International Journal on New Computer Architectures and Their Applications (IJNCAA) 2(3): 481-487
The Society of Digital Information and Wireless Communications, 2011 (ISSN 2220-9085)

FACE EXPRESSION RECOGNITION USING AUTOREGRESSIVE MODELS TO TRAIN NEURAL NETWORK CLASSIFIERS

M. Saaidia[1], A. Gattal[1], M. Maamri[1] and M. Ramdani[2]
[1]Dept. of electrical Engineering, University of Tebessa. Algeria. [2]University of Annaba. Algeria.
[1]{msaaidia, a.gattal, m.maamri}@mail.univ-tebessa.dz, [2]mes_ramdani@yahoo.com

ABSTRACT

Neural network classifying method is used in this work to perform facial expression recognition. The processed expressions were the six most pertinent facial expressions and the neutral one. This operation was implemented in three steps. First, a neural network, trained using Zernike moments, was applied to the set of the well known Yale and JAFFE database images to perform face detection. In the second step, Auto Regressive modeling (AR) using 2D- Burg and Levinson filters was used for facial parameterization. At the last step, neural networks, trained on a set of the AR models, were applied to the rest of the images models to test method's performances and to compare the efficiency of the model's representation (Burg and Levinson).

KEY WORDS
Image processing, Facial expression recognition, Face detection, Autoregressive modeling , Neural networks.

1 INTRODUCTION

The initial works on the human facial expression phenomenon were initiated by psychologists who have studied its individual and social importance. They showed that it plays an essential role in coordinating human conversation [1] through the multitude of information it carries. Moreover, Mehrabian [2] found that, while overall impact of the text content of a message is limited to only 7% and the intonation of the speaker's voice contributes by 38%, the facial expressions carry the most part of the message's information i.e. 55%. The recognition of any facial expression is linked to several semantic notions that make the problem difficult to manage given the relativism that it generates in terms of solutions found. Thus, it is quickly pointed out to distinguish between "expression" and "emotion". Indeed, the latter term represents only a semantic interpretation of the first one as "happy" to "smile". A facial expression may be the result of an emotion or not (expression simulated for example). So, a facial expression is a physiological activity of one or more parts of the face (eyes, nose, mouth, eyebrows,...) while an emotion is our semantic interpretation of this activity. However, given the difficulties still encountered in this area we can ignore this distinction.

The significant advances in several related areas such as image processing, pattern recognition, detection and face recognition have to come out studies of this phenomenon from the field of human psychology to the applied sciences domain such as analysis, classification, synthesis, and even the expressive animation control [3].

Different works that have been conducted to date are mostly oriented to the study and classification of the six so-called basic facial expressions (universally recognized): Smile, disgust, fear, surprise, anger and sadness. A multitude of methods which were developed, can be classified according to the parameterization step in the recognition process or to the classification one [4]. According to the first step, methods are "based motion extraction" [5], [6] or "based deformation extraction" [7], [8]. According to the classification step, methods can be "spatial methods"[9], [10], or "spatiotemporal methods" [11], [12]. Method proposed here, is a "spatial model based motion extraction" one.

Section 2, of this manuscript, contains an introduction to the face detection method and the modeling methods used to perform parameterisation. In section 3 we present the neural network classifier and the way to proceed. Section 4 contains the experiments carried out, the results obtained and the performances comparaison. Conclusions are given in section 5.

2 FACE DETECTION AND MODELISATION

Face expression recognition will be done on different types of information supports like images with single face, multi-face images, video, etc. Abstracting the semantic information, processed by human brain; a face in an image remains a common object with specific geometric and color characteristics. Thus, a direct expression processing will be obsolete and pre-processing operations have to be conducted.

- First, we need to isolate the target which will be subject to the expression processing ("face"). This will be done by performing face detection pre-processing operation.
- Secondly, dimensionality problem [13] rises when we try to directly process the delimited face. So we have to find an alternative representation of the face, other than the matrix of pixels, and which size is more reduced.

2.1 Face detection

To perform the first pre-processing operation, we found that several methods were developed to perform face detection [1], [14],… In this work, a NN trained with Zernike moments [13] is used to accomplish this process. The advantage of this method is the fact that it gives accurate faces contours which are well adapted to their shapes. Figure 1 gives some examples of the results given by this method.

