2010 International Conference on Cyber-Enabled Distributed Computing and Knowledge Discovery

Recognition of Facial Expression Using Centroid Neural Network

Dong-Chul Park, Huynh Thuy, and Dong-Min Woo
Dept. of Electronics Engineering
Myong Ji University
Yong In, KOREA
E-mail:parkd@mju.ac.kr

Yunsik Lee
System IC R&D Division
Korea Electronics Tech. Inst.
Songnam, KOREA
Email: leeys@keti.re.kr

***Abstract*—A novel approach to recognize facial expressions from static images is proposed in this paper. The local binary pattern (LBP) operator is adopted as an effective feature extraction tool for facial image data. An unsupervised competitive neural network, called a centroid neural network with χ^2 distance measure, CNN-χ^2, is then utilized as the classification tool for the histogram data obtained by the LBP operator on facial image data. The proposed recognition scheme is applied to the JAFFE database and compared with several conventional approaches to facial expression recognition problems. The results show that the proposed recognition scheme compares favorably with conventional approaches in terms of recognition accuracy.**

***Keywords*- facial expression; recognition; neural network; LBP**

I. INTRODUCTION

In recent years, efforts have been made to address the interesting and challenging problem of automatic facial expression recognition (FER). FER has potential applications in various recognition problems including video-conferencing, human-computer interaction (HCI) systems, and security systems. FER can also be an essential and indispensable component of technologies for the creation of human-like robots and machines. Many appearance-based approaches have been proposed to deal with FER problems and a survey of this body of research can be found in [1]. These approaches mainly differ in the choice of appropriate face description and corresponding similarity measure. An automatic FER system consists of three major components, a face detection procedure, a facial feature extraction method, and a facial expression classification algorithm. Unlike the face recognition problem, FER focuses on how to discern similar expressions from different persons. Once a face is detected in a complex scene with the cluttered background of an image, the corresponding region is extracted. It is then usually normalized to have the same size. The most widely used face detection method is the Adaboost algorithm [2]. Note that the face detection procedure will not be addressed in this paper. An important step for successful FER is to find an extraction method of features that can produce feature vectors to effectively discriminate different facial expressions while producing similar feature vectors for similar facial expressions. Most methods to extract facial features first detect some fiducial points and then calculate features around the chosen fiducial points instead of using the whole face image. With a set of manually detected fiducial points on a face image, there are two main approaches to extract facial features: The geometric positions of selected fiducial points and a set of Gabor wavelet coefficients at these points [3].

It has been shown that the Gabor wavelet coefficients show better performance than the geometric positions [4]. In using the Gabor wavelet coefficients as a tool for extracting facial features, the central problem is how to find Gabor wavelet representations efficiently [4], [5], [6]. That is, the computational complexity involved in convolving face images with a bank of Gabor filters to extract multi-scale and multi-orientation coefficients is another obstacle in using the Gabor wavelet method. Furthermore, the above approaches require manual selection of fiducial points. More recently, the local binary pattern (LBP) was proposed as a powerful local descriptor for the texture of images. It has been used with considerable success in a number of visual recognition tasks including face recognition [7], [8] and texture classification [11]. LBP has also been introduced to the field of facial expression analysis [9], [10].

The most important property of LBP features is its invariance to monotonic gray-level changes by design and as such it generally requires no preprocessing on images [8]. After the computation of feature vectors with the LBP operator on raw image data, the next step for a facial expression recognition system is to design a classifier that can determine the distance between a given face image and a certain facial expression model. Among several clustering algorithms such as the k-means algorithm, Self-Organizing Map (SOM), Centroid neural network (CNN), and Fuzzy c-means algorithm, we find CNN to be an appropriate candidate for the facial expression classifier. The advantages of employing the CNN over other clustering algorithms as a component for the classifier include its stable convergence without requiring predetermined parameters such as a schedule for learning gain or the total number of iterations for clustering [12]. In this paper, CNN-χ^2, an advanced form of the CNN suitable for classifiers using histogram features is adopted for the FER problem [7]. The CNN-χ^2 utilizes the χ^2 distance as its distance measure whereas the original CNN uses the

