A Simple Approach to Facial Expression Recognition

MU-CHUN SU, YI-JWU HSIEH, AND DE-YUAN HUANG
Department of Computer Science & Information Engineering
National Central University
TAIWAN, R.O.C.
http://cilab.csie.ncu.edu.tw/

Abstract: - Human face-to-face communication plays an important role in human communication and interaction. In recent years, several different approaches have been proposed for developing methods of automatic facial expression analysis. In this paper, a simple approach to automatic facial expression recognition is presented. The proposed system is able to automatically perform human face detection, feature point extraction and facial expression recognition from image sequences. The facial recognition procedures involves in two stages. At the first stage, three multi-layer perceptrons (MLPs) are separately trained to recognize action units involving in the eyebrows, the eyes and the mouth regions. Then individual expression network was trained to recognize five basic facial expressions based on the outputs computed from the aforementioned three MLPs. Experiments were conducted to test the performance of the proposed facial expression recognition system.

Key-Words: - facial expression analysis, emotion, face detection, facial action units, expression recognition

1 Introduction

Visual communication is very important for humans as social beings. The pioneering study on emotion messages revealed from human faces was from Darwin's work [1]. Then Ekman defined six basic emotions which are claimed to be universally associated with distinct facial expressions [2]. These six basic emotions are: happiness, sadness, surprise, fear, anger, and disgust. Although the question about whether these basic emotions are indeed universal still remains an open question, most of the vision-based facial expression studies rely on Ekman's definition about the universal categories of emotions.

The Facial Action Coding System (FACS) is a human-observer-based system that has been developed to facilitate objective measurement of subtle changes in facial appearance caused by contractions of the facial muscles [3]. Via 44 action units, FACS is able to give a linguistic description of all visibly discriminable expressions.

Automatic facial expression systems can be applied to human-computer interaction, stress-monitoring systems, low-bandwidth videoconferencing, human behavior analysis, etc [4]-[18]. Thus in recent years, the research of developing automatic facial expression recognition systems has attracted a lot of attention from many different fields. While an overview of the early works in facial expression analysis can be found in [19], a more recent and complete overview is referred to [20]. The approaches to facial expression recognition can be roughly divided into two classes: geometrical feature-based approaches and appearance-based approaches [16]. The geometrical feature-based approaches rely on the geometric facial features which present the shapes and locations of facial components such as eyebrows, eyes, canthus, nose, mouth etc. Experimental results exhibited that the facial features cannot always be obtained reliably because of the quality of images, illumination, and some other disturbing factors. As for the appearance-based approaches, the whole-face or specific regions in a face image are used for the feature extraction via optical flow or some kinds of filters. Some approaches focus on the discrimination of facial expression at the level of emotion prototypes but some other approaches are able to discriminate expressions at a fine-grained level via the recognition of action units [12]. Some approaches can fully automatically recognize expressions from image sequences but some approaches still need to manually label some feature points before the recognition procedure. With few exceptions, most proposed approaches have used relatively limited data sets. Detailed comparisons of the existing approaches were provided in the review article [20].

In this paper we propose a simple approach to automatic facial expression recognition. A detailed description of the proposed expression recognition algorithm is given in Section 2. Then simulation results are given in Section 3 and Section 4 concludes the paper.

2 Automatic Facial Expression Recognition Algorithm

The inputs to our 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. We assume the first frame of each image sequence is a neural expression. Our algorithm involves in the following three main steps.

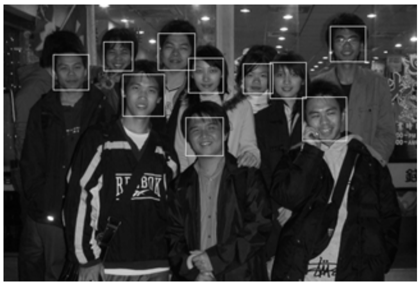
2.1 Face Detection

Face detection determines the locations and sizes of faces in an input image. They are easily located in cluttered scenes by infants and adults alike; however automatic human face detection by computers is a very challenging task because face patterns can have significantly variable image appearances. For example, human faces vary from genders, ages, hair styles and races etc. In addition, the variations of 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 [21]-[29]. 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 [30]. This face detection method can minimize computational time while achieving high detection accuracy. Fig. 1 shows some detection results. The faces in Fig. 1(a) were all detected but two faces in Fig. 1(b) were not detected and one region in Fig. 1(c) was falsely classified as a human face. The image sequences to be processed by our facial expression recognition algorithm are consisted of faces viewed from a near frontal view (as shown in Fig. 1(d)); therefore, the method achieved 100% correct detection rate for those image sequences.

