Kernel Similarity based AAMs for face recognition

Y. Zhang[1], Y. Benhamza[2], K. Idrissi[1], and C. Garcia[1]

[1]Universit de Lyon, CNRS

INSA-Lyon, LIRIS, UMR CNRS 5205, F-69621, France

[2] Laboratoire LARATIC, INTTIC Oran Algrie

{Yuyao.Zhang, khalid.Idrissi, Christophe.garcia}@insa-lyon.fr

ybenhamza@ito.dz

Abstract. Illumination and facial pose conditions have an explicit effect on the performance of face recognition systems, caused by the complicated non-linear variation between feature points and views. In this paper, we present a Kernel similarity based Active Appearance Models (KSAAMs) in which we use a Kernel Method to replace Principal Component Analysis (PCA) which is used for feature extraction in Active Appearance Models. The major advantage of the proposed approach lies in a more efficient search of non-linear varied parameter under complex face illumination and pose variation conditions. As a consequence, images illuminated from different directions, and images with variable poses can easily be synthesized by changing the parameters found by KSAAMs. From the experimental results, the proposed method provides higher accuracy than classical Active Appearance Model for face alignment in a point-to-point error sense.

Keywords: AAMs, Kernel methods, dimension reduction

1 Introduction

Face recognition has been a well investigated topic the in image processing and computer vision communities. In the last decade, large efforts have been done in searching for a face recognition system that is capable of working with "real-world" faces. Among these efforts, Active Appearance Model, first proposed in [1], is a non-linear, generative, and parametric model of a certain visual phenomenon [2]. AAMs is quite well-known and widely used in the face recognition field. The aim of the algorithm is to "explain" novel images by generating synthetic images that are as similar as possible, using a parameterized model of appearance. There are several major unsolved problems existing in AAMs like position variations of the faces and directions of the sources of illumination.

Within the last years, the problems of illumination independence and complex pose fitting have been addressed by different approaches. A common method to deal with variations of illumination is to supplement additional parameters during the construction of the model. As demonstrated explicitly in [3] and [4], this approach directly increase computational complexity and is more time consuming. According to W. Zhao and R. Chellappa [5], a method to normalize the variations in appearance, either by image transformation or by synthesizing a new image from the given image in the training set ([6,7]), is also popular for face recognition. However, such algorithms can hardly generate the faces with variations of illumination during the fitting procedure. Meanwhile, the pose problems are equally complex because of the non-linear variation caused by the rotation of the head and different positions of the camera. In particular, a single AAMs is able to cope with shape variations from a narrow range of face poses (turning and nodding of 20). It utilizes PCA to model shape and appearance variation across pose, expression, illumination and identity. But the linear assumption does not hold true when large rotations of the face exist. Several works [8,9,10] applied the non-linear statistic tool, Kernel-PCA instead of PCA to build shape models of wide range of face rotations. But these methods all suffered from the rough approximation reconstruction problem of kernel PCA.

In this paper, a non-linear statistic approach is supposed to build the shape, texture and appearance model of AAMs. It benefits from kernel similarity matrix, but avoids the reconstruction problem of kernel PCA. Our method efficiently functions with the face reconstruction in complex illumination environments, and works much more efficiently with the wide range face rotation case comparing with AAMs.

The organization of this paper is as follows. Section 2 presents the classical Active Appearance Models algorithm. Section 3 introduces how the proposed KSAAMs algorithm works. Section 4 presents results of the performance of the proposed algorithm and a statistical comparison with AAMs. Section 5 concludes the paper.

2 Active Appearance Model

Active Appearance Model is an algorithm which allows generating a synthetic image as close as possible to a particular target image by making use of constraints of the appearance models. An appearance model is combined by two linear subspaces, one for the object shape and one for the object texture which are both learnt from a labeled set of training images [11].

Interpreting a novel image is an optimization problem in which the method minimizes the difference between a new image and one synthesized by the appearance model. The difference vector δI can be defined as:

$$\delta I = I_{\mathrm{i}} - I_{\mathrm{m}} \tag{1}$$

where I_{i} is the vector of grey-level values in the image, I_{m} is the vector of grey-level values for the current model parameters.

