2017 14th International Conference on Electrical Engineering, Computing Science and Automatic Control (CCE)
Mexico City, Mexico September 20-22, 2017

LSM static signs recognition using image processing

Luis M. Pérez*, Alberto J. Rosales†, Francisco J. Gallegos‡ and Ana V. Barba§
*†‡ § Instituto Politécnico Nacional de México, S.E.P.I. Maestria en Ciencias en Ingeniería Electrónica
Av Luis Enrique Erro S/N, Unidad Profesional Adolfo Lopez Mateos, Zacatenco
Gustavo A. MAdero, C.P. 07738, Ciudad de México.
*Email: luisuacm1987@gmail.com
,†Email: arosaless@ipn.mx
‡Email: fgallegosf@ipn.mx
§Email: ana_cris814@hotmail.com

***Abstract*—Object recognition is a widely field in artificial vision application, because now the machines are intended to become autonomous. This article presents the methodology for recognizing objects in an image, tecniques used are: segmentation, feature extraction and classification object within the image. Fuzzy c-means algorithm was used for segmentation, which is a fuzzy classification algorithm in which a data can belong to multiple groups in different degree of membership. For feature extraction were used Hu moments as a geometrical descriptors, which are a mathematical tool that provides seven moments that identify geometric features of the objects, main characteristics of Hu moments is its invariance to rotation, scaling and translation. Finally, the geometric features that provide the seven moments are used as input to a classifier, delivering results: these moments are used to identify the signs of the alphabet of the Mexican Sign Language (LSM).**

I. Introduction

In Mexico there are 639,000 people diagnosed with hearing impairment [1], so it is necessary to design artificial vision systems to interpret the sign language known as Mexican Sign Language (*Lengua de Señas Mexicana, LSM*) to achieve a better inclusion to society of people with hearing impairment.

There were two works found in scientific literature about sign language detection, the first was developed in Colombia [2] in which an image containing the hand sign is captured by a camera, image preprocessing algorithms are applied for noise suppression, afterwards the hand edges are obtained resulting in a binary image, after, the image is converted to a row vector, which is the input for an artificial neural network (ANN) to perform the hand sign classification. The second work is proposed using a Kinect device [3] to perform an image segmentation of the hand in the image, it is followed by an image scaling obtaining a 5x5 matrix, which is the input for one layer ANN that classifies the hand sign, so a basic block diagram was found for hand sign detection , which consists in a block that performs the segmentation of the object to analyze, and once the object is isolated in the image, a feature extraction of the object is performed and finally a classification is performed within a predetermined number of classes to achieve a satisfactory detection of in the LSM. The primary objective consists in static signs reconition of the LSM using a series of image processing algorithms and comparing the results against the algorithms found in scientific literature.

The present work focuses on the LSM signs recognition using static signs, using preprocessing techniques, feature extraction and classification of patterns.

II. DEVELOPMENT AND THEORY

A. Hand segmentation

The primary task is to isolate the hand from the background of the image, for this a Fuzzy C means (FCM) is used, which is an unsupervised clustering algorithm based in fuzzy logic [4]. Equation that defines th following algorithm:

$$J_{FCM} = \sum_{k=1}^{n}\sum_{i=1}^{j}(W_{ik})^m \|x_k - c_i\|, \tag{1}$$

where,

c_i is the cluster center i where $1 \leq i \leq j$.
W_{ik} is the degree of membership of x_k in the cluster i [5].
m is the fuzzy factor $1.4 \leq m \leq j$.
n and j are the rows and columns of the image.
k and i are the pixel position of the image.

The variables used in the FCM algorithms are initialized: x is the matrix that contains the magnitude of the pixels that conform channel C_r of the image. c is the center of a class for C_b and C_r independently. For C_b a magnitude value of 110 is taken and 160 for C_r. J_{FCM} is the objective function, m is the weighing power with a value of $m = 2$, because there were no significant changes in the result by increasing the value of m. ε is the minimum difference between the value of the current objective function and the prior, if the value of ε is below 0.001 the algorithms stops. W is the membership matrix, which is initialized in a random fashion. Lastly there is the maximum number of iterations, in this case 5 is the number proposed as a higher number does not deliver a better result.

