Facial Expression Recognition based on Multiple Base Shapes

Lijun Cai, Lei Huang, and Changping Liu

Institute of Automation Chinese Academy of Sciences, Beijing, China

Abstract. Geometric variation is one of the important components deteriorating the facial expression recognition performance. Aligning the face image to a base shape is a commonly used preprocess step to alleviate the variation. However, the assumption of single base shape can not necessarily guarantee the best performance. In this paper, we propose for the first time a facial expression recognition framework based on multiple base shapes, which aims to minimize the geometric variation between face images with the same facial expression and retain the geometric shape difference between face images with different facial expressions. For a new sample, a weighed vote based criterion is used to give the final predicted facial expression given multiple base shapes. Experimental results on CK+ (Extended Cohn-Kanade) and JAFFE (Japanese Female Facial Expression databases) show the effectiveness of proposed method.

Keywords: Facial expression recognition, multiple base shapes, hybrid feature, weighted vote

1 Introduction

Facial expression is an effective way to express human emotion states, such as anger, sadness and happiness, therefore perceiving facial expression can be helpful to human interaction. With the development of artificial intelligence, researchers pay more and more attention on automatic facial expression recognition which plays an important role in human-machine interaction.

For two face images having the same facial expression, the differences include varying identities, gestures, illuminations, geometric variation and noise which is randomly distributed in face images. Geometric variation indicates that for a specific expression, face component displacements are not the same for different people, even for the same person at different time. Identity, geometric variation, gesture and illumination are four components deteriorating the facial expression recognition performance. Normally, noise is of small energy. In this paper, we focus on the geometric variation.

Face alignment is widely used to alleviate the geometric variation. Face alignment has been one of the most hot issues for many years and there are a lot of effective methods achieved so far [17]. When it comes to facial expression recognition, a common way for face alignment is to normalize images to a canonical

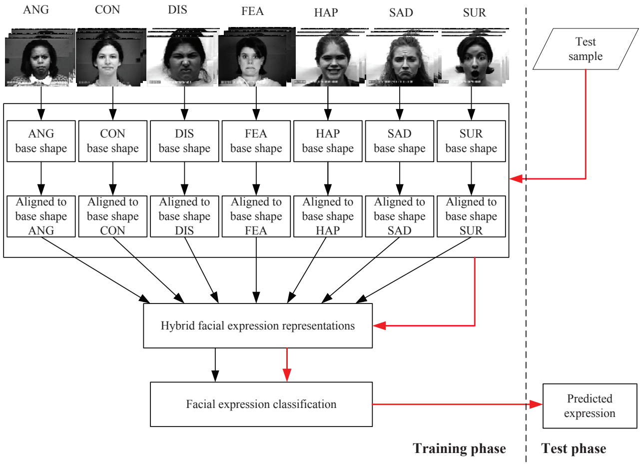


Fig. 1. The flowchart of our proposed method. ANG, CON, DIS, FEA, HAP, SAD and SUR are the acronym of Anger, Contempt, Disgust, Fear, Happiness, Sad and Surprise expressions.

template by an affine transformation. Particularly, face images are normalized by keeping the distances between two eye centers the same for all images [3–5, 13, 16]; face images are aligned to the mean shape of all facial expressions [7, 8, 14] or a fixed expression geometric shape with larger size in mouth/eyes regions [9].

To the best of my knowledge, all the current facial expression recognition methods simply adopt one base shape to be the canonical template, such as the mean shape of all facial expressions or a fixed expression geometric shape. However, from another perspective, decreasing the geometric variation between face images with different facial expressions can reduce the separability. In this paper, we propose a novel facial expression recognition method based on multiple base shapes aiming to decrease the geometric variation between face images having the same facial expression and retain the geometric variation between face images having different facial expressions. The generalized framework of proposed method is illustrated in Fig. 1 (take CK+ (Extended Cohn-Kanade) database [8] for example). It can be observed from the flowchart that each of multiple base shapes is generated based on samples with the same expression. In training phase, after aligned to its corresponding base shape, face image is described by the hybrid feature combining geometric and appearance feature. Then classifier is obtained based on SVM (Support Vector Machine) [18]. For a new sample, how to classify it into one category of expression given multiple base shapes? We will give a feasible solution to this question in the next section. Our contribution in this paper is to propose for the first time an assumption of one expression one base shape for facial expression recognition.