This method proceeds in three steps:

I) A Principal Component Analysis (PCA) is applied respectively on the shape training base and a shape-free texture training base. PCA created the statistical shape and texture model as the follows.

$$s = \bar{s} + Q_{\mathrm{s}} b_{\mathrm{s}} \tag{2}$$

$$t = \bar{t} + Q_{\mathrm{t}} b_{\mathrm{t}} \tag{3}$$

where $\bar{s}$ is the mean shape; $\bar{t}$ the mean texture in a mean shape patch; Q_{s} and Q_{t} are the matrices of eigenvectors of the shape and texture covariance matrices; b_{s} and b_{t} are vectors of coefficients in the Q_{s} and Q_{s} spaces which control the synthesis of shape and texture.

Another PCA is then applied on the samples of vector b, which is combined by b_{s} so as to construct the appearance parameter c:

$$b = Qc \tag{4}$$

with Q the matrix of PCA eigenvectors c is a vector controlling both b_{s} and b_{t} at the same time.

II) An experiment matrix creating procedure in which each control parameter c is disturbed from a known value and the residuals of each displacement in each image is measured to build a relationship between the parameter and the image variations. This relationship can be presented by:

$$\delta c = R * \delta I \tag{5}$$

Here R is the experiment matrix build in this step, δc and δI represent the parameter and the image variations respectively.

III) The fitting procedure in which by varying the model parameter c, the magnitude of the difference vector $\Delta = (\delta I)^2$ is minimized in order to find the best match between model and image.

3 Kernel Simlarity AAMs

A standard Active Appearance Model explains novel images by linear combination of statistic models which are build by applying Principle Component Analysis on training data. Therefore, PCA is not designed to extract non-linear features from the shape and texture of the non-frontal or non-uniformly illuminated faces. In general, both illumination and pose variations remain difficult to handle in face recognition.

In this paper, a non-linear component analysis method is considered to be more appropriate for handling the multiple variations which are caused by the changes of the light source. In this respect, a kernel method component analysis is employed instead of PCA to search more efficient components for generating new images in complex illumination and pose conditions. The following subsection aims at presenting the the proposed Kernel Similarity Active Appearance Model method.

3.1 PCA Trick in Feature Space

Consider a $N \times M$ observation matrix A, where each column is an observation and each row is the dimension of the observation. For example, in the context of this paper, each column is an image and each row are the image pixels. One observation is denoted as x_k, $k = 1, 2, \ldots\ldots, M$, $x_k \in R^N$, and $\sum_{k=1}^{M} x_k = 0$, which means that the data is centered. Normally, PCA diagonalizes the covariance matrix as shown in Eqn.(6).

$$C = \frac{1}{M} \sum_{j=1}^{M} x_j x_j^T \tag{6}$$

In some special case, for example the case in our paper we have much more dimensions than faces, that $N >> M$, so finding the eigenvectors of the large $N \times N$ matrix is computationally difficult. We apply a PCA trick: instead of (6), Eqn (7) is more computationally tractable.

$$\tilde{C} = \frac{1}{N} \sum_{i=1}^{N} y_i y_i^T \tag{7}$$

where y_i is the vector of each element of the observation x_j, N is the dimension of observation x_j.

To diagonalize it, one has to solve the following eigenvalue equation:

$$\tilde{\lambda} u = \tilde{C} u. \tag{8}$$

where $\tilde{\lambda} = \lambda$ represent to the eigenvalues of the matrix $\tilde{C}$; the eigenvectors $u = A^T v = \sum_{i=1}^{N} y_i v_i$.