B. Hand feature extraction

Feature extraction describes objects with sets of measurable quantities (descriptor). To carry out the hand feature extraction moments of Hu are applied, which are geometrical descriptors that provide a scalar quantity describing the geometry of the object[6]. These moments being invariant to translation, scaling and rotation, the central and normalized central moments can be derived. The moments in an image $f(x, y)$, are described by equation 2[7]:

$$m(p, q) = \sum_x \sum_y x^p y^q f(x, y), \tag{2}$$

where,
p and q indicate the moment order.
(x, y) is the spatial position of the pixel in the image.
m is the geometric moment of the hand.

In equation 2 low order moments are observed, where p and q take values between 0 and 2, which provides the area and geometric center of the object $(\bar{x}, \bar{y})$. The central moments of the image are expressed in equation 3:

$$\mu_{pq} = \sum_x \sum_y (x - \bar{x})(y - \bar{y}) f(x, y), \tag{3}$$

The central moments are invariant to the translation of the object inside the image as they are dependent of the geometric center of the object, where:

$$\bar{x} = \frac{m_{10}}{m_{00}}, \bar{y} = \frac{m_{01}}{m_{00}}. \tag{4}$$

Lastly there are the normalized central moments, which are invariant to object scaling in an image, described by equation (5).

$$\eta_{pq} = \frac{\mu_{pq}}{\mu_{00}^{\gamma}}, \tag{5}$$

where,

$$\gamma = \frac{p + q}{2}. \tag{6}$$

In 1962 Ming-Kuei Hu[6], performed algebra with the normalized central moments and determined seven moments that are invariant to translation, rotation and scaling of an object in an image. The seven Hu moments are derived from the normalized central moments of second and third order and the following equations describe it:

$$\begin{aligned}
\phi_1 &= \eta_{20} + \eta_{02}, \\
\phi_2 &= (\eta_{20} - \eta_{02})^2 + 4\eta_{11}^2, \\
\phi_3 &= (3\eta_{30} - 3\eta_{12})^2 + (3\eta_{21} - \eta_{03})^2, \\
\phi_4 &= (\eta_{30} - 3\eta_{12})(\eta_{30} + \eta_{12})((\eta_{30} + \eta_{12})^2 \\
&\quad -3(\eta_{21} + \eta_{03})^2) + (3\eta_{21} - \eta_{03})(\eta_{21} + \eta_{03}) \\
&\quad (3(\eta_{30} + \eta_{12})^2 - (\eta_{21} + \eta_{03})^2), \\
\phi_5 &= (\eta_{30} - 3\eta_{12})(\eta_{30} + \eta_{12})((\eta_{30} + \eta_{12})^2 \\
&\quad -3(\eta_{21} + \eta_{03})^2 + (3\eta_{21} - \eta_{03})(\eta_{21} + \eta_{03}) \\
&\quad (3(\eta_{30} + \eta_{12})^2 - (\eta_{21} + \eta_{03})^2), \\
\phi_6 &= (\eta_{20} - \eta_{02})[(\eta_{30} + \eta_{12})^2 - (\eta_{21} + \eta_{03})^2] \\
&\quad +4\eta_{11}(\eta_{30} + \eta_{12})(\eta_{21} + \eta_{03}) \\
\phi_7 &= (3\eta_{21} - \eta_{03})(\eta_{30} + \eta_{12})((\eta_{30} + \eta_{12})^2 \\
&\quad -3(\eta_{21} + \eta_{03})^2) - (\eta_{30} - 3\eta_{12})(\eta_{21} + \eta_{03}) \\
&\quad (3(\eta_{30} + \eta_{12})^2 - (\eta_{21} + \eta_{03})^2).
\end{aligned} \tag{7}$$

C. ANN for sign classification

En the present work multilayer perceptron ANN used, and as a classifier, its described by the following equation[8]:

$$j = \sum_{i=1}^{n} (\omega_{ij} x_i) - \theta, \tag{8}$$

where, x_i is the neuron inputs, $i = 1, 2, 3, ..., n$ and ω_{ij} is the synaptic weights of each of the inputs to the neuron j and θ is the neuron compensators.

In Fig. 1 shows the architecture of a multilayer perceptron ANN, the reason for using an ANN in this work is due to the capability of finding patterns in the input data each time that the training phase is done[8]. The preceding is performed in the hidden layers, transforming the input data into patterns that identify the network for an optimal classification[9].

