Unified Framework For Classifying Facial Images Based On Facial Attribute-Specific Subspaces And Minimum Reconstruction Error

Shiguang Shan[1], Wen Gao[1, 2], Yan Lu[2], Bo Cao[1], Xilin Chen[2], Debin Zhao[2], Wenbin Zeng[1]
[1] ICT-YCNC FRTJDL, Institute of Computing Technology, CAS, Beijing 100080,China
[2]Department of Computer Science, Harbin Institute of Technology, Harbin, 150001, China
{sgshan, wgao, ylu, bcao}@ict.ac.cn; xlchen@cs.cmu.edu, dbzhao@ict.ac.cn

Abstract

In this paper, a unified framework for classifying facial attributes is presented. Facial Attribute-Specific Subspace (FASS) is firstly proposed to represent each specific facial attribute. Then a framework is provided to classify facial images based on FASS and the Minimum Reconstruction Error (MRE) rule. The proposed framework is motivated by, but essentially different from the conventional Eigenface based methods, since, in our framework, similarity is measured by the reconstruction error. To evaluate the performance of the proposed method, it is applied to several face perception applications, such as face recognition, expression analysis, gender discriminating, and glasses detection. Extensive experiments in several face databases have demonstrated the impressive effectiveness and excellent robustness of the proposed framework against appearance variance due to changeable imaging conditions.

1. Introduction

To endow computers the ability to perceive facial attributes information, such as identity, emotional status, gender, race, age, and so on, it is essential to build more intelligent and intuitive Human-Computer Interface (HCI). Related research activities have significantly increased over the past few years as reviewed in [1,2,3,4].

Among these face perception tasks, face recognition is the most representative one and its development will greatly facilitate the resolution of other face perception tasks. Since the beginning of the 1990s, appearance based technologies have been the dominant methods, from which the two FRT categories were derived: one is holistic appearance feature based and the other is analytic local feature based. Popular methods belonging to the former paradigm include Eigenface [5], Fisherface [6], SVD and most NN based FRTs etc. Local Feature Analysis (LFA) [7] and Elastic Graph Matching (EGM) [8] are typical instances of the latter category. In recent years, Eigenface based methods [5,11], Gabor wavelet based Elastic Bunch Graph Matching (EBGM) technologies [8], active appearance model [9], and Fisherface [6]/LDA based approaches have attracted much attention. FERET evaluation has provided extensive comparisons among these algorithms [10]. More recently, SVM has been successfully applied to face recognition [11]. Representative resolutions to pose and illumination variation problems include Kriegman's illumination cone [12], Vetter's linear object class methods [13] and Shashua's "Quotient Image" [14].

This paper extends the Eigenface method by proposing the idea of representing each specific attribute by using a "Facial Attribute-Specific Subspace (FASS)". Then we propose a unified framework for facial attribute classification and apply it to several face perception tasks, such as face recognition/verification, expression recognition, gender classification, glasses detection, and pose estimation.

The remaining of the paper is organized as follows: In Section 2, some observations on Eigenface are presented, which lead to our FASS based method. Section 3 describes the FASS based framework in detail. Its applications and corresponding experiments in face recognition, expression analysis, gender classification, glasses detection, and pose estimation are presented in Section 4. A conclusion is drawn in the last section.

2. Observations On Eigenface Methods

As is well known in the face recognition community, Eigenface is essentially based on the idea that face images can be regarded as points in the high dimensional image space. They are believed to approximately form a subspace, so called "face subspace". Fig.1 visually illustrates the idea by describing an input face image as the linear combination of some leading Eigenfaces.

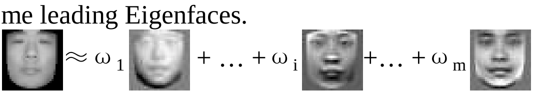

Fig.1 One face image is represented as the linear combination of leading Eigenfaces

A recognized nature of Eigenface method is the DFFS (Distance From Face Subspace), i.e. reconstruction error, which can be used to measure the extent of face pattern "hiding" in the input image or its similarity to face.

