2D-3D Face Matching using CCA

Weilong Yang, Dong Yi, Zhen Lei, Jitao Sang, Stan Z. Li
Center for Biometrics Security Research & National Laboratory of Pattern Recognition
Institute of Automation, Chinese Academy of Sciences, Beijing
http://www.cbsr.ia.ac.cn/

Abstract

In recent years, 3D face recognition has obtained much attention. Using 2D face image as probe and 3D face data as gallery is an alternative method to deal with computation complexity, expensive equipment and fussy pretreatment in 3D face recognition systems. In this paper we propose a learning based 2D-3D face matching method using the CCA to learn the mapping between 2D face image and 3D face data. This method makes it possible to match the on-site 2D face image with enrolled 3D face data. Our 2D-3D face matching method decreased the computation complexity drastically compared to the conventional 3D-3D face matching while keeping relative high recognition rate. Furthermore, to simplify the mapping between 2D face image and 3D face data, a patch based strategy is proposed to boost the accuracy of matching. And the kernel method is also evaluated to reveal the non-linear relationship. The experiment results show that CCA based method has good performance and patch based method has significant improvement compared to the holistic method.

1. Introduction

In the last decades, numerous face recognition methods have been proposed, such as Eigenface [13] and Fisherface [2]. Most of the proposed methods are based on the 2D appearance and thus sensitive to the varying illumination and pose. Because the shape of faces is independent of the illumination and pose, 3D face recognition has the potential to improve performance under the uncontrolled conditions. Some 3D face recognition algorithms have been proposed and very high recognition rates are reported [1].

However, due to the computational complexity, expensive equipment and fussy pretreatment, 3D technology is still not used widely in practical applications. Generally, 3D face recognition systems require that probe and gallery set are both 3D face data. However, in some application, there are only 2D images available for recognition (assuming the enrollment is done), such as the low resolution mugshot on ID card or the snapshot captured by video surveillance camera. The conventional 3D face recognition system cannot work under these circumstances. The second disadvantage of the 3D face recognition system lies in its 3D data acquisition equipment. To acquire the accurate 3D face data, some very costly equipment must be used, such as 3D laser scan or stereo camera system. They are not as stable and efficient as 2D cameras, and for some cases like the stereo camera system, calibration is needed before use. Moreover, both of them will take a longer time to acquire (or reconstruct) the 3D face data compared with the 2D camera only taking the 2D images. Besides, in some applications there is not so much time to capture user's 3D face data on-site, such as airport access control or E-passport. Respecting these facts, 3D face recognition is still not as applicable as 2D face recognition.

In order to overcome the restricts of 3D face recognition system while reserve its advantages, we can use 2D face image as probe and 3D face data as gallery. In this paper, we introduce a learning based method to match 2D probe against 3D gallery. From an intuitive viewpoint, we could find a very interesting relationship between 2D face image and 3D face data. Different 2D face images from one person (*i.e.* the faces under different illuminations) will always correspond to the same 3D face shape. However, different face shapes could never match with the same face image, which indicates the mapping between the 2D face image and 3D face shape is actually a many-to-one problem [5]. Based on this mapping, Lei *et al*. [7] propose to recovery the 3D face shape from one single 2D face image. Therefore, once we find the mapping between 2D face images and 3D face shapes, matching between 2D probe and 3D gallery will be achieved.

In this paper, we apply the canonical correlation analysis (CCA) [4] to learn the mapping between the 2D face image and 3D face data. The proposed method consists of two phases. In the learning phase, given the 2D-3D face data pairs of the subjects for training, PCA is applied on both 2D face image and 3D face data firstly to avoid the curse of dimensionality and reduce noise. Then the CCA regression is

978-1-4244-2154-1/08/$25.00

performed between features of 2D-3D in the previous PCA subspaces. In the recognition phase, given an input 2D face image as probe, the correlation between the probe and the gallery is computed as matching score using the learnt regression.

As reported in [12, 14], for face recognition task, different parts of face do not have the same contribution to the final recognition results. Patch based face recognition methods has shown the promising results. In this paper, we also propose a patch based method to deal with the 2D-3D face matching problem, in which the whole face is spatially partitioned into many patches. The CCA regression is performed on each patch and the matching score of each patch is computed. The experiment results show that the patch based method has significantly improved the 2D-3D face matching results. And due to the nonlinear relationship between 2D face images and 3D face data, CCA may be not the optimal descriptor of the mapping, therefore kernel CCA (KCCA) is also used to learn the regression.

