal fluctuations, and significantly complicates the image analysis and classification. An automatic image classification requires to determine the number of the classes of the objects that can be distinguished in the analyzed image. The statistical characteristics of the separable objects can be determined for their following recognition in the analyzed and similar images.

The automatic classification can be performed analyzing the local histograms calculated on the data within the sliding on the entire image window (see Fig. 1). The window size must be matched to the RS system resolution. Since in the separated image fragment several objects may be presented, the resulted histogram may be of

© L. Sánchez, O. Pogrebnyak and E. Rubio (Eds.)
Industrial Informatics
Research in Computing Science 31, 2007, pp. 115-124

the multimodal character (see Fig. 2). Analyzing the histogram multimodal content one can determine the number of the objects in the separated image fragment.

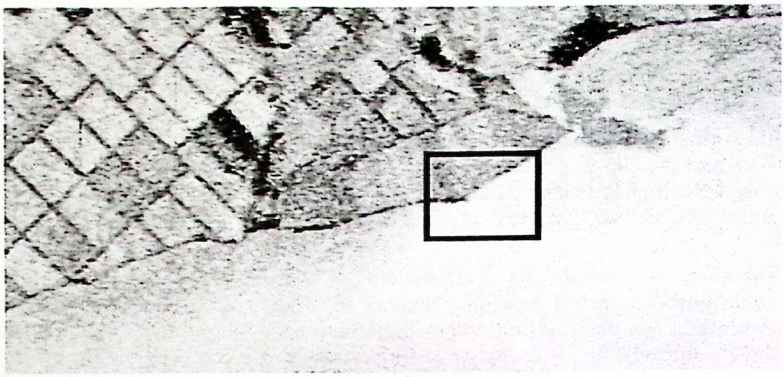


Fig. 1. Fragment of radar image of sea cost (negative).

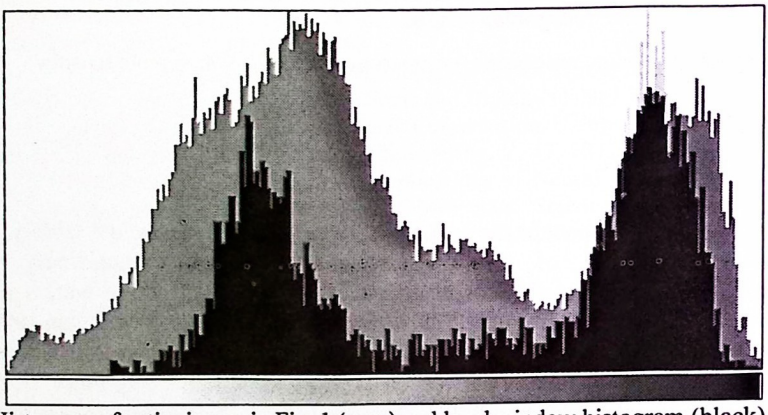


Fig. 2. Histogram of entire image in Fig. 1 (gray) and local window histogram (black).

2 Multimodal Statistical Model

One of the methods of multimodal distribution descriptions is the use of the normal distribution mixtures of the form $f(x)=\sum_{k=1}^{M} p_k \cdot \varphi_k(x)=\sum_{k=1}^{M} p_k \dfrac{\exp\left\{-\dfrac{(x-m_k)^2}{2\sigma_k^2}\right\}}{\sqrt{2\pi\sigma_k^2}}$ [4], where M is a number of the normal kernels $\varphi_k(x)$; m_k, σ_k are the parameters of k-th normal distribution $\varphi_k(x)$; p_k are the weight coefficients satisfying the condition $\int f(x)dx=1$.

The procedure of finding parameters M, m_k, σ_k, p_k is based on the minimization of the mean square error of the approximation. Since the true distribution density function is a priori unknown, the number of kernels M necessary to build model, the distribution parameters m_k, σ_k, and the weighting coefficients p_k are also unknown.

In reference [5] a probabilistic approach is proposed to find the estimates of the normalizing coefficients p_k, which consists of the determining of the probability of the appearance of each approximating kernel $\varphi_k(x)$ that can be evaluated on the distribution histogram. However, the implementation of such an approach requires the number of modes in the distribution $h(x)$ to be determined and the sample distribution to be separated onto the contented mixtures.

The determination of the number of modes for the analytical distributions is not a problem because one can find this number solving the following equations: $\frac{\partial h(x)}{\partial x} = 0$, $\frac{\partial^2 h(x)}{\partial x^2} > 0$.

Unfortunately, the non-smooth character of the histograms $h(x)$ that contained the source data for the approximation does not allow the use of the numerical differentiation methods directly for $h(x)$ [6].

Even if the number of modes in the distribution is known, its parameters m_k, σ_k, $k = 1...M$ of the normal kernels $\varphi_k(x)$ of the distribution mixture $f(x) = \sum_{k=1}^{M} p_k \varphi_k(x)$ are unknown. These parameters can be evaluated only in case when the source sample $h(x)$ can be divided by the data composed by each distribution mode $\varphi_k(x)$ that is equivalent to perform the data clustering in the terms of the recognition theory [7]. This way the sample estimates would contain a significant error.

3 Multimodal Statistical Model Design

The design of the data multimodal statistical model can be performed in several stages. At the first stage it is necessary to determine the number of components M in the distribution mixture. In the presence of several modes in the histogram $h(x)$ of the experimental data, for example, their quantity can be determined by calculating the crosscorrelation between the histogram and an etalon distribution. The crosscorrelation function is used for the reasons that, first, the correlation coefficient characterizes the similarity of one function to another, and second, the change in the parameters of the etalon function allows estimation of the histogram parameters.

If the normal distribution is adopted as an etalon function then the crosscorrelation function can be determined as

$$R(m) = \int_{X\min}^{X\max} h(x) \cdot \varphi(x,m)dx \,. \tag{1}$$

When the expectation m of the normal distribution $\varphi(x,m)$ varies from $x_{\min}$ to $x_{\max}$ at fixed variance σ^2 value, the maximums of the correlation function (1) correspond to the positions of the modes in the distribution histogram.

The analogous "scanning" can also be performed for the variance of the normal distribution. To this end, the crosscorrelation function at the fixed expectation m of the normal distribution is calculated as follows:

$$R(\sigma) = \int_{X\min}^{X\max} h(x) \cdot \varphi(x,\sigma)dx \tag{2}$$

The crosscorrelation function (2) is less sensitive to the variance changes because the analyzed distribution has several modes. It is supposed that the most accurate results can be found calculating a bidimensional correlation function. Fig. 3 shows an example of such a function.

$$R(m,\sigma) = \int_{X\min}^{X\max} h(x) \cdot \varphi(x,m,\sigma)dx \tag{3}$$

To determine the number of distribution modes it is necessary to determine the number of maximums in the bidimensional correlation functions. This problem can be solved differentiating numerically the function $R(m,\sigma)$ and solving the following equations:

$$\frac{\partial R(m,\sigma)}{\partial m} = 0 \,, \frac{\partial R(m,\sigma)}{\partial \sigma} = 0 \,, \frac{\partial^2 R(m,\sigma)}{\partial m \cdot \partial \sigma} > 0 \,. \tag{4}$$

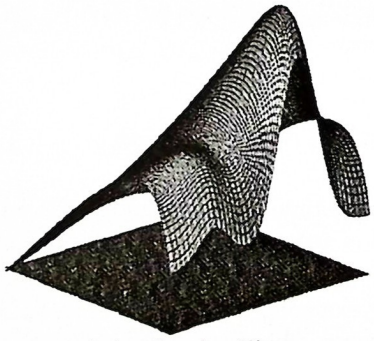


Fig. 3. Two-dimensional cross correlation function (3).

At the second stage the statistical estimates of the parameters m_k, σ_k are determined for each distribution mode $k = 1...M$. Such a possibility gives a sequential view of the function $R(m,\sigma)$ in the points that it satisfy to the conditions (4) and calculating the corresponding estimates of the values m_k, σ_k.

The third stage consists of finding the weighting coefficients for each distribution mixture components p_k, $k = 1...M$. The interpretation of the weighting coefficients p_k as probabilities of the data belong to the histogram clusters $\varphi_k(x)$ allows building the algorithm for the coefficients p_k determination using the method of the maximum a posteriori probability. For each data sample element x_i one can calculate the probability it drops in the class $\varphi_k(x)$, using the Bayes formula [7]:

$$Q_k(x_i) = \frac{\varphi_k(x_i)}{\sum_{j=1}^{M} \varphi_j(x_i)}. \quad (5)$$

As the parameters of the functions $\varphi_k(x)$ the previously found estimates of m_k, σ_k are used. Then, the maximal value of a posteriori probability $Q_k(x_i)$, $k = 1...M$ is found. The number k determines the number of the cluster $\varphi_k(x)$.

Counting the number of the elements n_k each cluster has the probability estimates that can be determined as follows:

$$p_k = \frac{n_k}{N},$$

where N is a total number of the sample elements.

Clearly, such an approach gives approximated estimates of the probabilities p_k because a priori it is assumed the equiprobable belonging of the element x_i to the class $\varphi_k(x)$. Besides, the considered approach automatically satisfies to the condition $\sum_{k=1}^{M} p_k = 1$.

The estimates of the parameters of the approximating multimodal distribution found according to the described method needs in a more accurate update. To this end, numerical optimization procedures [6] can be used. The cluster parameter estimates m_k, σ_k, p_k can be used as varied parameters in these procedures.

4 RESULTS

4.1 Simulation results on multimodal statistical modeling

To verify the elaborated method, samples were generated having random values of normal distribution with parameters N(10, 2) и N(0.0, 1.5). Then, the samples were mixed in proportion 1/8.

At the stage of the correlation analysis, a two-dimensional correlation function $R(m,\sigma)$ was obtained (see Fig. 3). The analysis of this function detected two clusters with the parameters N(10.02, 2.49) and N(0.26, 1.93) (see Fig. 4). A small bias in the estimates of the normal distributions is caused by mutual influence of the distribution modes at data correlation analysis.

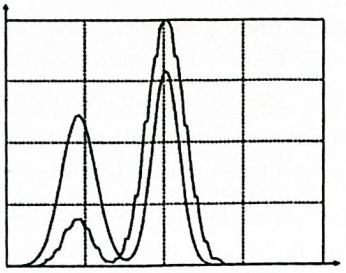


Fig. 4. Results of cluster analysis of bidimensional crosscorrelation function in Fig. 3.

To estimate the weighting coefficients p_k, the probabilities of classification for each element were calculated according to the method of the maximum of a posteriori probability (5). In such a manner, the statistical model was updated as it is shown in Fig. 5: p1 = 0.92, p2 = 0.08.

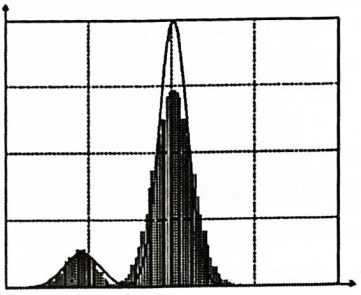


Fig. 5. Cluster probability estimation results.

The resulting values N(10.02, 2.49), p1 = 0.877 and N(0.26, 1.93), p2 = 0.123 were obtained applying a numerical optimization procedure. The shape of the approximating function is shown in Fig. 6.

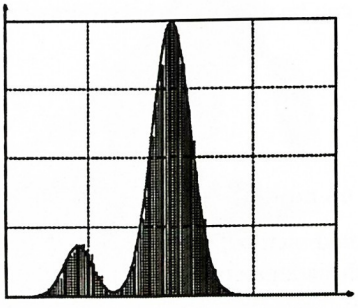


Fig. 6. Statistical model optimization results.

4.2 Results on application of automatic classification algorithm to real data

The proposed classification algorithm based on polynormal distribution was

applied to different images. For example, the result of the automatic classification the radar image in Fig. 1 is shown in Fig. 7. Here, the algorithm detected 16 different clusters. Each of them is represented in Fig. 7 by its intensity. Fig. 7 was obtained by application of the probabilistic filters [8], each filter was adjusted properly to its cluster.

The presence of the noise in the original image and the blurred object edges result in erroneous classification of the edges of the homogeneous areas, which may be classified as separate classes. Image prefiltering using low-pass smoothing filters [9] results in the reduction of the number of the separable object classes to 6 (see Fig. 8). It means that objects having different electrophysical nature (for instance, terrain and sea surface) may be classified as belonged to the same class. Therefore, the image preprocessing for subsequent classification needs in detail preserving filters [10] that do not smooth object edges.

The automatic classification of the images of better quality (sharper and less noisy, as, for example, image shown in Fig. 9) produces better results. In the image in Fig. 9 it was found 7 clusters shown in Fig. 10 by the areas of different intensity. The analysis of the classification results shows that in this case the areas of different electrophysical parameters practically are not unified.

5 Conclusions

The presented automatic image classification algorithm can be used to solve the problem of radar image classification in the conditions of a priori ambiguity about the number of the object classes and their statistical properties. The proposed classification algorithm permits to detect the distinguished object classes. The mathematical background for the proposed radar image classification algorithm is the image statistical model that assumes the polynormal distribution of the data. To approximate well the image histogram, it was solved the problem of the determination of the number of the kernels that are the centers of the histogram clusters. The statistical model updated recursively calculates more accurate statistical characteristics of the object classes. The additive components of the statistical model can serve as a description of the classes of the objects contained in the radar image and can be used for their recognition.

The application of the proposed algorithm to real radar images has demonstrated the stability of the proposed automatic classification technique. However, the number of the detected classes depends significantly on the quality of the original image and used preprocessing methods.

Fig. 7. Result of clustering radar image in Fig. 1.

Fig. 8. Result of clustering radar image in Fig. 1 after filtration.

Fig. 9. Image of river mouth.

Fig. 10. Result of classification of image in Fig. 9.

Acknowledgements

This work was partially supported by the research project SEMAR #11055.

References

1. Wolfgang-Martin Boerner, Eric Pottier, and Jong-Sen Lee. "Fundamentals of Radar/SAR Polarimetry." *Polarimetric and Interferometric SAR Techniques and Applications*, URSI-F Open Symposium on Propagation and Remote Sensing, Session 2AR, Wednesday, 2002 February 13, Germany.
2. Chipman, R. A., and J. W. Morris, eds. *Polarimetry:Radar, Infrared, Visible, Ultraviolet, X Ray*, Proc. SPIE-1317, 1990 (also see SPIE Proc. 891, 1166, 1746, 1988, 1989, and 3121).
3. A.C. van den Broek, A.J.E. Smith, A. Toet. "Visual Interpretation of polarimetric SAR Imagery." In *Image and Signal processing for Remote Sensing IV*, edited by Sebastiano B. Serpico, Proceedings of SPIE Vol. 4541 (SPIE, Bellingham, WA, USA, 2002), pp. 169-179. ISBN 0-8194-4266-6.
4. D. Middleton. "Non-Gaussian Noise Models in Signal Processing for Telecommunication." *IEEE Transactions. on Information Theory*, **Vol. 45, No. 4**, pp. 1129-1147, 1999.
5. A. Swami. "Non-Gaussian Mixture Models for Detection and Estimation in Heavy-Tailed Noise," *Proceedings of the ICASSP*, Vol. 6, Istanbul, Turkey, pp. 3802-3805, June 2000.
6. T. Shoup. *A practical guide to computer methods for engineers*. Prentice-Hall, inc. Englewood cliffs. NY, 1979.
7. K. Fukunaga. *Introduction to statistical pattern recognition*. Academic Press, NY, 1972.
8. Anatoliy V. Popov, Oleksiy Pogrebnyak. "Radar target recognition by probabilistic filtering." In *Earth Observing Systems IX*, edited by William L. Barnes, James J. Butler, Proceedings of SPIE Vol. 5542 (SPIE, Bellingham, WA, USA, 2004), p. 459-467. ISBN: 0-8194-5480-X.
9. Volodymyr V. Ponomaryov, Oleksiy B. Pogrebnyak. "Novel robust RM filters for radar image preliminary processing." *Journal of Electronic Imaging* **Vol. 8(4)**, USA, pp. 467-477, October 1999. ISSN 1017-9909.
10. Oleksiy B. Pogrebnyak, Juan Humberto Sossa Azuela, Pablo Manrique Ramírez. "Impulse rejecting image filter for efficient noise removal and fine detail preservation", In *Applications of Digital Image Processing IV*, Andrew G. Tescher, Chair/Editor, Proc. SPIE Vol.4472, 29 July - 3 August 2001, San Diego, USA, p.p. 546-554, ISBN: 0-8194-4186-4, ISSN: 0277-786X.

IIR lifting DWT for Lossless Image Compression

Oleksiy Pogrebnyak[1], Pablo Manrique Ramírez[2], Luis Pastor Sanchez Fernandez[3],

Instituto Politecnico Nacional, CIC-IPN, Av.Juan de Dios Batiz s/n,
Colonia Nueva Industrial Vallejo, C.P. 07738, Mexico D.F.
E-mail: [1]olek@pollux.cic.ipn.mx, [2]pmanriq@cic.ipn.mx, [3]lsanchez@cic.ipn.mx

Abstract. A data dependent wavelet transform based on the modified lifting scheme is presented. The algorithm is based on the wavelet filters derived from a generalized lifting scheme. The proposed framework for the lifting scheme permits to obtain easily different wavelet FIR filter coefficients in the case of the (~N, N) lifting and improve the performance by additional IIR filtering of the already calculated wavelet coefficients. The perfect image restoration in this case is obtained employing the particular features of the lifting scheme. Changing wavelet IIR filter coefficients, one can obtain the filter frequency response that match better to the image data than the standard lifting filters, resulting in higher data compression rate. The designed algorithm was tested on different images. The obtained simulation results show that the proposed method performs better in data compression for various images in comparison to the standard technique resulting in significant savings in compressed data length.

Key words: image processing, wavelets, lifting scheme, adaptive compression

1 Introduction

In the past decade, the wavelet transform has become a popular, powerful tool for different image and signal processing applications such as noise cancellation, data compression, feature detection, etc. Meanwhile, the aspect of wavelet decomposition/reconstruction implementation, especially for image compression applications, now continues to be under consideration.

The first algorithm of the fast discrete wavelet transform (DWT) was proposed by S.G.Mallat [1]. This algorithm is based on the fundamental work of Vetterli [2] on signal/image subband decomposition by 1-D quadrature-mirror filters (QMF), and orthonormal wavelet bases proposed by I.Daubechies [3]. Then, W.Sweldens [4] proposed the lifting scheme based on polyphase factorization of known wavelets that now is widely used (for example, in JPEG2000 standard) for lossless image/signal compression based on DWT. To enhance the energy compaction characteristics of the DWT, different methods basing on an adaptive lifting scheme [4 - 8], principal components filter banks [9] and signal-dependent wavelet/subband filter banks [10-12] were developed recently.

© L. Sánchez, O. Pogrebnyak and E. Rubio (Eds.)
Industrial Informatics
Research in Computing Science 31, 2007, pp. 125-132

In this paper, we present an algorithm for lossless image compression that is based on a subclass of IIR wavelet filters. These filters are derived from the generalized FIR wavelet lifting filters [8, 13, 14] introducing zeros in the prototype FIR filters. Performing causal filtering at the analysis and anticausal filtering of the time-inverted data at the synthesis stages, one can obtain the perfect image restoration with the presented IIR filters. Varying the order of the filter and the filter coefficients depending on the image data statistical/spectral properties, the decompositions can be optimized to achieve a minimum of the entropy in the wavelet domain.

2 DWT by Generalized Lifting Scheme

The lifting scheme [3] is widely used in the wavelet based image analysis. Its main advantages are: the reduced number of calculations; less memory requirements; the possibility of the operation with integer numbers for lossless data compression. The lifting scheme consists of the following basic operations: splitting, prediction and update.

Splitting is sometimes referred to as the lazy wavelet. This operation splits the original signal $\vec{x}$ into odd and even samples:

$$s_i = x_{2i}, \quad d_i = x_{2i+1}. \tag{1}$$

Prediction, or the dual lifting at the level k calculates the wavelet coefficients, or the details $\{d^{(k)}\}$ as the error of prediction of $\{d^{(k-1)}\}$ on $\{s^{(k-1)}\}$[14]:

$$d_i^{(k)} = d_i^{(k-1)} + \sum_{j=-\tilde{N}}^{\tilde{N}} b_j^p \cdot s_{i+j}^{(k-1)}, \tag{2}$$

where $\vec{b}^p$ are coefficients of the wavelet-based high-pass FIR filter and $\tilde{N}$ is the prediction filter order.

Update, or the primal lifting combines $\vec{s}^{(k-1)}$ and $\vec{d}^{(k)}$, and consists of low-pass FIR filtering to obtain a coarse approximation of the original signal $\vec{x}$:

$$s_i^{(k)} = s_i^{(k-1)} + \sum_{j=-N}^{N} b_j^u \cdot d_{i+j}^{(k)}, \tag{3}$$

where $\vec{b}^u$ are coefficients of the wavelet-based low-pass FIR filter and N is the update filter order.

The inverse transform is straightforward: first, the signs of FIR filter coefficients $\{u\}$ and $\{p\}$ are switched; the inverse update followed by inverse prediction is calculated. Finally, the odd and even data samples are merged.

In [13], the FIR filters that participate in the prediction and update operation were represented in the domain of Z-transform. Using this approach, we formulated the transfer function of the prediction FIR filter as follows [8, 14]:

$$H_p(z)=1+b_0^p\left(z+z^{-1}\right)+b_1^p\left(z^3+z^{-3}\right)+\ldots+b_{\tilde{N}-1}^p\left(z^{2\tilde{N}-1}+z^{-2\tilde{N}+1}\right), \tag{4}$$

The $H_p(z)$ must has zero at $\omega=0$, i.e., at $z=1$. It can be easily found [5] that this condition is satisfied when

$$\sum_{i=0}^{\tilde{N}-1} b_i^p=-\frac{1}{2}. \tag{5}$$

When the condition (5) is satisfied, $H_p(-1)=2$ and $H_p(0)=1$ that means the prediction filter has gain 2 at $\omega=\pi$ and unit gain at $\omega=\frac{\pi}{2}$.

Following this approach, the transfer function for update filter can be obtained in the terms of $H_p(z)$ [8, 14]:

$$H_u(z)=1+H_p(z)\left\{b_0^u\left(z+z^{-1}\right)+b_1^u\left(z^3+z^{-3}\right)+..+b_{N-1}^u\left(z^{2N-1}+z^{-2N+1}\right)\right\}. \tag{6}$$

Similarly, $H_u(z)$ must has zero at $\omega=\pi$, i.e., at $z=-1$. It can be easily found [8, 13] that this condition is satisfied when

$$\sum_{i=0}^{N-1} b_i^u=\frac{1}{4}. \tag{7}$$

When the condition (7) is satisfied, $H_u(1)=1$ and that means the prediction filter has gain 1 at $\omega=0$.

Using the generalization of the lifting scheme (4)- (7) and a simplified lifting coefficients' representation [13], we proposed in [14] the following recursive representation for lifting FIR filter coefficients of order $2\tilde{N}$

$$b_0^p=-\frac{128+B_p}{256},\ b_1^p=\frac{B_p}{256}-b_2^p,\ b_i^p=-\frac{b_{i-1}^p}{C_p}-b_{i+1}^p \tag{8}$$

$$b_0^u=\frac{64-B_u}{256},\ b_1^u=\frac{B_u}{256}-b_2^u,\ b_i^u=-\frac{b_{i-1}^u}{C_u}-b_{i+1}^u \tag{9}$$

where B_p, B_u, C_p, C_u are the parameters that control the DWT properties. The correspondences between these control parameters B_p, B_u, and the conventional (non-lifted) biorthogonal wavelet filters can be found in reference [13]. In the formulas (8), (9) the parameters B_p, B_u control the width of the transition bands and the parameters C_p, C_u control the smoothness of the pass and stop bands to prevent the appearance of the lateral lobes: with greater values of B_p, B_u the values of

C_p, C_u tend to be greater [8]. In practice, one can use predictor (4) and update filter (6) with $\tilde{N}=6$, $N=6$, $B_p=20, B_u=8$, $C_p=6, C_u=6$ to achieve narrow transition bands and good compression rate [8, 14].

3 IIR Lifting Scheme

Considering generalized lifting scheme (4), (6) that these all-zeros systems can be modified to obtain rational transfer functions of a special form containing zeros and poles as following [14]:

$$H_p(z)=\frac{1+b_0^p\left(z+z^{-1}\right)+b_1^p\left(z^3+z^{-3}\right)+\ldots+b_{\tilde{N}-1}^p\left(z^{2\tilde{N}-1}+z^{-2\tilde{N}+1}\right)}{1+a_2^p z^{-2}+a_4^p z^{-4}}, \tag{10}$$

$$H_u(z)=\frac{1+H_p(z)\left\{b_0^u\left(z+z^{-1}\right)+b_1^u\left(z^3+z^{-3}\right)+..+b_{N-1}^u\left(z^{2N-1}+z^{-2N+1}\right)\right\}}{1+a_2^u z^{-2}+a_4^u z^{-4}}. \tag{11}$$

A specific condition to lifting predictor is that it must have a fixed gain to fulfill condition (7), i.e., to prevent bias in the output of the update filer at $\omega=0$. This can be done introducing normalization by factor $1-a_{2p}-a_{4p}-\ldots$ in (11) [14]:

$$H_u(z)=\frac{1+H_p(z)\left\{u_0\left(z+z^{-1}\right)+u_1\left(z^3+z^{-3}\right)+..+u_{N-1}\left(z^{2N-1}+z^{-2N+1}\right)\right\}}{\left(1-a_{2p}-a_{4p}\right)\left(1+a_{2u}z^{-2}+a_{4u}z^{-4}\right)}. \tag{12}$$

Another problem arises when implementing inverse transform with IIR lifting. The wavelet analysis/synthesis filters must provide the perfect restoration of the original data that is especially important for lossless data compression. In the traditional dyadic wavelet decompositions/restorations technique, special care is took to design orthonormal filter banks where each filter satisfies Nyquist constraint $\left|H_k\left(e^{j\omega}\right)\right|_{\downarrow 2}=1$ [9]. In difference, the lifting scheme has a potential to design biorthogonal IIR wavelet filters in simpler way: in the restoration stage, one can use inverse predictor and inverse update filter that operates upon rearranging the input signal elements (wavelet coefficients) backward and then filtering them with the inverse filters for synthesis and next time performing rearranging of the data $\{\cdot\}^B$ [14].

Next, we want to proceed with integer calculus whereas it is possible. For this purpose, we use the representation of normalized coefficients a_i^p, a_i^u [14]:

$$a_i^p=\frac{A_i^p}{256},\ a_i^u=\frac{A_i^u}{256}. \tag{26}$$

Taking into account all before mentioned results and restrictions, we can formulate the integer-to-integer IIR lifting steps [14] tat uses sixth order FIR with $\tilde{N}=6$,

$N = 6$, $B_p = 20, B_u = 8$, $C_p = 6, C_u = 6$ and special form of second order IIR filters with $a_1 = 0$ as following.

Analysis stage:

- prediction:

$$d_i^{(k)} = d_i^{(k-1)} + \left\lfloor \frac{A_2^p d_{i-2}^{(k)} - 10\left(s_{i-1}^{(k-1)} + s_{i+1}^{(k-1)}\right) + 23.3\left(s_{i-3}^{(k-1)} + s_{i+3}^{(k-1)}\right) - 3.3\left(s_{i-5}^{(k-1)} + s_{i+5}^{(k-1)}\right) + \ldots}{256} \right\rfloor \quad (17)$$

- update:

$$s_i^{(k)} = s_i^{(k-1)} + \left\lfloor \frac{A_2^u s_{i-2}^{(k)}}{256} + \frac{56\left(d_{i-1}^{(k)} + d_{i+1}^{(k)}\right) + 9.3\left(d_{i-3}^{(k)} + d_{i+3}^{(k)}\right) - 1.3\left(d_{i-5}^{(k)} + d_{i+5}^{(k)}\right) - \ldots}{256 - A_2^u} \right\rfloor \quad (18)$$

In formulas (17),(18), $\lfloor \cdot \rfloor$ denotes the operation of rounding to the nearest lower integer value.

A_2^p, A_2^u are adjusted in such a manner that the filters (17), (18) match to the spectral properties of the image data to minimize the well known first order entropy of the wavelet coefficients [14]

$$\min_{\vec{A}^p, \vec{A}^u} \left\{ H(\vec{d}) = -\sum_i p_i \log(p_i) \right\}, \quad (19)$$

where p_i denotes the probability of the different values of wavelet coefficients $\vec{d}$. More specifically, this criterion must be modified to take into account the method of wavelet coefficient codification. Usually, to code the wavelet coefficients, some zero-free method, such as EZW, SPIHT etc. is used. It means that the total quantity of information defined by the first order entropy is reduced by $p_0 \log(p_0)$, and it makes DWT-based coders to be an efficient data compression tool. Thus, A_2^p, A_2^u must be adjusted to achieve the minimum of the modified entropy at each k-th level of DWT decomposition

$$\min_{A_2^{p(k)}, A_2^{u(k)}} \left\{ \tilde{H}(d^{(k)}) = -\sum_{i>1} p_i \log(p_i) \right\}; \quad (20)$$

Thus, one can obtain the optimal solution finding

$$\min_{\left\{A_2^{p(k)}\right\}, \left\{A_2^{u(k)}\right\}} \left\{ \sum_{k=1}^{K-1} \tilde{H}(\vec{d}^{(k)}) + H(\vec{d}^{(K)}) + H(\vec{s}) \right\}; \quad (21)$$

where K denotes the number of DWT decompositions, $H(\bar{d}^{(K)})$ is an entropy of wavelet coefficients at the highest level of decomposition, and $H(\bar{s})$ is an entropy of resulted approximations. Unfortunately, it is difficult to obtain an analytical solution for the optimal $\{A_2^{p(k)}\}\{A_2^{u(k)}\}$ and one can find these values by simulations.

4 Experimental Results

The described in the previous section algorithm were tested on a set of 512x512 standard images "Lena", "Baboon", "Barbara", "Boats", "Goldhill", "Peppers", "Bridge" shown in Fig. 1 (these images are available, for example, at http://sipi.usc.edu/database/).

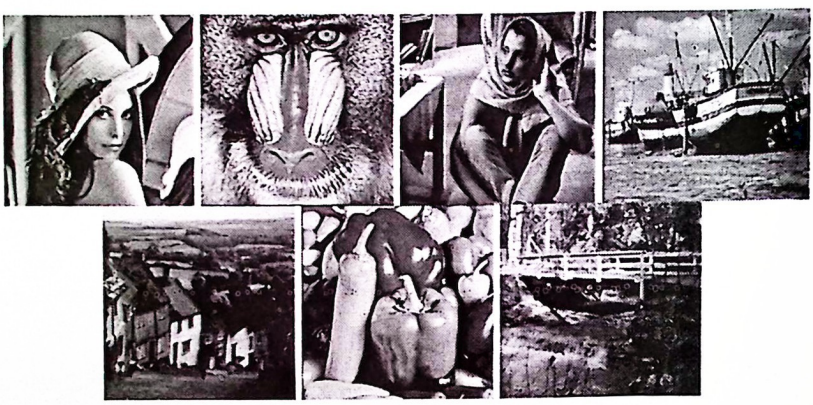


Fig. 1. Set of standard test images: "Lena", "Baboon", "Barbara", "Boats", "Goldhill", "Peppers", "Bridge"

Table presents the entropy values in bits per pixel (bpp) obtained for these images by applying standard lifting decomposition (1) – (3) and CDF(1,1) wavelet (Haar wavelet) with $\tilde{N}=1$, $N=1$, $B_p=0$, $B_u=0$, CDF(2,2) wavelet with $\tilde{N}=2$, $N=2$, $B_p=16$, $B_u=8$ (this wavelet is used by JPEG2000 for lossless image compression), and IIR lifting (17),(18) with optimal A_2^p, A_2^u values. In this paper, we assume that $\{A_2^{p(k)}=A_2^p,\ k=\overline{1,K}\}\{A_2^{u(k)}=A_2^u,\ k=\overline{1,K}\}$. In all cases, the number of decomposition level was $K=6$.

Analyzing the simulation results presented in Table, one can conclude that the proposed IIR lifting transform performs better, providing lower entropy values for all test images in comparison to the FIR lifting. Varying IIR coefficients A_2^p, A_2^u, one

can obtain higher data compression. In all cases, the IIR technique gives the best compression results.

Table. Entropy values in bpp for different lifting filters.

Technique	Image						
	Baboon	Lena	Barbara	Boat	Bridge	Peppers	Goldhill
CDF(1,1) lifting	5.896	3.803	4.566	4.044	4.232	4.237	4.427
CDF(2,2) Lifting	5.876	3.771	4.404	4.036	4.231	4.231	4.424
Proposed FIR only							
$A_2^p = 0, A_2^u = 0$	5.877	3.781	4.365	4.044	4.240	4.24	4.434,B_p=20 4.422,B_p=8 4.423,B_p=20
Proposed IIR Lifting	5.866	3.761	4.343	4.016	4.222	4.205	4.418,B_p=8
Optimal A_2^p	34	18	36	40	16	-26	17
Optimal A_2^u	11	12	12	2	11	17	-4

5 Conclusions

The algorithm of data-dependent DWT based on the generalized IIR lifting scheme was presented. The proposed algorithm requires only two additional integer sums and one floating point multiplication per pixel in comparison to the standard lifting decomposition. The presented results show that the derived algorithm provides lossless image compression and higher data compression rate comparing to the standard wavelet lifting technique. One can expect even better energy compaction, and, thus, higher compression rate with the presented algorithm using IIR filters of higher order and optimizing the IIR filter coefficients $\{A_2^{p(k)}\}\{A_2^{u(k)}\}$ at each level of decomposition and FIR filter parameters $\tilde{N}^{(k)}$, $N^{(k)}$, $B_p^{(k)}, B_u^{(k)}$, $C_p^{(k)}, C_u^{(k)}$ for each k -th level as well. This aspect is a subject of future work.

Acknowledgements

This work was supported by by Instituto Politécnico Nacional as a part of the research project SIP#20071380.

References

1. S.G.Mallat, A theory for multiresolution signal decomposition: The wavelet representation, *IEEE Trans. Pattern Rec ognition. Machine Intell., Vol. 1* , No. 7, p.p.674-693, 1989.
2. Martin Vetterli, Multi-dimensional sub-band coding: some theory and algorithms, *Signal Processing* , *Vol. 6*, pp.97-112, 1984.
3. I.Daubechies, Orthonormal bases of compactly supported wavelets, *Commun. Pure Appl. Math* ., *Vol. 41* , p.p. 909-996, Nov. 1998.
4. W.Sweldens, The lifting scheme: A new philosophy in biorthogonal wavelet constructions, *Wavelet Applications in Signal and Image Processing III* , *A.F.Laine and M.Unser, editors* ,*Proc. SPIE* p.p. 68-79, 1995.
5. R.Claypoole, R.Baraniuk, and R.Nowak, Adaptive wavelet transforms via lifting, *Proc. IEEE Int. Conf. Acoust., Speech, and Signal Proc.* , May 1998.
6. G. Piella and H. J. A. M. Heijmans, "An adaptive update lifting scheme with perfect reconstruction," *Proc. ICIP* , pp. 190 - 193, October 2001.
7. G.C.K. Abhayaratne, "Spatially adaptive wavelet transforms: An optimal interpolation approach", Third International Workshop on Spectral Methods and Multirate Signal Processing (SMMSP) 2003, pp. 155-162, Barcelona, Spain, 12-13 Sept. 2003.
8. Oleksiy Pogrebnyak, Pablo Manrique Ramírez "Adaptive wavelet transform for image compression applications" Proc. SPIE Vol.5203, Applications of Digital Image Processing XXVI, Andrew G. Tescher Chair/Editor, 5- 8 August 2003, San Diego, USA. p.p. 623-630, ISBN 0-8194-5076-6 , ISSN: 0277-786X.
9. P. P. Vaidyanathan and S. Akkarakaran, ``A review of the theory and applications of principal component filter banks," Applied and Computational Harmonic Analysis (invited paper), vol. 10, no. 3, pp. 254-289, May 2001.
10. J. O. Chapa and R. M. Rao, "Algorithms for designing wavelets to match a specified signal," *IEEE Trans. Signal Processing,* vol. 48, pp. 3395 - 3406, December 2000.
11. P. P. Vaidyanathan, ``Theory of optimal orthonormal subband coders", IEEE Trans. Signal Proc., vol. SP. 46, pp. 1528--1543, June 1998.
12. P. Moulin, M. Anitescu, and K. Ramchandran, "Theory of rate-distortion-optimal, constrained filterbanks--Application to IIR and FIR biorthogonal designs," *IEEE Trans. Signal Processing,* vol. 48, pp. 1120 - 1132, April 2000
13. Hoon Yoo and Jechang Jeong, A Unified Framework for Wavelet Transform Based on The Lifting Scheme, *Proc. of IEEE International C onference on Image Processing ICIP* Tessaloniki, Greece, October 7-10, p.p.793-795, 2001.
14. Oleksiy Pogrebnyak, Pablo Manrique Ramírez, Luis Pastor Sanchez Fernandez, Roberto Sánchez Luna. "Data Dependent Wavelet Filtering for Lossless Image Compression". Lecture Notes in Computer Science, Vol. 3773 "Progress in Pattern Recognition, Image Analysis and Applications" CIARP2005, p.p. 285-294, Germany, Springler-Verlag Berlin Heidelberg, 2005, **ISSN 0302-9743,** DOI:10.1007/11578079.

Attitude Subsystem Development for an Educative Satellite Based on Reaction/ Momentum Wheel and Magnetic Torquing Coils*

Esaú Vicente-Vivas[1], Emilio Jiménez[2], Rodrigo Alva[2], and Rodrigo Córdova[2].

[1]Instituto de Ingeniería, Universidad Nacional Autónoma de México, Coyoacan, México DF, evv@servidor.unam.mx
[2]Thesis students, Instituto de Ingeniería, Universidad Nacional Autónoma de México, Coyoacan, México DF,
{factore8, roro_send, jrodrigo_cordova }@hotmail.com

Abstract. This paper shows the obtained results on the development of the attitude subsystem for a human resources training system at laboratory level in the satellite technology field. The educative satellite (ES) integrates several subsystems such as: structure, power, short range wireless communications, digital voice, sensors, flight computer and attitude control. The last constitutes the subject of this paper and is required to demonstrate both automatic maneuvering and automatic stabilization functions capabilities on the educative satellite. Two active stabilization approaches are employed for this subsystem, the first one formed by a set of six magnetic torquing coils that would provide the ES with light stabilization forces, and a versatile second one constituted by a reaction/momentum wheel which render stronger forces for maneuvering control purposes. The exposed system was projected as an intelligent stabilization module and therefore includes a dedicated microcontroller to perform both stabilizations tasks and communications with the ES flight computer. A description of the subsystem components, architecture, operating modes, control functions and preliminary stabilization results obtained in laboratory are presented.

Keywords: Attitude control system, educative satellite, training system, momentum wheel, magnetic coils.

1 Introduction

The reaction/momentum wheels are attractive because by adding or removing energy from a reaction/momentum wheel (flywheel), torque is applied to a single axis of a spacecraft, causing it to react by rotating.

---

* Proyecto 52297 financiado por CONACYT

© L. Sánchez, O. Pogrebnyak and E. Rubio (Eds.)
Industrial Informatics
Research in Computing Science 31, 2007, pp. 133-142

By maintaining flywheel rotation, called momentum, a single axis of a spacecraft can be stabilized. Consequently, several reaction/momentum wheels can be used to provide full three-axis attitude control and stability in a space vehicle, [1].

The attitude control and stability in satellites is very important to perform payload operation. In LEO remote sensing satellites it is required to allow camera pointing to specific regions of our planet in order to acquire high resolution digital images. The bigger the capabilities of the satellite to fine control the spacecraft pointing the better the results from high resolution digital cameras will be obtained. In the case of GEO communications satellites its pointing capabilities define its potential to transmit high bandwidth data with an efficient use of electrical power.

In commercial satellites the use of reaction/momentum wheels is necessary, however, in spacecrafts with masses under 60 Kgs this approach was very rare in the past decade. Mainly because of financial reasons, because the space rated mechanical subsystems were very expensive for this type of projects, [2].

For those reasons, most of the real small satellites (nanosatellites and picosatellites) employ passive stabilization schemes such as magnetic torquing bars. Some of them use active stabilization through magnetic torquing coils. Though, complex stabilization schemes as reaction/momentum wheels are very few in this type of missions, [3].

On the other hand, the last years and particularly in the picosatellite field (1 Kg of mass), a couple of missions are being developed with reaction/momentum wheels employing commercial-off-the-shelf (COTS) general purpose motors [4]. Taking into consideration that the small satellite field is remarkable by its huge potential to validate new technologies, it is possible to affirm that in the near future more small satellite missions will be instrumented with this versatile stabilization scheme.

For those reasons it is very important for the ES to include intelligent satellite stabilization hardware to allow satellite maneuvering demonstration capabilities. This eventually would conduct to elaborate tasks such as payload pointing towards specific targets and more important, to allow the users of the educative system to learn and understand the use of this important satellite resource.

It is important to mention that the ES is intended to be used as an educative tool in laboratories from High schools, Technological Institutes and Universities. Moreover, it is also planed to be employed in research laboratories to develop new solutions for real satellite subsystems. In this sense research in fields such as three axis stabilization, digital communications, satellite sensors, power systems, payload validation, flight computers, navigation autonomy, and satellite constellations, would be addressed with the support of this laboratory tool, [5].

Up to now we have detected only one commercial educative satellite system. It was originally developed by the US Air force and commercialized by Colorado Satellite Services, [6], at a starting price of 8,000 US dollars.

2 Architecture of the stabilization subsystem for the educative satellite

The goal of the ES is to emulate a satellite stabilization system by means of two active methods. One is the reaction/momentum wheel and the second one is a set of magnetic torquing coils. They will allow the studying and experimentation of ES behavior when changes or control be applied in its system actuators. Then that experience and knowledge would take us to the design and exploration of different stabilization control schemes.

The stabilization subsystem (SS) is made up of a flywheel driven by a DC motor obtained from a videocassette player and a set of six magnetic torquing coils, two different coils (coarse and fine) for each one of the ES structure axis. Three coils apply a coarse momentum while the other three provide fine momentum forces. The dedicated control is given by a PIC18F4431 microcontroller that is connected through serial port with the ES flight computer, this microcontroller unit (MCU) receives and processes commands from the flight computer. The command and protocol software was inherited from software developed in our laboratory for a 50 Kg microsatellite mission, [7]. The chosen PIC device has enough resources and capabilities to accomplish the tasks for this subsystem. In addition, the SS contains the electronic control interfaces between the microcontroller and the active stabilization actuators. They are composed by an H bridge for the motor (TA7291S from Toshiba) and six further H bridges (3 L293DD integrated circuits from ST Electronics) driven as hardware interface to control the magnetic torquing coils. Those parts were employed in protoboard tests and will also be used in the PCB stabilizing system for the educative satellite.

In order to carry out the motor control it was added a 1 pulse per round-trip encoder mounted in the motor along with the flywheel. This serves as feedback to the microcontroller. However the goal for the project is to increase the capabilities of the encoder to at least 16 pulses per round-trip in order to increase the control capacities. Besides, in case the motor could become obstructed it was aggregated an overcurrent circuit to protect both the motor and its driving H bridge. It was also aggregated a set of LEDs that indicate the behavior of control signals applied to the stabilization actuators.

3 Stabilization microcontroller

The chosen microcontroller is the surface mount 44 pin, PIC18F4431. This is an 8 bit CPU, 10 MIPS (Million of instructions per second) with 4 crystal modes of up to 40 MHz, 16 Kb flash memory, 768 bytes of RAM, 256 bytes of EEPROM memory, 36 I/O pins, low power consumption mode, in-circuit programming capabilities, software programming mode under serial mode control and, capabilities to drive several levels of interruption priorities. Furthermore, the MCU was selected by its dedicated hardware to generate Pulse Width Modulation (PWM) signals that in our case are employed to control the motor speed as well as to regulate the magnetic field intensity applied to the magnetic torquing coils. The PWM module has a resolution of up to 14 bits and was designed to provide a pair of fault bits with input gates at 2 pins of the MCU. One of them is dedicated to the SS to protect the motor by overcurrent. In cases like this, when one or both of the input pins become activated, all of the PWM outputs become turned off automatically.