Further experiments are conducted to illustrate the effects of different Eigenfaces by reconstructing different input image patterns. As is shown in Fig.2, the patterns in each line, from left to right, are the original patterns and the patterns reconstructed by using the first 10, 30, 50, 70, 90, 100, 150, 200, 250, 300 Eigenfaces, respectively. In the first line, a face with salient characteristic (man-made "mole") is reconstructed, from which we can see that the "mole" cannot be portrayed when fewer leading Eigenfaces are used. When enough Eigenfaces are considered, the "mole" does emerge to some degree, however, much noise comes up together with the "mole". Line 2~4 illustrate the ability of the Eigenfaces to reconstruct non-face patterns, in which input patterns are monkey face, gates with flowers and tessellated windows respectively. It is interesting but understandable that they are reconstructed like common human faces when only a few Eigenfaces are considered. But the non-face patterns are also recovered better and better when more and more Eigenfaces are used. This suggests that fewer Eigenfaces provide favorable power to discriminate face patterns from non-face patterns, but the power declines with the increasing number of principal Eigenfaces involved.

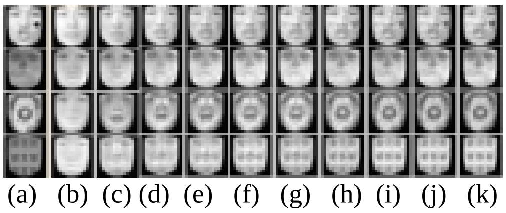

Fig. 2 Ability of different Eigenfaces to reconstruct face/non-face patterns

Based on the previous observations, we argue that PCA representation may mainly extract the features of the input pattern as a common face, but not individual features that discriminate different subjects. So the Eigenface method may be more suitable for the detection of face pattern, that is, less DFFS means more similarity to face pattern. Based on this point, we derive the idea that, if a subspace is learnt from the face examples of one specific attribute, correspondingly, it can be employed to detect the occurrence of face patterns with this specific attribute, i.e., less DFFS means more resemblance to the specific attributes. So the FASS based unified framework is proposed in the following section.

3. Unified Framework For Classifying Facial Attributes

Since all facial images with the same specific facial attribute (e.g. all male's facial images) have similar appearance, they form a subspace in the image space. So one private signal subspace can be used to model them, by which the invariant facial feature belonging to the same attribute is mostly retained as the expected signal, while most of the inter-attributes deformation that is useless for classification is thrown away into the "noise" subspace.

Let a facial attribute (e.g. gender) set with p physiologically separable attributes (for the "gender" case, p=2: male and female) be $C = \{\Omega_1, \Omega_2, \cdots, \Omega_p\}$. Each specific attribute $\Omega_k, k = 1,2,\cdots, p$ in C is analyzed using eigenvalue decomposition. Then the signal subspace for the k-th specific attribute is spanned by the leading eigenvectors:

$$S_{attri}^{(k)} = Span\{U_{attri}^{(k)}\} = span\{v_1^{(k)}, v_2^{(k)}, ..., v_{d_k}^{(k)}\}.$$

It is named the k-th FASS. Any face image Γ can be projected to the k-th FASS $S_{attri}^{(k)}$ by a matrix transform:

$$W^{(k)} = U_{attri}^{(k)\,T} \Phi^{(k)},$$

in which $\Phi^{(k)} = \Gamma - \Psi^{(k)}$ is the difference image and $\Psi^{(k)}$ is the mean image obtained from the training images corresponding to the k-th attribute. And the input image can be reconstructed by the linear combination of the leading eigenfaces:

$$\Phi_r^{(k)} = U_{attri}^{(k)} W^{(k)}.$$

Then the distance of any input face image from the k-th FASS, i.e. reconstruction error, can be computed as the Euclid distance between the original and the reconstructed pattern:

$$\varepsilon^{(k)} = \left\| \Phi^{(k)} - \Phi_r^{(k)} \right\|.$$

We denote the distance as DFFASS, which can measure the similarity between the image and the k-th attribute.

Fig. 3 visualizes the leading Eigenfaces of specific attributes, from which distinct facial characteristics of the corresponding attribute can be clearly seen. The images in the first line illustrate the leading Eigenfaces learnt from 15 example images with "surprised" attribute. And the Eigenfaces in the second line are learnt from 15 images all with glasses. The last line shows the leading Eigenfaces learnt from one frontal example image of the subject No.1 in Yale face database [6]. Obviously, the Eigenfaces of specific attribute distill the general characteristics of the corresponding attribute.