978-0-7695-4235-5/10 $26.00 © 2010 IEEE
DOI 10.1109/CyberC.2010.94

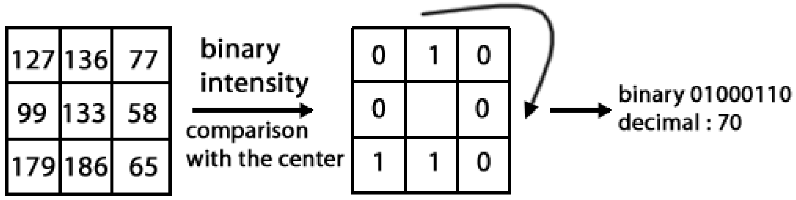

Fig. 1. Example of LBP operation

Euclidean distance as its distance measure. The remainder of this paper is organized as follows:

Section 2 briefly summarizes the LBP operator for extracting feature vectors from face data. The CNN and the CNN with χ^2 distance measure, which is used as a clustering algorithm in this work, are summarized in Section 3. Section 4 describes experiments involving the JAFFE database and presents the obtained results. Finally, conclusions are given in Section 5.

II. FEATURE EXTRACTION WITH LOCAL BINARY PATTERNS

Ojala et al. [8], [11] first introduced the basic LBP operator as a complementary measure for local image contrast. It is a gray-scale invariant texture primitive statistic and has shown excellent performance in the classification of various kinds of textures. The operator labels the pixels of an image by thresholding the neighborhoods of each pixel with the center value and considering the result as a binary number. As shown in Fig. 1, at a given pixel position (x_c, y_c), the LBP is defined as an ordered set of binary comparisons of pixel intensities between the center pixel and its predetermined S_p surrounding pixels (in this case, $S_p = 8$). The decimal form of the resulting S_p-bit word (LBP code) can be expressed as follows:

$$LBP\left(x_c, y_c\right) = \sum_{n=0}^{S_p-1} s\left(i_n - i_c\right)2^n \tag{1}$$

where i_c corresponds to the grey value of the center pixel (x_c, y_c) and the function $s(x)$ is defined as:

$$s\left(x\right) = \begin{cases} 1 & \text{if} \quad x \geq 0 \\ 0 & \text{if} \quad x < 0. \end{cases} \tag{2}$$

Based on the operator, each pixel of an image is labeled with an LBP code. The 256-bin histogram of the labels contains the density of each label and can be used as a texture descriptor of the considered region. Feature extraction is implemented as follows: First, the face image is divided into several non-overlapped blocks (16×18 in our method after experimenting with several different block sizes). Then, LBP histograms are computed for each block and the block LBP histograms are concatenated into a single vector. Therefore, the facial expression data are represented by concatenated LBP histograms and the shape of histogram is used as the feature in our FER system. Fig. 2 is an example of extracting LBP histograms.

III. CENTROID NEURAL NETWORK WITH χ^2 DISTANCE

A. Centroid Neural Network (CNN)

The CNN algorithm is an unsupervised competitive learning algorithm based on the classical k-means clustering algorithm. It finds the centroids of clusters at each presentation of the data vector. When an input vector x is presented to the network at epoch n, the weight update equations for winner neuron j and loser neuron i in CNN can be summarized as follows:

$$w_j\left(n+1\right) = w_j\left(n\right) + \frac{1}{N_j+1}\left[x\left(n\right) - w_j\left(n\right)\right] \tag{3}$$

$$w_i\left(n+1\right) = w_i\left(n\right) - \frac{1}{N_i-1}\left[x\left(n\right) - w_i\left(n\right)\right] \tag{4}$$

where $w_j(n)$ and $w_i(n)$ represent the weight vectors of the winner neuron and the loser neuron, iteration n, respectively. The CNN has several advantages over conventional algorithms such as SOM or k-means algorithm when used for clustering and unsupervised competitive learning. The CNN requires neither a predetermined schedule for learning gain nor the total number of iterations for clustering. It always converges to sub-optimal solutions while conventional algorithms such as SOM may give unstable results depending on the initial learning gains and the total number of iterations. More detailed description on the CNN can be found in [12], [13].