The rest of the paper is organized as follows: In Section 2, related work is briefly reviewed. CCA and KCCA methods are introduced in Section 3. The framework of CCA based 2D-3D face matching is presented in Section 4, whereas in Section 5 the patch based 2D-3D face matching method is proposed. The experiments and results analysis are given in Section 6 and we summarize this paper in Section 7.

2. Related Work

To our best knowledge, there exists very few method to tackle the 2D-3D face matching problem. Rama *et al.* [9] use Partial Principal Component Analysis (P^2CA) to match the 2D face image (probe) with 3D face data (gallery). They firstly project the 3D texture information in cylindrical coordinates. P^2CA is applied on both probe and gallery to reduce the dimension and form the P^2CA subspace. Because the 2D face image probe in P^2CA only contains partial information of the face, the vector representation of probe in P^2CA subspace will have a high correlation with a section of vector representation of one sample in the gallery.

In [10], D. Riccio *et al.* propose a particular 2D-3D face recognition method based on 16 geometric invariants, which are calculated from a number of “control points”. The 2D face images and 3D face data are related through those geometric invariants. This method is invariant to pose and illumination, but the performance of the method closely depend on the accuracy of “control points” localization. To our knowledge, there is no one method or model can locate those “control points” very well for face images of various pose and illumination.

3. Overview of CCA and KCCA

CCA is a powerful multivariate analysis method [3]. It has various applications in pose estimation [8] and face matching [15]. For two sets of variables, CCA is to construct the CCA subspace to mutually maximize the correlation between these two sets variables. As a non-linear variant of CCA, kernel CCA (KCCA) [8] uses *kernel trick* to transform the input variables into the space with high or even infinite dimensionality. In this high-dimensional space, the relationship of the two sets of variables could be considered as linear and thus able to be solved by linear CCA. In this section, we will give a brief overview of CCA, KCCA and discuss the regularization and its influence on CCA and KCCA.

3.1. CCA

Given N pairs of samples (x_i, y_i) of (X, Y), $i = 1, \ldots, N$, where $X \in \mathbb{R}^m, Y \in \mathbb{R}^n$. The mean of both X and Y is zero. The goal of CCA is to learn a pair of direction ω_x and ω_y to maximize the correlation between the two projections $\omega_x^T X$ and $\omega_y^T Y$, where T denotes the transpose, i.e. to maximize:

$$\rho = \frac{E[\omega_x^T X Y^T \omega_y]}{\sqrt{E[\omega_x^T X X^T \omega_x] E[\omega_y^T Y Y^T \omega_y]}} \tag{1}$$

where $E[f(x,y)]$ denotes the empirical expectation of the function $f(x,y)$.

The covariance matrix of (X,Y) is

$$\mathbf{C(X,Y)} = \mathbf{E}\left[\begin{pmatrix}\mathbf{X}\\ \mathbf{Y}\end{pmatrix}\begin{pmatrix}\mathbf{X}\\ \mathbf{Y}\end{pmatrix}^{\mathbf{T}}\right] = \mathbf{E}\left[\begin{pmatrix}\mathbf{C_{xx}}\\ \mathbf{C_{xy}}\end{pmatrix}\begin{pmatrix}\mathbf{C_{yx}}\\ \mathbf{C_{yy}}\end{pmatrix}^{\mathbf{T}}\right] \tag{2}$$

where $\mathbf{C_{xx}}$ and $\mathbf{C_{yy}}$ are within-sets covariance matrices; $\mathbf{C_{xy}}$ and $\mathbf{C_{yx}}$ are between-sets covariance matrices.

