

Research Article

Subject Independent Facial Emotion Classification Using Geometric Based Features

Suhaila N. Mohammed and Loay E. George

Department of Computer Science, College of Science, Baghdad University, Baghdad, Iraq

Abstract: Accurate emotion categorization is an important and challenging task in computer vision and image processing fields. Facial emotion recognition system implies three important stages: Prep-processing and face area allocation, feature extraction and classification. In this study a new system based on geometric features (distances and angles) set derived from the basic facial components such as eyes, eyebrows and mouth using analytical geometry calculations. For classification stage feed forward neural network classifier is used. For evaluation purpose the Standard database "JAFFE" have been used as test material; it holds face samples for seven basic emotions. The results of conducted tests indicate that the use of suggested distances, angles and others relative geometric features for recognition give accuracy about 95.73% when the seven emotion classes are tested and 97.23% when the 6 classes (except normal class) are only tested. These rates are considered high when compared with the results of other newly published works.

Keywords: Facial emotion, features reduction, geometric attributes, neural network, statistical classifier

INTRODUCTION

Recently, the use of machines to distinguish human expressions is more important than the mere use in research and analysis of the results. So, building a system for Human-Computer Interaction (HCI) is an essential task (Jain, 2011). Despite the importance of emotion recognition which demanded in various fields (like medicine, education, driver safety, games etc.) It still remains as the unsolved problem (Butalia *et al.*, 2010). Mehrabian explained that it can give the impression of person through words by 7 and 38% through tone of voice while the facial image gives the largest rate which is reach to 55% (Rani and Garg, 2014). Mehrabian indicated that facial expressions contain much information about mood and state of human impression (Saini and Rana, 2014). The number of the facial expressions cannot be strictly specified because of the difference in cultural and surrounding back ground that person has.

However, researches on the facial expression analysis have focused more on the seven basic emotional expressions (fear, anger disgust, happiness, surprise, sadness and neutral) (Iatraki, 2009). Facial emotions classification methods are mainly classified into two common types according to the features used namely: geometric feature-based methods and appearance feature-based methods (Chen *et al.*, 2014; Zhang *et al.*, 2012a). In geometric feature-based methods the facial components are extracted to form a feature vector that represents the face geometry such as shapes and facial fiducially points (Dutta and Baru,

2013; Sumathi *et al.*, 2012). Prasad and Danti (2014) identified expression using geometrical mouth features by applying Susan edge detector on this region to extract geometric features vector (like mouth width, mouth height etc.) (Prasad and Danti, 2014) have worked on Discriminant Laplacian Embedding (DLE) to extract feature points and use their displacement from emotion to another for the purpose of their classification (Wang *et al.*, 2013). Also, other class of feature-based face emotion methods had been developed; the methods belong to this class use features that appear temporarily in the face during any kind of facial expression (such as the presence of specific facial wrinkles, bulges). Image filters, such as Gabor wavelets, are applied to extract the features vector (Sonka *et al.*, 2008).

Surbhi and Arora (2013) proposed a system that extract features vector using optical flow, active shape model and Principle Component Analysis (PCA) techniques. They have used neural network for classification and the recognition rate was 90% (Surbhi and Arora, 2013). Local Binary Pattern (LBP) and Gabor wavelets representations have been proposed by Surbhi and Arora (2013) to perform facial expression recognition; their results indicated the system can lead to recognition rate 84.76% (Zhang *et al.*, 2012b).

Several attempts have also reported using both geometric and appearances based features to overcome the limitation of both types which are referred to as hybrid methods. RenuandSumeet proposed a system that extracts features by drawing Bezier curve on facial regions like eyes and lips based on points that represent

Corresponding Author: Suhaila N. Mohammed, Department of Computer Science, College of Science, Baghdad University, Baghdad, Iraq

This work is licensed under a Creative Commons Attribution 4.0 International License (URL: <http://creativecommons.org/licenses/by/4.0/>).