B. CNN with χ^2 distance measure

CNNs have been successfully applied to various clustering problems with deterministic data. CNNs, however, may not be appropriate for high dimensional data such as histograms. In order to measure the similarity of 2 histograms effectively, the following χ^2 distance measure is employed:

$$\chi^2\left(M, S\right) = \sum_{i=1}^{Q} \frac{\left(M_i - S_i\right)^2}{S_i} \tag{5}$$

where M and S correspond to the model and sample histograms, respectively, and Q represents the dimension of the histograms. For the CNN with the χ^2 distance measure, the objective function to be minimized is defined as:

$$J = \sum_{k=1}^{Q}\sum_{i=1}^{N} \frac{\left(w_k - x_i\left(k\right)\right)^2}{x_i\left(k\right)}, \quad x_i\left(k\right) \in \text{Group } k \tag{6}$$

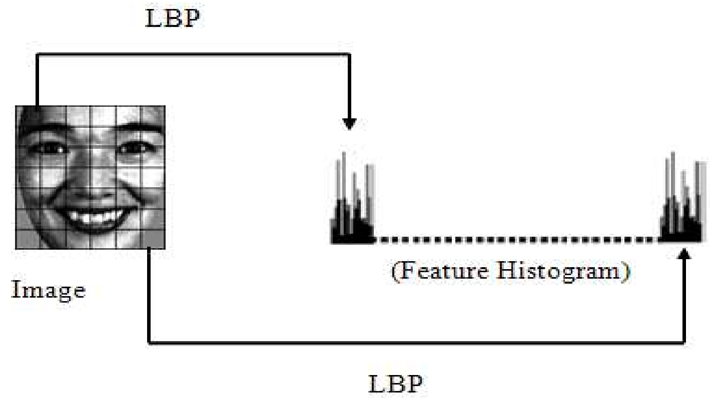

Fig. 2. Feature extraction from face image using LBP operator

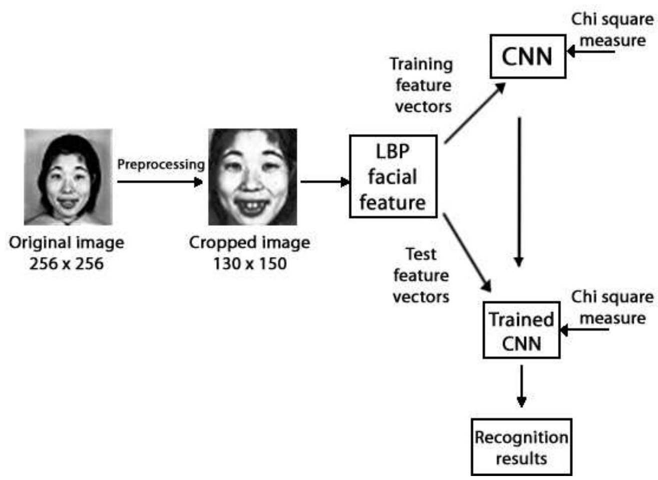

Fig. 3. General scheme of the proposed facial expression recognition system

where N denotes the number of data points in the Group k. A necessary condition for optimal position of the center for each group can be obtained as follows:

$$\frac{\partial J}{\partial w_k} = \sum_{i=1}^{N} \frac{2\left(w_k - x_i\left(k\right)\right)}{x_i\left(k\right)} = 0 \tag{7}$$

By solving Eq. (7) [14], the update equations for winner neuron j and loser neuron i of CNN with χ^2 can be summarized as follows:

$$\frac{1}{w_j\left(n+1\right)} = \frac{1}{w_j\left(n\right)} + \frac{1}{N_j+1}\left(\frac{1}{x\left(n\right)} - \frac{1}{w_j\left(n\right)}\right) \tag{8}$$

$$\frac{1}{w_i\left(n+1\right)} = \frac{1}{w_i\left(n\right)} - \frac{1}{N_i-1}\left(\frac{1}{x\left(n\right)} - \frac{1}{w_i\left(n\right)}\right) \tag{9}$$

where $w_j(n)$ and $w_i(n)$ represent the weight vectors of the winner neuron and the loser neuron at the iteration n, respectively.

