2014 4th International Conference on Artificial Intelligence with Applications in Engineering and Technology

Facial Expression Recognition Using Neural Network Trained with Zernike Moments

Mohammed Saaidia
Dept. Génie-Electrique
Université M.C.M Souk-Ahras
Souk-Ahras, Algeria
mohamed.saaidia@univ-soukahras.dz

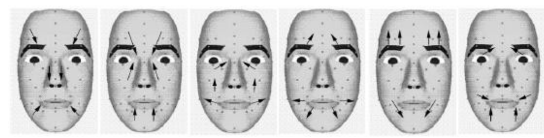
Narima Zermi and Messaoud Ramdani
Dept. Electronique
Université B.M Annaba
Annaba, Algeria
narima.naili@univ-annaba.dz

***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, detected faces are processed to perform the characterization phase through computed vectors of Zernike moments. At last step, a back propagation neural network was trained to distinguish between the seven emotion's states of a presented face. Finally, method performances were evaluated on the well known JAFEE and YALE database.**

Keywords-face detection; face expression recognition; image analysis; patern recognition

I. Introduction

The first works on this human phenomenon were initiated by psychologists who have studied the individual and social importance. They showed that it plays an essential role in coordinating human conversation [1] through the multitudes of information it conveys. Moreover, Mehrabian [2] showed that the textual content of a message is limited to only 7% of its overall impact, while the tone of the speaker's voice participates by 38% and facial expressions by 55 %. Recognition of any facial expression is linked to several semantic concepts that make this problem difficult to manage given the relativism that generates on solutions found. Thus, it is quickly pointed to distinguish between 'expression' and 'emotion'. Indeed, this last term is only a semantic interpretation of the first as the term 'happy' to 'smile'. A facial expression can be the result of an emotion or not (for example simulated expression). So a facial expression is simply a physiological activity of one or more parts of the face (eyes, nose, mouth, eyebrows,...) while an emotion is our semantic interpretation to this activity. However, given the difficulties still encountered in this area, we can still ignore this subtlety. The significant advances in several related fields such as image processing, pattern recognition, detection and face recognition, allowed to take of the studies of this phenomenon from the field of human psychology to the automatic analysis, classification, synthesis, and even expressive animation [3]. T The different works that have been carried out to date were all oriented to the study and classification of facial expressions called basic expressions (universally recognized); six in all and are summarized on Figure 1:



angry disgust happiness fear surprise sadness

Figure 1 : Direction of motion and deformation for each basic facial expression

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.

Another way to perform classification task is the way to characterize the face. Some methods process the face globally although some other methods extract face futures before performing characterization.

So, the recognition of facial expression can be approached in several ways. In this work, we propose to exploit the geometrical moments, especially those of Zernike, to perform facial expression recognition. This type of characterization was fully used in face recognition [13], [14]. The encouraging results obtained in these works allow to state that if this kind of characterization (by Zernike moments) enables efficient classification in inter-persons classification so we can pretend to better results in intra-persons classification problem like facial expressions recognition.

In second section of this paper we will present and explain the way to perform face detection and characterization. In third section, method implementation is explained. Experimental results are presented and discussed in fourth section. Section five will be reserved to the conclusions and future possible enhancements.

978-1-4799-7910-3/14 $31.00 © 2014 IEEE
DOI 10.1109/ICAIET.2014.39

II. FACE DETECTION AND CHARACTERIZATION

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 [15] 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.

A. Face detection

To perform the first pre-processing operation, we found that several methods were developed to perform face detection [16], [17],… In this work, a NN trained with Zernike moments [15] is used to accomplish this process.

Zernike moments form part of the general theory of the geometrical moments. They were introduced initially by F. Zernike [4]. At the difference of the general geometrical moments, those of Zernike are built on a set of orthogonal polynomials. These polynomials are the basic elements of the construction of an orthogonal base given by the equation (1)

$$V_{n,m}(x,y) = V_{n,m}(\rho,\theta) = R_{n,m}(\rho).e^{j.m.\theta} \qquad (1)$$

where :

$$\begin{cases} R_{n,m}(\rho) = \displaystyle\sum_{k=|m|}^{n} \frac{(-1)^{(n-k)/2}.(n+k)!}{(\frac{n-k}{2})!(\frac{k+m}{2})!(\frac{k-m}{2})!}.\rho^{k} \\ \rho=\sqrt{x^2+y^2} \quad \text{and} \quad \theta=arctg(y/x) \end{cases} \qquad (2)$$

with: $n \geq 0$, $m \neq 0$, $m < n$, $n - m < n$ and $(n-k)$ even.

