Expression Recognition using Elastic Graph Matching

Yujia Cao[1,2] , Wenming Zheng[1,2] , Li Zhao[1,2] , Cairong Zhou[2]

(1. Research Center for Learning Science, Southeast University, Nanjing 210096, China;
2. Department of Radio Engineering, Southeast University, Nanjing 210096, China)
Email: yujia_cao@seu.edu.cn

Abstract. In this paper, we proposed a facial expression recognition method based on the elastic graph matching (EGM) approach.The EGM approach is widely considered very effective due to it's robustness against face position and lighting variations. Among all the feature extraction methods which have been used with the EGM, we choose Gabor wavelet transform according to its good performance. In order to effectively represent the facial expression information, we choose the fiducial points from the local areas where the distortion caused by expression is obvious. The better performance of the proposed method is confirmed by the JAFFE facial expression database, compared to the some previous works. We can achieve the average expression recognition rate as high as 93.4%. Moreover, we can get face recognition result simultaneously in our experiment.

1 Introduction

Facial expression plays an important role in our daily face-to-face communication. Research on facial expression has already had a long history in the psychological field. In 1971, the famous psychologist Ekman defined six basic expressions: anger, disgust, fear, happy, sad and surprise [1]. This classification method is universal to ages, sex and race, therefore had been widely used in most of the facial expression related works. After 1980s, as the computer technology develops, facial expression analyse becomes more and more popular in pattern recognition and artificial intelligence field [2]. Many techniques are applied to facial expression recognition in recent years, for example: principal component analysis (PCA) [7],[10], Gabor wavelets [10],[17], neural network [8], Hidden Markov Models (HMM) [12], Point Distribute Model (PDM)[9], optical flow [11],[13], facial action coding system (FACS) [14], and so on. In this paper, we will propose a facial expression recognition method based on the elastic graph matching (EGM) approach.

The EGM method was first proposed by Lades et al. [3] and applied to face recognition. The original elastic graph is a rectangular lattice. Gabor wavelet transform is used for feature extraction. Based on the work of Lades et al, Wiskott [4] extracted more than one feature vectors on one fiducial point and called it elastic bunch graph (EBG). It is a net with fiducial points selected more wisely from the face image. Faces with large rotation angle and different size are also taken into consideration. In order to avoid the computational cost caused by the Gabor wavelet transform, Constantine [20] and Kun Ma [19] use morphological feature vectors and discrete wavelet graph to do the elastic graph matching. But they can't get performance as good as the Gabor wavelet. The EGM method with Gabor wavelet features is widely considered very effective due to it's robustness against face position and lighting variations. That is because both the algorithm and the Gabor kernel have some endurance to translation, distortion, rotation, and scaling.

In consideration of the particularity of facial expression recognition task, the fiducial points (nodes of the elastic graph) are defined wisely in our approach. They are put in the local areas where the information of expression is rich. Experiment is designed on the standard facial expression database JAFFE. At the same time of expression recognition, we can get person recognition result simultaneously in our experiment.

2 Gabor Feature Extraction

2.1 Gabor kernels

The two-dimension Gabor wavelet kernels have the characteristics similar to the mammalian cortical simple cells [5]. They have been found particularly suitable for extracting the local features, therefore are used widely for image analysis. The Gabor kernels take the form of a plane wave restricted by a Gaussian envelop function [3]:

$$\psi(\vec{x}) = \frac{\vec{k}^2}{\sigma^2}\exp(-\frac{\vec{k}^2\vec{x}^2}{2\sigma^2})\left[\exp(i\vec{k}\vec{x}) - \exp(-\frac{\sigma^2}{2})\right] \quad (1)$$

where $\vec{x}$ refers to the location of a given pixel $\vec{x}$ = (i, j); σ is a parameter which controls the width of the Gaussian ($\sigma = 2\pi$ in our experiment). The second exponent in the bracket makes the kernel DC-free, so they are robust against average brightness changes. The Gabor kernels are used at five frequencies index v=0,…,4, and eight orientations index μ=0,…,7, which is determined by vector $\vec{k}$:

$$\vec{k} = \begin{pmatrix} k_x \\ k_y \end{pmatrix} = \begin{pmatrix} k_v \cos\varphi_\mu \\ k_v \sin\varphi_\mu \end{pmatrix}, k_v = 2^{-\frac{v+2}{2}\pi}, \varphi_\mu = \mu\frac{\pi}{8} \quad (2)$$