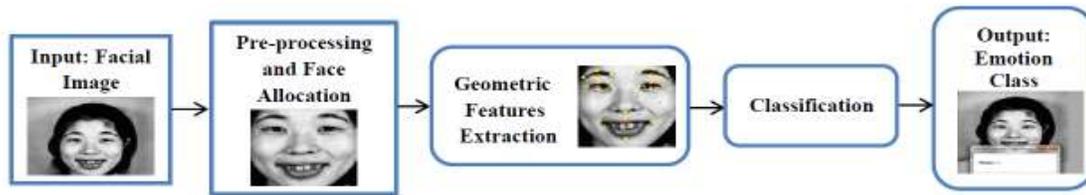


Fig. 1: Block diagram for overall system design

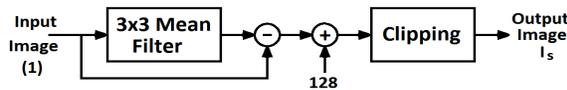


Fig. 2: Block diagram for subtraction process

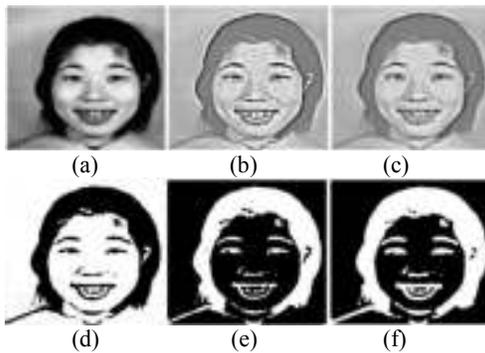


Fig. 3: (a): Original image; (b): Subtraction step result; (c): Smoothing step result; (d): Contrast stretching step result; (e): Thresholding step result; (f): Dilation step result

changes in muscles around them and achieve accuracy up to 70% (Shih, 2010).

MATERIALS AND METHODS

The overall design of the proposed facial emotion recognition system is shown in Fig. 1; it consists of three stages:

- Pre-processing and face area allocation stage
- Features vector extraction stage
- Classification stage

The main task of the pre-processing stage is image enhancement; this task is applied to improve the image data by suppressing the undesired illumination and noise effects and enhances some important image features for further processing (Sonka *et al.*, 2008). The task of face region (i.e., ROI) extraction is, also, performed as part of pre-processing stage.

The goal of feature extraction stage is to extract a set of strong discriminating features from the basic facial components to represent the emotion class state.

In classification stage the features' vectors extracted in the previous stage are populated into neural network classifier to perform final classification task.

Pre-processing and face area allocation stage: In pre-processing phase a sequence of image enhancement methods are applied to make the image (that loaded to the system as input) appropriate for related information extraction task.

First the original image is subtracted from the mean image. Mean image is constructed by applying mean filter of size 5×5 on the original image. After that the offset value 128 is added to all pixels of image subtraction step, this offset addition will make the mean of produced image is mid-gray (i.e., 128). Pixels' values bounding is done to prevent become less than 0 or greater than 255. The subtraction of mean value is applied to remove light shadow effect and make the basic facial components more isolated especially in eyebrow and eye regions. The subtraction process is illustrated in Fig. 2.

The result of subtraction process contains noise of salt type and this was handled by applying the smoothing average filter; because the average filter of size 3×3 intends to replace each pixel value in an input image by the average value of its neighbors, including itself and removing the noise (Shih, 2010).

Then, the contrast of produced image is improved using linear stretching for extending the dynamic range across the whole image spectrum range (0, L-1) (Kotkar and Gharde, 2013).

Thresholding operation is applied on the result of contrast stretching to convert a multilevel image into a binary image. In this step a proper Threshold (T) is selected and used to categorize image pixels values into two levels and then separate the objects from background by making each pixel value that is greater than T equal to 0 and otherwise make it equal to 1 (Rahini and Sudha, 2014).

However the image that is resulted from binarization process may contain gaps. So, dilation operation is applied for bridging gaps when the gaps contain small pixel numbers and making a connection between objects (Gonzalez and Woods, 2002).

Finally it is easy to allocate face region and remove background region by making four separated scans (i.e., along the four directions: right, left, up and down), to capture the first hit of white pixel along each direction. Then, the locations of the four hit points are used to define the face region coordinates. The results of pre-processing phase are shown in Fig. 3.