$R_{n,m}(\rho)$ is the orthogonal radial polynomial, n is the order of the moment and m the factor of repetition (the smoothness of the required details) at this order. ρ and θ are respectively the radius and the angle of treated point of the function.

To implement it, we use the fast algorithm developed by G. Amayeh et all [18] and given in (3) for face characterization through Zernike moments and a trained back-propagation neural network for the classification step.

$$\begin{aligned} Z_{n,m} &= \frac{n+1}{\pi} \sum_{x^2+y^2 \le 1} \sum \left(\sum_{k=|m|}^{n} \beta_{n,m,k}.\rho^{k} \right) e^{-j.m.\theta}.f(x_j,y_i) \\ &= \frac{n+1}{\pi} \sum_{k=|m|}^{n} \beta_{n,m,k}.\left(\sum \sum_{x^2+y^2 \le 1} e^{-j.m.\theta}.\rho^{k}.f(x_j,y_i) \right) \\ &= \frac{n+1}{\pi} \sum_{k=|m|}^{n} \beta_{n,m,k}.X_{m,k} \end{aligned} \qquad (3)$$

The advantage of this method is the fact that it gives accurate faces contours which are well adapted to their shapes. Figure 2 gives some examples of obtained results.

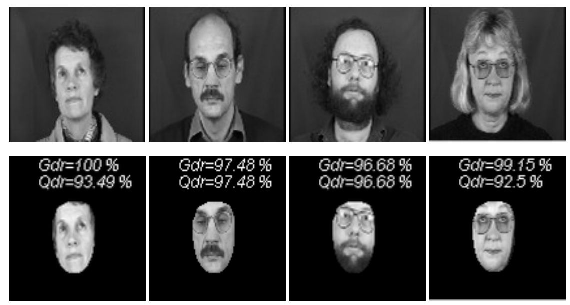


Figure 2: Face detection using NN and Zernike moment Top: original images; Bottom: detected faces.

B. Face characterization

One of main Characteristics known for Zernike moments is that they can compress the geometric information of the image into a vector of reduced dimensions depending on the parameters *m* and *n*.

Secondly, they allow a meaningful representation of the information contained on the face such as the surface, the vertical symmetry and distribution centers masses in the horizontal and vertical directions and other face characteristics which deals with information required for such type of classification problem.

In order to study the performances of the proposed method, experiences were conducted on two standard database of expressive faces namely Yale University database and JAFEE database. The Figure 3 gives representative images from these two database.

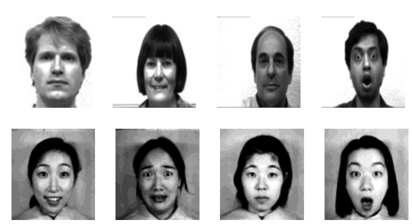


Figure 3: Sample pictures of faces from the expressive database of Yale university.

We therefore performed, to all the faces of the database, two kinds of segmentation. The first, manual, to evaluate the proposed method of classification and the second, through automatic face detection by location method proposed previously. This last way was intended to evaluate the classifier response to the limited faces which are not necessarily ideal and may contain degradation and occlusion on the areas necessary for the recognition process.

Figure 4 gives an example of two feature vectors for the same image manually then automatically processed.

For both sets of faces, we compute feature vectors with different values of parameters *m* and *n* with controls successively the smoothness of the representation and the length of the feature vectors.

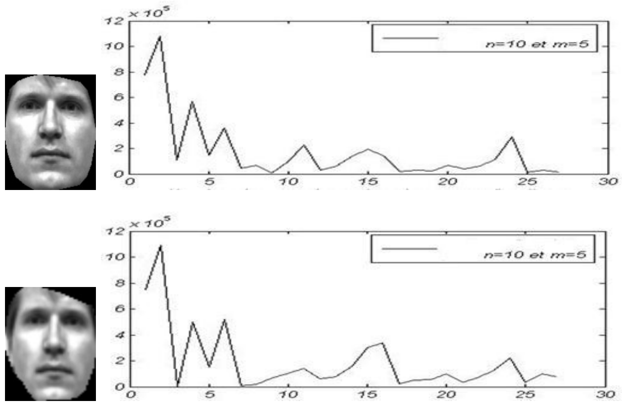


Figure 4: Sample pictures of faces from the expressive database of Yale university.

