An SOM-based Automatic Facial Expression Recognition System

Mu-Chun Su[1], Chun-Kai Yang[1], Shih-Chieh Lin[1],De-Yuan Huang[1], Yi-Zeng Hsieh[1], andPa-Chun Wang[2]

[1]Department of Computer Science &InformationEngineering,National Central University,Taiwan, R.O.C.
[2]Cathay General Hospital, Taiwan, R.O.C.
E-mail: muchun@csie.ncu.edu.tw

Abstract

Recently, a number of applications of automatic facial expression recognition systems havesurfaced in many different research fields. The automatic facial expression recognition problem is a very challenging problem because it involves in three sub-problems: 1) face detection, 2) facial expression feature extraction, and 3) expression classification. This paper presents an automatic facial expression recognition system based on self-organizing feature maps, which provides an effective solution to the aforementioned three sub-problems. The performance of the proposed system was computed on twowell-known facial expression databases. The average correct recognition rates were over 90%.

Keywords

Facial expression recognition, SOM algorithm, face detection.

1. INTRODUCTION

Automatic facial expression recognition systems have been applied to many practical application fields such as social robot interactions, human-computer interactions, human behavior analysis, virtual reality, etc. Thus in recent years, the study of automatic facial expression recognition has become a more and more important research topic for many researchers from different research fields [1]-[23].

The human face is an elastic object that consists of organs, numerous muscles, skins, and bones. When a muscle contracts, the transformation of the corresponding skin area attached to the muscle result in a certain type of visual effect. Although the claim that theredo exist universally basic emotions across genders and races has not been confirmed, most of the existing vision-based facial expression studies accept the assumption defined Ekman about the universal categories of emotions (i.e., happiness, sadness, surprise, fear, anger, and disgust) [24]. A human-observer-based system called the Facial Action Coding System (FACS) has been developed to facilitate objective measurement of subtle changes in facial appearance caused by contractions of the facial muscles [25]. The FACS is able to give a linguistic description of all visibly discriminable expressions via 44 action units.

DOI :10.5121/ijscai.2013.2305

The automatic facial expression recognition problem is a very challenging problem because it involves in three sub-problems: 1) face detection, 2) facial expression feature extraction, and 3) expression classification. Each of the sub-problems is a difficult problem to be solved due to many factors such as cluttered backgrounds, illumination changes, face scales, pose variations, head or body motions, etc. An overview of the research work in facial expression analysis can be found in [26]-[28]. The approaches to facial expression recognition can be divided into two classes in many different ways. In one way, they can be classified into static-image-based approaches (e.g., [10]) and image sequence-based approaches (e.g., [2], [7]-[8], [11],[17], [20]-[22], etc). While the static-image-based approach classifies expressions based on a single image, the image sequence-based approach utilizes the motion information in an image sequence. In another way, they can be classified into geometrical feature-based approaches (e.g., [1], [7], [15], etc) and appearance-based approaches (e.g., [12], [16], etc). The geometrical feature-based approach relies on the geometric facial features such as the locations and contours of eyebrows, eyes, nose, mouth,etc. As for the appearance-based approach, the whole-face or specific regions in a face image are used for the feature extraction via some kinds of filters or transformations. Some approaches can fully automatically recognize expressions but some approaches still need manual initializations before the recognition procedure.

In this paper we propose a simple approach to implement an automatic facial expression recognition system based on self-organizing feature maps (SOMs). The SOM algorithm is a well-known unsupervised learning algorithm in the field of neural networks [29]. A modified self-organizing feature map algorithm is developed to automatically and effectively extract facial feature points. Owing to the introduction of the SOMs, the motion of facial features can be more reliably tracked than the methods using a conventional optical flow algorithm. The remaining of this paper is organized as follows. A detailed description of the proposed expression recognition algorithm is given in Section 2. Then simulation resultsare given in Section 3.Finally, Section 4 concludes the paper.

The Proposed Facial Expression Recognition System

The proposed automatic facial expression recognition system can automatically detect human faces, extract facial features, and recognize facial expressions. The inputs to the proposed automatic facial expression recognition algorithm are a sequence of images since dynamic images can provide more information about facial expressions than a single static image.

2.1 Face Detection

The first step for facial expression recognition is to solve the face detection sub-problem. Face detection determines the locations and sizes of faces in an input image. Automatic human face detection is not a trivial task because face patterns can have significantly variable image appearances due to many factors such as hair styles, glasses, and races.In addition, the variations of face scales, shapes and poses of faces in images also hinder the success of automatic face detection systems. Several different approaches have been proposed to solve the problem of face detection [30]-[33]. Each approach has its own advantages and disadvantages. In this paper, we adopt the method proposed by Viola and Jones to detect faces from images [34]. This face detection method can minimize computational time while achieving high detection accuracy.

