Facial Emotion Recognition Using Nonparametric Weighted Feature Extraction and Fuzzy Classifier

Maryam Imani and Gholam Ali Montazer

***Abstract*—** The most facial emotion recognition methods use the geometrical features of face such as eye opening and mouth opening, which extraction of them is a hard and complicated task. But, in this paper, we propose to use the discriminant features simply extracted by nonparametric weighted feature extraction (NWFE). Moreover, to model the inherent uncertainties contained in the emotional features, we use the fuzzy measure in the classifier. To this end, a nearest neighbor classifier with fuzzy Euclidean distance is used to recognize emotions of face images. The experimental results on JAFFE database show the superiority of the proposed method compared to some other emotion recognition methods.

***Keywords*—** *fuzzy; nearest neighbor; nonparametric weighted feature extraction; emotion recognition.*

I. INTRODUCTION

The facial expressions are changes within the face due to human inner feelings. By evidence of Ekman, these facial expressions have the same meaning for all people [1]. Six basic human expressions presented by Ekman are: happiness, sadness, anger, fear, disgust and surprise. Facial emotion recognition is very important in human-machine interactions. In a human–machine Interaction, machine recognizes the facial expression and consequently the feeling of human. Then, machine performs a corresponding action to it for improvement of human-machine interaction.

Three main steps of each emotion recognition system are face detection, feature extraction, and classification. In the most of emotion recognition frameworks, after face detection, the next step is facial feature extraction. Feature extraction from the face image is one of the most time-consuming and complicated steps in an emotion recognition framework. The precision of this step largely affects the final precision of the emotion recognition method. Consequently, selection of a simple, efficient and precise algorithm for this step has high importance.

Maryam Imani is with the Department of Electrical and Electronics Engineering, Khatam University, Tehran, Iran (e-mail: m.imani@khatam.ac.ir).
Gholam Ali Montazer is with the Faculty of Information Technology Engineering, Tarbiat Modares University, Tehran, Iran (e-mail: montazer@modares.ac.ir).

The most of facial emotion recognition methods use the geometrical characteristics of face as features. For example in [2], the geometrical features of face are selected as the input of interval type-2 Mamdani fuzzy system. 17 points similar to MPEG-4 standard [3] are chosen. Nine features are formed by Euclidian distances between these points. These features consist of eye opening, mouth opening, mouth width, height of middle of eyebrows, height of inner eyebrows, and the means of distances between outer lip corners and inner eye corners. Uncertainty and vagueness are the indispensable and fundamental aspects of knowledge where in many of practical problems, there is vagueness in features and uncertainty in decision making. Fuzzy sets encounter the vagueness and uncertainty [4]-[5].

Two categories of facial features are used in [6]: 1- the primary features, which are needed to recognize every six basic emotions, and the secondary features, which are considered as the auxiliary attributes for accurate emotion recognition. Mouth opening, mouth corners displacement, eye opening, and eyebrow constriction are instances of primary features. The existence of tooth, lips thickness, mouth length, nose-side wrinkles, and eyebrows slope are instances of secondary facial features. According to [7], two levels of uncertainty can be considered in emotion recognition: intra-personal level uncertainty and interpersonal level uncertainty. A person may have variation in facial expression related to similar emotive experience, which results in intra-personal level uncertainty. On the other hand, different persons may have variation in facial expression for similar emotional experience, which results in interpersonal level uncertainty. To model these types of uncertainties, authors in [8] use the fuzzy approach. The skin-colored and eight facial components consist of eye brows, eyes, forehead, lips, teeth, chin, nose and cheeks are used in [9] for face expression detection. In [10], the fuzzy relational approach selects some facial features such as mouth opening, eye opening and the length of eyebrow constriction extracted from the localized regions. Then, the extracted features are fuzzified, and fed into a Mamdani-type relational model.