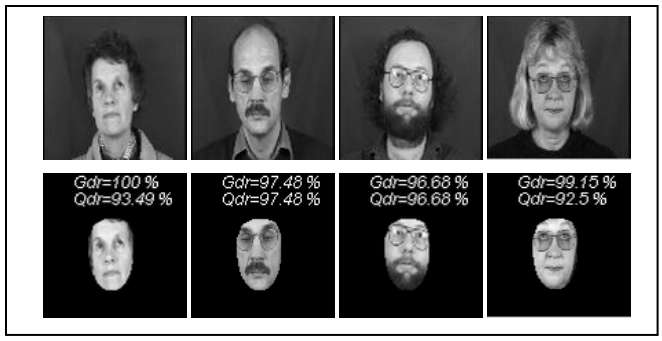

Figure 1: Face detection using NN and Zernike moment Top: original images; Bottom: detected faces

To implement it, we use the fast algorithm developed by G. Amayeh et all [15] and given in (1) for face characterization through Zernike moments and a trained back-propagation neural network for the classification step.

$$Z_{n,m}=\frac{n+1}{\pi}\sum_{x^2+y^2\le 1}\sum\left(\sum_{k=|m|}^{n}\beta_{n,m,k}.\rho^{k}\right)e^{-j.m.\theta}.f(x_j,y_i)$$

$$=\frac{n+1}{\pi}\sum_{k=|m|}^{n}\beta_{n,m,k}.\left(\sum_{x^2+y^2\le 1}\sum e^{-j.m.\theta}\cdot\rho^{k}.f(x_j,y_i)\right)$$

$$=\frac{n+1}{\pi}\sum_{k=|m|}^{n}\beta_{n,m,k}.X_{m,k} \qquad (1)$$

2.2 Face modelisation

The second pre-processing operation is performed in the goal of resolving dimensionality problem mentioned above.

To do so, we propose here to use autoregressive modeling to achieve the image characterization. In spite of being well known and very experienced in the different areas of the signal processing domain, this type of processing was never used for characterization of images in the goal to achieve this type of classification.

Our idea is that for each image or a part of image, we can find a unique model which represents the system producing that image in response to a source of excitation. So, instead of classifying images, we will proceed to the classification of the models which are supposed to be the producers. This will give a vector of parameters which dimensions are much reduced compared to those of the original image. This fact is the goal of the characterization step in all pattern recognition problems.

To do so, we have chosen to experiment the two well known 2D- Burg and Levinson AR models studied and enhanced by [16]. Burg algorithm is known for its simplicity that permits the model parameters estimation without need to calculate the

covariance matrix, and its efficiency when applied correctly with a suitable choice of the parameters' vector size. Despite its complexity, the Levinson one is known for its efficiency and precise results.

The 2D model represents the recursive solution to the mathematical problem posed through the equations:

$$X(k) = -\sum_{l=1}^{n1} A_l^{n1} . X(k-l).W(k) \qquad (2)$$

This model yields also the Multichannel Normal Equations, known as Multichannel Yule-Walker Equations:

$$A_{n1,n2}.R_{n1,n2} = r_{n1,n2} \qquad (3)$$

With:

$$A_{n1,n2} = [A_1^{n1}, A_2^{n1}, \dots, A_{n2}^{n1}]$$

$$R_{n1,n2} = \begin{bmatrix} \begin{bmatrix} R_0 & \cdots & R_{n1-1} \\ \vdots & \ddots & \vdots \\ R_{1-n2} & \cdots & R_0 \end{bmatrix} \end{bmatrix}$$

$$R_k = E_k(X(n+k).X^H(n), n = 0, \mp 1, \dots, \mp n1$$

$$r_{n1,n2} = [R_1, R_2, \dots, R_{n1}]$$

X is the signal to be processed, and E_k is the prediction error.