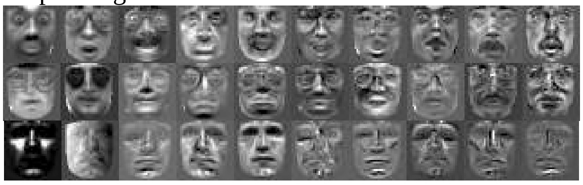

Fig.3 Leading ten Eigenfaces of surprising-FASS, glass-FASS, and subject-No-1-FASS from Yale face database

3.1 Minimum Reconstruction Error Classifier

As has been mentioned, DFFASS reflects the similarity of the input pattern to the facial attribute from which FASS is trained. It can be formally defined as follows:

Let Γ be any input image. It can be projected to the *k*-th FASS by:

$$W^{(k)} = U_{attri}^{(k)\,T} \Phi^{(k)},$$

and

$$\Phi^{(k)} = \Gamma - \Psi_k \ ,$$

where Ψ_k is the mean of *the k-th* attribute. Then $\Phi^{(k)}$ can be reconstructed by:

$$\Phi_r^{(k)} = U_{attri}^{(k)} W^{(k)} .$$

So, Γ 's distance from *the k-th* FASS, that is, the reconstruction error, can be computed as:

$$\varepsilon^{(k)} = \left\| \Phi^{(k)} - \Phi_r^{(k)} \right\| .$$

The DFFASS reflects the quantity of *the k-th* facial attribute "hiding" in the input image Γ, or in other words, the power of the *k*-th FASS to reconstruct the input pattern Γ. Thus, it can be regarded as the similarity of the input pattern Γ to the face samples that form the FASS. So the following minimal reconstruction error (MRE) classifier can be obtained:

$$\Gamma \in \Omega_m \ \text{ if } \ \varepsilon^{(m)} = \min_{1 \le k \le p} \{\varepsilon^{(k)}\} .$$

To demonstrate the rationality of the framework intuitively, further reconstruction experiments are conducted on FASS for different input patterns. To get comparable visual effects, the reconstruction is carried out based on the following formula:

$$\Gamma' = \left\| \Gamma - \Psi^{(k)} \right\| \cdot (\Phi_r^{(k)} + \Psi^{(k)}) .$$

As an example, Fig. 4 illustrates the power of one specific FASS to reconstruct various input patterns, in which the specific attribute is "identity". The first line shows the first 11 Eigenfaces of No.196 FASS trained from one face image of subject No.196 (Refer to section 4.1.4). In the first column from line 2 to line 6 are the input patterns. The subsequent pictures in each line illustrate the reconstructed patterns by using the leading 10~19 Eigenfaces of No.196 FASS. The input face in line 2 belongs to subject No.196, and we can see that the reconstructed faces are quite similar in appearance to the input face. The input faces in line 3~5 are non-No.196 faces. The reconstructed faces are quite different from the corresponding input face but still very similar to No.196's face. The last line illustrates the case when one non-face pattern is fed into the reconstructing procedure, where much more difference between the input pattern and reconstructed ones can be observed clearly.

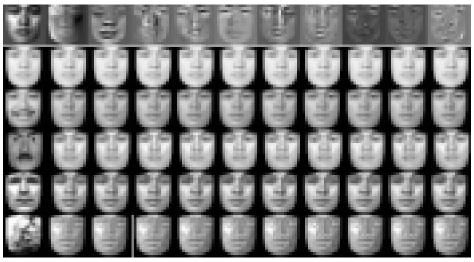

Fig.4 Eigenfaces of No.196 subject and its ability to reconstruct different patterns

Apparently, FASS possesses the favorable nature to reconstruct its own face patterns perfectly, while it is not the case for face patterns of other attributes. This strongly suggests that the FASS based face representation has outstanding class discriminating power.

4. Applications In Facial Perception

Based on the unified framework for facial attribute classification, we do abundant experiments to verify its effectiveness in five face analysis applications: face recognition, expression analysis, gender classification, and glasses detection. Similar methods can be used to categorize facial images according to race and age.

4.1 Face Recognition From A Single Face Image

As we know, to learn a FASS, multiple training example images with the specific attribute are required. For face recognition (i.e. classification according to different "identity"), it means more than one example per subject is needed to train his/her FASS. But for some other face recognition applications, such as mug shot matching, suspect identification etc., in which only a few (even single) face images are available for each subject involved, FASS based method cannot be applied directly. To solve this problem, we developed a simple technique to derive multiple samples from single example image. The technique is based on the following two intuitive propositions:

1. Proper geometric transforms, such as translation, rotation in image plane, scale variance etc., do not change the identity attribute of a face image visually.
2. Proper gray-level transforms, such as simulative directional lighting, man-made noise, etc., do not change the identity attribute of a face image visually.

In our system, the two kinds of transforms are combined to derive tens of training examples from single example image, which are then fed into the FASS learning procedure. Fig. 5(c) illustrates some normalized "virtual" examples derived from one face image as shown in Fig. 5(a) by utilizing our technique.