The MCU has also a dedicated 3 input capture module intended for movement feedback purposes, which in our case is employed to drive the motor feedback. In order to communicate the MCU with the ES flight computer a serial port from the MCU is utilized for command reception. In addition, the MCU contains further hardware such as an I2C and an SPI ports, 4 timers and a high speed analog to digital converter.

4 Stabilization control approach with a reaction/momentum wheel

A reaction/momentum wheel is a stabilization active element destined to operate at high speeds (several thousands of RPM). When operated into a satellite its goal is to generate a resistant force against position changes produced by external disturbances coming from solar wind, microgravity changes and so on.

With the resources implanted in the ES a closed-loop control mode can be driven and also three different signals are used to allow control in open-loop mode.
Figure 1 shows the laboratory tests applied to the ES reaction wheel, while figure 2 shows the mockup of the satellite educative system at real scale.

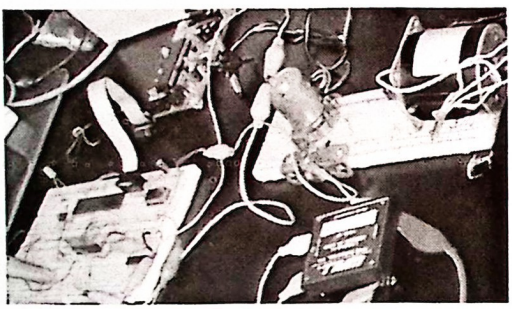


Fig. 1. Flywheel under laboratory Testing.

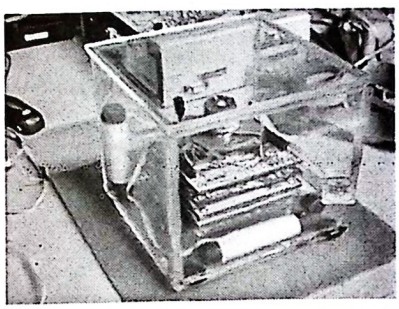


Fig. 2. EducativeSatellite mockup at 100 % scale.

4.1 Open-loop stabilization control with momentum wheel

The open-loop control for the reaction/momentum wheel, is performed without feedback to the MCU. In this case 3 different signals are generated (sine, triangular and swath tooth) by the MCU through a PWM channel to drive the motor speed. The signals will be employed later to characterize the ES attitude control behavior.

4.2 Closed-loop stabilization control with momentum wheel

The closed-loop control of the reaction/momentum wheel is performed to establish the direction of rotation as well as the speed of the flywheel according with commands sent by the flight computer. The digital control performed by the microcontroller gets a feedback signal

from the motor encoder. In addition, a proportional integral (PI) control algorithm specially designed for speed control is executed in the microcontroller.

The global closed-loop control scheme employed to control the flywheel is shown in figure, 3.

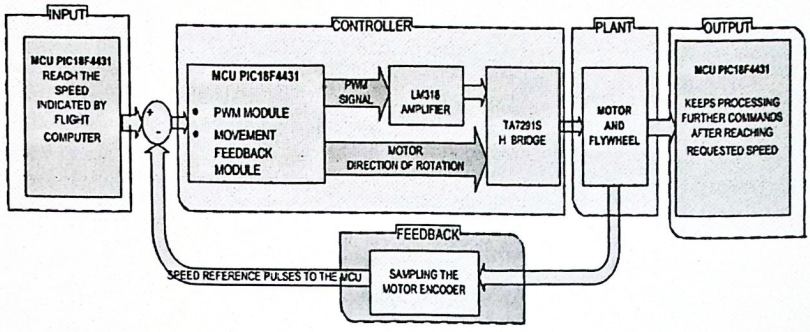


Fig. 3. Closed-loop scheme employed to control the reaction wheel.

Once the flight computer command reaches the MCU with specific data about flywheel speed, the command is executed by the MCU through a PI algorithm.

Testing procedures were applied with a personal computer running software to emulate the flight computer, figure 4. This software allows the user to select the command to be sent to the MCU (flight computer emulation) as well as the command preferences. The software admits the programming of control functions for the stabilizing system with only one command. This one includes specifications to drive the motor and therefore the flywheel, the magnetic torquing coils or any combination of them. As shown in figure 4 the stabilization command allows the user to define the motor control. For closed-loop operation the user selects the "velocidad" option and then the software enables the window titled "Velocidad (rpms)" where the user writes the RPM value to be controlled by the MCU. In the case of open-loop control the user selects the choice "Fijar CT a motor" and a left window receives the working cycle to be employed by the PWM module of MCU device. In the case of magnetic torquing coils the software can be used to control three axis of the educative satellite at the same time, but with only one coil per axis at a time.

In this way, with only one command the flight computer, and in this case the user, can test the whole stabilization system (flywheel and magnetic torquing coils) with any combination of preferences. Once the command is fully configured the user presses the button "ENVIAR" which sends the 14 byte command to the stabilization system. This command is then displayed in hexadecimals values in a window shown in the lower side of figure 4. In addition, with the window placed in the upper left side of figure 4, the software shows the answer delivered by the MCU as an answer to the received command.
From tests performed at laboratory, it is mentioned that the obtained closed-loop control is good when controlling the flywheel at high speeds. In addition, under this mode few other functions were integrated and worked successfully such as: turning off the motor, slowing

down the motor speed and, establishing a fixed working cycle in the direction of rotation. Every software function was individually tested and then the software was integrated to the global MCU software. It should be mentioned that the closed-loop operation at low speeds has deficiencies, therefore work is performed to enhance this behavior.

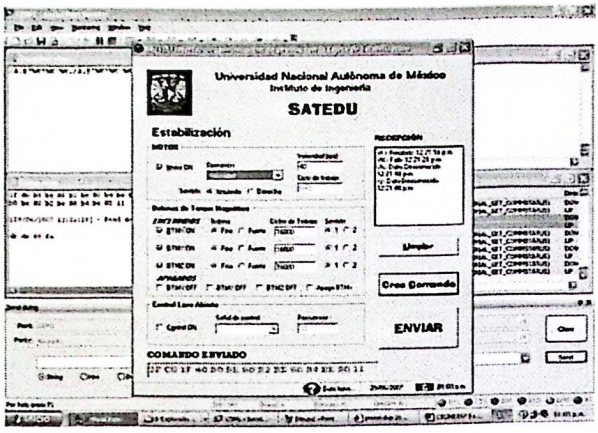


Fig. 4. Software employed to validate the control procedures elaborated for the flywheel, MTCs and its operative software.

In order to implement the PI control it is required to elaborate the mathematical model of the plant, to reach with this an optimal control for the system. In our case the plant is conformed by the reaction/inertial wheel and the motor. In addition, it should be added the required controller, seen this also as another well defined mathematical model. Therefore, with both of the mathematical models would be possible to achieve the corresponding stability and the frequency response analysis for tuning the controller. This process would take us to find values for the proportional and integral constants of the system.

However, in the case of the ES stabilization system we do not have enough information to create the suitable mathematical model. The experimental identification will be employed to obtain an empirical mathematical model. However as a beginning in this first working phase some values for the PI constants were assumed to implant the PI algorithm for the testing model within the MCU. This solution is acceptable taking into account that the flywheel will be employed mainly at high RPMs. Once the ES prototype is concluded, the experimental identification model will be used to obtain a mathematical model for the stabilization system. In this sense we will consider to employ an experimental method to adjust the controller constants. Specifically, the Ziegler-Nichols method to obtain a system response curve is being considered. With this data and based on certain characteristics of the curve it will be possible to obtain the value of the PI constants. However, to experimentally obtain this information, a strong experimental effort has to be made in order to get, acquire, store and process the information with the help of a MCU. Another solution to consider is related with the final motors that will be employed in the ES, they are general purpose motors. We will ask the manufacturer to provide the information to model the motor. Therefore this information will be discussed in a future publication.

Another reason by which the PI control does not work correctly is the flywheel encoder, because it only provides one pulse per round. This causes the MCU registers to overflow due to the existing long times between pulses. This problem can be solved by using a better flywheel with an encoder capable to generate at least 16 pulses per round, as those available for general purpose motors.

When the MCU finishes the PI calculation what it does is to fix a specific output to the PWM module based on that calculation. Then the motor will change its speed and therefore feedsback this process to the MCU through the encoder signal attached to the flywheel (which as said before is connected to the capture module of the MCU movement feedback). The last process closes the loop and takes the flywheel to the specified speed indicated by the flight computer command.

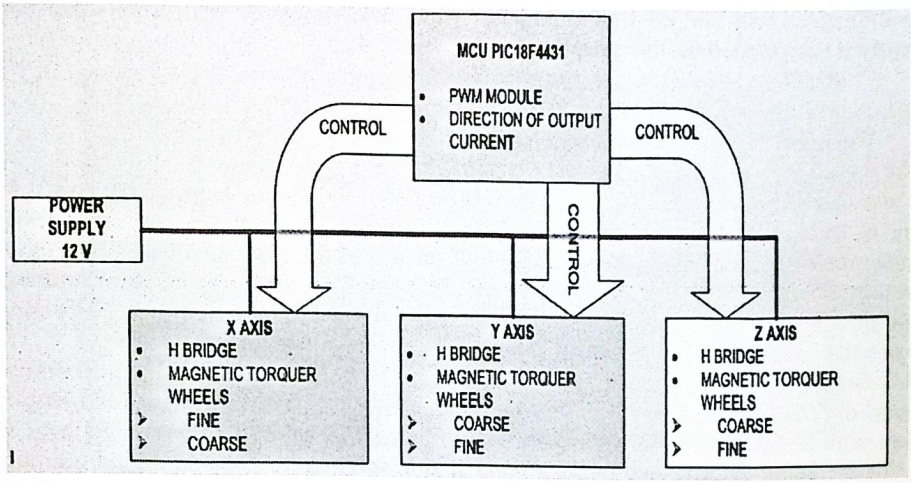


Fig. 5. Control of magnetic torquing coils through PWM performed by the ES MCU.

The magnetic torquing coils (MTC) as well as the momentum wheel are active stabilization elements. They are very important resources to develop complex 3-axis stabilized satellites as well as for stabilization of propelled space vehicles. In systems like these, the thrusting errors could cause the satellites to translate and this translation would be nearly impossible to correct, [8]. Both the reaction wheel and the MTC are highly controllable stabilizing media. One disadvantage of the reaction wheel is that it can become saturated, in which case momentum dumping would be required. However, this can be compensated by designing a wheel with a larger inertia than necessary. Another way to solve the problem would be the employment of the MTC to desaturate the reaction wheel. The MTC take advantage of the Earth magnetic fields to generate fine forces that orient a satellite in a desired position. Furthermore, the control forces are enhanced by employing ferrite nucleus coils, and when a magnetic field crosses them they generate a force according with the Faraday Law, emulating in certain way the principle of an electrical motor.

Under this stabilization method 6 coils from ES are driven by 6 PWM channels from MCU that control the magnetic field strength generated by them, as shown in figure 5.
In the case of the magnetic torquing coils, different tests were performed employing hard disk coils, since they closely emulate a real satellite MTC. The hard disk coils are attached to a good mechanical design and permanently exposed to a magnetic field generated by a magnet. The performed tests included: the establishment of the working cycle for each coil, the change of its working cycle, exchanging the use among them and the turning off procedure for each one of the coils, see figure 5.

5 Stabilization control with magnetic torquing coils

In addition, all the validated software processes were integrated in the main MCU software. Then the global stabilization software was successfully tested in laboratory employing communications with the ES flight computer which was emulated with a personal computer and software developed for this specific purpose, figure 4.

6 Printed circuit board for the stabilization subsystem

The two layer printed circuit for the stabilization board has dimensions of 89 mm x 89 mm, figure 6. It was elaborated with Protel DXP software, a powerful tool for design and manufacture of printed circuits boards. It allows the use of different data bases for components as well as the creation of new components and the possibility to aggregate them to the project. In addition, it enables the positioning of components and the automatic routing of connections between the components of the project in a very short time and in an efficient way.
In this design we had to consider several track widths because the stabilization components demand different intensities of current. Figure 7 shows the artistic view of the stabilization system with integrated flywheel elaborated with SOLID EDGE18 software. Several designs like this are being incorporated in the ground station software to provide digital animations in the visualization interface of the ES ground station.

7 Communications between stabilization board and the ES flight computer

The communications between the flight computer and the stabilization board are achieved by means of serial port. The flight computer sends command lines that contain 14 bytes, figure 8. The line includes a header, command specification and 1 byte of checksum, [7]. The last is aggregated to avoid information losses by effects of external noise or electrical interference. Once the command is received, the MCU becomes awaked and then stores it and verifies its integrity through the checksum. Then, it sends an acknowledge byte to the flight computer indicating if the command was okay or not. In this way the flight computer can decide, respectively, to finish or to resend the command to the MCU. Once the command is received by the MCU it proceeds to execute the orders given by the flight computer.

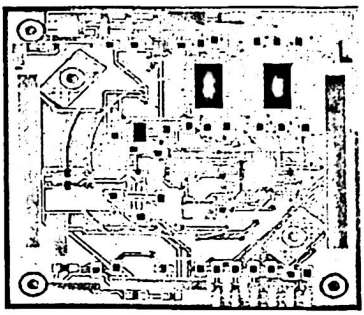


Fig. 6. PCB designed for the stabilization system.

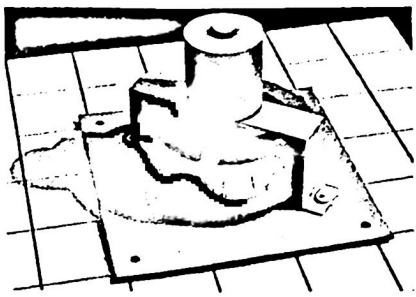


Fig. 7. Stabilization system view generated with SOLID EDGE18.

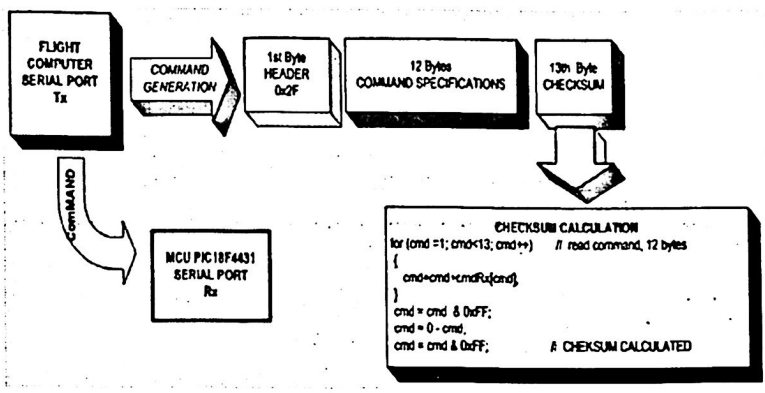


Fig. 8. Digital communications among MCU and flight computer.

8 Generated results and concluding remarks

The stabilization module for an educative satellite has been fully and successfully tested in laboratory protoboards. The control motor tests were performed in open-loop fashion applying 3 different type of signals (sine, triangular and swath tooth) to the motor. On the other hand the closed-loop control is good enough when controlling the flywheel at high speeds. In addition, under this mode few other functions were integrated and worked successfully such as: turning off the motor, slowing down the motor speed and, establishing a fixed working cycle in the direction of rotation. Every software function was individually tested and then the software was integrated to the global MCU software. It should be mentioned that the closed-loop operation at low speeds has deficiencies, therefore work is being performed to eliminate this behavior.

In the case of the magnetic torquing coils, different tests were performed employing hard disk coils, since they closely emulates a real satellite MTC. The hard disk coils are attached to a

good mechanical design and permanently exposed to a magnetic field generated by a magnet. The performed tests included: the determination of the working cycle for each coil, the change of its working cycle, exchanging the use among them and the turning off procedure for each one of the coils.

In addition, all the validated software processes were integrated in the main MCU software. Then the global stabilization software was successfully tested in laboratory employing communications with the ES flight computer which was emulated with a personal computer and software developed for this particular purpose. It is important to notice that the PCB for the ES flight computer is also about to be sent for fabrication and therefore was not ready for these testing procedures.

It is expected for the next months to have finished the PCB versions for at least: the stabilization subsystem, the flight computer, the voice card and the power subsystem. Therefore the integration procedures and preliminary ES tests and results are also expected to be finished for the end of 2007.

References

1. Xin-Sheng Ge and Li-Qun Chen, Attitude control of a rigid spacecraft with two momentum wheel actuators using genetic algorithm, Acta Astronautica Journal, Elsevier Science, Volume 55, Issue 1, July 2004, Pages 3-8.
2. "Reaction/Momentum Wheels", Commercial Benefits-Spinoffs, Goddard Space Flight Center, NASA Technical Report, July 2007. http://ntrs.nasa.gov/archive/nasa/casi.ntrs.nasa.gov/20020076147_2002119560.pdf .
3. Lappas et. Al., "Design, Analysis and In-orbit Performance of the BILSAT-1Microsatellite Twin Control Moment Gyroscope Experimental Cluster", AIAA Guidance, Navigation, and Control Conference and Exhibit, Providence, Rhode Island, Aug. 16-19, 2004.
4. Alminde, L., Bisgaard, M., Bhanderi, D., Nielsen, J., Experience and Methodology gained from 4 years of Student Satellite Projects, Proceedings of the Conference on Recent Advances in Space Technology, June, 2005.
5. Vicente-Vivas E., et Al., "Steps towards the development of a portable and cost-effective system for human resources trainning in small satellite technology", Research On Computer Science, Special Issue : Control, Virtual Instrumentation and Digital Systems, Vol. 24, ISSN: 1870-4069, pp. 117-130, Centro de Investigación en Computación, IPN, México, DF, November 2006.
6. D. Barnhart, et al., "Transforming the way students experience space systems Engineering", Proceedings of the 2004 American Society for Engineering Educational Annual Conference and Exposition. http://www.asee.org/acPapers/2004-1880_Final.pdf .
7. Vicente-Vivas E., López Estrada M., Mejía Galeana J.A., "Hardware and Communications Protocols for a Fault Tolerant Distributed Computer Network in the Satex1 Microsatellite System". Proceedings of the 3rd International Symposium on Small Satellites Systems and Services. Annecy, Francia, September 1996.
8. Roberts B., et al., "CamCube Satellite Final Report", ASEN 4018 – Fall 2003, http://www.colorado.edu/ASEN/SrProjects/Archive/2003-04/ReportFall/CamCube. pdf .

Diagonal Recurrent Neural Networks for Speed Control of a DC motor

Alfonso Noriega[1], Carlos A. Silva[1] and Jesús Pichardo[2]

[1] Facultad de Ingeniería, Universidad Autónoma de Querétaro, Querétaro, Qro., México
[2] CICATA-IPN, Querétaro, Qro., México
{Alfonso Noriega, anoriega@uaq.mx}

Abstract. A new control algorithm based on diagonal recurrent neural network (DRNN) is presented. The architecture of DRNN is a modified model of the fully connected recurrent neural network with one hidden layer [1], and the hidden layer is comprised of self-recurrent neurons. Two DRNN's are utilized in a control system, one as an identifier called diagonal recurrent neuroidentifier (DRNI) and the other as a controller called diagonal recurrent neurocontroller (DRNC). A controlled plant is identified by the DRNI which then provides the sensitivity information of the plant to the DRNC. A generalized dynamic backpropagation algorithm (DBP) is developed and used to train both DRNC and DRNI. Due the recurrence, the DRNN can capture the dynamic behavior of the system. The proposed DRNN paradigm is applied to numerical problems and the simulation results for speed control of a DC motor are included.

Keywords: Recurrent Neural Networks, Dynamic Back-propagation, DC Motor, Diagonal Recurrent Neural Networks, Adaptive Control.

I Introduction

The development in the control area has been fueled by three major needs: the need to deal with increasingly complex systems, the need to accomplish increasingly demanding design requirements, and the need to attain these requirements with less precise advanced knowledge of the plant and its environment [2], [3]. Increasingly complex dynamical systems with significant uncertainty have forced system designers to turn away from conventional control methods. However, the fundamental shortcomings of current adaptive control techniques, such as nonlinear control laws which are difficult to derive, geometrically increasing complexity with the number of unknown parameters, and the general unsuitability for real time applications have compelled researchers to look for solutions elsewhere [4].

Several neural network models and neural learning schemes were applied to system controller design during the last three decades, and many promising result are reported [4]-[8]. Most people used the feedforward neural network and the backpropagation training algorithm [4], [9] to solve the dynamical problems; however, the feedforward network is a static mapping. On the other hand, recurrent neural networks [1]-[2] and [7]-[11] have important capabilities not found in feedforward network, such as attractor dynamics and the ability to store information

© L. Sánchez, O. Pogrebnyak and E. Rubio (Eds.)
Industrial Informatics
Research in Computing Science 31, 2007, pp. 143-152

for later use. Thus the recurrent neural network is a dynamic mapping and is better suited for dynamic systems than the feedforward network.

In most control applications, the real-time implementation is very important, and thus the neurocontroller also needs to be designed such that it converge with a relative small number of training cycles. With the objective of a simple recurrent network and a shorter training time for a neural network model, a diagonal recurrent network (DRNN), as shown in figure 1, is developed. This model has considerably fewer weights and the network is simplified considerably.

This paper is organized as follows. In the section II, a DRNN model is developed and a dynamic backpropagation training algorithm is designed to train a DRNN based control system. Finally, in section III the practical relevance of the proposed control schemes is illustrated by simulation for speed control of a DC motor.

II Diagonal recurrent neural networks

Consider Fig. 1, where for each discrete time k, $I_i(k)$ is the ith input, $S_j(k)$ is the sum of inputs to the jth recurrent neuron, $X_j(k)$ is the output for the jth recurrent neuron, and $O(k)$ is the output of the network. Where W^I, W^O, W^D, represents input, output and diagonal weight vectors, respectively.

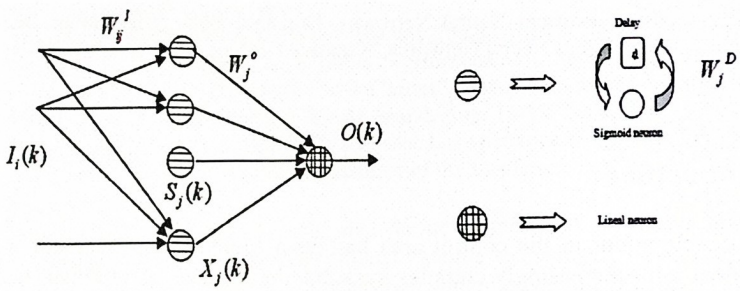


Fig1. Diagonal recurrent neural network

An approach for control and system identification using diagonal recurrent neural networks (DRNN) is presented in this section. An unknown plant is identified by a system identifier, called the diagonal recurrent neuroidentifier (DRNI), which provides information about the plant to a controller, called the diagonal recurrent neurocontroller (DRNC). The neurocontroller is used to drive the unknown dynamic system such that the between plant and desired output is minimized. A generalized algorithm, called the dynamic backpropagation (DBP), is developed to train both DRNC and DRNI. For simplicity, the plant is assumed to be single input/single output system.

Both DRNI and DRNC use the same DRNN architecture shown in Fig 1, which has only one hidden layer with sigmoid type recurrent neurons. The block diagram of the DRNN based control system is shown in Fig. 2. The inputs to the DRNC are the

reference input, the previous plant output, and the previous control signal, and the output of the DRNC is the control signal to the plant. By using the dynamic backpropagation (DBP) algorithm developed in this paper, the weights of the DRNC are adjusted such that the error between the output of the plant and the desired output from a reference model approaches a small value after some training cycles. When the DRNC is in training, the information on the plant is needed. Since the plant is normally unknown, the DRNI is used to estimate the plant sensitivity y_u for the DRNC.

The current control signal generated from de DRNC and previous output of the plant are used as the inputs to the DRNI. The error between the output of the DRNI and plant is computed for each iteration, and is used to adjust the weights of the DRNI. By training the DRNI and DRNC alternately, the weights of the DRNC can be adjusted more effectively.

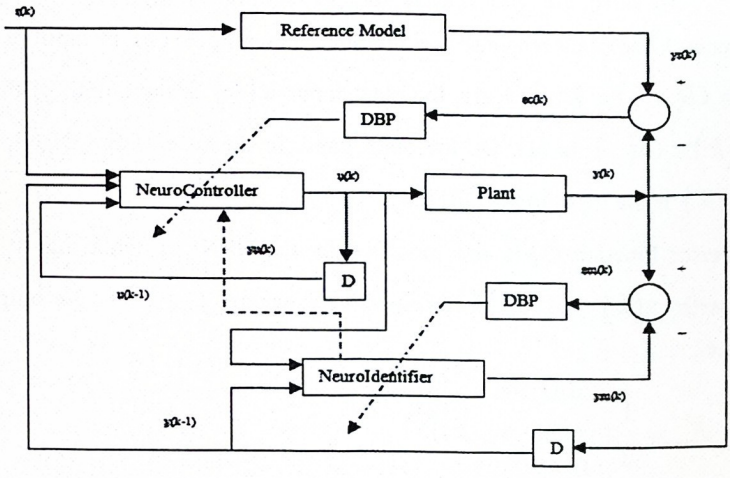


Fig 2. Block diagram of DRNN based control system

II.1 Dynamic backpropagation algorithm for diagonal recurrent neural networks

The mathematical model for the DRNN in Fig. 1 is shown below:

$$O(k) = \sum_j W_j^O X_j(k), \quad X_j(k) = f(S_j(k)) \tag{1}$$

$$S_j(k) = W_j^D X_j(k-1) + \sum_j W_{ij}^I I_i(k) \tag{2}$$

Where for each discrete time k, $I_i(k)$ is the ith input to the DRNN, $S_j(k)$ is the sum of inputs to the jth recurrent neuron, $X_j(k)$ is the output of the jth recurrent neuron, and $O(k)$ is the output of the DRNN. Here $f(\bullet)$ is the usual sigmoid function, and W^I, W^D, W^O are input, recurrent, and output weight vectors, respectively, in $\Re^{n_i}, \Re^{n_d}, \Re^{n_o}$.

Let $y_r(k)$ and $y(k)$ be the desired and actual responses of the plant, then an error function for a training cycle for DRNC can be defined as:

$$E_c = \frac{1}{2}(y_r(k) - y(k))^2 \tag{3}$$

In general, the plant response is a nonlinear mapping $G(\bullet)$ of input $u(k)$, i.e., $y(k) = G(i(), k)$. Here, the plant input $u(k)$ is the output of the DRNC, i.e., $u(k) = O(k)$ in (1). On the other hand, in the case of the DRNI, the plant input $u(k)$ is the input to the DRNI.

The error function (3) is also modified for the DRNI by replacing $y_r(k)$ and $y(k)$ with $y(k)$ and $y_m(k)$, respectively, where $y_m(k)$ is the output of the DRNI, i.e.,

$$E_m = \frac{1}{2}(y(k) - y_m(k))^2 \tag{4}$$

Where $y_m(k) = O(k)$ of (1).

The gradient of error in (3) with respect to an arbitrary weight vector $W \in \Re^n$ is represented by

$$\frac{\partial E_c}{\partial W} = -e_c(k)\frac{\partial y(k)}{\partial W} = -e_c(k)y_u(k)\frac{\partial u(k)}{\partial W} = -e_c(k)y_u(k)\frac{\partial O(k)}{\partial W} \tag{5}$$

Where $e_c(k) = y_r(k) - y(k)$ is the error between the desired and output responses of the plant, and the factor $y_u(k) \equiv \frac{\partial y(k)}{\partial u(k)}$ represents the sensitivity of the plant with respect to its input.

Since the plant is normally unknown, the sensitivity needs to be estimated for the DRNC. However, in the case of the DRNI, the gradient of error in (4) simply becomes

$$\frac{\partial E_m}{\partial W} = -e_m(k)\frac{\partial y_m(k)}{\partial W} = -e_m(k)\frac{\partial O(k)}{\partial W} \tag{6}$$

Where $e_m(k) = y(k) - y_m(k)$ is the error between the plant and the DRNI responses.

The output gradient $\frac{\partial O(k)}{\partial W}$ is common (5) and (6) and needs to be computed for both DRNC and DRNI. Its computation is summarized in the following paragraph: Given the DRNN shown in Fig 1 and described by (1) y (2), the output gradients with respect to output, recurrent, and input weights, respectively, are given by:

$$\frac{\partial O(k)}{\partial W_j^O} = X_j(k) \tag{7a}$$

$$\frac{\partial O(k)}{\partial W_j^D} = W_j^O P_j(k) \tag{7b}$$

$$\frac{\partial O(k)}{\partial W_{ij}^I} = W_j^O Q_{ij}(k) \tag{7c}$$

Where $P_j(k) \equiv \frac{\partial X_j(k)}{\partial W_j^D}$ and $Q_{ij}(k) \equiv \frac{\partial X_j(k)}{\partial W_{ij}^I}$ and satisfy:

$$P_j(k) = f'(S_j)(X_j(k-1) + W_j^D P_j(k-1)), \qquad P_j(0) = 0 \tag{8a}$$

$$Q_{ij}(k) = f'(S_j)(I_i(k) + W_j^D Q_{ij}(k-1)), \qquad Q_{ij}(0) = 0 \tag{8b}$$

From (1), the gradient with respect to the output weight is found as $\frac{\partial O(k)}{\partial W_j^O} = X_j$. Again, from (1), the gradient with respect to the recurrent weight is

$$\frac{\partial O(k)}{\partial W_j^D} = W_j^O \frac{\partial X_j(k)}{\partial W_j^D} = W_j^O P_j(k)$$

From (1) and (2),

$$\frac{\partial X_j(k)}{\partial W_j^D} = \frac{\partial X_j(k)}{\partial S_j(k)} \frac{\partial S_j(k)}{\partial W_j^D} = f'(S_j(k)) \frac{\partial S_j(k)}{\partial W_j^D}$$

and

$$\frac{\partial S_j(k)}{\partial W_j^D} = X_j(k-1) + W_j^D \frac{\partial X_j(k-1)}{\partial W_j^D}$$ Which lead to (8a).

The procedure of deriving the gradient with respect to input weight is similar to the above derivation, and the corresponding equations, (7c) and (8b) follow.

Equations (8a) and (8b) are nonlinear dynamic recursive equations for the state gradients $\frac{\partial X_j(k)}{\partial W}$, and can be solve recursively with given initial conditions. For the usual FNN, the current weight W_j^D is zero and the equations become algebraic.

II.2 Dynamic backpropagation for DRNI

From (6), the negative gradient of the error with respect to a weight vector in $\Re^n$ is

$$-\frac{\partial E_m}{\partial W} = e_m(k) \frac{\partial O(k)}{\partial W} \tag{9}$$

Where the output gradient is given by (7) and (8), and W represents W^D, W^O, or W^I in $\Re^{n_d}$, $\Re^{n_o}$, or $\Re^{n_i}$, respectively.

The weights can now be adjusted following a gradient method, i.e., the update rule of the weights become:

$$W(n+1) = W(n) + \eta\left(-\frac{\partial E_m}{\partial W}\right) \tag{10}$$

Where η is a learning rate. The equations (7)-(10) define the dynamic backpropagation algorithm (DBP) for DRNI.

II.3 Dynamic backpropagation for DRNC

In the case of DRNC, from (5), the negative gradient of the error with respect to a weight vector in $\Re^n$ is

$$-\frac{\partial E_c}{\partial W} = e_c(k) y_u(k) \frac{\partial O(k)}{\partial W} \tag{11}$$

Since the plant is normally unknown, the sensitivity term $y_u(k)$ is unknown. This unknown value can be estimated by using the DRNI. When the DRNI is trained, the dynamic behavior of the DRNI is close to the unknown plant, i.e., $y(k) \approx y_m(k)$ where $y_m(k)$ is the output of the DRNI.

Once the training process is done, we assume the sensitivity can be approximated as

$$y_u(k) \equiv \frac{\partial y(k)}{\partial u(k)} \approx \frac{\partial y_m(k)}{\partial u(k)} \tag{12}$$

Where $u(k)$ is an input to the DRNI.

Applying the chain rule to (12), and noting that $y_m(k) = O(k)$ of (1),

$$\frac{\partial y_m(k)}{\partial u(k)} = \frac{\partial O(k)}{\partial u(k)} = \sum_j \frac{\partial O(k)}{\partial X_j(k)} \frac{\partial X_j(k)}{\partial u(k)} = \sum_j W_j^O \frac{\partial X_j(k)}{\partial u(k)} \tag{13}$$

Also from (1),

$$\frac{\partial X_j(k)}{\partial u(k)} = f'(S_j(k)) \frac{\partial S_j(k)}{\partial u(k)} \tag{14}$$

Since inputs to the DRNI are $u(k)$ and $y(k-1)$ from Fig 2, (2) becomes

$$S_j(k) = W_j^D X_j(k-1) + W_{1j}^I u(k) + W_{2j}^I y(k-1) + W_{3j}^I b_I \tag{15}$$

Where b_I is the bias for DRNI. Thus

$$\frac{\partial S_j(k)}{\partial u(k)} = W_{1j}^I \tag{16}$$

From (13), (14) and (16),

$$y_u(k) = \frac{\partial y_m(k)}{\partial u(k)} = \sum_j W_j^O f'(S_j(k)) W_{1j}^I \tag{17}$$

Where the variables and weights are those found in DRNI.

Using the negative gradients in (11), the weights for DRNC can now be adjusted using the update rule similar to (10). The equations (7), (8), (19), (11), and (17) define the dynamic backpropagation algorithm for DRNC.

III SIMULATION RESULTS

Although the above described algorithm has been tested in many examples, here we will only show the results obtained in one simulated case corresponding to the following model [12] of a motor CD. The transfer function from the armature voltage to the angular velocity is:

$$\frac{W(s)}{V_a(s)} = G(s) = \frac{20.16}{s^2 + 4.244s + 14.34}$$

Discrete the plant G (s) the following is obtained:

Y=(-0.6542*Y_2)+(1.539*Y_1)+(0.0754*U_2)+(0.0869*U_1);

Period of sampling T = 0.1

After the reference is reached a change it is made in the constant of time of the motor obtaining the following model:

$$\frac{W(s)}{V_a(s)} = G(s) = \frac{20.16}{s^2 + 4s + 20}$$

Discrete the plant G (s) with T=0.1 the following is obtained:

Y=(-0.6703*Y_2)+(1.508*Y_1)+(0.07623*U_2)+(0.08716*U_1);

As conditions initials we have:

Ym = 0, Yu = 1, η_I = 0.2, η_C = 0.2

Where η_I, η_C are the constants of learning for neuroidentifier and neurocontroller respectively.

In the figure 3, it is shown the proposed schemes for DRNI and DRNC. The figure 4, illustrates the tracking-reference sequence $y_m(k)$, the plant output and the corresponding adaptive neural control input when the parameter's change occurs at

t_1. It can be observed that the tracking error converges to zero in few sampling periods. The transient period could be reduced by changing the learning rate η at the expense of more input energy.

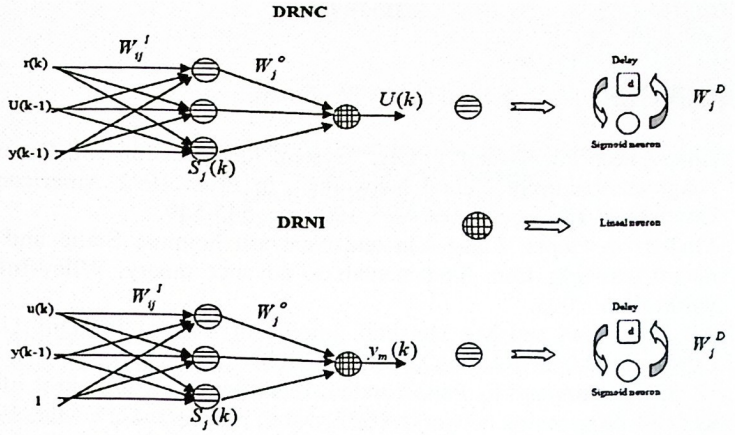


Fig. 3 Schemes DRNN for neuroidentifier and neurocontroller respectively.

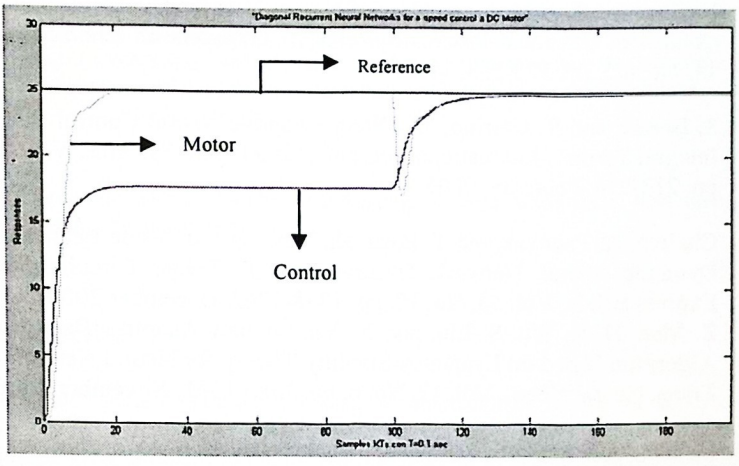


Fig. 4 Responses of the motor, signal of control and the reference

IV CONCLUSION

The control algorithm based on a diagonal recurrent neural network presented in this paper, promise to be a very interesting option for the control of processes with a difficult dynamics that could not be adequately controlled with PID regulators, even

in their self-tuning versions, as it was shown in the realized simulation cases. The class of processes in which the algorithm could be applied is very wide and it includes most of the cases that can appear in practice. In the near future, we plan to apply the algorithm to some real laboratory processes and to extend the obtained results to the case of multivariable and multiconnected systems.

REFERENCES

1. Chao-Cheen ku, Kwang Y. Lee, "System Identification and Control Using Diagonal Recurrent Neural Networks", in Proc. 1992 American Control Conference, Chicago, June 24-26, 1992, pp. 545-549.
2. Madan M. Gupta, Liang Yin and Noriyasu homma, Static and Dinamic neural networks from fundamentals to advance theory, Wiley-Interscience publication, 2003.
3. Ali Zilouchian and Mo Jamshidi, Intelligent Control Systems Using Soft Computing Methodologies, CRC press, 2001.
4. K. S. Narendra and K. Parthasarathy, Identification and control of dynamic systems using neural networks, IEEE, vol. I, No 1, pp. 4-27, Mar. 1990.
5. Richard P. Lippmann. "An Introduction to Computing Whit Neural Nets". IEEE ASSP Magazine, pages. 4-22, April 1987.
6. N. Bhat and T. J. McAvoy, "Use of Neural Nets for Dynamic Modelling and Control of Chemical Process Systems", Computers on Chem. Engng., Vol 14, No. 4/5, pp. 573-583, 1990.
7. S. Baruch and R. Garrido, "A Direct Adaptive Neural Control Scheme with Integral Terms", International Journal of Intelligent Systems, Vol. 20, No. 2, pp. 213-224, February, 2005.
8. Chairez, A. Poznyak and T. Poznyak, New Sliding-Mode Learning Law for Dynamic neural Network Observer, IEEE Trans. Circuits Systems-II: Express Briefs, Vol. 53, No. 12, pp. 1338-1342, December 2006.
9. Z. Man, H. R. Wu, S. Liu, and X. Yu, "A new Adaptive Backpropagation Algorithm Based on Lyapunov Stability Theory for Neural Networks", IEEE Trans. Neural Netw., Vol. 17, No. 6, pp. 1580-1591, November 2006.
10. O. De Jesús and M. T. Hagan, "Backpropagation Algorithm for a Broad Class of Dynamic Networks", IEEE Trans. Neural Netw., Vol. 18, No. 1, pp. 14-27, 2007.
11. Y. D. Jou and F. K. Chen, "Least-Squares Design of FIR Filters Based on a Compacted Feedback Neural Network", IEEE Trans. Circuits Syst. II, Exp. Briefs, Vol. 54, No. 5, pp. 427-431, May., 2007.
12. Khaled Nouri, Rached Dhaouadi, Naceur Benhadj Braiek, Nonlinear speed control of a dc motor drive system with online trained recurrent neural network, IEEE, AMC'06-Istanbul, Turkey, 2006.

A Modified Rete Match Algorithm for Predicate Management in Real Time Planning and Execution Systems

Marcelo Nicoletti Franchin

Electrical Engineering Department, Faculty of Engineering
UNESP – Sao Paulo State University
Av. Luis Edmundo Carrijo Coube 14-01
Bauru – Sao Paulo – CEP 17033-360 – Brazil
franchin@feb.unesp.br

Abstract. Automatic action planning for task accomplishment is one of the AI areas called *AI Planning*, where intense research occurred in the last decade. Among the many categories of planners and diverse problem domains, there are real time planners that consider external events and generate the plan interleaved with the execution system, processing real time collected data (i.e. sensory data) that have influences on the planning process. This paper presents a modified version of RETE match algorithm that is used to instantiate the parameters of prototype predicates and to verify if a predicate is true in a given world state. This modified version was implemented in a real time planner using $ALWAYS_{TRX}$ method. As a result, an efficient system for real time intelligent control was obtained and it is currently used in our research robots.

Keywords: RETE algorithm, AI/Planning, Artificial intelligence, Real time planning, Intelligent systems.

1 Introduction

Automatic action planning for task accomplishment is one of the Artificial Intelligence areas called *AI Planning*, where intense research occurred in the last decades. A great problem in automated reasoning is to design systems that can find automatically a sequence of actions (or a plan) that allow an agent to change the environment (called world) from an initial state to a desired state (goal state). This plan could be delivered, for example, to a manufacturing system, a robot, or any kind of actuator that, following the plan, execute the actions and make the world change to the desired goal state. In artificial intelligence terminology they can be called agents [1][2].

Early AI planning systems appeared in the sixties and at that time some assumptions were created to simplify the planning problem [3][4]. One of the

© L. Sánchez, O. Pogrebnyak and E. Rubio (Eds.)
Industrial Informatics
Research in Computing Science 31, 2007, pp. 153-162

assumptions was considering the agent (a robot, a manufacturing system or some another form of actuator) the only entity responsible for changing of the world state. Another supposition of the planning system was the belief that the changes made in the world by the agent were totally deterministic.

At that time the planning problem was to create techniques and heuristics to search in a tree of possible plans to achieve a desired goal. The complexity of this problem was analyzed years later and considered NP-Complete and NP-Hard [5][6]. Nowadays the complexity was demonstrated to be much harder with variations [2][7].

Considering applications in complex environments with uncertainty and dynamic changes, easily found in real world, like manufacturing systems, mobile robots, software agents, inspection, maintenance, surveillance, etc., the changes in world state are made by multiple agents. If the planner considers that only its agent can change the world, the resulting action plan may not represent the reality and the execution of this plan would be incompatible with the world where the agent is inserted. Therefore, the planning system must use its perception resources to adequate the plan to the eventual contingencies during the plan execution.

Beyond the dynamics and unpredictability of the environment, the agent owns sensors and actuators that do not work in an ideal way and the time available for decisions is very limited. The knowledge base of the agent therefore is incomplete and the planning system must consider this aspect to make possible agent interaction in unstructured environments [8][9][10].

Nowadays, the planning systems evolved to deal with real world applications. A planning system with treatment of external events integrated to the execution is a problem of enormous complexity and demands techniques that optimizes performance of execution to avoid system overhead, otherwise its use will be impracticable in real world environments because response time constraints [11].

Amongst the inherent problems of the planning methods there is the parameter instantiation of predicate prototypes that consists in a combinatorial task that can take a considerable processing time. Moreover, the process of verifying if a predicate is true in the world state consists of comparing the predicate with each one of the predicates that describe the state of the world. The computational cost of both comparisons increases linearly with the amount of facts and exponentially with the amount of changeable parameters of a predicate.

This paper presents an adaptation of RETE match algorithm that was implemented in a real time planner using the method ALWAYS $_{TRX}$.

Section 2 presents the complexity of the problem and related work. Section 3 presents a summary of ALWAYS$_{TRX}$ method where the adaptation of RETE algorithm was used successfully. Section 4 presents the adaptations made in RETE algorithm with an example case study in the blocks world. Finally, in section 5, the conclusions and future works are presented, followed by the bibliographical references.

