

Facial Expression Classification Method Based on Pseudo Zernike Moment and Radial Basis Function Network

Tran Binh Long, Le Hoang Thai, and Tran Hanh

Abstract—This paper presents a new method to classify facial expressions from frontal pose images. In our method, first Pseudo Zernike Moment Invariant (PZMI) was used to extract features from the global information of the images and then Radial Basis Function (RBF) Network was employed to classify the facial expressions, based on the features which had been extracted by PZMI. Also, the images were preprocessed to enhance their gray-level, which helps to increase the accuracy of classification. For JAFFE facial expression database, the achieved rate of classification in our experiment is 98.33%. This result leads to a conclusion that the proposed method can ensure a high accuracy rate of classification.

Index Terms—Facial expression classification, pseudo Zernike moment invariant, RBF neural network.

I. INTRODUCTION

Facial expressions deliver rich information about human emotions, and thus play an essential role in human communications. For facial expression classification, data from static images or video sequences are used. In fact, there have been many approaches for facial expression classification, using static images and image sequences [1],[2]. Those approaches first track the dynamic movement of facial features and then classify the facial feature movements into six expressions (i.e., smile, surprise, anger, fear, sadness, and disgust). Classifying facial expression from static images is more difficult than from video sequences because less information during the expression is available [3].

In order to design a highly-accurate classification system, the choice of feature extractor is very crucial. There are two approaches for feature extraction extensively used in conventional techniques [4]. The first approach is based on extracting structural facial features that are local structure of face images, for example, the shapes of the eyes, nose and mouth. This structure-based approach deals with local information. The second approach is based on statistical information about the features extracted from the whole image, so it uses global information. [5]

Our proposed facial expression classification system is composed of three stages (Fig.1). In the first stage, the location of face in arbitrary images was detected. To ensure a robust, accurate feature extraction distinguishing between face and non-face region in an image, exact location of the

face region is needed. We used a ZM-ANN technique which had been already presented in reference [6] for face localization and created a sub-image which contains information necessary for classification algorithm. By using a sub-image, data irrelevant to facial portion are disregarded. In the second stage, the pertinent features from the localized image obtained in the first stage were extracted. These features are obtained from the pseudo Zernike moment invariant. Finally, facial images based on the derived feature vector obtained in the second stage were classified by RBF network. Also, only automatic classification of facial expressions from still images in Japanese Females Facial Expression (JAFFE) database (Fig.2) [7] is discussed.

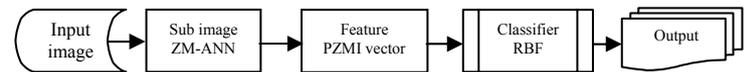


Fig. 1. The chart of PZMI-RBF system

The remainder of the paper is organized as follows: section 2 describes the preprocessing procedure to get the pure expression image; section 3 presents the pseudo Zernike feature extraction and our feature vector creation; section 4 discusses the classification based on RBF network, section 5 presents the experiments on the JAFFE facial expression database, and section 6 mentions our conclusions.



Fig. 2. Examples of seven principal facial expressions in JAFFE: smile, disgust, anger, surprise, fear, neutral, and sadness (from left to right)

II. FACE LOCALIZATION METHOD

Many algorithms have been proposed for face localization and detection, which can be seen from a critical survey [8]. Face localization helps find an object in an image to be used as the face candidate. The shape of this object resembles that of a face. Thus, faces are characterized by elliptical shape. In other words, an ellipse can approximate the shape of a face. A ZM-ANN technique presented in [6] has proven to be able to find the best-fit ellipse to enclose the facial region of the human face in a frontal pose image.

The operation of face detection is done in two phases:

- In the first phase, representative Zernike vector is extracted from a selected image by a proper algorithm.
- In the second phase, a three-layer perceptron neural network, beforehand trained, receives on its input layer the Zernike moments vector and then gives on its output layer 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 the interactions closely related to the determined face region of

Manuscript received May 22, 2012; revised June 25, 2012.