On figure5, we give the curves computed for three different expressions with two couples of (n,m) values.

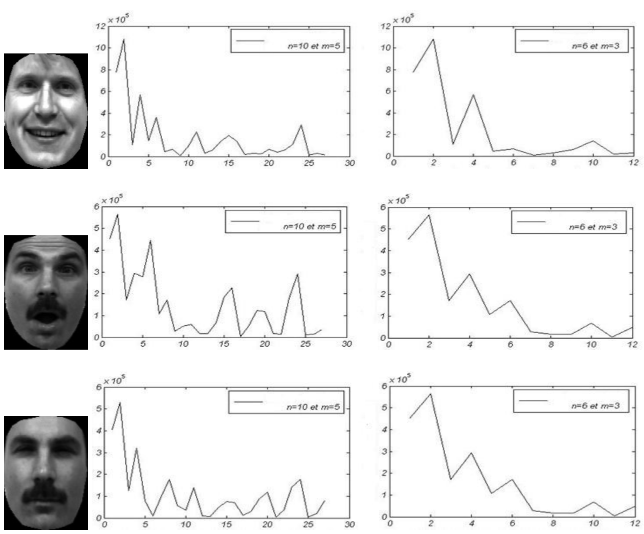


Figure 5: Tree different expressions of two different images with two couples of (n,m) values (a) : n=10 et m=5, (b) : n=6 et m=3.

We have also computed difference’s vector between feature vectors of the same person with different expressions. Figure 6 gives a sample representing differences between neutral and happy expressions of the same person.

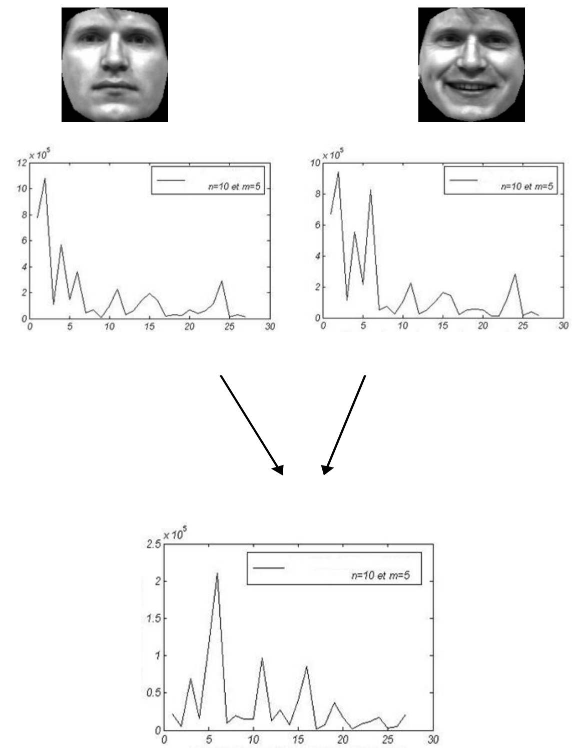


Figure 6: A sample representing differences between feature vectors of neutral and happy expressions of the same person

From these examples we can draw some useful observations:

- There are many apparent differences between the curves representing images of different persons. This justifies the different research that has been conducted on the use of characterization by such attributes for recognizing faces. [13], [14], [19], ...
- Differences in curves shapes associated with facial expression tend to become less detectable for small values of *n* and *m*. This Remarque will be commented in the results of different experiments related to influences of the parameters *n* and *m* on the recognition quality.
- The similarities of shape's curves of the same person are apparent but differences related to the expression on the face of that person are also largely highlighted on all the differences vectors computed between feature vectors of different expressions (On Figure 6, we give an example of that evidence). This is what justifies our attempt to use this type of characterization for facial expressions recognition.

III. NEURAL NETWORK CLASSIFIER

The second element of our classification system is the analysis and decision part. In the field of pattern recognition and especially that of the facial expression recognition, several types of classifiers have been proposed and used. It is based on the use of three main theories of classification, namely, the HMM (Hidden Markov Models) [20], [21], the SVM (Support Vector Machines) [22] and neural networks [23], [24]. These three systems can cope with the challenges of real classification problems that are non-linearity, dimensionality and generalization. Neural networks have proven effective in many areas of practice shape recognition such as handwriting recognition, speech analysis or robot control.

