Sign Language to Text by Feed Forward Back Propagation Neural Network

Shekhar Singh[*]

Assistant Professor

Department of Computer Science & Engineering

Panipat Institute of Engineering & Technology

Samalkha, Panipat , Haryana, India

shekharsinghsengar@gmail.com

Abstract: In this paper is presented an automatic deaf language to text translation and recognition system. The scheme is based on neural network (NN) classifier using a back propagation. The input parameter vector to neural network is the Fisher score, which represents the derivate of the matrix of symbol probability in hidden Markov model (HMM). The HMM, which needs a sequence to be trained and used, is fed by the hand contour chain code. Besides, an improvement on the calculation of Fisher score is introduced by means of reducing the kernel scores variance. The error ratio classifying hand text of the proposed system is less than 0.8% with our database. The objective of this effort was to explore the utility of a neural network based approach to recognition of the sign language.

Keywords: Computer vision, Human computer interaction, American Sign Language, Image processing, Chain Code, HMM, Fisher score, Neural network, Sign language recognition.

I. INTRODUCTION

The growing commitment of the society in reducing the barriers to persons with disabilities, added to the advances of the computers and form recognition methods, has motivated the development of the present system that transforms the sign language in the American text alphabet, because in each country the sign language is different. This system bases its operation in providing in the input an image of a sign alphabet American hand sign, and in their output the letter of the Latin alphabet that corresponds it, just as it is observed in the following figure 1, including the 25 American text letter. The used procedure in this automatic recognition is based in the adequate modeled of the hand sign by Neural Network**.** For it, it has been realized the calculation of Fisher score, which is extracted of HMM [2, 5]. This HMM is fed with the chain code, determined from the band sign image.
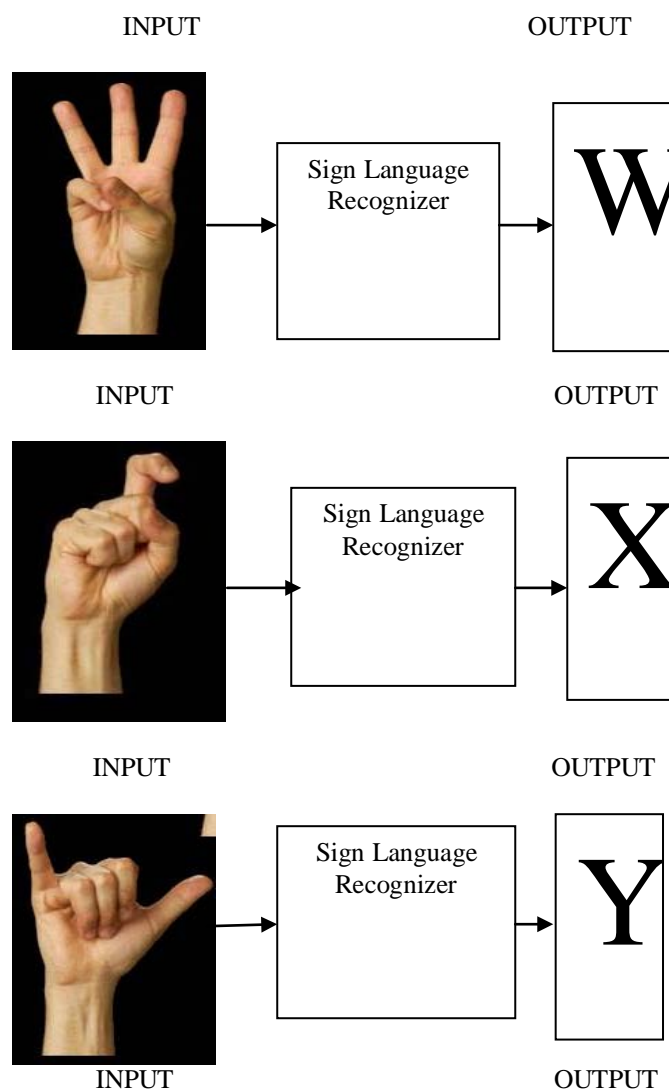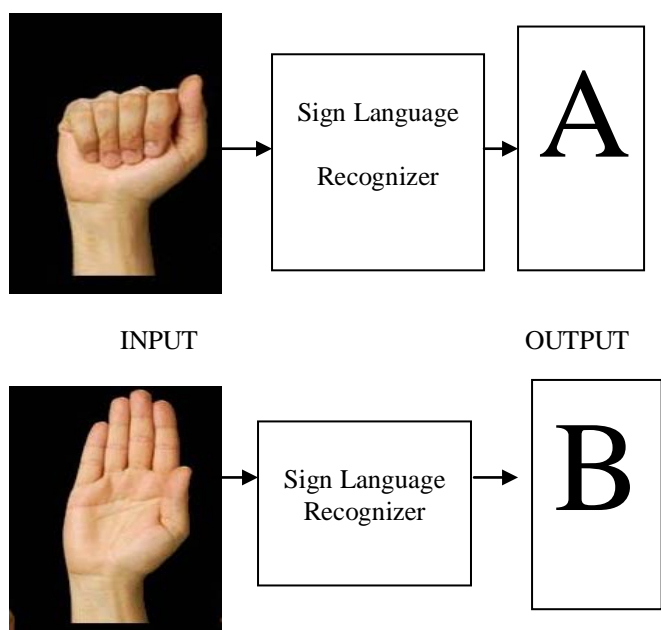