2 Problem Complexities and Related Work

Early work integrating planning and execution like STRIPS/PLANEX [12][13], Universal Plans [14], IPEM [15], BUMP [16], XII [9], SIPE-2/PRS-CL [17][3] e SAGE [11], as well early works in planning with incomplete information like UCPOP [18], BURIDAN and C-BURIDAN [19], among others, had been collaborated to the development of planning systems but partially solve the problem of plan in dynamic and unstructured environments because they do not consider, altogether, the aspects early presented. More detailed discussions can be seen in [20].

Research in planning systems that handle external events have been developed by our planning group since the eighties. Among the systems developed are: the PETRUS system [21], EXTEPS [22], PBE [23] and finally, the ALWAYS$_{TRX}$, in its early research stage being presented shortly in this paper. The ALWAYS$_{TRX}$ method [24][25] is an effort to create systems that can effectively plan in complex, dynamic and unstructured real world domains.

As cited previously, the problem of parameter instantiation of predicates and the process of logical verification of a predicate in a given state, both have a very high computational cost. Solutions have been found to make possible real time planning systems [26].

A similar problem to the described one in the previous paragraph is the work of find the match in the comparison of a set of patterns with a set of objects, made by the interpreters of rule production systems. They must determine which rules of production must be triggered through the comparison of its logical expressions (conditions to fire a rule) with the facts declared in the working memory.

A fast and efficient algorithm to execute this task is RETE algorithm [27]. RETE algorithm is, at least, many times more efficient than any another algorithm already developed for the task to compare objects with patterns [28][29][2].

Among the best algorithms that work with this kind of problem, there are TREAT [30] and GATOR [31]. The LANA algorithm [32] is a parallel version of RETE algorithm. An extension of RETE algorithm for the management of events with time information is presented in [33]. RETE algorithm still is the champion of efficiency being used in commercial and open systems as JESS (rule engine of the Java platform).

Therefore, RETE algorithm was chosen as a reference algorithm to be adapted to be used with parameter instantiation of predicates and in the verification and storage of predicates in a given world state, resulting in the performance enhancement of the ALWAYS real time planner.

3 ALWAYS$_{TRX}$ Method

The ALWAYS$_{TRX}$ method (ALWAYS Thinking, Reacting and eXecuting) [24][25] was designed to integrate the agent's planning task and execution task, including the external events treatment inside the planning phase, to give to the agent reactivity capabilities to the dynamic world changes in which it is inserted.

The integration approach requires that planning decisions followed by execution must be based in a common knowledge representation.

With ALWAYS$_{TRX}$ the planning system integration with the plan execution system considers that planning must be done concurrently with plan execution (the plan is made "on the fly"). The plan being created by the planner is represented by a tree where the nodes mean activities and events. Each path in the tree represents a sequence of activities with events that may occur. The execution module must follow one path in the tree. While the planner creates new paths, the execution module chooses one of these paths to execute, based on the actual occurrence of the external events associated with the chosen activities.

The planning system uses the information about which external events actually happened, obtained from the environment through the execution module, to adequate the paths to the real world state. As the execution module executes one step of the plan, the paths of the search tree not associated to the path that contains this step are excluded from the tree. One plan step means to observe the occurrence of an external event and initiate the execution of the correspondent activity. When a event occurs, the planner keep expanding paths below this event in the tree and exclude the paths that do not go through this event.

The ALWAYS$_{TRX}$ is domain independent and can be used in many engineering and automation applications such as intelligent manufacturing and robots control.

Fig. 1 presents how the method is inserted in the context "see, think, act" and its internal architecture that consists of two modules running concurrently: the planning module and the execution module. Both modules constantly communicate through a common knowledge base.

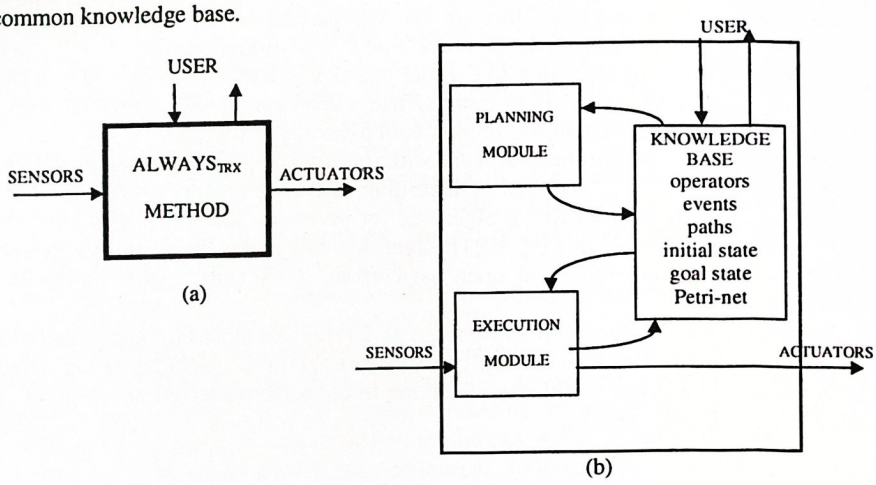


Fig. 1. *(a)* ALWAYS$_{TRX}$ method in the context "see, think, act"; *(b)* ALWAYS$_{TRX}$ method's internal architecture.

As the planning module generates the plan, the paths are stored in the knowledge base. The execution module reads the knowledge base and follows the plan to be executed. As the execution module works, the activity being executed and all the information about which external events are happening are stored in the knowledge base.

The planning module uses the information stored by execution module in the knowledge base in order to keep synthesizing the plan paths with updated world information.

4 RETE Algorithm Adaptation for the ALWAYS Method and Results

The RETE match algorithm is widely used in production rule systems. It works in two distinct phases: one of compilation (creation of the RETE net) and another one of matching check (operational). In the first phase, the algorithm compiles the set of rules inside the knowledge base (KB) in a net of linked nodes (an acyclic graph). Each node contains part of the conditions to fire a rule. In matching phase, the net processes each fact added or removed from the working memory. When a fact is added to the KB, it is presented to all nodes sons of root. Each node receives the fact, applies its test and, if the result of the test is true, the fact is stored and sent to the subsequent node (son nodes). If the test fails, the node simply ignores the presented fact and discards it. The last nodes (the leaves) in the net that receive a fact represent the rules to be fired.

In the implementation of the $ALWAYS_{TRX}$, RETE algorithm was adapted to make the analysis of the preconditions of operators and events, considering its performance and efficiency in finding equalities in multiple comparisons of facts with a high computational performance. The adapted RETE algorithm will be described in this section. The new adapted algorithm and new adapted RETE net will be referenced as RETEADAPT and RETEADAPT net, respectively.

Each change in world state (facts added or removed) calculated by the planning module is sent for the RETEADAPT and it indicates, as it was a combinational circuit, the operators and the events whose preconditions had been satisfied. With the operators and events enabled, the planner module expands each node of the planning tree.

RETEADAPT uses an acyclic graph where the nodes, except root node, represent the patterns that will be compared. The paths from root node until leaves represent the preconditions of the operators and of the events, which are the patterns to be analyzed. This graph is called RETEADAPT net. Inside RETEADAPT net, each node stores a pattern that will represent which elements from the world state satisfy the precondition of an operator or an event.

4.1 RETEADAPT Net Compiler

After the user insertion of operators and events with the corresponding preconditions, a procedure of creation of the net must be executed, that consists of compiling all the preconditions of operators and events, creating the nodes of the RETEADAPT net. Each node represents the test of a constant or variable declared as a parameter of a precondition. When there are preconditions with many parameters (constants or variables) and many predicates prototypes, a node will be created for each one of them.

RETEADAPT net has four types of nodes: the root node, one-input node, two-input node and enable-operator or enable-event node. Each one-input node can represent the name of the predicate prototype and the constants or variables that were declared as parameters. When a precondition has two terms linked by AND conjunction (e.g. clear(x)^ontable(y)), one-input nodes will be used to test the individual results and a two-input node will be used to join them.

Below there is an example of blocks world to illustrate how RETEADAPT net is created and used. It assumes the blocks world given by fig. 2. Three blocks are on a table and a manipulator robot must stack the blocks to reach a goal state. The definition of the $ALWAYS_{TRX}$ operators and events follows STRIPS definitions and can be seen in [24][25].

$ALWAYS_{TRX}$ operators:

```
operator     : pickup(x)
precondition: ontable(x)^handempty^clear(x)
del-list     : ontable(x),handempty,clear(x)
add-list     : holding(x),running(PICKUP)

operator     : putdown(x)
precondition: holding(x)
del-list     : holding(x)
add-list   :ontable(x),handempty,clear(x),running(PUTDOWN)

operator     : stack(x,y)
precondition: holding(x)^clear(y)
del-list     : holding(x),clear(y)
add-list     : handempty,on(x,y),clear(x),running(STACK)

operator     : unstack(x,y)
precondition: handempty^on(x,y)^clear(x)
del-list     : handempty,on(x,y),clear(x)
add-list     : holding(x),clear(y),running(UNSTACK)
```

$ALWAYS_{TRX}$ events:

```
event        : endactivity(x)
precondition: running(x)
del-list     : running(x)
add-list     : nil
```

In this example, the robot can execute four actions. The pickup(x) operator represents the robot action "to catch block x that is on the table and keep this block in its grip". The putdown(x) operator represents the robot action "to place block x, that is in its grip, on the table and release the block, freeing the grip". The stack(x, y) operator represents the robot action "to place block x, that is in its grip, on block y,

leaving the grip free". The unstack(x, y) operator represents the robot action "to catch a block x that is on block y and keep this block in its grip".

Initially RETEADAPT net must be created compiling the preconditions of the four operators. Analyzing the precondition of the pickup(x) operator, there are three terms: the predicate prototype *ontable* and *clear*, with x variable as parameter, and the predicate without parameters *handempty*. RETEADAPT net is assembled with three nodes below root node.

Initial State
clear(C)
clear(B)
on(C,A)
handempty
ontable(A)
ontable(B)

Goal State
on(B,C)
on(A,B)
ontable(C)
clear(A)

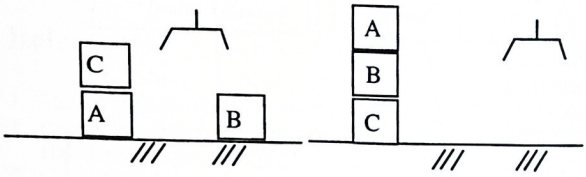


Fig. 2. Blocks world example. *Left* Initial State. *Right* Goal State.

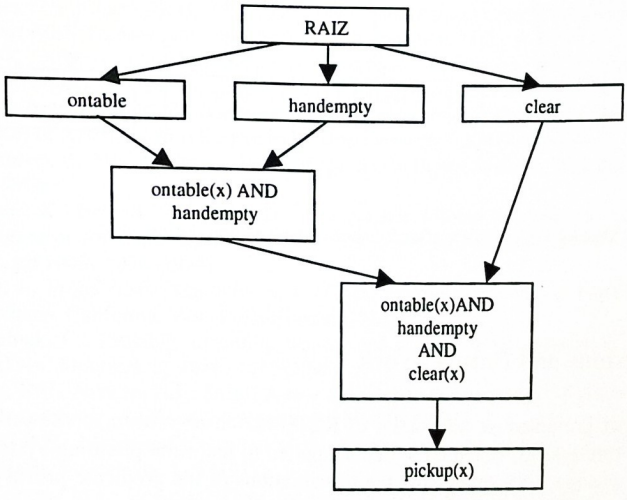


Fig. 3. RETEADAPT net for pickup(x) operator.

As the pickup(x) operator has a three term precondition with conjunctions AND, the algorithm creates two-input nodes to represent ontable(x)^handempty and after that ontable(x)^handempty^clear (x). Fig. 3 shows these two nodes with the enable operator node pickup(x). After the analysis of all preconditions of the others three operators, RETEADAPT net will be link fig. 4.

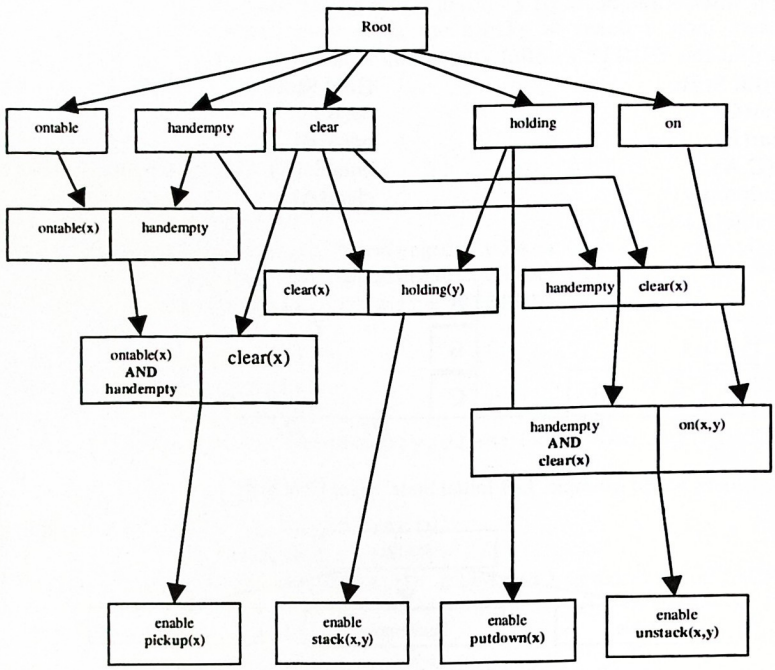


Fig. 4. RETEADAPT net complete after finished operator compilation.

5 Conclusions and Future Work

This paper presented an adaptation of RETE match algorithm, common used by production rule systems in expert systems, applied to real time planning systems. The adaptation was made for processing the instantiations of the predicate prototypes and testing of facts inside the Knowledge Base that describes the world state. This algorithm was implemented in a real time planner that uses $ALWAYS_{TRX}$ method. The planner works interleaved with execution and considers external events in the

planning phase. With this new algorithm the planning system increased the computational performance in analysis and synthesis of each state of the plan tree. Results show until 6 times less processing time of facts and overall planning time. The implementation of this suitable algorithm became possible the use of the real time planner to plan actions in many dynamic and unstructured environments. A detailed description of using this real time planning system on a manufacturing cell's manipulator robot, on a service mobile robot and in soccer robots, is being developed and will be published in the future.

Also as future work it is important to point out the performance optimization of the algorithm to better improve the response time of the system. The planner uses a RETEADAPT net in each node of the decision tree. Studies must be made to verify if it is possible to decrease the amount of information in each node and to use only one net. Another important aspect is in backtracking of the search tree where the corresponding predicates to the facts must be removed from the net in order to find new plans. Further studies must be made in compacting the data bytes used by the net.

References

1. Maes, P.: Designing Autonomous Agents. Mit/Elsevier (1991)
2. Russel, S., Norvig, P.: Artificial Intelligence: a modern approach. Second edition. Prentice Hall International (2003)
3. Georgeff, M.P. et al.: Reasoning about plans and actions. Exploring Artificial Intelligence. Morgan Kaufmann, AAAI, chap.5, pp.173--196 (1988)
4. Fikes, R.E., Nilsson, N.J.: Strips: a retrospective. Artificial Intelligence. 59, 227--232 (1993)
5. Noreils, F.R., Chatila, R.G.: Plan Execution Monitoring and Control Architecture for Mobile Robots. IEEE Transactions on Robotics and Automation. 11(2), pp.255--266 (1995)
6. Erol, K., Nau, D., Subrahmanian, V.S.: Complexity, decidability and undecidability results for domain independent planning. Artificial Intelligence. 76, 75--88 (1995)
7. Hoffmann, J. The FF planning system: fast plan generation through heuristic search. **The Journal of Artificial Intelligence Research.** 14, pp.253--302 (2001)
8. Abramson, B., NG, K.: Uncertainty Management in Expert Systems. IEEE Expert. 5(2), 29--48 (1990)
9. Golden, K., Etzioni, W., Weld, D.: Omnipotence without omniscience: Efficient sensor management for Planning. In: National Conference on Artificial Intelligence. 12, AAAI Press. pp.1048--1054 (1994)
10. Pearl, J.: Probabilistic Reasoning in Intelligent Systems: Networks of Plausible Inference. San Mateo, California, Morgan Kauffmann (1988)
11. Knoblock, C.: Planning, executing, sensing and replanning for information gathering. In: IJCAI, 14, Montreal. pp.1686--1693 (1995)
12. Fikes, R.E., Nilsson, N.J.: Strips: A new approach to the application of theorem proving to problem solving. Artificial Intelligence. 2, 189--208 (1971)
13. Fikes, R.E.: Monitored Execution of robot plans produced by Strips. In: IFIP Congress, Ljubljana, Yugoslavia. 1--5 (1971)
14. Schoppers, M.J.: Universal Plans for reactive robots in unpredictable environments. In: IJCAI, 10, Milan. pp.1039--1046 (1987)
15. Allen, J., Hendler, J., Tate, A.: Readings in PLANNING. Morgan Kaufmann, San Mateo, California (1990)

16.Olawsky, D., Gini, M.: Deferred planning and sensor use. In: Workshop on Innovative Approaches to Planning, Scheduling and Control. San Diego, CA, DARPA. pp.166--174 (1990)
17.Wilkins, D.E.: Practical Planning: extending the classical AI planning paradigm. Morgan Kauffmann (1988)
18.Pemberthy, J., Weld, D.: UCPOP: A sound, complete, partial order planner for ADL. In: International Conference on Principles of Knowledge Representation and Reasoning. 3, pp.103--114 (1992)
19.Kushmeric, N., Hanks, S., Wold, D.: An algorithm for probabilistic planning. Artificial Intelligence, 1-2(76), pp.239--286 (1995)
20.Ash, D.J., Dabija, V.G.: Planning for Real Time Event Response Management. Prentice Hall PTR (2000)
21.Rillo, M.: Aplicações de Redes de Petri em Sistemas de Manufatura. São Paulo, PhD Thesis, Departamento de Engenharia de Eletricidade, EPUSP – Polytechnic School of University of Sao Paulo (1988)
22.Rillo, M.: Expectation-Based Temporal Projection System. In: Annual Conference on AI, Simulation and Planning in High Autonomy Systems, 3, Perth, Australia. pp.276--281 (1992)
23.Lopes, C.: Planejamento Baseado em Expectativas. São Paulo, PhD Thesis, Departamento de Engenharia de Eletricidade, EPUSP - Polytechnic School of University of Sao Paulo (1998)
24.Franchin, M.N. et. al.: Planejamento de ações com tratamento de eventos externos integrado ao controle de execução para agentes inteligentes. In: IV SBAI – Simpósio Brasileiro de Automação Inteligente. São Paulo, SP. pp.233--238 (1999)
25.Franchin, M.N.: Um método de planejamento usando eventos externos e sua integração com o sistema de controle para agentes móveis. São Paulo, PhD Thesis, Departamento de Engenharia de Eletricidade, EPUSP – Polytechnic School of University of Sao Paulo (1999)
26.Ulusar, U.D.,Akin, H.L. Design and implementation of a Real Time Planning System for Autonomous Robots. In: IEEE ISIE International Symposium on Industrial Electronics, july 9-12, Montréal, Québec, Canada. pp.74--79 (2006)
27.Forgy, C.L.: RETE: A fast algorithm for the many pattern/many object pattern match problem. Artificial Intelligence. 19(1), 17--37 (1982)
28.Winston, P.H.: Artificial Intelligence. 3rd ed., Addison Wesley, Reading, MA (1992)
29.Giarratano, J.,Riley, G.: Expert System Principles and Programming. 2nd ed., PWS Publishing Company (1993)
30.Miranker, D.P.: TREAT a better match algorithm for AI production systems. In: AAAI87 Conference on Artificial Intelligence, pp.42--47 (1987)
31.Hanson, E.N.: Gator: A Generalized Discrimination Network for Production Rule Matching. In: IJCAI workshop on Production Systems and Their Innovative Applications. (1993)
32.Mostafa M. Aref, M.M., Mohammed A. Tayyib, M.A.: Lana-Match Algorithm: A Parallel Version of the Rete-Match Algorithm. Parallel Computing. 24(5-6), pp.763--775 (1998)
33.Berstel, B.: Extending the RETE algorithm for event management. In: Proceedings of the Ninth International Symposium on Temporal Representation and Reasoning (TIME'02), pp.49--51 (2002)

Flight Computer for a Human Resources Training System in Small Satellite Technology*

Esaú Vicente-Vivas[1], Emilio Jiménez[2], Francisco Gómez[2] and, Zaira Carrizales[2].

[1] Instituto de Ingeniería, Universidad Nacional Autónoma de México, Coyoacan, México DF, evv@servidor.unam.mx
[2] Thesis student, Instituto de Ingeniería, Universidad Nacional Autónoma de México, Coyoacan, México DF, {factore8, paco_84df, zairalilian}@yahoo.com.mx

Abstract. A detailed description of a small, versatile and capable flight computer (FC) developed to automate the operations of a small satellite educative system (SES) is presented. The computer architecture was designed with the required hardware interfaces to allow the automation of a laboratory satellite educative system. However, in order to widen the applications of this computer, radiation protections were added to allow its use in real small space vehicles. In addition, the computer board was developed to accomplish with size and mechanical restrictions from cubesat picosatellites. The paper shows the SES global description, the flight computer architecture, the local sensors integrated, the buses that provide connectivity with other cards, the two layer PCB generated to save production costs and the validation software developed to test its hardware interfaces.

Keywords: flight computer, satellite educative system, portable equipment, training system, university technology.

1 Introduction

Our group is developing a cost-effective training system in small satellite technology employing commercial-off-the-shelf (COTS) parts from the automotive and services industries. The system was planned to be affordable enough to be used in laboratories with the intention to offer attractive, fast, and versatile training practices and courses in satellite technology and related fields. Our goal is to use the system in High Schools, Technological Institutes and Universities, with the intention of approaching young people to the world of space applications, Science and Technology.

* Project 52297 financed by CONACYT

© L. Sánchez, O. Pogrebnyak and E. Rubio (Eds.)
Industrial Informatics
Research in Computing Science 31, 2007, pp. 163-172

In addition, the SES will also be useful in research laboratories to develop new solutions and modules for real satellite subsystems. In this sense research in fields such as three axis stabilization, digital communications, satellite sensors, power systems, payload validation, flight computers, navigation autonomy, and satellite constellations, would also be addressed with the support of this laboratory tool.

It must be mentioned that commercial availability of similar products to the SES is rarely seen in the global market. Right now the only satellite educative commercial product detected by the authors is the Eyassat educational system developed initially by the US Air Force and commercialized by Colorado Satellite Services, [1]. Besides, we found that few institutions have developed their own satellite prototypes to accomplish laboratory research in distributed space systems, as the case of the Israel Institute of Technology [2] and the case of the US Naval Academy Satellite Laboratory with its "LABsat" experimental hardware, [3]. The Eyassat basic equipment starts at 8,000 dollars. However, this price is difficult to be afforded in developing countries. This is why the SES is planned to be developed and offered for under 3,000 US dollars cost in order to be attractive for different schools, universities, and so on. This goal shaped the design and the main characteristics of the SES in order to achieve a cost-effective development.

It is also important to highlight that we are taking advantage of previous experiences in space projects, [4], [5] and [6], to fast track this project.

1.1 Further Applications of SES Computer

With the intention to increase the applications of the SES flight computer generated by our group, we planned to use it also in 1 kg picosatellites, like the one we are informally developing in our laboratory. Once the SES Project gets finished and the required documentation with CONACYT gets accomplished our goal is to start the picosatellite project.

1.1.1 Some characteristics of picosatellite missions

The cubesat picosatellite concept was developed by Dr. Robert Twiggs from Stanford University in 1999, [7]. Since then a great amount of universities from all over the world are developing, planning to build, or have developed a picosatellite project to enter the space activities. Among the countries involved in CubeSats projects are: USA, Canada, Germany, Japan, Denmark, Netherlands, Norway, Switzerland, Australia, Korea, Malasya, Argentina and Colombia, [8]. The last, through the "Universidad Sergio Arboleda", launched successfully the picosatellite called "Libertad 1" the 17th April 2007, [9].

The cost of this type of projects depends on the country that makes it as well as in the payload, and go from <100,000 to 1,500,000 US dollars including the costs of inception, launch, operation and end-of-life.
The picosatellite missions are launched in groups as secondary payloads on rockets from different countries. However, Russian rockets are often selected for launching purposes because of its attractive cost (40,000 US dollars in 2006). On the other hand, there are a couple of companies in USA that commercialize the fundamental subsystems to construct a picosatellite. They charge 6,000 US dollars for every subsystem such as: structure,

communications, electrical power, attitude determination & control, and flight computer. Nevertheless, it is important to notice out that additional investment has to be considered for experiments, development of operations software, environment tests, launching services and ground station.

Furthermore, it has to be taken into account that low-earth-orbit picosatellite missions have a typical lifespan of 3-9 months. In addition, it should be highlighted that picosatellite missions are attractive and demanded around the world because they are employed for technology demonstrations, proof-of-concepts, scientific experiments and human resources training, [10], [11]and [12].

2 Architecture of the satellite educative system

The SES is basically formed by a laboratory satellite prototype (LSP), operations software for LSP, and a laptop with executable software to monitor and control the LSP, figure 1. The last contains what is referred as the Ground Station Software (GSS) for Telemetry acquisition and Command Shipment.

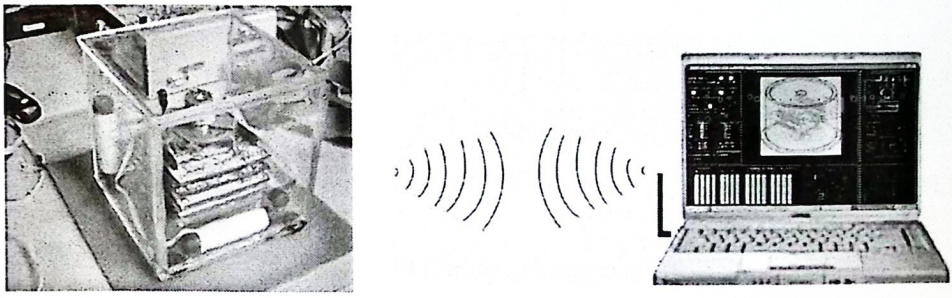


Fig. 1. The satellite educative system, on the left side the LSP mockup at real scale is shown.

3 The flight computer subsystem

The LSP has a single board flight computer (FC) with lateral connectors employed as electrical buses to interconnect electronic cards in tandem. The electrical connector offers mechanical attachments among printed circuit boards. In addition, each PCB contains holes at each corner to screw the whole printed board array to the LSP structure.

With the use of bus type connectors all the electrical signals from the flight computer are available for all the assembled cards, thus it is possible to interconnect cards without caring the order of them. Besides, it is important to notice out that all the bus type connectors in the cards are female type in the top side, whereas the bottom side is a wire-wrap connector. This allows the interconnection of cards in tandem either by the top or by the bottom sides.

The PDS single board computer is built around the 16 bit RISC SAB80C166 processor from Siemens with extended temperature, figure2, 40 Mhz oscillator, 256 kb of RAM memory where the SES operations software will be loaded, figure 2, hardware for automatic uploading of new programs to the flight computer, two main serial ports and 8 extended serial ports generated by multiplexing the second main serial port. The serial ports support full-duplex asynchronous communication up to 625 Kbaud and half-duplex synchronous communication up to 2.5 Mbaud.

On the other hand, the 100-pin SAB80C166 microcontroller internally contains important resources as follows: a watch dog timer, interrupt controller, 16-bit timers, 10-channel 10-bit A/D converters, two serial channels and several 16 bits I/O ports, with a total of 76 I/O lines.

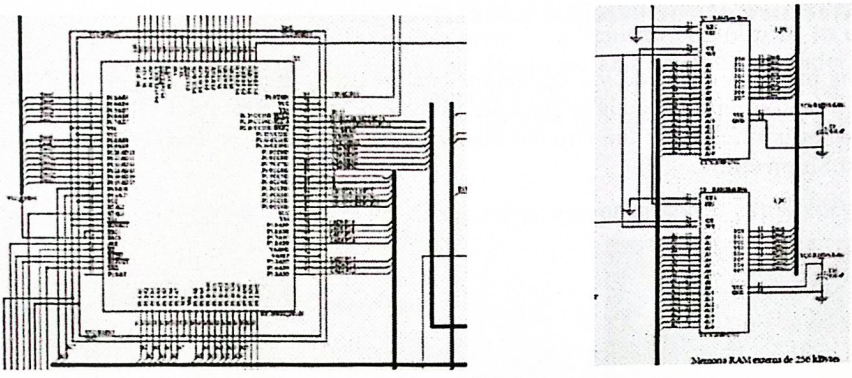


Fig. 2. Schematics details of RISC Siemens microcontroller and 256 Kb of SRAM.

The main specifications for the flight computer are the followings:

- 8.9 x 8.9 cm printed circuit board,
- Two layer PCB for easy and cheap fabrication,
- 16 bit RISC microcontroller, SAB80C166 from Siemens,
- Microcontroller PIC16F876A to control the process of new software uploading to the SAB processor,
- 256 Kb of SRAM for program execution,
- 32 Mb of Flash RAM for massive data storage,
- SPI port emulated with I/O lines and dedicated software,
- 3 local 1-wire temperature sensors,
- 3 local 1-wire current sensors,
- Latch-up protection for main resources from FC such as: SAB80C166 processor, SRAM and PIC16F876A,
- Multiplexing module to generate 8 extended serial ports to the SAB80C166 processor for communications with other boards from the satellite educative system.

3.1 Flash memory for massive data storage

The flight computer contains 4 AT45DB642D flash memories to offer a whole storage capacity of 32 Mb, each flash memory has a storage capacity of 8 Mb. The flash memories are accessed by SPI interface through 4 communications lines connected to the SAB microcontroller. The technical specifications for the AT45DB642D are the followings:

- Voltage supply from 2.7 V to 3.6 volts.
- SPI interface up to 66 MHz.
- 8 Mb Flash memory organized in sectors of 256 Kb, blocks of 8Kbytes and pages of 1 Kb.
- Two SRAM buffers each one of 1 Kb.
- Operation Temperature from – 40 °C to + 85 °C.

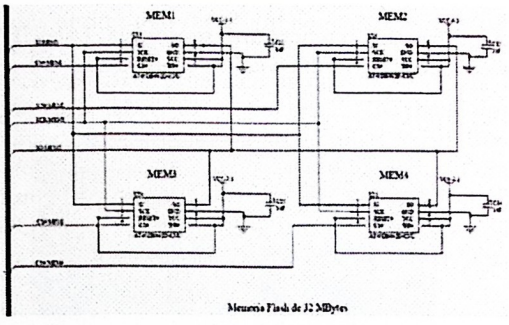


Fig. 3. Schematic circuit of the Flash memory module.

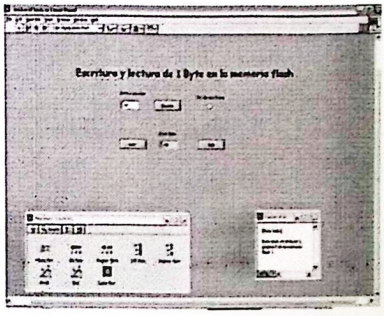


Fig. 4. LabVIEW software developed to validate basic operations with Flash memories.

Besides, it was developed a LabVIEW software to validate the programming and operation of Flash memories in laboratory. For this purpose a 1 layer PCB was developed in the laboratory to solder the surface mount flash memory. This process allowed the early validation of programmed software related to the Flash memories. Figure 3 shows the schematic diagram of the flash memory module while figure 4 shows the software interface employed to access the memories in laboratory.

3.2 Sensors installed on board the flight computer

The 1-wire sensors from the automotive industry are attractive because they are cheap and enable important reductions of line tracks in the PCBs and also reduce the amount of interconnection wires between system boards. Moreover, the 1-wire sensors let a microcontroller device to access an important amount of sensors by means of only two wires. This contrasts with conventional sensors, which require two wires for each included sensor.
By these reasons, the flight computer is instrumented with local sensors that are intended mainly to use the FC in a Picosatellite mission. However, the sensors will also be of important use in the satellite educative system to show the user the applications and importance of

telemetry systems. In addition, the sensors will make possible the interactions among user and SES to modify directly the readings of temperature sensors when they be touched, by instance, by the user's finger.

The FC has three DS18S20 1-wire temperature sensors. The three sensors are connected to the same input port from the SAB controller. The 1-wire temperature sensors have the following specifications:

- Every sensor has a unique ID serial number of 64 bits called ROM Code.
- Supply voltage from 3.0 V a 5.5 V.
- Temperature measurement from - 55 °C to + 125°C.
- ± 0.5 °C precision in the interval from – 10 °C to + 85 °C.
- 9 bit resolution.
- Temperature A/D conversion time of 750 ms (max.).

In order to validate the software programming for the temperature sensors a LabVIEW software was developed. Figure 5 shows the presentation of results with this software.

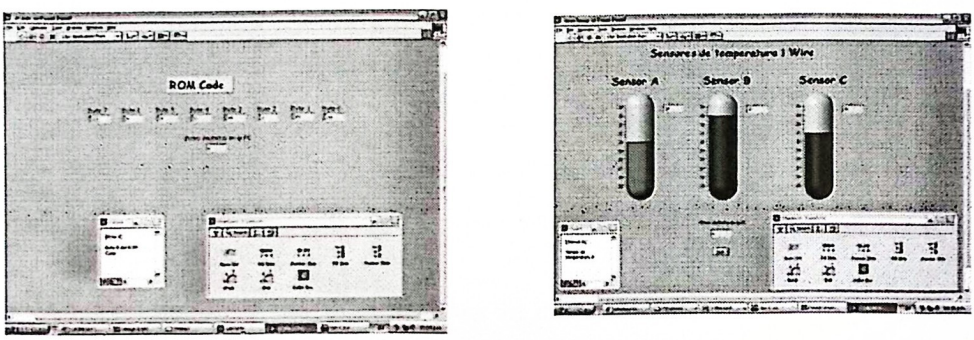


Fig. 5. LabVIEW software employed to display the ROM code and the temperature of the three 1-wire temperature sensors from FC.

3.3 Operations Software for the flight computer

The LSP will contain an operations control software loaded in the flight computer to carry out the following functions:

- Communications with the laptop software,
- Command requests from GSS,
- Reception and execution of new programs of up to 256 Kb,
- Acquisition, packing, data storage in flash memory, and transmission of telemetry by wireless means,
- Managing of protocols with GSS,
- Simulation of orbital times for LSP,
- Real time operations with LSP,

- Communications with SES subsystems (stabilization, power, digital audio and payload) through digital commands,
- Data acquisition of local sensors, among them: 3 1-wire temperature sensors and current sensors associated with latch-up sensors (SAB80C166, SRAM and PIC16F876A),
- Activation and deactivation of SES subsystems through the help of the power subsystem. This feature allows the system to save energy from the batteries,

Regarding the software development for the SAB80C166 it is written in standard "C" language, programs are compiled using the BSO Tasking family of tools for the SAB80C166 microcontroller.

4 Latch-up effect protections

Considering that the flight computer will be used as well in picosatellites to be launched into space orbit, the flight computer design includes hardware protections to prevent the destructive latch-up effect generated by cosmic radiation. This phenomenon affects electronic parts when they are taken to space orbit. The latch-up prevention hardware was installed in those components with higher possibilities to be affected, in this case the electronic parts with higher integration scales and associated to most important computer resources. Therefore these protections were placed in the SAB80C166 16 bit RISC microcontroller, its 256 Kb of external SRAM (from where the FC programs are executed) and the PIC16F876A (which controls the up-loading of new programs to the Siemens microcontroller).

The protection detects the electrical current consumption for every device and automatically generates a digital output pulse when any one of them reaches a maximum allowed current threshold level. These signals are taken to the lateral connectors and will be captured by the picosatellite power subsystem board. This subsystem will process the signals pulses and will turn off the power supply for the alarmed subsystems. This is the only mechanism that can stop the destructive latch-up effect.
The latch-up circuit placed in the FC is formed by the MAX4071 and the LM6511. In addition, an A/D DS2450 1-wire converter device was placed between the MAX4071 and the LM6511. The 1-wire voltage sensor will be used by the SAB80C166 to collect telemetry data from latch-up current sensors. The voltage sensor has the following specifications:

- 1-Wire interface,
- 64 bit ID number for every device,
- 4 A/D converters per chip,
- 16 bit resolution,
- 5V voltage supply,
- Operating temperature from – 40 °C to + 85 °C.

4.1 Latch-up current measurements

An additional LabVIEW software was developed to validate the basic operations of FC current sensors. This performs the programming and data reading of 1-wire current sensors. Figure 6

shows the graphical interface of the software, as can be seen the visualization of results is simple and straightforward. Validation tests were successfully performed in laboratory with this software and electronics assembled in 1 layer temporal PCB.

It is important to highlight that the flight computer PCB will be the same for both the SES and Picosatellite missions. However, the protection components will be installed exclusively in computers elaborated for Picosatellites. In other words, the SES computers will not install these electronic parts because they will be employed in terrestrial environment. The goal is to employ the same PCB computer for both applications to lower the development cost of our small satellites projects.

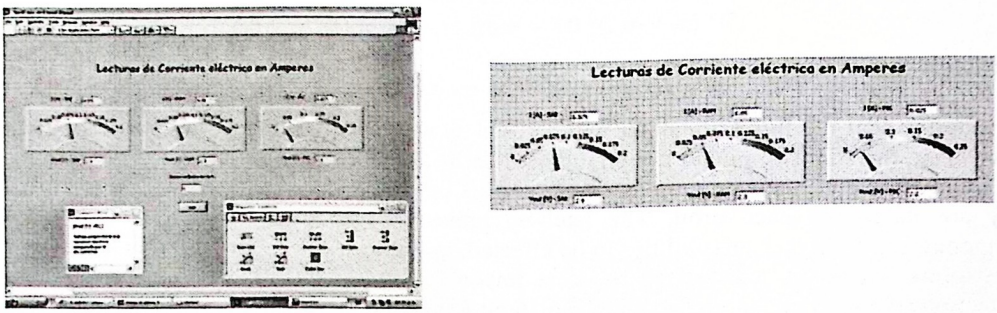


Fig. 6. LabVIEW interface developed to validate the programming and operation of FC current sensors.

5 Uploading new programs to the flight computer

The versatility and friendly operation of the satellite educative system will depend, among other things, on its flexibility to upload new programs to the LSP. This feature will allow the user to make changes, studies, practices and research activity with the educative system. Unfortunately, the SAB80C166 needs external control in order to achieve this operating mode. By this reason, a PIC16F876A microcontroller was added to the design. This processor is programmed to read the GSS serial port and continuously looks for the detection of the "upload new program" command. Once this is detected the PIC controller frees the communication channel among GSS and SAB80C166 and drives control signals from the SAB80C166 (ALE and NMI).

This process allows the SAB processor to execute its internal boot ROM which contains communications software to detect the baudrate speed and to receive the new software by serial port. The process takes three different steps, in the first one the SAB device receives a 32 byte program in its internal RAM and automatically executes it. This program allows the uploading of a second program of 256 bytes. The third step allows up to 256 Kb of software to be uploaded in the FC external SRAM. Every step is monitored by the PIC16F877 microcontroller, whenever the process is successfully concluded the PIC resets the SAB device in order to start automatically the execution of the new program with full control form the SAB processor.

In order to validate the new program uploading function for the flight computer LabVIEW software was developed. It allows the opening of a file with HEX extension and sends it through serial port.
Considering that the FC board is not yet ready for testing, the uploading software was tested with a SAB80C166 card developed in our lab for a previous project, figure 7. The software performed without problems and is ready for FC laboratory testing once the board is ready for assembly and testing. The LabVIEW software is shown in figure 8.

6 PCB design and fabrication

The two layer printed circuit for the flight computer board was elaborated with Protel DXP software, a powerful tool for design and manufacture of printed circuits boards. It allows the use of different data bases for components as well as the creation of new components and the possibility to aggregate them to the project. It also enables the positioning of components and the automatic routing of connections between the components of the project in a very short time and in an efficient way.

To elaborate the FC several component distributions on the board were practiced to allow the full automatic routing of the PCB. In our case the distribution is very important because the FC architecture includes many electronic components, most of them of surface mount type and few of them of dual in line type. However, the PCB area is too small to allocate the components in a suitable way to admit the automatic routing of electrical lines. This is the reason by which several chosen component distributions never reached the auto-routing completion. The final and successful distribution is shown in figure 9. This design is to be sent for manufacture in the next weeks, once this becomes manufactured, basic revision will take place and then parts assembly and partial testing will be applied.

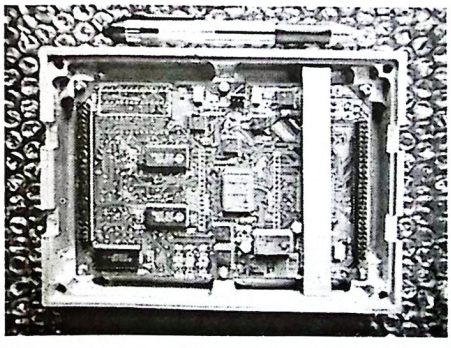


Fig. 7. Electronic board employed to validate the uploading software.

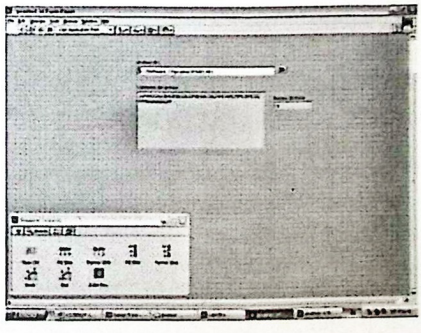
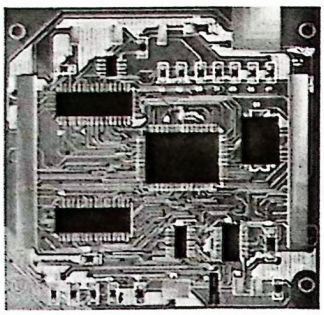


Fig. 8. LabVIEW software developed to send new programs to the flight computer.

7 Concluding Remarks

We have presented the FC computer design, the global architecture, the preliminary tests

performed with dedicated software specially developed for this purpose and the PCB development process. Few hardware interfaces such as 1-wire sensors, latch-up protection and new program uploading, required preliminary tests to validate the designs. This work will increase the success chances to generate an operative PCB from the first manufacturing run.
In this way, the final flight computer PCB is ready to be manufactured. We expect to have it ready in few weeks to proceed with its assembly and progressive testing in order to reach an operative flight computer by the end of 2007.



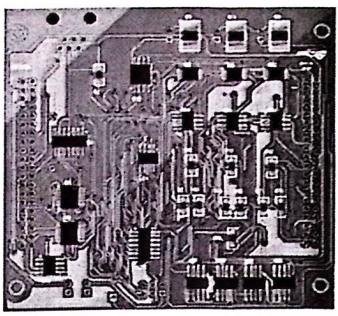


Fig. 9. Flight computer PCB ready for manufacturing process.

References

1. Eyassat, 2007, http://eyassat.com/tiki/tiki-index.php .
2. Israel Tecch, 2007, http://www.technion.ac.il/~pgurfil/projects.html
3. USNaval, 2007, http://www.wa8lmf.net/bruninga/ea467.html
4. Esaú Vicente-Vivas, Fabián García-Nocetti and Francisco Mendieta-Jiménez, "Automatic maintenance payload on board of a Mexican LEO microsatellite", Acta Astronautica Journal, Elsevier Science,, Volume 58, Issue 3, Pages 149-167, February 2006.
5. E. Vicente-Vivas y D.F. García Nocetti, "Computadora de Vuelo Triplex de Diseño y Manufactura Mexicana para el Microsatélite Satex", Revista Iberoamericana Información Tecnológica, ISSN 0716-8756, Vol. 17, No 1, pp. 69-76, Enero 2006, Chile.
6. E.Vicente Vivas, A.Espinoza M., C. Pineda F., J.R. Tórres F. Y A. Calvillo, "Software de Adquisición de telemetría y control de operaciones para microsatélites", 4° Conferencia Internacional en Control, Instrumentacion virtual y Sistemas Digitales "CICINDI 2002", Pachuca, Hidalgo, México, Agosto del 2002.
7. CubeSat Origins, http://directory.eoportal.org/pres_CubeSatConcept.html .
8. CubeSats Running Projects, http://en.wikipedia.org/wiki/CubeSat#Current_running_projects .
9. Libertad 1, 2007, http://www.usergioarboleda.edu.co/proyecto_espacial/flanzamiento.html
10. Montana Earth Orbiting Pico-Explorer, http://www.ssel.montana.edu/merope/ .
11. Y.Tsuda et al, "University of Tokyo's CubeSat "XI" as a Student-Built Educa tional Pico-Satellite -Final Design and Operation Plan", The 23rd International Symposium of Space Technology and Science, Matsue, Japan, 2002.
12. Miyamoto K. et al, "Tokyo Institute of Technology Small Satellite Projects", 7th University Space Systems Symposium, Waikoloa, Hawaii, November (2004).