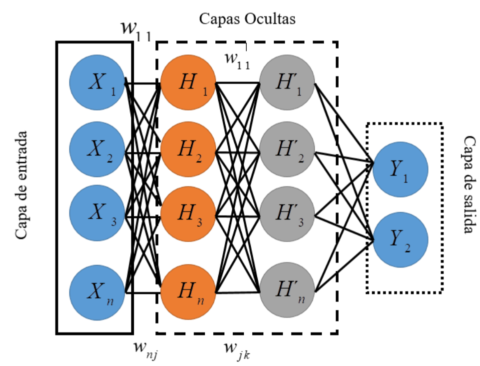

Fig. 1: Multilayer perceptron ANN general architecture diagram.

III. PROPOSAL FOR DETECTION OF LSM SIGNS

The proposal for LSM sign detection by image processing is show in Fig.2.The block diagram used in this work is shown in Fig. 2, where the image acquisition was obtained with a

Sony Cyber-Shot digital camera in a controlled environment, which consist in a black background and a shot distance of 25-35cm. A database was created consisting of 1680 images with dimensions of 640 x 480 pixels, belonging to 80 subjects.

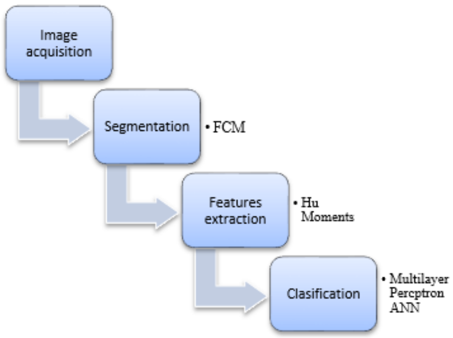

Fig. 2: Proposed block diagram.

In the segmentation shown in Fig. 3 there is digital shape of a hand with a background. A FCM algorithm was used to obtain the intensity values corresponding to skin pixels, achieving the segmentation of the hand.

Fig. 3: A letter sign segmented with FCM.

Equation (7) is used in the features extraction to get the seven moments that describe the object in question. The base 10 logarithm and absolute value were calculated for each moment so they could be compared. In Table I the seven Hu moments for letters A, B, C, L, W, and Y are shown.

The data in TABLE I shows that the seven moments describe in a particular way each of the static signs of the LSM, and for that they are applied as the input data for the neural network for pattern recognition providing an array that will be analyzed afterwards.

TABLE I: Seven Hu moments of letters A, B, C, L, W, Y

Hu Moments	A	B	C	L	W	Y
ϕ_1	0.69	0.70	0.52	0.68	0.64	0.65
ϕ_2	2.03	1.87	1.35	1.79	1.73	1.95
ϕ_3	2.83	3.55	2.94	3.01	3.72	2.49
ϕ_4	3.46	4.46	3.62	3.00	4.01	3.55
ϕ_5	6.68	8.67	7.12	6.16	8.24	6.87
ϕ_6	4.56	5.62	4.46	3.93	5.11	4.89
ϕ_7	7.44	9.15	7.27	6.50	8.21	6.97

The classification block analyses the seven column array of Hu moments via a neural network and then it decides to which static sign of the alphabet of the LSM does the hand sign of the processed image belongs to.

Then the used architecture of the multilayer perceptron is defined in the present work in Fig. 4.

- **Input Layer**: The seven Hu moments are used as a seven column array.
- **Hidden Layer**: Only one hidden layes is used, which contains 177 neurons nad each of them has a sigmoid activation fucntion.
- **Output Layer**: It´s a 21 element column array where each element represents a letter of the LSM alphabet.

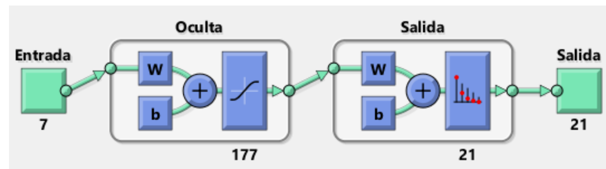

Fig. 4: MLP utilizado en el reconocimiento de seas estáticas de la LSM

The multilayer perceptron ANN was trained the back-propagation learning algorithm, consisting in minimizing equation (9), which describes the instantaneous energy of the error of the whole net.

$$\varepsilon(n) = \frac{1}{2} \sum_{j=C_0} e_j^2(n), \tag{9}$$

where, set C_0 includes all the neurons in the net and e_j^2 depends of the synaptic weights of each of the neurons of the ANN.