In addition, to alleviate the influence of translation, ration, lighting and scale variance, geometric and gray-level normalization are adopted. As to geometric normalization, the locations of the two irises are first localized manually and then fixed at specific locations by affine transformation. A mask, as shown in Fig.5 (b), is covered over the face region to eliminate the alterable background and hairstyle. Finally all faces are warped to the size of 32x32 as shown in Fig.5 (d). Histogram equalization is conducted to normalize illumination, and all the face data are vectorized to unit length before they are fed into the training or testing procedure.

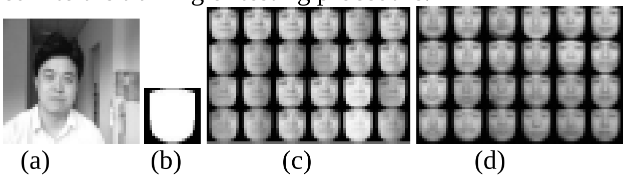

Fig.5 Deriving multiple samples from single image and normalization (a) input face image (b) mask (c) derived multiple examples from face in (a) (d) normalized faces

To verify the effectiveness of the proposed framework, we have also developed Eigenface method and template matching as benchmark algorithms. Extensive experiments are conducted on Yale face database, Bern face database, and our own face database containing 350 different subjects.

4.1.1 Benchmark Design And Performance Evaluation

Eigenface and template matching method are de facto the standard benchmarks in the face recognition community. We design the Eigenface method according to [5]. All faces are normalized as in Fig.5 (d). Template matching is operated on the normalized faces as shown in Fig.5 (d). Similarity between two faces is measured by using the cosine of the angle between the two vectors. Similar performance evaluation methodology as utilized in FERET evaluation [10] is adopted. The performance is evaluated and compared by using Cumulative Recognition Rate (CRR).

4.1.2 Experiments On Yale 15 Subjects Face Database

The Yale face database contains 165 images from 15 subjects, with 11 images per subject, among which there is a normal face with neutral expression, taken under ambient lighting conditions, while the left 10 images cover different cases including faces with/without glasses, images with basic expressions (happy, sad, sleepy, wink, surprised), images illuminated by center-light, left-light and right-light. All faces are frontal views. (Refer to [6] for details.)

In our experiment, all the 15 normal face images (one for each subject) are chosen to form the training set and gallery set, and all the other images (150 images) constitute the probe set for all the algorithms tested. The performance curves of different methods are plotted in Fig. 6. It is clear that our proposed method extraordinarily outperforms the other approaches. The top-ranking (first-choice) recognition ratio of our method is 95.33%, while that of the Eigenface method is 74.67%.

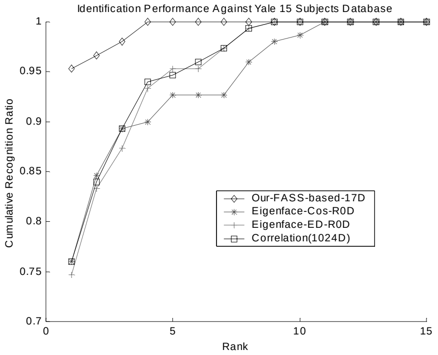

Fig.6 Performance comparisons on Yale face database

[6] is a well-known paper with experiments on Yale face database, in which Fisherface method is proposed. Note that the setup of our experiments is quite different from what is described in that paper, where error rates were determined by the "Leave-One-Out" strategy. But relative performance can still be compared as shown in Table 1. When comparing these data, readers must keep in mind that, in the "Leave-One-Out" strategy, ***ten*** training examples are learnt for each subject, except for the test person who is represented by ***nine*** [6]. However, in our methods only ***one*** example image for each subject is provided for the training procedure. It is obvious that our case is much more difficult than the "Leave-One-Out" strategy. Nevertheless our method outperforms all the other methods tested.

Table 1. Comparisons with methods in [6]

"Leave-one-out"		One example per subject	
Methods in [6]	Error Rate (Cropped)		Our Methods
Eigenface (W/O 1st 3)	15.3*	25.3	Eigenface
Correlation	23.9*	24.0	Correlation
Subspace	21.6*	24.0	Eigenface+Cos
Fisherface	7.3*	**4.7**	FASS based

*Note: Quoted data are from [6].