After a human face is detected, the proposed system adopts a composite method to locate pupils. First of all, we adopt the Viola-Jones algorithm discussed in [34] to locate the regions of eyes. One problem associated with the Viola-Jones algorithmis that it could effective locate the regions of eyes but could not precisely locate the centers of the pupils. Therefore, we need to fine-tune the eye regions to precisely locate the pupils. Weassume that the pupil regions are the darkest regions in the eye regions detected by the Viola-Jones algorithm.The segmentation task for locating the pupilscan be easily accomplished if the histogram of the eye regions presents two obvious peaks; otherwise, correct threshold selection is usually crucial for successful threshold segmentation. We adoptthep-tile thresholding technique to automatically segment the pupil regions from the eye regions [35]. From our many simulation results, we found that the ratio between the pupil regions and the remaining eye regions could be chosen to be 1/10 (i.e., p = 10).

After the pupils have been located, the face image is rotated, trimmed, and normalized to be an image with the size80×60.We rotate the detected face image to make the pupils lie on the horizontal axe. In addition, we normalize the distance between the two pupils to be 25 pixels. Fig. 1 illustrates an example of a normalized face image.

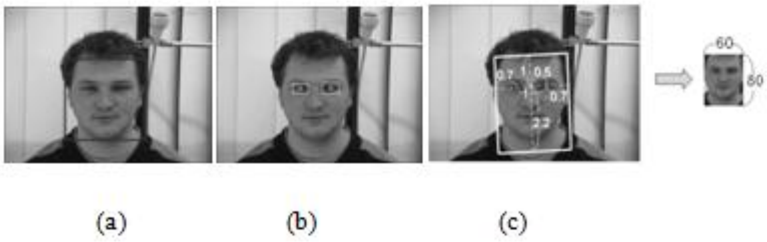


Fig. 1.An example of face detection and normalization. (a) Detected face. (b) Detected eyes. (c) Rotated, trimmed, and normalized face.

2.2Facial Feature Extraction

After the face in the first image frame has been detected, the next step is to extract necessary information about the facial expression presented in the image sequence. Facial features can be categorized into many different classes [26]-[28]. In general, there are two types of facial features can be extracted: geometrical features and appearance features [36]. While the appearance features can be extracted on either the whole face or some specific regions via some kinds of filters (e.g., Gabor wavelets filter), geometrical features focus on the extraction of shapes and locations of intransient facial features (e.g., eyes, eyebrows, nose, and mouth). In our system, geometrical features are extracted for facial expression recognition.

The movements of the facial featuressuch as eyebrows, eyes, and the mouth have a strong relation to the information about facial expressions; however, the reliable extraction of the exact locations of the intransientfacial features sometimes is a very challenging task due to many disturbing factors (e.g., illumination factor, noise). Even if we can accurately locate the facial features, we still encounter another problem about the extraction of the motion information of the facial features.

One simple approach to solvethe aforementioned two problems is to place a certain number of landmark points around the located facial feature regions and then use a tracking algorithm to track those landmark points to compute the displacement vectors of those points. However, this approach has to break some bottlenecks. The first bottleneck is that how and where to automatically locate the landmark points. Accurate locations of landmark points usually require intensive computational resources; therefore, some approaches adopted an alternative method to compute motion information in meshes or grids which cover some important intransientfacial features (e.g., potential net [9], uniform grid [20], Candide wireframe [22]). Another bottleneck is that the performance of the adopted tracking algorithm may be sensible to some disturbing factors such as illumination changes, head or body motions. This problem usually results in the phenomenon that severallandmarks points will be erroneously tracked to some far away locations.

To encounter the aforementioned two problems, we proposed the use of self-organizing feature maps (SOMs)[29]. The SOM algorithm is one of the most popular unsupervised learning algorithms in the research field of neural networks. Recently, numerous technical reports have been written about successful applications of the SOMs in a variety of problems.The principal goal of SOMs is to transform patterns of arbitrary dimensionality into the responses of one- or two-dimensional arrays of neurons, and to perform this transform adaptively in a topological ordered fashion.

In our previous work, we built a generic face model from the examinations of a large number of faces [20]. Based on the generic face model proposed in [20], we further proposed a normalized generic face model as shown in Fig. 2. Although geometric relations between the eyes and the mouth vary a little bit from person to person, the eyes, eyebrows, the nose, and the mouth are basically enclosed in the three rectangles and the pentagon as shown in Fig. 2.We adopt a pentagon instead of a rectangle is to make the recognition performance as insensible to beards as possible. This observation was concluded from our many simulation results.