Aircraft Noise Evaluation: First Stage*

Arturo Rojo Ruiz[1], Luis P. Sánchez Fernandez[1], Luis A. Sánchez Pérez[2]
[1]Center for Computing Research, National Polytechnic Institute, Mexico
[2]Higher School of Computing, National Polytechnic Institute, Mexico
Av. Juan de Dios Batiz s/n casi esq. Miguel Othon de Mendizabal, Col. Nueva Industrial Vallejo. CP 07738. Mexico City, Mexico
arturo.rojo@gmail.com
lsanchez@cic.ipn.mx
lalejandro@ipn.mx

Abstract. This paper presents the characteristics of aircraft noise and a brief description of some of its effects in human beings. The method uses frequency and octave analysis as well as indicators' calculation; programmed by virtual instruments. The obtained results show that the most significant frequencies are the low ones (around 100 Hz), and some of the effects caused by the noise are stress and sleeping problems.

1 Introduction

Nowadays noise has become a serious problem in all the big cities of the World. It is caused by a great variety of electronic machines, factories, etc. However, one of the more bothering noises that can be found is the one generated by means of transport. Among these, airplanes generate the biggest amount of acoustic energy; therefore, the closest areas to the airports are the most affected.

If the noise level generated by planes has been decreasing with the pass of time due to new technologies, it is also true that recent studies relate this kind of noise with annoyances to human beings, for example, difficulties when communicating, sleeping problems, etc. All this leads us to think about: How is the aircrafts' noise? Which characteristics does it have? Etc.

The purpose of this paper is to analyze the characteristics of aircrafts' noise; to make a comparison between aircrafts with propellers and jet aircrafts and to show brief some of the most common effects of these events.

It also important to mention that the sound patterns used to the elaboration of this research, were taken from the International Airport of Mexico City (AICM) during the takeoffs of the aircrafts. The takeoffs' sound was taken as reference because it is when the highest level of noise is produced; hence the bigger troubles for the nearby community.

* Project 51283-Y financed by CONACYT, Mexico

© L. Sánchez, O. Pogrebnyak and E. Rubio (Eds.)
Industrial Informatics
Research in Computing Science 31, 2007, pp. 173-182

2 Data acquiring

2.1 Measurement tools

The used aircrafts' noises in this work have been acquired by means of MP201 microphone. It is a good choice for use in IEC61672 class 1 sound level meters and other noise measurements requiring class 1 accuracy [1]. It is a 1/2" prepolarized free-field measurement microphone. The data acquisition card is USB-9233. It is a four-channel dynamic signal acquisition module for making high-accuracy measurements from IEPE sensors. The USB-9233 delivers 102 dB of dynamic range. The four USB-9233 input channels simultaneously acquire at rates from 2 to 50 kHz. In addition, the module includes built-in antialiasing filters that automatically adjust to your sampling rate. The USB-9233 uses a method of A/D conversion known as deltasigma modulation. If the data rate is 25 kS/s, each ADC actually samples its input signal at 3.2 MS/s (128 times the data rate) and produces samples that are applied to a digital filter. This filter then expands the data to 24 bits, rejects signal components greater than 12.5 kHz (the Nyquist frequency), and then digitally resample the data at the chosen data rate of 25 kS/s. This combination of analog and digital filtering provides an accurate representation of desirable signals while rejecting out-of-band signals. The built-in filters automatically adjust themselves to discriminate between signals based on the frequency range, or bandwidth, of the signal.

2.2 Measurement characteristics

In this paper, the noise samples were acquired with sampling frequencies of 25000 Hz (Samples/second: S/s), monophonic and during 24 seconds. In general, this interval is greater than the aircraft takeoff time, or greater than the time in which the produced noise affects the zones near an airport. The takeoff direction is always the same one and this reduces the disturbances of Doppler Effect.

According to [2], three points of reference must be taken in count to make measurements of airplanes' noises. The first one is a lateral point at 450m (it cans variate depending on the airplane that will be measured); the second is located at 6.5 Km. in straight line from the start of the takeoff path: the last one is placed at 2 Km. before the runway in order to measure the approximation noise.

Something that has to be noticed in the AICM is that at 130m from one side of the runway, there are houses. That is why it was decided to measure at this point to observe the noise level present in this homes and also because in takeoffs, the lateral point is the most significant.

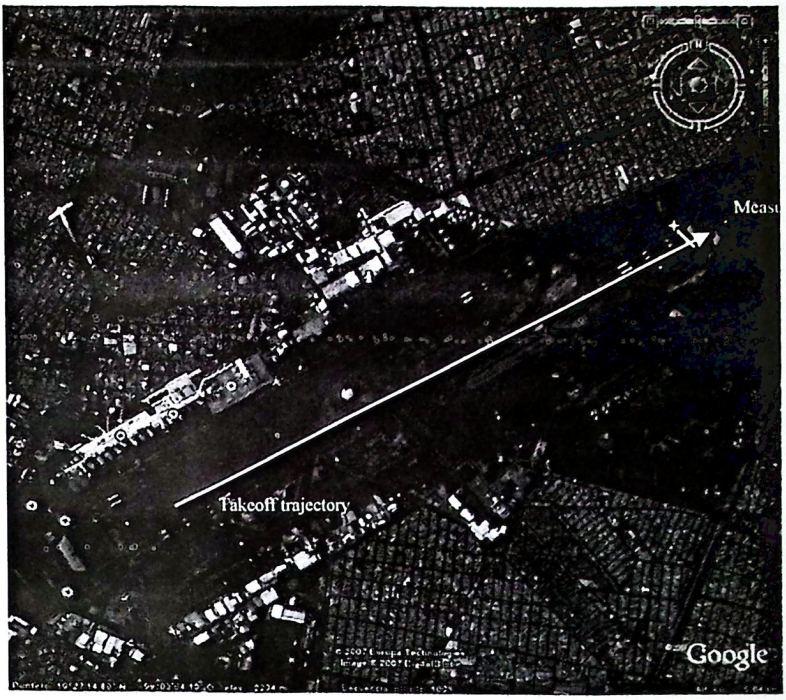


Fig 1 Map of AICM

The measurement point is placed at approximately 130m in a perpendicular way to the runway and is located at 19°26'41'' N, 99°3'44'' W. The meteorological conditions at the moment of the measurement were: temperature 19° Celsius and relative humidity of 64 %.

3. Data Processing

In order to obtain valuable information from the gathered information, it is necessary to process it. This can be done in several ways, one of them is to get statistic indicators which are used to discover the urban noise [3][4]; another way is to make a frequency analysis. This analysis allows us to get the spectral characteristics of the signal. Finally, the last option is to use an octave analysis, the information gotten from this analysis is the energy contained in each one of the frequency bands.

3.1 Sound Indicators [3][4][5]

There is a huge variety of indicators. However, the indicators used for monitoring aircraft's noise are: instant sound level (Lp), highest sound level (Lmax), the perceived noise level (PNL), the sound exposure level (SEL), the effective perceived noise level (EPNL), the equivalent sound level (Leq), the day/night sound level (DNL), the community noise equivalent level (CNEL), and the level of exceeded sound in a percentile x (Lx).

Most of these indicators require long term measurements, some of them even of 24 hours. They refer to the way the sound behaves through time; during measurement, a lot of information is collected from a wide variety of aircrafts; therefore, we cannot obtain the individual behavior of each one of them. Some other indicators are subjective values that may not reflect the characteristics of aircrafts' noise.

The selected indicators let us get information about the energy generated by aircrafts' noise in an individual way. They can be obtained from measurements of 24 seconds. The indicators used are the following:

- Equivalent sound level (Leq). It is the sound level that would have a constant noise in the same period of time than the measured noise. The Leq represents the sound energy contained in noise in a determined time.

$$Leq = 10 \cdot \log\left[\frac{1}{T} \cdot \int_0^T \frac{p^2(t)}{p_0^2} dt\right] \ (dB), \tag{1}$$

- Sound exposure level (SEL). It is the sound level that if kept constant during 1 second, would have the same energy that the measured sound event. It is used to measure isolated events such as the passing of airplanes, etc.

$$SEL = 10 \cdot \log\left[\int_{t_1}^{t_2} \frac{1}{T_{ref}} \cdot 10^{\frac{Lp(t)}{10}} dt\right] \ (dB), \tag{2}$$

- Maximum sound level (Lmax). It is the highest sound level that is registered during a period of time.
- L10, L50, L90. It is the sound level that is surpassed and is determined by a time percentage. For example, the L90 is the sound level that was surpassed during 90% of the measurement time.

Table 1. Indicators for the MD87 shown in Fig 2.

Indicator	Value
Leq	95.25 dB
SEL	109.056 dB
Lmax	114.79 dB
L_{10}	98.9744 dB
L_{50}	84.9091 dB
L_{90}	68.3066 dB

3.2 Frequencies Analysis

Frequencies analysis is one of the most used techniques for signal analysis. In this case, having discrete values, it is obtained by the FFT (Fast Fourier Transform). This lets us get the spectral components of the analyzed noise.

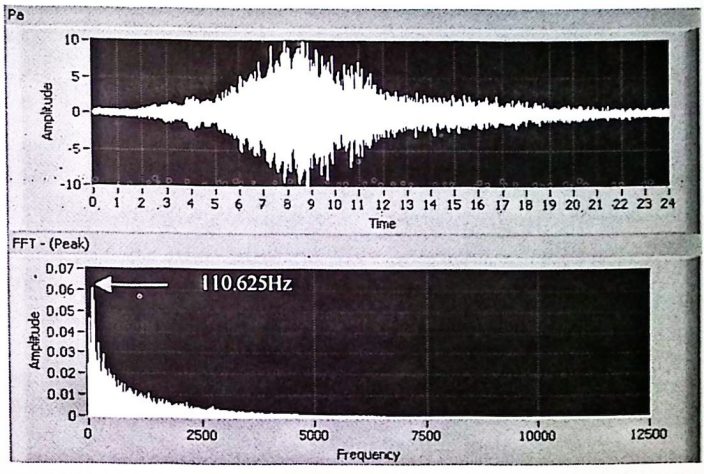


Fig 2 Noise of MD87 aircraft taking off, with sampling frequency of 25000 Hz.

As it can be seen in Fig 2, the noise produced by the plane MD87 during the takeoff reaches levels close to 10 Pa. In addition, the overtones of greater amplitude are contained between 30 and 210 Hz. (Counting the components that have at least half of the maximum amplitude), having the maximum value in 110.625 Hz.

3.3 Frequency weighting filters

The human ear is not equally sensitive to sound at different frequencies. To adequately evaluate human exposure to noise, the sound measuring system must account for this difference in sensitivities over the audible range. For this purpose, frequency weighting networks, which are really "frequency weighting filters", have been developed [6].

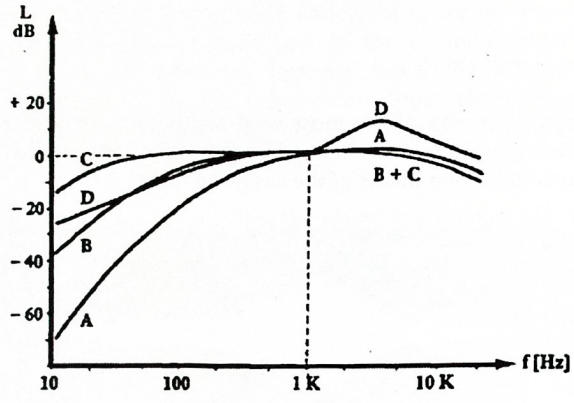


Fig 3 Frequency weighting filters

There is a large variety of frequency weighting filters; they evaluate different types of noise. There is also the frequency weighting filter D, which was developed to characterize the effects of aircrafts' noise. However, the two internationally standardized weighting networks in common use are the "A" and "C", which have been built to correlate to the frequency response of the human ear for different sound levels. Their characteristics are defined in [7].

- Frequency weighting filter A. It was created to model the human ear response to low intensities. Nowadays, almost all the laws and rules use it to limit the acceptable noise levels. The weighted A sound levels are called decibel A dB(A). The values of the frequency weighting filter A can be calculated basing on the next formula. [7]:

$$A(f) = 20\log\left(\frac{1{,}2588\times 12200^2 f^4}{\left(f^2+20{,}6^2\right)\sqrt{f^2+107{,}7^2}\,\sqrt{f^2+737{,}9^2}\,\left(f^2+12200^2\right)}\right) \qquad (3)$$

- Frequency weighting filter C. It was created to model the human ear response to great intensity sounds. It is used to evaluate ambient and low

frequency sounds in the band of audible frequencies. The values of the frequency weighting filter C can be calculated basing on the next formula [7]:

$$C(f) = 20\log\left(\frac{1{,}0071\times 12200\, f^2}{\left(f^2+20{,}6^2\right)\left(f^2+12200^2\right)}\right) \tag{4}$$

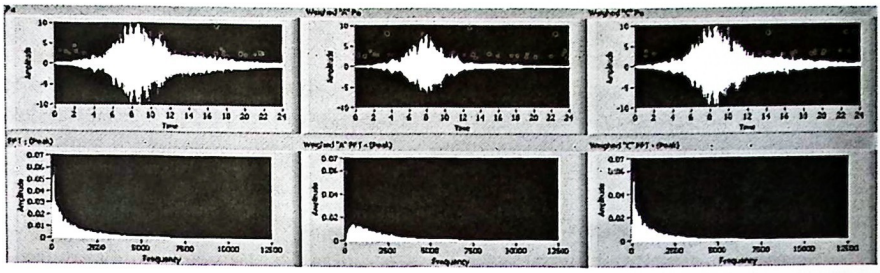


Fig 4. Left: Original Signal; Center: Weighted "A" Signal; Right: Weighted "C" Signal.

On Fig. 4 we can see that when the frequency weighting filter A is applied, it annuls the low frequencies, which in the case of aircrafts' noise are quite significant.

When weighting, the later calculations are affected, this can be seen in the sound indicators. This is really important because in many places the permitted limits of noise are defined in dB(A) because it is the way the ear responses. However, the noise not only affects the ear, but all the body, and it responses to the frequencies in a different way than the ear.

3.4 Octave Bands Analysis

The human auditory mechanism is more sensitive to proportions of frequencies than to frequencies. The frequency of a sound will determine its height as perceived by a hearer; a proportion of two times a frequency is heard as a change of height of an octave, no matter which were the frequencies. If for example, a sound goes up from 100 Hz to 200 Hz, its height will increase in one octave; when a 1000 Hz sound goes up to 2000 Hz, it will also increase in one octave of height. This fact is valid with so much precision in an important frequency range, that it is convenient to define an octave as a proportion of frequencies of two, although the octave itself is a subjective measure of change of height in a sound.

For the ear, an octave is an interval of frequencies. The octave analysis is defined as a rule for the acoustic analysis. In this case, the third of octave analysis was used to get a better resolution.

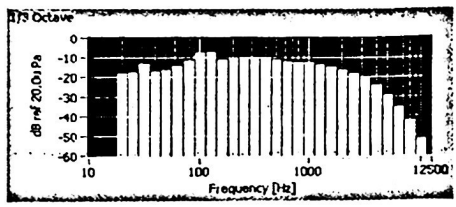


Fig 5 Noise produced by a MD87 in a third of octave analysis.

4. Comparison between propeller aircraft and jet aircraft

These days, there are mainly two types of aircrafts: with propellers and jets. The propeller aircrafts were the first ones to be used. However, the jets can carry more weight and therefore can transport bigger airplanes. The only problem they have is that they generate a lot more noise than propeller aircrafts, as can be seen in the following figure.

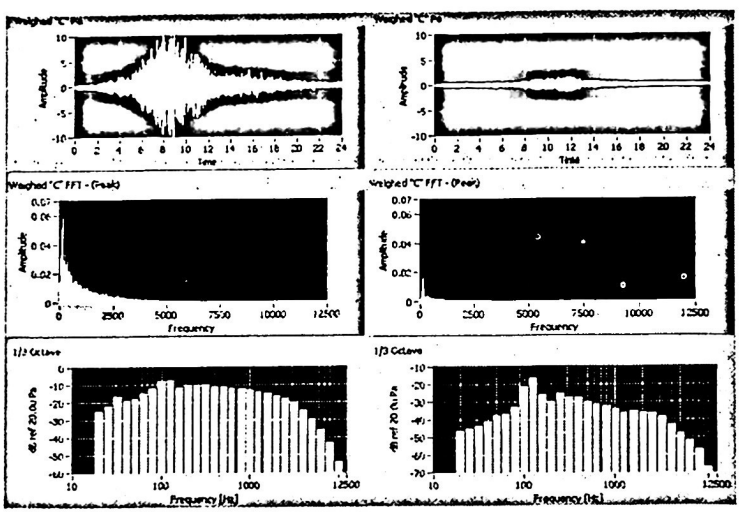


Fig. 6 Left: Noise of a MD87 (jet); Right: Noise of an ATR-42 (propellers).

The information shown in Fig. 6 was acquired by using a frequency weighting filter "C". In spite of all the differences between both airplanes, it can be seen that the amplitude of the overtones in both cases is bigger when it is near to 100 Hz.

Table 2. Comparison between noise indicators.

Indicator	MD87		ATR-42	
	dB(A)	dB(C)	dB(A)	dB(C)
Leq	89.86	94.87	71.24	80.24
SEL	103.659	108.671	85.0437	94.0436
Lmax	111.86	114.78	93.01	97.33
L_{10}	93.0506	98.5465	74.9248	84.2573
L_{50}	74.1949	83.7636	63.0512	70.0208
L_{90}	56.1661	67.1381	47.7169	54.6854

Although these a priori values do not cross the established limits, it is also interesting that the norm specifies that the lateral measurements have to be made between 450 and 650m, and in this zone, the closest houses are at 130 perpendicular meters from the runway, causing annoying problems to the habitants of this area.

5. Noise effects on humans.

We can find the following effects on humans of the frequencies contained inside the aircrafts' noise.

Tabla 3. Effects of the noise.

Effect	Causing frequencies	Sound pressure level	Exposition time.
A little pain in the ear [8][9]	50 - 8000 Hz	110 dB	Seconds
Interferences in communication [4][9]	30 - 100 Hz 100 - 4000 Hz	> 90 dB > 50 dB	Minutes
Body vibrations [10]	4 – 100 HZ	>105 dB	Seconds
Stress [8]	The whole spectrum	>105 dB	Minutes

Stress may have different physical or psychological manifestations. Some examples of this may be allergies, hives, back pain, muscular cramps, asthma, bronchitis, high pressure, migraine, duodenal ulcer, colitis, obesity, anxiety, fatigue, etc. [11].

Other effects of noise have been detected although there is not information as specific as the exposed on Table 3. Some of these are:

- Sleeping problems [12][13]
- Hypertension [11] [13]

Almost all of these problems are still objects of study.

Even during the measurements that lasted approximately 3 hours, some effects such as headaches, ears buzzing, and fatigue, were felt. These effects continued after hours of finishing with measurements.

6. Conclusions and Future Work.

The most significant frequencies in planes are the low ones (around 100 Hz). As a preliminary way, it can be said that frequency weighting filter "A", is not completely useful to measure airplanes' noise because it annuls low frequencies; although these frequencies cannot be detected by the human ear, the body responses to them in other ways such as stress. Considering that most of the legislations use frequency weighting filter "A" to establish the maximum noise levels, it is possible that all the noise effects are not considered because of the annulled frequencies.

For the future work, it is advisable to analyze the convenience of decreasing the spectral resolution in order to eliminate the noise of the patterns. Other indicators must be used; it can be the case of the EPNL, and other similar indicators. It is highly convenient to research more about noise's effects on health, specially the one generated by airplanes.

7. References

1. International Electrotechnical Commission (IEC): Standard IEC61672: Electroacoustics-sound level meters (2002).
2. Secretaría de Comunicaciones y Trasportes (SCT): "NOM-036-SCT3-2000": It establishes within the Mexican Republic the maximum permitted levels of noise emission produced by subsonic supersonic reaction aircrafts, propellers, and helicopters; its measurement method, as well as the requirements to fulfillment to these limits.
3. Kinsler, L. E.; et al: "Fundamentos de Acústica". Limusa, (1999).
4. Berglund, B; Lindvall, T; Schwela, D. H.: "Guidelines for Community Noise", World Health Organization (1999).
5. Crocker, M. J.: "Handbook of Acoustics". Wiley, (1998).
6. Hansen, C. H.; et al: "Occupational exposure to noise: evaluation, prevention and control", World Health Organization (2001).
7. International Electrotechnical Commission (IEC): Standard IEC651: Sound Level Meters (1979).
8. Berglund, B; Lindvall, T; Schwela, D. H.: "Community Noise", World Health Organization (1995).
9. Kryter, K.; "The Effects of Noise in Man", second edition, Academic Press, (1985)
10. Recuero, M.; "Ingeniería Acústica", Paraninfo, (1994).
11. Ostrosky—Solís: "Toc Toc, ¿Hay alguien ahí?", InfoRed, (2001)
12. Michaud, D. S.; et al: "Review of field studies of aircraft noise-induced sleep disturbance", The Journal of the Acoustical Society of America, Volume 121, pages 32-41, (January 2007).
13. Knipschild, P.: "Medical Effects of Aircraft Noise: Review and Literature", International Archives of Occupational and Environmental Health, Springer Berlin, Volume 40, Number 3, pages 201-204, (1977).

Color Image Segmentation Using Relation of Equivalence

Pablo Manrique Ramírez[1], Oleksiy Pogrebnyak[2]

Instituto Politécnico Nacional, Centro de Investigación en Computación,
Av. Juan de Dios Batiz S/N, Colonia Nueva Industrial Vallejo, C. P. 07738, México, D. F.
E-mail: [1]pmanriq@cic.ipn.mx, [2]olek@pollux.cic.ipn.mx

Abstract. An algorithm for color image segmentation is defined and implemented. It uses a binary relation of equivalence that is established with respect to different metrics defined in topological spaces. The designed algorithm works with RGB represented color images in BMP format. The algorithm is based on relating two points $\vec{x}$, $\vec{y}$ in a space or RGB image if and only if it is fulfilled that a metric $d(\vec{x}, \vec{y})$ is less or equal to a threshold $r>0$. Previously to the application of the relation of equivalence one fixed reference point is calculated on the expectation of a selected set of points. The posterior realized operations are dealing with different types of metrics in which different interesting results are observed comparing to other segmentation algorithms for this kind of images. The presented algorithm may be used to process satellite multichannel image data.

Key words: image segmentation, metric.

1 Introduction

The necessity to measure is evident in the majority of the technical and scientific activities. However, it is not interesting only to count with measurements, but also to know if these measurements are valid. To this end, one must remember the definition of measurement as the "process for which they are assigned numbers or symbols for attributes of entities from real world in such a manner that describes them according to the clearly defined rules" [1]. An assignment established between the real world and measurement values is usually denominated measuring scale.

The abstract concept of a metric space was introduced in 1906 by French mathematic M. Fréchet, and then was developed by German mathematic Felix Hausdorff. The Soviet mathematic J. V. Smirnov found the conditions necessary and suffiecient for the metricsability of a topologic space. Actually, the metric spaces are a particular case of the topological spaces.

In our application the segmentación is performed using the vectors of RGB colors on the base of a Cartesian coordinate system where each of its primary spectral components are red, green and blue. Hence, the images of RGB model consist of three planes of independent images, one for each primary color. Thus, the usage of RGB model for image processing has a sense when the proper images are expressed in a

© L. Sánchez, O. Pogrebnyak and E. Rubio (Eds.)
Industrial Informatics
Research in Computing Science 31, 2007, pp. 183-192

natural form in terms of three color planes. This condition permits that RGB images are employed in applications of multichannel data processing from airborne or spaceborne sensors.

The presented algorithm is defined on the pixels of a RGB image, which for entrance is considered as a metric space, and a formal tool named relation of equivalence that induce a partition on the metric. To develop the algorithm, in continuation important concepts y notations are briefly considered.

2 Theoretical aspects

Let us define a binary relation on a set X={ $\vec{x}$ / $\vec{x}$ is a point in RGB space} as following:

$$\vec{x} R \vec{y} \leftrightarrow d(\vec{x}, \vec{y}) \leq r, r > 0 \quad (1)$$

Clearly, R is reflexive, symmetric and transitive on metric space (X, d); to prove these properties one can use the properties that by definition a metric must fulfill [2]:

$$\begin{aligned} &\text{a) } d(\vec{x}, \vec{y}) \geq 0;\ d(\vec{x}, \vec{y}) = 0 \leftrightarrow \vec{x} = \vec{y} \\ &\text{b) } d(\vec{x}, \vec{y}) = d(\vec{y}, \vec{x}) \\ &\text{c) } d(\vec{x}, \vec{y}) \leq d(\vec{x}, \vec{z}) + d(\vec{z}, \vec{y}) \end{aligned} \quad (2)$$

Proof. Let $\vec{x} \in X$; by item a) from 2) one have that $d(\vec{x}, \vec{x}) = 0 \leq r \rightarrow \vec{x} R \vec{x}$ $\forall \vec{x} \in X$, say R is reflexive. Let $\vec{x} R \vec{y} \rightarrow d(\vec{x}, \vec{y}) \leq r$ and by item b) from 2) one have that $d(\vec{x}, \vec{y}) = d(\vec{y}, \vec{x})$. Hence, $d(\vec{y}, \vec{x}) \leq r \rightarrow \vec{y} R \vec{x}$, say R is symmetric. Let $\vec{x} R \vec{z}$ y $\vec{z} R \vec{y} \rightarrow d(\vec{x}, \vec{z}) \leq r_1$ and $d(\vec{z}, \vec{y}) \leq r_2$, and by item c) from 2) one have that $d(\vec{x}, \vec{y}) \leq d(\vec{x}, \vec{z}) + d(\vec{z}, \vec{y}) \leq r_1 + r_2 = r \rightarrow \vec{x} R \vec{z}$, say R is transitive. Thus, R is a binary relation of equivalence.

It is legal to measure in different manners with metrics both from the practical and mathematical points of view, because it corresponds to the general notion of distance. As it was before mentioned, different metrics [3] were employed in the algorithm, such as:

- City-block metric:

$$d_1(\vec{x}, \vec{y}) = |x_1 - y_1| + |x_2 - y_2| + |x_3 - y_3| \quad (3)$$

- Euclidean metric:

$$d_2(\vec{x}, \vec{y}) = [(x_1 - y_1)^2 + (x_2 - y_2)^2 + (x_3 - y_3)^2]^{1/2} \quad (4)$$

- Chessboard metric:

$$d_\infty(\vec{x}, \vec{y}) = \max\{|x_1 - y_1|, |x_2 - y_2|, |x_3 - y_3|\} \quad (5)$$

A relation exists between these metrics in the following manner:

$$d_\infty(\vec{x},\vec{y}) \le d_2(\vec{x},\vec{y}) \le d_1(\vec{x},\vec{y}) \quad (6)$$

The above described let us state that it is independent what metric is used in the application. However, in practice it is not true.

After the anterior definitions, we can start properly to design the topologic properties on the space (X, d). For each $\vec{x}_0$ that belongs to X we define:

$$\text{Open sphere: } B(\vec{x}_0, r) = \{\vec{x} / d(\vec{x}_0, \vec{x}) < r\} \quad (7)$$

$$\text{Closed sphere: } \overline{B}(\vec{x}_0, r) = \{\vec{x} / d(\vec{x}_0, \vec{x}) \le r\} \quad (8)$$

where d is any of defined before metrics.

In Fig. 1 shows graphically the relation between the unitary spheres, say r=1, for the different metrics d_1, d_2 y d_∞.

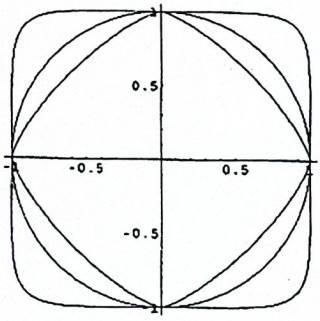


Fig.1 Relation between unitary spheres

Also, the first thing that one have to determine to carry out the application is calculate the point $\vec{x}_0$. It can be obtained calculating the expectation [5]:

$$\langle \vec{x}_0 \rangle = \sum_{i=0}^{N} \frac{x_i}{N} \quad (9)$$

3 Proposed segmentation algorithm

To process RGB images each point x that belong to X stays in a system of three-dimensional Cartesian coordinates and is denoted as:

$$\vec{x} = (\text{Red, Green, Blue}) = (x_R, x_G, x_B)$$

The segmentation algorithm of an RGB image [4], consists in dividing it according to a partition that induces the binary relation of equivalence R defined by (1). Each partition is constitutes by a region of common features that depend on the selection or subset of a set Y, y on fundamental problem that consists in finding the optimal threshold value r. Consider the original image in Fig. 2:

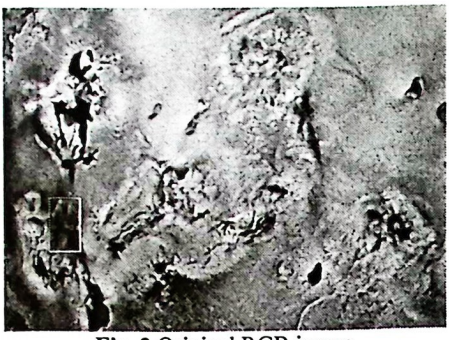


Fig. 2 Original RGB image

Using equation (9) one can obtain the expectation for each of its primary components x_R, x_G y x_B of a subset Y from X and determine $\vec{x}_0$. In Fig. 2 the considered subset Y is marked as a white rectangle. As it was mentioned above, the RGB images consist of three planes of independent images, one for each primary color [2] then the expectations are calculated for each of them:

$$\langle x_0 \rangle = \sum_{i=0}^{N} \frac{x_{R_I}}{N} \tag{10}$$

$$\langle y_0 \rangle = \sum_{i=0}^{N} \frac{y_{G_I}}{N} \tag{11}$$

$$\langle z_0 \rangle = \sum_{i=0}^{N} \frac{z_{B_I}}{N} \tag{12}$$

Once the expectations are calculated for each components of $\vec{x}_0$, i.e. (10), (11) and (12), we have that

$$\vec{x}_0 = (x_0, y_0, z_0)$$

Considering the image in Fig. 2 and applying the definition (8) together with the use of the metric $d_2(\vec{x}, \vec{y})$ with respect to the point $\vec{x}_0$ and calculating with respect to the selected subset Y, we have that one of the partitions or class $\overline{B}_2(\vec{x}_0, r) = \{\vec{x} / d_2(\vec{x}_0, \vec{x}) \leq r\}$ of the image in Fig. 2 for r=44 is:

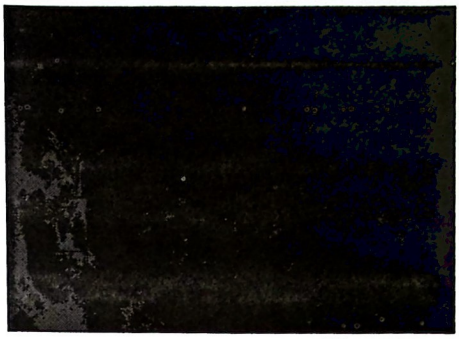


Fig. 3 Image of the class or sphere $\overline{B}_2(\vec{x}_0, r)$

Let us now apply in the same manner the same process with the same threshold r=44 to different images shown in Fig. 4, Fig. 5 but changing the metric.

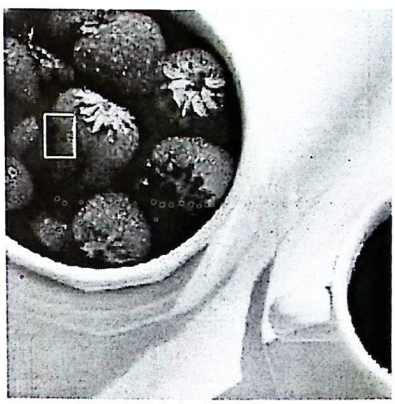


Fig. 4 Original RGB image

Now consider the image in Fig. 4. Applying the definition (8), together with the use of the metric $d_\infty(\vec{x}, \vec{y})$ with respect to the point $\vec{x}_0$ calculated with respect to the selected subset Y, we have that one of the partitions or class $\overline{B}_\infty(\vec{x}_0, r) = \{\vec{x} / d_\infty(\vec{x}_0, \vec{x}) \leq r\}$ from the image in Fig. 2 for r=44 is:

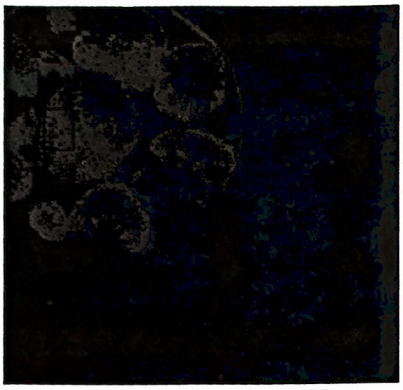


Fig. 5 Image of the class or sphere $\overline{B}_{\infty}(\vec{x}_0, r)$.

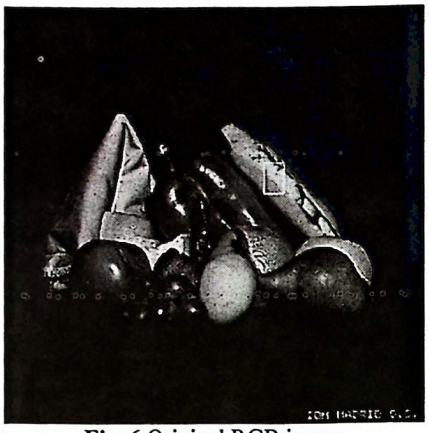


Fig. 6 Original RGB image

Finally, consider the image in Fig. 6. Applying the definition (8) together with the use of the metric $d_1(\vec{x}, \vec{y})$ with respect to the point $\vec{x}_0$ calculated with respect to the selected subset Y, we have that one of the partitions or class $\overline{B}_1(\vec{x}_0, r) = \{\vec{x} / d_1(\vec{x}_0, \vec{x}) \leq r\}$ of the image in Fig. 2 for r=44 is:

Fig. 7 Image of the class or sphere $\overline{B}_1(\vec{x}_0, r)$.

Comparing the images in Fig. 3, Fig. 5 and Fig. 7 it is clearly observed that the image in Fg.5 is defined better. In this case, the metric $d_\infty(\vec{x}, \vec{y})$ was used. Then it follows Fig. 3 with the metric $d_2(\vec{x}, \vec{y})$ and, finally, Fig.7 where $d_1(\vec{x}, \vec{y})$ was used.

4 Implementation

The presented algorithm was implemented in C++ and was applied to the common test images. Image analysis consists of image segmentation. The segmentation subdivides an image in its parts that it constitute of and objects that it conforms. Such a subdivision is performed until a level predetermined by the practical problem is reached, i.e., the segmentation must stop when the objects of interest for the particular application have been separated [9]. Steps to follow for image segmentation with the method of the binary relation of equivalence on X [10] defined en terms of a metric are:

- Select the image to analyze.
- Get a subset Y of points $\vec{x}$ of X (white rectangle that appear at the image).
- Determine the point $\vec{x}_0$ from the subset Y, and using the relations for the expectation (10), (11) and (12), determine $\vec{x}_0$.
- Select one of the metrics defined by (3) or (4) or (5).
- Define the threshold r.
- Apply the definition of the closed sphere (8), where the selected metric is compared wit respect to the threshold r having $\vec{x}_0$ as a fixed point.

5 Experimental Results

The described in the previous section algorithm were tested on a set of 512x512 images shown in Fig. 2, Fig. 4, Fig.6. In this paper we present the experimental results of the proposed algorithm for case when it was applied to each image changing the metric with a fixed threshold r=44. The results are observed in the following Table where as a measure of segmentation a function of normalized mean defined $\varphi: R^3 \to R$ is taken:

$$\varphi(\vec{x_0}) = \sum_{i=0}^{M} \frac{\rho(\vec{x_i})}{M} \qquad (13)$$

where M=(512)x(512) and $\rho: B(\vec{x_0}, r) \to N$, defined as

$$\rho(\vec{x_i}) = \begin{cases} \rho(\vec{x_{i-1}}) + 1, \text{ if } \vec{x_i} \in B(\vec{x_0}, r) \\ 0, \text{ otherwize} \end{cases}; \quad i=1,2,\ldots,M \qquad (14)$$

and $\rho(\vec{x_0}) = 0$.

Table. Normalized mean function.

$\varphi(\vec{x_0})$	Images		
	Fig. 2	Fig. 4	Fig.6
Unitary metric: $d_1(\vec{x}, \vec{y})$	0.013385	0.104860	0.023594
Euclidian metric: $d_2(\vec{x}, \vec{y})$	0.008249	0.068844	0.015865
Cubic metric: $D_\infty(\vec{x}, \vec{y})$	0.019259	0.098743	0.019767

6 Conclusions

Utilizing the proposed algorithm to realize the analysis of RGB images, employing a binary relation of equivalence defined in terms of some metric, one have to solve the problem or limit of the dependence of two parameters. One of these parameters have to be found averaging ($\vec{x}_0$) and another one, (r), in a random manner. Next, in practice the type of the used metric influences as well, until the objects are separated or until the problem of recognition is solved in some form. The latter is very subjective because depends on the observer of the segmented object. However, in the

various presented examples the way to determine the parameter $\vec{x}_0$ was found. It was found that the threshold value r=44 results in the desired segmentation for our particular application of recognition of forms and image analysis. Observing the Table, one can conclude that increasing the value of $\varphi(\overrightarrow{x_0})$, the segmentation of the analyzed image become to be better. Such a behavior depends only on the employed metric, because the threshold value was fixed for all metrics.

Acknowledgements

This work was supported by Instituto Politécnico Nacional as a part of the research project SIP#20071380.

References

[1] Fenton, N. E. y Pfllegeer, S.L., Software metrics. A rigorous and practical approach Pub. The Addison-Wesley Object Technology Series, 1997, p. 5.
[2] Rafael C. Gonzales y Richard E. Woods, Digital Image Processing. Second Edition, Pub. Prentice Hall, 2001 p. 49 y p. 290.
[3] Ward Cheney, Análisis for Applied Mathematics. Pub. Springer-Verlag New York, Inc. 2001, p. 8.
[4] Chaudhuri, B. B., A Note on Fast Algorithms for Spatial Domain Techniques in Image Processing. IEEE Trans. Syst. Man Cyb. Vol. SMC-13, No. 6, 1983, pp. 1166-1169.
[5] Dougherty, E. R. Random Processes for Imagen and Signal Processing, IEEE Press, New York, 1992.
[6] Haddon , J. F., and Boyce , J. F., Image Segmentation by Unifying Region and Boundary Information. IEEE Trans. Pattern Anal. Machine Intell., Vol. 12, No. 10, 1990, pp 929-948.
[7] Lee, S. U., Chung S. Y., and Park, R. H., A Comparative Performance Study of Several Global Thresholding Techniques for Segmentation. Comput Vision Graphics, Image Proc. Vol. 52, No. 2, 1990, pp. 171-190.
[8] Lim, J. S., Two-Dimentional Signal and Image Processing, Prentice Hall. Upper Saddle River, N. J., 1990.
[9] Liu, J., and Yang, Y. H. Multiresolution Color Image Segmentation. IEEE Trans. Pattern Anal. Machine Intell., Vol. 16, No. 7, 1994, pp. 689-700.
[10] Plataniotis, K. N., and Venetsanoapoulos, A. N., Color Image Processing and Applications, Springer-Verlag, New York, 2000.

Software Architecture Implementation of an e-HUB to offer e-Services for SMEs

Javier Espadas[1], David Concha[1], Teresa Najera[1], Nathalie Galeano[1], David Romero[1], Arturo Molina[2]

[1]CIDYT - ITESM Campus Monterrey, Monterrey, Mexico
mijail.espadas@itesm.mx, a00262912@itesm.mx, mtng@itesm.mx, ngaleano@itesm.mx, david.romero.diaz@gmail.com
[2]VIYD - ITESM Campus Monterrey, Monterrey, Mexico
armolina@itesm.mx

Abstract. The creation of industrial networks represents a value-added strategy to foster the individual competencies of Small & Medium Enterprises (SME). These networks require the support of e-services to enable the coordination and cooperation among different companies to integrate their capabilities in the configuration of virtual organizations. This paper present the integration of five e-services in an open technological platform named e-HUB. The technical design of its software architecture, the results of its implementation in a real project including the deployment of one e-service aimed to support e-supply chain processes in SMEs networks, are presented in this paper.

Keywords: e-HUB, e-Services, Business Process Management, Enterprise Architecture Integration, Software Architecture, Small and Medium Enterprises.

1. Introduction

The creation of value-added industrial networks that provide the basis for competitiveness, excellence, and agility in Small and Medium Enterprises (SMEs) requires the definition of new business models and supporting infrastructures to enable the coordination and cooperation among different SMEs, allowing them to share core competencies and resources during the creation of Virtual Organizations that responds to new global business opportunities [11] [4].

A key pre-requisite for the effective creation of value-added industrial networks such as Virtual Organizations (VOs) is the design and development of a transparent, easy-to-use, and affordable "plug-and-play" ICT-infrastructure playing the intermediary role as the enabler of interoperation among organizations [3]. Therefore, the implementation of collaborative network is facilitated by the existence of an ICT-infrastructure that allows different distributed/heterogeneous applications/actors to communicate with others transparently and seamlessly [12]. In this sense, PyME CREATIVA project intends to cover part of this gap based on the vision of an open and easy-to-access technological platform, known as "e-HUB" (Integrated e-Services Center for Virtual Business).

© L. Sánchez, O. Pogrebnyak and E. Rubio (Eds.)
Industrial Informatics
Research in Computing Science 31, 2007, pp. 193-202

PyME CREATIVA is a project funded by the Program of Multilateral Investment Found of the Interamerican Development Bank within ICT4BUS Program [7], aiming to produce a low cost infrastructure for value-added networks of SMEs through the construction of an e-HUB platform for the creation of industrial networks.

The main objective of the e-HUB is the creation of a business environment composed of value-added industrial networks collaborating around a particular technology and making use of a common architecture to deliver independent elements of value (e-services). The e-HUB intends to reduce critical troublesome that traditionally limits SMEs competitiveness, thus allowing the exploitation of new business opportunities [11]. SMEs can execute trading processes, purchase orders, supply chain management, request for quotations, and other types of e-businesses with others SME into the HUB [9]. The following e-services are being developed and integrated within the e-HUB architecture platform [11]:

- *e-Brokerage* . Integrates technologies to support the development of virtual businesses and the exploitation of business opportunities through VO creation.
- *e-Supply*. Implements technologies for the integration of manufacturing execution process, order processing tracking, and client/supplier relationship management.
- *e-Marketing*. Integrates different technologies for intelligent and customizable portals development, and customer relationship management.
- *e-Productivity* . Incorporates technologies for the diagnostic and monitoring of SME development.
- *e-Engineering*. A collaboration engineering environment that integrates design technologies for integrated product development.

The technological innovation in the e-HUB concept is achieved through three variants [11]:

- e-HUB development with Internet-based services, allowing SMEs access to a wide range of value added e-services.
- e-Services implementation methodology, to demonstrate its impact and benefit for the SMEs through an integrated process to achieve competitiveness.
- Creation and demonstration of the new SME business model based on value added industrial networks that enable the creation of virtual organizations.