For different pattern recognition problems, the varied feature extraction methods have been proposed [11]-[13]. The most popular supervised feature extraction method is linear discriminant analysis (LDA) [14]-[15]. In this method, the between-class scatter is maximized while simultaneously the within-class scatter is minimized. LDA has some difficulties. The within-class scatter matrix becomes singular when a small training sample is available. Moreover, because of limitation in the rank of between-class scatter matrix, LDA can extract maximum $c-1$ features where the number of classes is denoted by c. The principal component analysis (PCA) transform is the most popular unsupervised feature extraction method [15]. PCA produces a lower dimensional feature space where the direction with the largest variance for data is selected. In small sample size situations, PCA can outperform LDA because PCA has less sensitivity to different sizes of training sets. Because PCA does not consider the class separability, it may not be very efficient in classification problems. Locality preserving projection (LPP) is a manifold learning method which can be performed supervised or unsupervised [16]. LPP maintains the locality of data structure. While unsupervised LPP has good efficiency in data reconstruction, the supervised LPP can be a good choice in data classification. Nonparametric weighted feature extraction (NWFE), which is an extension of LDA, deals with difficulties of LDA, and so, is efficient using both the small and large training set from the classification accuracy point of view [17]. The nonparametric scatter matrices of NWFE calculated by the weighted mean allow to extract more than $c-1$ features.

While the most facial emotion recognition methods use the geometrical features of face for emotion recognition, which is complex and time consuming, we propose to use the discriminant features simply extracted by nonparametric weighted feature extraction. To improve the classification results by considering the inherent uncertainty present in facial emotion data, the fuzzy measure is used for classification. In other words, a nearest neighbor (NN) classifier with fuzzy Euclidean distance is used for classification of extracted features. The proposed method is experimented on a popular facial expression database. The results show the superiority of the proposed method compared to some other emotion recognition methods.

The reminder of this work is continued as follows. The proposed method is introduced in section II and the experimental results are reported in section III. Finally, section IV represents the conclusions.

II. PROPOSED METHOD

An emotion recognition method from facial features is proposed in this section. So far, many methods have been proposed for face detection. In this work, the assumption is that the face detection step is done and the part of image that contains the frontal face is provided and forwarded to the next step. Almost, all facial emotion recognition methods use the geometrical features of face images such as eye opening, height of middle of eyebrows, and mouth opening as the discriminant features for separating different emotion classes. The extraction of facial geometrical features is complicated and time consuming, and the final precision of classification is largely dependent to the precision of detection of them.

The proposed emotion recognition method in this paper uses the NWFE for feature extraction and the NN classifier with Euclidean distance for classification. A nonparametric extension of scatter matrices is used in NWFE which provides some advantages compared to parametric discriminant analysis methods. The nonparametric scatter matrices in NWFE are full rank, and so, can provide the desired number of extracted features. In addition, because of the nonparametric nature of scatter matrices, NWFE is efficient even for non-normal datasets. It assigns greater weight to sample points near the expected decision boundary, and so, the classification accuracy is improved. Moreover, NWFE deals with the singularity problem of within-class scatter matrix using the regularization approach. To compute the weighted mean, NWFE assigns different weights on every sample. The nonparametric within-class scatter matrix ($\mathbf{S}_w$) and between-class scatter matrix ($\mathbf{S}_b$) are calculated by:

$$\mathbf{S}_b = \sum_{i=1}^{c} p_i \sum_{\substack{j=1 \\ j\neq i}}^{c} \sum_{l=1}^{n_i} \frac{\lambda_l^{(i,j)}}{n_i}\left(x_l^{(i)} - M_j\left(x_l^{(i)}\right)\right)\left(x_l^{(i)} - MjxliT \right. \quad (1)$$

$$\mathbf{S}_w = \sum_{i=1}^{c} p_i \sum_{l=1}^{n_i} \frac{\lambda_l^{(i,i)}}{n_i}\left(x_l^{(i)} - M_i\left(x_l^{(i)}\right)\right)\left(x_l^{(i)} - MixliT \right. \quad (2)$$

where $x_l^{(i)}$ is lth sample of class i, n_i is the number of training samples in class i, p_i represents the prior probability of class i and c is the number of classes. The scatter matrix weight $\lambda_l^{(i,j)}$ and the weighted mean $M_j\left(x_l^{(i)}\right)$ are defined as follows:

Anger Disgust Sad Neutral Happy Surprise Fear

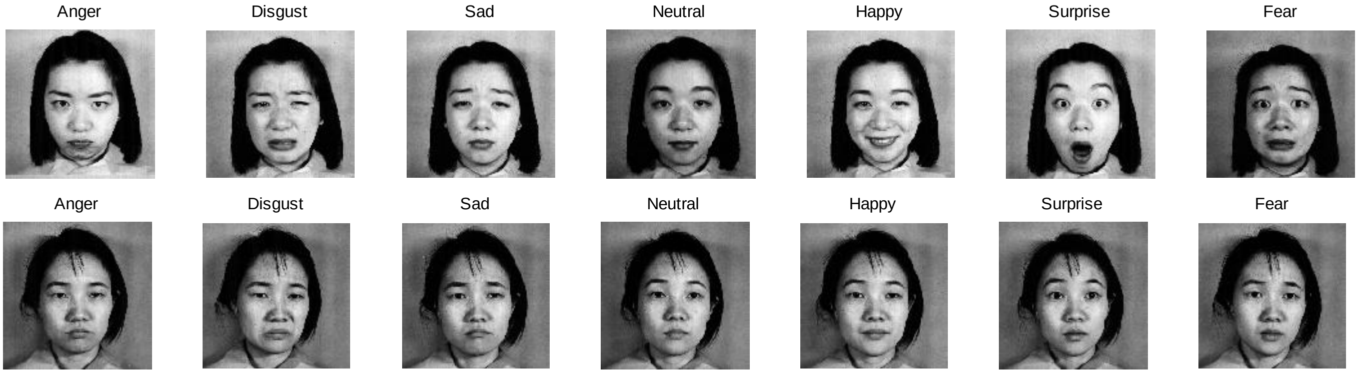

Fig. 1. The sample images of JAFFE database.

$$\lambda_l^{(i,j)} = \frac{dist\left(x_l^{(i)}, M_j\left(x_l^{(i)}\right)\right)^{-1}}{\sum_{t=1}^{n_i} dist\left(x_t^{(i)}, M_j\left(x_t^{(i)}\right)\right)^{-1}} \quad (3)$$

$$M_j\left(x_l^{(i)}\right) = \sum_{k=1}^{n_i} w_{lk}^{(i,j)} x_k^{(j)} \quad (4)$$

$$w_{lk}^{(i,j)} = \frac{dist\left(x_l^{(i)}, x_k^{(j)}\right)^{-1}}{\sum_{t=1}^{n_j} dist\left(x_l^{(i)}, x_t^{(j)}\right)^{-1}} \quad (5)$$

To deals with singularity of $\mathbf{S}_w$, the following regularization method is used:

$$\mathbf{S}_w = 0.5\mathbf{S}_w + 0.5diag(\mathbf{S}_w) \quad (6)$$

where the diagonal parts of matrix $\mathbf{A}$ is denoted by $diag(\mathbf{A})$. To obtain the projection matrix, the Fisher criterion, $\max tr\left(\boldsymbol{S}_{\mathrm{w}}{}^{-1}\boldsymbol{S}_{\mathrm{b}}\right)$ is used.

For facial feature extraction using NWFE, each face image with r rows and c column is firstly reshaped to a $n \times 1$ vector where $n = r \times c$. Then, the projection matrix is calculated using the training samples vectors. By applying the transformation on n-dimensional vector, m discriminant features are extracted. Let $\boldsymbol{F} = [f_1, f_2, \dots, f_m]^T$ be the feature vector extracted from a sample image of dataset. The membership value of each feature is considered to be simply the normalized feature value as follows:

$$\mu_{\boldsymbol{F}}(f_i) = \frac{f_i - min(\boldsymbol{F})}{max(\boldsymbol{F}) - min(\boldsymbol{F})};\ i = 1,2, \dots, m \quad (7)$$