Hence, ρ could be rewritten as

$$\rho = \frac{\omega_x^T C_{xy} \omega_y}{\sqrt{\omega_x^T C_{xx} \omega_x \omega_y^T C_{yy} \omega_y}} \tag{3}$$

Let

$$\mathbf{A} = \begin{pmatrix} 0 & \mathbf{C_{xy}} \\ \mathbf{C_{yx}} & 0 \end{pmatrix}, \quad \mathbf{B} = \begin{pmatrix} \mathbf{C_{xx}} & 0 \\ 0 & \mathbf{C_{yy}} \end{pmatrix} \tag{4}$$

It can be shown that the solution $\mathbf{W} = (\omega_x^T, \omega_y^T)^T$ amounts to the extremum points of the *Rayleigh quotient*:

$$r = \frac{\mathbf{W^T A W}}{\mathbf{W^T B W}} \tag{5}$$

The solution ω_x and ω_y can be obtained as solutions of the generalized eigenproblem:

$$\mathbf{AW} = \mathbf{BW}\lambda \tag{6}$$

As a subspace learning method, CCA is inclined to overfit to the training data, especially when the sample size is small [8]. Here, we add regularization term $\lambda_x I$ and $\lambda_y I$ to C_{xx} and C_{yy} respectively to alleviate overfitting, where λ_x and λ_y are regularization parameters and $\lambda_x > 0, \lambda_y > 0$. Let $\lambda_x = \lambda_y = \lambda$, we can get the object function of regularized CCA (RCCA). RCCA is to maximize ρ, where

$$\rho = \frac{\omega_x^T C_{xy} \omega_y}{\sqrt{\omega_x^T (C_{xx} + \lambda I)\omega_x \omega_y^T (C_{yy} + \lambda I)\omega_y}} \tag{7}$$

3.2. KCCA

Because of the nonlinear relationship between 2D face images and 3D face data, CCA may not extract useful descriptors of the data. Therefore, here we also introduce the Kernel CCA (KCCA).

Firstly, KCCA projects the data into a higher even infinite dimensional feature space F through kernel trick as

$$\begin{aligned} \phi : & X \in \mathbb{R}^m \rightarrow \phi(X) \in F \\ \phi : & Y \in \mathbb{R}^n \rightarrow \phi(Y) \in F \end{aligned} \tag{8}$$

From the theory of reproducing kernel, we know that $\omega_x \in span\phi(X)$ and $\omega_y \in span\phi(Y)$. So, we can rewrite ω_x andω_y as the project of the data onto the direction α and β.

$$\begin{aligned} \omega_x &= \phi(X)\alpha \\ \omega_y &= \phi(Y)\beta \end{aligned} \tag{9}$$

Thus, the goal of KCCA is actually to maximize

$$\begin{aligned} \rho &= \frac{\alpha^T \phi(X)^T \phi(X) \phi(Y)^T \phi(Y) \beta}{\sqrt{\alpha^T \phi(X)^T \phi(X) \phi(X)^T \phi(X) \alpha \beta^T \phi(Y)^T \phi(Y) \phi(Y)^T \phi(Y) \beta}} \\ &= \frac{\alpha^T K_x K_y \beta}{\sqrt{\alpha^T K_x K_x \alpha \beta^T K_y K_y \beta}} \end{aligned} \tag{10}$$

Where $K_x(i,j) = \phi(x_i)^T \phi(x_j) = k(x_i, x_j)$, $K_y(i,j) = \phi(y_i)^T \phi(y_j) = k(y_i, y_j)$, $k(.\ ,.)$ is the kernel function. Therefore, KCCA equals to find optimal linear projection pairs α and β from the corresponding data (K_x, K_y), which can be solved as CCA.

As similar to RCCA, Regularized KCCA (RKCCA) applies $\lambda \|\omega_x\|^2$ and $\lambda \|\omega_y\|^2$ as regularization terms, where λ is regularization parameter. Thus, Equ.10 could be rewritten as:

$$\begin{aligned} \rho &= \frac{\alpha^T K_x K_y \beta}{\sqrt{(\alpha^T K_x K_x + \lambda \|\omega_x\|^2)(\beta^T K_y K_y \beta + \lambda \|\omega_y\|^2)}} \\ &= \frac{\alpha^T K_x K_y \beta}{\sqrt{(\alpha^T K_x K_x \alpha + \lambda \alpha^T K_x \alpha)(\beta^T K_y K_y \beta + \lambda \beta^T K_y \beta)}} \end{aligned} \tag{11}$$

4. CCA based 2D-3D face matching

In this section, we will introduce the proposed CCA based 2D-3D face matching method, in which the CCA is used to learn the mapping between 2D face image and 3D face data.