T. B.Long is with Lac Hong University, Dong Nai, 71000, Viet Nam (e-mail: tblong@lhu.edu.vn).

L. H.Thai is with the University of Science, Ho Chi Minh city, 70000, Viet Nam (e-mail:lhthai@fit.hcmus.edu.vn).

T. Hanh was with Lac Hong University, Dong Nai, 71000, Viet Nam. (e-mail: tranbinhlong97@gmail.com).

the selected image. (Fig.3)

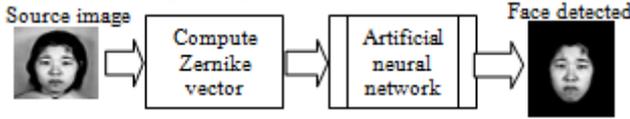


Fig. 3. General diagram of the system detection

Generally, the implementation of our method can be briefly described as follow:

- Computing vectors of Zernike moments for all the images (N) in the work database.
- Constructing training database by randomly choosing M images from the work database (M<<N) and identifying Zernike moment vectors Z_i corresponding to M images.
- Manually delimiting face area in each image of the training database based on a set of points representing the contour C_i of each treated face. The points include the top, bottom, left and right of identified image and they form an ellipse whose semi-major axis $a=45$, semi-minor axis $b=40$, rate $b/a \approx 8/9$ (see fig.5.a).
- Training neural network on the set of M couples (Z_i, C_i).

The test and measurement of the performance of the network obtained after training operation were done on (N-M) the other images in the work database.

III. FEATURE EXTRACTION TECHNIQUE

Feature extraction is defined as a process of converting a captured biometric sample, i.e. face expression, into a unique, distinctive and compact form so that it can be compared to a reference template. According to [9], moment sequence, M_{pq} is uniquely determined by the image $f(x,y)$ and conversely, $f(x,y)$ is uniquely described by M_{pq} . The uniqueness of the moment method has prompted us to its suitability in face feature extraction. Furthermore, the orthogonal property of the PZM enables redundancy reduction among their respective description and thus helps to improve the computation efficiency.

A. Pseudo Zernike Moment Invariant

The kernel of pseudo Zernike moments is the set of orthogonal pseudo Zernike polynomials defined over the polar coordinates inside a unit circle. The two dimensional pseudo Zernike moments of order p with repetition q of an image intensity function $f(r,\theta)$ are defined as [10]:

$$PZ_{pq} = \frac{p+1}{\pi} \int_0^{2\pi} \int_0^1 PV_{pq}(r, \theta) f(r, \theta) r dr d\theta \quad (1)$$

where Zernike polynomials $PV_{pq}(r,\theta)$ are defined as:

$$PV_{pq}(r, \theta) = PR_{pq}(r) e^{-jq\theta}; \quad j = \sqrt{-1}$$

and

$$r = \sqrt{x^2 + y^2}, \theta = \tan^{-1}\left(\frac{y}{x}\right), \quad -1 < x, y < 1$$

The real-valued radial polynomials are defined as:

$$PR_{pq}(r) = \sum_{s=0}^{p-|q|} (-1)^s \frac{(2p+1-s)!}{s!(p+|q|+1-s)!(p-|q|-s)!} r^{p-s} \quad (2)$$

and $0 \leq |q| \leq p, p \geq 0$

Since it is easier to work with real functions, PZ_{pq} is often split into its real and imaginary parts, PZ_{pq}^c, PZ_{pq}^s as given below:

$$PZ_{pq}^c = \frac{2(p+1)}{\pi} \int_0^{2\pi} \int_0^1 PR_{pq}(r) \cos(q\theta) f(r, \theta) r dr d\theta \quad (3)$$

$$PZ_{pq}^s = \frac{2(p+1)}{\pi} \int_0^{2\pi} \int_0^1 PR_{pq}(r) \sin(q\theta) f(r, \theta) r dr d\theta \quad (4)$$

where $p \geq 0, q > 0$

Since the set of pseudo Zernike orthogonal polynomials is analogous to that of Zernike polynomial, most of the previous discussion for the Zernike moments can be adapted to the case of PZM.