The previous part of this section is devoted to a straightforward translation to a non-linear scenario. We shall now describe this computation in a Hilbert space H, which is introduced via a mapping Φ.

$$\Phi : R^N \to H, x \to X. \tag{9}$$

In the feature space H, we assume that $\Phi(x)$ has an arbitrarily large, possibly infinite dimensionality. Again, in feature space, the data should be centered, $\sum_{k=1}^{M} \Phi(x_k) = 0$. Applying the PCA trick in feature space H,

$$\bar{C} = \frac{1}{N} \sum_{i=1}^{N} \Phi(y_i) \Phi(y_i)^T \tag{10}$$

Now one has to extract eigenvalues satisfying

$$\bar{\lambda} U = \bar{C} U \tag{11}$$

The solutions U lies in the span of $\Phi(y_1), \Phi(y_2), \cdots, \Phi(y_N)$. As shown in [13], this has two useful consequences: first, we can consider the equivalent equation

$$\bar{\lambda}(\Phi(y_k)^T U) = (\Phi(y_k)^T \bar{C} U) \tag{12}$$

for all $k = 1, 2, \cdots, N$ and second, there exist coefficients α_i $(i = 1, \cdots, N)$ such that

$$U = \sum_{i=1}^{N} \alpha_i \Phi(y_i) \tag{13}$$

Combining (12) and (13), we get

$$\bar{\lambda} \sum_{i=1}^{N} \alpha_i (\Phi^T(y_k) \cdot \Phi(y_i)) = \frac{1}{M} \sum_{i=1}^{N} \alpha_i (\Phi^T(y_k) \cdot \sum_{j=1}^{N} \Phi(y_j))(\Phi^T(y_j) \cdot \Phi(y_i)) \tag{14}$$

Defining a $N \times N$ matrix K by

$$K_{i,j} = (\Phi^T(y_i) \cdot \Phi(y_j)) \tag{15}$$

which lead (13) to:

$$M\bar{\lambda}K\alpha = K^2\alpha \tag{16}$$

where α denotes the column vector with entries $\alpha_1, \cdots, \alpha_N$. As K is symmetric,

$$M\bar{\lambda}\alpha = K\alpha \tag{17}$$

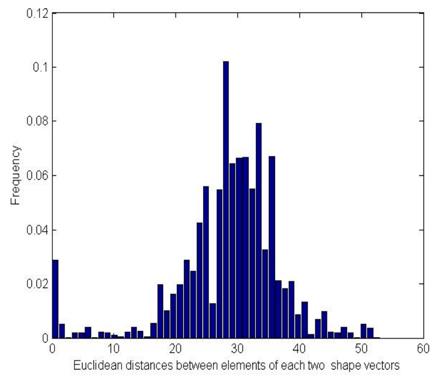
Note that K is positive semi-definite, which can be seen by noticing that it equals

$$K_{i,j} = k(y_i, y_j) = e^{-\frac{\|y_i - y_j\|^2}{2\delta^2}} \tag{18}$$

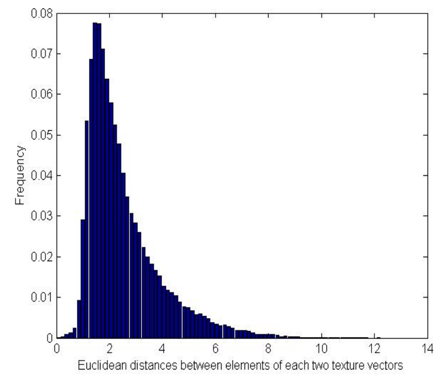
Then, for the extraction of eigenvalues in feature space, we therefore only need to diagonalize the kernel similarity matrix $K_{i,j}$. Let $\bar{\lambda}_1 \geq \bar{\lambda}_2 \geq \cdots \geq \bar{\lambda}_N$ denote the eigenvalues, and $\alpha^1, \alpha^2, \cdots, \alpha^N$ the corresponding complete set of eigenvectors.

3.2 Parameter estimation

As described previously, kernel method component analysis deals with nonlinear transformation via nonlinear kernel functions. In kernel the functions, there is a parameter σ that must be predetermined, knowing that it has a significant impact on image representation in feature space. As the kernel function is defined with the Gaussian function $k(a, b) = e^{-\frac{d(a,b)^2}{2\delta^2}}$, in which $d(a, b)^2$ represent Euclidean distances between elements contained in each vector; $k(a, b)$ can be considered as a zero mean Gaussian distribution of $d(a, b)^2$. So if $d(a, b)^2$ follows Gaussian distribution, then σ represented the variance of $d(a, b)^2$. With respect to this assumption, we built histograms of the Euclidean distances between elements contained in each vector to study the distribution of observed variables. As illustrated in Figure 1, for shape vectors and texture vectors from the illumination database (described in the Experiment result section), the Euclidian distances between each observed variable follows a Gaussian distribution. The parameters σ is estimated.