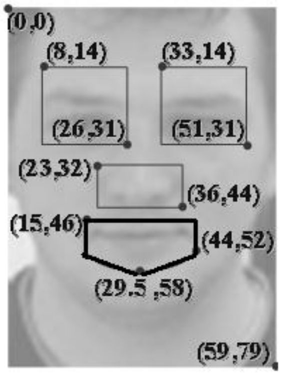


Fig. 2.A normalized generic face model.

After we have located the four critical regions (i.e., the eyes, the nose, and the mouth), the next step is to extract the motion information of these facial features.

To encounter the aforementioned two problems, we proposed the use of SOMs. Four nets with the sizes,6×6, 6×6, 7×7, and 4×4, are placed around the regions of the two eyes, the nose, and the mouth, respectively. In total, there are 137 neurons in the four SOMs. A modified SOM algorithm is developed to automatically and effectively extract facial feature points. The approximated gradient of each pixel inside the corresponding facial region is used as an input pattern to the modified SOM algorithm. The modified SOM algorithm is summarized as follows:

Step 1. Initialization:In the conventional SOM algorithm, weight vectors are usually randomly initialized. Instead of adopting the random initialization scheme, we initialize the weight vectors, $\vec{w}_j = (w_{j1}, w_{j2})^T$,$j = 1, \ldots, M \times M$, to lie within a rhombus as shown in Fig. 3(a). From many simulations, we found that the use of rhombuses was more efficient than the use of rectangles, not to mention the initialization scheme.

Step 2. Winner Finding:Instead of directly presenting the gray level of each pixel, we present the approximated gradient of each pixel,,$\vec{x}_j = (G_x(j), G_y(j))^T$, to the network and find the winning neuron. The two gradients,,$G_x(j)$and $G_y(j)$, represent the row edge gradient and the column gradient at the jth pixel, respectively. In our system, the Sobel operator was adopted for the computation of the gradients. The neuron with the largest value of the activation function is declared the winner for the competition. The winning neuron j^* at time *k* is found by using either the maximum output criterion or the minimum-distance Euclidean criterion:

$$j^* = Arg \min_{1 \le j \le M \times M} \left\| \vec{x}(k) - \vec{w}_j(k) \right\| \tag{1}$$

where $\vec{x}(k) = [x_1(k), x_2(k)]^T = [G_x(k), G_y(k)]^T$ represents the kth input pattern corresponding to a pixel located in the face feature region (i.e., the eye region, nose region, and the mouth region), $M \times M$ is the network size, and $\|\cdot\|$ indicates the Euclidean norm. For example, we input the gradients of the pixels inside the rectangle (i.e., the region defined by $[23,36] \times [32,44]$ in Fig. 2) which enclose the nose to the neural network with the size 7×7.

Step 3. Weight Updating: Adjust the weights of the winner and its neighbors using the following rule:

$$\vec{w}_j(k+1) = \vec{w}_j(k) + s_j \eta(k) \Lambda_{j^*,j}(k)[\vec{x}(k) - \vec{w}_j(k)] \quad for\, 1 \le j \le M \times M \tag{2}$$

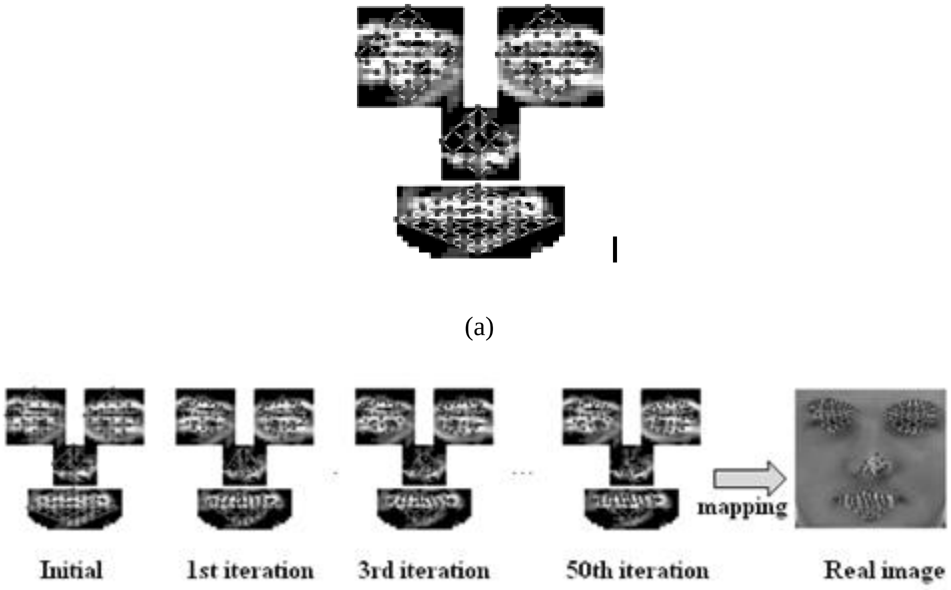
$$\Lambda_{j^*,j}(k) = \exp\left(-\frac{d_{j^*,j}^2}{2\sigma^2(k)} \right) \tag{3}$$