It can be seen that Zernike moment in below equation

$$Z_{pq} = \frac{p+1}{\pi} \int_0^{2\pi} \int_0^1 V_{pq}(r, \theta) f(r, \theta) r dr d\theta; \quad |r| \leq 1 \quad (5)$$

will become pseudo Zernike moments if the radial polynomials, R_{pq} , defined as in below equation

$$R_{pq}(r) = \sum_{k=0}^{\frac{p-|q|}{2}} (-1)^k \frac{(p-k)!}{k! \left(\frac{p+|q|}{2} - k\right)! \left(\frac{p-|q|}{2} - k\right)!} r^{p-2k} \quad (6)$$

with its condition $p-|q| = \text{even}$, are eliminated [11]. Hence, pseudo Zernike moments offer more feature vectors than Zernike moments as pseudo Zernike polynomial contains $(p+1)^2$ linearly independent polynomials of order $\leq p$, where as Zernike polynomial contains only $\frac{1}{2}(p+1)(p+2)$ linearly independent polynomials due to condition of $p-|q| = \text{even}$.

B. Feature Vector Creation

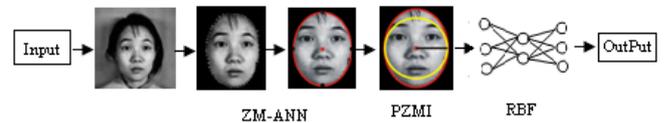


Fig. 4. Schematic block diagram of the proposed PZMI model

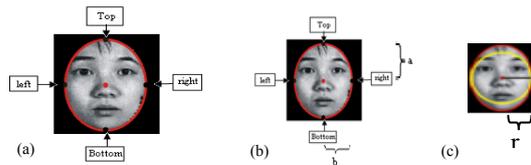


Fig. 5. Center of ellipse, circle determined by basing on the top, bottom, left and right.

The computation of the vectors of pseudo Zernike moments for all the images in the work database includes two stages. The first stage is selecting the image region to compute the pseudo Zernike vector. It is noticed from the analysis of facial expressions that when the emotion changes, the primary changing face areas are likely to be the eyes, the mouth, and the eyebrows (fig.5.c). The research on PZMI shows that the farther a position is away from the center of the circle, the larger the PZM coefficient at that position. Through the analyses, based on prior studies, we propose a technique to extract the selected image area to calculate PZM vector as follows: First, we determine the circle which is the typical area to compute the PZM vector- illustrated in Fig.5.c. The center of the circle coincides with that of the ellipse with semi major axis $a=45$, semi minor axis $b=40$, rate $b/a \approx 8/9$.

The ellipse itself is the border area surrounding the face region (fig.5.b). Our experimental results have proved that the proposed technique enables a full collection of eye and mouth features in Jaffe database (Fig.4). Then, we identify the characteristic pseudo Zernike vectors in the selected images.

With this technique, the center of the circle PZMI is placed in such a position that it coincides with the center of the images identified in phase 1 (where $r=b$)

In the second stage, the feature vector was obtained by calculating the PZMI of the derived sub-image. According to selecting PZMI as face feature, we defined four categories of feature vectors based on the order (p) of the PZMI. In the first category with $p=1, 2, \dots, 6$, all moments of PZMI were considered as feature vectors elements. The number of the feature vector elements in this category is 26. In the second category, $p=4, 5, 6, 7$ were chosen. All moments of each order included in this category were then summed up to create feature vectors of size 26. In the third category, $p=6, 7, 8$ were considered. The feature vector for this category has 24 elements. Finally, the last category with $p=9, 10$ was considered with 21 feature elements.[12]