(a)Histogram of Euclidean distances between elements of shape vectors



(b)Histogram of Euclidean distances between elements of texture vectors

Fig. 1. Histogram of Euclidean distances between elements of vector of illumination database

3.3 Feature extraction

In classical AAMs, we seek a parameterized model (the parameter c) used to control variations if both shape and texture, which is extracted by Principal Component Analysis. In our work, the kernel similarity matrix $K_{i,j} = k(y_i, y_j) = e^{-\frac{\|y_i - y_j\|^2}{2\delta^2}}$ replaces the covariance matrix (used in PCA).

As explained in section 2., features are extracted to control variation on shape, texture and appearance. The parameter c, which represents the parameter of appearance is build to control both shape and texture variation simultaneously. Figure 2 illustrates the first three modes of variations of c for classical AAMs and the proposed method. One can observe that the model built by kernel similarity matrix is able to take into account more efficiently the variations of illumination.

Fig. 2. Variations in the appearance parameter c; the first row presents the first three modes learnt by KSAAMs; the second row presents the first three modes learnt by PCA)

4 Experimental Results

We evaluated the proposed method on the CMU Pose, Illumination, and Expression (PIE) database of human faces [14]. For the experiments on the variation of illumination and pose, the training database is built from a subset of the CMU database as illustrated in Figure 3. The test set is built from the images of the persons shown in the last row of Figure 3. We manually labeled 1200 images of size 640×486 pixels. To train the models, 58 landmarks were placed on each face image: 8 points for the mouth, 11 points for the nose, 16 points for both eyes, 10 points for both eyebrows, and 13 points for the chin. The warped images have approximately 7325 pixels inside the facial mask.

To evaluate the performance of the proposed algorithm, the manually annotated landmarks are considered as the ground truth shape information. For each image the landmarks re-labeled by the methods are compared with the ground truth landmarks. A distance measure, $D(x_{gt}, x)$, gives a interpretation of the fit between two shapes, the ground truth, x_{gt} and the actual shape x. Point-to-point error E_{pt-pt} is defined as the Euclidean distance between each corresponding landmark in Eqn.(19). To interpret a novel image, an optimization is performed in which the method minimizes the error between the pixels contained in a new image and the pixels synthesized by the appearance model. The pixel-to-pixel error $E_{pix-pix}$ can be defined as in Eqn. (20).

$$E_{pt-pt} = \frac{1}{n} \sum \sqrt{(x_i - x_{gt,i})^2 + (y_i - y_{gt,i})^2} \tag{19}$$

$$E_{pix-pix} = |\delta I|^2 = |I_i - I_m|^2 \tag{20}$$

where (x_i, y_i) are the coordinates of the re-labeled landmarks, I_i is the vector of grey-level values in the image and I_m is the vector of grey-level values for the current model parameters.

To evaluate the superiority of the proposed method, Eqn.(21) is used to compute the gain in precision.

$$gain\% = \frac{E_{pt-pt}(kernel) - E_{pt-pt}(AAMs)}{E_{pt-pt}(AAMs)}\% \tag{21}$$

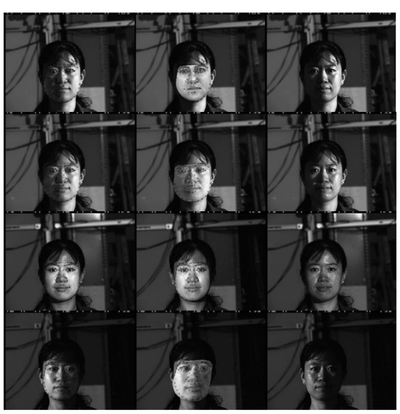
Fig. 3. Persons in the training and test databases; the first two rows present the persons in the training database; the last row present the persons in the test database (not present in the training database).

4.1 Sensitivity to the Illumination

The database for training is built by all the frontal faces which are captured by camera number 27. Each person involved in the training set (shonw in Figure 3) has 20 frontal face images under 20 different illumination conditions. The training database contains 16 people and the test database contains 10 people as shown in Figure 3.