4.1. Framework of CCA based 2D-3D face matching

The essential idea of the proposed framework is to learn the mapping between 2D face image and 3D face data.In the learning phase, the 2D-3D mapping is learnt from the training set which consists of K pairs of 2D-3D face image. In the recognition phase, the most correlative 3D face shape to the input 2D facial image is found by the trained 2D-3D mapping. The general framework is illustrated in Fig.1.

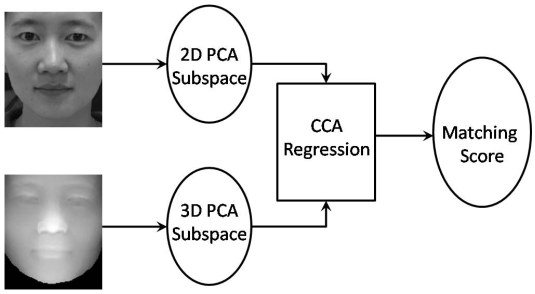

Figure 1. CCA based 2D-3D face matching

In the learning phase, N pairs of 2D-3D face are given as $(X, Y) = \{(x_k, y_k)\}, (k = 1, 2, \cdots N)$, where (x_k, y_k) is a corresponding pair of 2D and 3D. The training process is divided into the following two steps:

- **Dimension Reduction**.

  To reduce the computational complex of the following step, principle component analysis(PCA) is applied firstly to transform x_k and y_k into the lower dimensional space. PCA transform matrices P_x and P_y are learnt from 2D and 3D training sets respectively. The PCA projections are computed as

  $$X' = P_x^T (X - \overline{X}),\ Y' = P_y^T (Y - \overline{Y}) \tag{12}$$

  where the $\overline{X}$ and $\overline{Y}$ are the mean faces of 2D face and 3D face respectively.

- **CCA Regression**.

  For linear CCA, by performing CCA on X' and Y', two further projective directions ω_x and ω_y are learnt. $\omega_x^T X'$ and $\omega_y^T Y'$ are best correlated.

  For KCCA, by performing KCCA on X' and Y', projective direction α and β are learnt to maximize the correlation between $\omega_x^T \phi(X')$ and $\omega_y^T \phi(Y')$. Due to $\omega_x = \phi(X')\alpha, \omega_y = \phi(Y')\beta$, it is equivalent to maximize the correlation between $\alpha^T K_x'$ and $\beta^T K_y'$.

In the recognition phase, we also need to project new pair of images X_{ts} and Y_{ts} into PCA subspace firstly, that is

$$X'_{ts} = P_x^T(X_{ts} - \overline{X}),\ Y'_{ts} = P_y^T(Y_{ts} - \overline{X}) \tag{13}$$

For linear CCA, X'_{ts} and Y'_{ts} are projected into CCA subspace, that is

$$X_{output} = \omega_x^T X'_{ts}, Y_{output} = \omega_y^T Y'_{ts} \tag{14}$$

For KCCA, test data is needed to be projected into the CCA subspace in the kernel space.

$$\begin{aligned} X_{output} &= \alpha^T \phi(X')^T \phi(X'_{ts}) = \alpha^T K_x^{ts} \\ Y_{output} &= \beta^T \phi(Y')^T \phi(Y'_{ts}) = \beta^T K_y^{ts} \end{aligned} \tag{15}$$

Both K_x^{ts} and K_y^{ts} are kernel matrices, and

$$\begin{aligned} K_x^{ts}(i,j) &= \phi(X'_i)^T \phi((X'_{ts})_j) = k(X'_i, (X'_{ts})_j) \\ K_y^{ts}(i,j) &= \phi(Y'_i)^T \phi((Y'_{ts})_j) = k(Y'_i, (Y'_{ts})_j) \end{aligned} \tag{16}$$

We then compute the matching score between X_{output} and Y_{output} through following score function:

$$\mathit{score}(x,y) = \frac{x \cdot y}{(\|x\|\ \|y\|)} \tag{17}$$

5. Patch based extension

The human face consists of different parts, such as eyes, nose, and mouth. The holistic based methods consider the face as a whole part and aim to find a single mapping to describe the relationship between 2D-3D data. However, the relationships of 2D-3D face parts may be not always identical. A single mapping maybe too simple to describe the relationship between 2D and 3D face data. In order to get a more accurate model of 2D-3D face mapping, we extend the CCA regression method to a patch based CCA regression. The experiments show this method works much better than the normal CCA based method. The process is shown in Fig.2:

1. All 2D face images and 3D face data are divided into several patches.
2. For each patch, CCA based method is performed to learn the mapping respectively.
3. In the recognition phase, the testing images are also departed into several parts in the same way. Then we can get a score for each patch.
4. The final machining score is calculated by fusing the scores of all patches.