The following sections will describe the architecture and technologies being used in PyME CREATIVA project to implement the e-HUB based on a selection of open source technologies and e-business models, necessaries to produce a low cost infrastructure.

2. e-HUB Software Architecture Analysis & Design

Different steps were applied to build the e-HUB infrastructure defined by a set of e-services in PyME CREATIVA project. The first step was the definition of an e-business model focused in the creation of value-added industrial networks. Secondly, the design of the e-HUB architecture reaching the requirements of service oriented economy of the future industrial networks. Thirdly, the e-services design for supporting the operation of those networks. And finally, the combination of open

source technologies with the aim of having low costs in implementing the e-HUB architecture.

Different approaches have been identified in order to design the e-HUB architecture, such as: customized corporate portals, enterprise applications integration, SOA, workflows, and business process management. These approaches try to fulfill non functional requirements such as: high level of scalability, robustness, and integration, including easy maintenance and adaptation with external systems.

The first approach is the need to have customized corporate portals; an implementation of a customized corporate portal for each enterprise that contracts the e-services offers a unified access point to its clients and suppliers besides marketing services through configurable contents. The second approach is the technologies for application integration that access remote applications and integrate different services into composite applications through service oriented architectures (SOA) such as web services. Another quite important approach is the Workflow concept and Business Process Management (BPM) in order to maintain diverse activities within the enterprises business logic or inter-enterprises business (B2B) logic. These technological approaches used in the development of the e-HUB architecture are defined and explained in the e-HUB implementation section.

Furthermore, e-services design was approached as an overarching service-centric concept based on Internet customer service and online account management services with the aim of providing an integrated solution for customized functionalities that are delivered through the Internet. The fundamental objective of e-services in PyME CREATIVA project is to have a collection of network resident software applications accessible via standardized Web protocols (HTTP, HTTPS), whose functionality can be easily discovered and integrated into applications or composed to form more complex systems. At a fundamental level, the e-services were considered as an emerging confluence of three distinct technologies: (a) process description formalisms, including automata and workflow; (b) data management (including transforms, mediation, transactions), and (c) distributed computing middleware [6] to create business modules that represent basic independent e-business service processes or functions such as authentication, authorization, advertisement, negotiation, and process integration.

Finally, the combination of these open source technologies outcomes into an innovative e-HUB software architecture that could be considered as the new generation of business process oriented architecture.

3. e-HUB Software Architecture Implementation

As stated before, different approaches supported the implementation of the e-HUB architecture: customized corporate portals, EAI, SOA, workflows and BPM. Below a brief description of each one is presented.

Corporate Portals. Portals allow users easy access to information by integrating heterogeneous applications or data sources in a consistent way under a friendly Web environment [2] [5]. Portal pages may have different sets of portlets creating content for different users. A portlet is a Java technology based on a Web component,

managed by a portlet container that processes the requests and generates dynamic content. Portlets are used by portals as pluggable user interface components that provide a presentation layer to information systems [1].

EAI & SOA. Enterprise Architecture Integration (EAI) is the key technical enabler for managing business processes. One approach to solve the integration issues is the Service Oriented Application Integration (SOAI) that allows enterprises to share common application services as well as information. Enterprises accomplish this sharing either by defining application services that they can integrate, or by providing the infrastructure for such application service sharing. Application services can be shared either by hosting them on a central server or by accessing their inter-application through distributed objects or Web services [10].

Workflow & BPM. Workflow is the automation of processes, totally or partially, during which documents, information or tasks are passed from one participant to another for a specific action that will be accomplished according to a set of procedural rules [13]. Moreover, as the e-services are becoming the focus in BPM; Workflow technology is facing the challenge to support e-services in a proper way [14]. BPM has proved to be valuable in the definition of effective business processes as workflows for everything from a single department to an entire enterprise and its associated value networks.

After the technological approaches were identified the implementation of PyME CREATIVA e-HUB was done, next paragraphs will detail this implementation.

3.1 PyME CREATIVA e-HUB Model: Architecture & Technological Platform

The intention behind the software architecture implementation in PyME CREATIVA project is to have a portal platform accessible through standard Web protocols (HTTP, WSDL) running e-services in a robust platform (J2EE). To access the e-services, SMEs must be able to integrate to this e-HUB by deploying their own Web portal (a sub-portal for each SME) that serves as a single access point to all their e-services.

Fig. 1 shows the e-HUB architecture, a technological implementation of a robust platform that involves many components and technologies. This platform includes a set of components to deploy a number of e-services and sub-portals for each SME. At top of the architecture are the clients or consumers, these can be Web browsers, mobile devices with Web browser capabilities, accessing with HTTP standard protocol or even 3rd party applications through Web services technologies.

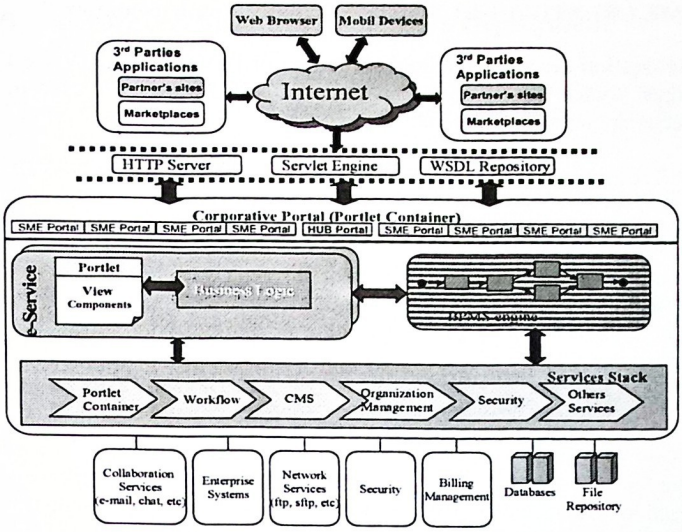


Fig. 1. e-HUB Architecture (e-Services Platform)

The main component within the platform is a corporate portal with a portlet container (Java Specification Request-168) in order to achieve hot-deployment behavior of e-services; each e-service is a portlet-application containing several view components like Java Server Pages (JSPs) or static HTML content. An e-service implementation could be defined as a composite component from a combination of user interfaces (views) and portlets (controllers) that interact with the services stack (model) explained later.

Another component is the business process system engine integrated to the platform to manage many workflows like quotation requests, work orders, purchase orders, among others; it interacts with other components through its Application Programming Interface (API) and the services stack. The component that serves as broker of the entire platform is a services stack containing the definition and implementation of applications, database connections (JDBC pools), content management system (CMS), security (LDAP, certificates), organization management (roles, users, permissions), and others. Below this platform is sited the technological physical infrastructure that supports the services stack consisting in collaboration services (SMTP, POP3), file repository, network services (FTP, SFTP), security (Kerberos, SSL), and database management systems (PostgreSQL).

At infrastructure level, security concerns are supported by J2EE technologies such as JAAS (Java Authentication and Authorization Service) combined with physical certificates through Kerberos and SSL configurations. In the business level, the security model is based on roles and access levels defined within portal platform and all user accesses are filtered for each page request. The combination of both security levels provides the e-HUB architecture with a strong robustness of information protection and isolation.

3.2 PyME CREATIVA e-HUB Model: Technical Implementation

This sub-section describes the implementation of PyME CREATIVA e-HUB model using open source technologies and open source software tools used to implement these technologies as listed in Table 1.

Table 1. Technologies and Open Source Software Tools used to implement the e-HUB

Technology	Free Software
J2EE Web Container	Apache Tomcat
Portal Platform	eXo Platform
Web Services	Apache Axis
Business Processes	jBPM
MVC framework	Apache Struts
Database Management	PostgreSQL

Portal Platform. The portal platform is eXo Platform, an open source portal technology that includes a portlet container and a services stack by providing several components like content management, LDAP and database integration, security, organization management, XML processing, and BPM implementation. Its main features are [14]: (a) Services Container that provides the inversion of control pattern design to allow services dependencies resolutions; (b) Content Management System (CMS) that provides a hierarchical organization of binary objects that can be stored in different databases or file systems with a small amount of code; (c) Portlet Container that is an open source implementation of the portlet API specification (JSR 168). This module manages the life cycle and lazy instantiation of portlet components, and the portal platform implements; and (d) User Management based on roles, groups and memberships by allowing to have access control of portlets and portals.

e-Services. In general, an e-service is an integrated solution for customize services that are delivered through the Internet, enabling their dynamic discovery, composition and delivery [8]. For its implementation, each e-service is composed by a set of components using the eXo portal platform core API and components.

Business Processes Implementation. The BPM engine integrated to eXo platform is JBoss' jBPM a flexible and extensible workflow management system. The first step was to identify the occurrence of the business process and how it helps to achieve its main goals. Once this step is performed, it is necessary to model the business process using a formal notation (e.g. UML activity diagram). Once the business process is modeled, the next step is to transform the process model into a process definition in XML language. jBPM uses a process definition language (jPdl) for process description and deployment. Each activity, decision, fork or join from the activity diagram is transformed into XML elements with its attributes.

Web Services. The integration of SMEs systems is a common need in the implementation of an e-HUB. A good example is when a legacy application or an ERP system must create a report of work orders, so it needs to check their status and tracking them. A Web service interface could achieve this kind of integration; and an open source software to develop Web services is Apache Axis, which is an implementation of the Simple Object Access Protocol (SOAP), based on XML to allow applications communicate through Internet protocols like HTTP.

4. e-HUB Software Architecture: e-Services Deployment

The e-HUB implementation comprises also the details involved in each e-service; the purpose of each e-service results in different features and needs. The e-brokerage service was selected to describe the implementation details, because it business process definition is a good example to show the mapping of e-services implementation in the e-HUB. The purpose of e-Brokerage analysis is to show the tracking of negotiation business process, starting from the business process analysis and ending in the e-service implementation.

The e-Brokerage service objective is to underpin business development through SMEs collaboration for the creation of virtual organizations (VO) by identifying the competencies needed to participate in a particular business opportunity. Therefore, collaboration between SMEs, especially those participating in a VO require the rapid establishment of such relation supported by a high level of scalability, robustness, internal/external integration and ease of maintenance ICT-infrastructure.

SMEs negotiation process is the starting point to define the roles and functionalities to be implemented by the e-Brokerage service. Even though the e-Brokerage service was defined using a more complex process model than the one presented in Fig. 2, this diagram can be used to show some points related to the e-Brokerage service design. Fig. 2 shows a business process diagram represented in Business Process Management Notation (BPMN), where every activity in this diagram ended with a corresponding use case in the e-service design and the swim lanes ended also as actors of a UML use case diagrams. Having such a mapping ensures that at least some part of the process requirements are reflected in the e-service design. Each one of the use cases functionalities can be exposed as Web Services allowing a process level integration with external systems.

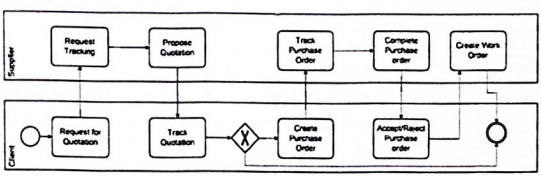


Fig. 2. e-Brokerage Simplified Process Diagram

In the e-Brokerage service context there are basically three user's types, first is the anonymous user who has not been registered in any portal and is not a member of any SME belonging to the e-HUB but is looking for a product or service in the e-HUB. The second user type is the one who belongs to an SME registered in the e-HUB and who is trying or is participating in a business transaction as a client. Finally there is a third user type; an SME registered in the e-HUB and who is acting in a business transaction as a provider. It is important to remark the fact that a user can only be a client or provider in the context of a single business transaction, but the same user can take both roles in a global context.

The functions that a user can perform are related to the role in the business transaction and also to the subscription in the e-HUB. The use case diagram presented

in Fig. 3 shows the functions that the e-service allows to the users given their user type. For example, a client can perform several actions related to the access point he gets: (1) the enterprise portal he belongs to, (2) the enterprise portal he is client of, or (3) the e-HUB's main portal. A provider can only use his own portal to get to the e-negotiation service. An anonymous user is restricted to the posting of information requests through the public section of any HUBs' enterprise portal.

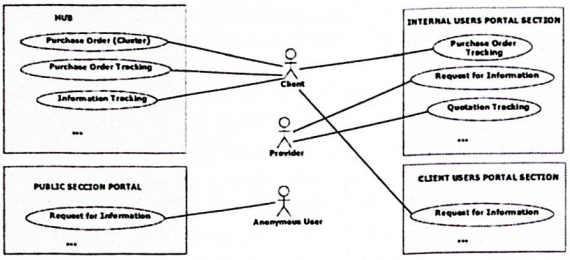


Fig. 3. e-Brokerage Service Use Cases

The BPM engine integrated to eXo is JBoss' jBPM. The language used by jBPM is jBPM Process definition language (jPdl); a jPdl process definition is an XML file containing process information. JBoss jBPM maintains the state, logs, and performs all automated actions of the process instances. The jPdl file that serves as input to the jBPM engine provides all the required information required by the engine. An example of these input file can be seen in the following XML snippet.

```
<process-definition name="RFQ-process-cluster">
  <swimlane name="client"> <delegation
class="org.jbpm.delegation.assignment.ActorAssignmentHa
ndler">client</delegation> </swimlane>
  <swimlane name="supplier"> <delegation
class="org.jbpm.delegation.assignment.ActorAssignmentHa
ndler">supplier</delegation> </swimlane>
  <start-state name="generate_request"
swimlane="client"> <transition to="evaluate_request">
    <action> <delegation
class="org.jbpm.ehub.RequestForQuotationActionHandler">
</action> </transition> </start-state>
  <state name="evaluate_request"> <assignment
swimlane="supplier"/> <transition name="accept"
to="make_quotation"/> <transition name="cancel"
to="end"/> </state>
  <state name="make_quotation"> <assignment
swimlane="supplier"/> <transition
to="evaluate_quotation"> <action> <delegation
class="org.jbpm.ehub.GenerateQuotationActionHandler"/><
/action> </transition> </state>
  <state name="evaluate_quotation"> <assignment
swimlane="client"/> <transition name="accept"
```

```
to="generate_purchase_order"/> <transition
name="negotiate" to="make_quotation "/> <transition
name="cancel" to="end"/> </state>
   <state name="generate_purchase_order">
<assignment swimlane="client"/> <transition to="end">
<action> <delegation
class="org.jbpm.ehub.POGenerationActionHandler"/></acti
on> </transition> </state> <!-- END-STATE --> <end-
state name="end"/> </process-definition>
```

Once the engine has been started a process instantiate a process activity resulting in the execution of an action class. These action classes have to be coded to achieve the functionality of each business process activity. Inside these action handlers it is possible to implement any action in order to achieve the process goal, such as send mail, change a database or call a Web service.

An actor (e.g. a client or supplier), can perform an action associated to an activity when a process arrives to this activity. In this way, all variables and business process tokens can be read into the action class and execute its code.

The user types detected were mapped to eXo Platform roles, eXo memberships are used to assign roles to new users, and finally the configuration of navigation and pages available were defined in the eXo Platform configuration files. This configuration was defined according to the eXo platform features before the deployment of each enterprise portal.

5. CONCLUSIONS & FURTHER RESEARCH

ICT-infrastructures are very important enablers for effective implementation and operation of collaborative networks as a new business model for SMEs competitiveness. PyME CREATIVA project conscious of SMEs limited resources has designed and implemented a new ICT-infrastructure (the e-HUB) supported by a scalable, robust and quite complete IT architecture able to satisfy SMEs technological requirements (low cost, easy access and operational infrastructure) for the developing of businesses opportunities over the Internet.

This paper presented the e-services HUB software architectural approaches to allow SMEs share their competences towards the creation of virtual organizations, supporting its architecture in three main IT approaches: corporate portals, enterprise application integration, and business process management.

The software components implemented for the PyME CREATIVA e-services HUB were constructed by integrating open source information technologies: a portal platform (eXo platform), portlet container (JSR-168), BPM engine (JBoss' jBPM), user interfaces components (Java Server Pages) and Web services (Apache) for Enterprise Application Integration. By developing this software implementation, the e-HUB architecture can be accessible through a Software-as-a-Service (Saas) supply model in order to deliver on-demand services based on actual and potential customers.

The PyME CREATIVA implementation is now under deployment to a large group of SMEs, personalizing their individual portals and enabling collaboration inside the marketplace defined by the e-HUB of engineering services. Future implementation of

the new e-HUB version will consider the deployment of a Content Management System, new e-services such as e-Quality based on quality models and the integration of a usability layer in the platform in order to develop new Web 2.0 business models.

6. ACKNOWLEDGEMENT

This work was possible in part by PyME CREATIVA project, supported by a grant from ICT4BUS Program of the Multilateral Investment Found (Interamerican Development Bank), and by ECOLEAD Project, funded by the European Community, FP6 IP 506958.

7. REFERENCES

1. Abdelnur, A., and Hepper, S. (2003). JSR 168 Portlet Specifications - Version 1.0
2. Bellas, F. (2004). Standards for Second Generation Portals. In Internet Computing, IEEE, Volume 8, Issue 2, pp. 54-60.
3. Camarinha-Matos, L.M. and Afsarmanesh, H. (2004). Support Infrastructures for New Collaborative Forms. In Collaborative Networked Organizations: a research agenda for emerging business models, Kluwer Academic Publishers, pp. 175-192.
4. Camarinha-Matos, L.M. and Afsarmanesh, H. (2006). Collaborative Networks: Value Creation in a Knowledge Society. In K. Wang et al (Eds.). Knowledge Enterprise, IFIP, New York: Springer Publisher, Volume 207, pp. 26-40.
5. Diaz, O. and Paz, I. (2005). Turning Web Applications into Portlets: Raising the Issues. In SAINT'05, pp. 31-37.
6. Hull, R., Benedikt, M., Christophides, V., and Su, J. (2003). e-Services: A look behind the curtain. In 22th ACM SIGMOD-SIGACT-SIGART Symposium on Principles of Database System, pp. 1-14.
7. ICT4BUS Program. (2003). PyME CREATIVA Project. Proposal for Inter-American Development Bank.
8. Kim, D.J, Manish, A., and Jayaraman, B. (2003). A Comparison of B2B E-Service Solutions. In Communications of the ACM, pp. 317-324.
9. Jimenez, G. and Espadas, J. (2006). Implementation of an e-services HUB for SMEs. In the Advanced International Conference on Telecommunications and the International Conference on Internet, Web Applications and Services, IEEE.
10. Linthicum, D.S. (2004). Next Generation Application Integration: From Simple Information to Web Services. In Addison-Wesley Information Technology Series.
11. Molina, A., Mejía R., Galeano, N., Nájera, T., and Velandia, M. (2006). The HUB as an Enabling Strategy to Achieve Smart Organizations. In István Mezgár (Ed.). Integration of ICT in Smart Organizations. Hungary IDEA group publishing, pp 68-99.
12. Rabelo, R.J., Gusmeroli, S., Arana, C., Nagellen, T. (2006). The ECOLEAD ICT Infrastructure for Collaborative Networked Organizations. In Collaborative Networks and their Breeding Environments, IFIP, Volume 224, Network-Centric Collaboration and Supporting Frameworks, (Boston: Springer), pp. 103-110, 2006.
13. Van Der Aalst, W.M.P. and M. Weskez. (2003). Advanced Topics in Workflow Management. In Journal of Integrated Design and Process Science.
14. Wu, Z., Deng, S., and Li, Y. (2004). Introducing EAI and Service Components into Process Management. In IEEE International Conference on Services Computing, pp. 271-276.

FPGA Implementation of Turbo Codification Technique for Error Control in Communication Channels

Pablo Manrique Ramírez [1], Rafael Antonio Márquez Ramírez[2], Oleksiy Pogrebnyak[3],

Instituto Politécnico Nacional, CIC-IPN, Av.Juan de Dios Batiz s/n,
Colonia Nueva Industrial Vallejo, C.P. 07738, México D.F.
E-mail: [1]pmanriq@cic.ipn.mx, [2]ramr@super.unam.mx, [3]olek@pollux.cic.ipn.mx

Abstract. A hardware implementation of turbo coding technique for error detection and correction for data transmission is presented. The designed coder generates a more efficient transmission code with a high error correction capacity at decoder in real time. The implemented techniques provides the minimum possible energy consumption and is oriented to reach a transmission speed similar to the theoretical capacity of the communication channel "The Shannon's Limit". The turbo encoder and decoder were implemented in FPGA Spartan development system. With the implemented turbo coding, the number of erroneous bits for low signal to noise ratios decreases increasing the number of iterations starting with the second iteration.

Key words: error correction code, turbo code, FPGA implementation

1 Introduction

In the codification and information theory, an error correction code (ECC) is a code in which each one of the data signals observes specific rules of "construction", so that, depending on this construction, the received signal can be detected and be corrected automatically. It is used commonly in storage of computer data, for example in dynamic RAM, and data transmission. Some examples of ECC are: algebraic codes (block codes), Hamming code, Reed-Solomon code, Reed-Muller code, binary Golay code, convolutional code, turbo code, low density parity check code (LDPCC), etc. [1].

The simplest ECC can correct single bit errors (Single Error Correction) and detect double bit errors (double error detection). Other codes can detect or correct multiple bits errors. The two main classes of ECC are block codes and convolutional codes. In 1993 Claude Berrou, Alain Glavieux and Punya Thitimajshima from the Superior National School of Telecommunications of Bretagne, France developed the Turbo Codes, the most powerful ECC at the moment. They are a class of convolutional codes whose performance in terms of binary error rate (BER) approaches to the Shannon's limit [2].

The turbo codification techniques are based on convolutional algorithms that strengthen the data to be sent adding redundant information to data at the transmitter side,

© L. Sánchez, O. Pogrebnyak and E. Rubio (Eds.)
Industrial Informatics
Research in Computing Science 31, 2007, pp. 203-212

and using iterative probability estimation algorithms at the receiver side [2],[3],[4]. They are those that more approach the theoretical limit of maximum rate of information transference on a channel with noise. We are able to aproach the AWGN capacity to a few fractions of a decibel [5]. Turbo codes does possible to increase the rate of data without the necessity to increase the transmission power, reason why also they can be used to diminish the amount of used energy to transmit to a certain rate of data.

Turbo codes can be used in various applications where one look for to obtain a maximum information transference in a limited bandwidth channel with the presence of noise. Such applications, for example, are: standards of cellular telephony of 3rd generation, satellite communications, future standards of satellite and mobile television, DVB-S (digital video broadcasting - satellite), DVB-H (digital broadcasting video - handheld) [6]. Now, turbo codes replace the Reed-Solomon codes used in telecommunications. In some future missions of the NASA space reconnaissance, the turbo codes will be used as a standard replacing RSV concatenated codes.

Actually, the refinement and implementation of the turbo codes are an active area of research in many universities and research centers.. In the paper, an attempt of efficient FPGA implementation of a turbo coder and decoder is presented.

2 Design of turbo coding

The Shannon's theorem is important in error correction. It describes the maximum attainable efficiency of an error correction scheme against the awaited noise interference levels as:

$$C = \lim_{T\to\infty} \frac{\log N(T)}{T}$$ were N(T) is the number of allowed signals of duration T . (1)

If the number of errors is smaller or equal to the correctable maximum threshold of the code, all the errors will be corrected. Thus, the error correction codes require more signal elements than necessary ones to transport the basic information [7], [8].

2.1 Encoder

The encoder is designed using a parallel concatenation of two recursive systematic encoders (RSC) and the decoder uses decoding rules on the base of iterative estimations of probabilities using identical blocks of decoding [9].

Fig. 1 shows an example of two RSC encoders that use a parallel concatenation scheme. Both elementary encoders (C_1 and C_2) use the same input signal d_k but in different sequence due to the presence of an interleaver. Thus, the turbo encoder can input the original data X_k followed by Y_{1k} sequence and later by Y_{2k} sequence. This way, the redundant parity information is added to the transmitted data making it more robust under the noise effects that will be added in the channel.

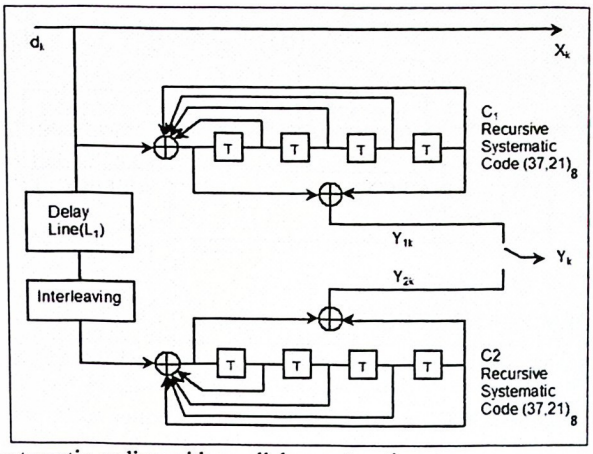


Fig. 1 Recursive systematic coding with parallel concatenation

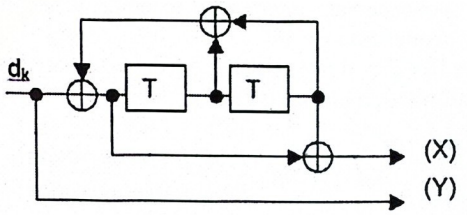


Fig. 2 RSC encoder with generators $[1, 05/07]_8$

The RSC encoder in Fig 2 was used as encoder component for the implementation. The interleaver process consists in generating a square matrix and arranging all the bits by row and reading by column in bit-reversing order. Although there are other ways and other pattern for the interleaver could be used, different patterns provide different results with significant BER differences, reason why the design of the interleaver significantly contributes in the general performance of a turbo code system.

2.2 Decoder

The original turbo decoder, is an array of decoders in a serial concatenation scheme united by interleaver [9].

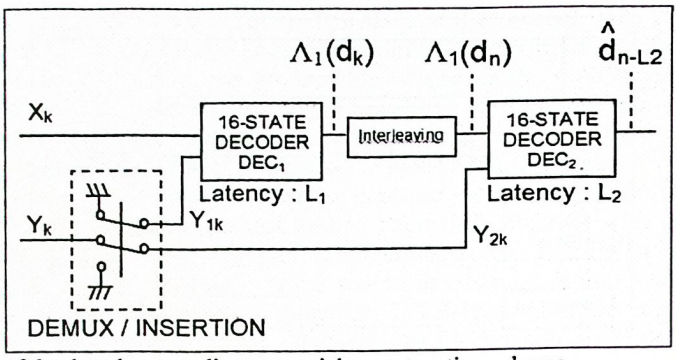


Fig. 3 Principle of the decoder according to a serial concatenation scheme

In Fig. 3, the decoder DEC_1 is associated with the sequence Y_1 and generates a soft decision. The decoding is made simultaneously for the full scope of the code word sequence. Such decoding provides good results, but it is not always feasible.

The output DEC_1 sequences are interleaved to enter at second decoder DEC_2 that is associated with the Y_2 sequences.

The logarithm of the likelihood ratio (LLR), $\Lambda_1(d_k)$ associated with each decoded bit by DEC_1 is an relevant information piece for DEC_2.

$$\Lambda_1(d_k) = \log \frac{P_r\{d_k = 1 \mid observation\}}{P_r\{d_k = 0 \mid observation\}} \tag{2}$$

where $P_r\{d_k=d \mid observation\}$, d=0,1 is the d_k data bit of the aposteriori probability (APP), and the observed data set is received from the transmission channel (y_1^N). Thus, the final decision of the decoded data becomes to be based on the sign of Λ_1.

As decoders are convolutional, these will have a certain amount of memory reason why they can be considered like "state machines", then, at time of deciding if d_k data is a 1 or a 0 the decoder will be in some of those states. $S_k = s$. Considering this already can be defined the conditional probabilities of the possible d_k value since the data set (y_1^N) has been received; and if in addition we considered that the received data set can be separated in the observed data before moment k, the present observation at moment k and the future observation after moment k

$$y_1^N = \left\{ y_1^{k-1}, y_k, y_{k+1}^N \right\} \tag{3}$$

it is possible to pass (2) to

$$\Lambda(d_k) = \log\left(\frac{P(d_k = 1, y_1^N)/P(y_1^N)}{P(d_k = 0, y_1^N)/P(y_1^N)}\right) = \log\left(\frac{\sum_{s'} P(s_{k-1} = s', d_k = 1, y_1^N)/P(y_1^N)}{\sum_{s'} P(s_{k-1} = s', d_k = 0, y_1^N)/P(y_1^N)}\right).$$

With a little manipulation of course, and using some of the BAHL's concepts in the algorithm [4], LLR becames to

$$\Lambda(d_k) = \log\left(\frac{\sum_{d=1} P(s_{k-1} = s', d_k = d, y_1^{k-1}, y_k, y_{k+1}^N)/P(y_1^N)}{\sum_{d=0} P(s_{k-1} = s', d_k = d, y_1^{k-1}, y_k, y_{k+1}^N)/P(y_1^N)}\right),$$

$$\Lambda(d_k) = \log\left(\frac{\sum_{d=1} P(s_{k-1} = s', y_1^{k-1})P(y_k^N / s_{k-1} = s')P(d_k = d, y_k / s_{k-1} = s')/P(y_1^N)}{\sum_{d=0} P(s_{k-1} = s', y_1^{k-1})P(y_k^N / s_{k-1} = s')P(d_k = d, y_k / s_{k-1} = s')/P(y_1^N)}\right).$$

With a little more manipulations and doing:

$$\alpha_k(s) = P(s_k = s, y_1^k)$$

$$\beta_{k-1}(s') = P(y_k^N / s_{k-1} = s')$$

$$\gamma_k(s', s) = P(d_k = d, y_k / s_{k-1} = s')$$

one can finally obtain LLR as

$$\Lambda(d_k) = \log\left(\frac{\sum_{d=1} \alpha_{k-1}(s')\beta_k(s)\gamma_k(s', s)/P(y_1^N)}{\sum_{d=0} \alpha_{k-1}(s')\beta_k(s)\gamma_k(s', s)/P(y_1^N)}\right) \quad (5)$$

where: $\alpha_k(s)$ is the probabilities join in s state in a k time given the last observations (fordward metric [2]), $\beta_{k-1}(s')$ is the conditional probability of the future observations given s' state in a k-1 time (reverse metric [2]), $\gamma_k(s', s)$ is the transition probability that the s_{k-1} state transits to s_k in a time k caused by d_k input.

Making simplifications and a little more of manipulation, a recursive version of $\alpha_k(s)$ and $\beta_{k-1}(s')$ can be obtained

$$\alpha_k(s) = \sum_{s'/s_k = s} P(s_{k-1} = s', y_1^{k-1})P(d_k = d, y_k / s_{k-1} = s')$$

$$\alpha_k(s) = \sum_{s'/s_k = s} \alpha_{k-1}(s')\gamma_k(s', s) = \sum_{s'} \alpha_{k-1}(s')\gamma_k(s', s),$$

$$\beta_{k-1}(s') = \sum_{s/s_{k-1}=s'} P(y_{k+1}^{N}/s_k = s)P(d_k = d, y_k / s_{k-1} = s')$$

$$\beta_{k-1}(s') = \sum_{s/s_{k-1}=s'} \beta_k(s)\gamma_k(s',s) \tag{6}$$

with the initial conditions $\alpha_0(s) = \begin{Bmatrix} 1 & s=1 \\ 0 & s \neq 1 \end{Bmatrix}, \beta_0 = \begin{Bmatrix} 1 & s=1 \\ 0 & s \neq 1 \end{Bmatrix}$.

Since the forward and backward probabilities are expressed in a recursive form, it is possible to think in a recursive scheme that could be of two forms: with feedback loop and concatenated. Fig. 4 [9] shows the feedback decoder, where W_{2k} represents the forward and backward probabilities that now are feedback to DEC_1 like a third input parameter (z_k).

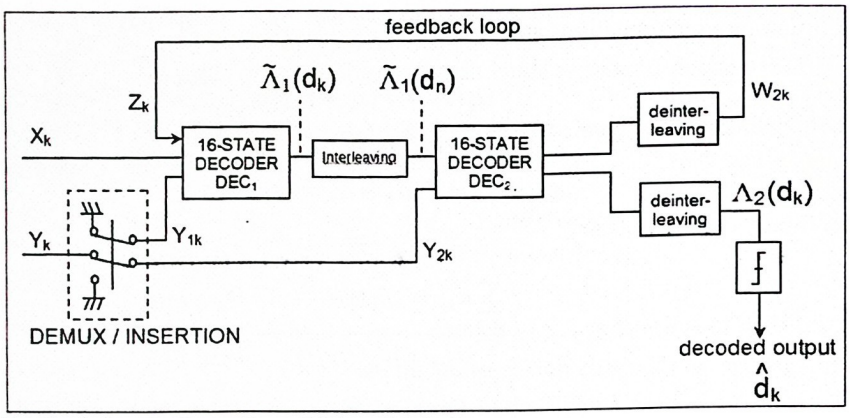


Fig. 4 Feedback decoder (assuming a 0 internal delay)

Because the first decoder DEC receives additional redundant information (z_k), this can improve the performance significantly. The turbo code term arises from this iterative scheme of decoder, remembering the turbo ignition principle in motors. Note this additional information (*extrinsic* [2], [9]) comes from an iterative previous step.

For a "modular concept" the decoder generates a delay caused by DEC_1 and DEC_2. The interleaver and deinterleaver imply that z_k information must be used through an iterative process as shown in Fig. 5, where global decoder circuit is compound of P serially concatenated identical elementary decoders. The pth DEC decoder input is formed by the output sequence of demodulator $(y)_p$ through a delay line and an extrinsic information $(z)_p$ generated by the (p-1)th DEC decoder.

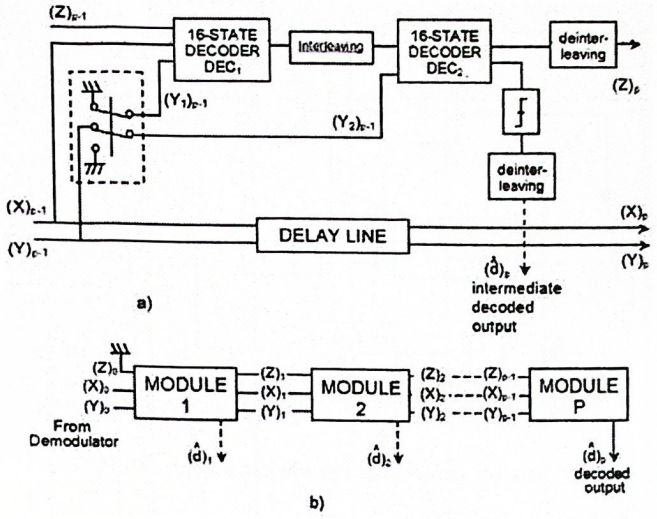


Fig. 5 a) Module decoder (level p); b) modular decoder corresponding to an iterative process of feedback decoding

3 Implementation and results

The turbo encoder and decoder was implemented in Xilinx FPGA Spartan3 development system. The turbo encoder was integrated with both encoders RSC in Fig 2 concatenated in parallel and separated by the interleaver as in Fig 1. At the output of the encoder, the original data X_k are followed by the first parity sequence Y_{1k} next by the second parity sequence Y_{2k}. The original information and the parity information redundant data are passed through the line encoder to the power line transmission channel as in Fig 7, and are transmitted in a manner more robust to the noise effects that will be added by the channel. At the channel output, the signal is mixed with additive channel noise, this signal passes through the line decoder to determine the metric that will be submited to the statistical evaluation of the turbo decoders ("maximum a priori" MAP decoder, see Fig 6).

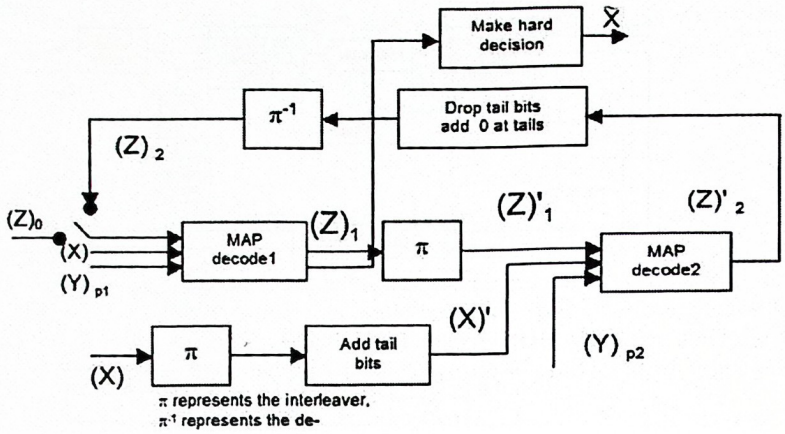


Fig. 6 Turbo decoder with iterative serial concatenation implemented in FPGA

Figure 6 shows the block diagram of the turbo decoder implemented, the most complex component is the MAP decoders, because there occurs the most intensive computing process. This process implies sums, multiplications, divisions, logarithms and exponents. Due to the nature of the MAP process the forward, backward and transition metrics can be estimated in parallel with the respective limitations of hardware resources. The logarithms and exponents are computed with iterative algorithms that restrict the performance.

The results of Berrou et al.[9] reproduced in Fig 8 show that for any signal to noise ratio greater than 0 dB, BER decreases as function of iterations p. The codification gain is high for the first values of p (p = 1, 2, 3). Thus, for p=18, for example, BER is smaller than 10^{-5} with a signal to noise ratio E_b/N_0= 0,7 dB. Shannon establish that, for a binary modulation with .5 transmission rate, P_e=0 for E_b/N_0=0 dB (several authors take $P_e = 10^{-5}$ as a reference). With a parallel concatenation of encoders RSC and feedback decoding, the performance is at 0,7 dB from Shannon's limit. Modifying the amount of memory of encoders and decoders one can obtain some degradation and inefficiencies in BER and the correction capacity added by the encoders C_1 and C_2 [9].

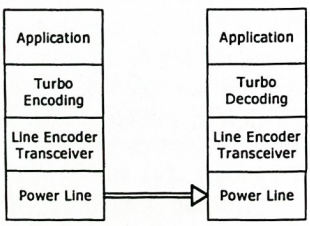


Fig. 7 Block diagram of turbo encoder-decoder test system

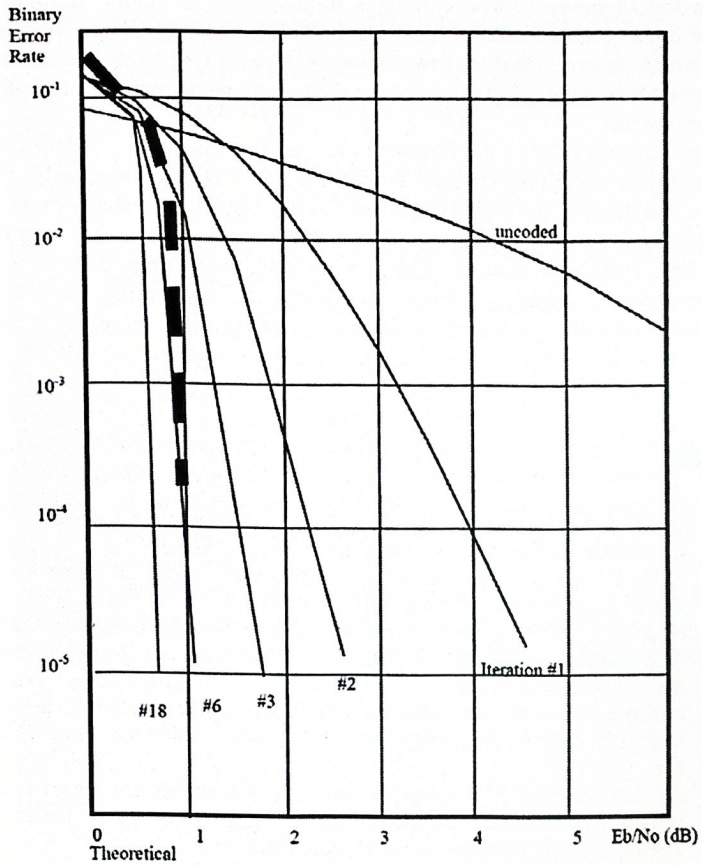


Fig. 8 Binary error rate given by iterative decoding (p=1,...,18) Of code of Fig. 1 (rate 1/2) interleaving (256x256).

Fig. 8 shows the obtained at moment by author proper results presented by the segmented bold line, for block sizes of 64 bit in approximately 6 iterations.

5 Conclusions

The development of turbo coding techniques is an intense activity in different academic and research centers. Recently, much of improvements, such as inventions, algorithms for encoders and decoders, save energy, diminution of components etc. are

presented in the literature. Turbo coding is implemented in many of high technology products that imply great amounts of data manipulation, not only in telecommunications.

The presented turbo coding implementation is approaching to the possible theoretical limit, although it needs the efforts to improve the energy consumption and optimization of computing resources.

The initial purpose of this development was an improvement of data transmissions between controllers using the electrical power line as a channel. Obviously, there are too many factors that degrade the signal quality. With this encoding system, we hope decrease the power consumed for line coding with the decrement of necessary components and energy for line coding of cousrse. The line encoder transceiver was developed with discrete components: opams, network transformers, capacitors, fast commutation transistors, etc. (very cheap). The final application depends on the existed demands that are out of our objectives now.

References

1. David J.C. MacKay, Information Theory, Inference and Learning Algorithms, Cambridge University Press, ISBN 0-521-64298-1 (2003)
2. Sklar, B., Digital Communications: Fundamentals and Applications,Second Edition, Upper Saddle River, NJ: Prentice-Hall, 2001.
3. Nguyen, Quang. High speed turbo codes decoder for 3G using pipelined SISO Log-map decoders architecture. *United States ICOM Technol ogies, Inc. (Wilmington, D)*, 2004, Patent 6813742.
4. McGraw-Hill Dictionary of Scientific and Technical Terms. Sci-Tech-Dictionary Copyright © 2003, 1994, 1989, 1984, 1978, 1976, 1974 by Mc-Graw Hill Companies Inc.
5. A. Graell i Amat, F. Brännström, and L. K. Rasmussen, Design of rate-compatible serially concatenated convolutional codes, *Int. Symp. on Turbo Codes and Related Topics*, Munich, Germany, paper 13, Apr. 2006.
6. Xilinx Inc. 3GPP2 Turbo Decoder v2.1. Product Specification LogiCORE Facts DS275, February 15, 2007.
7. Claude E. Shannon: A Mathematical Theory of Communication, *Bl System Technical Journal, Vol. 2* , pp. 379–423, 623–656, 1948.
8. C.E. Shannon, Communication in the presence of noise. *Proc. of the IRE, Vol. 37* pp.1021, Jan. 1949
9. C. Berrou, A. Glavieux, and P. Thitimajshima. Near Shannon Limit Error-Correcting Coding and Decoding: Turbo-Codes. *In Proc. 9 International Conference on Communications (ICC9)* , Geneva, Switzerland, 1993.
10. L.R.Bahl, J.Cocke, F.Jelinek, and J.Racic. Optimal decoding of linear codes for minimizing symbol error rate. *IEEE Trans, Inform. Theory, Vol. IT-0* pp.284-287, 1974.

Experiences concerning the application of electric machines steady state models for diagnostic purposes

Angel Costa Montiel[1], Miriam Vilaragut Llanes[1], and Rodrigo López Cárdenas[2]

[1] Instituto Superior Politécnico José Antonio Echeverría, Cuba
{aacm, miriamv}@electrica.cujae.edu.cu
[2] Centro de Investigación en Computación
Instituto Politécnico Nacional, México
rodrigo@sagitario.cic.ipn.mx

Abstract. In this paper, the application of steady state models of electric machines for detecting and tracking incipient faults is reported. After highlighting the simplicity and feasibility of these methods, two examples are explained: detecting inter-turn stator windings short circuits in induction motors and detecting this kind of fault, but at synchronous generators excitation windings. In the last case, the practical difficulties for fulfilling the necessary tests, advise to employ Artificial Neural Networks. The successful application of this emergent technology is demonstrated by means of a case study.

1 Introduction

Electric machines are one of the most important components of all kinds of industrial plants. Currently, all the electricity that feeds these plants is produced in electric generators and more than 60% of the electric power consumed by them is by means of electric motors.

It is evident then, the importance of keeping these machines operating in a reliable and efficient way and the necessity of reducing to a minimum its outages that are very harmful from the economical point of view.

