IMPROVED 3D ASSISTED POSE-INVARIANT FACE RECOGNITION

Liting Wang, Liu Ding, Xiaoqing Ding, Chi Fang

Department of Electronic Engineering, Tsinghua University, Beijing, China
E-Mail: wanglt@ocrserv.ee.tsinghua.edu.cn

ABSTRACT

Recent face recognition algorithm can achieve high accuracy when testing face samples are frontal. However, when face pose changes largely, the performance of existing methods drop drastically. In this paper, we propose an improved algorithm aiming at recognizing faces of different poses when each face class has only one frontal training sample. For each sample, a 3D face is constructed by using 3D Morphable Model (3DMM). The shape and texture parameters of 3DMM are recovering by fitting the model to the 2D face sample which is a non-linear optimization problem. The virtual faces of different views are generated from the 3DMM to assist face recognition. Different from the conventional optimization energy function, proposed energy function takes not only image intensity but also shape constraint into account. In this paper, we locate 88 sparse points from the 2D face sample by automatic face fitting and use their correspondence in the 3D face as shape constraint. We experiment proposed method on the publicly available CMUPIE database which includes faces viewed from 11 different poses and the results show that proposed method is effective and the face recognition results towards pose-variant are promising.

Index Terms— Face Recognition, 3D Morphable Model, Virtual Images, 2D face fitting

1. INTRODUCTION

Pose variation is one of the major difficulties for face recognition. The performance drops dramatically, when large pose variations are presented in the input images, especially when each face class has few non-frontal training face samples. A reasonable way to improve multi-view recognition is to use multiple training views. As far as we know, view-based method has shown its efficiency when the gallery samples are sufficient. However, efficient and robust algorithm is needed when each face class has only one frontal training sample. The main contribution of this paper is to propose a reasonable, efficient and fully automatic pose-invariant face recognition algorithm when each face class has only one frontal training sample.

Recently, many research works towards pose-invariant face recognition when each face class has only one frontal training sample have been developed. Among them, the virtual view image based method is one of the most effective and important. The seminal work is proposed by Blanz and Vetter (1999) [1]. They presented a 3D face reconstruction algorithm by using 3D morphable model. A drawback of this fitting algorithm is it estimates the shape, texture and image conditions from the pixel intensity only. As a result, the energy function presents many local minima and the speed is not able to meet the requirement of most real face recognition systems. Daniel and Jose [2] proposed a pose-invariant 2D face recognition algorithm by using the point distribution models and facial symmetry but did not validate their methods with automatic face fitting. Chai [3] and Hu [4] both proposed an efficient method only by accurate face fitting of 83 facial points which estimates 3D shape from 2D sparse points only. It is fast but ill-posed. In this paper, we develop an improved 3D assisted pose-invariant face recognition method which takes not only the image intensity but also the shape constraint into account. 88 located facial feature points are utilized to obtain a more accurate estimate of the correspondence. Recognition experiments are reported on the publicly available CMU-PIE database which includes faces viewed from 11 different poses.

2. ALGORITHM OVERVIEW

As illustrated in Fig.1, proposed 3D assisted pose-invariant face recognition algorithm consists of three modules, namely 2D face fitting, 2D-3D face reconstruction and face recognition using virtual images.

The basic idea to deal with pose changes which is in the same direction with Franco [5] is to synthesis the multi-view faces with the help of 3D face model. The difference is they need three training images, but our method needs only one frontal training sample and face fitting is implemented before 3D reconstruction. Our method is also different with [3,4]. They estimate 3D shape from 2D sparse points only, but in this paper, we estimate 3D shape from both image intensity and shape constraint by modifying the energy function.

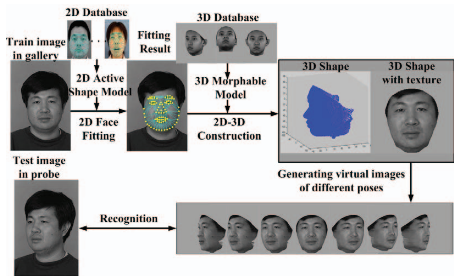

Figure.1. Algorithm overview

The next section of this paper will briefly review 2D face fitting algorithm. Section 4 will motivates the use of shape constraint to fit 3D Morphable Model and presents the formula that combines shape constraint to find the best shape and texture parameters. Then, face recognition algorithm with virtual images are introduced in Section 5. Finally, Section 6 concludes the paper with future direction.

3. 2D FACE FITTING

Frontal face fitting algorithm is mature and the results are satisfactory. In this paper, we adopt an improved active shape model by using random forest with pair-compare features [6]. We adopt this method to precisely locate 88 facial points. The mean point to point error reported is 3.6 pixels when the distance of two eyes is 60 pixels. The face recognition results with automatic fitting and with manually annotated landmarks were also compared.

4. 2D TO 3D RECONSTRUCTION

4.1. 3D Morphable Model

We now describe the 3D model that we use in this paper. We call it 3D morphable model. But note that the model here is a subset of the model in Blanz and Vetter (1999) not includes a reflectance model, like the Phong model.