A. Neural network structure

For our present work, we choose to use a multilayer back propagation neural network which parameters will be subject to study according to experiences related to the dimensionality of the feature vectors compiled during the characterization step.

The system is formed by three layers of neurons; a hidden layer and two standard input and output layers. The structure used is:

- The number of neurons in the input will be the same as the length of the feature vectors which influences the quality of recognition results. The number of neurons in the hidden layer will also be part of the parameters whose influence will be studied during the different experiences we will present later. The number of neurons of the output layer depends on the number of expressions (number of clusters) the network is required to recognize. In the most case, seven output neurons will be used (6 standard expressions and the neutral one). According to our experimental database, which contains only examples for four facial expressions, we will only use four neurons on the output layer of the network.
- After several tests on different functions for driving the network we have chosen to retain the 'Resilient propagation' function that provides the best results. Moreover, even in the literature, it is the most used in real classification problems.
- For activation functions in the three layers we selected the combination: Linear-logSig-Linear in the case of the module of Zernike moments and the combination: Linear-tanSig-Linear in the case of attributes vectors based on the real and the imaginary of Zernike moments.

B. Implementation of the recognition process

Training process will be achieved in four stages:

- Computation of Zernike moments feature vectors 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 feature vectors
- 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 (*Zn,m* , T).

Testing step will be achieved in two stages:

- During the first stage, a Zernike moments feature vector is compiled for the detected face for which expression recognition will be performed.
- At the second stage, the back-propagation neural network, beforehand trained, receives on its input layer the Zernike moments feature vector. Then, on its output layer, it gives a probabilistic vector for expressions subject to be recognized.

IV. 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 [25][26]. 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 7 and figure 8 give examples of images with different expressions from the two databases.

Experiences were carried out separately on each database. 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 (*Zn,m* , 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 80 couples (Zi,Ci) and (*Zn,m*, T) examples.

Obtained results will be detailed in the following subsections.

Figure 7: Two subjects from Yale database with three different expressions for each one. Neutral, Sad, surprised and Happy.

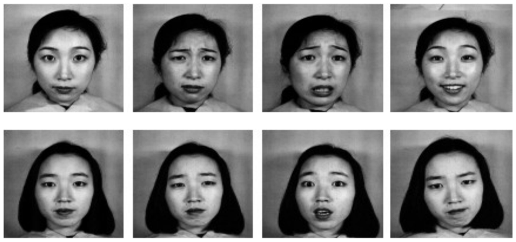


Figure 8: Example of two subjects from JAFFE database with three different expressions for each one. Up: Neutral, Disgusted, Afraid and Happy. Down: Neutral, Sad, surprised and Angry.

A. General results

Table I and table II give the results obtained applaying the previously described method to Yale and JAFFE database with a randomly chosen parameters *m* and *n*.

TABLE **I**. RESULTS OBTAINED FOR ‘YALE’ DATABASE.

real Expression / Detected Expression	***Neutral***	***Hap.***	***Surp.***	***Angry***	***TPR %***	***FPR %***
Neutral	**4**	1	1	2	40	40
Happy	1	**7**	1	1	70	30
Surprise	2	2	**6**	1	60	50
Angry	3	0	2	**6**	60	50
	Global TPR				**57.5**	

TABLE II. RESULTS OBTAINED FOR ‘JAFFE’ DATABASE.

Expression real / Expression Detected	***Neutral***	***Hap.***	***Surp.***	***Angry***	***TPR %***	***FPR %***
Neutral	**8**	0	0	2	80	20
Happy	0	**9**	1	0	90	10
Surprise	0	2	**7**	2	70	40
Angry	1	0	1	**8**	80	20
	Global TPR				**80**	

Results reported in table I and table II, using the the sensitivity measurement (TPR: True Positive Rate) and FPR (False positive Rate), were obtained with parameters n=10 and m=5. Recorded TPR's show that the 'smile' is an expression that seems best represented in the vectors attribute since it is best distinguished (70% for Yale and 90% for JAFFE) and less present in the misclassification of other expressions (30% for Yale and only 10% for JAFFE). Also, the 'smile' is totally distinct from the 'Angry' expression and closer to the expression of 'surprise' than as to the ‘neutral’ one.