In order to achieve these goals, it is essential to count on a predictive maintenance system based upon a continuous tracking of any incipient fault. This system makes it possible to take out the machine when a maintenance period is scheduled or, in the worse case, without any surprise [1].

The main faults that occur in an electric machines are, firstly, the mechanical ones such that rotor eccentricities or misalignment and bearings problems, as follows, the insulation

© L. Sánchez, O. Pogrebnyak and E. Rubio (Eds.)
Industrial Informatics
Research in Computing Science 31, 2007, pp. 213-222

failures that causes inter-turn short-circuits and finally the presence of broken rotor bars in squirrel cage induction motors or in synchronous generators damper winding [2] [3].

One of the most important types of method for detecting motor faults is the *Frequency Domain* that relies on a spectral frequency analysis of the stator current, the rotor vibration amplitudes, the axial flux or the machine electromagnetic torque [1-5]. Another type of method is the *Model Based*, which compares the actual behaviour of the machine with its healthy performance based on its models [1] [2] [6-14].

This paper addresses the principal aspects of this kind of method applied to electrical machines, and explains some author's experiences on its applications.

2 Electric machines models for diagnostic

The rotating electrical machines are electromechanical energy conversion devices that convert electric energy to mechanical energy or vice versa. This process is characterized for the existence of two well defined and different states or modes of operation: the *Steady State* and the *Dynamic or Transient State* [15] [16] [17].

At the steady state, the principal variables of the machine remain constant. The most important performance variables of the machine in this state are the current (rotor, stator or both), the speed, the active and reactive power consumption and the power factor. These variables are greatly influenced by the existence of any fault at the mechanical or at the electric system. For example, a damage bearing increases the current and the active power consumption, the presence of inter-turn short circuit in the rotor of a synchronous generator increases its excitation current for delivering the same active and reactive power to the mains and so on.

On the other hand, the transient or dynamic state occurs when the machine is transitioning between two steady states or is submitted to a mechanical or electric disturbance. The more important variables to be analyzed in this case are the currents, the speed and the electromagnetic torque as a function of time. The dynamic performance is also affected by a fault, generally, the transient phenomena are delayed and the oscillatory behaviour is different.

This permit to state a *Normal Operation Model,* this is the model of a "healthy machine", and comparing it with the actual machine performance (steady or transient) is possible to detect the presence of a failure and its level or depth.

As it is well known, the first stage in establishing a model is to identify adequately its parameters. Because of that, another kind of motor method for diagnostic purposes is the so called *Parameters Detection Method* since a failure situation always changes some machine parameter [2] [18-21].

Generally, it is more difficult to identify and obtain a dynamic model than a steady state model, because of that, this paper concentrates in this kind of models reporting some author's experiences. First, a method to detect incipient faults at stator winding induction

motor is treated and finally a method to detect excitation winding failures in a synchronous generator applying Artificial Neural Networks (ANN) is reported.

3 Steady state models for failure diagnostic in induction motor stator

This method is based upon the well known induction motor equivalent circuit shown in Fig. 1. With this model, after the parameter identification, it is possible to carry out the characteristics of the motor. This characteristic expresses that, in a healthy motor, for any load torque, there will be a specific speed of rotation, current and power factor.

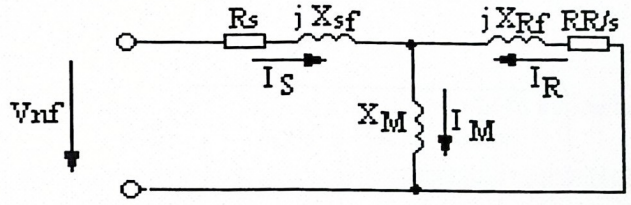


Fig 1. Induction motor equivalent circuit for a healthy motor

The parameters of this model are:

- R_S, the stator winding resistance per phase
- X_{Sf} and X_{Rf}, the leakage reactance of stator winding and rotor winding per phase.
- RR, the rotor winding resistance per phase and
- X_M, the magnetizing reactance

These parameters are identified with the no load test, the blocking rotor test at variable frequency and the measurement of the stator resistance and its temperature correction [22]. Due to the skin effect, the rotor resistance and leakage reactance are function of the slip, the corresponding equations are:

$$RR = K_1 + K_2 . \sqrt{s}$$
$$X_{Rf} = K_3 + \frac{K_4}{\sqrt{s}} \qquad (1)$$

The parameters of these equations are identified utilizing the blocking rotor test at variable frequency. On the other hand, if inter-turn short-circuits are present, these short circuited turns can be considered as a new closed winding similar to the single phase induction motor. The current that circulates through this winding creates a pulsating flux

that can be decomposed into two components: a positive sequence and a negative sequence, that are equal in value and speed of rotation but one rotates in same direction of motor rotation and another in opposite direction

This situation translated to the equivalent circuit model looks as depicted in Fig. 2.

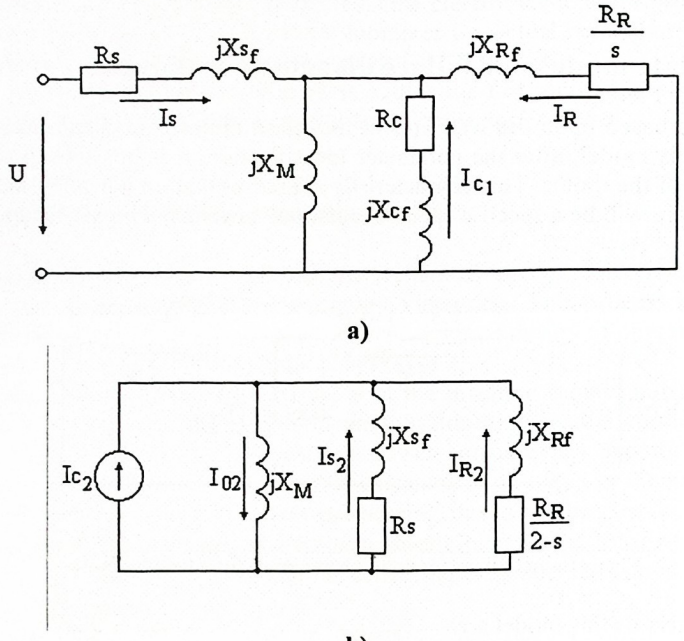


Fig 2. Induction motor equivalent circuit for a motor with short circuited turns. a) Positive sequence b) Negative sequence.

I_{C1} and I_{C2} are the positive and negative current components, and the parameters of the new branch in parallel with the magnetizing reactance are:

$$R_c = \frac{R_s}{\alpha} \qquad X_{cf} = \frac{X_{sf}}{\alpha} \tag{2}$$

In this equation, α is the percent of turns in short circuit. A MATLAB program was elaborated, such that, varying the slip from 1 to 0, calculates the current, the speed, the power factor and the torque with and without short circuited turns and finally, plots the characteristics for both cases. The results of current and power factor, applied to a 6 kV, 1

MW, 8 poles, 60 Hz motor (belonging to a feed water pump of a thermal power plant), are shown in Fig. 3 and Fig. 4.

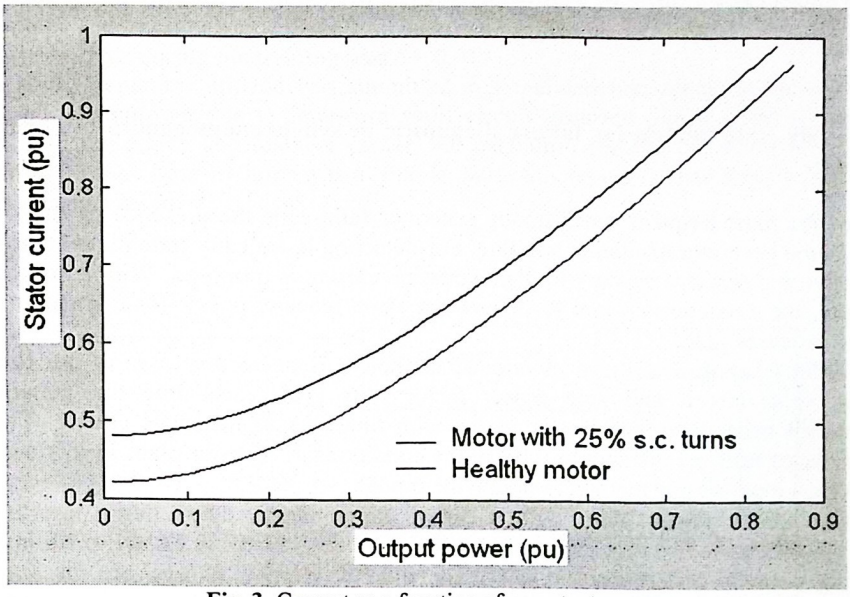


Fig. 3. Current as a function of an output power

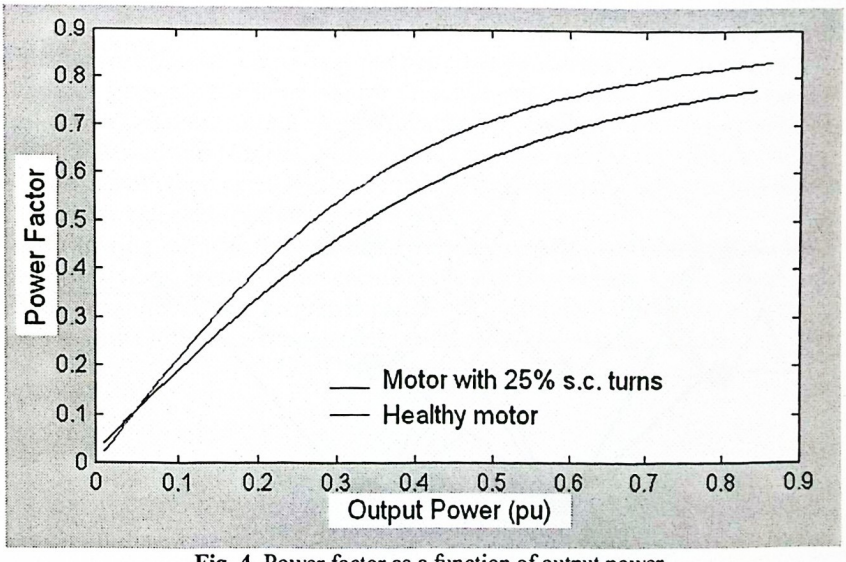


Fig. 4. Power factor as a function of output power

Since is easy to realize analyzing the figures, by measuring the differences between the motor and the model current and power factor, it is possible to detect and incipient fault and track its development because these differences increases with the depth of the failure.

4 Steady state models for failure diagnostic in synchronous generators excitation winding

One of the most frequent synchronous generator failures is the presence of short circuit turns in the generator excitation winding. For detecting it, a steady state model that relates the active and reactive power with the excitation current is proposed. With a faulted rotor winding, the excitation current to deliver the same reactive power for a constant active power increases [23].

In order to apply the above mentioned method it is necessary to carry out the open circuit, short circuit and zero power factor tests [17], each time the generator is maintained because its parameters varies with time due to materials ageing. The zero power factor tests are difficult to fulfil in the conditions of a power plant so it is advisable to devise a more simple method to obtain the model.

The solution of this problem could be carried out employing a Function Approximation. As it is well known, the best choice is an Artificial Neural Network (ANN) profiting by its learning capacity and robustness [24]. The training patterns chosen are the values of active power P and reactive power Q corresponding to a specific excitation current I_{exc}. This method was applied to a 13, 2 kV, 60 MW, turbo generator using the information collected during two years in Antonio Maceo Power Plant located at Havana, Cuba [25] [26].

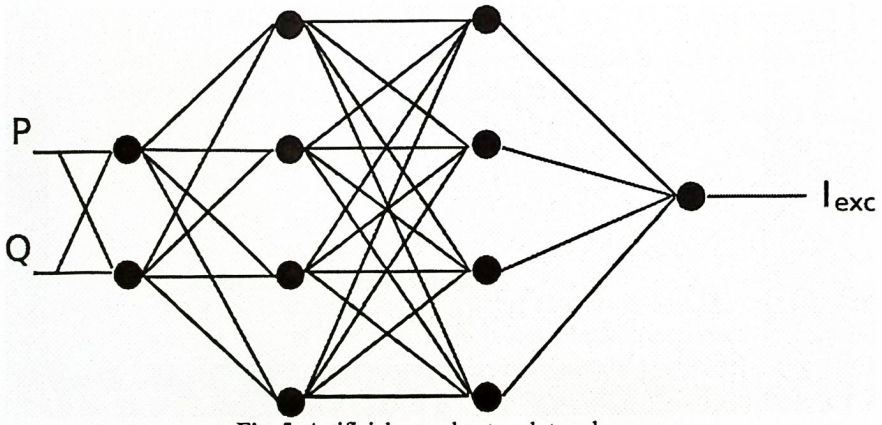


Fig. 5. Artificial neural network topology

The Levenberg Marquadt backpropagation algorithm was utilized. By means of the trail and error method the network topology and the values of the training coefficient η and momentum M were determined. In this case, a network with two hidden layers of five neuron (see Fig 5) and a value of $\eta = 0.2$ and $M = 0.9$ was the best choice. The maximum permissible error during the training was $\varepsilon = 0.01$.

The turbo generator utilized was submitted to a maintenance outage at the middle of 1996 and a shutdown due to inter-turn short circuit occurred during 1998 summer. A report with the active and reactive power, the excitation current, the stator current and excitation voltage hour by hour was available. With this report, a Data Base at Microsoft Access was prepared.

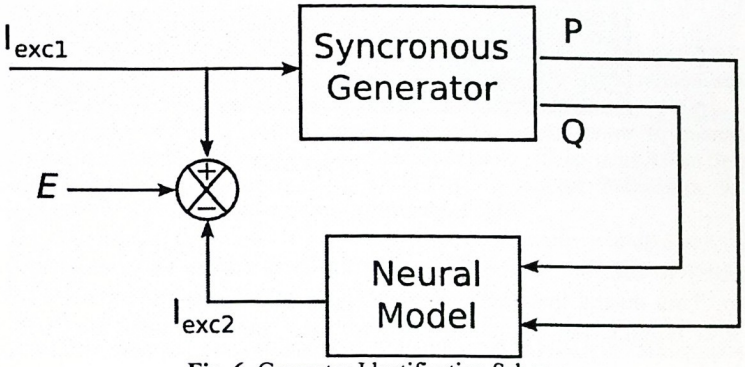


Fig. 6. Generator Identification Scheme

Due to the instruments errors and the perturbation caused by the variation of voltage and frequency, not all the same values of active and reactive power correspond to the same value of excitation current. A statistical and probabilistic analysis demonstrated that the values followed the Normal Probabilistic Law was carried out in order to justify the utilization of an average excitation current. The data appertaining to the first months after the outage were selected for training the ANN.

At the following months, the trained ANN was compared with the data taken each hour and the errors were averaged for each month. This value was called *Monthly Average Error* and was utilized as a diagnosis parameter. In figure 7 the values of this parameter for the 14 months following the selected for the ANN training, are shown.

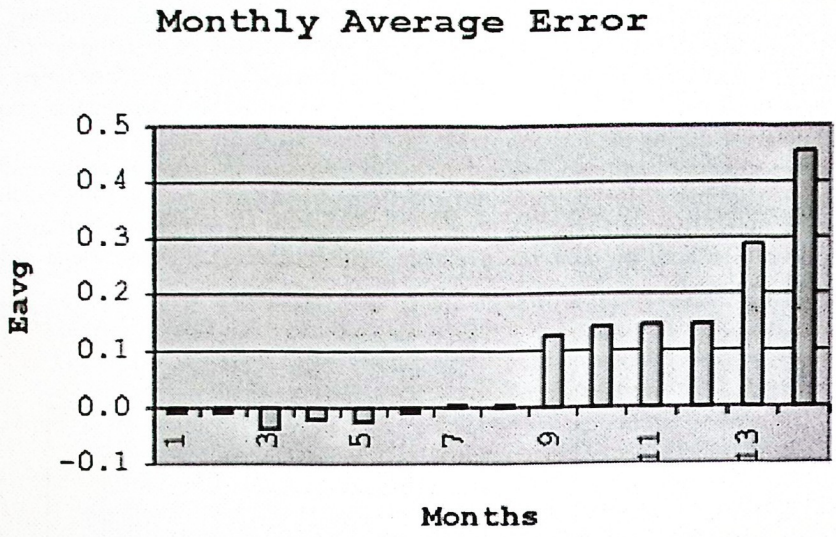


Fig. 7. Monthly average error.

Since is easy to see, before the 9th month, the errors were very small and oscillating around zero. This means that there was no appreciable inter-turn short-circuits. At the following four months, the errors are positive, their values greater than 0.1 and almost constant. At the 13th month, the error is more than doubled and at the 14th month reached a value equals to 0.45. At this time, the generator was stopped, the rotor was extracted and a great deal of the excitation winding was with inter-turn short circuit.

6 Conclusions

By means of these two examples, it is easy to realize the advantages of applying steady state models of electric machines to the diagnostic of incipient faults in order to adopt a predictive maintenance policy.

In both cases, well-known machines characteristics were utilized and, to determine them, only are necessary to fulfil conventional tests and when it is difficult or impossible to identify the needed parameters, some Artificial Intelligence (AI) Technique could be applied successfully.

These methods could be applied to the detection of another electric machine failure situation such as induction motor broken rotor bars, machine mechanical damage and other related faults.

References

1. Fernández Cabañas, M., et. al.: Técnicas para el mantenimiento predictivo y diagnóstico de maquinas eléctricas rotativas. Gijón, España, 1997.
2. Vas, P.: Parameter estimation, condition monitoring and diagnosis of electrical machine. Oxford, United Kingdom, 1993.
3. Nandi, S., Toliyat, A. H., Li, X.: Condition monitoring and fault diagnosis of electric motors, a review. IEEE Transactions on Energy Conversion, Vol 20, No. 4, pg 719-729, December 2005.
4. Cameron, J. R., Thomson, W. T., Dow, A. B.: Vibration and current monitoring for detecting air gap eccentricity in large induction motors. Procc. of the Institution of Electrical Engineers, B, vol 133, No 3, pp 155-163, May 1986.
5. Benbouzid, M. E. H., Vieira, M., Theys, C.: Induction motors fault detection, and localization using stator current advanced signal processing techniques. IEEE Transactions on Power Electronics, Vol 14, No. 1, pp 14-22, January 1999.
6. Costa, A., Dorrbercker, S.: Modelo dinámico y simulación del motor de inducción de gran potencia considerando la variación de la resistencia del rotor. Seminario Anual de Automática, Electrónica Industrial e Instrumentación, SAAEI'01. Universitat Politécnica de Catalunya, Universidad de Matanzas, Matanzas, Cuba, Septiembre 2001.
7. Soler, O., Choong J., Costa, A.: Comportamiento estacionario y transitorio del motor de induccion trifasico con espiras en cortocircuito en el estator. II Taller sobre Computación aplicada a la Ing. Eléctrica, C.Hab., Cuba, Noviembre 2002.
8. Lopez, X. M., Costa, A., Lima, M.: Dynamic induction motor model considering skin effect and thermal performance. Proceedings of the International Conference on Electrical Machines, Brugues, Belgium, June 2002.
9. Nandi, S., Toliyat, H. A.: Novel frequency domain based technique to detect incipient stator interturn faults in induction machines using stator induced voltages after switch off. IEEE Transactions on Industry Applications, Vol 38, No 1, pp 101-109, January/February 2002.
10. Mc-Inerny, S. A., Dai, Y.: Basic vibration signal processing for bearing fault detection, IEEE Transactions on Education, Vol 46, No 1, pp 149-156, February 2003.
11. Henao, H., Demian, C., Capolino, G. A.: A frequency domain detection of stator winding faults in induction machines using an external flux sensor. IEEE Transactions on Industry Applications, Vol 39, No 5, pp 12721279, Sep./Oct. 2003.
12. Toliyat, H.A., Lipo, T. A.: Transient analysis of cage induction machines under stator, rotor bar and end ring faults. IEEE Transactions on Energy Conversion, vol 10, No 2, pp 241-247, June 1995.
13. Xiangrong, C., Cocquempot, V., Christophe, C.: A model of asynchronous machines for stator fault detection and isolation. IEEE Transactions on Industrial Electronics, Vol 30, No. 5, pp 578-584, June 2003.
14. Cruz, S. M. A., Cardoso, A. J. M., Toliyat, H. A.: Diagnosis of stator, rotor, and airgap eccentricity faults in three phase inductrion motor based on the multiple reference frame theory. Proceedings of the 38th Industry Application Annual Meeting Conference, Salt Lake City, UT, pp 1340-1346, Oct. 2003.
15. Kovacs, P.: Transient Phenomena in Electrical Machines. Elsevier Publications, Amsterdam, 1986

16. Fitzgerald, A. E., Kingsley, C., Umans, S.: Electric Machinery, Ed. Mc Graw Hill, New York, 1992.
17. Sanz Feito, J.: Maquinas Eléctricas. Editorial Prentice Hall, 2002.
18. Cho, R. K., Lang, J. L., Umans, S.: Detection of broken rotor bars in induction motors using state and parameters estimation. IEEE Transactions on Industry Applications, Vol 30, pp 905-911 July/Aug 1992.
19. Sottile, J., Kohler, J. L.: An on line method to detect incipient failure of turn insulation in random wound motors. IEEE Transactions on Energy Conversion, Vol 8, No 4, pp 762-768, December 1993.
20. Huang, Ch. T., et al.: On line measurement based model parameter estimation for synchronous generators: model development and identification schemes. IEEE Transactions on Energy Conversion, Vol 9, No. 2, pp 330-336, June 1994.
21. Tsai, H., Keyhani, A., Demcko, J. A., Farmer, R. G.: On line synchronous machine parameter estimation from small disturbance operating data. IEEE Transactions on Power Systems, Vol 9, No 4, November 1994.
22. Costa, A., Vilaragut M.:, Modelos Matemáticos y Simulación del Motor de Inducción Trifásico. Monografía, CIPEL, Mayo 2005.
23. Zlatanovici, D., Velicu, S., Engster, F., Micçhailescu, D., Moraru, A.: Technologies for on line diagnosis, datta logging and prediction at turbogenerators. CIGRE 11-204, 1996.
24. Vas, P., Artificial Intelligence Based Electrical Machines and Drives. Oxford Science Publications, 1999
25. Vilaragut, M., Costa, A., López Fernández, X. M.: Diagnostic of turbogenerator rotor shortcircuit turns utilizing Artificial Neural Networks. Proceedings of the International Conference on Electrical Machines, Brugues, Belgium, June 2002.
26. Vilaragut, M., Costa, A.: Comparación de redes inteligentes aplicadas al diagnóstico de espiras en cortocircuuito en el rotor de los generadores sincrónicos, Anales del Congreso Venezolano de Ingeniería Eléctrica, CVIE 2004, Caracas, Septiembre 2004.

Development of Mobile Robots - A Proposal of an Open and Flexible Control Architecture

Wilson Massashiro Yonezawa[1], Humberto Ferasoli Filho[1], Renê Pegoraro[1], Marco Antonio Corbucci Caldeira[1] and João Maurício Rosário[2]

[1] Unesp – Universidade Estadual Paulista – campus de Bauru, Departamento de Computação Bauru – SP, Brazil
[2] Unicamp – Universidade de Campinas, Laboratório de Automação Integrada e Robótica - FEM Campinas– SP, Brazil
{ yonezawa@fc.unesp.br, ferasoli@fc.unesp.br, pegoraro@fc.unesp.br, caldeira@fc.unesp.br, rosario@fem.unicamp.br }

Abstract. This work presents a model of an open control architecture for the development of applications with mobile robots. The software explores the computational power of mobile devices (i.e. PDA) for the embedded system and the system of wireless communication for access to a remote computer and the Internet. The software architecture, with hierarchic characteristics, distributed in increasing levels of complexity with resources of reconfigurable computation and embedded systems exploring the technology of mobile devices, directed to the cooperative development of applications of the Mobile Robotics. The hardware is based on a net of sensors and actuators. Each sensor and each actuator are controlled individually by microcontrollers. The modules of the developed hardware and software are structuralized in independent parts, considering architecture of open implementation and allowing the easy expansion of the system (software and hardware), better adapting the platform to the diverse tasks associated to it.

Keywords: Embedded Systems, Control Architecture, Reconfigurable Systems, PDA devices, Wireless networks.

1 Introduction

The development of control architectures of mobile robots faces some difficulties as of hardware as of software. Each application can demand particular situations of development either for the mechanical aspect, for the electronic or the computational one. Thus, an interesting model of development is well characterized for the easiness of adaptation to a new situation and allowing the reutilization of parts or the whole of other initiatives. This means that the demanded flexibility requires a model primitively well established of hardware and software that offers an environment of open development, flexible and scaleable. Possibly, this model can be obtained based in an operational net of sensors and actuators and by a management protocol of resources not centered and with computational embedded power that allows the

© L. Sánchez, O. Pogrebnyak and E. Rubio (Eds.)
Industrial Informatics
Research in Computing Science 31, 2007, pp. 223-232

implementation of embedded control loops supported by the one of the state of the art technology.

The existing computational capacity in the PDA current allows its use in different domain of application, for example the use of PDA as embedded computer for use in mobile robots, which used as computational device with communication easiness, allows the reduction of the complexity in the construction of mobile robotic systems.

The use of PDA, with wireless communication, allied with the architecture concept client/server (C/S) facilitates the construction and extends the universe of use of mobile robotic systems, eliminating the designer of the necessity of conception of the whole system of communication between the mobile robot and the central station of control, inasmuch as most of the functionalities is implemented in the operational system and in the hardware integrated to the PDA.

At the same time can assist the designer offering an open platform for software development, this means, most of the time the operational system of the PDA offers a set of functions for development of applications known as API (Application Programming Interface). It extends due to the offering of a solid and flexible base for the project of robots, making possible to concentrate only in the construction of the application of the robot in specific domains.

Proposals and projects of robotic systems based in architecture C/S are not new features, [1], [2] and [3] are argued in this work, and systems of controls of long-distance robots using Internet technology can be found in diverse works such as [4] and [5].

On the other hand, as cited, the flexibility offered by the software architecture must be followed by the hardware architecture. This way, the hardware architecture is based on a net of sensors and actuators. Each point of access, a sensor or an actuator, is carried through by a microcomputer, forming for this set, a module. Modules can be inserted or removed from the net according to the necessity of each application. Moreover, the standard of connection established, followed by preset characteristics of software, becomes the net formed for the modules open and reusable. Each microcomputer presents characteristics according to requirements of the sensor or actuator by it controlled. The microcontrollers communicate through the physical layer available by its manufacturer. A central module of control does not exist, but a module of entrance and exit of the net that communicates with the PDA to gain wireless access to external computer network. Moreover, the PDA is hold as the element of opened and flexible processing, of high level, for the system of embedded control.

The objective of this work is to present a proposal for construction of robotic systems mobile autonomous worker on the basis of the concepts of architecture of software based on the concept Client and Server, in a net for slide bars of sensors and actuators, open and flexible, and nets of wireless communication that support the some architectural approach and applications.

2 Architecture of Control

The project of mobile robots requires the evaluation of a considerable set of variables. Questions such as degree of autonomy, mobility and navigation through the environment directly influence the conception and construction of the robot. The growth of the functionalities of a robot is directly proportional to the complexity of the same. However, nor always it is possible to accommodate in a mobile robot all the desired functionalities. To balance desires and restrictions in a project of mobile robots is, most of the time, a complex task.

Unlike a software project, a mobile robot project requires the integration of elements of software electronic and mechanic engineering. Many times is necessary to choose or even to build all constituent elements of a mobile robot, which is, since the hardware until the software components, as operational system or monitorial program, of the robot and the application properly said.

The development of a mobile robot certainly is guided by the application. The environment determines the physical characteristics and the form of robot's navigation.

However, besides the navigation, other questions must be managed to keep the integrity and the survival of the robot. To guarantee the navigation of the robot for the environment allows the accomplishment of the task demanded for the application.

To minimize the increasing complexity of construction of mobile robotic systems, it can opt to the job of partition strategies and abstraction. The partition consists on the division of a bigger problem in smaller parts (divide and conquest) and the abstraction, in determining different levels of services among a function.

The use of architecture client/server (C/S) is a form of partition and abstraction. With this architecture it is possible to divide and to distribute the functions of a mobile robotic system between two separate entities. The client side is implemented in the mobile robot and the server side in a fixed computational system, that is, in a computer. The communication between the parts client and server is carried through a computer network. The main objective of the choice of an architecture C/S is to gain flexibility in the project of mobile robots.

A flexibility example is in the implementation of the process of decision of action of the robot. One part of the process of decision can be implemented in the proper mobile robot and another one in the server, as it shows Fig. 1. Questions related to the survival of the robot must be processed quickly and locally. More complex decisions can be passed to an up level coordinator.

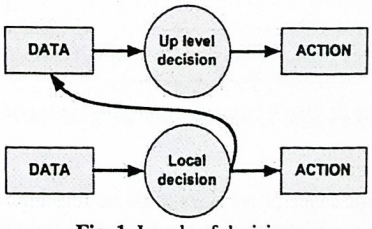


Fig. 1. Levels of decision.

The proposal contemplates the abstraction by means of the use of a structure of layers of services as in the side client as in the server side. The main objective is to facilitate the development of applications by means of the fast prototype with reuse of components, as much of software as of hardware. Each layer offers a set of specific services for the immediately superior layer. Inside of this model of construction of mobile robotic system, whoever develops the application of use of the robot does not need to construct everything from the zero, is only enough to know the functionalities offered for the inferior layers. Moreover, the considered model can be extended, that is, new functions in any layer, can be created and offered.
The proposed structure is composed of three layers: layer of application, layer of services and layer of the robot, as shown in Fig.2. In the first layer, inferior level, is the layer of the hardware of the robot. This layer supplies the structural base (mechanics) of the mobile robot, presenting the mechanical characteristics and the system of sensors, as well as the firmware that implements the local control of the robot. The intermediate layer is the layer of services. This layer is divided in two sub layers: sub layer of the client and sub layer of the server. In the sub layer of the client are implemented the functionalities that allow the construction of the application that controls the robot directly, as well as the functions that allow the communication between the sub layer of the client and the sub layer of the server. One of the objectives of the sub layer of the client is to allow the implementation of reactive architectures, behavioral or hybrid in local level, that is, next to the robot. The incorporation of processing power in the robot or next to the robot is a form of supporting this. The sub layer of the server has as objective to extend the functionalities of a mobile robot, allowing an additional level of implementation of architectures with more deliberative profile. With the sub layer of the server it is possible to generate a superior level of taking decision of the robot, besides allowing the construction of a central repository of information captured by the robot.

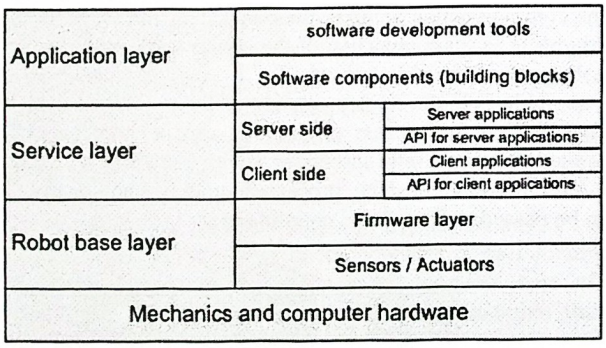


Fig. 2. Block Diagram of the proposed architecture.

All the functions of the layer of services must be implemented and packed in an API or even, in two distinct APIs: one of the sub layer of the client and another one of the sub layer of the server. In the superior layer, it is the application layer. In the

application layer there are the tools of development and the components of software for construction of the applications of or for the layer of services, as much of the server side as the client side. The components of software of the application layer are abstractions in the class form and object of the API of the layers of the client and the server. The application layer must favor the construction of tools of development that use the software components and facilitate the development of applications.

In the proposed architecture, the interface between the layers is implemented through software components (objects and classes) that carry through calls to an API specifically constructed to offer support to each layer. Each API in each layer implements primitive functions of operation of the layer, as well as services of communication with the adjacent layers.

The sub layer of the client is embedded physically in the robot and its implementation consists of the use of architecture of flexible control, in the case of this work, of a device PDA. The decision for the use of a PDA is based on the fact that it is a ready hardware that uses standardized interfaces, as serial interface and wireless communication, and an incorporated operational system. This set of characteristics favors and reduces the effort for the development of the components of software and the API necessary for implementation of the sub layer in question.

It is relevant to detach that the PDA could be substituted by specific hardware, for example, a plate of reduced size of a same PC architecture or for a customized device based in microcontroller. However, such solutions would demand greater efforts to configure the environment, since the hardware until the installation of an embedded operational system. In the case of the use of a micro controlled device, the necessary efforts for implementation of such easiness would still be bigger mainly considering the versatility searched in this proposal of architecture.

The layer of the robot or mobile robotic base is notoriously directed to the hardware and requires, also, cares due to its mechanic construction. The function of the robotic base is to allow to the mobility of the robot and the acquisition of information of the operation environment. In the proposal of this work, the base of the robot must be open enough to receive expansions, such as, inclusion of new types of sensors and/or actuators. For being highly dependent of the hardware, the layer of the robot requires special considerations related to its construction. A proposal would be to consider each sensor or actuator to be used in this layer, as a complete subsystem, which means, that it encapsulates all the functionalities and descriptions of functioning of the same. Instead of using a luminosity sensor, it would be necessary to construct a small system of capitation of luminosity. In the specification of this module the interfaces for configuration and use of the data collected for the system would be described. Easiness in the interconnection of these sensory modules or actuators in the architecture as a whole would be the main advantage in the use of this type of approach. The modules sensory /actuators could be integrated in a robotic base for an internal net of data based, for example, in standard I_2C or even USB. Fig. 3 shows the block diagram of the components of the internal net of data of the robot layer.

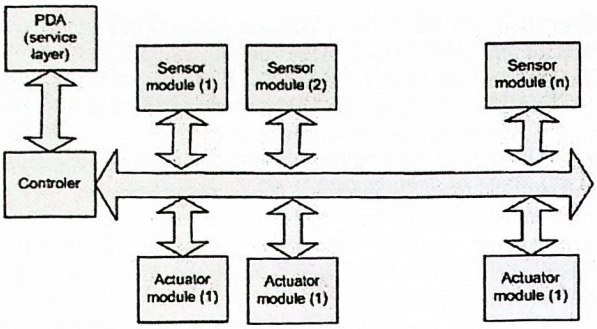


Fig. 3. Internal net of the robot layer.

3 Construction of the proposed architecture

The development of the proposed architecture requires the passage for different levels. The first one is associated with the construction of the proper architecture, which means, in the development of the API in the part of client as well as the part of the server; in the construction of a mobile base with a set of sensory modules and actuators; in the implementation of a protocol of control of the internal net of data of the robot; configuration of the elements of the wireless communication network that binds the robot to the server; in the offering of a minimum set of classes and objects for the construction of applications for the mobile robot and in the proper validation of the considered model.

The second level consists on the improvement and expansion of the considered model. Improvements through the addition of new functionalities and construction of environment of development of applications but directed to the final user.

The last stage is in the use of the architecture (Fig. 4) by the final user. In this point, the architecture must offer the necessary abstraction so that a user can create an application with what is available for it.

The mobile robot implements part of the layer of services, more specifically, the sub layer of the client, as well as all the layer of the robot, as presented in Fig.5. The sub layer of the client is contained on the PDA, implemented in the form of a set of functions on the proper API of the operational system, for example, Windows CE or Palm OS. This way the API client carries through calls to the API of the operational system of the PDA. The communication between the sub layer of the client and the module of the robot is carried through by a direct connection between the PDA and the mobile base. The direct connection can be implemented through the use of serial communication standard, USB or IR (infrared) depending on the implementation of the control module of the internal net of data of the robot. The internal net of data of the robot, binds to the controlling element with the sensory modules and actuators (Fig. 4).

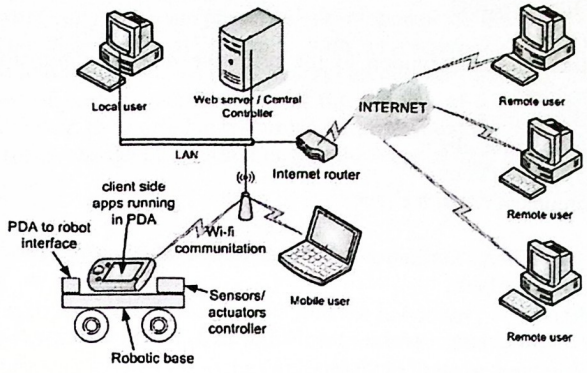


Fig. 4. System architecture.

All of the implementation of the module of control of the internal net of data of the robot is contained in firmware incorporated to the mobile base of the robot. In such a way, such firmware is responsible for the interface between the layer of the robot as the layer of services, in this case, the sub layer of client.

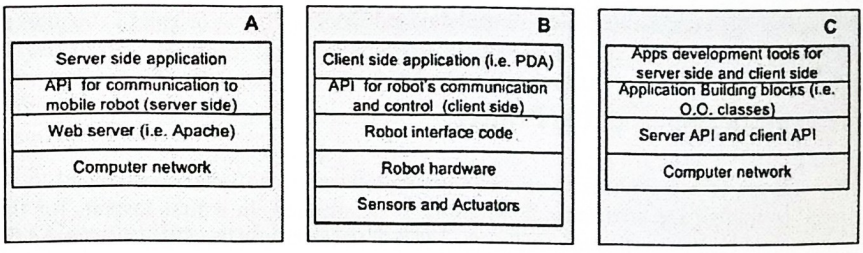


Fig. 5. Server, Client and Application.

The communication between the sub layer of the client and sub layer of the server is carried through by Internet technology, more specifically, through protocol TCP/IP. Functions of communication between sub layers of services are implemented as in the API of the sub layer of the client as in the API of the sub layer of the server. To facilitate the implementation of the side of the server, an HTTP server, for example, Apache, can be used to handle the communication. The application layer can be also divides between client and server. An application can be constructed to be executed only in the client, which means, loaded in the PDA for direct control of the mobile robot or can be divided between part in the client and part in the server. Functions of the API of the layers of the client and server offer support to this type of application. The functions of the API of the layer of services, are encapsulated in classes and

objects, called construction blocks of application, that provide a base construction of applications, Fig. 6, part C.

A- Side of the server: Definition of the primitive of communication between the server and the client.
- Send data to the client
- Receive data from the client
- Obtain status of the client
- Initiate communication with client

B - Side of the Client: Definition of the primitive of communication between the client and the robotic base.
- Main functions:
    - i) movement of the base: Right, Left, Front, Reverse Stopped;
    - ii) activation of devices
    - iii) reading of data and status of the sensors
- Definition of the primitive of communication between the client and the server.
    - i) To send data for the server
    - ii) It receives data from the server
    - iii) To obtain status of the server
    - iv) To initiate communication with server

C - Development of applications (Component Building Blocks)
- Robotic base
- Interface between the Robotic base and device PDA

4 Considerations of the Project

The success of a project of an autonomous mobile robot is directly associated to its mobility though the environment. Mobility is determined, in a first instant, for the mechanical characteristics inferred to the robot referred to the limitations imposed by the environment. In a second instant, the inserted sensors and actuators contribute to determine and to refine situations that better contribute for the mobility and survival of the robot in the environment. This entire situation is directed to the navigation of the robot in the environment.
The navigation difficulty is related with mobility and the quality attributed to the robot of feeling and reacting in the environment. Increased to this, weather the previous knowledge of the environment for the robot, falls again on as the architecture of control will treat the set of sensorial information and as to act to adapt itself in this environment. Certainly, the fact of the environment be dynamic propitiates some situations that determine the architecture of control, more reactive or more deliberative options, for the decisions taken during the navigation or the survival of the robot in the environment.
Historically, some architectural approaches register the search for the negotiation between the reactive extremity and the deliberative one in the control of the mobile

robots. Anyway, solutions found for a situation hardly are adapted to others. This evidences the difficulty of conception and development of architecture of control. Therefore, a development environment, with characteristics that facilitate the implementation and modifications in the control architecture and with modularity that allows the reuse of parts of architecture for the creation of a new one is extremely interesting. Moreover, the available resources follow an established and open standard that facilitates the exchange of information and the reuse among developed projects.

5 Proposal of Implementation

The described architecture is in development in the Laboratory of Systems Integration and Intelligent Devices - LISDI of UNESP campus of Bauru - Brazil. The versatility of the architecture can in such a way be explored in the development and validation of diverse architectures varying from the most reactive until most deliberative, [6], [7] and [8] as a group of benches of development for industrial mobile robotic applications in environment or specific applications as, for example, mobile robots for didactic use. Architecture open as proposed in this work, when adopted for a community, tends to grow, as much in the addition of functionalities as in the number of applications.

The model considered in open architecture offers different fronts of research. Research in the application layer can be directed to the construction of components of software and environments of software development. The layer of services can be extended, adding to new functions such as in the serving sub layer as in the client sub layer. In the first layer, layer of the robot, studies on new protocols of control of the internal net of data and also proposals for standardization for specification and construction of sensory modules and actuators, are subjects that deserve greater inquiry.

6 Conclusions and Future Works

This work presents a model of architecture for the project of mobile robots using PDA and devices of wireless net. The control structure is carried through using concepts of reconfigurable embedded electronics and considering the division of the system in diverse functional blocks, allowing a fast adaptation of the system to the new dedicated applications (service robots). This implementation was carried through by a multidisciplinary team, allowing to the consolidation of diverse works of undergraduation and graduation. This way, the implemented platform became a powerful environment for diverse activities of research, as for example, the validation of control strategies and supervision, generation of models of communication protocols, fusion and treatment of sensors, among others.

To facilitate the future modifications, implementations and tests, the use of reconfigurable embedded electronics in diverse blocks of the project is strongly emphasized. In fact, it was observed that systems of development based in reconfigurable logic present as characteristic a great capacity of adaptation to new

demands of the project. Another desired objective with this work is to get knowledge on this type of manipulator, aiming to provide alternative solutions for practical industrial problems, in the areas of maintenance, supervision, simulation and manufacture.

Some promising aspects of the resultant product are:

- It presents flexibility, because it has a great variety of possible configurations for solution of diverse types of practical problems.
- Is a powerful tool for validation of prototypes, allowing a simple solution for integration of sensors and actuators.
- Due to the use of reconfigurable logic, presents the possibility of modification of the strategies of control in real time.
- The implemented open architecture allows its use in academic activities.

The proposals of future activities are presented as:

- Optimization of the current prototype in terms of project and modularization.
- Validation of the architecture of control for other tasks, with new combination of sensors and actuators.
- Study and implementation of strategies of cooperation between different industrial mobile robots and services.

References

1. Brugali, D., Fayad, M.E.: Distributed Computing in Robotics and Automation. IEEE Transaction on Robotics and Automation. Vol. 18, n. 4, (2002) 409-420.
2. Kubitz, O., Berger, M.O., Stenzel, R.: Cliente-Server-Based Mobile Robot Control. IEEE/ASME Transactions on Mechatronics, vol 3, n. 2, (1998) 82-90.
3. Berger, M.O.; et al. A Modular Layered Client-Server Control Architecture for Autonomous Mobile Robots. In Proceedings of the IEEE International Symposium on Industrial Electronics. Vol. 2 , (1997) 7-11.
4. Han, K.H.; et al. Implementation of Internet-Based Personal Robot with Internet Control Architecture. In: Proceedings of the 2001 IEEE International Conference on Robotics & Automation, Seoul, (2001) 21-26.
5. Ohchi, M.; et al. Development of Autonomously Móbile Robot with Control System Connected by TCP/IP Network. In: Proceedings of IECON'01: The 27th Annual Conference of the IEEE Industrial Electronics Society, (20010 311 – 316.
6. Brooks, R.A. A Robust Layered Control System for a Mobile Robot. IEEE Journal of Robotics and Automation, Vol. RA-2, n. 1, (1986)14-23.
7. Gat, E.: Integrating Planning and Reacting in a Heterogeneous Asynchronous Architecture for Controlling Real-World Mobile Robots. Proceedings of the AAAI92, (1992).
8. Elfes, A.: A Distributed Control Architecture for an Autonomous Mobile Robot. Artificial Intelligence, Vol.1, No 2, (1986) 135-144.

