Faical Expression Recognition by Combining Texture and Geometrical Features

Renjie Liu, Ruofei Du, Bao-Liang Lu*

Abstract Most of the existing facial expression recognition methods are based on either only texture features or only geometrical features. In this paper, we propose to improve the performance of facial expression recognition by combining both types of features using fuzzy integral. The geometric features used are the displacements of positions of feature points on the face. We first embed them in a lower dimensional manifold space, then use a modified version of Support Vector Machine (SVM) as the classifier. The texture features are boosted Gabor features. Since the dimension of Gabor features is quite high, we use Adaboost to select the most important features and then use SVM to classify them for different emotions. Finally, we combine these two methods using fuzzy integral. The experiment results show that our method significantly improves the performance of facial expression recognition.

1 Introduction

Facial expression is one of the most powerful, natural and immediate means for human beings to communicate their emotions and intentions. Automatic facial expression analysis plays a vital role in a wide range of applications such as human-computer interaction, data-driven animation and so on. Due to its wide application, it has drawn much attention and interests in recent years. Though much effort has been made, automatic recognition of facial expression remains difficult. The facial action coding system (FACS) is an objective and comprehensive coding system used in the

Renjie Liu[1], Ruofei Du[1], and Bao-Liang Lu[1,2,3]
[1]Center for Brain-Like Computing and Machine Intelligence
Department of Computer Science and Engineering
[2]MOE-Microsoft Key Laboratory for Intelligent Computing and Intelligent Systems
[3]Shanghai Key Laboratory of Scalable Computing and Systems
Shanghai Jiaotong University
No. 800, Dongchuan Rd., Shanghai, China
*Corresponding author, e-mail: bllu@sjtu.edu.cn

behavioral science. Any facial movements can be decomposed into 46 component movements, which roughly correspond to the individual facial muscles [1]. Much research has shown the importance of automatic facial action recognition [2, 3, 4].

Much research effort has been applied to facial expression recognition in the last two decades. The facial expressions under examination were defined by psychologists as a set of six basic facial expressions (anger, disgust, fear, happiness, sadness and surprise) [5]. A survey about recently proposed approaches can be found in [6]. Features employed by most of the existing methods are of two types: geometric features and appearance features. In [7], a shape model of 58 facial landmarks was used to give a set of features. In [8], a set of facial feature points around facial components such as mouth and eyes were used as features. They are both typical examples of methods based on geometric features. In [9] and [10], features were taken from a Gabor wavelets representation. In [11], a comprehensive study on expression recognition based on Local Binary Patterns (LBP) features was conducted. These are all typical examples of methods based on appearance features.

For facial action unit (FAU) detection, pattern recognition techniques such as P-CA, LDA and ICA have been used and compared on different features such as Gabor Wavelet and gray scale histogram [2]. In [12], optical flow and facial feature point tracking were used to perform facial expression information tracking. The extracted features were fed into a HMMs system for facial action detection. In [13], a facial image exhibiting a combination of FAUs is represented as a sparse linear combination of vectors from a basis constituting an over-complete dictionary. By solving an L_1-norm minimization, the detection problem is simplified to a rank maximization problem.

In this paper we first investigate the performance of two facial expression recognition methods which use either only texture features or only geometric features. In the first method, we extract the displacements of ASM feature points and embed them in a manifold space with lower dimension, and then we use a modification to the SVM proposed in [14] as the classifier. The modification stems from the fact that the Fisher's discriminant optimization problem for two classes is a constrained least-squares optimization problem. In the second method, we apply Gabor filters to face images and use Adaboost to select the most important features, which are then classified by SVM. Finally, we combine the two methods using fuzzy integral, which gives better performance than either classifier alone.

The remainder of this paper is organized as follows. In Section 2, we describe expression recognition with geometric features. In Section 3, we describe expression recognition with texture features. In Section 4, we introduce fuzzy integral and show how to use it to combine the two classifiers. In Section 5, we compare the results of four methods — two with one type of features, one with both types, one the state-of-the-art method - on the extended Cohn-Kanade database [18]. The paper is concluded in Section 6.