IV. EXPERIMENTS AND RESULTS

Fig. 3 presents a block diagram of the FER system based on the CNN with χ^2 distance and the LBP. The facial images are extracted by employing a LBP operator to produce facial feature vectors. The training samples are clustered according to the similarity of their features and the classifier utilizes clustering information in recognizing facial expressions.

The experiments are conducted on the Japanese Female Facial Expression (JAFFE) database, which contains 213 images of 7 facial expressions posed by 10 Japanese female models. Ten expressers posed 3 or 4 examples of each of six facial basic expressions (happiness, sadness, surprise, anger, disgust, and fear) and a neutral expression. Images from the JAFFE database are cropped to eliminate most of the background and some parts of the hair and chin. The size of the images ranges from 256×256 to 130×150. For details on the collection of these images, the reader is referred to [15]. Fig. 4 shows examples from the original and preprocessed images in the JAFFE database. In the real-world environments, the rotation of the camera axis and the head pose variations often exist. The JAFFE database we are using for experiments includes these images with minor rotation of camera axis and variations in head poses. Fig. 5 presents some examples of these cases. As a result, the robustness of the proposed method for these cases is evaluated in the experiments

After the input image is transformed to LBP codes, sun-window histograms (spatial histograms) are used to model the face. However, the size of the sun-window has to be determined in order to balance the spatial locality and compactness of the model. Therefore, experiments are performed to examine the influence of the number of sun-windows on the recognition rate. Seven different cases of sun-windows, i.e., 6×7, 8×10, 14×16, 16×18, 18×21, 21×25, and 26×30, are evaluated.

Classification is performed by the CNN-χ^2 algorithm with 4 code vectors. This number of vectors is small enough to achieve a high-speed classification rate and large enough to achieve good discrimination between classes of used data. The $LBP(8,2)$ uniform is specifically employed to reduce the length of feature vectors. Throughout the experiments, the 10-fold cross-validation method is adopted to deal with the small sample size. That is, the database was divided randomly into ten roughly equal-sized parts. Nine parts were used for training the classifiers and the last part was used for testing. The above process was repeated ten times so that each part was used once as the test set. This is a standard methodology used in machine learning [16].

Changes in the parameters of LBP descriptors may cause variations in the overall performance. For example, the change of window sizes from 6×7 to 16×18 for an uniform $LBP(8,2)$ increases the histogram length from 2478 to 16992 while the mean recognition rate improves from 85.71 % to 91.43 %. Some parameters need to be optimized for LBP feature selection. The first is choosing the LBP operator. Choosing a LBP operator that produces a large amount of different labels makes the histogram long and thus calculation of the distance matrix becomes more time-consuming. On the other hand, using a small number of labels makes the feature vector shorter, but it also increases the likelihood of losing some important information. Ojala et al. acknowledged that the vast majority of the uniform LBP could be found in texture images and facial images. With the uniform LBP, the length of the feature vector can be much shorter and a simple version of rotation invariant LBP can be obtained [11]. As a compromise, the uniform $LBP(8,2)$ operator, which has 59 labels, is selected for our experiments. The number of labels for a neighbor of $Sp = 8$ pixels is $2^{S_p} = 256$ for standard LBP and $Sp \times (Sp - 1) + 3 = 59$ for the uniform LBP. For example, by changing from LBP(8,2) to the uniform LBP(8,2) for the case of 42 sub-windows (6×7 sub-window case), the histogram length can be reduced from 10752 (256 $\times$ 6 $\times$ 7) to 2478 (59 $\times$ 6 $\times$ 7).