Resolving (3), will be done using the Levinson algorithm (equation (4)) or the enhanced Burg algorithm (equation (5))

2D Levinson algorithm:

Starting from $n > 0$, with the initial condition $P_0 = R_0$, and using a recursive process we can obtain the coefficients matrices according to:

$$A_n^n = \Delta_n (P_{n-1})^{-1} \qquad (4)$$

With:

$$\Delta_n = R_n + \sum_{l=1}^{n-1} A_l^{n-1} . R_{n-1}$$

$$P_{n-1} = R_0 + \sum_{l=1}^{n-1} J.(A_l^{n-1})^*.J.R_1$$

Mij*=Mji and *J* is the Exchange Matrix given by:

$$J = \begin{bmatrix} 0 & . & . & . & 0 & 1 \\ . & . & . & 0 & 1 & 0 \\ . & . & . & . & 0 & . \\ . & . & . & . & . & . \\ 0 & 1 & 0 & . & . & . \\ 1 & 0 & . & . & . & 0 \end{bmatrix}$$

2D Burg algorithm:

The following set of equations (Equation 5 to Equation 7) gives the mathematical way to implement this type of modeling.

Let x(k1,k2) a size-limited 2D signal;

$$x(k1,k2); k1 = 0,..., N_1 - 1, k2 = 0,..., N_2 - 1 \qquad (5)$$

Backward and forward errors estimation can be written as follows;

$$\begin{aligned} & e_n^b(-1)z = 0; e_n^f(N_1 + n) = 0 \\ & e_{n+1}^f(k) = e_n^f(k) + A_{n+1}^{n+1}.e_n^b(k-1) \\ & e_{n+1}^b(k) = e_n^b(k-1) + J.(A_{n+1}^{n+1})^*.J.e_n^f(k) \end{aligned} \qquad (6)$$

Where: k Є [0,N_1+n],

The matrix parameters are compiled using the relation (7) below;

$$A_{n+1}^{n+1} = -\left[\sum_{k=1}^{N_1+n} e_n^f(k).\left(e_n^b(k-1)\right)^H \right] . \left[\sum_{k=0}^{N_1+n} e_n^b(k).\left(e_n^b(k)\right)^H \right]^{-1} \qquad (7)$$

Implementing these equations and applying them on an image, gives us a reduced–size matrix of parameters which represents the filter model supposed to be the generator of the treated image. These parameters are then the new face characterization vector, on which the classification step will be done.

3 Neural Network Classifier's Implementation

It is clear that the implementation of our method is mainly based on training phase which we

summarize here for first and second pre-processing operations

3.1 Face detection

It is accomplished in four stages:

- Computation of the vectors of Zernike moments for all the images (N) in the work database.
- Construction of the training database by randomly pulling up M images from the work database (M<<N) and their corresponding Zernike moments vectors Zi.
- Manual delimitation of the face area in each image of the training database by a set of points representing the contour Ci of each treated face.
- Training of the neural network on the set of M couples (Zi,Ci).

To test and measure the performances of the network obtained after training operation, we proceed, according to Figure 2, on the hole (N-M) images remaining in the work database.

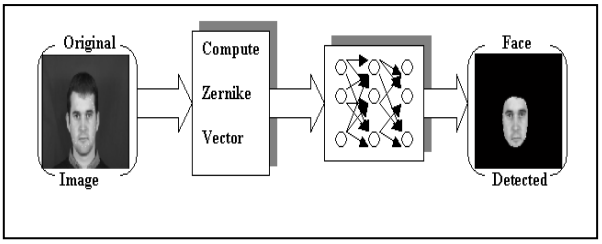

Figure 2: Face detection system

The operation of face detection is thus realized in two steps:

- During the first step, an image is presented to an algorithm which extracts the representative Zernike vector.
- At the second phase, a back-propagation neural network, beforehand trained, receives on its input layer the Zernike moments vector. Then, on its output layer, the neural network gives a set of points representing the probable contour of the face contained in the original image.

The neural network is used to extract statistical information contained in the Zernike moments and in their interactions which are closely related to the area of the required face.

3.2 Expression recognition

It is achieved in four stages:

- Computation of Levinson or Burg matrix for all the detected faces (N) in the work database.
- Construction of the training database by randomly pulling up MM detected faces from the work database (MM<<N) and their corresponding Levinson or Burg matrices A_{n+1}^{n+1}.
- Manual construction of the target matrix T used as the predefined response of the neural network to the MM training faces.
- Training of the neural network on the set of MM couples (A_{n+1}^{n+1}, T).