2 Facial Expression Recognition with Geometric Feature

2.1 Active Shape Model

It has been shown that the active shape model [15] is a good method for locating facial feature points. Generally speaking, ASM fits the shape parameters using optimization techniques such as gradient descent. The shape S of ASM comprises a set of landmarks, namely $S = [x_1, y_1, x_2, y_2, ..., x_n, y_n]$, where n is the number of landmarks and x_i, y_i are the coordinates of the ith landmark. In our method, we use the displacements of landmarks in the initial and last frame in a series of expression images, where the initial frame is a neutral expression and the last frame is the peak of one of the seven emotions. That is, if S_{k1} is the shape vector of the first frame in the kth series of expression images, and S_{kj} is the shape vector of the last frame in the same series, then $\Delta S_k = S_{kj} - S_{k1}$ is the displacements of landmarks in the series, and can be used as the kth feature vector for facial action detection and expression recognition. In our method, n is set to 68, so the ASM feature vector has 136 dimensions. Fig. 1 illustrates the landmarks on different expressions.

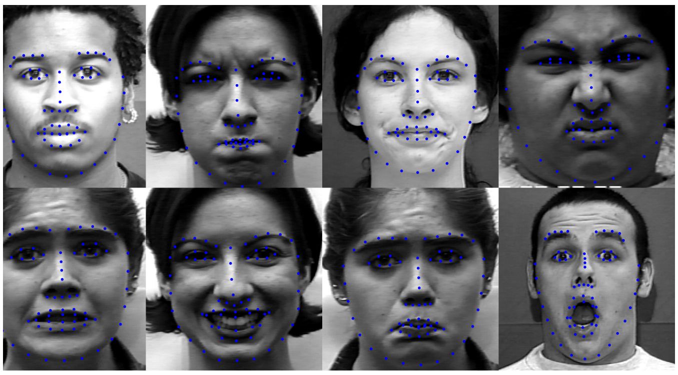
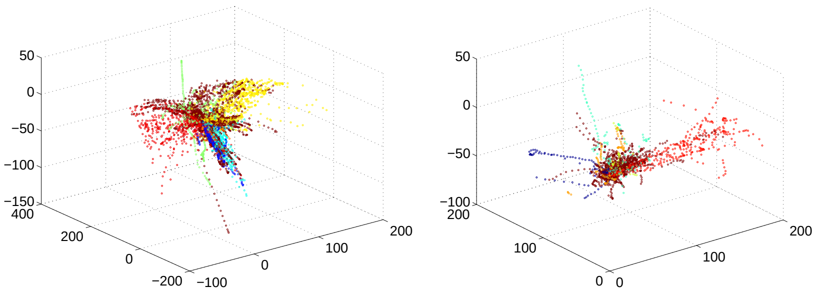


Fig. 1: Examples of landmarks on different expressions.

2.2 Manifold Learning

Classical dimensionality reduction methods such as Principal Component Analysis (PCA) are linear and efficient, but they are not suitable for facial expression recognition since the changes in facial expressions are inherently non-linear. In this paper, we proposed to embed the features in a manifold space with a lower dimension using the Isomap [19] algorithm. For the sake of illustration, we use PCA and Isomap to embed the differential ASM features to a 3-D space and show them in Fig. 2. It can be seen that the result of Isomap is much more separable than that of PCA.



(a) (b)

Fig. 2: Training data in the embed space: (a) result of Isomap; (b) result of PCA.

2.3 *Modified Support Vector Machine*

In [14], it has been shown that by combining statistical pattern recognition and Support Vector Machine (SVM), Fisher's discriminant ratio can be reformulated to a quadratic optimization problem subject to a set of inequality constraints. The detailed motivation can be found in [14]. When applying this approach to expression recognition, we need to define the scatter matrix on the training set first. Let $\mathbf{K}$ denote the number of classes, where $\mathbf{K} = 7$ for facial expression recognition, ω_k denote the set of training examples in the $\mathbf{k}$th class, and μ_k denote the mean vector of the $\mathbf{k}$th class. Then the scatter matrix is defined as follows:

$$\mathbf{S} = \sum_{i=1}^{\mathbf{K}} \sum_{g_i \in \omega_k} (g_i - \mu_k)(g_i - \mu_k)^T \tag{1}$$