$$s_j = \begin{cases} 1 & \text{if } |G_x(j)| + |G_y(j)| \ge 255 \\ 1 \big/ |G_x(j)| + |G_y(j)| & \text{if } |G_x(j)| + |G_y(j)| < 255 \end{cases} (4)$$

where $\eta(k)$ is a positive constant, $d_{j*,j}$denotes the lateral distance of neuron j from the winning neuron j*, $\sigma(k)$ is the “effect width” of the topological neighborhood , and $\Lambda_{j^*,j}(k)$ is the topological neighborhood function of the winner neuron j^* at time k. The parameter sj is a weighting factor for the learning rate $\eta(k)$. It was introduced to make the learning rate larger when the absolute value of theapproximated gradient of each pixel is large (e.g., larger than or equal to 255).We assume that pixels with high gradient convey more facial featureinformation. Due to the introduction of the weighting factor, the weight vectors of the network can quickly converge to important pixels on the corresponding facial regions as shown in Fig. 3(b).

Step 4.Iterating: Go to step 2 until a pre-specified number of iterations is achieved or some kind of termination criterion is satisfied.

After sufficient training, each weight vector in a trained SOM corresponds toa landmark point in the corresponding facial region as shown in Fig. 4.



(a)

(b)

Fig. 3 TheSOM training procedure. (a) The use of rhombus for the initial weight vectors. (b) The example of a trained SOM in 50 iterations.

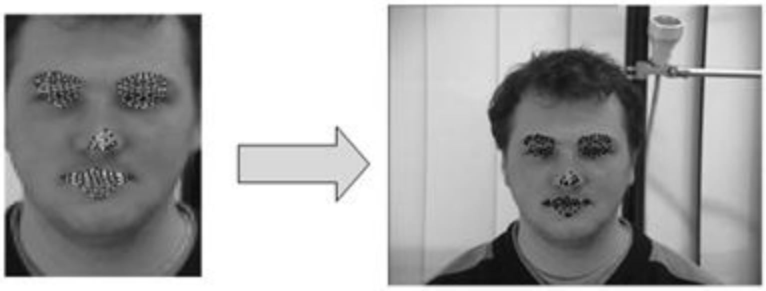


Fig. 4.The correspondence between the trained SOMs and the landmark points in the facial regions.

2.3 Landmark Point Tracking

To track landmark points, we adopt a two-stage neighborhood-correlation optical flow tracking algorithm. At the first stage, we adopt the optical flow method to automatically track the 137 landmark points in the image sequence. Similar to the article [7], we adopt the cross-correlation optical flow method. Cross-correlation of a T×T template in the previous image, and a W×W searching window at the present image iscalculated and the position with the maximum cross-correlation value which is larger than a pre-specified threshold is located at the present image. The accuracy of the cross correlation method is sensitive to the illumination change, noise, template size, moving speed, etc. Due to these disturbing factors, landmark points with correlation values smaller than the pre-specified threshold are apt to result in tracking errors; therefore, we cannot use the positions directly computed by the cross correlation method. To provide an acceptable solution to the prediction of the positions of those points with small correlation values, we propose to fully use the topology-preserving property of the SOM.The assumption made by us is that nearby landmark points in a facial region move, to a certain extent, coordinately. For a landmark point with a low correlation value, we use the average location of the positions of its neighbors with correlation values which are larger than the threshold. For each landmark, the information of its neighbors is already embedded in the trained SOMs as shown in Fig. 4.