Another parameter to determine is the number of regions to be divided. A small number of windows induces a decrease in the recognition rate due to partial loss of information while using a large number of windows produces long feature vectors, leading to greater computation time and slow classification. Through experiments on several numbers of sun-windows, the uniform $LBP(8,2)$ operator in a 16×18 pixel window was selected because it represents a good trade-off between recognition performance and feature vector length. Table I provides a comparison among different FER methods using the JAFFE database. The following important observations were obtained from the performance comparison of the different methods:

The method based on labeled elastic graph matching, 2D Gabor wavelet representation, and a linear discriminate analysis [4] reports a recognition rate of 75 %. In the experiments, face images of only nine persons, that is, 193 of the 213 images, were used. The downside of this method is the 2D Gabor wavelet representation is insufficient to overcome the illumination problem. Moreover, a set of 34 fiducial points is used to model the face geometry in combination with a Principal Component Analysis (PCA), which created some redundant information in the construction of filter banks. To avoid redundant information, a reduced set of 19 fiducial points is used to model the face geometry in [21].

In [18], 34 landmark points were manually located from each facial image and then converted into a labeled graph (LG) vector using the Gabor wavelet transformation method to represent the facial features. The nearest neighbor classifier is then used for classification. The best classification performance was compared with that of other methods: Canonical correlation analysis (CCA) and Generalized discriminate analysis (GDA). A recognition rate of 77.05 % was reported, where all neutral images were excluded and only 183 images were used. A major drawback on this approach is high computational complexity as the training data size increases.

The technique based on linear programming for both feature selection and classifier training in the FER system [6] reports an accuracy of 91.4 %. However, the computation of Gabor-wavelet representations is very intensive; Bartlett et al. show that the Gabor-wavelet representation derived from each 48×48 face image has a high dimensionality of O(10^5)[19]. Furthermore, Shan et al. compared the time and memory costs of two feature extraction processes, LBP features and Gabor-wavelet features, as presented in Table II[20]. The results show that the FER system using the LBP histogram allows for very fast feature extraction whereas the other method requires a high computational cost in extracting a large set of Gabor wavelet coefficients.

The experimental results show that the proposed FER system using LBP features for facial information representation and the CNN-χ^2 algorithm for classification provides 91.43 % recognition accuracy. In addition, it should be pointed out that 34 fiducial points are manually selected in [4], [6], [18]. The preprocessing step of face cropping is performed manually in [17] and face cropping is obtained by coordinates of the eyes in [9]. In contrast, the procedures such as the preprocessing images method, the facial feature extraction method and the facial expression classification method are

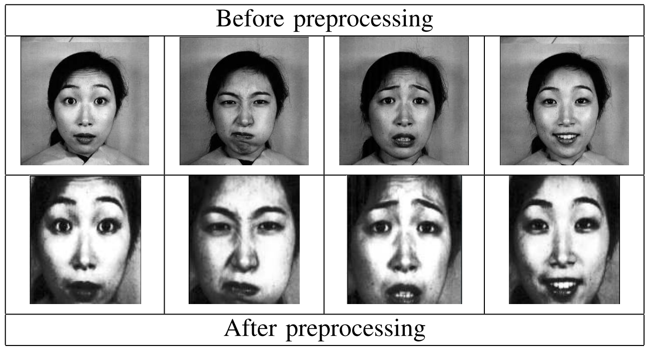

Fig. 4. Samples from original JAFFE images before and after preprocessing.

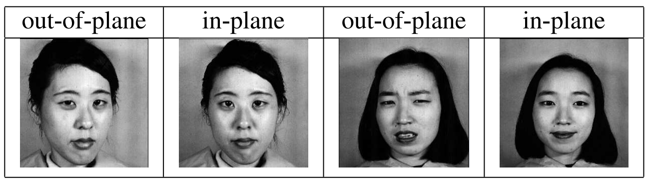

Fig. 5. Examples of images with head pose variations

TABLE II
COMPARISON OF TIME AND MEMORY COSTS FOR EXTRACTING FEATURES

Methods	Memory (feature dimension)	Time (feature extraction time)
LBP	2478	0.03 s
Gabor	42,650	30 s

completely automatic in the proposed FER system. Fig. 6 and Fig. 7 show some examples of test results with successful recognitions and with failed recognitions, respectively.