Here we discuss only the two-class case in detail, since we can solve the multi-class problem using techniques such as "one-against-one" or "one-against-the rest". As proposed in [14], the optimization problem of the modified SVM is expressed as

$$\begin{aligned} &\min_{\mathbf{w},b,\varepsilon} \mathbf{w}\mathbf{S}\mathbf{w}^{\mathbf{T}} + \mathbf{C}\sum_{i=1}^{N} \varepsilon_i \\ s.t. \quad & y_i(\mathbf{w}^{\mathbf{T}}\mathbf{g_i} + \mathbf{b}) \geq 1 - \varepsilon_i, \varepsilon_i \geq 0 \end{aligned} \tag{2}$$

According to the Kuhn-Tucher (KT) conditions, we can derive the solution

$$\mathbf{w} = \frac{1}{2}\mathbf{S}^{-1}\sum_{i=1}^{N}\alpha_i y_i g_i \tag{3}$$

The resulting decision function is

$$\begin{aligned} f(g) &= sign(\mathbf{w}g^T + b) \\ &= sign(\frac{1}{2}\sum_{i=1}^{N}\alpha_i y_i g_i^T \mathbf{S}^{-1} g + b) \end{aligned} \tag{4}$$

There are many ways to solve the problem defined in Eq. (2). For example, you can use the *"quadprog"* function provided in MATLAB.

3 Facial Expression Recognition with Texture Features

3.1 Gabor Wavelet

Gabor-based features have been widely used in facial analysis tasks such as emotion recognition [9] and facial action unit detection [2]. In this paper, we convolve face images with a bank of Gabor filters at five spatial frequencies and eight orientations as proposed in [2]. Though we downsample the resulting image by a factor of sixteen, each feature still remains a vector with 81000 dimensions, which is much larger than the number of training samples. In this paper, we select the most important Gabor features using multi-class Adaboost.

3.2 Multi-class Adaboost

Adaboost [20] is a boosting algorithm that constructs a strong classifier by combining several weak classifiers. The traditional two-class Adaboost algorithm has been successfully applied in many areas. Schapire and Singer [20] have extended Adaboost to multi-class multi-label problems using a measure of Hamming loss. Here we denote the sample space by $\mathbf{X}$ and the label set by $\mathbf{Y}$. Each sample of the multi-class multi-label problem is a tuple $(\mathbf{x}, y)$, where $\mathbf{x} \in \mathbf{X}$ and $y \subseteq \mathbf{Y}$. For each label l in $\mathbf{Y}$, we define the following

$$y[l] = \begin{cases} 1 & if\ l \in y \\ -1 & if\ l \notin y \end{cases} \tag{5}$$

Algorithm 1 Multi-class Adaboost

Given:$((\mathbf{x_1}, y_1), \ldots, (\mathbf{x_m}, y_m))$
Initialize $D_1(i,l) = 1/(mk)$
for $t = 1$ *to* T **do**
Select a weak classifier h_t that maximizes the error measure

$$r_t = \sum_{i,l} D_t(i,l) Y_i(l) h_t(\mathbf{x_i}, l) \tag{6}$$

Let

$$\alpha_t = \frac{1}{2} ln \left(\frac{1 + r_t}{1 - r_t} \right) \tag{7}$$

Update:

$$D_{t+1}(i,l) = \frac{D_t(i,l) \exp(-\alpha_t Y_i[l] h_t(\mathbf{x_i}, l))}{Z_t} \tag{8}$$

where Z_t is a normalization factor.
end for

After selecting the most important features using Adaboost, we then feed them into SVM and classify them into different emotions.

4 Fuzzy Integral

Fuzzy integral is integrals of real functions with respect to a fuzzy measure. One of the most representative definitions is Choquet [21].
The discrete Choquet integral of function $f\colon X \to \mathbb{R}^+$ is defined by

$$C_u(f(x_1), \ldots, f(x_n)) = \sum_{i=1}^{n} f(x_{(i)}) - f(x_{(i-1)}) \mu(A_{(i)}) \tag{9}$$

where the subscript i indicates that the indices have been rearranged so that $0 \le f(x_{(1)}) \le f(x_{(2)}) \cdots f(x_{(n)}) \le 1$, and $A_{(i)} = \{x_{(i)}, \ldots, x_{(n)}\}$, and $f(x_{(0)}) = 0$.