Figure 1: Translation from sign alphabet automatic recognition system, until the standard Latin alphabet

For this process, the first step has been arranged the capture of samples. It is necessary to create a database with hand sign pictures of the sign text of numerous persons, being the ideal infinite. In our case, it was enclosed for sixty different people (25 hand signs each one). Subsequently, it has been applied a image preprocessing, transforming the color image in a white and black one, with more reduced size, which defines the hand outline. With this outline, we are going to extract a series of parameter, associated in a vector that defines the contour of the hand sign with chain code [3, 6, 19, 20], and it is classified and parameterized by HMM. From these HMM, some parameters are going to extract, based on the emission probability of the vectors, which are going to determine the score Fisher [2, 6], for classifying it with Neural Network [10, 11, 23, 24]. This system will he supervised with a training process, in which is learned to differentiate some hand signs of other; and a other test process, where the models will be verified. This process is resumed in the following figure2. In the second section the creation of the database. In the third, the applied image processing. Subsequently, the calculation of Fisher score by HMM. In the fifth section the neural network classifier. The sixth sections are all the realized experiments and finally the conclusions and references.

II. BUILDING OF DATABASE

A database has been created for the execution of this paper that is enclosed of 25 hand signs or patterns, collecting 60 samples of each one all of different peoples. These images have been captured by a digital photograph camera. The creation of this database has had a series of difficulties, among them is emphasized the communications barrier of working with people with auditory disability. By another side, it has been worked with images static, without movement.

Some statistic data from database are the following:
a. A 60% of males and a 40% of female.
b. The 30% of the peoples are deaf and the remainder not.
c. The distribution in ages varies since the 20 to the 60 years old.
d. The images have a size of 640x480 pixels (wide x high), with a resolution 24 bits of color, in a format REG, that is to say, 8 bits for each one of the fundamental colors.

Finally, the hand signs are presented in the figure 3, in particular, a sample of each one of the hand signs to recognize.

III. SIGN LANGUAGE IMAGE PREPROCESSING

In this process are going to take the color images for transforming to binary images of hand sign shape (white and black) with a fixed height of 400 pixels, conserving the relation with respect the width.
The following steps are;
a. Filter RGB image for eliminating the noise, hue, saturation effects.
b. Convert RGB image in YCBCR color image by eliminating the hue and saturation [24].
c. Convert YCBCR image in grayscale by eliminating the hue and saturation [24].
d. Filter grayscale image for eliminating the noise, hue and saturation.