2.2 Wavelet transform

A wavelet transform is defined as a convolution of an image with the Gabor kernels:

$$C(\vec{x}) = \int \psi(\vec{x} - \vec{x}')I(\vec{x}')d^2\vec{x}' \quad (3)$$

In order to accelerate the computation, we can complete the convolution by the Fast Fourier Transform (FFT) and inverse Fast Fourier Transform (IFFT):

$$C(\vec{x}) = F^{-1}\{F\{\psi(\vec{x})\}F\{I(\vec{x})\}\} \quad (4)$$

Where F and F^{-1} refers to FFT and IFFT, respectively. The convolution result is called wavelet transform coefficients [4].

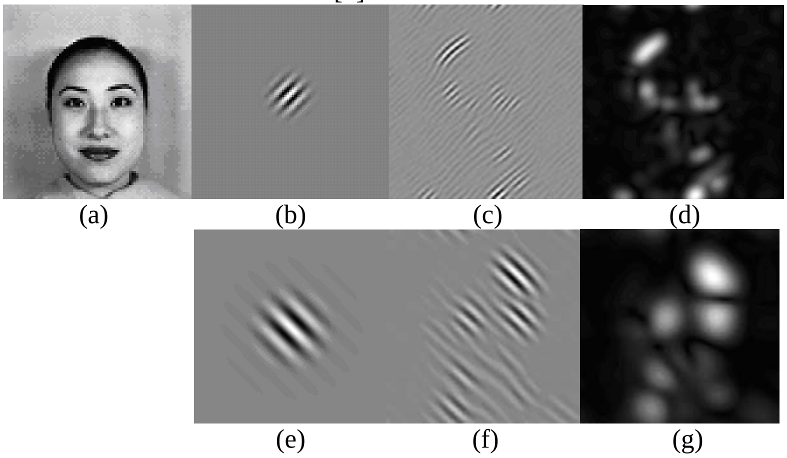

Fig. 1. (a) original image (b) The real part of Gabor kernels at v=2, u=2 (c),(d) the real part and the magnitude of wavelet transform coefficients (e) The real part of Gabor kernels at v=4, u=6 (f),(g) the real part and the magnitude of wavelet transform coefficients

From Fig.1, we can see that the wavelet transform coefficients are relatively small if the grey value in the image changes gently. When the kernel meets the edges in the image, the real part (and the imaginary part) of the coefficient oscillates (with frequency decided by μ), the magnitude grows. The response is especially strong if the direction of the edge and the kernel is identical. If the fiducial points are defined along edges, the real part changes remarkably even within several pixels, however, the magnitude provide a smooth peak [3]. So the magnitude of Gabor wavelet transform coefficients will be used for feature representation.

2.3 Feature extraction

Applying the convolution with all possible $\vec{k}$, we obtain 40 complex coefficients at one image pixel $\vec{x}$. A vector is set up with the magnitude of these coefficients and is called a jet (5). A set of jets taken from the same fiducial point in different images is called a bunch.

$$\vec{J} = \{abs\,\{C_i(\vec{x}_0)\}\}, \quad i = 1,\dots,40 \tag{5}$$

Distortions caused by facial expressions always occur on some specific areas, e.g. the eyebrow, the upper and lower eyelid, the mouth, the cheek, the chin and the area around the nose. In order to extract features effectively, we define fiducial points at these specific areas (Fig2 (a)), but ignore the forehead, the hair and the face profile [4]. With this development, we can get better performance with less fiducial points, therefore accelerates the matching process.

Some points are connected together to form some metric structure (Fig2 (b)). After extracting one jet at each point, an elastic graph is generated. When matching, the sub graphs (eye model, nose model, mouth model et. al) are allowed to move respectively. If we extract one bunch at each fiducial point, an EBG can be set up (Fig2 (c)).