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

Expression recognition will be also done, according to Figure 3, in two steps:

- During the first step, a Levinson or Burg matrix is compiled for the detected face for which expression recognition will be performed.
- At the second step, the back-propagation neural network, beforehand trained, receives on its input layer the Levinson or Burg matrix. Then, on its output layer, the neural network gives a probabilistic vector for expressions subject to recognition.

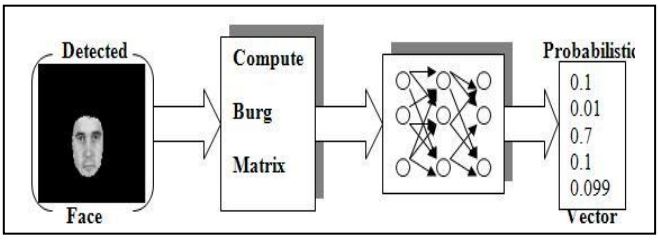

Figure 3: Face expression recognition system

4 EXPERIMENTAL RESULTS

In order to check the validity of our proposed method, experimental studies were carried out on the well known Yale and JAFFE images databases [17][18]. Yale database contains **4** recordings of **15** subjects taken for three different expressions (Happy, Sad and Surprise) and the neutral expression. Instead, JAFFE database contains only female subjects with the six well known and most studied expressions (Happy, Fear, Sad, Surprise, Disgust and Anger) in addition to the neutral expression. Figure 4 and figure 5 give examples of images with different expressions from the two databases.

Figure 4: Example of two subjects from Yale database with three different expressions for each one. Neutral, Sad, surprised and Happy

Figure 5: Example of two subjects from JAFFE database with three different expressions for each one. Up: Neutral, Disgusted, Afraid and Happy. Down: Neutral, Sad, surprised and Angry

Firstly, the efficiency of the two modelling algorithms (Burg and Levinson) was compared on the JAFFE database. Secondly, experiences were carried out separately on each database for Burg algorithm and then on mixed database containing images with the four common expressions (Neutral, Happy, Sad and Surprise).

To obtain the training database for Yale images we have take randomly **10** images of different people, each one with **4** different recordings, so that it gives us **40** couples (**Zi**,**Ci**) and (A_{n+1}^{n+1}, T) examples for training the neural networks. For JAFFE database we took randomly 2 images for each person with each expression so we obtain a training database with 140 couples (**Zi**,**Ci**) and (A_{n+1}^{n+1}, T) examples.

For the mixed training database, we took all the already used Yale training examples and their corresponding examples in JAFFE database which give us **120** training couples.

Obtained Results, for each experience were given respectively in tables Table1, Table2, Table3 and Table4.

4.1 Burg and Levinson performances comparison

The first experience was carried out on the JAFFE database to compare the efficiency of Levinson and Burg algorithms. The comparison results were reported on Table1.

Table1: Comparison results for Levinson and Burg algorithms on JAFFE database

	Expr	Neu.	Sad	Surp.	Hap.	Fear	Dis.	Ang.
Burg	TPR %	**90**	**90**	**90**	**81.81**	**90**	**77.78**	**70**
	FPR %	**10**	**10**	**10**	**18.19**	**10**	**22.22**	**30**
	Time	**1**	**1**	**1**	**1**	**1**	**1**	**1**
Levinson	TPR %	**90**	**81.81**	**90**	**81.81**	**90**	**77.78**	**81.81**
	FPR %	**10**	**18.19**	**10**	**18.19**	**10**	**22.22**	**18.19**
	Time	**70**	**90**	**75**	**80**	**70**	**50**	**80**

Performances comparison were recorded using TPR (True Positive Rate) and FPR (False Positive Rate) to measure the efficiency of recognition process and Time to measure the time taken to compile the model parameters vector.

The results demonstrate that the recognition efficiency is comparable for the two algorithms with a slight advantage for the Levinson one. However,

the time consumption parameter gives the advantage to the Burg algorithm which is faster.