Results recorded for JAFFE database are better than those obtained for ‘Yale’ database. This will be due to the homogeneity of the persons (all females) and the greater number of training images.

B. Sensitivity to m and n parameters

As already indicated, the values of parameters *n* and *m* not only influences the number of elements of the feature vector but, more importantly, it influences the discriminative ability between different expressions of the same face. This remark pushed us to study the influence of these parameters on the quality of the classification process. Table III and table IV report the recorded results for different values of the couple (*m*,*n*).

TABLE **III**. RESULTS OBTAINED FOR ‘YALE’ DATABASE.

Expression / Couples (*n,m*)	*Recognition rate %*				Global TPR %
	Neutral	***Hap.***	***Surp.***	***Angry***	
(*3,1*)	10	30	20	10	17.50
(*4,2*)	10	30	20	10	17.50
(*4,3*)	10	30	30	10	20.00
(*5,1*)	10	30	20	20	20.00
(*6,3*)	10	30	30	20	22.50
(*6,5*)	20	40	40	20	30.00
(*9,5*)	40	60	60	50	52.50
(*10,5*)	40	70	60	60	57.50
(*12,5*)	60	70	70	70	67.50
(*12,7*)	60	80	70	70	70.00
(*12,10*)	70	80	70	70	72.50
(*12,11*)	60	80	80	70	72.50
(*15,11*)	60	80	70	60	67.50
(*17,11*)	40	70	70	50	57.50
(*20,11*)	20	60	60	30	42.50

TABLE IV. RESULTS OBTAINED FOR ‘JAFFE’ DATABASE.

Expression / Couples (*n,m*)	Global TPR %				Global TPR %
	Neutral	***Hap.***	***Surp.***	***Angry***	
(*3,1*)	50	50	43	50	48,25
(*4,2*)	50	55	47	53	51.25
(*4,3*)	50	57	51	53	52.75
(*5,1*)	63	63	60	60	61.50
(*6,3*)	70	75	68	69	70.50
(*6,5*)	71	83	68	71	73,25
(*9,5*)	80	90	73	78	80.25
(*10,5*)	80	90	70	80	80.00
(*12,5*)	85	92	73	86	84.00
(*12,7*)	90	92	73	89	86.00
(*12,10*)	92	95	81	90	89.50
(*12,11*)	87	92	78	90	86.75
(*15,11*)	87	92	78	83	85.00
(*17,11*)	71	87	70	78	76.50
(*20,11*)	61	76	63	70	67.50

Recorded results in the two tables allow us to draw some remarks:

- Here also, best results were recorded for JAFFE database. Although, for both database, the best results were obtained for (n,m)=(12,10).

- Detailed results indicate that greater are the (n,m) values, better are the performances of the neural classifier. Although, beyond a certain limit, the increase of the (n,m) values bring performance degradation. This fact is due to the exponentially increase in the neural network classifier complexity which influences the convergence process and then the quality of the classification (see Figure 9).
- For values low and increasing values of the pair (n, m) the performances of the method are ascending. This is a direct consequence of the increase of the discriminating capacity of the feature vectors with their increased size.
- By setting *m* to 5 and increasing the values of *n* or vice versa, by setting *n* to 12 and increasing the values of *m*, we obtain improvement in classifier performances and thus the recognition rate. However, this improvement is much more remarkable in the case of an increase of *n* (On table III: an improvement of 37.5% in the first case and only 5% in the second. On table IV: an improvement of 10.75% in the first case and only 2.75% in the second.).

V. COCLUSION

Facial expression recognition method was proposed. Recognition process was achieved in two principal steps; face detection and face expression recognition. The study was especially focused on the second step. Arguments were presented to justify characterization choices and the way to implement proposed method was detailed. Practical study was carried out on the well known Yale and JAFFE database. Recorded results were presented and commented.

More detailed studies have to be achieved to explore all classifier parameters and to improve method performances.

REFERENCES

[1] Jedynak, B.[Bruno], Zheng, H.C.[Hui-Cheng], Daoudi, M.[Mohamed], “Skin detection using pairwise models”, IVC(23), No. 13, 29, pp. 1122-1130, November 2005.

[2] N. F. TROSE, H. H. BULTHOFF, “Face Recognition Under Varying Poses: The Role of Texture and Shape”, Elsevier, Vol. 36. No. I, pp. 1761-1771.