2 Generalized Framework of Facial Expression Recognition based on Multiple Base Shapes

2.1 AAM Derived Representations

Representations derived from AAM (Active appearance model) [10] used in [7] will be adopted in this paper. Here, we briefly explain these representations as follows.

Shape $\boldsymbol{s}$: the shape $\boldsymbol{s}$ of AAM is described by a 2D triangulated mesh. In particular, the coordinates of the mesh vertices define the shape $\boldsymbol{s}$ (see Fig. 2(a)), which correspond to a source appearance image (see Fig. 2(d)).

Base shape $\boldsymbol{s}_0$: shape $\boldsymbol{s}$ can be described as a base shape $\boldsymbol{s}_0$ plus a linear combination of finite shapes. Procrustes alignment [10] is used to estimate the base shape $\boldsymbol{s}_0$ (see Fig. 2(c)).

Rigid normalized shape $\boldsymbol{s}_n$: $\boldsymbol{s}_n$ gives the vertex locations after all rigid geometric variation (translation, scale, rotation), relative to the base shape $\boldsymbol{s}_0$, has been removed (see Fig. 2(b)).

Rigid normalized appearance $\boldsymbol{a}_n$: it represents the appearance after all rigid geometric variation removed (see Fig. 2(e)) and is obtained by affine warping pixels of the source appearance image into $\boldsymbol{s}_n$.

Non-rigid normalized appearance $\boldsymbol{a}_0$: we can obtained $\boldsymbol{a}_0$ by affine warping pixels of the source appearance image into $\boldsymbol{s}_0$ (see Fig. 2(f)).

Different from [7] using a single base shape $\boldsymbol{s}_0$, we propose a method based on multiple base shapes, each of which is generated by the Procrustes alignment from the shapes belong to the same expression (see Fig. 1). To emphasize the different expressions, in this paper, we use $\boldsymbol{s}_i^c$, $\boldsymbol{s}_{i,n}^c$, $\boldsymbol{a}_{i,n}^c$, $\boldsymbol{a}_{i,0}^c$, $c = 1, \cdots, C$, to indicate above mentioned AAM derived representations of sample i belonging to expression c, and $\boldsymbol{s}_0^c$ to indicate the base shape of expression c, where C is the total number of expressions.

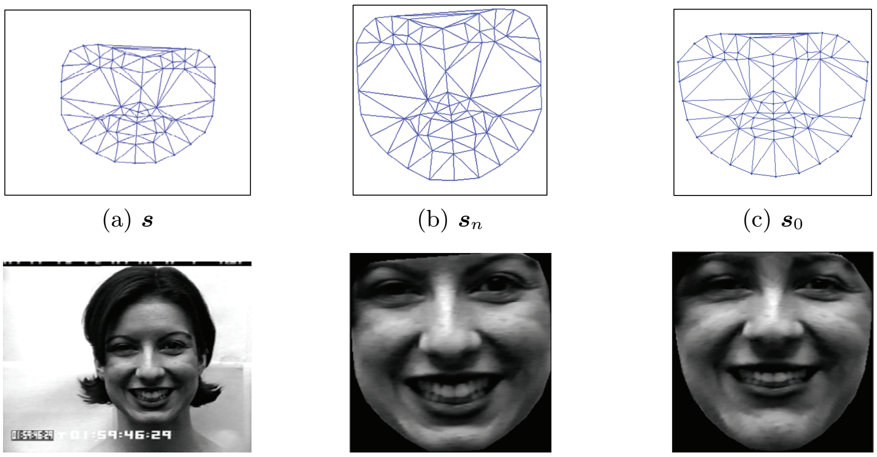
2.2 Hybrid Feature

In this paper, hybrid feature concatenating normalized geometric and appearance feature is adopted to represent the facial expressions.