2.4Expression Recognition

There are total 137 neurons in the four regions. Basically, the displacement vectors of these 137landmark points located on the SOMs are used for the facial expression recognition. The displacement of each landmark point is calculated by subtracting its original position in the first image from the final position in the last image of the image sequence. We cannot directly feed these 137 displacement vectors into a classifier for facial expression because the sizes of facial feature regions vary from person to person.In addition, head movement may affect the displacements. The 137 displacements have to be normalized in some way before they are inputted to a classifier. To remedy the problem of head movement, we use the average displacement vector of the 16 landmark points corresponding to the 16 neurons in the network with the size in the region of nose to approximate the head displacement vector. Then all displacement vectors are subtracted from the head displacement vector. In the following, the remaining 136 displacement vectors are re-sampled to 70 average displacement vectors (as shown in Fig. 5) in order to make the recognition system be person independent. We take the left eye region for example to illustrate how we re-sample the 36 displacement vectors located in the left eye region. First of all, we find a rectangle to circumscribe the 36 landmark points in the left eye region. Then we dichotomize the rectangle into 20 small rectangles. The average displacement vector of those landmark points lying in the same small rectangle is computed. Therefore, there are 20 displacement vectors to represent the left eye region. The same re-sampling procedure is applied to the right eye region and the mouth region. Since the mouth region is larger than the eye region, we use 30 small rectangles in the mouth region. Finally, there are totally 70 normalized displacement vectors as shown in Fig. 5.

Finally, a multi-layer perceptron (MLP) with the structure140×10×10×7was adopted for the classification of the seven expressions including six basic facial expressions (i.e., happiness, sadness, surprise, fear, anger, and disgust) and a neutral facial expression. The structure of the MLP was chosen from many simulation results.

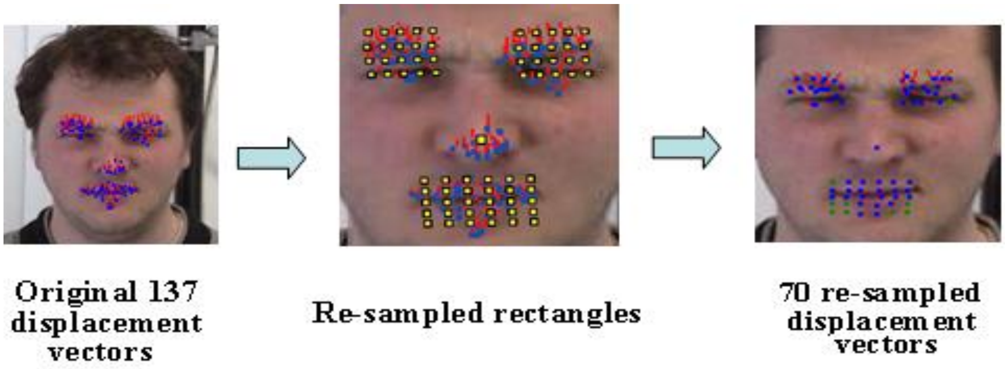

Original 137 displacement vectors
Re-sampled rectangles
70 re-sampled displacement vectors

Fig. 4. The final 70displacement vectors re-sampled from 137 displacement vectors.

3.SIMULATION RESULTS

The performance of the proposed system was tested on the well-known Cohn-KanadeDatabase [37]-[38] and FG-NETdatabase from the Technique University of Munich [39]. The Cohn-KanadeDatabasecurrently contains 2105 digitized image sequences performed by 182 adult subjects.This database has been FACS (Facial ActionCoding System) coded.The FG-NET databaseis an image database containing face images showing a number of subjects performing the six different basic emotions. The database contains material gathered from 18 different individuals. Each individual performed all six desired actions three times. Additionally three sequences doing no expressions at all are recorded. In total, there are 399 sequences in the database.

To provide accurately labeled sequences for training the MLP to recognize 7 facial expressions (i.e., happiness, sadness, surprise, fear, anger, disgust, and neutral), we asked 13 subjects to visually evaluate the two databases and then label each sequence to a certain expression. Via the majority consensus rule, we finally selected 486 image sequences from the Cohn-KanadeDatabase and 364 sequences from the FG-NET database, respectively.

The training data set was consisted of the 75% of the labeled data set and the remaining data was used to generate the testing data.For the Cohn-Kanade database, the recognition results were tabulated in Tables1-2. The average correct recognition ratios were 94% and 91% for the training data and testing data, respectively. The recognition results for FG-NET database were tabulated in Tables3-4.The average correct recognition ratios were 88.8% and 83.9% for the training data and testing data, respectively. Comparisons with other existing methods are shown in Table 5. Table 5 shows that the performance of the proposed SOM-based facial expression recognition system was comparable to those existing methods.

