Single and Composite Action Units Classification in Facial Expressions by Feature-Points Tracking and RBF Neural Network

*Ali Aghagolzadeh *, Hadi Seyedarabi **, Sohrab Khanmohammadi ***
* Islamic Azad University of Tabriz, Tabriz, Iran
** Elect. Eng. Faculty, Tabriz University, Tabriz, Iran
{aghagol, seyedarabi, khan} @tabrizu.ac.ir

Abstract

Face plays an essential role in interpersonal communication. .Most of the current works on facial expression recognition attempt to recognize a small set of prototypic expressions such as happy, surprise, anger, sad, disgust and fear. However the most of human emotions is communicated by changes in one or two of discrete features. In this paper, we develop a facial Action Units (AU) Classification system, based on the facial features extracted from facial characteristic points in frontal image sequences. Selected facial feature points were automatically tracked using a cross-correlation based optical flow, and extracted feature vectors were used to classify Action Units, using RBF neural networks. Proposed classifier showed good results for classifying single and composite AUs.

1. Introduction

Face expression recognition is useful for designing new interactive devices offering the possibility of new ways for human to interact with computer systems. Automating facial expression analysis could bring facial expressions into man-machine interaction.

In 1971 Ekman and Friesen [1] postulated six primary emotions that each posses a distinctive content together with a unique facial expression. These prototypic emotional displays are also referred to as basic emotions. They seem to be universal across human ethnicities and cultures and comprise happiness, sadness, fear, disgust, surprise, and anger. In recent years, there has been great amount of research on face recognition and facial expression recognition.

Most computer-vision based approaches to facial expression analysis attempt to recognize only these prototypic emotions. These expressions involve simultaneous changes in facial features in multiple regions of the face. In everyday life, however, such prototypic expressions occur relatively infrequently. Instead, emotion is communicated by changes in one or two discrete features, such as lowering the lip corners only in sadness.

Table 1 shows Facial Action Coding System (FACS) AUs used in this work that occur in the upper and lower face.

Table 1: Some of FACS AUs used in this work

AU	NAME
1	Inner brow raise
2	Outer brow raise
4	Brow corrugators
5	Upper lid raise
6	Cheek raise
12	Lip corner pull
15	Lip corner depress
17	Chin raise
23	Lip tighten
24	Lip press
25	Lips part
26	Jaw drop
27	Mouth stretch

Hara and Kobayashi [2] apply a $234\times50\times6$ back-propagation neural network for classification of expressions into one of the six basic emotion categories. The units of the input layer correspond to the number of the brightness distribution data extracted from an input facial image, while each unit of the output layer correspond to one emotion category. The neural network has been trained by 90 images of six basic facial expressions shown by 15 subjects and it has been tested on a set of 90 facial expressions images shown by another 15 subjects. The average recognition rate was 85 percent.

Beat Fasel [3] proposed a data-driven face analysis approach that was capable of extracting features relevant to a given face analysis task. His proposed system was more robust with respect to the face location changes and the scale variations when compared to the classical methods such as MLPs. His approach was based on the Convolutional Neural Networks (CNN) that used the multi-scale feature extractors, which allow for improved facial expression recognition results with faces subject to in-plane pose variations. He used a total of 140 images from JAFFE database to train neural networks and 70 images for testing and obtained recognition rate of 91.4%.

Yacoob and Davis [4] used optical flow and tracked the motion of the surface regions of the facial features (brows, eyes, nose and mouth).

Zhang [5] investigated the use of two types of facial features: the geometric position of 34 fiducial points on a face and a set of multiscale, multiorientation Gabor wavelet coefficients at these points for facial expression recognition.

In this paper we proposed a facial Action Units classification system, based on the facial features extracted from facial characteristic points in the frontal image sequences. Selected facial feature points were automatically tracked using a cross-correlation based optical flow, and extracted feature vectors were used to classify Action Units, using RBF neural networks. The proposed classifier uses fewer feature points compared to the related works and so the optical flow algorithm has low complexity. The classifier uses a smaller size neural network compared to the other types of classifiers. Experimental results showed good results for classifying single and composite AUs.

2. Image Database

In this work, we used Cohn-Kanade database [6] that consists of expression sequences of subjects, starting from a neutral expression and ending in the peak of the facial expression. There are 104 subjects in the database. Subjects sat directly in front pf the camera and performed a series of facial expressions that included single AU (e.g. AU25) and combinations of AUs (e.g. AU6+12+25). Since for subjects, not all of the single and composite Action Units sequences were available to us, we used a subset of subjects for which selected single and composite AUs were available. For each person there are on average 12 frames for each expression. Image sequences for frontal views were digitized into 640×490 pixel array with 8 bits grayscale. Table 2 shows Cohn-Kanade database specifications and Fig. 1 shows one of the subjects showing some facial expressions.