4.1.3 Experiments On Bern 30 Subject Multi-Pose Face Database

To further verify the effectiveness of the proposed framework on multi-pose face recognition problem, comparative experiments are conducted on Bern 30 subjects multiple-pose face database. The Bern database consists of 300 example images of 30 subjects, for each subject 10 gray-level images with slight variations of the head positions (1,2 facing the camera, 3,4 facing right, 5,6 facing left, 7,8 downwards, 9,10 upwards)*.

In our experiments, the No.“1” examples (looking right into the camera) of each subject in the database are chosen as the example images to form the training set (30 examples totally). The performance curves of different methods are plotted in Fig. 7. It is clear that our proposed method extraordinarily outperforms the other approaches.

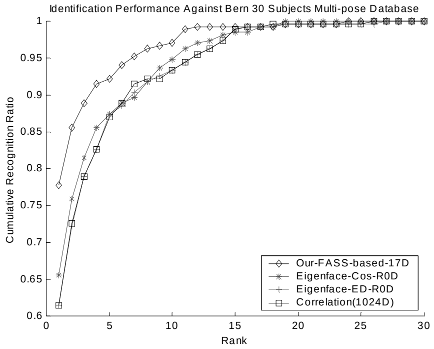

Fig. 7 Performance comparison on Bern face database

4.1.4 Experiments On Our 350 Subjects Face Database

To further demonstrate the performance and scalability of our FASS based method on larger database, more detailed experiments are conducted on a 350-subject face database. For the 350 subjects, 1750 images are acquired, with 5 images per subject. All images are taken with a general USB camera. For each subject, 1 normal face (nearly frontal, neutral expression and ambient lighting condition) is chosen as the training example; therefore a training/gallery set containing 350 faces is constructed. All the remaining 1400 images (4 examples per subject) constitute the probe set, which cover face images with different expressions, lighting conditions and minor pose variance. Apparent difference can be observed between the images in the gallery set and the probe set. The performances of different methods are compared in Fig. 8. Notably the proposed method outperforms all other algorithms. The Rank-1 (first-choice) recognition rate of our method is 88.36%, while that of the improved Eigenface method is 61.57%.

4.1.5 Observations on Experimental Results

These experiments clearly indicate the outstanding performance of our FASS based method compared with the two benchmarks, in which over 20% improvement is achieved in the same training/testing context. The results sufficiently demonstrate the adaptability and scalability of our method to expression, lighting, and slight pose variance, in that the Yale database covers various expressions and varying illuminations, Bern database involves faces with pose variance, and our database contains 350 subjects with various expressions and varying poses.

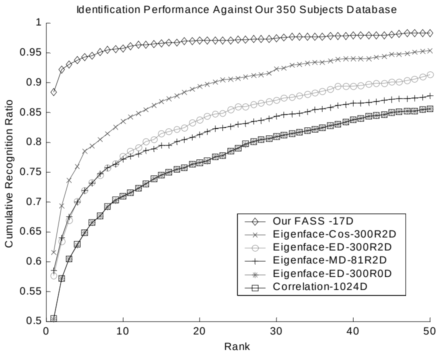

Fig. 8 Performance comparison on our face database

4.2. Facial Expression Classification

Facial expression has been studied for quite a long time. Ekman classified the human facial expressions into six main categories: happiness, sadness, anger, disgust, surprise and fear. In this paper, the method based on the FASS is used to recognize Ekman’s six static facial expressions by constructing six subspaces specific to the six facial expressions respectively and applying the proposed unified classifying framework. Details can be found in [15]. Table2 shows our recognition results.

Table 2. Expression recognition results

	Hap	Surp	Fear	Sad	Ang	Dis	Neut	Total
Sample Number	77	80	32	51	98	68	64	470
Correctly recognized	71	73	25	44	91	57	60	421
Ratio(%)	92.2	91.2	78.1	86.2	92.9	83.8	93.8	89.6

4.3. Gender Classification

To further verify the proposed framework, we have done experiments for gender classification as well. Two linear subspaces respectively specific for male and female are constructed based on 550 male facial image and 550 female facial images. Then we test the algorithm on a testing set containing 1048 faces (596 females and 452 males, none of them are in the training set), the recognition ratio is up to 87.1%. Table.3 shows the detailed results of these experiments.