For the geometric feature, based on the assumption that a data can be locally approximated by linear Euclidean subspace, Roweis et al. [11] present to describe each data point by coefficients that linearly reconstruct the data point from its neighbors, and they gave an algorithm to solve these linear coefficients. For an face image i with rigid normalized shape $\boldsymbol{s}_{i,n}$ and the set of its neighbors N^K (K-nearest neighbors), the reconstruction error on the cth expression space is calculated by

$$e_i^c = \min_{w_{i,j}^c} ||\boldsymbol{s}_{i,n} - \sum_{\boldsymbol{s}_{j,n}^c \in N^K} w_{i,j}^c \boldsymbol{s}_{j,n}^c||^2 \quad (1)$$

It is commonly used that image i is classified as class $\hat{c}$ if $\hat{c} = \min_c e_i^c$ [3, 4]. Therefore, we believe that $\boldsymbol{w}_i^c = [w_{i,1}^{\hat{c}}, \cdots, w_{i,K}^{\hat{c}}]^{\mathrm{T}}$ satisfying equation (1) can



(a) $\boldsymbol{s}$ (b) $\boldsymbol{s}_n$ (c) $\boldsymbol{s}_0$

(d) Source image (e) $\boldsymbol{a}_n$ (f) $\boldsymbol{a}_0$

Fig. 2. AAM derived representations. (a)Face shape $\boldsymbol{s}$; (b)Rigid normalized shape $\boldsymbol{s}_n$; (c)Base shape $\boldsymbol{s}_0$; (d)Source image; (e) Rigid normalized appearance $\boldsymbol{a}_n$; (f) Non-rigid normalized appearance $\boldsymbol{a}_0$.

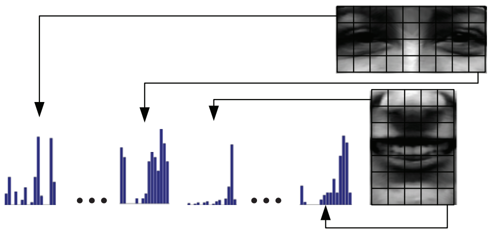


Fig. 3. Appearance feature.

be taken as the geometric feature to describe the shape information of samples belonging to expression $\hat{c}$.

For the appearance feature, according to the experimental results in [7], poor performance was gained by using rigid normalized appearance, thus only non-rigid normalized appearance is used in this paper. Considering that the micro-texture information of eye, nose and mouth regions plays an important role in facial expression recognition [9], patch-based local binary pattern histograms are extracted and concatenated to a long feature to form the appearance feature (see Fig. 3). As we know that LBP (Local Binary Pattern) [5] is an effective texture descriptor, to improve the generalization of training model, we use an improved version of LBP, DH-LBP (Dual Histogram Local Binary Pattern) [19], which has lower feature dimensions and remains the discriminative ability of LBP. In this paper, there are $36 + 35 = 71$ subresions and each subregion has 16-dimensional DH-LBP feature, totally $71 \times 16 = 1136$ dimensions.

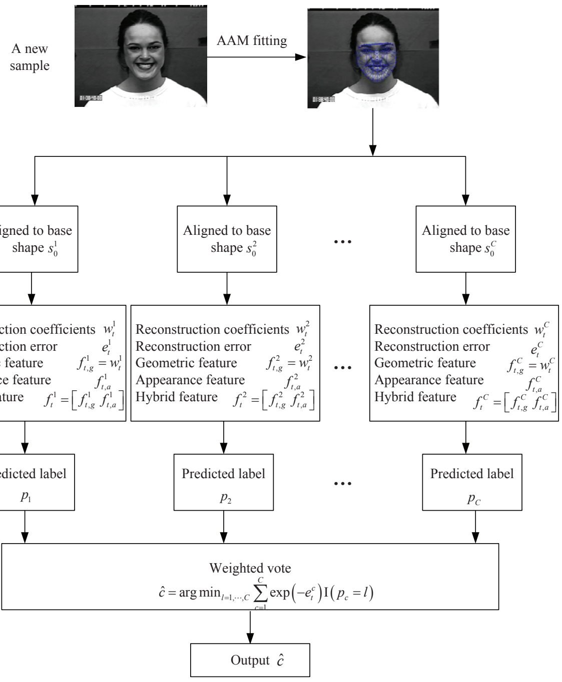


Fig. 4. Test phase of proposed method.

2.3 Classification based on Multiple Base Shapes

Given multiple base shapes, we propose a weighted vote based criterion to fulfill the classification of a new sample. The flowchart is illustrated in Fig. 4.