Features vector extraction stage: This stage consists of two sub-stages, the first one is dedicated to calculate the

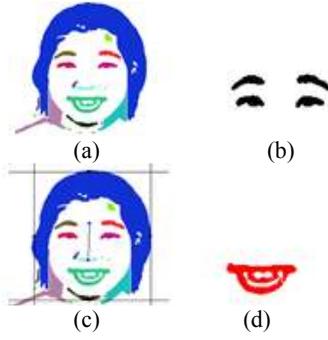


Fig. 4: (a): Segmentation process result; (b): Eyebrow and eye segments; (c): Face symmetry property and nose position; (d): Mouth segment

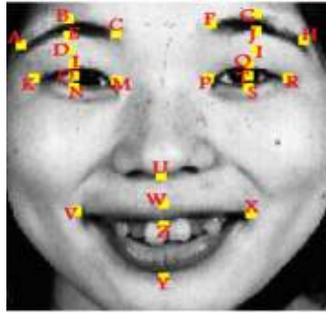


Fig. 5: Points of interest for basic facial components

best geometrical discriminating features (i.e., distances and angles). The second sub-stage is to make normalization for the determined geometrical features.

Features vector generation: In this study, we suggest the use of a set of geometric features extracted from the face regions to recognize emotion type. First, it is necessary to extract the five important facial regions: two eyebrows, two eyes, nose and mouth) in order to generate the feature vector. The binary image is segmented into N number of isolated segments using region growing method. By taking advantage of natural face symmetry and natural top-to-bottom and left-to-right order in which the features appear in the human face, we find rules to describe the shape, size, texture and other characteristics of facial features. Figure 4 illustrates the segmentation results of the binary image and the extracted basic facial components regions.

Then a set of 26 points of interest (i.e., five points for each segment except nose region) are extracted from these segments. In nose region a point that represents the light reflection area is extracted. The most left point, the most left point, lowest point in Y-access, the highest point in Y-access and the center point (i.e., the mean these four points) are generated for the four remaining regions as shown in Fig. 5.

The Distance (D) between two points $P = (x_p, y_p)$, $Q = (x_q, y_q)$ is calculated using the following equation (Weir *et al.*, 2008):

Table 1: The proposed geometric-based features vector

F _{No.}	Description	F _{No.}	Description
F ₁	∠BGV	F ₁₂	∠CKM
F ₂	∠NQX	F ₁₃	D _{SS}
F ₃	∠ABV	F ₁₄	D _{IP}
F ₄	∠BKN	F ₁₅	D _{LN}
F ₅	∠FGR	F ₁₆	∠NIX
F ₆	∠IRT	F ₁₇	∠AEY
F ₇	∠BHY	F ₁₈	∠EKM
F ₈	D _{VX}	F ₁₉	D _{CZ/D_{NW}}
F ₉	∠KNY	F ₂₀	∠LGQ
F ₁₀	D _{MG/D_{SW}}	F ₂₁	∠CKR
F ₁₁	D _{BN/D_{KX}}	F ₂₂	∠BEX

$$d_{PQ} = \sqrt{(x_q - x_p)^2 + (y_q - y_p)^2} \quad (1)$$

The angle (θ) that is enclosed between the three points P (x_p, y_p), Q (x_q, y_q), R (x_r, y_r) can be founded using the following:

$$\theta_{PQR} = \tan^{-1} \left(\frac{\Delta x_{PR} \Delta x_{QR} + \Delta y_{PR} \Delta y_{QR}}{\sqrt{\Delta x_{PR}^2 + \Delta y_{PR}^2} \sqrt{\Delta x_{QR}^2 + \Delta y_{QR}^2}} \right) \quad (2)$$

where,

$$\Delta x_{PR} = x_p - x_r \quad (3)$$

$$\Delta y_{PR} = y_p - y_r \quad (4)$$

$$\Delta x_{QR} = x_q - x_r \quad (5)$$

$$\Delta y_{QR} = y_q - y_r \quad (6)$$

The relative distance between two distances can be found as the division of one distance by another as shown in the following equation:

$$DR_{1,2} = \frac{d_1}{d_2} \quad (7)$$

To extract a new features vector that can recognize emotions effectively we generate all possible different distances, (d ()), that can be exist between any two points, without duplication. Also, all possible relative distances, (d_r ()), between the distances (d ()) are generated. The possible angles (θ ()) among any three points are also calculated. Totally, 58175 possible features of these three types are extracted.