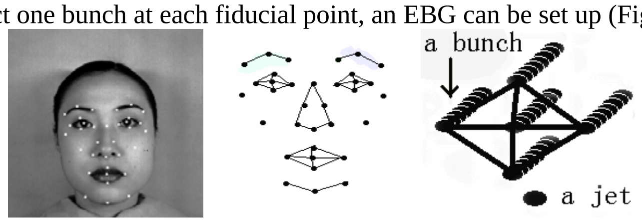

Fig. 2. (a) fiducial points (b) an elastic graph (c) a part of EBG

3 Elastic Graph Matching

As the elastic graph is already generated, we will design a coarse-to-fine process, which guides the elastic graph to match the face. The matching process will be controlled by some similarity functions which are related to the jets and the metric structure.

3.1 The similarity function

The similarity function between two jets is defined as:

$$S(\vec{J}, \vec{J}') = \frac{< \vec{J}, \vec{J}' >}{\| \vec{J} \| \cdot \| \vec{J}' \|} \tag{6}$$

Suppose that G and G' are two elastic graphs, each has N fiducial points. The jets similarity is defined as follows:

$$S_j = \frac{1}{N} \sum_{i=1}^{N} S(\vec{J}_i, \vec{J}_i') \ , \quad \vec{J}_i \in G, \vec{J}_i' \in G' \tag{7}$$

$\vec{J}_i$ and $\vec{J}_i{}'$ refers to a jet extracted from the i th fiducial point in G and G', respectively. The value of S_j is between -1 and 1. More similar the two model graphs are, the closer S_j is to 1.

The metric structure similarity is defined as:

$$S_e = \frac{1}{E} \sum_{i=1}^{E} | d_i - d_i' | \tag{8}$$

$$\text{where,} \quad d_i = \| \vec{e}_i \|, \vec{e}_i \in \mathrm{G}, \quad d_i' = \| \vec{e}_i' \|, \vec{e}_i' \in \mathrm{G}' \tag{9}$$

E is the total number of edges in the elastic graph. $\vec{e}_i$ and $\vec{e}_i$ represents the i th edge in G and G', respectively. d_i and d_i' refer to the length of $\vec{e}_i$ and $\vec{e}_i$. The value of S_e is positive. The greater the difference between two elastic graphs is, the greater the value is. We define a parameter λ to determine the relative weight of jets and metric structure. It should be negative. Then there comes the total similarity function:

$$S_{total} = S_j + \lambda\, S_e \qquad (10)$$

3.2 The Matching Process

The elastic graph matching will be completed from coarse to fine. The matching procedure we designed has the following three stages:

Stage 1: We use the model graph as a whole, which means that the relative positions of fiducial points are not allowed to change. Only S_j is used to calculate the similarity. The elastic graph goes through the whole test image with a step of 8 pixels, searching for a position which has the maximum S_j, then searches in a 9×9 square area around the position with a step of 2 pixels and find the best fitting position again. At the end of this stage, the elastic graph is approximately put on the face.

Stage 2: We match the sub graphs separately. For each sub model, check the ±1, ±2, ±3 position in the x and y dimensions, respectively. At each position, try three different size and $\pm5°$ rotation. In order to prevent the sub graphs from departing too far away from their original positions and maintain the general shape of the elastic graph, the metric structure is also taken into consideration. Therefore, S_{total} is used to measure the similarity. After this stage, the elastic graph has already matched the face, but still need some refined adjustment.

Stage 3: In order to further enhance the matching precision, each fiducial point moves in a 3×3 square area, scanning for a better fitting position. S_{total} is still used to control the displacements. Some matching results are shown in Fig3.