4.2 Yale database

After training neural networks to perform initially face detection and expression recognition, we proceed to test the performances of the expression recognition neural network on the rest of images of Yale database. So, the 20 detected faces were pre-processed to obtain AR-model parameters of each face. Model matrices were presented to the inputs of the trained neural network. Obtained results are reported on table2.

Table2: Expression recognition results obtained on Yale database

expression	Neutral	Sad	Surprise	Happy
Neutral	**5**	**1**	**0**	**0**
Sad	**0**	**4**	**0**	**0**
Surprised	**0**	**0**	**4**	**2**
Happy	**0**	**0**	**1**	**3**
TPR %	**100**	**80**	**80**	**60**
FPR %	**0**	**20**	**20**	**40**

Although, there are not a lot of test examples, the results obtained demonstrate the validity of the applied algorithm. Recorded TPR (True Positive Rate) and FPR (False Positive Rate) show that confused decisions were held between Surprised and Happy expressions. This may be due to the way that a person expressed them especially at the mouth region.

4.3 JAFFE database

As it was done with images of Yale database, 73 detected faces were pre-processed to obtain AR-model parameters of each face. Model matrices were presented to the inputs of the trained neural network. The results obtained are reported in table3.

Instead of Yale database, JAFFE one presents more examples and therefore the validation results are more credible. Treated expressions are also more complete.

Table3: Expression recognition results obtained on JAFFE database

Expr	Neu.	Sad	Surp.	Hap.	Fear	Dis.	Ang.
Neu.	**9**	**0**	**0**	**0**	**0**	**0**	**0**
Sad.	**0**	**9**	**0**	**0**	**1**	**1**	**0**
Surp.	**0**	**0**	**9**	**2**	**0**	**0**	**1**
Hap.	**0**	**0**	**1**	**9**	**0**	**0**	**0**
Fear	**0**	**0**	**0**	**0**	**9**	**1**	**1**
Dis.	**0**	**0**	**0**	**0**	**0**	**7**	**1**
Ang.	**1**	**1**	**0**	**0**	**0**	**0**	**7**
TPR %	**90**	**90**	**90**	**81.81**	**90**	**77.78**	**70**
FPR %	**10**	**10**	**10**	**18.19**	**10**	**22.22**	**30**

For the common expressions to the two databases, results obtained are comparable to those reported in table2. Conflict decisions are also done by the trained neural network in the case of couple expressions Happy-Surprised. Confused decisions are especially obtained between Fear, Disgust and Anger expressions

4.4 Mixed database

Neural network trained with 120 images of the mixed database was tested on a set of 62 images (20 images from Yale database and 42 images from the JAFFE database). The results obtained are given in table 4.

Table4: Expression recognition results obtained on mixed (Yale-JAFFE) database

expression	Neutral	Sad	Surprise	Happy
Neutral	**13**	**2**	**0**	**0**
Sad	**2**	**14**	**0**	**0**
Surprise	**0**	**0**	**12**	**4**
Happy	**0**	**0**	**3**	**12**
TPR %	**86.67**	**87.50**	**80.00**	**75.00**
FPR %	**13.33**	**12.50**	**20.00**	**25.00**

Combined database let to worst results for all common expressions. This may be due to the different ways that subjects, of the two databases, express their emotions. Another reason will be the difference in gender and ethnicity of the subjects of the two sets.

5 CONCLUSION

Expression recognition system was proposed. It was implemented in three steps; face detection by training neural network, face modeling according to AR models and expression recognition using trained neural network. The study was especially focused on the second and the third step. Practical study was carried out on the well known Yale and JAFFE databases. Simulation results were compiled on a set of testing examples taken first from each database alone then on a mixed database containing the images with the same expressions. The two well known modeling algorithms of Levinson and Burg were tested and there performances were compared. Obtained results demonstrate the validity of the proposed technique. However, confused decisions were obtained between some expressions especially the couple Happy-Surprise and the triplet, Fear, Disgust and Anger.

Results demonstrate also, that the efficiency of the two algorithms is comparable but the Burg one is more faster.

The study of modeling parameter's influence was started but not yet finished. This will be the continuation of this work.