For a new sample, its shape $\boldsymbol{s}_t$ is firstly obtained by AAM fitting. Given the multiple base shape $\{\boldsymbol{s}_\mathbf{0}^\mathbf{c}\}_{c=1}^{C}$ obtained in the training phase, $\boldsymbol{s}_t$ is aligned to each of them. Then multiple groups of reconstruction coefficients $\{w_t^c\}_{c=1}^{C}$, reconstruction error $\{e_t^c\}_{c=1}^{C}$, geometric feature $\{f_{t,g}^c\}_{c=1}^{C}$, appearance feature $\{f_{t,a}^c\}_{c=1}^{C}$ and hybrid feature $\{f_t^c\}_{c=1}^{C}$ are obtained. For each group of hybrid features, a predicted label p_c is obtained by support vector machine with Radius basic function kernel. Then the following weighted vote based criterion determines the finally predicted class.

$$\hat{c} = \text{argmax}_{l=1,\cdots,C} \sum_{c=1}^{C} \exp(-\mathrm{e}_\mathrm{t}^\mathrm{c}) * \mathrm{I}\{p_c = l\} \tag{2}$$

where $\exp(-e_t^c)$ is the weight, the smaller the reconstruction error is, the bigger the weight. $\mathrm{I}(p_c = l)$ is 1 if $p_c = l$, else is 0. Equation (2) allows each base shape contributes a weighted vote to a specific class $l, l = 1, \cdots, C$, and the new sample is assigned to the class with the most weighted vote.

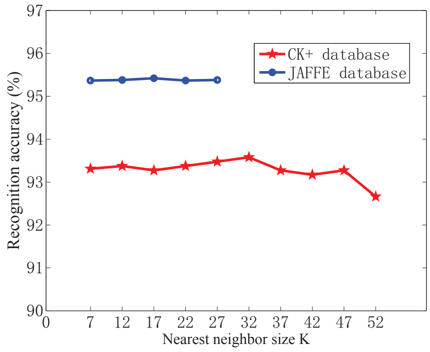


Fig. 5. Sensibility of proposed method to the choice of nearest neighbors size K.

3 Experiments

3.1 Databases

Two commonly used databases for facial expression recognition, CK+ [8] and JAFFE (Japanese Female Facial Expression) [20], are used in this paper.

CK+ database is the extended version of original CK (Cohn-Kanade) database. Aside from six expressions in CK database: Anger, Disgust, Fear, Happy, Sadness and Surprise, Contempt expression is added to CK+. There are totally 593 sequences from 123 subjects. Image sequences vary in duration and each sequence starts from the neutral face to peak frame of a specific expression. In this paper, three peak frames of every sequence are used for our experiments, and leave-one-subject-out cross-validation configuration is adopted. what's more, confusion matrix and recognition accuracy are uses as evaluation metrics, which are consistent with the baseline of CK+ database [8].

JAFFE database contains 213 images with 256×256 pixels of 7 facial expressions (6 basic facial expressions plus 1 neutral) posed by 10 Japanese female models. Each subject has 3 or 4 images for each facial expression. 2 images from each expression for each subject are selected in training step, and the rest of images from each expression are used as test images [21].

3.2 Experimental Results

Experimental section consists of three parts: 1) the sensibility of proposed method to the choose of nearest neighbors size K in equation (1). 2) Comparison with facial expression recognition based on single base shape. 3) Comparison with the state-of-the-arts.

Firstly, we will demonstrate the sensibility of proposed method to the nearest neighbors size K. The range of K is determined by the least number of samples for each expression. For CK+, $K \in [1, 54]$ and for JAFFE, $K \in [1, 29]$. Experimental results on two databases with step 5 are shown in Fig.5. It can be seen that proposed method is not sensible to the size K over a range of values. The result is not surprising because the idea of reconstruction coefficients and errors comes from [11], in which the authors have obtained the similar conclusion. In this paper, considering both the recognition accuracy and computation complexity, $K = 12$ is adopted in the following experiments.

Table 1 and Table 2 illustrate the confusion matrix on two databases. It can be observed that proposed method achieves good performance on both two public facial expression databases. However, recognition of Sadness expression in CK+ database is still difficult because of the database characteristics.

Table 1. Confusion matrix of proposed method on CK+ database(%).