Nonlinear Modelling of Water Quality in Shrimp Ponds based on Artificial Neural Networks

José Juan Carbajal[1], Luís Pastor Sánchez [1], Jorge Hernández[2]

[1]Centro de Investigación en Computación, Instituto Politécnico Nacional,
Av. Juan de Dios Batiz s/n, Colonia Nueva Industrial Vallejo, México D.F.
juancarvajal@sagitario.cic.ipn.mx, lsanchez@cic.ipn.mx
[2]Laboratorio de Análisis Integral Acuícola, Centro de Investigaciones Biológicas del Noreste,
Centenario Nte. 53, P. del Centenario, Hermosillo, Sonora, México
Jhlopez04@cibnor.mx

Abstract. Water quality is an important factor for a good maturing and reproduction of the organisms in a water environment. A model that describes its behaviour in shrimp ponds is used to determine good or bad situations that could generate a low growing or lethal problems. This work proposes a non-linear model of water quality using various types of architectures of artificial neural networks (ANN) for recognizing the status into the pond. ANN's are used to establish a relationship among environmental variables that affect the shrimp habitat. The results show a good performance of our approach, giving an alternative tool for the aquaculture field.

Keywords: Water quality, neural networks, aquaculture.

1 Introduction

The water quality in oceanic research is a problem that affects daily the activities of many people that practice fishing activities. The quantity of biological data is increasing each day and it is needed to create models and use some functional features.

The shrimp farming is an important activity in many countries because their economy, that is why there exists many companies that practice it. The farming shrimp is made in different ways; in intensive, semi intensive or extensive ponds. The shrimp production is determined by two main factors: 1) the capacity of maturing in the organisms and 2) the capacity of the environment. The capacity of the environment is referred to the conditions that allow a growing and reproduction, whose would be the best in good environment conditions. The water quality into the farming ponds determines the capacity of the environment that makes influence in the life of the organism [1], [2].

Water quality indicators have been grouped in three categories: physical, chemical and biological, each of them containing a significant number of water quality variables. There are variables that have more significance because they can affect the growing and the surviving of the organism, for example: temperature, salinity, dissolved oxygen, pH, Alkalinity, ammonia and nitrites. The values and the combinations of these variables, allow creating normal patterns. The Table 1 shows the range of the environmental

© L. Sánchez, O. Pogrebnyak and E. Rubio (Eds.)
Industrial Informatics
Research in Computing Science 31, 2007, pp. 233-242

variables and its classifications, these are the most common ranges used by the biologists in the Gulf of California (Páez, O. 2001).

Table 1. Classification of the water quality status using the combination of environmental variables.

GOOD	
Temperature	23 – 30 °C
Salinity	15 – 25 (mg/l)
PH	7.6 – 8.6
Dissolved oxygen	6 – 10 (mg/l)
Turbidity	35 – 45 (mg/l)
RISK	
Salinity	Greater than 25 (mg/l)
PH	Less than 7.6 or greater than 8.6
Dissolved oxygen	Less than 6 or greater than 10 (mg/l)
Turbidity	Less than 35 or greater than 45 (mg/l)

In general, the environmental variables have nonlinear relations, which have been observed and proven experimentally. The equations that represent them, have been formulated, which is very hard to do, for example in the Gulf of California the pH in seawater is a function of total CO_2 (*Ct*), total alkalinity (*At*) and pressure (*P*):

$$\partial \mathrm{pH} = \left(\frac{\partial \mathrm{pH}}{\partial \mathrm{Ct}}\right)\partial \mathrm{Ct} + \left(\frac{\partial \mathrm{pH}}{\partial \mathrm{At}}\right)\partial \mathrm{At} + \left(\frac{\partial \mathrm{pH}}{\partial \mathrm{P}}\right)\partial \mathrm{P} \tag{1}$$

Where the quantities in brackets are the partial derivatives of pH with respect *Ct*, *At* and *P*, respectively [2], [3], [4], [5].

We can represent these nonlinear relations following the equation 2:

$$\text{Habitat status (water quality)} = f(\text{Temp, Sal, Turb, At, Do } \ldots) . \tag{2}$$

Where:

f: nonlinear function; *Temp*: temperature; *Sal*: salinity; *Turb*: turbidity and *Do*: dissolved oxygen.

The need for appropriated techniques to manage the importance of water quality variables, the interpretation and the method used to integrate those variables involved in the evaluation is recognized (Fig. 1). In this sense some methodologies have been developed from the artificial intelligence. These methodologies have been based mainly on artificial neural networks (ANN), fuzzy logic and mathematical reasoning. The main problem with the mathematical reasoning is that an equation that describes the behavior of the water quality is very hard to develop, and some techniques like ANN are a great solution [6], [7].

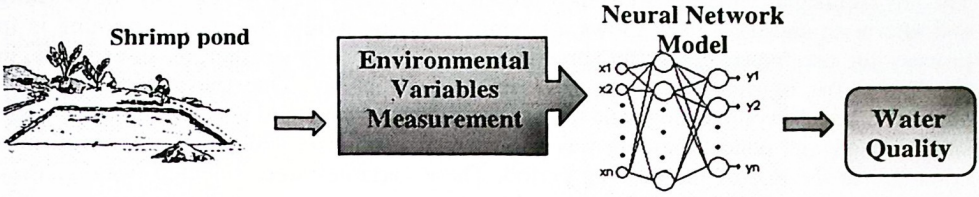


Fig. 1. The measurements of the environment will be analyzed by a model that describes the relationship among them, obtaining a status of water quality.

The objective of this study was to develop a model that describes a relationship between the environmental variables through the use of non-linear system identification techniques. Some variables are measured more frequently than other because they have more impact in the status of the habitat; these variables were taken as an input to the system to asses the water quality (temperature, salinity, pH, dissolved oxygen and turbidity). Since there are not methodologies for water quality with ocean water in shrimp ponds, this work proposes the use of a neural network for recognizing problems in the ponds related with the environment.

2 Methods

2.1 Data Collection

The shrimp ponds are systems where the environment must to be in equilibrium, a controlled habitat generates a lower stress level, high growing and high resistance for sickness; for example, for shrimp maturation and spawning [8], [9] almost all hatcheries require availability of oceanic-quality water on a 24-h basis. In the habitat, there exist environmental variables that make a bigger degradation of the water quality than others salinity and temperature are important water parameters impacting production of shrimp in the hatchery, and must be maintained in a narrow range, between 27 and 36 ppt salinity and 28 C (82 F) plus or minus two degrees for most Shrimps penaeids. According to information obtained from the shrimp ponds located in Sonora, Mexico, the environmental variables with a bigger impact are Temperature, PH, Salinity, Dissolved Oxygen and Turbidity [10], [11]. A bad control of this set of variables will generate chemical reactions that affect other variables. Although there are a big set of variables that affects the water quality, they no generate a bigger impact as the first, in this work the data collecting is based in those environmental variables.

Data collection was made on the farming shrimp ponds from Sonora, Mexico. Almost 400 measurements were obtained in a farming period using manual techniques for example, the DO, pH, salinity and temperature were measured with electronic sensors by an operator and turbidity was measured with laboratories instruments. For each pond, the environmental variables measured were pH, temperature, salinity, turbidity and dissolved oxygen.

The environmental variables were measured in a frequency of twice per day: at morning and afternoon and the values shows a random behavior. While the shrimp farming is in process, the parameters can suffer some variations because the weather, as an example, in the Fig. 3 the oxygen has a big variation during the day. Due these variations this parameter are always in change and the combination of them could generate a risk or good situation into the pond. The data was collected by biologists and were transferred to a computer at the end of the farming period. These variables were selected because they produce the biggest impact and they could generate conditions not appropriated for the good maturing and reproduction of the organisms. The nature of this parameter and their relationship with the water quality make them an excellent set for their processing with an ANN.

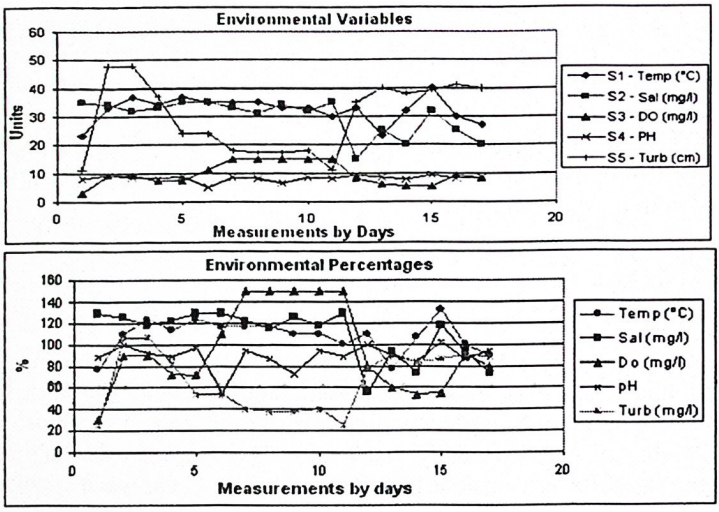


Fig. 2. Measurements and percentages in a month of the environmental variables into a shrimp pond in a farming , S1- Temperature, S2-Salinity, S3-DO, S4-pH and S5-Turbidity .

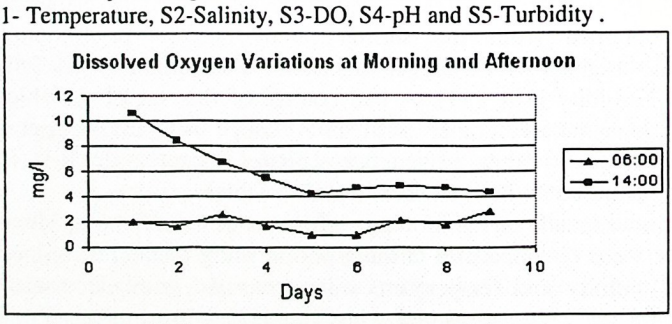


Fig. 3. The measured values of DO during a day show some variations at morning and afternoon, for example in the day one, there is a variation of 8 units, and DO with 2 mg/l generate hypoxia to the organisms.

The water quality status is obtained using a vector of the environmental variables, for example in the day one (Fig. 2) we have the vector v[Temp, Sal, Do, Turb]=[23, 35, 3, 8, 11], this vector is used as an input of the ANN for obtaining the status in the moment of the measurement. In new researches we are using a matrix of values per week as an input of the neural network.

2.2 Pre-processing

Neural networks training can be made more efficient if certain preprocessing steps are made to the data collected. The preprocessing was applied to each of the input variables (pH, temperature, dissolved oxygen, salinity) before were used for the ANN, this way allows transforming the data [12], [13]. All variables has a different range of values, for example, the temperature is given in Celsius (25°, 30°, 45°), the pH is given in a lower range for 6-9 units, and the oxygen is given from 2 to 10 ml/g. It is not convenient to use the data without a pre-processing, because the data with bigger values will have more impact in the result [7], [14]. For standardization, the data were transformed to achieve zero man and unity standard deviation following the equation (3):

$$X_1 = \frac{X - \overline{X}}{\sigma} \tag{3}$$

All variables were scaled to [0, 1] range according to:

$$\text{var} = \frac{x - \min(x)}{\max(x) - \min(x)} \tag{4}$$

Where x is the environment variable and var is the new normalized variable.

The universe of data information that involves all cases of water quality (good and risk) is extremely huge as it is shown in Table 2, but only a smaller set is measured in real time, in order to have a complete set, an artificial test patterns was made, additionally the measured patterns was added for creating a new complete set, more than 1000 patterns were used for training an ANN.

Table 2. Number of patterns used for defining only the set of a good water quality.

Variable	Temp °C	Sal (mg/l)	DO (mg/l)	PH	Turb (mg/l)
Range	23 – 30	15 – 25	6 – 10	7.6 – 8.6	35 – 45
Minimal value measured	.5	1	.1	.1	1
Total measurements per Variable	14	10	40	10	10
Total measurement per all variables	**560000**				

The patterns classification is made with an ANN to recognize and to quantify two statuses within the pool:

a) **Good**: Everything is well, the parameters are on range or some of them are out or range but it do not generate any problem, for example when temperature is one or two grades below the range.

b) **Risk**: One, some or all environmental variables are out of range, which generates problems into the pond.

2.3 Neural Network Model

Neural networks have been used as a powerful tool in the analysis of complex and uncertain data, like non-linear models. A neural network is a system of interconnected processing elements (neurons), based on the structure of the brain. The networks are constructed by nodes connected together by parameters called weights. The units in an ANN are usually accommodated in layers [7], [15].

The artificial neural network theory was applied in this study providing a non-linear relationship between input sets (Environmental variables) and output targets (water quality status). The environmental variables have dependences among them, and trying to represent them with a mathematical equation is a very hard task [6]. There are other methods for classifying patterns, in this study the classification of the status is given by a ANN, another way is the regression linear method, but it can not be applied in this work because its goal is to find parameters of the best linear approximation to the input and the desired response, and the goal of a neural network is to separate the data as well as possible into classes. Other feature of an ANN is its power to generalize, which means when a value that never was learned by the ANN, the network can give an approximation about a desired response. In this work the ANN recognize the status and quantify it, in others words, how well or risky it is, which allows finding trend of the water quality [14], [7].

In this study we use a Multilayer Perceptrons (MLPs). The number of input is given by the number of environmental variables: pH, temperature, salinity, dissolved oxygen and turbidity. The number of outputs is given by the number of status of the water quality: good and risk, and therefore the size of this layer depend on the data representation. The topology used in this work was a feed forward network; it is shown in Fig. 4.

There are many learning algorithms for the ANN; the gradient descent algorithm is very slow because it requires small learning rates for stable learning, the momentum algorithm is faster but it is still too slow for many practical applications. This study uses the Levenberg-Marquardt training algorithm because it gives a better performance for small and medium size networks with enough memory available. If memory is a problem, then there are a variety of other fast algorithms available like Resilient Backpropagation algorithm or the Scaled Conjugate Gradient [7], [16].

3 Results

The ANN was tested with different sets of measurements. Two topologies were used for this work, one of them was a 5-4-1 layered ANN, the output of this ANN responds with a 1 if water quality is good and a -1 if water quality status is bad (risk). The second network is a 5-4-2 layered ANN's, and the output is given by two neurons, whose the first represents the good status when its output is 1 and the output of the second is -1; and the second neuron represents the risk status when its output is 1 and the output of the first is -1, when one neuron has a value of 1 the other has -1 and viceversa.

The training of the MPL's was carried out with the software package MATLAB. The patterns classification given by the environmental variables confirms its nonlinear interrelation. The trained feed forward neural network represents suitably the nonlinear relations between the environmental variables. The good pattern classification of the ANN is shown in the Fig 6.

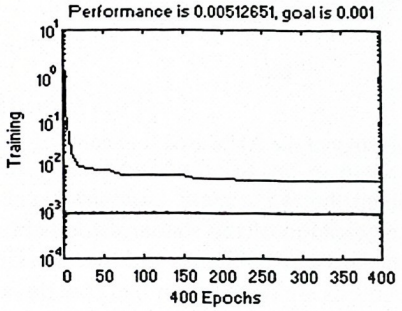


Fig. 4. Performance of the ANN (mse) and epochs obtained on the training.

There were some cautions in the Training of the ANN, when a learning rate is too large, the learning could be unstable or memorize too much and do not generalize. In other hand, a learning rate too small could not memorize. The performance of the training is showed n the Fig. 4, where the curve shows that the maximum performance obtained at 400 epochs, without an overfitting [7].

ANN's are sensitive to the number of neurons in their hidden layers. Too few neurons can lead to underfitting. Too many neurons can contribute to overfitting, in which all training points are well fit, but the fitting curve takes wild oscillations between these points. We used many topologies, training ANN with one, two and three hidden layers and with different number of neurons; however these topologies did not show a better performance, also the ANN was trained with different activations functions as tangential, logarithmic, exponential and lineal functions, finding a better performance with a tangential activation function in the first and the hidden layer, and a linear activation function in the output layer [7], [16].

The 5-4-1 ANN showed a good performance with more epochs for training but the response was not effectively, thus the 5-4-2 layered ANN showed the best performance

with low epochs obtaining the best classification pattern. One computational system was designed using this ANN. Fig. 5 presents the topology with better results for this application, and the operation of output neurons with linear activation function in the output neurons.

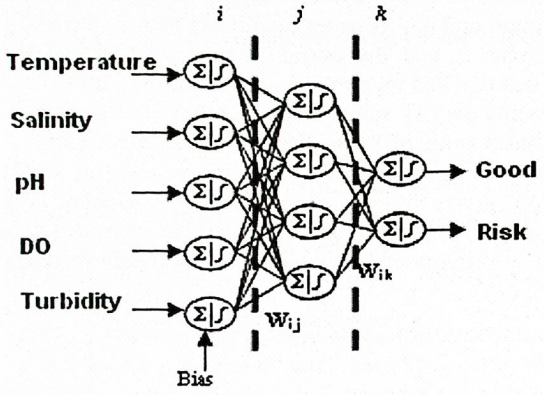


Fig. 5. Topology of the ANN used for analyzing water quality.

Some patterns are given by an expert biologist, such patterns build a set of measurements with its classification of the status of the habitat, in order to compare the classification of the ANN, the figure 6 shows the two classifications showing a similar behavior, however some patterns are not good or bad exactly, and this is because the ANN gives a response of patterns that were not included into the training set (generalize) and this result is closer to a similar patter that exist in the training set.

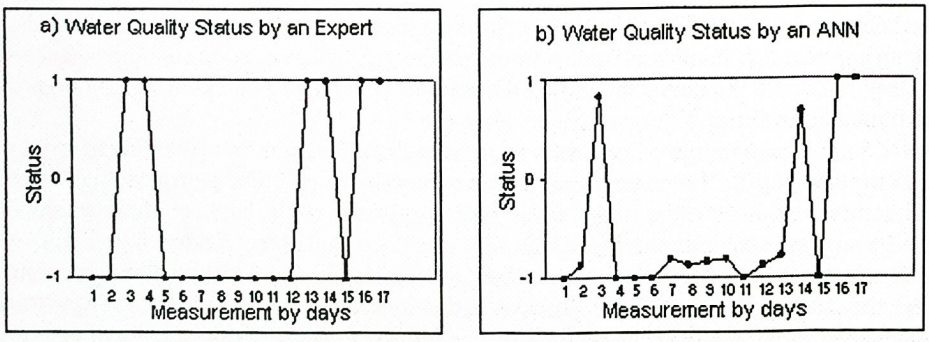


Fig. 6 a) water quality status defined by an expert, when status is 1, means good, and when is -1 means bad water quality (risk); and b) water quality status defined by the ANN., the classification gives a deteriorated status, for example when the variables are not exactly in the normal range but they do not represent a bad status the classification is not entirely 1 or -1.

3.1 Post-processing

Post-processing steps are needed for measuring the performance of the trained ANN. We used a regression analysis between the response and the targets, using the linear equation $y = mx + b$ [7], [16].

For the first neuron (good) the values obtained were: $m = 0.9548$, $b = 0.0077$, $r = 0.9771$ and, for the second neuron (risk) the values obtained were: $m = 0.9675$, $b = 0.0181$ and $r=0.9836$. Where m corresponds to the slope, b the intercept with the y axis and r is the correlation coefficient between outputs and targets. A perfect fit is obtained when the slope is 1, the y-intercept is 0 and the correlation coefficient is 1. We can see that the numbers are very close.

The validation of the ANN was made using experimental data information taken over a three-month period of farming shrimp and using some experimental ponds in laboratories, demonstrating the reliability of the network. The results demonstrated that the ANN model trends very well when all the environmental variables are measured in real time or when a set of information is analyzed.

4 Conclusions and Future Work

In this paper, we present a decision model in the form of neural network for water management in shrimp ponds located in Sonora, Mexico. A mathematical relationship between the environmental variables is too hard to develop. A big research about the shrimp habitat was made for determining what kind of environmental variables affect the water quality into the farming ponds.

An advantage for using the ANN's as classifiers is their sensibility to recognize patterns that were not used in the training set, that is called generalize. A disadvantage is when the classification depends of amount of patterns in the training set, and if this set is incomplete the ANN will not recognize a status in the right way.

According with the results shown in Fig. 6, the ANN shows a good classification of the status of water quality; also it gives a degradation of the status depending of the conditions into the pond.

From the results obtained, we can conclude that our model is effective for resolution of problems related with water quality into shrimp farming ponds; also it is a powerful tool for the analysis that are practiced all days by biologists that are searching a better way for farming shrimp.

Based on tour experimental results, the main advantage of our model over other evaluating methods are: speed on the assessment of water quality and, its portability and adaptability to others aquaculture systems.

As future work, we are researching in other way to obtaining more effectively a status of the water quality. Also we are working in a way for predicting the water quality status into the pond using ANN's and fuzzy systems during a farming period, working with the best model developed to assess the water quality in the present. The model for evaluating and

for predicting the water quality will be integrated to a computational system that will give a complete analysis for the ponds.

References

1. Casillas, R., Magallón, F., Portillo, G., Osuna, P.: Nutrient Mass Balances in Semi-intensive Shrimp Ponds from Sonora, Mexico Using Two Feeding Strategies: Trays and Mechanical Dispersal. Aquaculture, Vol. 258. Elservier (2006) 289-298.
2. Páez O, F.: Camaronicultura y Medio Ambiente. Instituto de Ciencias del mar y Limnología. UNAM. pp. 271-298, México, 2001.
3. Angulo Corrales, Adán y Angulo Corrales, Uriel, "Estudio de Calidad del Agua y su Relación con el Crecimiento del Camarón Blanco (Litopenaeus Vannamei), en la Granja Camaronera Agua Verde, S.A. de C.V. en Rosario, Sin.", Tesis, Universidad Autónoma de Sinaloa, 2003.
4. Martínez Córdova Rafael M., "Cultivo de Camarones Pendidos, *Principios y Practicas*", Ed. AGT Editor S.A., 1994.
5. Hernández, J., Zirino, A., Marione, S., Canino, R., Galindo, M.: PH-density relationships in seawater. Ciencias Marinas, Vol. 29 (2003) 597-508.
6. Ocampo, W., Ferré, N., Domingo, J., Schuhmacher, M.: Assessing water quality in rivers with fuzzy inference systems: A case study. Environment International, Vol 32. Elservier (2006) 733-742.
7. Romero, C., Shan, J.: Development of an artificial neural network-based software for prediction of power plant canal water discharge temperature. Expert Systems with Application, Vol 29. Elservier (2005) 831-838.
8. Granvil D. Treece, Shrimp Maturation And Spawning, Texas A&M University, Sea Grant College Program. UJNR Technical Report No. 28, 2001
9. Granvil D. Treece, Shrimp Maturation and Spawning. In Proceedings of the 28th US-Japan Natural Resources Aquaculture Panel. UJNR Technical Report No. 28. pp 121-134. TAMU-SG-01-814. 2001
10. Li, Y., Li, J., Wang, Q.: The effects of dissolved Oxygen Concentration and Stocking Density on Grown and Non-Specific Immunity in Chinese Shrimp, Fenneropenaeus Chinensis. Aquaculture, Vol. 256. Elservier (2006) 608-616.
11. Wahab, A., Bergheim, A., Braaten, B.: Water Quality and Partial Mass Budget in Extensive Shrimp Ponds in Bangladesh. Aquaculture, Vol. 218. Elservier (2006) 413-423.
12. Campos, M., López, D., "Neural Network Approach to Locate Motifs in Biosequences", Progress in Pattern Recognition, Image Analysis and Applications, Springer (2005), LNCS 3773, 214-221.
13. Lek, S., Guiresse, M., Giraudel, J.: Predicting stream nitrogen concentration from watershed features using neural networks. Water Resources, Vol. 33. Pregamon (1999) 3469-3478.
14. Principe, José, Luliano, Neil y Lefebvre W Curt, 2000, Neural and Adaptative systems: *Fundamentals through Simulations,* 1ª Edition, John Wiley & Sons.
15. Sharma, V., Negi, S., Rudra, R., Yang, S.: Neural network for predicting nitrate-nitrogen in drainage water. Agricultural Water Management, Vol 63. Elservier (2003) 169-183.
16. Chacón, M., Blanco, C., Panerai, R. and Evans, D., "Non linear Modelling of Dynamic Cerebral Autoregulation Using Recurrent Neural Networks", Progress in Pattern Recognition, Image Analysis and Applications, Springer (2005), LNCS 3773 205-213.

Mechanical Faults Diagnosis in Induction Motors based on Neural Networks and Vibration Patterns

Beatriz A. Jaime Fonseca[1], Luis P. Sánchez Fernández[1] and Rodrigo López Cárdenas[1]

[1]Center for Computing Research, National Polytechnic Institute
Av. Juan de Dios Batiz s/n casi esq. Miguel Othon de Mendizabal, Col. Nueva Industrial Vallejo, CP 07738. Mexico City, Mexico
betyjf08@yahoo.com.mx, lsanchez@cic.ipn.mx, rodrigo.lc@gmail.com

Abstract. The importance of this project takes root in the aptitude to recognize some types of mechanical faults across a new algorithm as neural networks are, with little margin of mistake that will allow the users of small, medium and big companies, to become free of the technician specializing in vibrations and signs, and to realize an alone investment for the constant monitoring of their engines. The patterns generated for the training and operation of the back propagation neural network, not only is able to also diagnose the type of fault but that so severe is this one, existing the possibility of predicting future behavior of the machine with the obtained results.

1 Introduction

In the industrial automatic processes, one of the electric parts of major application, they are the engines of induction of fall, medium and high tension. If the used engine was presenting a sudden fault, there might be generated serious consequences that would cause losses in the production. For the previous reason, it turns out necessary to assure the operative continuity of the engine by means of the detection of incipient faults generated by the efforts to which it is submitted, and to give it maintenance of an opportune way [1].

Anyone that is the strategy used in the maintenance is necessary to identify the variable of operation that will provide to us the readings of the machine's condition, to be able to generate a diagnosis by means of an automatic algorithm. In this article, the variable to measuring will be the mechanical vibrations produced in different points of the engine and the algorithm to analyze the vibration readings involves the neural networks.

Traditionally in industry, the maintenance has developed on the base of two fundamental strategies [2]. On one hand, the machinery; as part of an automated process, it works in a constant way and its breakdowns are attended until these appear [2], [3]. This strategy is known as Corrective Maintenance. In other types of machines the maintenance is applied in certain intervals of time that can be established on the basis

© L. Sánchez, O. Pogrebnyak and E. Rubio (Eds.)
Industrial Informatics
Research in Computing Science 31, 2007, pp. 243-252

of the recommendations of the manufacturer or the frequency of failures, which is known as strategy of Preventive Planned Maintenance [4].

Nevertheless, from the 40's, the world has come implementing and perfecting the predictive technologies, whose conceptual base takes root in the intervention of the machinery when it is justified by the presence of objective symptoms of the deterioration of its mechanical condition [1], [5], [6].

The PPM (Program of Predictive Maintenance), it is that one who contemplates in an effective way three indispensable stages: Detection, Identification and Correction [6]. In general, the PPM helps at first, to incipient faults detect simultaneously that allows to have the necessary tools to analyze the reason of the problem that is developing, being achieved to determine finally, the opportune moment to correct effectively detected problem . Across the years, the operators of rotating machines in the different industrial sectors have used technologies of auditory checking to verify if the behavior of their engine is normal or not [7]. Traditionally and probably in unconscious form, the vibrations have been used as an indicator of the technical condition of the engines, being able by means of the measurement and analysis of these, detect and identify different types of faults. Besides, the frequencies and magnitudes of the signs of vibration of the engine, they will allow identifying what this one badly, and so severe what is the problem, to take opportune action before the damage produces irreversible losses of production.

2 Mechanical Vibration

The forces that cause the vibration are usually generated on the parts of the engine that they are in movement, or in which ones this movement is transmitted. Due to the fact that these forces change direction or magnitude in agreement to the rotational speed (RPM) of the machine, it is possible to deduce that the majority of the problems of vibration will have frequencies directly related to its speeds [1].

The frequency response of vibration has been so useful, that could have rejected many faults so much electrical as mechanics of the engine, the Fig. 1 shows the frequencies of the characteristics harmonics signals of different kind of mechanical faults [1], [2], [3], [8].

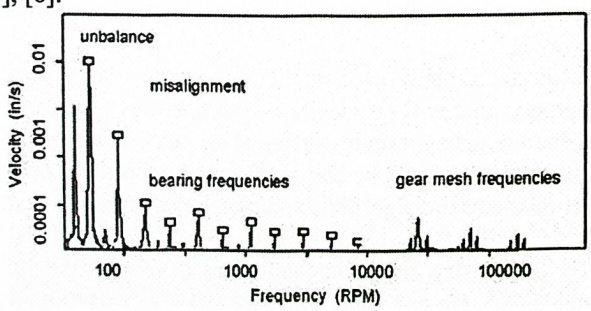


Fig. 1. Frequency response of mechanical faults in induction motors. This shows the characteristic harmonics of different kind of faults in electrical engine.

The diagnosis of the type of present fault in the engine is realized searching harmonic in frequencies multiple of its speed. This way, the degree of severity of the fault is evaluated by the magnitude of the harmonic one and the comparison by international procedure of monitoring of electrical machines [8].

3 International Standards of Mechanical Condition

International organizations as the American National Standards Institute (ANSI), Engineers' German Association VDI 2056 or International Standards Organization ISO 10816-1 1995 (replaces to the ISO norm 2372 and ISO 3945), and the BS 4675, they all stipulate the positions of measurement as well as the maximum permissible limits for class of equipment. These indicators contemplate the measurement of the total level of RMS speed inside a specific range of frequencies. The vibrations considered by these standards to evaluate the severity of the vibrations, are those measures in the surface of the machines, on its rests or in the points of assembly, in the range of frequencies of 10 Hz to 1000 Hz [9], [10].

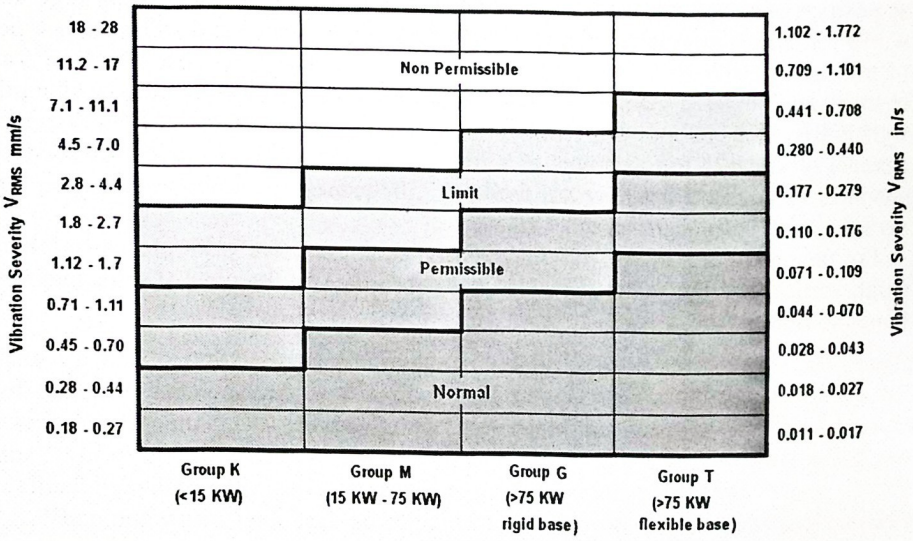


Fig. 2. Standards levels of mechanical vibrations. The figure shows normal, permissible, limits and non permissible levels of mechanical vibration in different kind of engine. Motors of the group K has power level's lower than 15 000 Watts, group M greater than 15 000 Watts and lower than 75 000 Watts and group G and T for power level's greater than 75 000 Watts with rigid and flexible bases respectively.

It thinks that a significant change takes place in the vibratory response, when this one changes in the reason 1:1.6 meaning that the vibratory level changes in 60%. Increases of the levels of vibration in 2.5 times indicate a change of condition Me-

chanics of the engine. Increases of the levels of vibration in 10 times constitute an alarming change.

4 Data Acquisition

The piezoelectric accelerometer can be fixed to the surface where it is desired to carry out the measurement with the aid of different elements such as: the steel bolt, the wax of bee, the permanent magnet, glues and the sharp shooting well-known. The sensor used for the construction of the System of Predictive Maintenance of Motors of Induction of this project, were the MA11 of the Honeywell mark, with one frequency of operation located in the rank of 2Hz - 10kHz. The MA11 is a piezoelectric accelerometer with stainless cover and system of magnetic fixation and a Sensitivity of 100 mV/g.

For the acquisition of the mechanical measures of vibrations in the induction motor, the tool of graphical programming LabVIEW 7 Express was used, with the use of card PCI - 4452 of National Instruments. This card specifically is designed for the acquisition of signals of sound and vibrations. It counts on Applications package that works for signals associated with accelerometers, microphones and some other types of transducers with an ample dynamic range. Card PCI – 4452 also count on four channels of integrated analogical entrances in the accessory of acquisition of dynamic signals BNC – 2140 and whit the frequency of sampling equals to 204.8 kS/s, each channel of entrance can sampling to a frequency of 51.2 kS/s in case that the four channels of simultaneous way are used. This frequency is sufficient since he was necessary that outside greater to 10 kHz.

For the acquisition of the vibration signals, accelerometers were placed on guard axial, vertical and horizontal [11], [12], [13] according to the type of fault that was hoped to find, as it is in Fig 3.

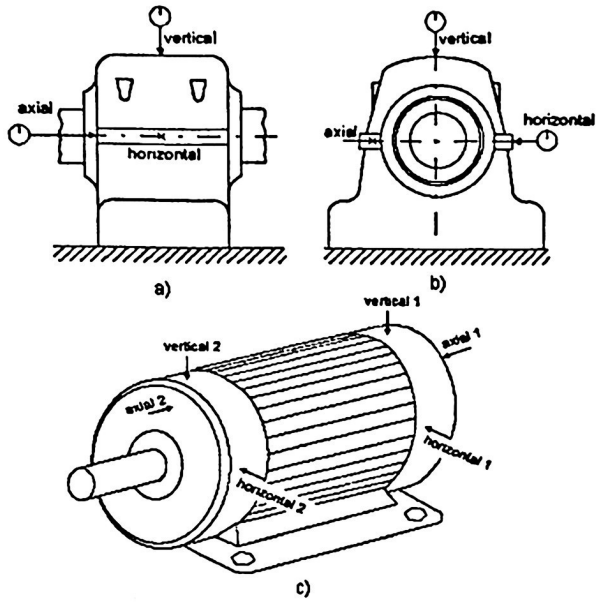


Fig. 3. Location of the piezoelectric accelerometers. The accelerometers must be placed simultaneously in the positions that the figure shows, of simultaneous way or one depending on the type of fault

5 Fault Recognition Using Neural Networks

The artificial neural network that was used in this project is based on the architecture of back propagation. This is a learning rule that is used for the training of the network, has like main characteristics, more able containing one or intermediate between the layer of entrance and the one of exit, and its operation consists of a learning predefined by a data set of entrance related to its corresponding exit [14].

When one works with back propagation neural networks, first a pattern of entrance like excitation is applied to the network, the data propagate layer after layer, and a percentage of calculated error is returned, to readjust the training patterns, in order that the difference between the wished values "targets" and the values of exit, are minimum. This procedure is made in cycles known like epochs [14].

Although the network is only trained to recognize four types of mechanical faults, the patrons of entrance to undercoat were planned to detect any type of well-known mechanical fault at the present time.

5.1 Generation of Patterns

In order to train to the neural network 50 signals in LabVIEW of each one of the following mechanical faults were simulated: unbalance, bent shaft, angular misalignment and parallel misalignment; with Normal, Permissible, in the Limit and Non Permissible level vibration.

Table 1. Mechanical Faults. It shows the characteristic harmonics for the different types of faults.

Frequency (RPM)	FAULT
< 0.5 x RPM	Oil whirl instability.
1x RPM	Unbalance.
1x RPM	Weakness of machine feet.
1x , 2x RPM	Bent Shaft.
0.5x, 1x, 2x, 3x RPM	Mechanical Looseness.
1x, 2x, 3x RPM	Misalignment.
1x, 2x, 3x, 4x RPM	Mechanical Looseness.
0.5x, 1x, 1.5x, 2x, 3x, 4x, 5x, 6x, 7x, 8x, 9x, 10x	Mechanical Looseness.

Reason why altogether, 12 overtones looked for: 0.5x, 1x, 1.5x, 2x, 3x, 4x, 5x, 6x, 7x, 8x, 9x, and 10x RPM. Each significant overtone of a fault is observable in a bandwidth of approximately 10 Hertz, reason why the vectors of training of the neural network were constructed as it is in Fig 4. What means that, the vector of the vibration signal is of length 132, independently of the frequency or engine's speed.

The samples for unbalance, for example, generate an overtone in frequency 16 of the training adjustment, in amplitude that vary of 0,28-28 mm/s equivalent to the amplitude of the speed.

The neural network that was used contains 132 neurons of entrance for each one of the elements of the generated patterns, two intermediate ones, one that the entrance transmits and a second that classifies the twelve harmonics characteristic of the faults, and finally layer of exit of seven neurons one for each one of the four faults detected by the three different levels from severity (Fig. 5).

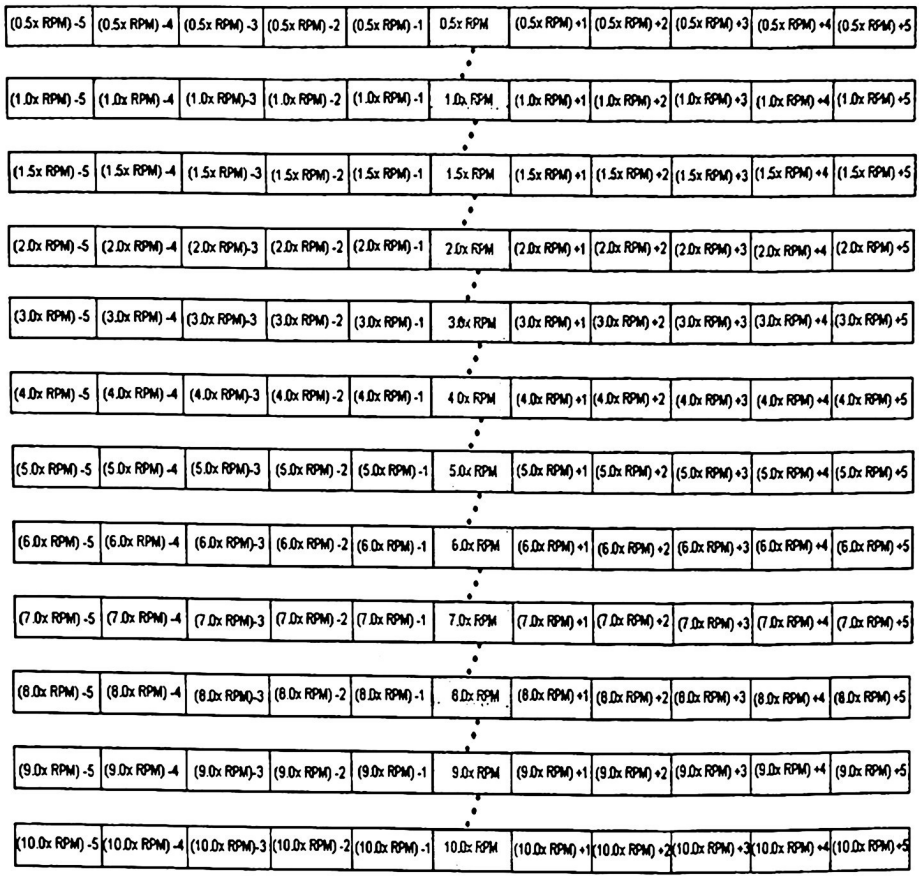


Fig. 4. Patterns of entrance to the neural network. Figure shows the 132 significant readings to identify all kind of mechanical fault.

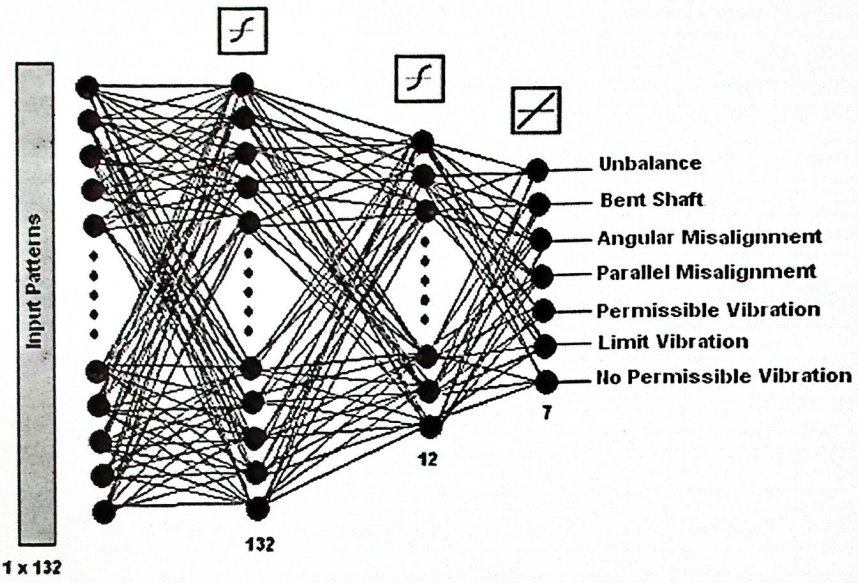


Fig. 5. Neural network topology. The neural network has the input layer with 132 neurons, 2 hidden layers with 132 and 12 neurons and the output layer with 7 neurons, 4 for fault's kind and 3 for severity level.

6 Results

The training of the described neural network was reached at 5682 epochs with algorithms established in MATLAB under the following code:

```
p=[PP1 PP2 PP3 PP4];
t=[tt2];
net=newff(minmax(p),[132,12,7],{'tansig','tansig','pure
lin'},'trainrp');
net=init(net);
net.trainParam.show = 10;
net.trainParam.goal = 1.4e-4;
net.trainParam.epochs = 6000;
net.trainParam.delt_inc=1.2;
net.trainParam.minstep =1.0000e-4;
net.trainParam.min_grad=1.0000e-18;
[net,tr]=train(net,p,t);
```

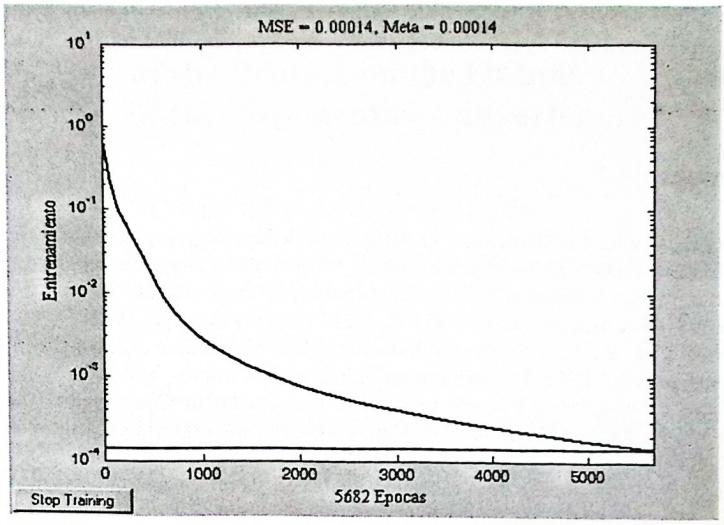


Fig. 6. Training of the neural network. Number of epochs x-axis against error in the y-axis.

The results of only some of tested with the neural network are in Table 2.

Table 2. Neural Networks Results. The most of the experiments does almost 100% recongnition.

Pattern	% Recognized
Permissible Unbalance	99.3966
No permissible Parallel Misalignment	99.6066
Permissible Bent Shaft	99.7288
No Permissible Angular Misalignment	99.3653
Permissible Parallel Misalignment	99.1234
Limits Angular Misalignment	97.8126

Conclusions

Not only incipience's faults of a motor are detected of induction by means of the implementation of neural networks, but also it is possible to emit objective diagnoses, with low levels of error, on the existence of more severe faults.