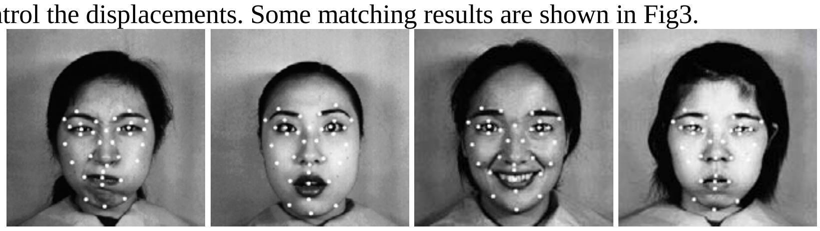

Fig. 3. Some matching results

4 Experiments and Results

The JAFFE database [6] contains 213 images of Japanese women. There are 10 people in the database; each person has 7 expressions (6 basic expressions and neutral face). The image size is 256×256. The faces are all frontal but the illumination intensity is not the same.

The training process has two tasks: 1) Use all the neutral faces to generate an average elastic graph. It will be used for matching a test image. 2) For each expression, use 1/3 training images to generate an EBG. The remaining 2/3 images will be used for testing.

When a test image is input, we firstly use the average elastic graph to do the matching (the process is described in 3.2). When the elastic graph has matched the face precisely, we calculate its graph similarity with the 6 EBGs. The graph similarity function is defined as (11). It is an average of M S_{total} in one EBG. The expression (EBG) with the maximum similarity is identified to be the recognition result. Therefore, a test image is recognized correctly if the EGB of the correct expression yields the highest graph similarity.

$$S_g = \frac{1}{M}\sum_{n=1}^{M}\left[\frac{1}{N}\sum_{i=1}^{N} S\left(\vec{J}_{ni}, \vec{J}_{ni}^{(k)}\right) + \frac{\lambda}{E}\sum_{j=1}^{E} | d_{nj} - d_{nj}^{(k)} |\right], \quad k = 1,..., 6 \qquad (11)$$

We can also identify the person in the test image to be the one with the maximum S_{total} in the best matching EBG. In this way, we realize the expression recognition and the person recognition simultaneously.

We use the cross-validation technique [21] to yield an average recognition rate (the same experiment procedure was repeated three times). The results are shown in Table1-3. According to the results, we can see that the recognition rate of 'Surprise' and 'Happy' is high, that is because these expressions cause obvious distortion on the face. The muscle movement of the other four expressions is relatively week, therefore they are harder to recognize. We also notice that the person with distinct expression is easy to recognize.

Table 1. Expression average recognition results

Expression	Anger	Disgust	Fear	Happy	Sad	Surprise
Number of test images	20	19	22	21	21	20
Correct recognition	18	16	20	21	20	20
Recognition rate	90.0%	84.2%	90.9%	100%	95.2%	100.0%
Average recognition rate	93.4%					

Table 2. Person average recognition results

person	1	2	3	4	5	6	7	8	9	10
Test images number	14	13	13	11	12	12	11	12	12	13
Correct number	12	13	13	10	10	11	10	11	10	12
Average correct recognition rate: 91.1%										

Table 3. total results

Total number of test images	Expression √ Person √	Expression × Person ×	Expression × Person √	Expression √ Person ×
123	109	5	3	6

Table 4. Comparison with previous work

Method	Training image ratio	Test image ratio	Recognition rate
LBP+Coarse-to-Fine classification in[15]	8/9	1/9	77%
HLAC+ Fisher weight maps in [16]	8/9	1/9	69.4%
Wavelet +PCA+LDA in [17]	8/9	1/9	75%
	9/10	1/10	92%
Multi-Layer Perceptron in [18]	9/10	1/10	90.1%
Our elastic graph matching method	1/3	2/3	93.4%

We compare our work to some previous work with different methods on the same database. Though the numbers of training images are different, all these recognition results are gained with the cross- validation technique. So they are still comparable. The result (Table 4) shows that the proposed approach achieve higher recognition rate with fewer test images.

5 Conclusion and Future work

We proposed a facial expression recognition approach based on the elastic graph matching method. Experiment on the JAFFE database shows the performance of our approach. The average recognition rate is 93.4%, which is higher than previous work on the same database. The highest single expression recognition rate reaches 100%.

In the future, we need more studies to test the performance of our approach on images with more challenging variations, such as side lighting, profile faces and glasses. We will try other feature extraction method in aim of accelerate the matching process. If the facial expression recognition technology can be combined with psychology judgement, the performance will be essentially improved to reach the practical applying level.

