

Image Recognition System for the Mexican Sign Language

Fausto Pavel Priego Pérez¹, Jesús Manuel Olivares Ceja¹,
José Félix Serrano Talamantes², Diana Naim Rivera Aguilar³

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

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

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

pavelpriego@gmail.com, jesuso@acm.org,
jfserrano@ipn.mx, dr_euro7@hotmail.com

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

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

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

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

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

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

1 Introduction

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

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

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

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

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

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

2 State of art

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

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

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

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

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

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

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

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

3 Mexican Sign Language (LSM)

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

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

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

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

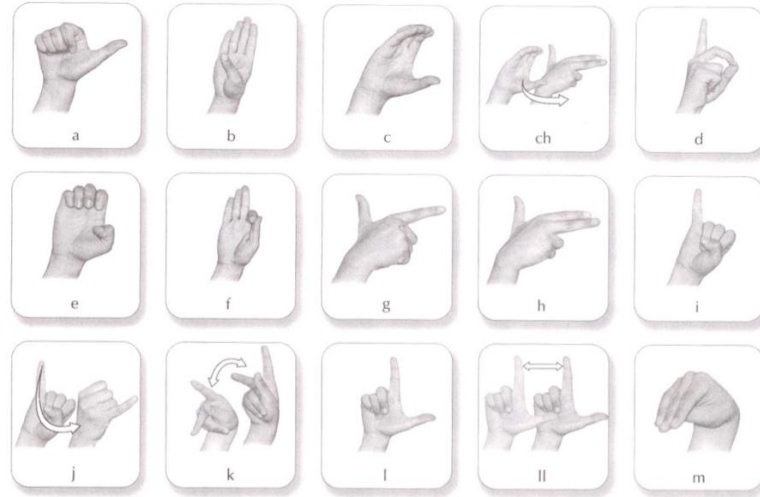


Fig. 1. Examples of Mexican Sign Language [15]

4 Proposed system

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

4.1 Architecture

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

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

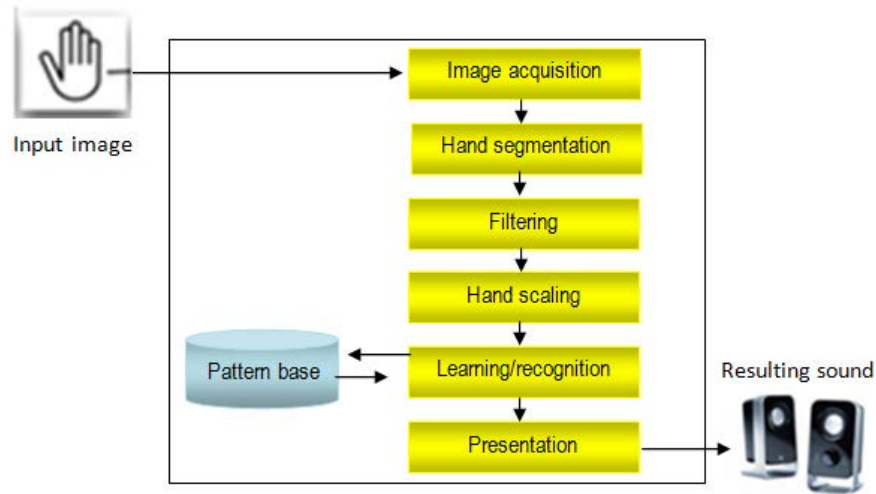


Fig. 2. Main modules of the proposed system

4.2 Image acquisition

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

4.3 Hand segmentation

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

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



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

4.4 Filtering

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

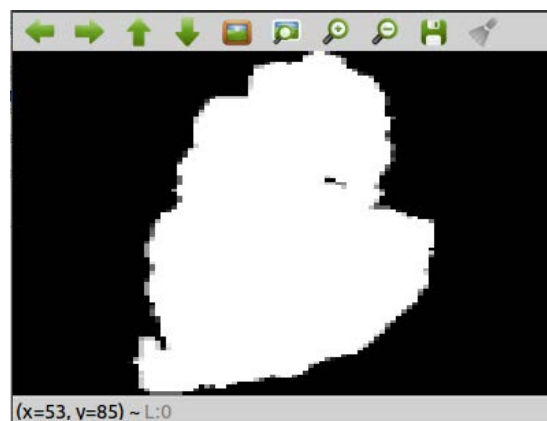


Fig. 4. Example of image segmented using a binary representation

4.5 Hand scaling

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

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

4.6 Learning/recognition

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

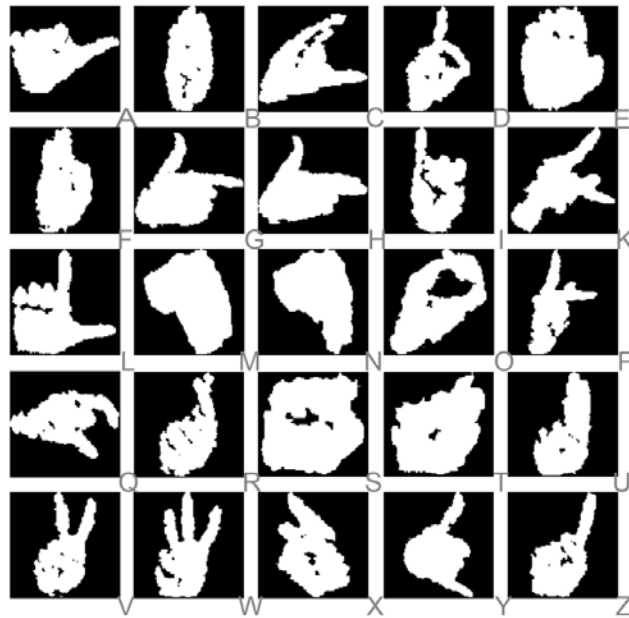


Fig. 5. Usual stored sign patterns example

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

Table 1. Example of the pattern base

y_0	y_1	y_2	y_3	y_4	$\dots$	y_{99}	Label
1	1	0	1	0		0	A
1	1	1	0	0		0	B
1	1	0	0	0		1	C
							.
							.
							.

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

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

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

where:

S: similarity percentage in range $[0, 1]$

x_i : each pixel in the image to be recognized

y_i : each pixel in the pattern contained in the pattern base

$\text{compare}(x, y)$: function that returns 1 if $x = y$, 0 otherwise

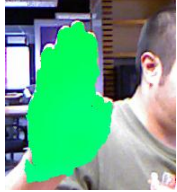
4.7 Presentation

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

Table 2. Distance between patterns

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

Table 3. Example of the recognition process

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

5 Test and results

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

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

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

6 Conclusions and future work

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

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

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

7 Acknowledgements

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

8 References

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