Then, this big feature pool is reduced by selecting only the best features which show lowest within-class variations for the seven classes of expressions and have highest between class variations. Features reduction can speed up the classification process by keeping the most important classes relevant features (Reif and Shafait, 2014). Statistical Analysis process had used for feature dimensionally reduction and it is found that the features listed in Table 1 lead to the best emotions classification performance.

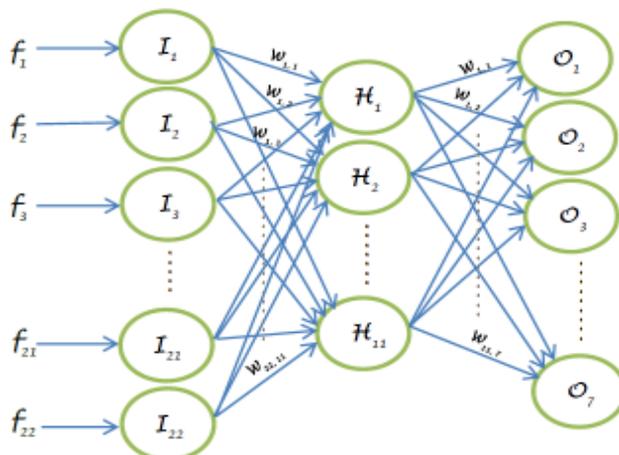


Fig. 6: FFNA model of neural network

As shown in Table 1 the best features are mostly of angle type. The main reason behind this is the distance-based features effect by size of face and this lead to size-variant representations of facial data, while angle metric saves the effort for face normalization, which is necessary for distance-based features.

The cluster analysis shows that the extracted features have more than one template for each emotion class. This is due to variability in expressing emotions from person to person and even for the same person there may be different ways to express same emotion. *ISO_DATA* clustering algorithm is used to cluster each class into K numbers of sub-classes in clusters analyzing step.

Feature vector normalization: If different features have considerably different magnitude scales then feature normalization may be necessary. Otherwise, the features on larger scales will dominate the cost-function and prevent the other features from being "effective".

One simple way to map all features into a similar scale is through using the following equation (Cowan, 2013):

$$f_n = \frac{f_r - \mu}{\sigma} \quad (8)$$

where, μ and σ are the mean and standard deviation of all classes' samples, f_r is the raw value of the considered feature (i.e., its value before normalization) and f_n is its normalized result.

Since the extracted features are of different scales then we have performed normalization step in which similar normalization range (-1, 1) is done for all features.

Classification stage: Finally, classification stage is performed to classify the extracted features vector into

appropriate emotion class. Since each class has more than one features template and the number of emotion classes is limited; the neural network classifier is adopted because it is easy and suite for such classification task, because it has good data clustering abilities and high robustness against partial input variability. Feed forward neural network with the structure shown in Fig. 6 had been used. The network is learned using supervised learning to change weights values of the features vector for achieving the best representation of emotions classes.

RESULTS AND DISCUSSION

The performance of the proposed system was tested on JAFFE (Japanese Female Facial Expressions) database. The database contains ten Japanese females' images. There are seven different facial expressions on JAFFE database, such as Neutral (NE), Happy (HA), Angry (AN), Disgust (DI), Fear (FE), Sad (SA) and Surprise (SU). Each female has posed from two to four times for each expression. Totally, there are 213 gray scale facial expression images in this database. The size of image is 256×256. However there is a problem of incorrect labeling of some class samples in this database and this degrade the recognition rate for researches that has worked on it for testing the system performance (Ahmed *et al.*, 2014). The statistical analysis results indicate that the samples shown in Fig. 7 have wrong labeling, so for this reason we have exclude them from the experiments and use only samples with true labeling which is in total 211 samples.