The results obtained in the training phase where obtained with a confusion matrix (Fig.5), which provides information of the training phase. Some of its primary characteristics are that the matrix diagonal shows the number of correct calssification of each of proposed classes, the right lower part of the matrix shows the total percentage of correct classifications.

Then some results are shown from the training phase, you can observe that in the first row, the ANN output has 74 predictions for the A sign, with a 96% correct and only 4% incorrect classifications, providing the false positives.

The first column of the matrix provides the false negatives being that of the 80 cases of the A sign introduced to the algorithm, 93% was correct prediction and 7% a wrong one.

For the letter S row, of the 80 signs, 72% was of correct classification and the algorithm confused the sign with one corresponding to the letter B eight times.

For the F sign row, there is an 84% correct classification and 5% of the time the algorithm confuses it with the letter N sign.

IV. RESULTS AND DISCUCCION

An evaluation was performed condifering the same condition that were used in the algorithms found in the literature. the first algorithm reports 20 of 21 correctly classified, the evaluation conditions were with two subjects, providing a total of 42 sign images, the sign language is the colombian

A	74	0	0	1	0	0	0	0	0	0	0	1	0	0	1	0	0	0	0	0	0	96%
B	0	54	0	0	1	1	0	0	0	0	4	0	2	0	1	8	3	0	0	0	0	73%
C	0	0	79	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	100%
D	1	0	0	78	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	99%
E	0	2	0	0	73	2	0	0	0	0	2	0	0	0	0	4	1	0	0	0	0	87%
F	0	1	0	0	2	67	0	0	0	0	3	7	0	0	0	1	0	2	0	0	1	80%
G	0	0	0	0	0	0	76	0	0	0	0	0	0	0	0	0	0	0	0	0	0	100%
H	0	0	0	0	0	0	3	80	0	1	0	0	0	0	1	0	0	0	0	0	0	94%
I	0	0	0	0	0	0	0	0	80	0	0	0	0	0	0	0	0	0	0	0	0	100%
L	0	0	0	0	0	0	0	0	0	77	0	0	0	0	0	0	0	0	0	0	0	100%
M	0	10	0	0	2	1	0	0	0	0	64	0	0	0	0	5	1	0	0	1	0	76%
N	1	0	0	0	0	2	0	0	0	0	0	56	0	0	3	1	0	9	0	0	0	78%
O	0	1	1	0	0	1	0	0	0	0	2	0	76	0	1	0	0	0	0	0	0	93%
P	0	0	0	0	0	0	0	0	0	1	0	0	0	78	0	0	0	0	0	0	0	99%
R	3	0	0	1	0	0	0	0	0	0	0	3	0	0	70	0	0	5	0	0	0	86%
S	0	8	0	0	2	0	0	0	0	0	5	0	1	0	0	58	7	0	0	0	0	72%
T	1	4	0	0	0	0	0	0	0	0	0	0	0	0	0	3	68	0	0	0	0	90%
U	0	0	0	0	0	4	0	0	0	0	0	13	0	0	3	0	0	64	0	0	0	76%
V	0	0	0	0	0	0	0	0	0	1	0	0	0	1	0	0	0	0	80	0	0	98%
W	0	0	0	0	0	1	1	0	0	0	0	0	1	1	0	0	0	0	0	79	0	95%
Y	0	0	0	0	0	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	79	99%
	93%	68%	99%	98%	91%	85%	95%	100%	100%	96%	80%	70%	95%	98%	88%	73%	85%	80%	100%	99%	99%	90%
	A	B	C	D	E	F	G	H	I	L	M	N	O	P	R	S	T	U	V	W	Y	

Fig. 5: Training phase confusion matrix.