4. CONCLUSION

In this paper, anSOM-based automatic facial expression recognitionis presented. The proposed system is able to automatically detect human faces, extract feature points, and perform facial expression recognition from image sequences. First of all, the method proposed by Viola and Jones was used to detect a face from an image. After a human face is detected, a composite method was proposed to locate pupils so that the located face image can be rotated, trimmed, and normalized to be an image with the size 80×60. To alleviate the computational load for extracting facial features, we propose the use of SOMs. In the following section we adopt a two-stage neighborhood-correlation optical flow tracking algorithm to track facial features. Finally, a multi-layer perceptron (MLP) with the structure 140×10×10×7 was adopted for the classification of the seven expressions including six basic facial expressions (i.e., happiness, sadness, surprise, fear, anger, and disgust) and a neutral facial expression.Simulation results showed that the performance of the proposed SOM-based facial expression recognition system was comparable to those existing methods.

ACKNOWLEDGMENTS

This paper was partly supported by the National Science Council, Taiwan, R.O.C, under NSC 101-2221-E-008-124-MY3 and NSC 101-2911-I-008-001, CGH-NCU Joint Research Foundation 102NCU-CGH-04, and LSH-NCU Joint Research Foundation 102NCU-CGH-04. Also, the authors would like to thank Dr. Cohn et al. for providing us their image sequences in the Cohn-Kanade AU-Coded Facial Expression Database.

REFERENCES

[1] G. W. Cottrell and J. Metcalfe, "EMPATH: Face, gender, and emotion recognition using holons," Advances in Neural Information Processing Systems, vol. 3, pp. 564-571,1991.

[2] K. Mase, "Recognition of facial expression from optical flow," IEICE Trans., vol. E74, no. 10, pp. 3474-3483,1991.

[3] D. Terzopoulos and K. Waters, "Analysis and synthesis of facial image sequences using physical and anatomical models," IEEE Trans. Pattern Anal. Machine Intell., vol. 15, pp. 569-579,1993.

[4] I. A. Essa and A. Pentland, "A vision system for observing and extracting facial action parameters," in Proc. Computer Vision and Pattern Recognition, pp. 76-83, 1994.

[5] K. Matsuno, C. Lee, and S. Tsuji, "Recognition of human facial expressions without feature extraction," ECCV, pp. 513-520, 1994.

[6] T. Darrel, I. Essa, and A. P. Pentland, "Correlation and interpolation networks for real-time expression analysis/synthesis,"Advancesin Neural Information Processing Systems (NIPS) 7, MIT Press,1995.

[7] Y. Yacoob and L. D. Davis, "Recognizing human facial expressions from long image sequences using optical flow," IEEE Trans.on Pattern Analysis and Machine Intelligence, vol. 18, no. 6, pp. 636-642, 1996.

[8] M.Rosenblum,Y. Yacoob,L.S.Davis, "Human expression recognition from motion using a radial basis function network architecture,"IEEE Trans. on Neural Networks, vol. 7, no. 5, pp. 1121-1138, 1996.

[9] S. Kimura and M. Yachida, "Facial expression recognition and its degree estimation," in Proc. Computer Vision and Pattern Recognition, pp. 295-300, 1997.

[10] C. L. Huang and Y. M. Huang, "Facial expression recognition using model-based feature extraction and action parameters classification," Journal of Visual Communication and Image Representation, vol. 8, no. 3, pp. 278-290, 1997.

[11] T. OtsukaandJ.Ohya, "Spotting segments displaying facial expression from image sequencesusing HMM," in Proc. IEEE Conf. onAutomatic Face and Gesture Recognition, pp. 442-447, Apr. 1998.

[12] M.S. Bartlett, J.C. Hager, P. Ekman, and T.J. Sejnowski, "Measuring Facial Expressions by Computer Image Analysis," Psychophysiology, vol. 36, pp. 253-263, 1999.

[13] G. Donato, M.S. Bartlett, J.C. Hager, P. Ekman, and T.J. Sejnowski, "Classifying Facial Actions," IEEE Trans. on Pattern Analysis and Machine Intelligence, vol. 21, no.10,1999.

[14] J. J. Lien, T. Kanade, J. Cohn, and C. Li, "Detection, tracking, and classification of action units in facial expression,"Journal of Robotics and Autonomous Systems, vol. 31, no. 3, pp. 131-146, 2000.

[15] Y. l. Tian, T. Kanade, and J. F. Cohn, “Recognizing Action Units for Facial Expression Analysis,” IEEE Trans. on Pattern Analysis and Machine Intelligence, vol. 23, no. 2, 2001.

[16] Y. l. Tian,T. Kanade, and J. F. Cohn, “Evaluation of Gabor-Wavelet-Based Facial Action Unit Recognitionin Image Sequences of Increasing Complexity,” in Proc. of the Fifth IEEE International Conference on Automatic Face and Gesture Recognition, pp. 229-234, 2002.