e. Enhance contrast using histogram equalization [3, 19, 20]. To realize this process, it is equalized the histogram of the different levels of gray, under a lineal function (see the figure 4), but without affecting to the darkest points, only for the clearest parts (the hand sign), marking differences among the shadow of the hand and the background; and the own hand.
f. Take the frame out to eliminate some border effects (see figure 5).
g. Convert image to binary image by Thresholding. The Thresholding is computed by means of Otsu's method, which chooses the threshold to minimize the interclass variance of the Thresholding black and white pixels. With this step is finished for determining the hand as an object.
h. Morphologic operators [4, 20]. It is applied the dilatation operations in a first place, and after the erosion, as effect to recover the conditions of size of the image. The elimination of the noise is other desired effect. The dilatation is going to unite the holes that can remain in the line of the contour. Subsequently the hand is filled as object, and finally recovers the original dimension, by means of the erosion, that is able to eliminate the possible noises happened by the own photo.
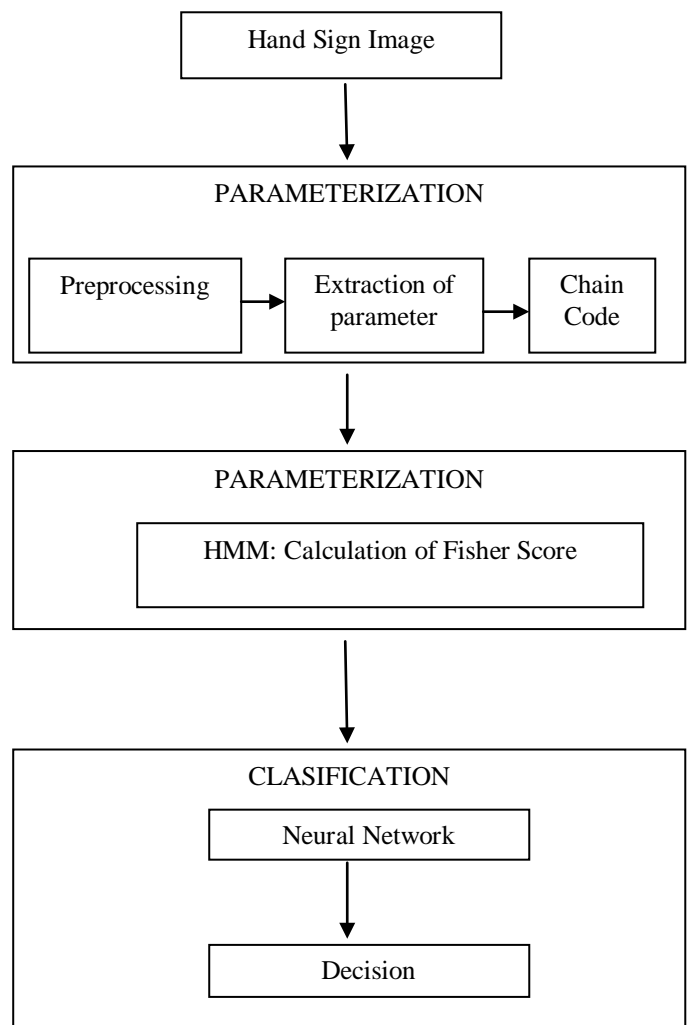

Figure 2: Procedure of automatic hand sign recognition system

i. Reduction. The size of image is reduced in one quarter, for reducing the size of the final vector.

j. Calculation of the contour [5, 6]. It is calculated the hand contour, which determines the hand sign on the background of the image, with the particularity that the connection among pixels is alone one pixel, with a connectivity among neighbors of 8, that is chain code.

k. Adjustment of the wrist. The image is trimmed slightly with the intention to determine the side of the one that the hand sign arises. Adjustment of high of the image. Finally it is fixed the value of the height, maintaining the proportionality with regard to the width. In this way the information is not loosen, indistinctly that the hand sign is horizontal, vertical up or vertical downward.

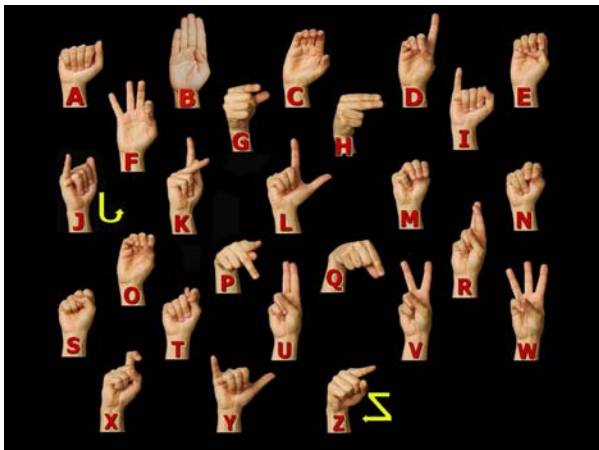l. In the figure 5, this process of processing image is shown, in particular, for the hand sign of the latter A and W.