4.1 Learning Fuzzy Measure

We use fuzzy integral as the classifier combination mechanism for the following two reasons:

- It is representative, since a lot of combination mechanisms, such as weighted sum, min or max rules, are special cases of it.
- We can represent the importance of an individual classifier and interactions among any subset of the classifiers using an appropriate fuzzy measure.

In this section, we will describe the application of fuzzy integral in combining classifiers. Here we assume that the number of classes is m, and $T = \{t_1, \ldots, t_m\}$ is the set of given labels. Let $X = \{x_1, \ldots, x_n\}$ be the set of classifiers and n be the number of classifiers. For an observed example A, we let $h_i^j(A)$ denote the probability that A belongs to class j given by classifier x_i. Then the probability that A belongs to class j is defined by the following:

$$C_{u_j}(A) = C_{u_j}(h_1^j(A), \ldots, h_n^j(A))$$

Therefore we can determine that A belongs to the class with highest probability. The last and most important step before applying a fuzzy integral to classifier combination, is to determine the importance of classifiers and the interaction among classifiers, namely the fuzzy measure. For a problem with n-classifiers and m-classes, we need to determine $m(2^n - 2)$ variables (since the fuzzy measures of $\emptyset$ and X are known to be 0 and 1, respectively). For the sake of simplicity, we assume here that $m = 2$ and the number of examples of the kth class is l_k. So the criterion for identifying the best fuzzy measure is to minimize the following squared error:

$$J = \sum_{i=1}^{l_1} (C_{u_1}(A_1^i) - C_{u_2}(A_1^i) - 1)^2 + \sum_{i=1}^{l_2} (C_{u_2}(A_2^i) - C_{u_2}(A_2^i) - 1)^2 \tag{10}$$

where A_k^i is the ith example in class k. This problem is a quadratic programming problem which can be solved using packages provided in MATLAB. As mentioned above, a fuzzy integral requires that the output of a classifier is in the form of a probability, so we choose SVM-PO with RBF kernel [17].

5 Experiments

We compare the performance of four algorithms for facial expression recognition on the extended Cohn-Kanade database [18]. In the first algorithm, we extract the Gabor representation of the facial image for five spatial frequencies and eight orientations and downsample it by a factor of sixteen. Then we use Adaboost to select the most important 600 features and feed them into and SVM classifier for expression recognition. We will refer to it as 'Adaboost+Gabor'. For the second algorithm, we extract the displacements of ASM feature points between the neutral expression and the peak frame of one emotion and embed them in a lower manifold space with the a dimension of 100, then we use a modified SVM classifier for expression recognition. We will refer to this as 'ASM+MSVM'. In the third algorithm, we combine the two classifiers using fuzzy integral; this will be referred to as 'Fuzzy'. Also, we list the results reported in [18], which will be referred to as 'SPTS+CAPP'. The confusion matrices of the four methods are listed in Tables 2-4. From these tables, we can see that the method proposed in this paper outperforms the existing approaches in facial expression recognition. We also compare the performance of combining the two classifiers using different combination strategies in [21], which will be referred to as 'Max', 'Sum', and 'Product'. The result shows that fuzzy integral outperforms all these classifier combination methods.

Table 1: Confusion Matrix for Facial Expression Recognition Using ASM+MSVM Method

	an	co	di	fe	ha	sa	su
an	**93.33**	0.00	0.00	0.00	6.67	0.00	0.00
co	0.00	**100.00**	0.00	0.00	0.00	0.00	0.00
di	10.00	0.00	**90.00**	0.00	0.00	0.00	0.00
fe	0.00	0.00	0.00	**87.50**	0.00	12.50	0.00
ha	0.00	0.00	0.00	0.00	**95.65**	0.00	4.35
sa	0.00	0.00	0.00	0.00	0.00	**100.00**	0.00
su	0.00	3.57	0.00	0.00	0.00	0.00	**96.43**