In figure 4(a), images in the left column are synthesized by the proposed method, and compared with fitting results of Standard AAMs in the central column. An increased precision has been obtained due to the extraction of non-linear features. The gain in precision on point-to-point errors is reported in table 1, computed by eqn. (19)-(21).

Figure 5 show the errors obtained in the "Standard AAMs experiment" (square dotted curve) and "Kernel Similarity AAMs experiment" (asterisk curve) on both training (subfigure (a)) and test (subfigure (b)) databases. Errors are normalized by the Euclidian distance between the eyes ($E_{pt-pt}/Deye$, where E_{pt-pt} represented point-to-point error, and $Deye$ represented the distance between the centre of the eyes of each person). This normalization is done to eliminate the effect of varying size of faces on the point-to-point error. Each curve point in Figure 5 is the mean error made by the model in the database under the same illumination conditions. The number of illumination from 1 to 20 are the 20 different illuminations contained in database. Illuminations from numbers 1 to 4 correspond to a light source from the left side of the face, illuminations from 12 to 17 correspond to a light source from the right side of the face. The other illuminations (from number 5 to 11 and number 18 to 20) correspond to different light sources in front of the face. The error curves depict the robustness of the KSAAMs method since it makes it possible to find non-linear facial features.



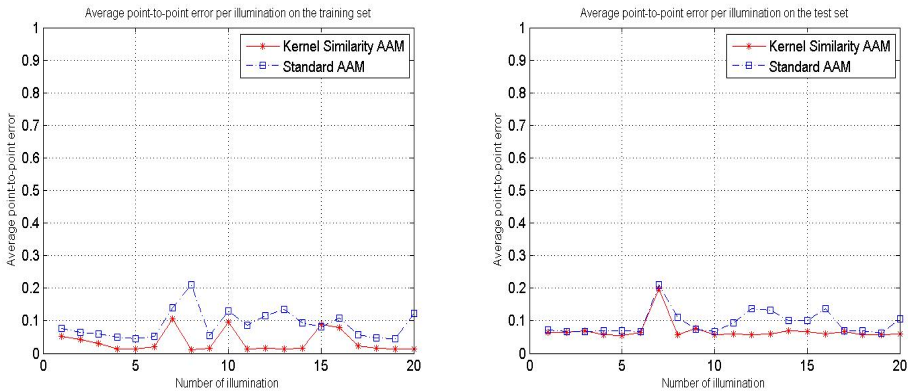
(a) Fitting results on different for the different illuminations

(b)Fitting results for different poses

Fig. 4. Fitting result on the PIE facial images with the proposed method in left column, classical AAMs in middle column, and Input images in the right column.

Table 1. Gain in percentage for the KSAAMs method for the illumination problem (Eqn.(21))

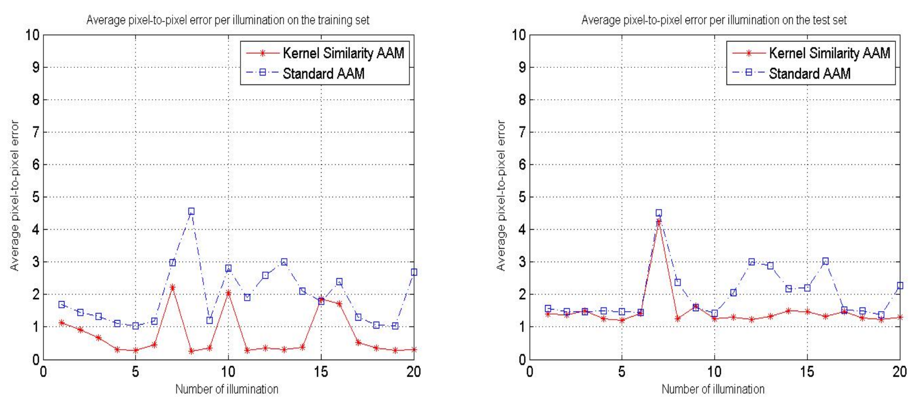
	$E_{pt-pt} gain\%$	$E_{pix-pix} gain\%$
Training database	87.85%	61.23%
Test database	76.76%	27.02%