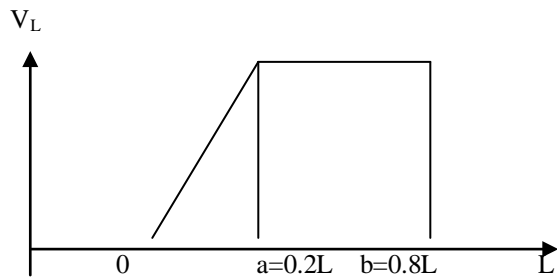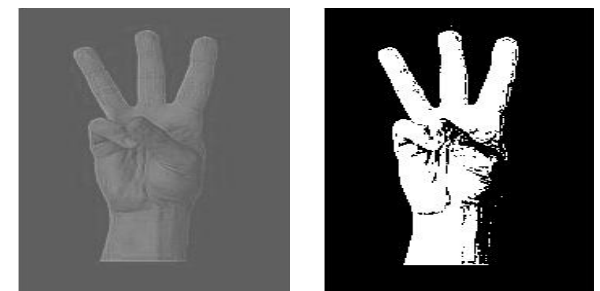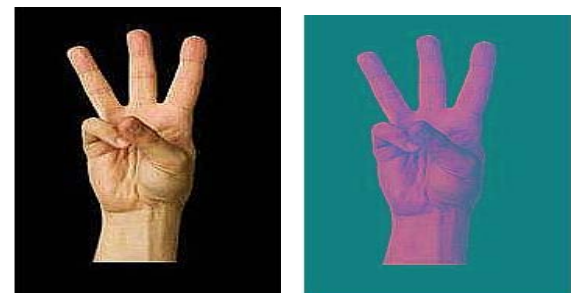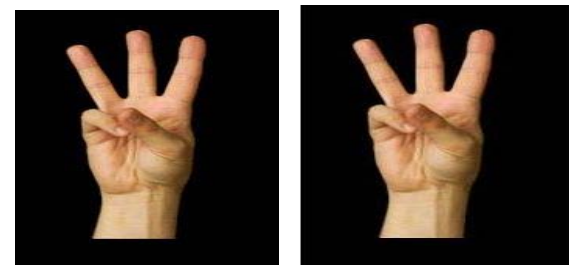Figure 3: Hand signs that conform the database

Figure 4: Applied equalization to the intensity
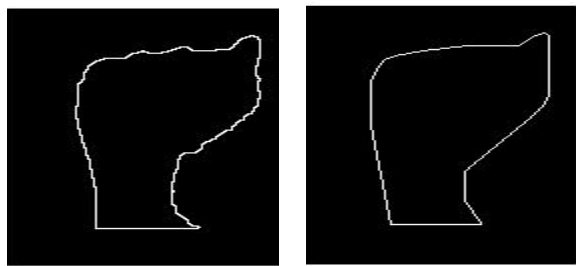
Original Image   (a) Resize Image   (b) Filter Image(c) YCBCR Color Image (d) Gray Image   (e) Binary Image (f) Morphologic (g) Close (h) Outline (i) Crop Image   (j) Adjustment   (k) minimum polygon adjustment

Figure 5: Steps of the applied processing hand signs image

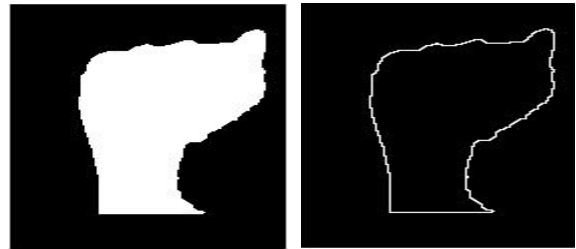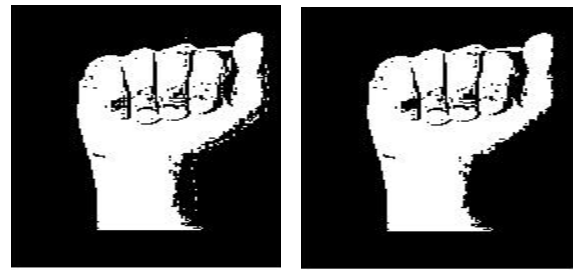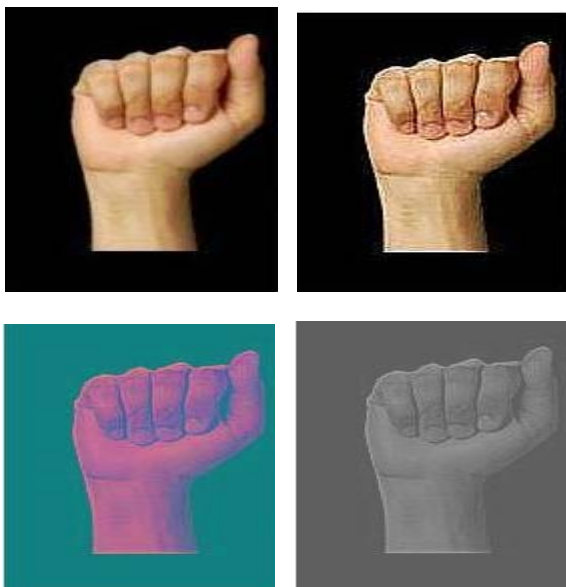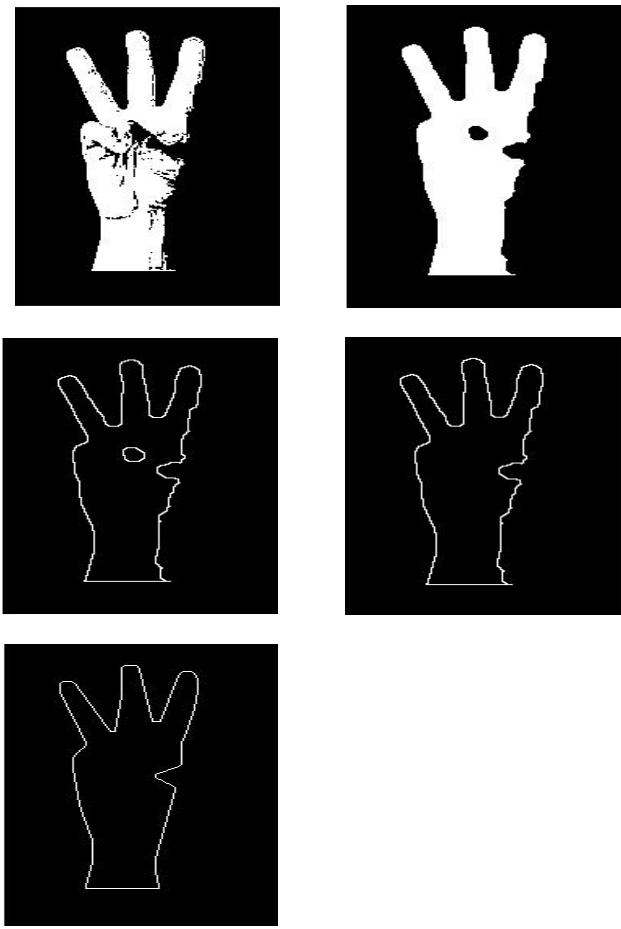In the figure 5, this process of processing hand signs image is shown, in particular, for the hand signs of the latter A and W.