Table 2: Confusion Matrix for Facial Expression Recognition Using Adaboost+Gabor Method

	an	co	di	fe	ha	sa	su
an	**93.33**	0.00	0.00	0.00	0.00	6.67	0.00
co	16.67	**83.33**	0.00	0.00	0.00	0.00	0.00
di	0.00	0.00	**94.74**	0.00	0.00	5.26	0.00
fe	0.00	0.00	0.00	**87.50**	12.50	0.00	0.00
ha	0.00	0.00	8.70	0.00	**91.30**	0.00	0.00
sa	0.00	0.00	0.00	0.00	0.00	**100.00**	0.00
su	3.57	0.00	0.00	0.00	0.00	0.00	**96.43**

Table 3: Confusion Matrix for Facial Expression Recognition Using Fuzzy Method

	an	co	di	fe	ha	sa	su
an	**100.00**	0.00	0.00	0.00	0.00	0.00	0.00
co	0.00	**100.00**	0.00	0.00	0.00	0.00	0.00
di	0.00	0.00	**94.74**	0.00	0.00	5.26	0.00
fe	0.00	0.00	0.00	**87.50**	12.50	0.00	0.00
ha	0.00	0.00	0.00	0.00	**95.65**	0.00	4.35
sa	0.00	0.00	0.00	0.00	0.00	**100.00**	0.00
su	3.57	0.00	0.00	0.00	0.00	0.00	**96.43**

Table 4: Confusion Matrix for Facial Expression Recognition reported in [18]

	an	co	di	fe	ha	sa	su
an	**75.00**	5.00	7.50	5.00	0.00	5.00	2.50
co	3.10	**84.40**	3.10	0.00	6.30	3.10	0.00
di	5.30	0.00	**94.70**	0.00	0.00	0.00	0.00
fe	4.40	8.70	0.00	**65.20**	8.70	0.00	13.00
ha	0.00	0.00	0.00	0.00	**100.00**	0.00	0.00
sa	12.00	8.00	4.00	4.00	0.00	**68.00**	4.00
su	0.00	0.00	0.00	0.00	0.00	4.00	**96.00**

Table 5: Comparison of Different Methods

	an	co	di	fe	ha	sa	su	Avg
SPTS+CAPP	75.00	84.40	94.70	65.20	**100.00**	68.00	96.00	88.30
ASM+MSVM	93.33	**100.00**	90.00	**87.50**	95.65	**100.00**	**96.43**	94.50
Adaboost+Gabor	93.33	83.33	**94.74**	**87.50**	91.30	**100.00**	**96.43**	93.52
Max	93.33	33.33	80.00	**87.50**	53.33	80.00	92.86	77.48
Product	93.33	16.67	90.00	**87.50**	83.00	88.89	96.29	86.30
Sum	93.75	33.33	90.00	70.59	80.00	75.00	96.00	88.31
Fuzzy	**100.00**	**100.00**	**94.74**	**87.50**	**100.00**	**100.00**	**96.43**	**97.20**

6 Conclusion and Future work

In this paper, we have proposed a method for improving the performance of facial expression recognition by combining two methods which use only geometrical features or only texture features. The first one uses a modified SVM as the classifier and ASM feature points displacements embedded in a lower dimensional manifold space as features. The second one uses Adaboost to select the most important features from the result of Gabor filtering and uses SVM as the classifier. Then we combine the two methods using fuzzy integral. The experimental results show that our method achieves better performance in facial expression recognition, which is encouraging.

7 Acknowledgments

This work was partially supported by the National Natural Science Foundation of China (Grant No. 90820018), the National Basic Research Program of China (Grant No. 2009CB320901), and the European Union Seventh Framework Programme (Grant No. 247619).