Table 2: Cohn-Kanade database

Age	18 to 50
Female	69%
Male	31%
Euro-American	81%
Afro-American	13%
Other groups	6%
Resolution	640×490 Grayscale

Figure 1:Facial expressions in Cohn-Kanade Database

3. Facial Feature Point Tracking using Optical Flow

In the first digitized frame, 21 key feature points were manually marked with a computer-mouse around facial landmarks (Fig. 2). Among 44 AUs, 39 AUs are directly associated with movement of eyes, eyebrows and mouth. That's why the information expressing movement of eyes, eyebrows and mouth is desirable for machine recognition of facial expressions. We are confined in these three components and then determine facial feature points which are representative of the boundary between these components and skin.

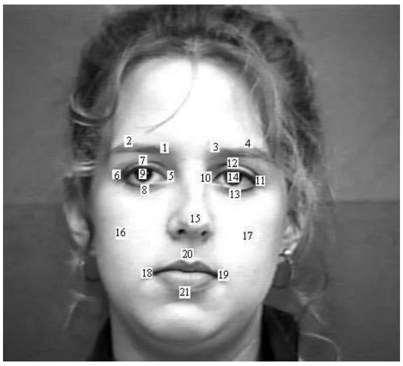

Figure 2: Selected 21 facial feature points.

Each point is the center of a 11×11 flow window that includes horizontal and vertical flows. A cross-correlation based optical flow method is used to automatically track feature points in the image sequence. Fig. 3 shows the implementation of this method in two subsequent frames:

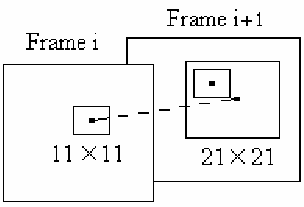

Figure 3: Cross-correlation optical flow calculation

Cross-correlation of 11×11 window in the first frame, and a 21×21 window at the next frame were calculated and the position with maximum cross-correlation of two windows, were estimated as the position of feature point at the next frame. Each feature point is calculated by subtracting its normalized position in the first frame from its current normalized position (The position of all feature points was normalized by position of the tip of nose) [7, 8, 9].

4. Feature Extraction from Feature Points

7 features were extracted from the feature point positions in the first frame and the last frame. These features form the feature vector for each expression, and were used to classify that expression to AUs, using neural network [7, 8, 9].

Extracted features are as fellows:

Width of eye:

$$we=\frac{(x_{11}-x_{10})+(x_5-x_6)}{2} \tag{1}$$

Height of eyebrows 1:

$$he1=\frac{(y_{15}-y_4)+(y_{15}-y_2)}{2} \tag{2}$$

Height of eyebrows 2:

$$he2=\frac{(y_{15}-y_3)+(y_{15}-y_1)}{2} \tag{3}$$

Width of mouth:

$$wm=x_{19}-x_{18} \tag{4}$$

Openness of mouth:

$$om=y_{21}-y_{20} \tag{5}$$

Nose tip-lip corners distance:

$$nl=\frac{(y_{18}-y_{15})+(y_{19}-y_{15})}{2} \tag{6}$$

Eye- cheek distance:

$$ec=\frac{(y_{16}-y_9)+(y_{17}-y_{14})}{2} \tag{7}$$

(x_i, y_i) Represents the coordinate of feature points which the origin of X-Y coordinate system is assigned at the tip of nose.

5. RBF Neural Network Classifier

RBFN is a class of single hidden layer feedforward networks where the activation functions for hidden units are defined as radially symmetric basis functions such as the Gaussian function. The fraction of the overlap between each hidden unit and its neighbors is decided by the width of Gaussian function (sigma) such that a smooth interpolation over the input space is allowed. The whole architecture is therefore fixed by determining the hidden layer and the weights between the middle and the output layers [10].

The number of input layer units must be equal to 7, equal to the number of extracted features, and that of output layers is 3, in single AU recognition which corresponds to three kinds of the selected single AUs and is 14, in composite AU recognition which corresponds to 14 kinds of AUs in the selected composite AUs. The network training is carried out by back propagation algorithm.

6. AUs Recognition

Fig. 4 shows our proposed AUs classifier.

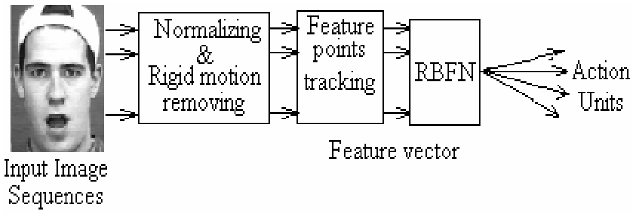

Figure 4: AUs Classifier

The results of feature extraction from an image sequence of facial expression are now represented by a 7×1 feature vector. . We want to classify each sequence into one of the predetermined AU in the single AU images and combinations of AUs in the composite AU images. Training target vector for a single AU images is a 3×1 vector which contains one for the existing AU and zeros for the others. For composite AU images, training target vector is a 13×1 vector which contains ones for the existing AUs and zeros for the others. Manual FACS codes are used to form this vector. We used introduced RBF neural networks as a decision making mechanisms.