To avoid missing the whole information of face, the size of the patch must be chose suitably. If the patch is too big, the local mapping will be difficult to learn as the whole face, otherwise the information will be too local to reserve discriminant power.

For simplicity we fuse the scores of each patch at decision level. There are many fusion schemes which have been proposed such as sum rule, product rule, max rule, min rule, median rule and majority voting [6]. In this paper, we employ the weighted sum fusion. The final matching score is given by

$$F = \sum_{i=1}^{P} w_i * s_i \tag{18}$$

where s_i is the output score for the *ith* patch, w_i is the weight of the *ith* patch, P is the number of the pathes.

How to choose the weight for the sum rule is still an open question. In practical applications, 2D and 3D faces are usually not captured simultaneously. It will result in misalignment between 2D probe and 3D gallery. Patch based method might have a better alignment than the others. For example, the alignment is done according to the eye coordinates. The patches near to the two eyes will have better alignments than other patches. In [11], the curse of misalignment in face recognition has been presented. Therefore, it is reasonable to assign smaller weights to those parts far from the two eyes, in order to reduce the negative effect of misalignment. And the parts near to the eyes or nose are less affected by facial expression. Besides, we also need to take the texture information into account, *e.g.* the parts containing eyes, eyebrow, nose, or mouth should be assigned large weights.

Furthermore, we can apply kernel and patch method together. In the patch based KCCA face matching, the specific kernel space for each 2D-3D face patch is constructed, thus the nonlinear relationship of each 2D-3D face patch could be modeled accurately in their own kernel space respectively.

6. Experiments

In this section, we evaluate the performance of our algorithms for 2D-3D face matching. The description of the face database is presented, as well as the details of experiments. Then, the results of the experiments are concluded.

6.1. Face database

The face database consists of the 2D face images and 3D face data of 200 subjects. The 2D facial images are captured by an ordinary camera. The 3D faces are acquired with a Minonlta 910 laser scanner. The laser scanner provides the range image the faces which are actually 2.5D data. Regions not belonging to the face are discarded. To simulate

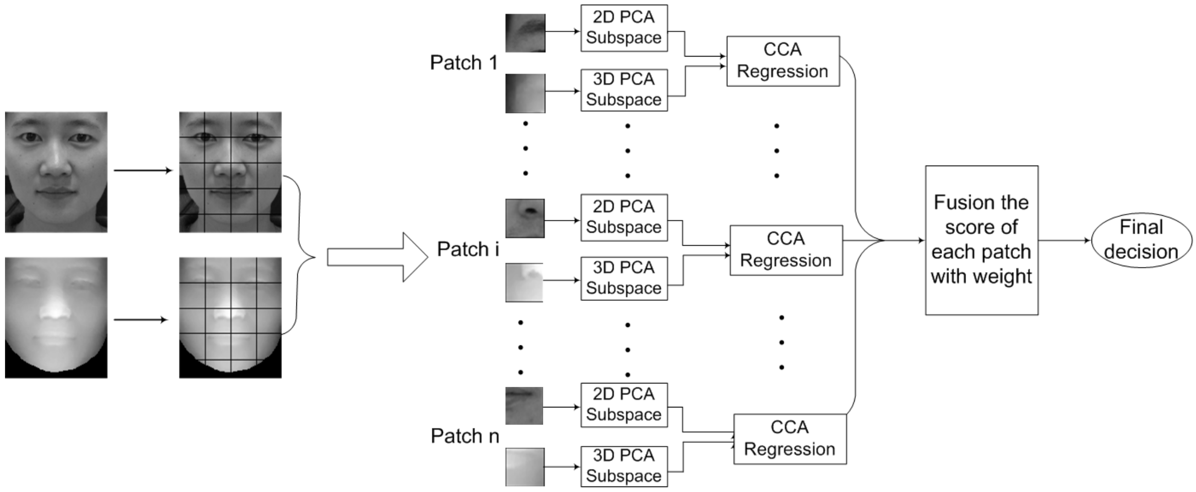

Figure 2. Patch based CCA in 2D-3D face matching

the real-word application, the 2D face image and 3D face data were not acquired simultaneously. As shown in Fig.3, the 2D face images and 3D range image have some slightly facial expression and head pose changing.