References

1. P. Ekman and W. V. Friesen, Facial Action Coding System: A Technique for the Measurement of Facial Movement. Consulting Psychologists Press, San Francisco (1978).
2. G. Donato, M. S. Bartlett, J. C. Hager, P. Ekman, T. J. Sejnowski, Classifying Facial Actions. IEEE Trans. on Pattern Analysis and Machine Intelligence, vol. 21, no. 10, pp. 974-989, 1999.
3. M. Pantic and J. M. Rothkrantz, Facial action recognition for facial expression analysis from static face images. IEEE Trans. on Systems, Man and Cybernetics, vol. 34, no. 3, pp. 1449-1461, 2004.
4. 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, pp. 97-116, 2001.
5. P. Ekman and W. V. Friesen, Emotion in the Human Face. Pergamon Press, Oxford England (1975).
6. Zhihong Zeng, Glenn I. Roisman, Thomas S. Huang, A Survey of Affect recognition Methods: Audio, Visual, and Spontaneous Expressions. IEEE Trans. on Pattern Analysis and Machine Intelligence, vol. 31, no. 1, pp. 39-58, 2009.
7. Y. Chang, C. Hu, R. Feris, and M. Turk, Manifold Based Analysis of Facial Expression. CVPRW'04, pp. 81-81, 2004.
8. M. Pantic and L. J. M. Rothkrantz, Facial Action Recognition for Facial Expression Analysis from Static Face Images. IEEE Trans. on Systems, Man and Cybernetics, vol. 34, no. 3, pp. 1449-1461, 2004.
9. M. S. Bartlett, G. Littlewort, P. Braathen, T. J. Sejnowski, and J. R. Movellan, A Prototype for Automatic Recognition of Spontaneous Facial Actions. Advances in Neural Information Processing Systems, vol. 15, pp. 1271-1278, 2003.
10. M. S. Bartlett, G. Littlewort, M. G. Frank, C. Lainscsek, I. Fasel, and J. Movellan, Fully Automatic Facial Action Recognition in Spontaneous Behavior. International Conference on FGR, pp. 223-230, 2006.
11. C. Shan, S. Gong, P. W. McOwan, Facial expression recognition based on Local Binary Patterns: A comprehensive study. Image and Vision Computing, vol. 27, no. 6, pp. 803-816, 2009.
12. J. J. Lien, T. Kanade, J. F. Cohn, and C. Li, Detection, tracking and classification of action units in facial expression. Robotics and Autonomous Systems, vol. 31, no. 3, pp. 131-146, 2000.
13. M. H. Mahoor, Mu Zhou, K. L. Veon, S. M. Mavadati, and J. F. Cohn, Facial Action Unit Recognition with Sparse Representation. IEEE International Conference on FGR, pp. 336-342, 2011.
14. A. Tefas, C. Kotropoulos, and I. Pitas. Using support vector machines to enhance the performance of elasticgraph matching for frontal face authentication. IEEE Trans. on Pattern Analysis and Machine Intelligence, vol. 23, no. 7, pp. 735-746, 2001.
15. T. F. Cootes, C. J. Taylor, D. H. Coper, and J. Graham, Active Shape Models-Their Training and Application. Computer Vision and Image Understanding, vol. 61, no. 1, pp. 38-59, 1995.
16. C. Chang and C. Lin, LIBSVM: a library for support vector machines. ACM Transactions on Intelligent Systems and Technology, vol 2, no. 3, pp. 1-27, 2011. Software available at http://www.csie.ntu.edu.tw/ cjlin/libsvm
17. T. F. Wu, C. J. Lin, and R. C. Weng, Probability Estimates for Multi-class Classification by Pairwise Coupling. Journal of Machine Learning Research, vol. 5, pp. 975-1005, 2004
18. P. Lucey, J. F. Cohn, T. Kanade, J. Saragih, Z. Ambadar and I. Matthews, The Extended Cohn-Kanade Dataset (CK+): A complete dataset for action unit and emotion-specified expression. CVPRW, pp. 94-101, 2010.
19. Joshua B. Tenenbaum, Vin de Silva and John C. Langford, A Global Geometric Framework for Nonlinear Dimensionality Reduction. Science, vol. 290, no. 5500, pp. 2319-2323, 2000.
20. R. E. Schapire and Y. Singer, Improved boosting algorithms using confidence-rated predictions, Machine Learning, vol. 37, NO. 3, pp. 297-336, 1999.
21. Bing Li, Xiao-Chen Lian, Bao-Liang Lu, Gender classification by combining clothing, hair and facial component classifiers. Neurocomputing, vol. 76, No. 1, pp. 18-27, 2012.