	ANG	DIS	FEA	HAP	SAD	SUR	CON
ANG	**94.07**	2.96	0	0	0	0.74	0.22
DIS	2.96	**96.05**	0	0	0	1.09	0
FEA	5.33	0	**81.34**	5.33	4	2.67	1.33
HAP	0	0.96	0	**99.03**	0	0	0
SAD	10.71	5.95	0	0	**78.57**	1.19	3.57
SUR	0	0	0	0	0	**98.90**	1.20
CON	5.56	0	0	0	0	0	**94.44**
Average	**93.37**						

Table 2. Confusion matrix of proposed method on JAFFE database(%).

	ANG	DIS	FEA	HAP	SAD	SUR
ANG	**100**	0	0	0	0	0
DIS	0	**100**	0	0	0	0
FEA	0	0	**91.67**	0	0	8.33
HAP	0	0	0	**90.91**	0	9.09
SAD	0	0	0	9.09	**90.91**	0
SUR	0	0	0	0	0	**100**
Average	**95.38**					

Secondly, comparison between proposed method based on multiple base shapes and method based on single base shape is performed. Single base shape based method has only one base shape (mean shape) which is acquired according to the Procrustes alignment based on all training samples. Given the same facial expression representation and classifier training, recognition accuracies on two databases are shown in Table 3. It can be observed that proposed method based on multiple base shapes achieves better performance than the method based on single base shape. Therefore, retaining the geometric shape difference between face images with different facial expressions is helpful to facial expression recognition.

Table 3. Compared recognition accuracies of method based on single base shape and proposed method based on multiple base shapes(%).

Method	CK+	JAFFE
Single base shape	90.32	93.33
Proposed method	**93.37**	**95.38**

Lastly, compared results with the state-of-the-arts on two databases are given in Table 4 and Table 5. For CK+ databse, to compare with existing methods, proposed method is performed using different class numbers and cross validations. Class number is 6 means that all expressions except Contempt are involved into the experiment. It is demonstrated in Table 4 and 5 that compared with existing popular methods, proposed method obtains a comparable performance given the same class number and cross validation. All the experimental results illustrated above shows that our proposed method based on multiple base shapes are effective.

Table 4. Compared accuracy with State-of-the-Arts.

Methods	Class number	Accuracy(%)	Cross validation
Proposed method	7	**93.37**	leave-one-subject-out
Proposed method	6	**94.31**	leave-one-subject-out
Proposed method	7	**96.22**	10-fold
Proposed method	6	**96.33**	10-fold
Proposed method	7	**95.82**	5-fold
Proposed method	6	**96.01**	5-fold
Lucey[8]	7	88.33	leave-one-subject-out
Islam[12]	7	90.10	10-fold
Ptucha[13]	7	91.40	leave-one-subject-out
Shan[5]	7	88.9	10-fold
Shan[5]	6	92.6	10-fold
Sadeghi[9]	6	94.48	10-fold
Sadeghi[9]	6	94.16	5-fold
Jain[14]	6	95.79	4-fold
Khan[15]	6	95.3	10-fold

Table 5. Compared accuracy with State-of-the-Arts on JAFFE database.

Methods	Class number	Accuracy(%)
Proposed method	6	**95.38**
Sadeghi[21]	6	91.23

4 Conclusion

In this paper, we propose a novel facial expression recognition method based on multiple base shapes. Compared with most existing methods utilizing single base shape, this method performs face shape alignment on the assumption of one facial expression one base shape. In the training phase, each of multiple base shapes is generated based on the shapes with the same facial expression, which can minimize the geometric variation between intra-expression face images and retain the geometric shape difference between extra-expression face images. In the test phase, a new sample are firstly aligned to all the base shapes respectively and then weighted vote based criterion is used to determine the final predicted facial expression. Experimental results on CK+ and JAFFE databases show the effectiveness of proposed multiple base shape based method. However, this method is not limited to facial expression recognition and will be applied to other face attributes recognition in our next work, for example the age estimation.