V. CONCLUSIONS

In this paper, we have developed a facial expression recognition method based on the LBP operator and a CNN with the χ^2 distance measure. The combination of the CNN and the χ^2 distance provides an efficient approach to deal with histograms obtained by LBP for facial expression analysis. Because of the LBP operator, the proposed FER method is invariant to monotonic gray-level changes in image data and the feature extraction procedure is very fast when compared to other methods using Gabor wavelet coefficients. Another advantage over conventional FER methods is that the proposed approach requires no manual operation whereas conventional approaches require some manual operations for face cropping and/or selection of fiducial points on face image data. The proposed FER method was applied to the JAFFE database with 213 images in total. Through experimentation with different parameters for the proposed FER method, we noticed relative sensitivity to the choice of the number of sun-windows and LBP operators for extracting feature vectors. A comparison of mean recognition rates among conventional methods and the proposed FER method shows that the proposed FER method outperforms conventional methods in terms of recognition accuracy. The proposed FER method also has advantages in offering fully automatic operation. The experimental results demonstrate the effectiveness of the combination of the CNN-χ^2 algorithm and LBP features for FER. Future research should focus on conducting more experiments with larger data sets and reducing the length of the feature vector so that the system can run faster and more efficiently.

ACKNOWLEDGMENT

We would like to thank Dr. Michael J. Lyons for the use of the JAFFE database. This work was supported by National Research Foundation of Korea Grant funded by the Korean Government (2010-0009655) and by MKE "Configurable Device and SW R&D"(No.: KI002168). The author would like to thank Intelligent Research Laboratory (ICRL) members including J.-Y. Kim for their help in preparing this manuscript.

REFERENCES

[1] B. Fasel and J. Luettin. Automatic facial expression analysis: A survey. *Pattern Recognition*, 36:259–275, 2003.

[2] Y. Freund and R.E. Schapire. A Decision-Theoretic Generalization of On-Line Learning and an Application to Boosting. *J. Comput. Syst. Sci*, 55(1):119–139, 1997.

[3] Z. Zhang. Feature-Based Facial Expression Recognition: Sensitivity Analysis and Experiment with a Multi Layer Perceptron. *Int. J. Pattern Recog. Artif. Intel.*, 13(6):893–911, 1999.

[4] M.J. Lyons, J. Budynek, and S. Akamatsu. Automatic classification of single facial images. *IEEE Trans. Pattern Analysis and Machine Intelligence*, 21(12):1357–1362, 1999.

[5] G. Donato, M. Bartlett, J. Hager, P. Ekman, and T. Sejnowski. Classifying facial actions. *IEEE Trans. Pattern Analysis and Machine Intelligence*, 21(10):974–989, 1999.