A	6	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	100%
B	0	6	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	100%
C	0	0	6	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	100%
D	0	0	0	6	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	100%
E	0	0	0	0	5	1	0	0	0	0	0	0	0	0	1	0	0	0	0	0	1	62%
F	0	0	0	0	0	4	0	0	0	1	0	0	0	0	0	0	0	0	0	0	0	80%
G	0	0	0	0	0	0	5	0	0	0	0	0	0	0	0	0	0	0	0	0	0	100%
H	0	0	0	0	0	0	0	6	0	0	0	0	0	0	0	0	0	0	0	0	0	100%
I	0	0	0	0	0	0	0	0	6	0	0	0	0	0	0	0	0	0	0	0	0	100%
L	0	0	0	0	0	0	0	0	0	5	0	0	0	0	0	0	0	0	0	0	0	100%
M	0	0	0	0	0	0	0	0	0	0	6	0	0	0	0	0	0	0	0	0	0	100%
N	0	0	0	0	0	0	0	0	0	0	0	4	0	0	0	0	0	0	0	0	0	100%
O	0	0	0	0	0	0	0	0	0	0	0	0	5	0	0	0	0	0	0	0	0	100%
P	0	0	0	0	0	0	0	0	0	0	0	0	1	5	0	0	0	0	0	0	0	83%
R	0	0	0	0	0	0	0	0	0	0	0	0	0	1	4	1	0	1	0	0	0	57%
S	0	0	0	0	1	0	0	0	0	0	0	0	0	0	0	4	0	0	0	0	0	80%
T	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	6	0	0	0	0	100%
U	0	0	0	0	0	1	0	0	0	0	0	0	0	0	0	0	0	7	0	0	0	87%
V	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	6	0	0	100%
W	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1	0	0	0	0	5	0	83%
Y	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	6	100%
	100%	100%	100%	100%	83%	67%	100%	100%	100%	83%	100%	100%	83%	83%	67%	80%	100%	87%	100%	100%	86%	91%
	A	B	C	D	E	F	G	H	I	L	M	N	O	P	R	S	T	U	V	W	Y	

Fig. 6: Proposed algorithm results confusion matrix.

language. In the second algorithm found in the literature, the results were 20 of 21 static signs of the LSM. They did not have a database to compare results.

Afterwards the algorithm is evaluated using the proposed database. 124 images from the static signs of the LSM were used for the test, emphasizing that these images were not used for the training phase.

The confusion matrix indicates that the accuracy is of 91%, recognizing 111 signs in a correct manner, however for class six, belonging to the letter F, there was a rate of 20% of false positives due to the algorithm confusing it with the letter L.

In another matter, Fig. 6 shows that the algorithm classified wrongly the sign for the letter R, due to the algorithm confusing it 3 times, in a rate of 80% or lower along with the letter F sign.

V. Conclusion

The LSM can be applied for an automatic sign recognition system via computer vision algorithms, presenting optimal results for recognition and improving on other works related to the scientific literature. This algorithm recognizes static signs in the LSM correctly at a rate of 91%. This indicates that the result is optimal being that 9 out of 10 signs are recognized correctly by the algorithm, the experimentation conditions were that the process was done with six subjects equivalent to 124 images that were not trained in the ANN.

It can be observed that this algorithm has some problems with letter R sign being that the correct classification rate is below 70% but even so its an optimal solution as the process is of low cost and noninvasive.

Acknowledgment

The authors would like to thank the *Instituto Politécnico Nacional*, and *Consejo Nacional de Ciencia y Tecnología of México* for the provided support for the production of this research.

References

[1] Inegi: Personas con discapacidad y estadística de las personas sordas, `http://cuéntame.inegi.org.mx/población/discapacidad.aspx?tema=P. [Acceso 13 octubre 2015]`

[2] L. Vargas, L. Barba, L. Mattos *Sistema de identificación de la lengua de señas usando redes neuronales artificiales*, Revista Colombiana de Física, vol. 402, 2010.

[3] P. Pérez,*Reconocimiento de Imágenes del Lenguaje de Señ Mexicano*, Revista Colombiana de Física, vol. 402, 2010.

[4] J. Bezdek, R. Ehrlich, W. Full *FCM:The Fuzzy C-Means Clustering Algorithm*, Computer and Geosciences, vol. 10, pp.191-203, 1984.

[5] I. Ozkan, I. Turksen *Upper n Lower Values for the Level of Fuzziness in FCM*, , 2009.

[6] H. Ming-kuei *Visual Pattern Recognition By Moment Invariant*, IRE Transactions on Information Theory, vol. 8, no. 2, pp. 179-187, 1962.

[7] J. Fluster, T. Suk, B. Zitova *Moments and moment invariants in pattern recognition* , United Kingdom: Wiley, 2009.

[8] S. Haykin *Neural Network And Learning Machines*, New Jersey: Pearson Education, 2009.

[9] M. Bonifacio, A. Sanz *Redes Neuronales y Sistemas Borrosos*, Ciudad de México: Alfaomega, 2007.