Acknowledgment

This work was supported in part by the Jiangsu nature science foundations under grants BK2005407.

References

1. P.Ekman, W.V.Friesen, Constants across cultures in the face and emotion, J.Personality Social Psychol. 17 (2) (1971) 124-129.
2. B. Fasel, Juergen Luettinb Automatic facial expression analysis: a survey Pattern Recognition 36 (2003) 259-275
3. Martin Lades et al, Distortion Invariant Object Recognition in the Dynamic Link Architecture, IEEE Trans. Computers, vol. 42, no. 3, March 1993

4. Laurenz Wiskott, Jean-Marc Fellous et al, Face Recognition by Elastic Bunch Graph Matching, IEEE Trans. Pattern Analysis and Machine Intelligence, vol. 19, no. 7, July 1997
5. D.A.Pollen, and S.F.Ronner, Visual Cortical Neurons as Localized Spatial Frequency Filters, IEEE Systems, Man, and Cybernetics, vol.13, no.5, pp.907-916, Sept.-Oct. 1983.
6. Michael J. Lyons, Shigeru Akamatsu et al, Coding Facial Expressions with Gabor Wavelets, Proceedings, Third IEEE International Conference on Automatic Face and Gesture Recognition, April 14-16 1998, Nara Japan, IEEE Computer Society, pp.200-205.
7. Andrew J. Calder, A, Mike Burton, Paul Miller, Andrew W. Young. A Principal Component Analysis of Facial Expressions Vision research 41(2001) 1179-1208
8. C. Lisetti, D. Rumelhart, Facial expression recognition using a neural network, Proceeding of the 11th International Flairs Conference, AAAI Press, New York, 1998
9. C. Huang, Y, Huang, Facial expression recognition using model-based feature extraction and action parameters classification, J. Visual Commun. Image Representation 8(3) (1997) 278-290
10. M. Dailey, G. Cottrell, PCA Gabor for expression recognition, Institution UCSD, Number CS-629, 1999
11. K. Mase, A. Pentland, Recognition of expression from optical flow, IEICE Trans. E 74 (10) (1991) 3474-3483.
12. Ira Cohen, Ashutosh Garg, Thomas S. Huang, Emotion Recognition from Facial Expressions using Multilevel HMM, http://citeseer.ist.psu.edu/502003.html
13. Y. Yacoob, L.S. Davis, Recognition human facial expression from long image sequences using optical flow, IEEE Trans.PAMI, 18(6) (1996) 636－642
14. P. Ekman, E. Rosenberg, J. Hager, Facial action coding system affect interpretation database(FACSAID), www.yahoo.com
15. Xiaoyi Feng，Facial Expression Recognition Based on Local Binary Patterns and Coarse-to-Fine Classification，Proceedings of the Fourth International Conference on Computer and Information Technology (CIT'04)
16. Yusuke Shinohara and Nobuyuki Otsu, Facial Expression Recognition Using Fisher Weight Maps, Proceedings of the Sixth IEEE International Conference on Automatic Face and Gesture Recognition (FGR'04)
17. Michael J. Lyons, Julien Budynek, and Shigeru Akamatsu, Automatic Classification of Single Facial Images, IEEE Trans.PAMI, vol. 21, no. 12, DECEMBER 1999
18. Zhengyou Zhang et al. Comparison Between Geometry-Based and Gabor-Wavelets-Based Facial Expression Recognition Using Multi-Layer Perceptron Automatic Face and Gesture Recognition, 1998. Proceedings. Third IEEE International Conference on 14-16 April 1998 Page(s):454 - 459
19. Kun Ma and Xiaoou Tang, Discrete Wavelet Face Graph Matching, Image Processing, 2001. Proceedings. 2001 International Conference on Volume 2, 7-10 Oct. 2001 Page(s):217 - 220 vol.2
20. Constantine L. Kotropoulos, Anastasios Tefas et al, Frontal Face Authentication Using Discriminating Grids with Morphological Feature Vectors, IEEE Trans. Multimedia, vol. 2, no. 1, March 2000
21. C. Bishop. Neural Networks for Pattern Recognition.Clarendon Press, Oxford, 1995.