IV. FISHER SCORE CALCULATION

Once it is obtained all the outline images of the hand signs, it is realized the calculation of Fisher score. This process comprises in three steps;

a. Extraction of parameter from outline: chain code [3, 20].
b. Creation HMM with chain code as input [5, 6].
c. Calculation of Fisher score from gradient of logarithm of the observation symbol probability distribution [2, 6].

A. Chain Code from Outline Hand:

This vector of the contour hand sign is obtained with the mask of the figure 6, observing the position of each pixel with their adjacent one. This vector is formed by numbers from the 1 to the 8 that describes the outline of the hand sign. The information that is extracted, describes the sequence of the hand, accompanied by temporary information, because all the hand signs are acquired in the same order and sense. This information is very important for the recognizer based on HMM, therefore utilizes this information to distinguish the different hand signs. It is fixed a criterion of start for the obtaining of this vector,

verifying first if is a vertical, horizontal since down or vertical since up hand sign, for this is not rotation invariant. It is begun for seeking write pixels with the following order of priority: first, the first column of the left (horizontal hand sign), the last row (vertical since down) or the first row (vertical since up). This order of priority will depend; evidently, of the probability to be produced each type of hand sign, thus, the vertical hand signs since up are the last, since only there are two.

1	2	3
8	X	4
7	6	5

Figure 6: Mask of composition of the vector of the chain code

B. Transformation of Parameter with HMM:

It is going to determine by the supervised classification of chain code using HMM, which is the maximum rate of success, for extrapolating the forward and backward parameter of HMM in the calculation of Fisher score. Therefore, the HMM employed is a Bakis, and is trained with the procedure of Baum-Welch, to maximize the probabilities of success [2]. Besides, 8 symbols by state have been utilized. The creation of the HMM models has two phases, the training and the test. Finally, the number of states (N) and the percentage of training samples have utilized like parameters to find the highest rate of success.

C. Fisher Score:

Finally, it is proposed the transformation that provides the HMM probabilities relating to the approach of the Fisher score [2, 6]. With this goal, it intends to unite the probability given by the HMM to the given discrimination of the neural network, whose tie of union **is** this Fisher score. This score calculates the gradient with respect to the parameters of HMM, in particular, on the probabilities of emission of a vector of data x, while it is found in a certain state q $\in$ {1,……., N}, given by the matrix of symbol probability in state q($b_q(x)$), just as it is indicated in the following equation ;

$$P(x/ q, \lambda) = b_q(x) \qquad \text{(Eq. 1)}$$

If it is realized the derivate of the logarithm of the above probability, with the purpose to calculate its gradient, it is obtained the kernel of Fisher, whose expression comes given by;

$$\partial\log p(x/q, \lambda)/\partial p(x, q) = \zeta(x, q)/b_q(x) - \zeta(q) \qquad \text{(Eq. 2)}$$