[17] M. Yeasin, B. Bullot, and R. Sharma, “Recognition of facial expressions and measurement of levels of interest from video,”IEEE Trans. on Multimedia, vol. 8, pp. 500-508, June 2006.

[18] S. Kumano, K. Otsuka, J. Yamato, E. Maeda, and Y. Sato, “Pose-invariant facial expression recognition using variable-intensity templates,” inProc. ACCV'07, 2007, vol. 4843, pp.324-334, 2007.

[19] J. Wang and L. Yin, “Static topographic modeling for facial expression recognition and analysis,” Computer Vision and Image Understanding, vol. 108, no. 1-2, pp. 19-34, Oct. 2007.

[20] M. C. Su, Y. J. Hsieh, and D. Y. Huang, "Facial Expression Recognition using Optical Flow without Complex Feature Extraction," WSEAS Transactions on Computers, vol. 6, pp. 763-770, 2007.

[21] I. Kotsia and I. Pitas, “Facial expression recognition in image sequences using geometric deformation features and support vector machines,” IEEE Trans. On Image Processing, vol. 16.No. 1, pp. 172-187, 2007.

[22] P. Wanga, F. Barrettb, E. Martin, M. Milonova, R. E. Gur, R. C.Gur, C. Kohler, and R.Verma, “Automated video-based facial expression analysis of neuropsychiatric disorders,” Neuroscience Methods, vol. 168, pp. 224-238, Feb. 2008.

[23] I.Kotsia, I.Buciu, and I. Pitas, “An analysis of facial expression recognition under partial facial image occlusion,”Image and Vision Computing,vol. 26, no. 7, pp. 1052-1067, July 2008.

[24] P. Ekman and W. V. Friesen, “Constants across cultures in the face and emotion,” Journal of Personality and Social Psychology, vol. 17, pp. 124-129, 1971.

[25] P. Ekman and W.V. Friesen, Facial Action Coding System(FACS), Consulting Psychologists Press, 1978.

[26] M. Pantie and L. J. M. Rothkrantz, “Automatic analysis of facial expressions: the state of the art,” IEEE Trans. on Pattern Analysis and Machine Intelligence, vol. 22, no. 12, pp. 1424-1445, 2000.

[27] B. Fasel and J. Luettin, “Automatic facial expression analysis: a survey,” Pattern Recognition, vol. 36, no. 1, pp. 259-275,2003.

[28] Z. Zeng, M. Pantic, G. I. Roisman, T. S. Hung, “A survey of afftect recognition method: audio, visual, and spontaneous expressions,”IEEE Trans. on Pattern Analysis and Machine Intelligence, vol. 31, no. 1, pp. 39-58, 2009.

[29] T. Kohonen, Self-Organizing Maps. Springer-Verlag, Berlin, 1995.

[30] S.H. Lin, S.Y. Kund, and L.J. Lin, “Face recognition / detection by probabilistic decision-based neural network,”IEEE Trans on Neural Networks, vol. 8, no. 1, pp. 114-132, 1997.

[31] H. A. Rowley, S. Baluja, and T. Kanade, “Neural network- based face detection,”IEEE Trans. on Patt. Anal.And Mach. Intell., vol. 20, no. 1, pp. 23-38, 1998.

[32] K. K. Sung and T. Poggio, “Example-based learning for view-based human face detection,”IEEE Trans on Patt.Anal.And Mach. Intell., vol. 20, no. 1, pp. 39-51, 1998.

[33] M. C. Su and C. H. Chou, “Associative- memory-based human face detection,” IEICE Trans. on Fundamentals of Electronics, Communications and Computer Sciences, vol. E84-D, no. 8, pp. 1067-1074, 2001.

[34] P. Viola and M. J. Jones,“Rapid object detection using a boosted cascade of simple features,”in Proc. of the IEEE Computer Society International Conference on Computer Vision and Pattern Recognition,vol. 1, pp. 511-518,2001.

[35] M. Sonka, V. Hlvac, and R. Boyle, Image Processing, Analysis, and Machine Vision, PWS Publishing, 1999.

[36] Y. l. Tian, T. Kanade, and J. F. Cohn, “Facial expression analysis,” Handbook of Face Recognition, S. Z. Li and A. K. Jain, eds., Chap. 11, pp. 247-276, 2001.

[37] T. Kanade, J. Cohn,and Y. Tian, “Comprehensive Database for Facial Expression Analysis,” in Proc. of the Fourth IEEE International Conference on Automatic Face and Gesture Recognition, pp. 45-63, 2000.