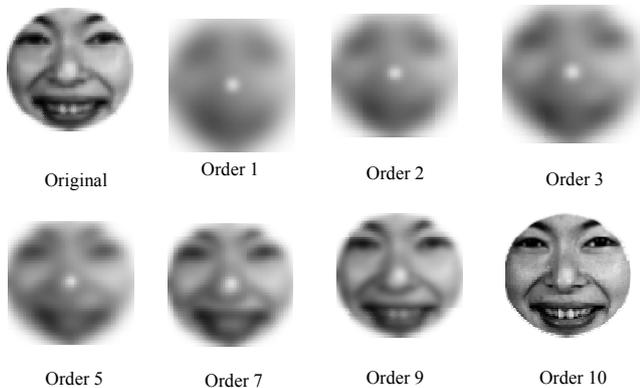


Fig. 6. Original and reconstructed with different order face images

With the results based on the value of $p = 10$, our experimental study indicates that this method of selecting the pseudo Zernike moment order as the feature elements allows the feature extractor to have a lower-dimensional vector while maintaining a good discrimination capability. (Fig.6)

IV. CLASSIFIER DESIGN

The major advantages of RBFN over other models such as feed-forward neural network and back propagation are its fast training speed and local feature convergence [13]. Thus, in this paper, RBF neural network is used as a classifier in a facial expression classification system where the inputs to the neural network are feature vectors derived from the proposed feature extraction technique described in the previous section.

A. RBF Neural Network Description

The radial basis function neural network (RBFN) theoretically provides such a sufficiently- large network structure that any continuous function can be approximated within an arbitrary degree of accuracy by appropriately choosing radial basis function centers [13]. The RBFN is trained using sample data to approximate a function in

multidimensional space. A basic topology of RBFN is depicted in Fig. 7. The RBFN is a three-layered network. The first layer constitutes input layer in which the number of nodes is equal to the dimension of input vector. In the hidden layer, the input vector is transformed by radial basis function as activation function:

$$\varphi(\bar{x}; \bar{c}_j) = \exp\left(-\frac{1}{2\sigma^2} \|\bar{x} - \bar{c}_j\|^2\right), \quad (7)$$

where $\|\bar{x} - \bar{c}_j\|$ denotes a norm- (usually Euclidean distance)- of the input data sample vector \bar{x} and the center \bar{c}_j of radial basis function. The k th output is computed by equation

$$F_k(x) = \sum_{j=1}^m w_{kj} \varphi(\bar{x}; \bar{c}_j) \quad (8)$$

where w_{kj} represents a weight synapse associates with the j th hidden unit and the k th output unit with m hidden units.

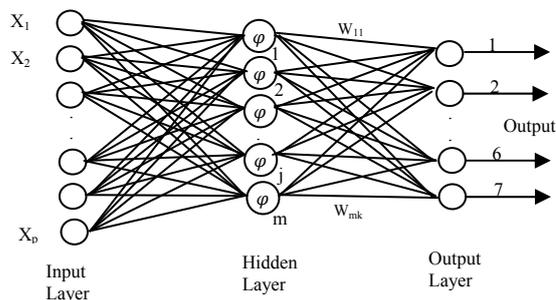


Fig. 7. Basic topology of RBFN

We employed the RBFN to classify the facial expressions from images in the Eigen-space domain extracted via PZMI as described in the previous section. The architecture was depicted in Fig. 7.

B. RBF Neural Network Classifier Design

To design a classifier based on RBF neural networks, in the input layer of the neural network, we set an amount of input nodes which are as many as feature vector elements. The number of nodes in the output layer is 7, equivalent to 7 facial expressions of image classes. Initially, the RBF units are equal to the number of output nodes, and these RBF units increase if classes are overlapped.

V. EXPERIMENTAL RESULTS

In this section, we demonstrate the capabilities of the proposed PZMI-RBFN approach in classifying seven facial expressions. The proposed method is evaluated in terms of its classification performance using the JAFFE female facial expression database [14], which includes 213 facial expression images corresponding to 10 Japanese females. Every person posed 3 or 4 examples of each of the seven facial expressions (happiness, sadness, surprise, anger, disgust, fear, neural). Two facial expression images of each expression of each subject were randomly selected as training samples, while the remaining samples were used as test data, without overlapping. We have 140 training images and 73 testing images for each trial. To investigate the local effect of the source images, we used Images size: 80×80 . Since the size of the JAFFE database is limited, we had performed the trial over 3 times to get the average classification rate. Our obtained classification rate is 98.33% (Table I).