Table 3. Gender Classification Results

	Male	Female	Total
Training Samples	550	550	1100
Testing Samples	452	596	1048
Correctly recognized	390	523	913
Ratio(%)	86.3	87.8	87.1

* Copyright 1995 University of Bern All Rights Reserved

4.4. Glasses Detection

Facial images can be categorized into two classes on the basis of wearing glasses or not. Then two FASS can be learnt respectively from examples with or without glasses. In the Bern face database (Refer to 4.1.3), there are 15 subjects wearing glasses among all the 30 subjects. 50 examples from 5 subjects (10 for each) wearing glasses are chosen as training samples to learn glasses-FASS. And similarly, 50 examples from 5 subjects (10 for each) without glasses are used to learn non-glasses-FASS. All the remaining images are used as testing examples. Table.4 shows the experimental results, where an average correct rate of 77% is achieved. Note that in the Bern database, many glasses of some images are quite thin-edge glasses, which increases the difficulty of the detection.

Table 4. Glasses Detection Results

	glasses	Non-glasses	Total
Training Samples	50 (5x10)	50(5x10)	100
Testing Samples	100(10x10)	50(10x10)	200
Correctly recognized	74	80	154

5. Conclusion

In the paper, a general concept named "Facial Attribute-Specific Subspaces (FASS)" is proposed and a unified framework to tackle face perception tasks is presented. Extensive experiments have shown the effectiveness and robustness of our method, especially its robustness for face recognition against variance due to expression, lighting and pose changes compared with benchmark algorithms.

Future efforts will be devoted to research on the adaptability and invariance of the proposed framework to varying illuminations. More deliberate virtual view synthesis algorithms should also be developed to derive multiple samples from single example view.

6. Acknowledgments

This research is sponsored partly by Natural Science Foundation of China (No.69789301), National Hi-Tech Program of China (No.863-306-ZD03-01-2), 100 Talents Foundation of Chinese Academy of Sciences, and SiChuan Chengdu Yinchen Net. Co.

References

[1] 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

[2] R. Brunelli and T. Poggio, “Face Recognition: Features versus Template”, *IEEE Transaction on PAMI*, Vol.15, No.10, pp1042-1052, 1993.10

[3] R.Chellappa, C.L.Wilson and S.Sirohey, “Human and Machine Recognition of faces: A survey”, *Proc. of the IEEE*, vol.83, No.5, 1995.5

[4] M.Pantic, Leon J.M. Rothkrantz, “Automatic Analysis of Facial Expressions: The State of the Art”, *IEEE TPAMI*, Vol.22, No.12, pp1424-1445, 2000.12

[5] M.Turk and A.Pentland. “Eigenfaces for Recognition” *Journal of cognitive neuroscience*, 3(1), pp71-86, 1991.1

[6] P.N.Belhumeur, J.P.Hespanha and D.J.Kriegman. “Eigenfaces vs Fisherfaces: recognition using class specific linear projection”. *IEEE TPAMI*, vol.20, No.7, 1997.7

[7] P.Penev and J.Atick, “Local Feature Analysis: A General Statistical Theory for Object Representation,” *Network: Computation in Neural Systems*, vol.7, pp.477-500, 1996

[8] L Wiskott, J.M.Fellous, N.Kruger and C.V.D.Malsburg, “Face Recogniton by Elastic Bunch Graph Matching”, *IEEE Trans. On PAMI*, 19(7), pp775-779, 1997.7

[9] T.F.Cootes, G.Edwards and C.J.Taylor “Active Appearance Models”. *Proc. ECCV, vol.2*, pp484-498, 1998

[10] P.J.Phillips, H.Moon, etc. “The FERET Evaluation Methodology for Face-Recognition Algorithms”, *IEEE TPAMI*, Vol.22, No.10, pp1090-1104, 2000.10

[11] G.Guo, S.Z.Li and K.Chan, “Face Recognition by Support Vector Machines”, *Proc. of the 4th Int. Conf. on Auto. Face and Gesture Recog.*, pp.196-201, Grenoble, 2000.3

[12] A.S.Georghiades, P.N.Belhumeur and D.J.Kriegman, “From Few to Many: Illumination Cone Models for Face Recognition under Differing Pose And Lighting”, *IEEE TPAMI, Vol.23, No.6, pp643-660, June 2001*

[13] T.Vetter and T.Poggio, “Linear Object Classes And Image Synthesis From A Single Example Image”, *IEEE Trans. On PAMI*, Vol.19, pp733-742, 1997

[14] A.Shashua and T.Riklin-Raviv, “The Quotient Image: Class-Based Re-Rendering And Recognition With Varying Illuminations”, *IEEE Trans. on PAMI*, pp.129-139, 2001.2

[15] Xilin Chen, Sam Kwong and Yan Lu, Human Facial Expression Recognition Based on Learning Subspace Method, Proceedings of ICME, 2000, pp403-406