[38] Cohn-Kanade AU-Coded Facial Expression Database. [Online]. Available: http://vasc.ri.cmu.edu/idb/html/face/facial_expression/index.html July 2008 [date accessed]

[39] Databasewith Facial Expressions and Emotions from the Technical University Munich. [Online]. Available: http://www.mmk.ei.tum.de/~waf/fgnet/ July 2008 [date accessed]

[40] S. Hadi, A. Ali, and K. Sohrab, "Recognition of six basic facial expressions by feature-points tracking using RBF neural network and fuzzy inference system," in Proc. of the IEEE Int. Conf.on Multimedia and Expo, 2004,vol. 2, pp. 1219-1222.

[41] F. Wallhoff, B. Schuller, M. Hawellek, and G. Rigoll, "Efficient recognition of authentic dynamic facial expressions on the Feedtum Database," in IEEE Int. Conf. on Multimedia and Expo, July 2006, pp. 493-496.

Table.1. The recognition performance for the training data set from the Cohn-KanadeDatabase. Su: Surprise, F: Fear, H: Happy, Sa: Sad, D: Disgust, A: Angry, and N: Neutral.

NN Real	Su	F	H	Sa	D	A	N	Recognition rate
Su	**70**	1	0	0	0	0	0	98.6%
F	0	**24**	0	0	1	0	2	88.9%
H	0	0	**77**	0	0	0	1	98.7%
Sa	0	0	0	**52**	1	2	1	92.9%
D	0	1	0	0	**33**	5	2	80.5%
A	0	1	0	1	0	**38**	1	92.7%
N	0	0	0	0	1	1	**48**	96.0%
Average	94%							

Table.2. The recognition performance for the testing data set from the Cohn-KanadeDatabase. Su: Surprise, F: Fear, H: Happy, Sa: Sad, D: Disgust, A: Angry, and N: Neutral.

NN Real	Su	F	H	Sa	D	A	N	Recognition rate
Su	**20**	0	0	0	0	0	2	91.0%
F	0	**7**	0	0	0	2	0	77.8%
H	0	0	**25**	0	0	0	2	92.6%
Sa	0	0	0	**18**	1	0	0	94.7%
D	0	1	0	0	**12**	1	0	85.7%
A	0	0	0	0	0	**12**	2	85.7%
N	0	0	0	0	0	0	**17**	100.0%
Average	91%							

Table.3. The recognition performance for the training data set from the FG-NETDatabase. Su: Surprise, F: Fear, H: Happy, Sa: Sad, D: Disgust, A: Angry, and N: Neutral.

NN Real	Su	F	H	Sa	D	A	N	Recognition rate
Su	**37**	1	0	0	0	0	2	92.5%
F	0	**12**	0	1	0	0	9	54.5%
H	0	0	**42**	0	0	0	0	100.0%
Sa	0	0	0	**36**	0	0	3	92.3%
D	0	2	0	1	**32**	2	2	82.1%
A	0	0	0	0	0	**30**	6	83.3%
N	0	0	0	0	0	0	**42**	100.0%
Average	88.8%							

Table.4. The recognition performance for the testing data set from the FG-NETDatabase. Su: Surprise, F: Fear, H: Happy, Sa: Sad, D: Disgust, A: Angry, and N: Neutral.

NN Real	Su	F	H	Sa	D	A	N	Recognition rate
Su	**9**	0	0	0	0	0	1	90.0%
F	0	**6**	0	0	0	0	2	75.0%
H	0	0	**12**	0	1	0	2	80.0%
Sa	0	0	0	**12**	0	0	2	85.7%
D	0	0	1	0	**10**	0	2	76.9%
A	0	0	0	1	0	**11**	1	84.6%
N	0	0	0	1	0	0	**13**	92.9%
Average	83.9%							

Table 5.Comparisons with other existing methods.

Method	**Recognition Results**		**Database**	**Features**	**Classifier**
	Expressions	**Recognition**			
Our method	7	93.2%	Cohn-Kanade (486 sequences:3/4 for training and 1/4 for testing)	Modified SOM	MLP
		87.6%	FG-NET (364 sequences)		MLP
Su et al. [20]	5	95.1%	Cohn-Kanade (486 sequences)	Uniform grids	MLP
Yeasin et al. [17]	6	90.9%	Cohn-Kanade (488 sequences)	Grid points	HMMs
Kotsia et al. [21]	6	91.6%	Cohn-Kanade and JAFFE (leave- 20% out cross-validation)	Texture model	SVM
Seyedarabiet al. [40]	6	91.6%	Cohn-Kanade (43 subjects for training and 10 subjects for testing)	Manually label	RBF
Wallhoff et al. [41]	7	61.7%	FG-NET (5-fold cross-validation)	2D-DCT	SVM