TABLE I: CLASSIFICATION RATE (%) OF THE PROPOSED PZMI-RBF MODEL

Test	Sadness	Smile	Disgust	Neutral	Surprise	Fear	Anger
1	97.98	97.58	98.95	98.01	98.68	97.88	98.45
2	98.8	98.88	98.76	98.87	98.64	98.46	98.95
3	98.7	96.85	96.92	97.42	98.84	98.45	98.86

For the classification performance evaluation, a False Acceptance Rate (FAR) and a False Rejection Rate (FRR) test were performed. These two measurements yield another performance measure, namely Total Success Rate (TSR):

$$TSR = \left(1 - \frac{FAR+FRR}{total\ number\ of\ accesses}\right) \times 100\% \quad (9)$$

The system performance can be evaluated by using Equal Error Rate (EER) where FAR=FRR. A threshold value is obtained based on Equal Error Rate criteria where FAR=FRR. Threshold value of 0.2954 is gained for PZM as measure of dissimilarity.

Table II shows the testing result of verification rate with order moments from setting 10 (moments order 10) for PZM based on their defined threshold value.

The results demonstrated that the application of pseudo Zernike moments as feature extractor can best perform the classification.

TABLE II: TESTING RESULT OF VERIFICATION RATE OF PZM

moment	thres	FAR(%)	FRR(%)	TSR(%)
PZMI	0.2954	2.7998	3.1674	98.33

We have compared our proposed with some of the existing facial expression classification techniques on the same Jaffe database. This comparative study indicates the usefulness and the utility of the proposed technique

The three other methods taken for the comparison were HRBFN+PCA [15], Gabor + PCA+LDA [16], GWT+DCT+RBF [17] (see in Table III)

TABLE III: COMPARATIVE RESULTS OF THE CLASSIFICATION RATE (%) OF DIFFERENT APPROACHES

Methods	Rate
Gabor + PCA+LDA [16]	97.33%
GWT+DCT+RBF [17]	89.11%
HRBFN+PCA [15]	95.68%
Proposed method	98.33%

VI. CONCLUSIONS

The performance of orthogonal pseudo Zernike moment invariant (PZMI) and radial basis function neural network (RBFN) in the facial expression classification system was presented in this paper. It was seen from the performance that higher orders of orthogonal moment contain more information about face image and this improves the classification rate. The pseudo Zernike moments of order 10 has the best performance. An RBF neural network was used as a classifier in this classification system. The highest classification rate of 98.33%, FAR = 2.7998% and FRR = 3.1674% with Jaffe database was achieved using the proposed algorithm, which represents the overall performance of this facial expression classification system. The proposed algorithms, orthogonal PZMI+RBF N, possess some advantages: orthogonality and geometrical invariance. Thus, they are able to minimize information redundancy as

well as increase the discrimination power.