7. Experimental Results

Selected facial feature points were automatically tracked using a cross-correlation based optical flow, and the extracted feature vectors were used to classify AUs, using RBF neural networks.

AUs that are important to the communication of the emotion and that occurred a minimum of 25 times in our data base were selected for analysis. This frequency criterion ensured sufficient data for training and testing. Therefore 35 subjects were selected for the single AU and 53 subjects for the composite AUs classification. We used the sequence of some subjects as test sequences (7 subjects for the single AU and 10 subjects for the composite AUs classifier), and the sequence of the remaining subjects as training sequence. This test is repeated five times, each time leaving different subjects out.

7.1 Single AUs

Selected single AUs are shown in figure 5.

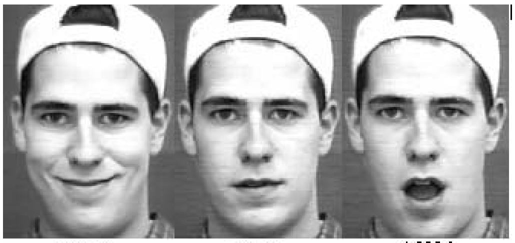

Figure 5: Selected single AUs.

Single AUs recognition results from proposed RBF classifier are shown in Table 3.

Table 3: Single AUs recognition results

Recognition	AU12	AU25	AU26
AU 12	**83%**	17%	0%
AU 25	0%	**84%**	16%
AU 26	0%	11%	**89%**

7.2 Composite AUs

Selected composite AUs are shown in figure 6.

Figure 6: Selected composite AUs.

Composite AUs recognition results from proposed RBF classifier are shown in Table 4.

Table 4: Composite AUs recognition results

AU	Recognition Rate
1	98%
2	98%
4	73%
5	98%
6	91%
12	91%
15	88%
17	91%
23	73%
24	73%
25	91%
7	98%

8. Conclusions

In this research we presented a system for classifying of single and composite AUs, from continuous video inputs. In the RBFN classifier, 7 features extracted from 21 feature points were used as training and test sequences. The trained Neural Network was tested by features that were not used in training and we have obtained a good recognition rate. Recognized AUs in this analysis procedure can be apply to a 2D or 3D synthetic face model to obtain the same expression, consisting single AU or combinations of AUs, for facial animation, video-conferencing or facial image coding applications.

9. Acknowledgement

The authors appreciate the support of Iranian Telecommunication Research Center (ITRC) for this research and would like to thank J.F. Cohn and T. Kanade from Pittsburgh University for their kindly providing the image database.

10. References

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

[2] H. Kobayashi and F. Hara, "Facial Interaction between Animated 3D Face Robot and Human Being," *Proc. Int'l Conf. Systems, Man, Cybernetics*, pp. 3732-3737, 1997.

[3] Beat Fasel, "Head-pose invariant Facial Expression Recognition using Convolutional Neural Networks." *Proc. Of the fourth IEEE International Conference on Multimodal Interfaces (ICMI'02)*, Pittsburgh, Pennsylvania, 2002.

[4] Y. Yacoob and L. S. Davis, "Recognizing Human Facial Expression from Long Image Sequences Using Optical Flow," *IEEE Trans. Pattern Analysis and Machine Intelligence*, vol 18, no. 6, pp. 636-642, June 1996.

[5] Z. Zhang, "Feature Based Facial Expression Recognition: Sensitivity Analysis and Experiments with a Multilayer Perceptron," *Int'l J. Pattern Recognition and Artificial Intelligence*, vol. 13, no. 6, pp. 893-911, 1999.

[6] T. Kanade, J. Cohn and Y. Tian. *Comprehensive database for facial expression analysis*, 2000.

[7] H. Seyedarabi, A. Aghagolzadeh and S. Khanmohammadi, "Facial Expressions Recognition from Static Images using Neural Networks and Fuzzy logic," *The 2nd Iranian Conference on Machine Vision and Image processing (MVIP2003)*,vol.1 pp 7-12, Tehran, 2003.

[8] H. Seyedarabi, A. Aghagolzadeh and S. Khanmohammadi, "Facial Expressions Recognition from Image Sequences using Cross-correlation Based Optical-Flow and Radial Basis Neural Networks," *The 12th Iranian Conference on Electrical Engineering,(ICEE 2004)*, vol.1 pp 165-170,Mashhad, 2004 .

[9] H. Seyedarabi, A. Aghagolzadeh and S. Khanmohammadi, "Recognition of Six Basic Facial Expressions by Feature-Points Tracking using RBF Neural Networks and Fuzzy Inference System," *The IEEE International conference on Multimedia & Expo (ICME2004)*, Taipei, Taiwan ,June 2004.

[10] J. Moody and C. Darken, "Learning with Localized receptive fields" *in proc, 1988 Connectionist Models Summer School*. San Matco. CA: Morgan-Kaufmann 1988.