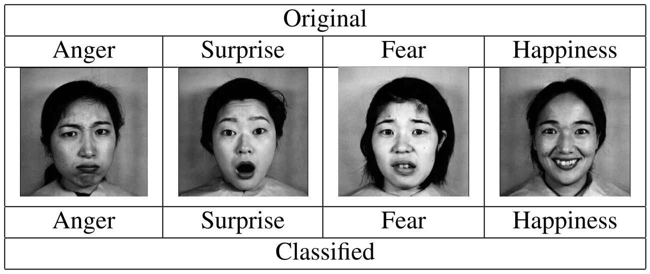

Fig. 6. Examples of test results with successful recognition results

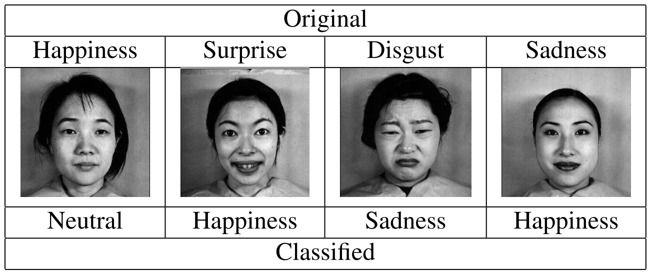

Fig. 7. Examples of test results with failed recognition results

TABLE I
COMPARISON OF VARIOUS ALGORITHMS USING JAFFE DATABASE

Method [Reference]	Preprocessing & Data Used			Accuracy
	Image Cropping	# of Data	Fiducial Points	
HLAC + Fisher weight map [17]	Manual	193	None	% 69.4 %
Gabor Wavelet + PCA + LDA [4]	None	193	Manually selected 34 points	75 %
LBP + Coarse-to-Fine method [9]	Manual (Position of eyes)	193	None	77%
Gabor wavelet + KCCA [18]	None	193	Manually selected 34 points	77.05%
Gabor Filter + LP [6]	None	213	Manually selected 34 points	91.4%
LBP + CNN (Proposed)	Automatic	213	None	91.43%

[6] G. Guo and C. Dyer. Learning from examples in the small sample case: face expression recognition system. *IEEE Trans. Sys. Man and Cyber.- Part B*, 35(3):477–488, 2005.

[7] T.L.H. Vu and D. Park. Centroid neural network for face recognition. *Proc. of IEEE IJCNN*, 1:1304–1309, 2009.

[8] T. Ahonen, A. Hadid, and M. Pietikinen. Face description with local binary patterns: Application to face recognition. *IEEE Trans. Pattern Analysis and Machine Intelligence*, 28(12):1134–1146, 2006.

[9] X. Feng. Facial expression recognition based on local binary patterns and coarse-to-fine classification. *Proc. of IEEE Conf. on Computer and Information Technology*, 178 - 183, 2004.

[10] G. Zhao and M. Pietikinen. Dynamic texture recognition using local binary patterns with an application to facial expressions. *IEEE Trans. Pattern Analysis and Machine Intelligence*, 29(6):915–928, 2007.

[11] T. Ojala, M. Pietikainen, and T. Maenpaa. Multiresolution gray-scale and rotation invariant texture classification with local binary patterns. *IEEE Trans. Pattern Analysis and Machine Intelligence*, 24(7):971–987, 2002.)

[12] D.C. Park. Centroid Neural Network for Unsupervised Competitive Learning. *IEEE Trans. on Neural Networks*, 11(2):520–528, 2000.

[13] D. Park and Y. Woo. Weighted centroid neural network for edge reserving image compression. *IEEE Trans. on Neural Networks* 12(5):1134–1146, 2001.

[14] L. Vu Thi, D.-C. Park, D.-M. Woo, and Y. Lee. Centroid neural network with chi square distance measure for texture classification. *Proc. of IJCNN*, 1310-1315, 2009.

[15] M. Lyons, S. Akamatsu, M. Kamachi, and J. Gyoba. Coding facial expressions with gabor wavelets. *Proc. of Third IEEE Int. Conf. on Automatic Face and Gesture Recognition*, 200 - 205, 1998.

[16] S.M. Weiss and C.K. Kulikowski. *Computer systems that learn*, San Mateo, CA: Morgan Kaufmann, 1991.

[17] Y. Shinohara and N. Otsu. Facial expression recognition using fisher weight maps. *Proc. of IEEE Conf. on Automatic Face and Guesture Recognition*, 499–504, 2004.

[18] W. Zheng. Facial expression recognition using kernel canonical correlation analysis. *IEEE Trans. on Neural Networks*, 17(1):233–238, 2006.

[19] M.S. Bartlett, et al. Recognizing facial expression: machine learning and application to spontaneous behavior. *Proc. of IEEE CVPR*, 2:568–573, 2005.

[20] C. Shan, S. Gong, and P.W. McOwan. Facial expression recognition based on Local Binary Patterns: A comprehensive study. *Image and Vision Computing*, 27:803–816, 2009.

[21] A.C. Koutlas and D.I. Fotiadis. A region based methodology for facial expression recognition. *Proc. of Int. Conf. on Bio-inspired Systems and Signal Processing (BIOSIGNALS)*, 2:218–223, 2008.