REFERENCES

- [1] I. A. Essa, A. P. Pentland, "Coding, Analysis, Interpretation, and Recognition of Facial Expressions," *IEEE Trans. Pattern Analysis and Machine Intelligence*, Vol. 19, pp.757-763, 1997.
- [2] B. Fasel, J. Luetttin, "Automatic facial expression analysis: a survey," *Pattern Recognition*, Vol. 36, pp.259-275, 2003.
- [3] X. W. Chen and T. Huang, "Facial expression recognition: a clustering-based approach," *Pattern Recognition Letter*, Vol. 24, pp. 1295-1302, 2003.
- [4] J. Daugman, "Face Detection: A Survey," *Computer Vision and Image Understanding*, Vol. 83, No. 3, pp. 236-274, Sept. 2001.
- [5] L. F. Chen, H. M. Liao, J. Lin and C. Han, "Why Recognition in a statistic-based Face Recognition System should be based on the pure Face Portion: A Probabilistic decision-based Proof," *Pattern Recognition*, Vol. 34, No.7, pp. 1393-1403, 2001.
- [6] D.T.Hai, L.H.Thai, L.H.Bac, "Facial boundary detect in images using Moment Zernike and Artificial Neural Network," in *Proc of Dalat University's Information technology Conference 2010*, pp. 39-49, DaLat, Vietnam (in Vietnam)
- [7] www.kasrl.org/jaffe.html.
- [8] J. Daugman, "Face Detection: A Survey," *Computer Vision and Image Understanding*, Vol. 83, No. 3, pp.236-274, Sept. 2001
- [9] Hu. M.K, "Visual pattern recognition by moment invariants." *IRE Trans. On Information Theory*, vol. 8,No. 1, pp. 179-187, 1962.
- [10] R. Mukundan and K.R. Ramakrishnan, *Moment functions in image analysis – theory and applications*. World Scientific Publishing, 1998.
- [11] C.H. Teh and R.T. Chin. "On image analysis by the methods of moments." *IEEE Trans. Pattern Anal. Machine Intell.*, vol. 10, pp. 496-512, July 1988.
- [12] J. Haddadnia, M. Ahmadi, K. Faez, "An Efficient Human Face Recognition System Using Pseudo Zernike Moment Invariant and Radial Basis Function Neural Network," *International Journal of Pattern Recognition and Artificial Intelligence* Vol. 17, No. 1,pp. 41-62, 2003
- [13] S. Haykin, *Neural Networks: A Comprehensive Foundation*, Macmillan College Publishing Company, New York, 1994.
- [14] M. J. Lyons, S. Akamatsu, M. Kamachi, J. Gyoba, "Coding Facial Expressions with Gabor Wavelets," In: *Proceedings of the 3th IEEE International Conference on Automatic Face and Gesture Recognition*, Nara, Japan, 1998, pp.200-205
- [15] V. P. Lekshmi, M.Sasikumar,"A Neural Network Based Facial Expression Analysis using Gabor Wavelets," *World Academy of Science, Engineering and Technology* 42, 2008, pp 563-567.
- [16] H.B. Deng, L.W. Jin, L.X. Zhen, J.C.Huang, "A New Facial Expression Recognition Method Based on Local Gabor Filter Bank and PCA plus LDA," *International Journal of Information Technology*, vol. 11, no.11, 2005.
- [17] D.T. Lin,"Facial Expression Classification Using PCA and Hierarchical Radial Basis Function Network," *Journal of Information Science and Engineering* 22, 1033-1046, 2006.

Tran Binh Long received B.S degree from Lac Hong University, Dong Nai, Viet Nam, in 2002 and M.S degree in Ho Chi Minh University of Information Technology, Viet Nam, in 2007. Since 2002, he has been a lecturer at Faculty of Information Technology, Lac Hong University, Dong Nai, Viet Nam. His research interests include soft computing, pattern recognition, image processing, biometric and computer vision. Mr. Tran Binh Long is co-author of four papers in the international conferences and a national journal.

Le Hoang Thai received B.S degree and M.S degree in Computer Science from Ha Noi University of Technology, Viet Nam, in 1995 and 1997. He received Ph.D. degree in Computer Science from Ho Chi Minh University of Natural Science, Viet Nam, in 2004. Since 1999, he has been a lecturer at Faculty of Information Technology, Ho Chi Minh University of Natural Science, Viet Nam. His research interests include soft computing, pattern recognition, image processing, biometric and computer vision. Dr. Le Hoang Thai is co-author over twenty five papers in international journals and international conferences.

Tran Hanh received B.S degree in Mathematics from Ha noi University of Synthesis, Viet Nam, in 1972. He received Ph.D degree in Mathematics from the Calculated Center, Academy of National Science USSR, in 1978. From 2000 – 2011, he had been a lecturer at Faculty of Information Technology, Lac Hong University, Dong Nai, Viet Nam. His research interests include soft computing, biometric and computer vision. Mr. Tran Hanh is co-author over twenty papers in the international and national conferences.