SIGN LANGUAGE TO TEXT BY SVM

Carlos M. Travieso, Jesús B. Alonso, Miguel A. Ferrer
Dpto. de Señales y Comunicaciones, Universidad de Las Palmas de Gran Canaria
Campus de Tafira, 35017 Las Palmas de Gran Canaria, Spain
Phone: +34 928 452 864 Fax: +34 928 451 243 e-mail: travieso@dsc.ulpgc.es

ABSTRACT

In this paper is presented an automatic deaf language to text system. The scheme is based on support vector machines (SVM) classifier using a gaussian kernel. The input parameter vector to SVM 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 letter of the proposed system is less than 0.4% with our database.

1 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 spanish 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 spanish hand sign, and in their output the letter of the latin alphabet that corresponds it, just as it is observed in the following figure, including the Ñ spanish letter.

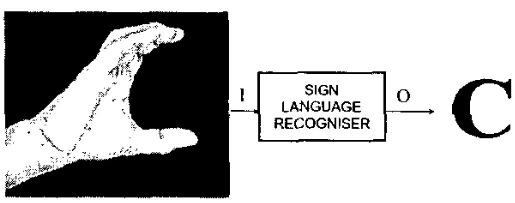

Figure 1. Translation from sign alphabet automatic recognition system, until the standard latin alphabet.

The used procedure in this automatic recognition is based in the adequate modelled of the hand sign by SVM [2]. For it, it has been realised the calculation of Fisher score, which is extracted of HMM [3]. This HMM is fed with the chain code, determined from the hand sign image.

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 alphabet of numerous persons, being the ideal infinite. In our case, it was enclosed for fifty different people (25 hand signs each one). Subsequently, it has been applied a image pre-processing, transforming the colour 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 [4][6], and it is classified and parameterised 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 [1], for classifying it with SVM. This system will be 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 figure;

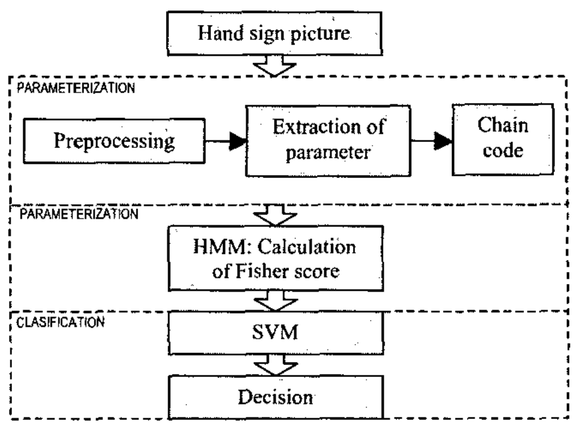

Figure 2. Procedure of automatic hand sign recognition system

The continued structure in this article is the one that continues. 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 SVM classifier The sixth section are all the realized experiments and finally the conclusions and references.

2 BUILDING OF DATABASE

A database has been created for the execution of this paper, that is enclosed of 25 hand signs or patterns, collecting 50 samples of each one, all of different peoples. These images have been captured by a digital photograph camera.

0-7803-7946-2/03/$17.00 ©2003 IEEE

The creation of this database has had a series of difficulties, among them is emphasised the communications barrier of working with people with auditory disability. By another side, it has been worked with images static, without movement, since the sign alphabet has some hand signs that are carried out movement, as CH, H, Ñ, V, X, Y and Z.

The solutions contributed have been two, the first is abandoned the movement as CH, H, X, and Y. The second solution is eliminated some hand signs, because there is some couples very similar with and without movement, as I-Z, N-Ñ and U-V. In this case, it has been eliminated the Z, Ñ and V. The use the movement in the pictures is left for a future language processor.

Some statistic data from database are the following:

- A 74% of males and a 26% of female.
- The 30% of the peoples are deaf and the remainder not.
- The distribution in ages varies since the 20 to the 60 years old.
- The images have a size of 312×232 pixels (wide×high), with a resolution 24 bits of color, in a format RBG, 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.