The conducted experiments comprise of two modules: training and testing.

The recognition rate reached to 99.05% for the seven emotion classes when all database samples were used for training neural network and then for testing

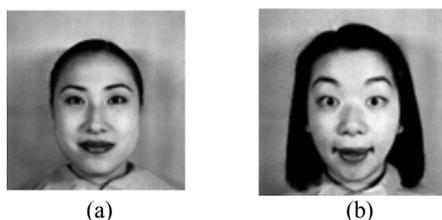


Fig. 7: Samples with incorrect labeling; (a): Happy sample in sad class; (b): Incorrect surprise sample in surprise class

Table 2: Confusion matrix for seven emotion classes classification when all database samples used for learning

Class	AN	DI	FE	HA	NE	SA	SU
AN	30	0	0	0	0	0	0
DI	0	28	1	0	0	0	0
FE	0	0	32	0	0	0	0
HA	0	0	0	31	0	0	0
NE	0	0	0	0	30	0	0
SA	0	0	0	0	0	30	0
SU	0	0	0	1	0	0	28

Table 3: Confusion matrix for seven emotion classes classification

Class	AN	DI	FE	HA	NE	SA	SU
AN	28	0	1	1	0	0	0
DI	1	26	1	1	0	0	0
FE	0	0	30	0	0	2	0
HA	0	0	0	30	0	1	0
NE	0	0	0	0	30	0	0
SA	0	0	0	0	0	30	0
SU	0	0	0	1	0	0	28

Table 4: Confusion matrix for six emotion classes classification

Class	AN	DI	FE	HA	SA	SU
AN	30	0	0	0	0	0
DI	2	26	0	1	0	0
FE	1	0	31	0	0	0
HA	0	0	0	30	1	0
SA	0	0	0	0	30	0
SU	0	0	0	0	0	29

Table 5: Comparison of recognition rate with some other persons in depended systems on JAFFE database

Author	Features extraction method	RR (%)
Juxiangand Tianwei (Zhou <i>et al.</i> , 2013)	LDP	92.14
Surbhi and Arora (2013) and Zhang <i>et al.</i> (2012a)	LBP	90.70
Castillo <i>et al.</i> (2012)	LSDP	89.20
Porawat (Visutsak, 2013)	Facial points of a lower face region	74.50
Our method	Distances and angles	95.73

system recognition abilities. Table 2 shows the confusion matrix for this case.

In order to test the system performance using part of the database images as training material; a set consists of 23 samples taken, randomly, from each class was used as part of training dataset and the remaining samples were assembled as testing sample performance (Table 2).

We have made two experiments. The first one is made to test system recognition capabilities on all the

seven emotions classes exist in the JAFFE database. The attained recognition rate reached 95.73% and the confusion matrix for this experiment is shown in Table 3.

The second conducted experiment was made after excluding neutral class and working only on the six remaining classes. The recognition rate was increased by 1.5% and reached to 97.23% as shown in Table 4.

Table 5 shows the results of comparison that is made with some other researches in a person independent environment on JAFFE database with seven emotions classes and the results show the effectiveness of the proposed system.

CONCLUSION

A new geometric based features vector has been proposed in this study. The five basic facial regions have been used to drive the points of interest through which the distances and angles metrics are applied to extract new geometric features which are populated into feed forward neural network classifier. Feature normalization step had been used to prevent small scale features from not being learnt and allow features to prove their effectiveness in classification of emotions regardless of their scales. The experiment results have been clearly showed the effectiveness of the proposed system on JAFFE database.