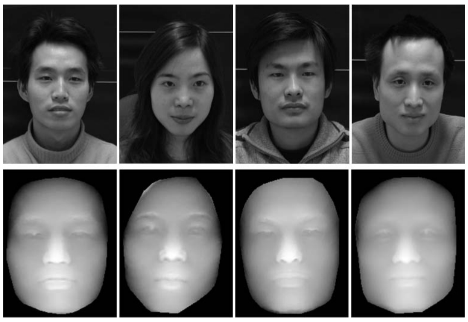

Figure 3. 2D-3D face samples

Firstly, we align the 2D image and 3D range image according the eye coordinates. And then all face images are cropped into 142×120 pixels. For the experiments, the database is divided into two sets, training set and test set. The training set contains the 2D-3D pairs of 172 subjects, and the test set contains the 2D-3D pairs of the other 28 subjects. Each subject has seven 2D face images and two 3D face data. Note that training set and test set are not overlapped.

6.2. Experiment and Results

In experiments, PCA was employed to reduce the space dimension and preserves 98% of the energy in the principal space for both linear CCA and KCCA. Gaussian kernel was chosen as the the kernel function in KCCA algorithm as

$$k(x, y) = e^{-\frac{\|x-y\|}{2\sigma^2}} \tag{19}$$

Table 1. The weight of each patch

Patch	1	2	3	4	5	6	7	8	9
Weight	0.5	0.7	0.5	0.4	0.6	0.4	0.2	0.3	0.2

For patch based 2D-3D face matching, we divide face into nine patches empirically as Fig.4.

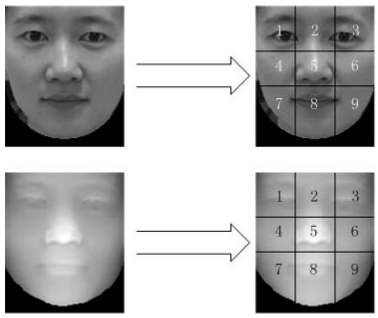

Figure 4. Dividing scheme of the face patches

As mentioned in the last section, the 2D face image and 3D shape were aligned according to the coordinates of the eyes. The patches around the eyes obtain the better alignment. They were supposed to be assigned larger weight. But the nose part contains rich texture information, so the weight is also relative large. In our experiments, the weight of each patch was shown in Table 1. The index of the patch is in row-major order.

After training, the proposed approaches were evaluated on the test set. In this set, the probe consists of 196 facial images of 28 subjects, and the gallery consists of 56 face shape of the same 28 subjects. The comparative resutls are shown in Fig.5 in terms of cumulative match curves (CMC).

The best Rank-1 recognition rate is 87.04% achieved by

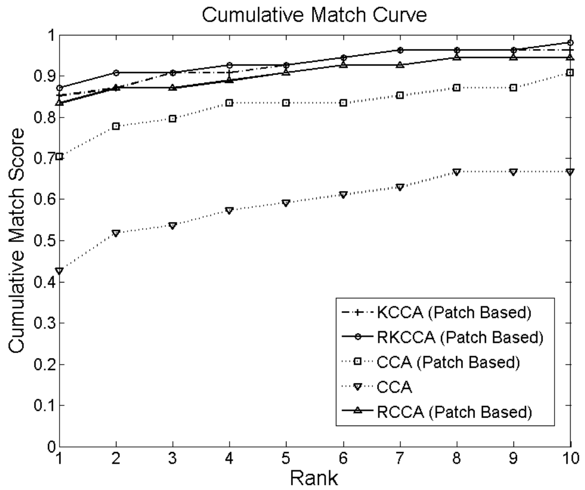

Figure 5. Cumulative Match Curve for 2D-3D face matching

patch based RKCCA. For linear CCA, we can see that the patch based method has significant improved the matching result from only 42.59% to 70.37%. It proves the effectiveness of our patch based methods for 2D-3D face matching. Compared with patch based linear CCA, the path based kernel method also improves the recognition rate from 70.37% to 85.19%.