Where in [2, 5, 6], it has been found the approximations and the calculation of above equation. Besides, $\zeta(x, q)$ represents the number of times, that is localized in a state q, during the generation of a sequence, emitting a certain symbol x [2, 5]. $\zeta(q)$ represents the number of times that has been in q during the process of generation of the sequence [2, 5]. These values are obtained directly and of form effective, from the forward backward algorithm, applied to the HMM [2, 5]. The application of this score (U_x) to the neural network, comes given by the expression of the equation 2, utilizing the techniques of the natural gradient, from the following equation [1];

$$U_x = \nabla_{p(x, q)} \log p(x / q, \lambda) \qquad \text{(Eq. 3)}$$

Where U_x define the direction of maximum slope of the logarithm of the probability to have a certain symbol in a state.

V. CLASSIFICATION METHOD

The objective of this study is to classifying Fisher kernel data of sign language latter symbol using feed forward back propagation neural network and Levenberg-Marquardt (LM) as the training algorithm. LM algorithm has been used in this study due to the reason that the training process converges quickly as the solution is approached. For this study, sigmoid, hyperbolic tangent functions are applied in the learning process. Feed forward back propagation neural network use to classify sign language text symbol according to fisher score characteristic [10, 11, 23, 24]. Feed forward back propagation neural network is created by generalizing the gradient descent with momentum weight and bias learning rule to multiple layer networks and nonlinear differentiable transfer functions. Input vectors and the corresponding target vectors are used to train feed forward back propagation neural network. Neural network train until it can classify the defined pattern. The training algorithms use the gradient of the performance function to determine how to adjust the weights to minimize performance. The gradient is determined using a technique called back propagation, which involves performing computations backwards through the network. The back propagation computation is derived using the chain rule of calculus. In addition, the transfer functions of hidden and output layers are tan-sigmoid and tan-sigmoid, respectively.

A. Training and Testing:

The proposed network was trained with fisher score data cases. When the training process is completed for the training data, the last weights of the network were saved to be ready for the testing procedure. The time needed to train the training datasets was approximately 2.60 second. The testing process is done for 60 cases. These 60 cases are fed to the proposed network and their output is recorded.

a. **Performance plot:** Performance plot show the training errors, validation errors, and test errors appears, as shown in the training process. Training errors, validation errors, and test errors appears, as shown in the following figure 7.

b. **Receiver Operator Characteristic Measure (ROC) Plot:** The colored lines in each axis represent the ROC curves. The *ROC curve* is a plot of the true positive rate (sensitivity) versus the false positive rate (1 - specificity) as the threshold is varied. A perfect test would show points in the upper-left corner, with 100% sensitivity and 100% specificity. For this problem, the network performs very well. The results show very good quality in the following figure 8.

VI. EXPRIMENT RESULTS

The experiments have been realized with independent samples for training and for test. These have been repeated in five times to take averaged values, expressed them by their average and their variance. The execution of the same have been done sequential mode, first in the HMM to achieve its maximum success, varying the number of states

and the percentage of samples in the training. The results are the following:

Table 1: Rate of success of the HMM in function of the percentage of samples for training and of the number of states

Samples Training	Number of state				
	20	45	45	65	100
20%	75.66% ±22.82	86.60% ±3.33	87.74% ±2.47	85.38% ±7.67	64.40% ±45.65
40%	78.20% ±12.83	88.70% ±3.50	88.58% ±7.24	92.10% ±1.82	73.20% ±26.50
60%	76.78% ±11.20	91.90% ±2.95	87.30% ±3.90	87.98% ± 3.10	68.10% ±25.87
80%	77.31% ±31.42	92.75% ±14.19	93.20% ±8.30	92.50% ±14.15	75.65% ±16.80

Of the above table is deduced that the best percentage of training is fur 92.75% and 45 states, with a rate of 91.20%, presenting a small variance. Of this model is going to generate the kernel of Fisher, and it is going to apply to the neural network. The results of these tests are collected in the following table1.