2.2 Feature Points Tracking

After the face in the first frame has been detected the next step is to extract necessary information about the facial expression presented in the image sequence. When facial muscles contract, the transformation of the corresponding skin areas attached to the muscles produces changes in the appearance of facial features and results in a certain type of visual effect. The movements of facial points (eyebrows, eyes, and mouth) have a strong relation to the information about the shown facial expression. Therefore, many approaches greatly depend on the tracking of permanent facial features (eyebrows, eyes, mouth, and furrows that have become permanent with age) and/or transient facial features (facial lines and furrows that are not present at a neutral state). In fact, the extraction of facial features sometimes is a very challenging task. Facial features cannot always be obtained reliably because of the quality of images, illumination, and some other disturbing factors. Furthermore, it usually takes a lot of computations to extract precise facial features.



(a)

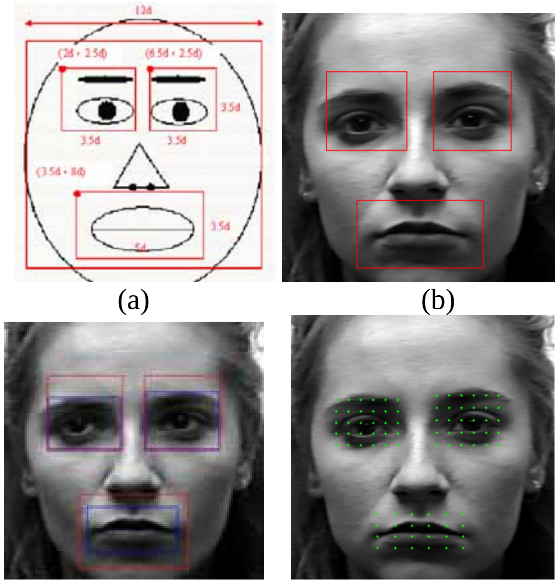
(b)

(c)

(d)

Fig. 1 Examples of faces detected by the method proposed in [30].

To alleviate unnecessary computational load we decided to compute optical flow within three rectangles which include the action units having high correlations with facial expressions. The idea of the use of optical flow in tracking feature points have been developed in several different approaches such as [11], [12], [31]-[32]. However, there are major differences between our approach and those existing optical-flow-based approaches. For example, the user has to manually mark key feature points in the first frame [12]. A more efficient and automatic way is to compute optical flow at the points with high gradient at each frame without the need of extracting precise facial features [11]. However, it still wastes computations to choose feature points with high gradients. In order to further improve the computational efficiency, we proposed to uniformly distribute 84 feature points in three rectangles which enclose the regions of interest.



(a) (b) (c) (d)

Fig. 2 The regions of interest in a face. (a) The geometric face model. (b) The three initial rectangles on a detected face. (c) The three refined rectangles. (d) The 84 feature points uniformly distributed on the three rectangles.

These three rectangles are consisted of the upper left rectangle enclosing the left eye and the brow, the upper right rectangle enclosing the right eye and the brow, and the lower middle rectangle enclosing the mouth since these three regions have high correlations with facial expressions. We built a geometric face model as shown in Fig. 2(a) to represent the geometrical relations of those three regions. Based on the face model, three initial rectangles as shown in Fig. 2(b) can be quickly located from the face detected by the previous step. Since geometric relations between the eyes and the mouth vary a little bit from person to person, we have to refine these three initial rectangles to fit their correct regions. We use the information of the horizontal and the vertical edges of the regions enclosed by the three initial rectangles to refine the locations and sizes of the rectangles. Fig. 2(c) illustrates three refined rectangles. While there are 30 feature points uniformly distributed in each of the upper rectangles, 24 feature points are uniformly distributed in the mouth rectangle, as shown in Fig. 2(d).

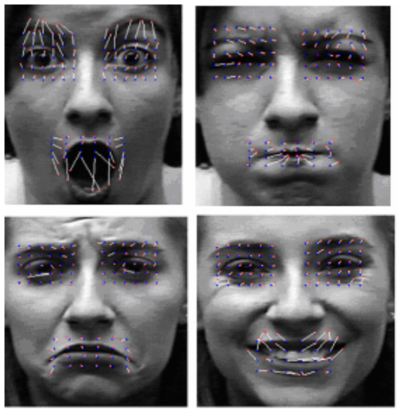


Fig. 3 Some tracking results of the 84 feature points.

Then a pyramidal implementation of a hierarchical optical flow method [33]-[34] is used to automatically track the 84 feature points in the image sequence. The displacement of each feature point is calculated by subtracting its original position in the first frame from the final position in the last frame of the image sequence. Since the size of the face varies from person to person, the computed displacements are normalized by dividing the displacements by the face width. The 84 flow vectors are used as an input pattern to neural networks for the recognition of action units. Fig.3 illustrates some examples with optical flow vectors superimposed on the faces.