7. Conclusion

2D-3D face matching is a feasible approach to complement the conventional 3D face recognition systems. In 2D-3D face matching, we need only 2D face images to do recognition, which can be acquired very easily on-site. 2D-3D face matching could serve as an independent biometrics system, or an auxiliary system for the 3D-3D face recognition system. In this paper, we propose a CCA based 2D-3D face matching method. In our method, only several matrix multiplication is needed in the recognition process, so it decreases the computation complexity drastically compared to the conventional 3D-3D face matching. The experiments show that the CCA based 2D-3D face matching have achieved relatively high recognition rate. Moreover, introducing the patch based method and kernel method, the performance is improved further.

Acknowledgements

This work was supported by the following funds: Chinese National Natural Science Foundation Project #60518002, Chinese National 863 Program Projects #2006AA01Z192, #2006AA01Z193, and #2006AA780201-4, Chinese National Science and Technology Support Platform Project #2006BAK08B06, and Chinese Academy of Sciences 100 people project, and AuthenMetric R&D Funds.

References

[1] R. O. Ajmal S. Mian, Mohammed Bennamoun. An efficient multimodal 2d-3d hybrid approach to automatic face recognition. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 29(11):1927–1943, 2007.

[2] P. N. Belhumeur, J. P. Hespanha, and D. J. Kriegman. “Eigenfaces vs. Fisherfaces: Recognition using class specific linear projection”. In *Proceedings of the European Conference on Computer Vision*, pages 45–58, 1996.

[3] D. R. Hardoon, S. R. Szedmak, and J. R. Shawe-taylor. Canonical correlation analysis: An overview with application to learning methods. *Neural Comput.*, 16(12):2639–2664, 2004.

[4] H. Hotelling. “Relations between two sets of variates”. *Biometrika*, 28:321–377, 1936.

[5] X. Hou, S. Li, H. Zhang, and Q. Cheng. Direct appearance models. In *Proc. of IEEE Computer Society Conference on Computer Vision and Pattern Recognition*, pages I:828–833, 2001.

[6] J. Kittler, M. Hatel, R. P. W. Duin, and J. Matas. “On combining classifiers”. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 20(3):226–239, January 1998.

[7] Z. Lei, Q. Bai, R. He, and S. Z. Li. Face shape recovery from a single image using cca mapping between tensor spaces. In *Proc. of IEEE Computer Society Conference on Computer Vision and Pattern Recognition*, 2008.

[8] T. Melzer, M. Reiter, and H. Bischof. Appearance models based on kernel canonical correlation analysis. *Pattern Recognition*, 36(9):1961–1971, September 2003.

[9] A. Rama, F. Tarres, D. Onofrio, and S. Tubaro. Mixed 2d-3d information for pose estimation and face recognition. In *Proceedings of the International Conference on Acoustic, Speech and Signal Processing*, 2006.

[10] D. Riccio and J.-L. Dugelay. Geometric invariants for 2d/3d face recognition. *Pattern Recognition Letters*, (28):1907–1914, 2007.

[11] S. Shan, Y. Chang, W. Gao, B. Cao, and P. Yang. Curse of mis-alignment in face recognition: Problem and a novel mis-alignment learning solution. In *Proc. IEEE International Conference on Automatic Face and Gesture Recognition*, pages 314–320, 2004.

[12] Y. Su, S. Shan, X. Chen, and W. Gao. Patch-based gabor fisher classifier for face recognition. In *Proc. of International Conference on Pattern Recognition*, pages II: 528–531, 2006.

[13] M. A. Turk and A. P. Pentland. “Eigenfaces for recognition”. *Journal of Cognitive Neuroscience*, 3(1):71–86, March 1991.

[14] B. Weyrauch, B. Heisele, J. Huang, and V. Blanz. Component-based face recognition with 3d morphable models. In *Proceedings of the IEEE Computer Society Conference on Computer Vision and Pattern Recognition Workshop*, 2004.

[15] D. Yi, R. Liu, R. Chu, Z. Lei, and S. Li. Face matching between near infrared and visible light images. pages 523–530, 2007.