It is observed as for the neural networks are presented better results, arriving to an average rate of 99.30%, with the smallest variance.
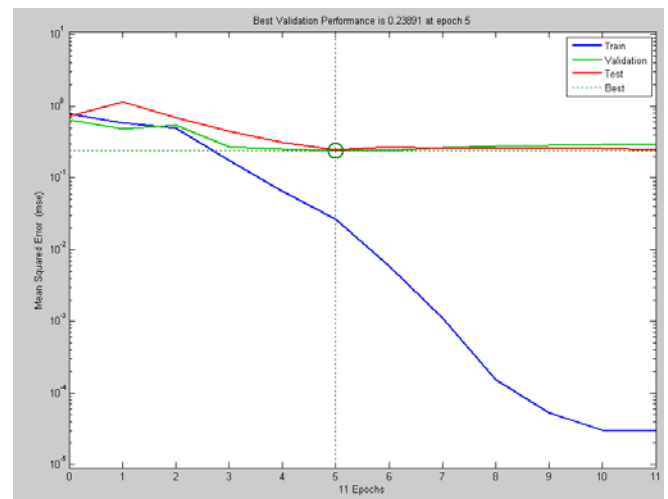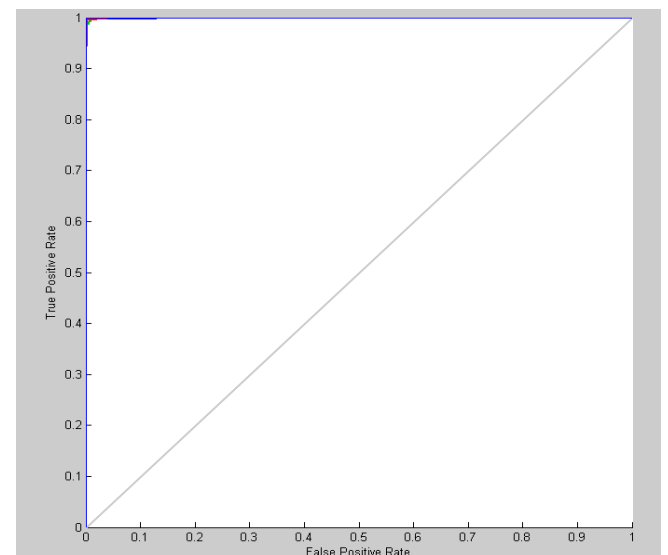

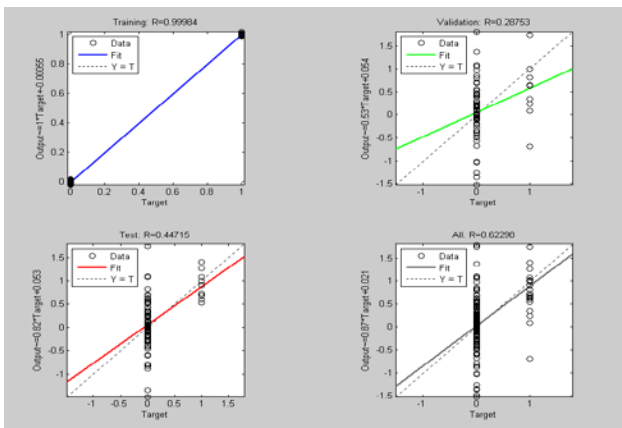Figure 7: Performance plot

Figure 8: ROC plot
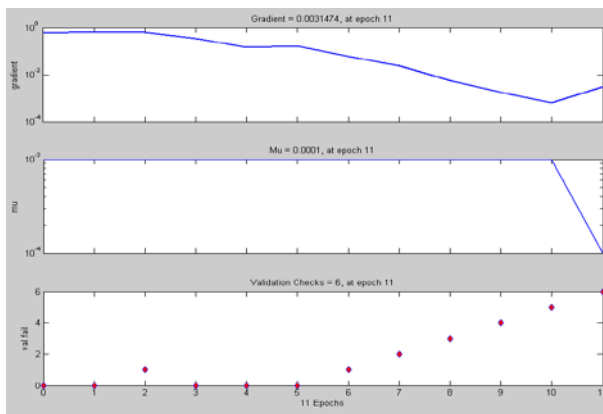
Figure 9: Regression plot

Figure 10: Training State Plot

VII. CONCLUSION

In this article a robust and novel automatic sign language recognition system has been presented, from the transformation of chain code of the outline by HMM until Fisher score, related to the forward-backward probabilities into HMMs. System is improved with the decreasing of the variance of the data with respect to the initial approach of the kernel of fisher, arriving to a rate of 99.30 % of success. The error ratio classifying hand text of the proposed system is less than 0.8% with our database. The objective of this effort was to explore the utility of a neural network based approach to recognition of the sign language. The major goal of this research is to develop a system that will aid in the interaction between human and computer through the use of sign language as a text or control commands.