2.3 Facial Expression Recognition

In this step, we are to classify the facial display conveyed by the face. Some approaches directly use flow vectors to recognize facial expressions [6], [11]; however, Cohen et al claimed that those direct approaches can’t capture the full range of emotion expression [12]. Therefore, they suggested that it is better to recognize action units first and then facial expressions. Following the idea, our facial recognition procedure involves in two stages. At the first stage, three multi-layer perceptrons are trained to recognize the action units in the eyebrows, the eyes and the mouth regions. Then another five single-layer perceptrons are used to recognize facial expressions based on the outputs computed from the aforementioned three MLPs.

In our algorithm, three MLPs are trained to recognize three action units in the brows (AU1, AU2, and AU4), three action units in eyes (AU5, AU6, and AU7), and eleven action units in the mouth (AU10, AU12, AU15, AU16, AU17, AU20, AU23, AU24, AU25, AU26, and AU27). These 17 action units are chosen because they have strong relations with facial expressions. For example, action unit 4 (AU4) means that brows are lowered and drawn together. These three MLPs are separately trained based on the 12 flow vectors, 48 flow vectors, and 24 flow vectors shown in Fig. 4, respectively. Then the outputs from the trained three MLPs are used as the inputs to five single-layer perceptrons for the recognition of facial expressions, as shown in Fig. 5. The goal of this paper is to recognize the following five facial expressions: surprise, happiness, sadness, anger, and neural.

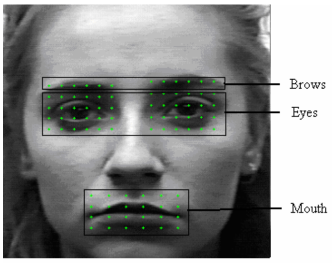


Fig. 4 The feature points used for the recognition of action units in the brows, the eyes, and the mouth.

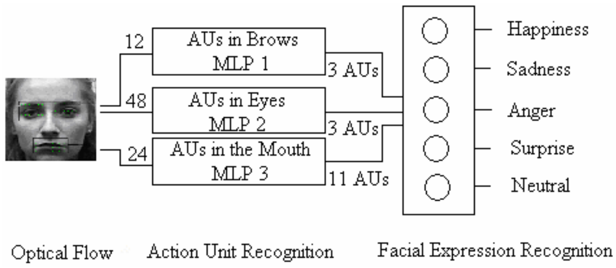


Fig. 5 The neural networks for the recognition of action units and facial expressions.

3 Simulation Results

The Cohn-Kanade AU-Coded Facial Expression Database [18] was used to test our algorithm. The database was consisted of approximately 2000 image sequences from 210 adults between the ages of 18 and 50 years. They were 69% female, 31% male, 13% Afro-American, 81% Euro-American, and 6% other groups. Subjects were instructed by an experimenter to perform a series of 23 facial displays which included single AUs and AU combinations. The image sequences began with a neutral face and were digitized into either 640 x 490 or 640 x 480 pixel arrays with 8-bit gray-scale or 24-bit color values. Face size varies between 90 x 80 and 220 x 200 pixels.

There were 486 image sequences which were used in our experiments. In this database, the image sequences were not attached with expression labels. Therefore, we asked 13 subjects to judge facial expression presented at the final frame of each image sequence. We found that there were only 415 image sequences which could be labeled to corresponding expressions based on the following majority consensus rule. If there were more than or equal to 7 subjects who gave the same votes then this image sequence was labeled to that particular expression.

Table 1 tabulates the recognition results for 17 action units. The ratio of the training data set to the testing data set was 74 to 26. For the testing set, the average correct recognition ratios were 86.36% and 85.23% for the upper face and the lower face, respectively.

Table 1 Recognition results of action units

AU	Training Set	Testing Set
Upper Face (Brows and Eyes)	91.56%	86.36%
Lower Face	92.87%	85.23%

The recognition results for five facial expressions were shown in Tables 2-3. The ratio of the training data set to the testing data set was 74.9 to 25.1. The average correct recognition ratios were 93.89% and 93.27% for the training set and the testing set, respectively.

Table 2 Recognition results of facial expressions for the training data set.

Real \ NN	Surprise	Happiness	Sad	Anger	Neutral
Surprise	63	0	0	1	2
Happiness	0	72	0	4	1
Sad	0	0	47	2	1
Anger	0	0	3	72	0
Neutral	1	0	1	3	38
Average	93.89%				

Table 3 Recognition results of facial expressions for the testing data set.