REFERENCES

- Ahmed, F., H. Bari and E. Hossain, 2014. Person-independent facial expression recognition based on Compound Local Binary Pattern (CLBP). *Int. Arab J. Inform. Technol.*, 11(2): 195-203.
- Butalia, A., A.K. Ramani and P. Kulkarni, 2010. Emotional recognition and towards context based decision. *Int. J. Comput. Appl.*, 9(3): 42-54.
- Castillo, J.A.R., A.R. Rivera and C. Oksam, 2012. Facial expression recognition based on local sign directional pattern. *Proceeding of the 19th IEEE International Conference on Image Processing (ICIP, 2012)*, pp: 2613-2616.
- Chen, J., Z. Chen, Z. Chi and H. Fu, 2014. Facial expression recognition based on facial components detection and HOG features. *Proceeding of the International Workshops on Electrical and Computer Engineering Subfields.*, Koc University, Istanbul/Turkey.
- Cowan, M.K., 2013. *Machine Learning*. Creative Commons Attribution. Retrieved from: <http://ctan.imsc.res.in/graphics/pgf/contrib/neuralnetwork/examples/neural-networks-ebook.pdf>.
- Dutta, S. and V.B. Baru, 2013. Review of facial expression recognition system and used datasets. *Int. J. Res. Eng. Technol.*, 2(12): 641-645.

- Gonzalez, R.C. and R.E. Woods, 2002. Digital Image Processing. 3rd Edn., Pearson Education, Prentice Hall, USA.
- Iatraki, G., 2009. Emotional facial expression recognition and classification. M.A. Thesis, Department of Media and Knowledge Engineering, Delft University of Technology.
- Jain, G., 2011. Emotion recognition from eye region signals using local binary patterns. M.A. Thesis, Department of Electrical and Computer Engineering, University of Toronto.
- Kotkar, V.A. and S.S. Gharde, 2013. Review of various image contrast enhancement techniques. *Int. J. Innov. Res. Sci. Eng. Technol.*, 2(7): 2786-2793.
- Prasad, M. and A. Danti, 2014. Classification of human facial expression based on mouth feature using SUSAN edge operator. *Int. J. Adv. Res. Comput. Sci. Softw. Eng.*, 4(12): 440-443.
- Rahini, K.K. and S.S. Sudha, 2014. Review of image segmentation techniques: A survey. *Int. J. Adv. Res. Comput. Sci. Softw. Eng.*, 4(7): 842-845.
- Rani, J. and K. Garg, 2014. Emotion detection using facial expressions: A review. *Int. J. Adv. Res. Comput. Sci. Softw. Eng.*, 4(4): 465-467.
- Reif, M. and F. Shafait, 2014. Efficient feature size reduction via predictive forward selection. *Pattern Recogn.*, 47: 1664-1673.
- Saini, R. and N. Rana, 2014. Facial expression recognition techniques, database & classifiers. *Int. J. Adv. Comput. Sci. Commun. Eng. (IJACSCE)*, 2(2): 15-20.
- Shih, F.Y., 2010. Image Processing and Pattern Recognition: Fundamentals and Techniques, Wiley-IEEE Press, New York.
- Sonka, M., V. Hlavac and R. Boyle, 2008. Image Processing, Analysis and Machine Vision. 3rd Edn., Thomson/West, Toronto.
- Sumathi, C.P., T. Santhanam and M. Mahadevi, 2012. Automatic facial expression analysis a survey. *Int. J. Comput. Sci. Eng. Survey (IJCSES)*, 3(6): 47-59.
- Surbhi and V. Arora, 2013. The facial expression detection from human facial image by using neural network. *Int. J. Appl. Innov. Eng. Manage. (IJAIEM)*, 2(6): 126-130.
- Visutsak, P., 2013. Emotion classification through lower facial expressions using adaptive support vector machines. *J. Man Mach. Technol.*, 2(1): 12-20.
- Wang, H., H. Huang and F. Makedon, 2013. Emotion detection via discriminant laplacian embedding. *Univ. Access. Inf. Soc.*, 13: 23-31.
- Weir, M.D., J. Hass and F.R. Giordano, 2008. Thomas Calculus. 11th Edn., Pearson Education.
- Zhang, S., X. Zhao and B. Lei, 2012a. Facial expression recognition using sparse representation. *WSEAS Trans Syst.*, 11(8): 440-452.
- Zhang, S., X. Zhao and B. Lei, 2012b. Facial expression recognition based on local binary patterns and local fisher discriminant analysis. *WSEAS T. Signal Process.*, 8(1): 21-31.
- Zhou, J., T. Xu and J. Gan, 2013. Facial expression recognition based on local directional pattern using SVM decision-level fusion. *ASTL*, 17: 126-132.