References

1. Kotsia, I., Pitas, I.: Facial expression recognition in image sequences using geometric deformation features and support vector machines. IEEE Trans. Image Proc. 16, 172–187 (2007)
2. Pai, N.S., Chang, S.P.: An embedded system for real-time facial expression recognition based on the extension theory. Comput. Math. Appl. 61, 2101–2106 (2011)
3. Xiao, R., Zhao, Q.J., Zhang, D. Shi, P.F.: Facial expression recognition on multiple manifolds. Pattern Recog. 44, 107–116 (2011)
4. Zafeiriou, S., Petrou, M.: Sparse representations for facial expressions recognition via l1 optimization. In: IEEE Computer Society Conference on Computer Vision and Pattern Recognition Workshops, pp. 32–39. IEEE Press, San Francisco (2010)
5. Shan, C.F., Gong, S.G., McOwan, P.W.: Facial expression recognition based on local binary patterns: A comprehensive study. Image Vision Computing 27, 803–816 (2009)
6. Buciu, I., Kotropoulos, C., Pitas, I.: ICA and Gabor representation for facial expression recognition. In: International Conference on Image Processing, pp. 855–858. IEEE Press (2003)
7. Ashraf, A.B., Lucey, S.C., Jeffrey F., Chen, T., Ambadar, Z.P., Kenneth M. Solomon, P.E: The painful face–pain expression recognition using active appearance models. Image Vision Computing 27, 1788–1796 (2009)
8. Lucey, P. Cohn, J. Kanade, T. Saragih, J. Ambadar, Z. Matthews, I.: The Extended Cohn-Kanade Dataset (CK+): A complete dataset for action unit and emotion-specified expression. In: IEEE Computer Society on Computer Vision and Pattern Recognition Workshops, pp. 94–101. IEEE Press, San Fransico (2010)

9. Sadeghi, H., Raie, A.A., Mohammadi, M.R.: Facial expression recognition using geometric normalization and appearance representation. In: 8th Iranian Conference on Machine Vision and Image Processing, pp. 159–163. IEEE Press, Zanjan (2013)
10. Cootes T. Edwards G. Taylor, C.: Active appearance models. IEEE Trans. Pattern Anal. Machine Intell. 23, 681–685 (2001)
11. Roweis, S.T.,Saul, L.K.: Nonlinear dimensionality reduction by locally linear embedding. Science 290, 2323–2326 (2000)
12. Islam, M.S., Auwatanamongkol, S.: A novel feature extraction technique for facial expression recognition. Int. J. Comput. Science 10, 9–14 (2013)
13. Ptucha, R.,Savakis, A.: Manifold based sparse representation for facial understanding in natural images. Image Vision Computing 31, 365–378 (2013)
14. Jain, S., Hu, C., Aggarwal, J.: Facial expression recognition with temporal modeling of shapes. In: IEEE International Conference on Computer Vision Workshops, pp. 1642–1649. IEEE Press, Bacilona (2011)
15. Khan, R.A., Meyer, A., Konik, H., Bouakaz, S.: Human vision inspired framework for facial expressions recognition. In: International Conference on Image Processing, pp. 2593–2596. IEEE Press, Orlando (2012)
16. Jeni, L., Girard, J., Cohn, J., Fernando D.L.T.: Continuous AU intensity estimation using localized, sparse facial feature space. In: 10th IEEE International Conference on Automatic Face and Gesture Recognition, pp. 1–7. IEEE Press, Shanghai (2013)
17. Cao, X.D., Wei, Y.C, Wen, F., Sun, J.: Face alignment by explicit shape regression. Int. J. Comput. Vision 107, 177–190 (2014)
18. Chang, C.C., Lin, C.J.: LIBSVM: a library for support vector machines. ACM Trans. Intell. Syst. Techno 2, 27 (2011)
19. Ma, W.H., Huang, L., Liu, C.P.: Advanced local binary pattern descriptors for crowd estimation. In: Pacific-Asia Workshop on Computational Intelligence and Induatrial Application, pp. 958–962. IEEE Press, Wuhan (2008)
20. Lyons, M., Akamatsu, S., Kamachi, M., Gyoba, J.: Coding facial expressions with Gabor wavelets. In: Third IEEE International Conference on Automatic Face and Gesture Recognition, pp. 200–205. IEEE Press, Nara (1998)
21. Sadeghi, H., Raie, A.A., Mohammadi, M.R.: Facial Expression Recognition Using Texture Description of Displacement Image. J. Inf. Syst. Telecom. 2, 205–212 (2014)