At the present time, there are systems of no detection of mechanical faults based on algorithms of artificial intelligence, so this system can be implemented with facility and good results, reflecting its utility in economic gains and improvements in the production levels, as long as opportune action of maintenance is taken, when some fault is detected. This system is limited the detection of mechanical faults in induction

motors, so that although it is possible to detect some electrical faults by means of he himself procedure, has been verified that exist greater success in the use of other techniques.

References

1. Palomino, Evelio: La Medición y el Análisis de Vibraciones en el Diagnóstico de Máquinas Rotatorias. División de Ingeniería de las Vibraciones y Diagnóstico, Cuba
2. Cameron, J. R.: Vibration and Current Monitoring for Detecting Airgap Excentricity in Large Induction Motors, *IEEE,* Vol. 133, Pt. B, Num. 3, Mayo de 1986.
3. Benbouzid, M. E. H.: A Review of Induction Motors Signature Analysis as a Medium for Faults Detection", IEEE Transactions on Industrial Electronics, Vol. 47, No. 5 (2000)
4. Saavedra, P.: Análisis de Vibraciones nivel II. Universidad de Concepción. Chile (1997).
5. Saavedra, P.: Evaluación de la Severidad Vibratoria. Laboratorio de Vibraciones Mecánicas. Chile (2002).
6. Jiménez, O.: Casos Reales de Análisis de Vibración. 1er. Congreso Mexicano de Confiabilidad y Mantenimiento. México (2003).
7. Ramírez Cruz, J.M., Carvajal Martínez, F.A., Campos Hernández, M.: Diagnóstico en Sitio de Motores de Inducción Mediante el Análisis del Espectro en Frecuencia de las Corrientes de Fase, Octavo Congreso Nacional de Máquinas Eléctricas Rotatorias, AMIME
8. Berry, James E.: Vibration Signature Analysis I, (1993).
9. International Standards Organization, ISO Standard 2372: Mechanical Vibrations Of Machines With Operating Speeds From 10 to 200 Rev/Sec - Basis For Specifying Evaluation Standards (1974).
10. International Standards Organization, ISO Standard 2954: Requirements for Instruments for Measuring Machinery Vibration (1974).
11. Forland, Clair: Why Phase Information is Important for Diagnosing Machinery Problems. Orbit. Bently Nevada Corporation (1999) 29-31.
12. Saavedra, P.: La Medición y Análisis de las Vibraciones como Técnica de Inspección de Equipos y Componentes, Aplicaciones, Normas y Certificación. Universidad de Concepción. Chile (1997).
13. Carvajal, F., Ramírez, J., Arcos L.: Diagnóstico en Línea y Fuera de Línea de Motores de Inducción de Baja, Mediana y Alta Tensión. Boletín IIE. Marzo (1999). 90 – 96.
14. Chow, M.: Methodologies of Using Neural Network and Fuzzy Logic Technologies for Motor Incipient Fault Detection. World Scientific.
15. Príncipe, José C., Euliano, Neil R.: Neural and Adaptive Systems: Fundamental through Simulations. John Wiley and Sons, Inc., New York.
16. Su, H., Chong, K.T. Condition monitoring for electrical failures in induction machine using neural network modeling of vibration signal, Proceedings of the 2005 IEEE International Conference on Computational Intelligence for Measurement Systems and Applications, CIMSA 2005, art. no. 1522851. (2005) pp. 156-161
17. Han, T., Yang, B., Yin, Z. Feature-based fault diagnosis system of induction motors using vibration signal, Journal of Quality in Maintenance Engineering 13 (2). (2007) pp. 163-175

Two Reconstruction Algorithms of the Process on the Output of the Exponential Converter

Vladimir Kazakov[1], Mónica Sedeño[2]

[1] Department of Telecommunications, ESIME Zac-IPN, Mexico D.F., Mexico
{Vladimir Kazakov}, vkazakov41@hotmail.com,
{Mónica Sedeño}, msedeno@ipn.mx

Abstract. The two extrapolation algorithms of the Sampling – Reconstruction Procedure of the non Gaussian random process at the output of the exponential converter are considered. The input process is the call the Gaussian process. The conditional mean rule is used for the investigation. The principal SRP characteristics (the reconstruction function and the error reconstruction function) are obtained. The comparison of the optimal and non optimal algorithms is given in this investigation.

Keywords: Conditional Mean, Conditional Variance, Covariance Function, Extrapolation.

1 Introduction

The classical Sampling Theorem associated with the names of Whittaker, Kotelnikov and Shannon has been proved for deterministic functions.

The Balakrishnan's theorem [1] has been proved for stationary random processes with limited spectrum. This theorem has some important disadvantages [2]:

i) The principal characteristic - probability density function (pdf) - is not taken into account.
ii) The Reconstruction Function is considered as a linear function with the samples for any kind of stochastic process.
iii) The Basic Function sin x /x is the same for any kind of stochastic process.
iv) The error reconstruction function is cero for any kind of stochastic processes.
v) The number of samples is equal to infinite.
vi) It is impossible to describe the optimal extrapolation procedure for arbitrary stochastic processes.

In fact, it is very important to consider the Sampling- Reconstruction Procedure (SRP) for stochastic processes with a finite number of samples. We want to obtain the two principal characteristics of SRP: 1) the reconstruction function and 2) the error

© L. Sánchez, O. Pogrebnyak and E. Rubio (Eds.)
Industrial Informatics
Research in Computing Science 31, 2007, pp. 253-260

reconstruction function for some arbitrary stochastic processes. This problem can be resolved by the application of the famous statistical rule of random variables - the Conditional Mean Rule [2].

If we know an arbitrary number of samples, it is possible to say that the Conditional Mean Function is the Reconstruction Function and the Conditional Variance is the Error Reconstruction Function. On the basis of this approach the SRP of different types of stochastic processes have been investigated [3 - 8].

The present paper is devoted to the investigation of the SRP of the non Gausssian process at the output of the exponential converter, driven by the Gaussian process. We analyze optimal and non optimal SRP algorithms of the mentioned process and make the comparison of these two algorithms.

The optimal SRP algorithm is based on the knowledge of the exact expression of pdf of the process under consideration. The non optimal SRP algorithm is characterized by the covariance function of this process. In both cases we use the extrapolation case of the reconstruction.

2 Gaussian Case

Let us consider the general case of the stationary Gaussian process $x(t)$ with the mathematical expectation m, the variance σ^2, and the covariance function $K(t_i\text{-}t_j)$. On the basis of the covariance function one can write the covariance matrix:

$$\left|K(t_i-t_j)\right|=\begin{vmatrix} K(t_1-t_1) & K(t_1-t_2) & \cdots & K(t_1-t_m) \\ \cdots & \cdots & \cdots & \cdots \\ \cdots & \cdots & \cdots & \cdots \\ K(t_m-t_1) & K(t_m-t_2) & \cdots & K(t_m-t_m) \end{vmatrix}. \tag{1}$$

These characteristics are complete because they determine the well known analytical expression of the multidimensional Gaussian pdf. There is the expression for the conditional Gaussian multidimensional pdf. On the basis of this conditional pdf one can determine the expressions for the conditional mathematical expectation $\tilde{m}(t)$ and for the conditional variance $\tilde{\sigma}^2(t)$. We fix an arbitrary set of N samples $X, T=\{x(T_1), x(T_2), ..., x(T_N)\}$ where $x(T_i)(i=1,...,N)$ the value of the sample is at the time moment T_i. The expressions for $\tilde{m}(t)$ and $\tilde{\sigma}^2(t)$ have the view:

$$\tilde{m}(t)=m+\sum_{i=1}^{N}\sum_{j=1}^{N}K(t-T_i)a_{ij}\left[x(T_j)-m(T_j)\right]. \tag{2}$$

$$\tilde{\sigma}^2(t) = \sigma^2 - \sum_{i=1}^{N}\sum_{j=1}^{N} K(t-T_i)a_{ij}K(T_j-t). \tag{3}$$

where a_{ij} is the element of the inverse covariance matrix (1) with the change of the designation - T_i, T_j instead t_i, t_j:

$$\left|a_{ij}\right| = \left|K(t_i - t_j)\right|^{-1}, \quad \sum_{j=1}^{m} a_{ij}K(t_j - t_i) = 1. \tag{4}$$

Let us consider the simple case $m=0$ and $\sigma^2=1$, then instead of (2) we have:

$$\tilde{m}(t) = \sum_{i=1}^{N}\sum_{j=1}^{N} K(t-T_i)a_{ij}x(T_j) = \sum_{j=1}^{N} x(T_j)b_j(t) . \tag{5}$$

where the basic function $b_j(t)$ is determined by the following expression:

$$b_j(t) = \sum_{i=1}^{N} K(t-T_i)a_{ij} . \tag{6}$$

From (5) and (6) one can see that the reconstruction function $\tilde{m}(t)$ is the linear function of samples and the basic function is not the function $Sin x/x$. The optimal basic function $b_j(t)$ depends on the number of the sample j, it depends on the quantity of samples N, and it depends on the different covariances between the sections of the process in T_i and T_j moments (the elements of the inverse covariance function a_{ij} includes the corresponding covariance moments $K(T_i-T_j)$, the covariance $K(t-T_i)$) between the current section in the time t and the section in the time of the samples T_i.

The formula (3) shows that the error reconstruction function does not depend on the values of samples, but it depends on the time moments of sample T_i.

One example of the easiest case is the Markov process, which is generated at the output of an integrator circuit RC with the parameter $\alpha=1/RC$. And its own covariance function is expressed with the exponential form:

$$K(\tau) = \sigma^2 \exp(-\alpha|\tau|). \tag{7}$$

The covariance time $\tau_c = 1/\alpha$ will be chosen in this case as $\tau_c = 1$. The interpolation process depends on the samples which are closer $x(T_i)$ and $x(T_{i-1})$ $(T_i < t < T_{i-1})$ and the extrapolation process is described by the latest sample $x(T_N)$, $(t>T_N)$. These characteristics are determined by the Markov properties.

3 The output process of the exponential converter

Let us consider a non linear exponential converter with the characteristic:

$$\eta(t) = a_0 \, e^{\beta \xi(t)} \ . \tag{8}$$

where a_0 and β are constants, $\xi(t)$ is an input process, $\eta(t)$ is an output process.

We suppose that the process $\xi(t)$ is Gaussian Markov process with the covariance function (7) and we put $\sigma^2 = 1$.

Firstly, we determine the unconditional characteristics of the process $\eta(t)$. Using the known methodology:

$$F(\eta) = f(h(\eta)) \frac{1}{\left| \frac{d\eta}{d\xi} \right|} \ . \tag{9}$$

where $h(\eta)$ is an inverse function with respect to (8).

Then, we can obtain the inverse unconditional characteristics of the process $\xi(t)$

$$\xi(t) = \frac{1}{\beta} \ln \left(\frac{\eta(t)}{a_0} \right). \tag{10}$$

After that, we derivate the expression (8) with respect of $\xi(t)$

$$\left| \frac{d\eta}{d\xi} \right| = a_0 \beta \exp\left(\beta \xi(t) \right). \tag{11}$$

We get an one-dimensional pdf $F(\eta)$.

$$F(\eta) = \frac{1}{a_0 \beta \sqrt{2\pi}} \exp\left[-\frac{1}{2} \left(\frac{1}{\beta} \ln \frac{\eta}{a_0} \right)^2 - \ln \frac{\eta}{a_0} \right]. \tag{12}$$

The graph of this function is presented in Fig. 1.

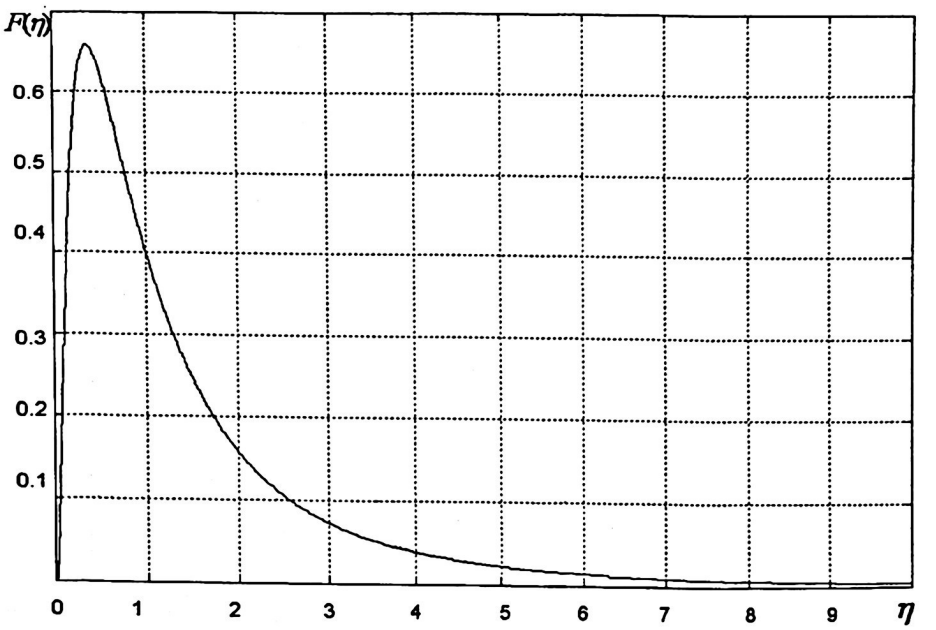


Fig. 1. Pdf of the output process with the parameters $a_0=1$, $\beta=1$

On the basis (10) we can determine the mathematical expectation m_η

$$m_\eta = m_1^{\eta} = a_0 \exp\left(\frac{\beta^2}{2}\right). \tag{13}$$

and the variance

$$\sigma_\eta^2 = a_0^2\left[\left(\exp 2\beta^2\right) - \left(\exp \beta^2\right)\right]. \tag{14}$$

Using (11) and (12) for $\beta = a_0 = 1$ we obtain the concrete values of the mean and the variance: m_η=1.6487, σ_η^2= 4.6708.

Writing (8) for two time moments t and $t + \tau$, multiplying both expressions and applying the average operation, one can find the covariance function of de output process $\eta(t)$:

$$K_\eta(\tau) = a_0^2 e^{2\alpha a_0} \exp[\alpha^2 \sigma_\xi^2 (1 + K_\xi(\tau))] \ . \tag{15}$$

The algorithm of extrapolation only depends on the sample $\eta(T_N)$ because $\eta(t)$ is a Markov process.

4 The optimal reconstruction

Let us consider the optimal extrapolation procedure. Using the method [6] on the basis (10) – (13) we obtain the required expressions:

$$\tilde{m}_1^\eta(t) = a_0 \exp\left\{\beta \tilde{m}_1^\xi(t) + \frac{1}{2}\beta^2\right\} \ . \tag{16}$$

$$\tilde{m}_2^\eta(t) = a_0^2 \exp\left\{2\beta \tilde{m}_1^\xi(t) + \beta^2 + \beta^2 \tilde{\sigma}_\xi^2(t)\right\} \ . \tag{17}$$

Here the first conditional moment function $\tilde{m}_1^\xi(t)$ is determined by the formula (5) with $N=1$:

$$\tilde{m}_1^\xi(t) = \xi(T_N) \exp(-\alpha(t - T_N)) \ . \tag{18}$$

The value of the input sample $\xi(T_N)$ is connected with the value of the output sample $\eta(T_N)$ by the inverse expression

$$\xi(T_N) = \frac{1}{\beta} \ln \frac{\eta(T_N)}{a_0} \ . \tag{19}$$

We obtain the conditional variance $\tilde{\sigma}_\eta^2(t)$ by the formula

$$\tilde{\sigma}_\eta^2(t) = \tilde{m}_2^\eta(t) - [m_1^\eta(t)]^2 \ . \tag{20}$$

where the right part is determined by the expressions (14) – (17).

We can see that the conditional variance $\tilde{\sigma}_\eta^2(t)$ depends on the values of samples in the contrast with the Gaussian case.

The results of the calculations of the error reconstruction function are presented in Fig. 2. Here there is the family of the curves $\tilde{\sigma}_\eta^2(t)$ for some different values of

output samples $\eta(T_N)$ = 0.25, 1, 2, 3, 4, 5. As one can see any chosen sample has its own special curve. All the curves converge at the value of the variance, which is $\sigma_\eta^2 = 4.67$. These curves characterize the minimum error reconstruction function for the optimal reconstruction algorithm of the extrapolation type.

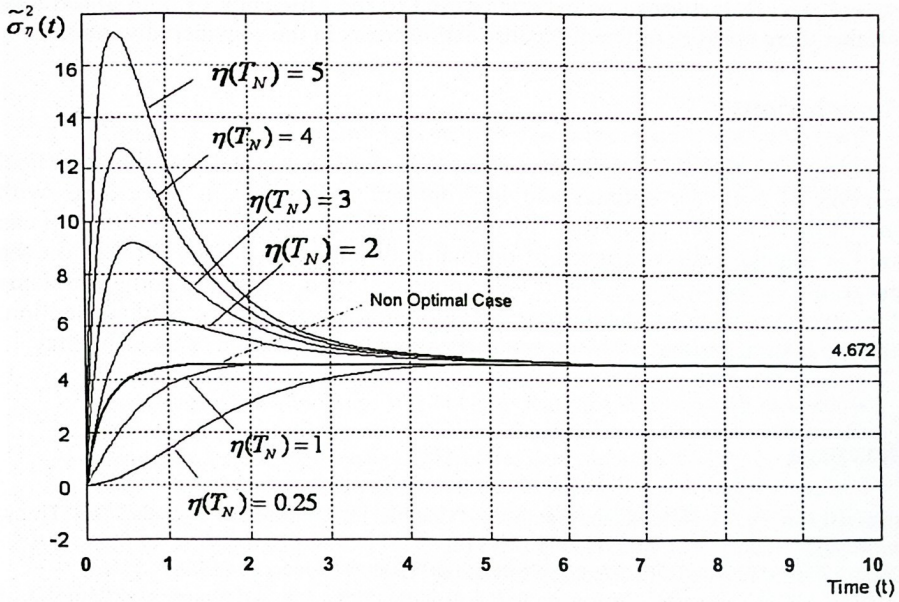


Fig. 2. The graphs of the Reconstruction Error Function. Comparison between the Optimal Case and the Non-optimal Case.

5 The non Optimal Case

Let us consider another extrapolation algorithm of the reconstruction. This algorithm is based on the knowledge of the covariance output function (13) only. It means that we must use the Gaussian approximation for the SRP description automatically, i.e. it is necessary to apply the formulas (2) and (3) with $N = 1$. Then we have:

$$\tilde{m}_\eta(t) = m_\eta + K_\eta(t - T_N)\sigma_\eta^{-2}\left[\eta(T_N) - m_\eta\right]. \tag{21}$$

$$\tilde{\sigma}^2{}_\eta(t) = \sigma_\eta{}^2\left[1 - \sigma_\eta^{-2}K_\eta^2(t - T_N)\right]. \tag{22}$$

Substituting (11) – (13) we obtain the reconstruction function $\tilde{m}_{\eta}(t)$ and the error reconstruction function $\tilde{\sigma}_{\eta}^{2}(t)$. The graph of the function $\tilde{\sigma}_{\eta}^{2}(t)$ is shown in Fig. 2 for the parameters $\beta = a_0 = 1$ (see the bold curve). The asymptote of this curve is equal to $\sigma_{\eta}^{2} = 4.6708$ as well. We can see that the application of the non optimal Gaussian Algorithm give us acceptable results for the samples which are in the interval $[1 < \eta(T_N) < 2]$. It means that the comparison of the efficiency of both algorithms shows that there are very different reconstruction errors in the considered variants.

6 Conclusions

We demonstrate that the conditional mean rule can be applied for the statistical description of both the optimal and non optimal reconstruction algorithms with respect of non Gaussian processes. The results of this investigation show that we can obtain the principal characteristics of optimal and non optimal SRP process on the output of the converter non lineal of the exponential type. There are some problems for the next investigations: the average operation of the error reconstruction function, the interpolation algorithm of the reconstruction and the evaluation of its efficiency.

References

1. Balakrishnan, A.: A Note on the Sampling Principle for Continuous Signals, IRE Trans, Information Theory, Vol. IT-3. (1957) 143-146.
2. Cramer, H.,: Mathematical Methods of Statistics, Prinston University (1946)
3. Kazakov, V.: 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 System (2005)
4. Kazakov, V., Rodriguez, D.: Study of the Sampling-Reconstruction Procedure of the Gaussian Processes with a Number of Limited Samples: Master Degree Thesis IPN, Mexico (2003)
5. Kazakov, V., Sánchez, S.: Sampling and Reconstruction Procedure of Gaussian Processes with an Arbitrary Number of samples and Uniform Jitter, Electromagnetic Waves and Electronic System, Vol. 9. (2004) 108-114
6. Stratonovich R.L.: Topics in the theory of random noise. Vol.1. Gordon and Breach, N.Y. (1963).
7. Lacaze B.:Periodic Bi-sampling of Stationary Processes. Signal Processing (1998).
8. Marvasti F.:Non Uniform Sampling Theory and Practice Information Technology Processing and Storage (2001) 361-391

Robust Telecontrol with Output Feedback

Luis A. Gonzalez, Jose L. Rodriguez
Centro de Investigaciones y Desarrollo de Tecnologia Digital
Instituto Politecnico Nacional
Tijuana, Mexico
lgonzal@citedi.mx

Abstract. Close loop control systems with time delay on either the feedforward or feedback channels have been under investigation for a long time. Teleoperated close loop systems belong to this class of systems. Different from the classical teleoperated structure where the close loop is located close to the plant, and the communication channels are only either used for sending command signals and/or monitoring output responses, the teleoperated close loop system has the controller located in the command center and the communication channels form part of the close loop system..The new technologies on wireless communication used with sensors and actuators, are making possible the construction of .less expensive wireless teleoperated close loop systems in the industry. It is obvious that time delay on both channels is the must important parameter that limit the stability and performance of the teleoperated close loop system. Therefore, one important issue is to design a robust close loop teleoperated system that kept stability and performance in a wide range of time delay. In this work an Hoo optimization technique was applied in the design of a close loop telecontroled system with uncertaint time delay. The results show that the designed robust controlled system under the Hoo technique allows a significant extention of performans and stability for a wider range of time delay.

Keywords: Robust Telecontrol, Time Delay, Hoo Design

1 Introduction

Telerobotics is a technology for operating and monitoring movil or fixed robots from a distance. We can found multiple examples of applications as in nuclear plants, and medicine. Design of a teleoperated system requires tracking of reference commands, regulation with respect to output disturbances, immunity to sensor and communication channels noise, and performance and stability robustness to parameter variations. All these features can only be satisfied by use of feedback. There exist two feedback telecontrol schemes that satisfy the above requirements, the one that has the feedback control loop in the same site as the plant, and the scheme where the controller is at the command center with the direct and the feedback communication channels as part of the close loop. The former scheme is the must applied, but there exists applications where the second one is the only option, as in bilateral telerobotics, [1], [2] or because of dangerous environments that limit the use of electronic

© L. Sánchez, O. Pogrebnyak and E. Rubio (Eds.)
Industrial Informatics
Research in Computing Science 31, 2007, pp. 261-269

hardware. Another powerful reason is the advent of the new inexpensive wireless telecommunication technology use with sensors and actuators making it possible to built this type of feedback loops in industrial environments. Therefore, study and design of telecontrol feedback loops has become indispensable. A main problem present in the second scheme is the presence of time delay in the control loop that may seriously jeopardize the system performance and stability. There exist several methods and techniques that can extend, the performance and stability of the telecontroled system, to ample values of the time delay. The robust Hoo optimization [3], [4], has proved to be a good method to improve the system performance when there exist model and parameter uncertainty. Here, by taking the time delay as an uncertaint parameter, we apply such method to the telecontrol of a three degree of freedom robot manipulator. The controlled system shows robustness to ample changes on time delay in the communication channels as compared with other simpler schemes, besides it was able also to attenuate the noise present in the sensors and induced in the communication channels. The paper is divided as follows: Section II is dedicated to the description of the controlled system, and the modeling of the time delay taken as a multiplicative uncertainty. Then Section III, briefly presents the Hoo design problem. Section IV shows how the weighting matrices are chosen for this design, and the results obtained. Finally, in Section V some conclusions are given.

2 Multivariable modeling of the time delay

The structure of the telecontrol system considered for this case is shown in Fig.1.

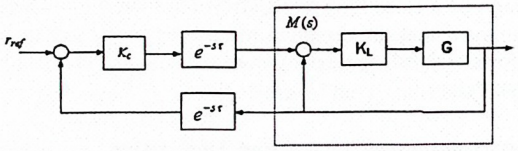


Fig.1.- Structure of the telecontroled system.

The robot was taken as three by three multivariable system G(s) to which a robust local regulator KL(s) has previously been designed and applied to form the regulated system M(s). The output vector is composed of the three angular positions of the rotational arms of the robot. Each of the two multivariable communication channels were modeled as a diagonal matrix with time delay $e^{-s\tau}$ as entries. The six time delays were taken as unknown, independent, but bounded, that is $\tau_i<\tau_{ci}$ for i = 1,2,...,6. In order to facilitate the analysis and design the six uncertaint time delay were taken as a single value τ bounded by τ_c where $\tau_c = \max_i(\tau_{ci})$. Kc(jw) is the multivariable controller for design. Fig.2 shows an equivalent system more appropriate for analysis and design where the direct and feedback time delays are put in a single expression.

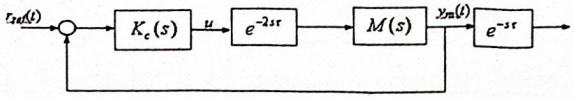


Fig.2. Equivalent system

From an analysis of the above system the loop gain L(s) is obtained as:

$$L(s) = e^{-2s\tau} M(s) K_c(s)$$

and the output sensitivity matrix function So(s) that relates the vector measured output with the disturbances at the output of the plant is given by:

$$S_o(s) = \left(I + e^{-2s\tau} M(s) K_c(s)\right)^{-1} \tag{1}$$

It is clear from the above expression how the time delay enters and affects the performance and stability of the close loop system by changing the corresponding characteristic equation. Another useful expression for design is the complementary sensitivity matrix function T(s) = I - So(s) given as:

$$T(s) = \left(I + e^{-2s\tau} M(s) K_c(s)\right)^{-1} e^{-2s\tau} M(s) K_c(s) \tag{2}$$

and, from Fig. 2 the achieved output of the system to a reference command is given as:

$$e^{-s\tau} y_m(s) = T(s) r_{ref}(s) \tag{3}$$

indicating that the actual output will show a τ. secs. direct delay with respect to a command signal and not 2τ. secs., as it could be expected.

Generally speacking, random delays are present in every communication channel with magnitudes depending on the characteristics of the channel. Stochastic analysis has been the general tool applied to analyse system with random delays, here we use the Hoo optimization method based on the "worst case", policy in order to design a controller that assert stability and performance in a wide range of time delay values. While the results may be conservative, this can be seen as an advantage if seen from the security point of view. The model uncertainty of an actual plant P(s) may be represented by normalising the plant with respect to a nominal plant Pn(s), as:

$$P(s) = P_n(s)(I + \Delta(s)) \tag{4}$$

where $\Delta(s)$ is a non-structured multiplicative uncertainty such that $\|\Delta(s)\|_\infty < \beta$. In a block diagram the above equation is represented as in Fig.3

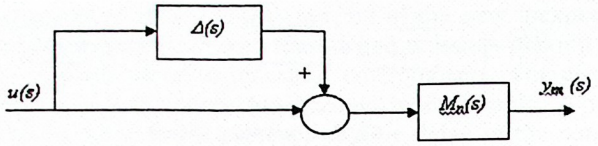


Fig.3. Multiplicative uncertainty modeling.

Now, back to the equivalent system of Fig. 2, if the actual plant P(s) is taken as $e^{-2s\tau}M(s)$ and M(s) is taken as the nominal plant, we will obtain the following identity:

$$e^{-2s\tau}M(s) = M(s)(I + \Delta(s)) \tag{5}$$

From Eq.(5) the following non-structured multiplicative uncertainty is derived:

$$\Delta(s) = e^{-2s\tau}I - I \tag{6}$$

Clearly, the parametric uncertainty of the time delay τ has been incorporated in the multiplicative uncertainty Δ(s). Hence, a class of uncertainty Δ_m is defined such that for every time delay $\tau < \tau_c$ there is a $\Delta \in \Delta_m$ and β so that $\|\Delta(jw)\|_\infty < \beta$. In this case

$\|\Delta(j\omega)\|_\infty = \max_\omega \left|e^{-2j\omega\tau} - 1\right|$. Fig. 4 shows several frequency response curves of the above multiplicative uncertainty for different values of the time delay.

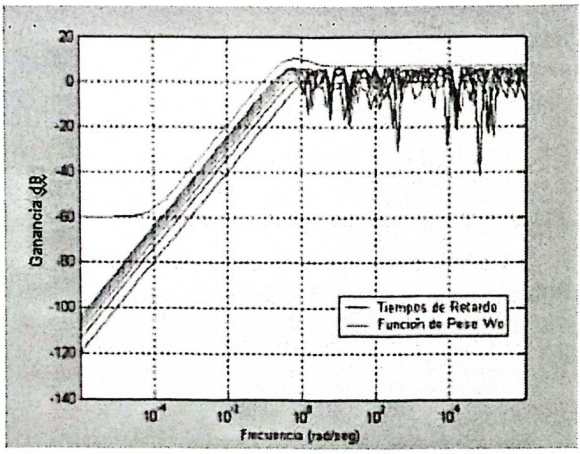


Fig. 4. Frequency response of the multiplicative uncertainty $\Delta(jw)$.

3 Design Problem

The objective of the design is to compute a controller by only using output feedback such that the controlled system remains stable on ample values of the time delay and satisfies the following design specifications:

-Tracking of a step reference signal with a steady state error less than 2%

-Sensor and communication channels noise attenuation.

-No actuator saturation.

In order to apply Hoo optimization design method to derive a controller to satisfy the above specifications we use the augmented system shown in Fig. 5.

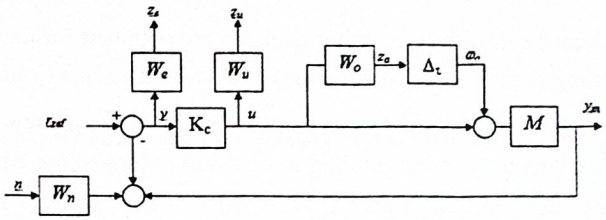


Fig. 5. Augmented system

Here the *Wi(jw) i=1,2,..4* are weighting 3x3 matrices where *Wo(jw)* is designed to normalize the multiplicative uncertainty, that is if $\Delta \in \Delta_m$, then $\|W_o(j\omega)\Delta(j\omega)\| < 1$, *Wu(jw)* models or normalizes the control signals amplitude under different performance conditions such that actuators saturation be avoid, *Wn(jw)* emulates sensor noise, and *We(jw)* gives as output the vector signal $z_e(t)$ that indicates the system performance when subject to reference, and noise inputs as well as modeling uncertainty. The external signals to the system are: the referente and noise signals , *r(t),* and *n(t),* and $w_a(t)$ signal that represents the way the uncertainty affects the system. y*(t)* is the input to the controller that for this case is the error signal of the system, *u(t)* the control signal, and *ym(t)* is the measured output signal. The output signals taken for design are $z_u(t)$ the normalized control signal, $z_e(t)$ the normalized error signal, and finally $z_a(t)$ the equivalent weighting response to model uncertainty. From a simple analysis of the diagram of Fig.5 the following transfer function matrix between the external vector signal $\overline{w} = \begin{bmatrix} w_a^T & r^T & n^T \end{bmatrix}^T$ and the output vector signal $\overline{z} = \begin{bmatrix} z_a^T & z_e^T & z_u^T \end{bmatrix}^T$ of the augmented system is derived:

$$\Im_{wz} = \begin{bmatrix} -W_a S_o KM & W_a S_o K & -W_a S_o K W_n \\ -W_e S_o M & W_e S_o & -W_e S_o W_n \\ -W_u S_o KM & W_u S_o K & -W_u S_o K W_n \end{bmatrix} \tag{7}$$

Hence, the design H_∞ optimization problem is stated as follow: *Compute a controller Kc(jw)) such that the close loop system became robust stable in ample*

values of the communication channels time delays, with robust performance with respect to design specifications, and

$$\min_{K(s)} \left\| \Im_{wz} \right\|_{\infty}$$

4 Weighting Matrices

Next, we describe how the weighting matrices are obtained for design purpose.

The weighting matrix $W_o(j\omega)$ is chosen such that its maximum singular value for all frequencies $\bar{\sigma}(W_o(j\omega))$ covers the family of frequency responses of the multiplicative uncertainty corresponding to different values of the time delay in order to satisfy the condition $\left\| W_o(j\omega)\Delta(j\omega) \right\| < 1$. The matrix transfer function obtained, in this case, to satisfy the above condition was:

$$W_o(s) = \frac{2.2387(s + 0.089 \times 10^{-3})(s+2)}{(s+0.4)(s+1)} I_3 \tag{8}$$

which frequency response is shown in Fig. 4. Wn(jw), the noise weighting function, models the frequency spectrum of the noise present in the system. It was found that the servopotenciometer, used for position sensing, were the source of high intensity noise at low frequencies as compared with the communication channel noise, so, a high-pass filter model was proposed with 35 dbs attenuation up to two octabs from the bandwith of the nomial plant Mn(s), (0.42 rad/sec), and with a 6 dbs constant attenuation at high frequencies. Then, the corresponding weighting transfer matrix obtained was:

$$W_n(s) = \frac{2(s+4)}{s+80} I_3 \tag{9}$$

The weighting matrix Wu(jw) was mainly applied in such a way as to avoid as much as possible the actuators saturation under different performance. The control input *u(t)* in this particular case represents the angular movements of the robot, so Wu(jw) was taken as a scalar attenuation of 1/170 in all control channels. As with respect to the error weighting matrix We(jw), this models the performance of the system to referente and disturbance inputs of low frequency. Therefore, in order to obtain a good time responce to step inputs, an integrator is a good option. Here instead of using this model an approximation was used by the application of a lag filter in the three channels of the error vector. As a rule this last weightig matriz is the

main function for tuning the controller. Thus, after several simulations, it was obtained the following matrix:

$$W_e(s) = \frac{10\left(s/5 + 1\right)}{\left(s/0.0021 + 1\right)} I_3 \tag{10}$$

The final step of the Hoo design is to find out if the controlled system satisfies the nominal and robustness conditions which are:

For robust stabilty we look for:

$$\left\| -W_a S_c K_c M \right\|_\infty < 1 \tag{11}$$

For a good tracking, and rejection of referente and disturbance respectively as well of high noise attenuation we need:

$$\left\| \begin{matrix} W_e S_M & -W_e S_M W_n \\ W_u S_c K_c & -W_u S_c K_c W \end{matrix} \right\|_\infty < 1 \tag{12}$$

which is the condition for a good nominal performance.

Finally, for robust performance the controlled system must satisfy the following condition:

$$\left\| \mathfrak{I}_{wz} \right\|_\infty < 1 \tag{13}$$

From application of the Robust Toolbox of MATLAB, we obtain the following results: A high order controller Kc(jw) which was reduced to a lower order controller (10th order), without much deterioration of the performance and stability conditions of the controlled system. The controller shows a high gain at low frequencies (20dbs) with a cross frequency around 0.08 rad/sec, and with a roll-off of -20 dbs/dec at intermediate frequencies. The frequency responses of equations (11), and (12) are shown in Fig.6 below. It is clear from this graph that the conditions for robust stability and nominal performance were well satisfied by the design.

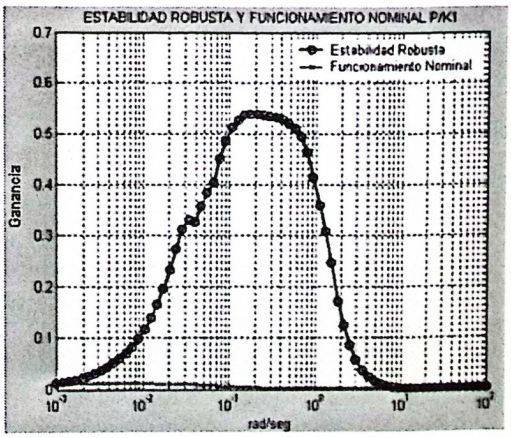


Fig. 6. Robust stability and Nominal performance requirement.

Robust performance condition Eq. (13), was examined by computing the corresponding structural singular value μ of the equation. When this value is less than one, then the system becomes robust on performance. For this design a value of $\mu = 0.79$, was obtained, meaning that the controlled system stands up to 126% non structured modeling variations.

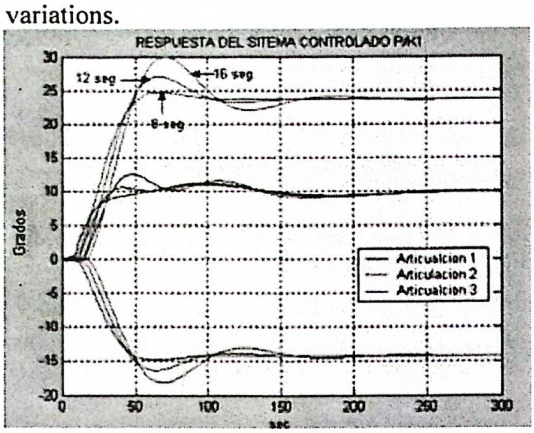


Fig. 7. Step responses of the controlled system.

The time response of the controlled system to step reference inputs of 10, -15, and 25 degrees respectively, on the three joints of the robot are shown in Fig.7 for different values of the time delay.

Steady state errors of up to 6% were obtained for all values of the time delay. Therefore more subtle tuning is needed to reach the specifications on this performance. The specification on noise attenuation was ample satisfied as can be seen in Fig.8 below

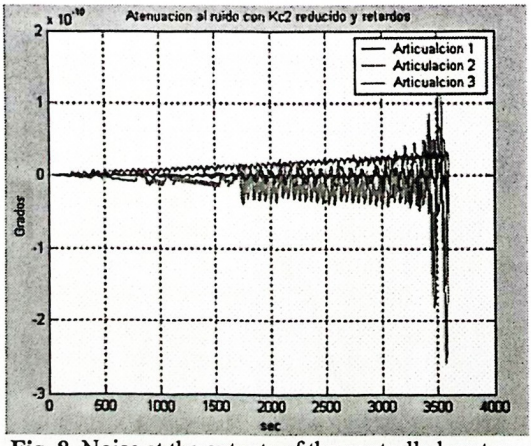


Fig. 8. Noise at the outputs of the controlled system.

Before the Hoo design was applied, a PI controller design was carried out on each of the loops of the telecontrolled system. It was found that for such controllers the system became unstable when a 3.4 secs time delay was present in the communication channels while for the Hoo design this time was extended almost 7 times to 21 secs., and good time responses were obtained for up to 16 secs. as can be seen in Fig.7. Therefore, the design condition of extending the stability and performance of the system, to ample values of the time delay in the communication channels was reached.

5 Conclusions

Due to the new technologies on integrated circuits for wireless communication the cost, and size of the main elements, transmitters and receptors, of the communication system have been reduced and now it is possible to consider the application of feedback structure of close loop telecontrol systems. As is well know the main problem of such structure is the introduction of uncertainty on the time delays present in the communication channels. There exist control design techniques that can extend the performance and stability of this system to ample values of the time delay. Hoo is one of such techniques. Here, it has been applied to a close loop telerobotic system and has proved to introduce significant improvement with respect to extend the performance and stability to ample values of time delay in the communication channels, as compared with other controllers. Also, the controlled system shows great noise attenuation and exelent disturbance rejection .

References

1. M. Cenk Cavusoglu, "Bilateral Controller Design for Telemanipulation in Soft Enviroment", Procc. IEEE Int. Conf. on Robotics and Aut. (ICRA 2001), Seoul Korea, May 2001.
2. G.M.L. Leung, Bruce A. Francis, and J. Apkarin, " Bilateral Controller for Teleoperators with Time Delay via μ - synthesis ", Dpt. Elect., and Comp. Eng., University of Toronto, April 22, 1997.
3. Zhou, K., Doyle J. " Essentials of Robust Control", Prentice Hall 1996.
4. Doyle J., Glover K., Khargoneckar P.P.," State Space Solutions to Standar H2, and Hoo Control problems.", IEEE Transc. Aut. Contr. Vol 34, No.8, pp 831-847,1989.

Author Index

Abdi, Hamid 25
[illegible].....25
Ahmadi, Abolfazl 25
Albertos, Pedro 1, 73
Alva, Rodrigo 133
Angulo, Modesto 73

Brashevan, Alexandra N. 115

Córdova, Rodrigo 133
Caldeira, Marco 223
Carbajal Hernández, José Juan ..233
Carrizales Muñoz, Zaira 163
Chile, Rajan 95
Concha, David 193
Costa Montiel, Angel 213
Cuesta, Alfredo 63

de Madrid, Angel P. 85
Domínguez, Ariel 73

Espadas, Javier 193

Ferasoli, Humberto 223
Franchin, Marcelo 153

Gómez Islas, Francisco 163
Galeano, Nathalie 193
Gallegos-Funes, Francisco Javier ...43
Gonzalez, Luis 261
Grau, Luis 63
Gutierrez Tornes, Agustin Francisco 53

Hernández, Jorge 3
Hernandez, Roberto 85

Jaime, Beatriz 243
Jiménez Madrigal, Emilio ... 133, 163

Kazakov, Vladimir 33, 253

López Cárdenas, Rodrigo.....23 243
Lopez, Ignacio 63
Lopez, Maximo 53
Lukin, Vladimir 105

Márquez Ramírez, Rafael Antonio 203
Mañoso, Carolina 85
Manrique Ramírez, Pablo ...125, 183, 203
Molina, Arturo 193

Najera, Teresa 193
Noriega, Alfonso 143

Orantes Jiménez, Sandra Dinora ..53

Pérez -Aguila, Ricardo 15
Patil, Nitin 95
[illegible] , I. 1
Pegoraro, Renê 223
Pichardo, Jesús 143
Pogrebnyak, Oleksiy 43, 105, 115, 125, 183, 203
Ponomaryov, Volodymyr 43
Popov, Anatoliy 105, 115

Ramírez, Mercedes 73
Rodriguez, Jose 261
Rojo, Arturo 173
Romero, David 193
Romero, Miguel 85
Rosário, João 223
Rosales-Silva, Alberto 43

Sánchez Fernández, Luis Pastor . 125, 173, 233, 243
Salami, Abolfazl 25
Sanchez, Luis Alejandro 173
Sanchis, R. 1
Sedeno, Monica 253
Silva, Carlos 143

Tejeda, Francisco 33

Vicente Vivas, Esaú 133, 163
Vilaragut Llanes, Miriam 213

Waghmare, L. M. 95

Yonezawa, Wilson 223

Zriakhov, Mikhail 105

Impreso en los Talleres Gráficos
de la Dirección de Publicaciones
del Instituto Politécnico Nacional
Tresguerras 27, Centro Histórico, México, D.F.
Noviembre de 2007.
Printing 500 / Edición 500 ejemplares.

We are privileged to present this special issue that contains the best works presented at the First International Conference on Industrial Informatics CICINDIN 2007, which is organized by the Center of Computing Research belonging to National Polytechnic Institute. The conference will be held in México City. CICINDIN is an international conference that joins contributions obtained by specialists in areas such as: Automation, Virtual Instrumentation and Signal Processing, Real Time Systems, Infrastructure, Technology and Data Industrial Network, Enterprise Management and Applications of Industrial Informatics, specialized areas that are having a high impact on contemporary societies. We received several papers from researches of many parts of the world. We accepted only the best papers. Every paper was evaluated by three reviewers with the common objective of select the papers with the highest quality.

We want to express our acknowledgement to all the people that contribute with this magazine. We are specially indebted with editors, reviewers and program committee members for their outstanding work and advice. Finally we would also like to thank to the National Polytechnic Institute for facilities given in order to achieve the success of this Conference, which we believe, is so much important for the academic and scientific life in this country.

Luis Pastor Sánchez Fernández
Oleksiy Pogrebnyak
Elsa Rubio Espino

ISSN: 1870-4069
www.cic.ipn.mx/rcs
www.ipn.mx

SEP

INSTITUTO POLITÉCNICO NACIONAL
"La Técnica al Servicio de la Patria"