Then, the NN classifier with fuzzy Euclidean distance is used to compute the label (emotion) of each testing sample (a non-seen face image). Let $\mu_{\boldsymbol{F}}^{x_j}(f_i)$ be the membership degree of feature f_i in the feature vector of jth training sample $(\boldsymbol{x}_j)$ and $\mu_{\boldsymbol{F}}^{\boldsymbol{y}}(f_i)$ be the degree of membership of f_i in the feature vector of testing sample $(\boldsymbol{y})$. The fuzzy Euclidean distance between training sample $\boldsymbol{x}_j$ and testing sample $\boldsymbol{y}$ is calculated by:

$$d_j = \sqrt{\sum_{i=1}^{m} \left(\mu_{\boldsymbol{F}}^{x_j}(f_i) - \mu_{\boldsymbol{F}}^{\boldsymbol{y}}(f_i)\right)^2};\ \ ; j = 1, \dots, N \quad (8)$$

where $N = \sum_{k=1}^{c} n_k$ is the number of total training samples. The label of testing sample $\boldsymbol{y}$ is obtained by:

$$l_{\boldsymbol{y}} = \arg\min_{j=1,\dots,N} d_j \quad (9)$$

III. EXPERIMENTS

In this section, the performance of the proposed method is compared to some facial emotion recognition methods. The Japanese Female Facial Expression (JAFFE) Database, which contains 213 images of 10 Japanese female models, is used in experiments [18]. The models posed 7 facial expressions consist of anger, disgust, sadness, neutrality, happiness, surprise, and fear. The JAFFE images were taken at the Psychology Department in Kyushu University. Some samples of this database are shown in Fig. 1.

For doing experiments, we used 15 images from each class for training and the remained samples are used as testing samples. To assess the performance of different methods, we use several measures such as average recognition rate (mean of class specific accuracies), overall recognition rate (the percentage of correctly classified samples), and Kappa coefficient [19]. The Kappa coefficient is defined by:

$$KC = \frac{L\sum_{i=1}^{c} t_{ii} - \sum_{i=1}^{c} t_{i+}t_{+i}}{L^2 - \sum_{i=1}^{c} t_{i+}t_{+i}} \quad (10)$$

where L is the number of testing samples, c is the number of classes. t_{ii} denotes the number of testing samples correctly recognized in class i, t_{i+} is the number of testing samples recognized as class i, and t_{+i} representes the number of samples predicted as belonging to class i. In addition, the McNemars test is done to assess the statistical significance of differences in the achieved recognition rates [20]. The parameter Z_{12} in McNemars test is defined as:

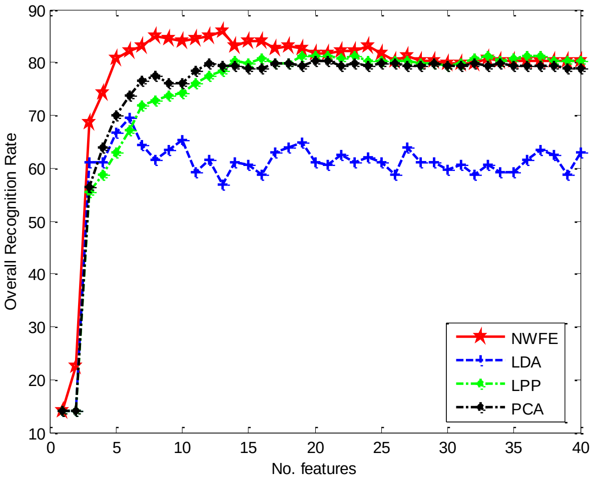

Fig. 2. The average recognition rate versus the number of extracted features.

$$Z_{12} = \frac{f_{12}-f_{21}}{\sqrt{f_{12}+f_{21}}} \quad (11)$$

where f_{12} denotes the number of samples correctly recognized by method 1 and incorrectly by method 2. The difference in the recognition rate between methods 1 and 2 is statistically significant if $|Z_{12}| > 1.96$. The sign of Z_{12} is used to say whether method 1 is more accurate than classifier 2 ($Z_{12} > 0$) or vice versa ($Z_{12} < 0$).