REFERENCES

1. B. Jedynak, H.C. Zheng and M. Daoudi, "Skin detection using pairwise models", IVC(23), No. 13, 29 November 2005, pp. 1122-1130.
2. N. F. Trose and H. H. Bulthoff, "Face Recognition Under Varying Poses: The Role of Texture and Shape", Elsevier, Vol. 36. No. I, pp. 1761-1771.
3. Sh. Wu, W. Lin and Sh. Xie, "Skin heat transfer model of facial thermograms and its application in face recognition", Elsevier Pattern Recognition, Vol. 41, Issue 8, pp. 2718-2729, August 2008
4. C. Padgett and G. W. Cottrell, "Representing Face Image for Emotion Classification" In M. Mozer, M. Jordan, and T. Petsche, editors, Advances in Neural Information Processing Systems, volume 9, pages 894-900, Cambridge, MA, 1997. MIT Press.
5. Z. Zhang, M. Lyons, M. Schuster and S. Akamatsu, "Comparision between Geometry-based and Gabor-Waveletes- based Facial Expression Recognition Using Multi-layer Perceptron", Proceedings, Third IEEE International Conference on Automatic Face and Gesture Recognition, April 14-16 1998, Nara Japan, IEEE Computer Society, pp. 454-459.
6. W. K. Teo, L. C. De Silva and P. Vadakkepat, "Facial Expression Detection and Recognition", Journal of the Institut of Ingineers, Singapor, Vol. 44, Issue 3, 2004.
7. Y.Tian, T. Kanade, and J. F. Cohn. "Recognizing Action Units for Facial Expression analysis", IEEE Transactions on Pattern Analysis and Machine Intelligence, 23(2), Feb 2001.
8. M. Wang, Y.I. wai, and M. Yachida, "Expression Recognition from Time-Sequential Facial Images by use of Expression Change Model", In IEEE Proceedings of the Second Int. Conf. on Automatic Face and Gesture Recognition, 324-329, Japan, April 14-16 1998.
9. M. Rosenblum, Y. Yacoob, and L. Davis, "Human Expression Recognition from Motion using a Radial Basis Function Network Architecture" IEEE Transactions on Neural Networks, 7(5):1121-1138, 1996.
10. H. Van Kuilenburg, M. Wiering and M. den Uyl, "A Model Based Method for Automatic Facial Expression Recognition", The 16th European Conference on Machine Learning (ECML), pp. 194 – 205, Porto, Portugal, October 3-7, 2005
11. J. L. Crowley and F. Berard, "Multi-model tracking of faces for video communications", in IEEE Int, Conf. on Computer Vision and Pattern Recognition, Puerto Rico, Jun. 1997.
12. R. J. Prokop and A. P. Reeves, "A survey of moment-based techniques for unoccluded object representation and recognition", CVGIP Graphical models and Image Processing, 54(5):pp. 438-460, 1992.
13. M. Saaidia, A. Chaari, S. Lelandais, V. Vigneron and M. Bedda, "Face localization by neural networks trained with Zernike moments and Eigenfaces feature vectors. A comparison", AVSS2007, pp. 377-382, 2007
14. E. Hjelmas and B. K. Low. "Face detection: A survey" Computer Vision and Image Understanding, vol. 83, no. 3, pp. 236-274, 2001.
15. G. Amayeh, A. Erol, G. Bebis, and M. Nicolescu, "Accurate and efficient computation of high order zernike moments", First ISVC, Lake Tahoe, NV, USA, pp. 462–469, 2005.
16. R. Kanhouche, "Méthodes Mathématiques En traitement Du Signal Pour L'estimation Spectrale", Doctorate thesis in Applied Mathematics; Ecole Superieur de Cachan. Dec 2006.
17. P. N. Bellhumer, J. Hespanha, and D. Kriegman, "Eigenfaces vs. fisherfaces: Recognition using class specific linear projection", IEEE Transactions on Pattern Analysis and Machine Intelligence, Special Issue on Face Recognition, 17(7):711--720, 1997.
18. M. J. Lyons, Sh. Akamatsu, M. Kamachi and J. Gyoba, "Coding Facial Expressions with Gabor Wavelets", Proceedings of the Third IEEE International Conference on Automatic Face and Gesture Recognition, April 14-16 1998, Nara Japan, IEEE Computer Society, pp. 200-205.