(a) Average Point-to-point Error on training dataset

(b)Average Point-to-point Error on the test dataset

Fig. 5. Average point-to-point error versus illumination variations



(a) Average Pixel-to-pixel Error on the training dataset

(b)Average Pixel-to-pixel Error on the test dataset

Fig. 6. Average pixel-to-pixel error versus illumination conditions

Table 2. Gain in percentage for the KSAAMs method for the variation of pose problem (Eqn.(21))

	$E_{pt.-pt.}Avd\%$	$E_{pix.-pix.}Avd\%$
Training database	16.63%	14.05%
Test database	22.44%	18.67%

We can see that with the proposed method, some errors are still made, but not as strong as with the standard method with an increased robustness to side illuminations. Figure 6 present the average pixel-to-pixel error for each illumination condition.

4.2 Sensitivity to the poses

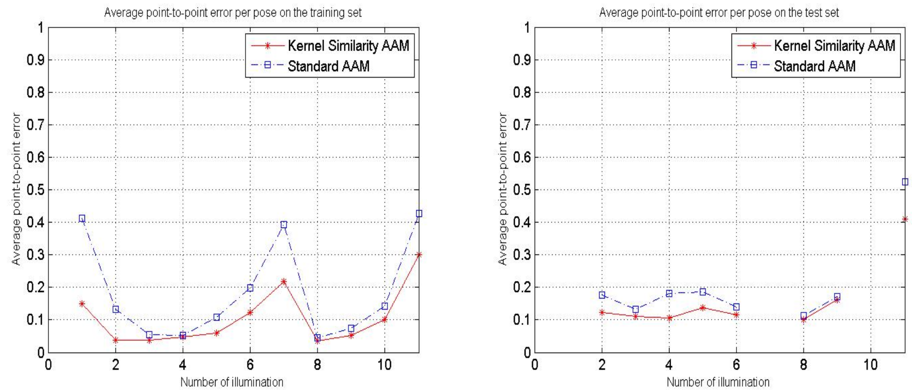
The training set is built with the same 16 persons, each person having 11 different pose captured by different cameras, the test set containing images from 10 persons.

As presented in Figure 4(b), images in the left column are synthesized by the proposed method, compared with fitting results of classical AAMs in the central column. An increased precision has been obtained due to the extraction of non-linear features. The gain in precision on point-to-point errors is reported in table 2, computed by eqn. (19)-(21).

Poses numbers 1, 6, 7, 10 are profile faces which are hard to synthesize, while the other poses are less complicated. The curves in Figure 7 and Figure 8 which present the point-to-point errors and the pixel-to-pixel errors for each pose respectively give a consistent result. As illustrated by the curves, for the test on the training database, the proposed method is more efficient except for poses number 1, 7 and 10. On the test set, the results of poses number 1, 7 and 10 are missing, because the fitting procedures have problem to converge for both KSAAMs and AAMs. As a consequence, the proposed kernel method gives better fitting results in the conditions that the out-of-plane rotations of face are in a the range of $\pm 60°$. The problem of the complete profile faces is still waiting to be solved.

5 Conclusion

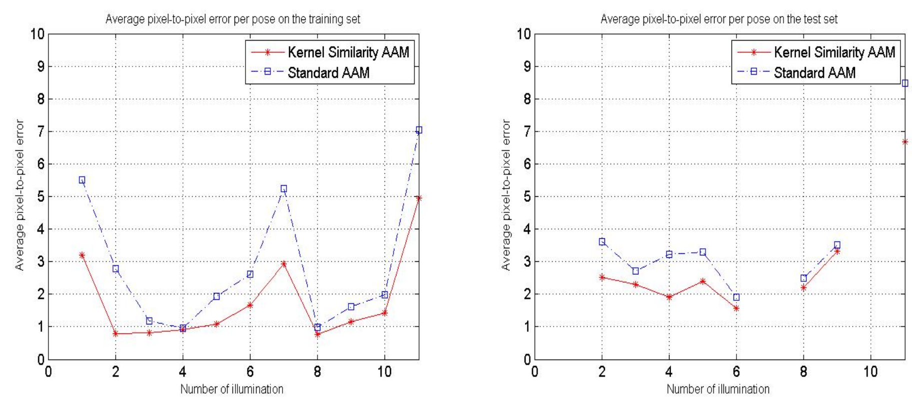
In this study, we have proposed a Kernel method combined with the AAMs fitting algorithm that is robust to illumination and pose changes of face images. Instead of the covariance matrix used in Principal Component Analysis of classical Active Appearance Model, we use eigenvectors of the kernel similarity matrix to build the deformable model.