Real \ NN	Surprise	Happiness	Sad	Anger	Neutral
Surprise	19	0	0	0	3
Happiness	1	25	0	0	0
Sad	0	0	15	1	0
Anger	1	1	0	23	0
Neutral	0	0	0	0	15
Average	93.27%				

4 Conclusions

In this paper, a simple approach to automatic facial expression recognition is presented. The proposed system is able to automatically perform human face detection, feature point extraction and facial expression recognition from image sequences. The extraction of facial features sometimes is a very challenging task and it usually takes a lot of computations to extract precise facial features. To alleviate the computational load, we propose to uniformly distribute 84 feature points over the three automatically located rectangles instead of extracting precise facial features. The average recognition performance for facial expressions could be achieved to 93% correct. This result was very encouraging compared to some existing approaches.

Acknowledgements

This work was partly supported by the NSC Program for Promoting Academic Excellent of Universities (Phase II) under the grant number NSC-95-2752- E - 008-002-PAE, the National Science Council, Taiwan, R.O.C, under the NSC-95-2221-E-008-128 and the NSC-95-2524-S-008-001, and the NSC - 95 - 2524-S-008-005-EC3, the Ministry of Economic Affairs under the 95-EC-17-A- 02-S1-029 and the NCU Project of Promoting Academic Excellence & Developing World Class Research Centers - Applied Informatics and Creative Contents: Service-Oriented Information Platform. 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] C. Darwin, *The Expression of Emotions in Man and Animals*, University of Chicago Press, 1965.

[2] P. Ekman, *Emotion in the Human faces*, Cambridge University Press, 1982.

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

[4] 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.

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

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

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

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

[9] 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.

[10] 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.

[11] 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.

[12] J. F. Cohn, A. J. Zlochower, J. J. Lien, and T. Kanade, "Feature-point tracking by optical flow discriminates subtle differences in facial expression," in *Proceedings of the 3rd IEEE International Conference on Automatic Face and Gesture Recognition*, pp. 396-401, 1998.

[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, Issue: 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 Recognition in Image Sequences of Increasing Complexity," in *Proceedings of the Fifth IEEE International Conference on Automatic Face and Gesture Recognition*, pp. 229-234, 2002.

[17] 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.

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

[19] A. Samal and P.A. Iyengar, "Automatic Recognition and Analysis of Human Faces and Facial Expressions: A Survey," *Pattern Recognition*, vol. 25, no. 1, pp. 65-77, 1992.

[20] 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.

[21] T. Sakai, M. Nagao, and S. Fujibayashi, "Line extraction and pattern recognition in a photograph," *Pattern Recognition*, vol. 1, pp. 233-248, 1969.

[22] P. Sinha, "Object recognition via image invariance a case study," *Investigative Ophthalmology and Visual Science*, vol. 35, pp. 1735-1740, 1994.

[23] V. Govindaraju and M. Rajapakse, "Verification of faces in cluttered photographs," *IEEE International Conference on Systems, Man, and Cybernetics*, pp. 2028-2033, 1997.

[24] A. L. Yuille, P. W. Hallinan, and D. S. Cohen, "Feature extraction from faces using deformable templates," *International Journal of Computer Vision*, vol. 8, no. 2, pp. 99-111, 1992.

[25] P. Seitz and G. K. Lang, "Using local orientation and hierarchical spatial feature matching for robust recognition of objects," *Proceedings of SPIE*, pp. 252-259, 1991.

[26] 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.

[27] 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.

[28] 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.

[29] 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.

[30] P. Viola and M. J. Jones, "Rapid Object Detection using a Boosted Cascade of Simple Features," in *Proceedings of the IEEE Computer Society International Conference on Computer Vision and Pattern Recognition*, vol. 1, pp. 511-518, 2001.

[31] M. J. Black and Y. Yacoob, "Recognizing Facial Expressions in Image Sequences Using Local Parameterized Models of Image Motion," *International Journal of Computer Vision*, vol. 25, no. 1, pp. 23-48, 1997.

[32] T. Otsuka and J. Ohya, "Spotting Segments Displaying Facial Expression from Image Sequences Using HMM," in *Proc. International Conf. Automatic Face and Gesture Recognition*, pp. 442-447, 1998.

[33] B. D. Lucas and T. Kanade. "An investigation of smoothness constraints for the estimation of displacement vector fields from image sequences," *IEEE Trans. on Pattern Analysis and Machine Intelligence*, vol. 8, pp. 565-593, 1986.

[34] J. Y.Bouguet, "Pyramidal Implementation of the Lucas Kanade Feature Tracker Description of the algorithm," Intel Corporation Microprocessor Research Labs.