VIII. REFERENCES

[1] Christopher J. C. Burges, "A Tutorial on Support Vector Machines for Pattern Recognition", Data Mining and Knowledge Discovery, vol. 2, number 2, pp, 121-167, 1998.

[2] Lawence R. Rabiner, "A tutorial on Hidden Markov models and Selected Applications in Speech Recognition", in Proceedings of the IEEE, vol. 77, no. 2,IEEE, pp. 257-286, 1989.

[3] L. O'Gorman and R. Kasturi, Document Image AnalJsis, IEEE Computer Society Press, 1995.

[4] J. Serra, Image Analysis and Mathematical Morphology, Academic Press, 1982.

[5] C. Travieso, C. Morales, I. Alonso y M. Ferrer, "Handwritten digits parameterisation for HMM based recognition", Proceedings of the Image Processing and its Applications,vol.2, pp. 770-774, julio de 1999.

[6] E. Gomez, C.M. Travieso, J.C. Briceiio, M.A. Ferrer, "Biometric Identification Svstem by Lip Shape", in Proceeding of 361h International Carnahan Conference on Security Technology, Atlantic City, October 2002,pp. 39-42.

[7] Carlos M. Travieso, Jesus B. Alonso, Miguel A. Ferrer,"Sign language to text by SVM", 0-7803-7946-2/03, IEEE , 2003.

[8] [Bauer & Hienz, 2000] Relevant feature for video-based continuous sign language recognition. Department of Technical Computer Science, Aachen University of Technology, Aachen, Germany, 2000.pages 440 – 445.

[9] Starner, Weaver & Pentland, 1998] Real-time American Sign Language recognition using a desk- and wearable computer-based video. In proceedings IEEE transactions on Pattern Analysis and Machine Intelligence, 1998, pages 1371-1375.

[10] L. Fausett, "Fundamentals of Neural Networks, Architectures, Algorithms, and Applications", Prentice-Hall, Inc. 1994, pp-304-315.

[11] K.Murakami, H.Taguchi: Gesture Recognition using Recurrent Neural Networks. In CHI '91 Conference Proceedings (pp. 237-242). ACM. 1991.

[12] G. Joshi, and J. Sivaswamy, "A simple scheme for contour detection", 2004 pages 236-242.

[13] [Bowden & Sarhadi, 2000] Building temporal models for gesture recognition. In proceedings British Machine Vision Conference, 2000, pages 32-41.

[14] J.Maclean, R.Herpers, C.Pantofaru, L.Wood, K.Derpanis, D.Topalovic,J.Tsotsos, ―Fast hand gesture recognition for real-time teleconferencing application‖, Proc. Of the IEEE ICCV Worksop on Recognition, Analysis and Tracking of Faces and Gesture in Real-Time Systems, RATFE-RTS 2001, pp. 133-144, 2001

[15] Chang, J. Chen, W. Tai, and C. Han, ―New Approach for Static Gesture Recognition", Journal of Information Science and Engineering22, 1047-1057, 2006.

[16] S. Naidoo, C. Omlin and M. Glaser, "Vision-Based Static Hand Gesture Recognition Using Support Vector Machines", 1998.pages 88 – 94.

[17] J. Driesch, and C. von der Malsburg, "Robust classification of hand postures against complex backgrounds", Proc. Second International Conference on Automatic Face and Gesture Recognition, pages 170 – 175.

[18] Gallaudet University Press (2004). 1,000 Signs of life: basic ASL for everyday conversation. Gallaudet University Press.http://venturebeat.com

[19] Hans-Jeorg Bullinger (1998). Human-Computer Interaction.www.engadget.com.

[20] G. Awcock, and R. Thomas, "Applied Image Processing", 1995, pp-1281 - 1284 Acmillan press LTD, London.

[21] Vladimir I. Pavlovic, Rajeev Sharma, Thomas S Huang, ―Visual Interpretation of Hand Gestures for Human-

Computer Interaction: A review‖ IEEE Transactions of pattern analysis and machine intelligence, Vole 19, NO 7, July 1997, 677 - 695

[22] Visual gesture recognition. In proceedings IEEE Visual Image Signal Process, by Davis & Shah, 1994 vol.141,Issue:2, 101 – 106.

[23] Hand gesture Recognition of Radial Basis Functions (RBF) Networks and Decision Trees ―International Journal of Pattern Recognition and Artificial Intelligence. Volume: 11, Issue: 6(1997) pp. 845-850.

[24] Face detection in complicated backgrounds and different illumination conditions by using YCbCr color space and neural network Pattern Recognition Letters, Volume 28, and Issue 16, 1 December 2007, Pages 2190-2200.