The proposed approach uses the NWFE method for feature extraction from face images and the NN classifier with fuzzy Euclidean distance for classification. The performance of the proposed method is assessed when other feature transformations such as LDA, LPP, and PCA are used instead of NWFE. The comparison results are shown in Fig. 2. This figure illustrates the average recognition rate versus the number of extracted features. As we can see, NWFE is superior to other methods especially when a low number of features (3 to 16 features) are extracted. The classification results obtained by 13 features are represented in Table I. The results of McNemars test are represented in Table II. In all methods except PCA the fear class is recognized with the lowest accuracy. The supervised feature extraction methods (NWFE and LDA) recognize angry with the highest accuracy and unsupervised feature extraction methods (LPP and PCA) classify the neutral class with the highest recognition rate. The performance of the proposed method is also compared to the fuzzy relational approach [10], multilayer perceptron (MLP) [21]-[22], and radial basis function network (RBFN) [21]-[22] methods in terms of average recognition rate. The results are reported in Table III.

Table I. The classification results obtained by 13 features.

Class	NWFE	LDA	LPP	PCA
Anger	96.67	66.67	80.00	83.33
Disgust	82.76	62.07	86.21	79.31
Sadness	80.65	58.06	74.19	80.65
Neutral	93.33	60.00	90.00	93.33
Happiness	93.55	48.39	74.19	70.97
Surprise	83.33	56.67	76.67	76.67
Fear	71.88	46.88	68.75	71.88
Average Recognition rate	**86.02**	56.96	78.57	79.45
Overall recognition rate	**85.92**	56.81	78.40	79.34
Kappa coefficient	**0.8357**	0.4962	0.7481	0.7590

Table II. The McNemars test results.

	NWFE	LDA	LPP	PCA
NWFE	0	7.52	2.83	2.56
LDA	-7.52	0	-6.04	-6.30
LPP	-2.83	6.04	0	-0.53
PCA	-2.56	6.30	0.53	0

Table III. Comparison of face emotion recognition methods.

Method	Average recognition rate
Proposed	**86.02**
Fuzzy relational approach	82.32
MLP	67.58
RBFN	60.73

IV. CONCLUSION

An emotion recognition method was proposed in this paper that uses the NWFE transformation for feature extraction from face images. The proposed method uses the NN classifier for classification of extracted features and includes the fuzzy measure in the classifier to model the uncertainties contained in the emotional extracted features. The experimental results show that the proposed method outperforms some other feature extraction methods such as LDA, LPP, and PCA and also is superior to some other facial emotion recognition methods such as fuzzy relational approach, MLP, and RBFN.