[3] Shiqian Wu, Weisi Lin, Shoulie 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. Kotropoulos and I. Pitas, Rule-based face detection in frontal views, in Proc. Int. Conf. on Acoustic, Speech and Signal Processing, 1997.

[5] H. Graf, E. Cosatto, and T. Ezzat, “Face analysis for the synthesis of photo-realistic talking heads”, Proc. Fourth IEEE Int. Conf. Automatic Face and Gesture Recognition, Grenoble, France, IEEE Computer Society, Los Alamitos, pp. 189-194, 2000.

[6] J. L. Crowley and F. Berard, Multi-model tracking of faces for video communications, in IEEE Conference on Computer Vision and Pattern Recognition, CVPR '97, Puerto Rico, June 17-19, 1997.

[7] D. Reisfeld and Y. Yeshurun, Preprocessing of face images: Detection of features and pose normalization, Comput. Vision Image Understanding, CVIU, 71(3), pp. 413–430, September 1998.

[8] R. Herpers, H. Kattner, H. Rodax, and G. Sommer, Gaze: An attentive processing strategy to detect and analyze the prominent facial regions, in IEEE Proc. of Int. Workshop on Automatic Face- and Gesture-Recognition,, Switzerland, pp. 214–220, Jun. 1995.

[9] W. Huang, Q. Sun, C. P. Lam, and J. K. Wu, A robust approach to face and eyes detection from images with cluttered background, in Proc. of International Conference on Pattern Recognition, 1998.

[10] T. Yokoyama, Y. Yagi, and M. Yachida, Facial contour extraction model, in IEEE Proc. of 3rd Int. Conf. On Automatic Face and Gesture Recognition, 1998.

[11] A. L. Yuille, P. W. Hallinan, and D. S. Cohen, Feature extraction from faces using deformable templates, Int. J. Comput. Vision 8, ppp. 99–111, 1992.

[12] T. F. Cootes and C. J. Taylor, Active shape models—‘smart snakes,’ in Proc. of British Machine Vision Conference, pp. 266–275, 1992.

[13] Javad Haddadnia, Majid Ahmadi, and Karim Faez,: “An Efficient Feature Extraction Method with Pseudo Zernike Moment in RBF Neural Network-Based Human Face Recognition System”, EURASIP, JASP, 2003, vol. 9, pp. 890-891.

[14] S. Alirezaee, M. Ahmadi, H. Aghaeinia, and K. Faez, “A weighted pseudo-zernike feature extractor for face recognition”, IEEE, ICSMC, vol. 3, pp. 2128-2132, 2005.

[15] 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.

[16] B. Jedynak, H.C. Zheng and M. Daoudi, “Skin detection using pairwise models”, IVC(23), No. 13, pp. 1122-1130, 29 Nov. 2005.

[17] E. Hjelmas and B. K. Low. “Face detection: A survey” Computer Vision and Image Understanding, vol. 83, no. 3, pp. 236-274, 2001.

[18] 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.

[19] A. Ono, “Face recognition with Zernike moments”, Systems and Computers, vol. 34, No 10, 2003.

[20] Vezhnevets V., Szonov V., Andreeva A., “A Survey on Pixel-Based Skin Color Detection Techniques". Proc. Graphicon-2003, pp. 85-92, Moscow, Russia, September 2003.

[21] Terrillon, J.-C., Shirazi, M. N., Fukamachi, H., and Akamatsu, “Comparative performance of different skin chrominance models and chrominance spaces for the automatic detection of human faces in colour images”, In Proc. of the nternational Conference on Face and Gesture Recognition, pp. 54–61, 2000.

[22] Kovac, J.; Peer, P.; Solina, F, “Human skin color clustering for face detection”, EUROCON 2003. Computer as a Tool. The IEEE Region 8, Volume 2, pp. 144-148, Sept. 2003.

[23] Guerfi, J.-P. Gambotto et S. Lelandais, “Ligne de Partage des Eaux pour l'extraction de visage dans l'espace de couleurs TLS”, CORESA, Nov. 2005.

[24] D Brown, I Craw, J Lewthwait, “A SOM based approach to skin detection with application in real time systems”, Proc. of the British Machine Vision Conference, 2001.

[25] 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), pp. 711--720, 1997.

[26] 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, Nara Japan, IEEE Computer Society, pp. 200-205, April 14-16 1998.