3 IMAGE PRE-PROCCESING

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 50 pixels, conserving the relation with respect the width. The following steps are;

1. Convert RGB image in grayscale by eliminating the hue and saturation.
2. Enhance contrast using histogram equalization [4]. To realize this process, it is equalized the histogram of the different levels of gray, under a lineal function

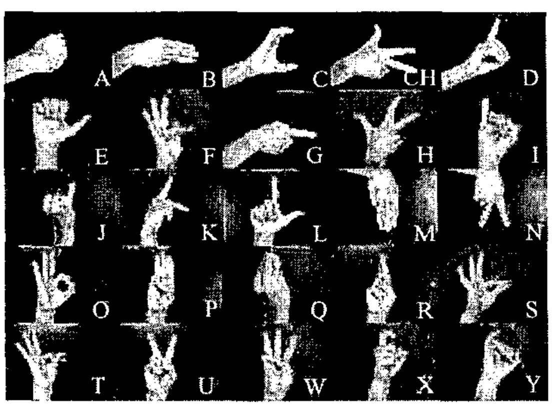

Figure 3. Hand signs that conform the database.

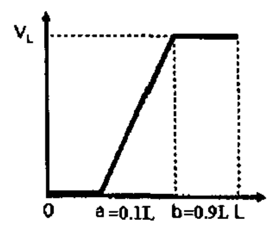

Figure 4. Applied equalization to the intensity.

(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.

3. Take the frame out to eliminate some border effects (see figure 5).
4. 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.
5. Morphologic operators [5]. 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.
6. Reduction. The size of image is reduced in one quarter, for reducing the size of the final vector.
7. Calculation of the contour [6][7]. 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.
8. Adjustment of the wrist. The image is trimmed

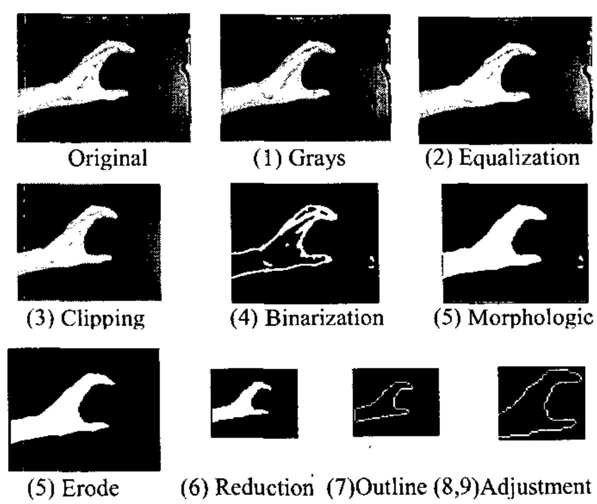

Figure 5. Steps of the applied processing image.

slightly with the intention to determine the side of the one that the hand sign arises.

9. 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.

In the figure 5, this process of processing image is shown, in particular, for the hand sign of the letter C.

4 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;

- Extraction of parameter from outline: chain code [4].
- Creation HMM with chain code as input [6][7].
- Calculation of Fisher score from gradient of logarithm of the observation symbol probability distribution [1][3].

4.1 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 ('M' and 'N').

1	2	3
8	X	4
7	6	5

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

4.2 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 [3]. 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.

4.3 Fisher score

Finally, it is proposed the transformation that provides the HMM probabilities relating to the approach of the Fisher score [1]. With this goal, it intends to unite the probability given by the HMM to the given discrimination of kernel of the SVM, 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 [1];

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

If it is realized the derivade 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 [1];

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

where in [1], it has been found the approximations and the calculation of above equation.

Besides, $\xi(x,q)$ represents the number of times, that is localized in a state q, during the generation of a sequence, emitting a certain symbol x [1][3]. And $\xi(q)$ represents the number of times that has been in q during the process of generation of the sequence [1][3]. These values are obtained directly and of form effective, from the forward backward algorithm, applied to the HMM [3].

The application of this score (U_X) to the SVM, comes given by the expression of the equation 2, utilizing the techniques of the natural gradient, from the following equation [2];

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

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

5 SUPPORT VECTOR MACHINES (SVM)

In the SVM, it is calculated the separate between patterns, by means of the calculation of the natural distance among the score of two sequences X and Y [2];

$$D^2(X,Y) = \frac{1}{2}(U_X - U_Y)^T F^{-1}(U_X - U_Y) \qquad \text{(Eq. 4)}$$

and where F is the matrix of information of Fisher, and is equivalent to the matrix of covariance of the vectors U_X and U_Y.