REFERENCES

[1] P. Ekman and W. V Friesen, Manual for the facial action coding system, Consulting Psychologists Press, 1978.
[2] M. Safayani,V. Farmani, M. Ahmadi, A. Mirzaei, "Adaptive Interval Type-2 Fuzzy Inference System for Facial Expression Recognition," Journal of Computing and Security, vol. 2, no. 3, pp. 195-207, July 2015.
[3] A Murat Tekalp and Joern Ostermann. Face and 2-d mesh animation in mpeg-4. Signal Processing: Image Communication, vol. 15, no. 4, pp. 387-421, 2000.
[4] B. Ashtiani, F. Haghighirad, A. Makui, G. A. Montazer, " Extension of fuzzy TOPSIS method based on interval-valued fuzzy sets," Applied Soft Computing, vol. 9, pp. 457‑461, 2009.
[5] G. A. Montazer, H. Q. Saremi, V. Khatibi, "A neuro-fuzzy inference engine for Farsi numeral characters recognition," Expert Systems with Applications, vol. 37, pp. 6327–6337, 2010.
[6] M. Ilbeygi, H. Shah-Hosseini, " A novel fuzzy facial expression recognition system based on facial feature extraction from color face images," Engineering Applications of Artificial Intelligence, vol. 25, pp. 130–146, 2012.
[7] J. M. Mendel, R. I. John, and F. Liu, "Interval type-2 fuzzy logic systems made simple," IEEE Trans. Fuzzy Syst., vol. 14, no. 6, pp. 808–821, Dec. 2006.
[8] A. Halder, A. Konar, R. Mandal, A. Chakraborty, P. Bhowmik, N. R. Pal, and A. K. Nagar, " General and Interval Type-2 Fuzzy Face-Space Approach to Emotion Recognition," IEEE Transactions on Systems, Man, and Cybernetics: Systems, vol. 43, no. 3, pp. 587-605, May 2013.
[9] K. azadmanesh, R. javidan, S. E. Alavi, "Improvement of Facial Emotion Recognition Using Skin Color and Face Components," International journal of Computer Science & Network Solutions, vol. 2, no. 4, pp. 55-65, April 2014.
[10] A. Chakraborty, A. Konar, U. K. Chakraborty, and A. Chatterjee, "Emotion recognition from facial expressions and its control using fuzzy logic," IEEE Trans. Syst., Man, Cybern. A, Syst., Humans, vol. 39, no. 4, pp. 726–743, Jul. 2009.
[11] M. Imani, H. Ghassemian, "Feature Extraction Using Weighted Training Samples," IEEE Geoscience and Remote Sensing Letters, vol. 12, no. 7, pp. 1387-1391, July 2015.
[12] J. Zabalza, J. Ren, Z. Wang, H. Zhao, J. Wang, and S. Marshall, "Fast Implementation of Singular Spectrum Analysis for Effective Feature Extraction in Hyperspectral Imaging," IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing, vol. 8, no. 6, pp. 2845 – 2853, June 2015.
[13] M. Imani, H. Ghassemian, "Binary coding based feature extraction in remote sensing high dimensional data," Information Sciences, vol. 342, pp. 191–208, 2016.
[14] G. Baudat and F. Anouar, ''Generalized discriminant analysis using a kernel approach,'' Neural Comput., vol. 12, pp. 2385–2404, 2000.
[15] K. Fukunaga, Introduction to Statistical Pattern Recognition, San Diego: Academic Press Inc., 1990.
[16] X. F. He and P. Niyogi, "Locality preserving projections," in Proc. Adv. Neural Inf. Process. Syst., vol. 16, pp. 153–160, 2004.
[17] B. C. Kuo and D. A. Landgrebe, "Nonparametric weighted feature extraction for classification," IEEE Trans. Geosci. Remote Sens, vol. 42, no. 5, pp. 1096-1105, May 2004.
[18] Michael J. Lyons, Shigeru Akemastu, Miyuki Kamachi, Jiro Gyoba, "Coding Facial Expressions with Gabor Wavelets," 3rd IEEE International Conference on Automatic Face and Gesture Recognition, pp. 200-205, 1998.
[19] J. Cohen, A coefficient of agreement from nominal scales, Edu. Psychol. Meas., vol. 20, no. 1, pp. 37–46,1960.
[20] G. M. Foody, "Thematic map comparison: Evaluating the statistical significance of differences in classification accuracy," Photogramm. Eng. Remote Sens., vol. 70, no. 5, pp. 627–633, 2004.
[21] M. Gargesha and P. Kuchi, "Facial expression recognition using a neural network," Artif. Neural Comput. Syst., pp. 1–6, 2002.
[22] D.-J. Kim, S.-W. Lee, and Z. Bien, "Facial emotional expression recognition with soft computing techniques," in Proc. 14th IEEE Int. Conf. Fuzzy Syst., 2005, pp. 737–742.

Maryam Imani received the B.Sc. and M.Sc. degrees in electrical engineering from Shahed University, Tehran, Iran, and the Ph.D. degree in electrical engineering from Tarbiat Modares University, Tehran, Iran in 2009, 2011, and 2015 respectively. She continued her research in Tarbiat Modares University as a postdoc. She is an Assistant professor of the Department of Electrical and Electronics Engineering, Khatam University, Tehran, Iran. Her research interests include pattern recognition, signal and image processing, and remote sensing.

Gholam Ali Montazer received his B.Sc. degree in Electrical Engineering from Kh.N. Toosi University of Technology, Tehran, Iran, in 1991, his M.Sc. degree in Electrical Engineering from Tarbiat Modares University, Tehran, Iran, in 1994, and his Ph.D. degree in Electrical Engineering from the same university, in 1998. He is an Associate Professor of the Department of Information Engineering at Tarbiat Modares University, Tehran, Iran. His areas of research include Information Engineering, Knowledge Discovery, Intelligent systems, E-Learning and Image Mining.