(a) Average Point-to-point Error on faces known

(b)Average Point-to-point Error on faces unknown

Fig. 7. Average point-to-point error per pose



(a) Average Pixel-to-pixel Error on faces known

(b)Average Pixel-to-pixel Error on faces unknown

Fig. 8. Average Pixel-to-pixel error per pose

It is shown that the model build by the proposed kernel method is less sensitive to the illumination variations. With this novel method, the fitting procedure can accurately synthesize faces semi-bright-semi-dark affected by the illumination. Meanwhile, conditions with a variety of poses also benefit from the proposed algorithm; the ability of synthesizing faces with shape variations from a wide range of face poses has been improved.

References

[1] T.F. Cootes, G.J Edwards and C.J. Taylor: Active Appearance Models. Proc. European Conference on Computer Vision. (1998)

[2] Iain Matthews and Simon Baker: Active Appearance Models Revisited. INTERNATIONAL JOURNAL OF COMPUTER VISION. **60** (2003) 135–164

[3] F. Kahraman et al. An Active Illumination and Appearance (AIA) Model for Face Alignment. IEEE Conference on Computer Vision and Pattern Recognition. CVPR. **07 60** (2007) 1–7

[4] A. S Georghiades, P. N Belhumeur, and D. J Kriegman: From few to many: Illumination cone models for face recognition under variable lighting and pose. IEEE Transactions on Pattern Analysis and Machine Intelligence. **23** no. 6 (June 2001) 643–660

[5] WenYi Zhao and Rama Chellappa: SFS Based View Synthesis for Robust Face Recognition. Proceedings of the Fourth IEEE International Conference on Automatic Face and Gesture Recognition (2000).FG.**00** (Washington, DC, USA: IEEE Computer Society, 2000), 285–291

[6] Shiguang Shan et al: Illumination normalization for robust face recognition against varying lighting conditions. IEEE International Workshop on Analysis and Modeling of Faces and Gesture. **60** (2003) AMFG 2003, 157–164

[7] Jiali Zhao et al: Illumination ratio image: synthesizing and recognition with varying illuminations. Pattern Recognition Letters 24. no. 15 (November 2003), pages 2703–2710

[8] Sami Romdhani et al: A Multi-View Nonlinear Active Shape Model Using Kernel PCA. T.Pridmore and D. Elliman, editors, Proceeding of the 10th British Machine Vision Conference. (BMVC**99**) (1999) 483–492

[9] C. J Twining and C. J Taylor: Kernel Principal Component Analysis and the Construction Of Non-Linear. (BMVC**01**) (2001), pages 23–32.

[10] Fernando De la Torre and Minh Hoai Nguyen: Parameterized Kernel Principal Component Analysis: Theory and applications to supervised and unsupervised image alignment. IEEE Conference on Computer Vision and Pattern Recognition, June, 2008

[11] T F Cootes and C J Taylor: Statistical models of appearance for computer vision. World Wide Web Publication. February 2004

[12] T. Sim, S. HAL: Combining visual dictionary, kernel-based similarity and learning strategy for image category retrieval. Computer Vision and Image Understanding. Volume 110, Issue 3, June 2008, Pages 403-417

[13] B. Scholkopf and A. Smola: Nonlinear component analysis as a kernel eigenvalue problem. Neural computation. 1998 Vol. 10, No. 5, Pages 1299-1319

[14] Baker, and M. Bsat: The CMU Pose, Illumination, and Expression (PIE) database. Fifth IEEE International Conference on Automatic Face and Gesture Recognition (2002)Proceedings. IEEE. 2002 46–51