Finally, different types of functions can be utilized with for classify the Fisher kernel, particularly we are going to work with a lineal and gaussian kernel, although this last produces better discriminations among pattern, being shown this kernel in the following equation [2];

$$K(X,Y) = e^{-D^2(X,Y)} \qquad \text{(Eq. 5)}$$

6 REALIZED EXPERIMENTS

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;

Samples	Number of state				
Training	20	45	55	65	100
20%	71.64% ± 21.81	82.58% ± 2.32	83.72% ± 1.46	82.36% ± 6.65	60.2% ± 45.64
40%	74.1% ± 11.85	85.65% ± 2.64	87.43% ± 6.22	87.77% ± 0.82	68.1% ± 25.2
60%	75.76% ± 9.2	87.8% ± 1.95	88.32% ± 3.24	87.84% ± 3.76	64.08% ± 24.86
80%	75.32% ± 31.41	89.6% ± 13.17	90.18% ± 7.25	89.48% ± 13.14	70.64% ± 15.97

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

Of the above table is deduced that the best percentage of training is for 80% and 55 states, with a rate of 90.18%, presenting a small variance. Of this model is going to generate the kernel of Fisher, and it is going to apply to the SVM, so much with a lineal function as with a gaussian function (RBF), with a trade-off between training error of 10; and in the RBF kernel a variable gamma (g). The results of these tests are collected in the following table;

Type of SVM	Parameter	Average±Variance
Lineal	---------	94.56% ± 11.65
RBF	$g=1.28\cdot10^{-14}$	92.08% ± 3.15
RBF	$g=6.28\cdot10^{-16}$	97.52% ± 1.87
RBF	$g=8.28\cdot10^{-16}$	98.34% ± 0.37
RBF	$g=1.28\cdot10^{-17}$	97.36% ± 2.13
RBF	$g=1.28\cdot10^{-18}$	93.44% ± 1.67

Table 2. Rate of success of the kernel of Fisher modelled with lineal and gaussian (RBF) SVM .

It is observed as for the RBF are presented better results, arriving to a averaged rate of 98.34%, with the smallest variance. Finally, it has been raised to decrease the variance of the obtained data through the kernel of Fisher, eliminating the contribution of F, given in the equation 4, because this factor F disperses the data of the different patterns, in this way the rates of calculated success are;

Type of SVM	Parameter	Average±Variance
Lineal	---------	93.6% ± 12.8
RBF	$g=1.28\cdot10^{-8}$	98.32% ± 0.11
RBF	$g=8.28\cdot10^{-9}$	99.2% ± 0.16
RBF	$g=1.28\cdot10^{-9}$	99.68% ± 0.11
RBF	$g=8.28\cdot10^{-10}$	99.56% ± 0.33
RBF	$g=1.28\cdot10^{-10}$	97.84% ± 3.25

Table 3. Rate of success of the kernel of Fisher modified, classified them with SVM.

The improvement with respect to the table 2 from this table 3 is very significant, giving a better result the RBF when the variance of the data is decreased, achieving a rate of success of the 99.68%. Too, the variance of the rates is minor, until a value of 0.11.

7 SUMMARY

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. Besides this 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 99.68% of success.

8 REFERENCES

[1] Jaakkola, T., Diekhans, M., and Haussler, D. (1998). *"A discriminative framework for detecting remote protein homologies"*. Unpublished, available from http://www.cse.ucsc.edu/ research/compbio/research.html

[2] 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.

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

[4] L. O'Gorman and R. Kasturi, *Document Image Analysis*, IEEE Computer Society Press, 1995.

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

[6] 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.

[7] E. Gomez, C.M. Travieso, J.C. Briceño, M.A. Ferrer, *"Biometric Identification System by Lip Shape"*, in Proceeding of 36th International Carnahan Conference on Security Technology, Atlantic City, October 2002,pp. 39-42.