The main idea behind the morphable face model approach is that given a sufficiently large database of 3D face models, any arbitrary face can be generated by morphing between the ones in the database.

The geometry of a face is represented with a shape-vector $S=(X_1,Y_1,Z_1,X_2,\cdots,Y_n,Z_n)^T \in R^{3n}$ that contains the X,Y,Z coordinates of its n vertices. The texture of a face is presented with a texture-vector $T=(R_1,G_1,B_1,R_2,\cdots,G_n,B_n)^T \in R^{3n}$ that contains the R,G,B color values of the n corresponding vertices. A morphable face model was then constructed using a data set of m exemplar faces, each represented by its shape-vector S_i and texture-vector T_i. New shape S_{new} and new texture T_{new} can be expressed:

$$S_{new}=\sum_{i=1}^{m} a_i S_i, \quad T_{new}=\sum_{i=1}^{m} b_i T_i, \quad \sum_{i=1}^{m} a_i=\sum_{i=1}^{m} b_i=1 \qquad (1)$$

We define the morphable model as the set of faces ($S_{new}(\vec{a})$, $T_{new}(\vec{b})$), parameterized by the coefficients $\vec{a}=(a_1,a_2\cdots a_m)^T$ and $\vec{b}=(b_1,b_2\cdots b_m)^T$. Arbitrary new faces can be generated by varying the parameters $\vec{a}$ and $\vec{b}$ that control shape and texture.

BJUT-3D-R1 database which includes 487 laser-scanned heads is used in this paper to build 3DMM.

4.2. 2D to 3D face reconstruction

Fitting 3DMM to a 2D face image is a non-linear optimization problem. The general strategy is to define an energy function between the novel image and the current guess for the closest model image. It has some advantages, for example, it only needs one frontal face image. The geometry and texture are both considered into optimization. On the other hand, it has some disadvantages. First, such algorithm estimates the shape, texture and image conditions from the pixel intensity only. As a result, the energy function presents many local minima. Second, it is not efficient enough to meet the requirement of most real face recognition systems.

Conventional method defines the sum of squared differences error as:

$$E(\vec{a},\vec{b})=\frac{1}{2}\sum_{x,y}\left[I^{novel}(x,y)-I^{model}(x,y)\right]^2 \qquad (2)$$

The sum is over all pixels (x,y) in the images, I^{novel} is the novel gray level image being matched and I^{model} is the current guess for the model gray-level image.

Our method improves the energy function, which take both image intensity and shape constraint into account. The automatic 2D face fitting result is utilized to obtain a more accurate estimate of the correspondence.

$$E(\vec{a},\vec{b})=\frac{1}{2}\sum_{x,y}\left[I^{novel}(x,y)-I^{model}(x,y)\right]^2 + k\bullet\sum_{i=1}^{J}[P\circ(S^{3DModel}(x_i,y_i))-S_i^{2DPoints}]^2 \qquad (3)$$

The energy function Eq.(3) is different with the conventional energy function Eq.(2). $S^{2DPoints}$ is the 88 automatically located facial points vector. $S^{3DModel}(x_i,y_i)$ is the point to point correspondence between 2D face deformable model and 3D face deformable model as illustrated in Fig 2. P is the projection matrix from 3D to 2D frontal face.

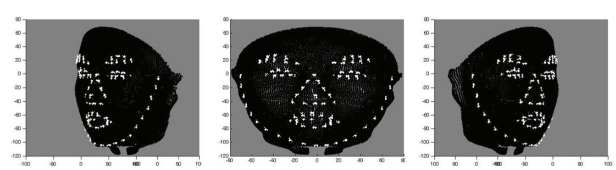

Figure.2. Point correspondence between 3D and 2D facial points.

Now, the added shape constraint part will be expressed in detail.

Let $S=\left(X_1,Y_1,Z_1,X_2,\ldots,Y_{\mathrm{N}},Z_{\mathrm{N}}\right)^T$ be the 3D shape. S_i is the PCA shape-vector, $\overline{S}$ is the mean shape of training database. $\vec{a}=\left(a_1,a_2\cdots a_m\right)^T$ is the coefficients of shape PCA. Then

$$S=\overline{S}+\sum_{j=1}^{M}a_j s_j=\overline{S}+Q\cdot a^T \tag{4}$$

Q is PCA shape eigen-vector matrix. Let $(x_i,y_i)\in\{(x_1,y_1)\cdots(x_l,y_l)\};\quad 1\le i\le J\ \ (J=88)$ be the 88 facial points in 2D image. The shape coefficient $\vec{a}$ convergence to the optimization:

$$\min\sum_{j=1}^{M}a_j^{\,2}/\sigma_j^{\,2}$$

$$\text{s.t. } S_i^{2DPoints}=\overline{S}(x_i,y_i)+Q(x_i,y_i)\cdot a^T \tag{5}$$

The shape constraint is depended on the 2D facial feature points fitting result and geometric distribution of 3D training samples.

In order to minimize the energy function (3), we utilize a dynamic iterative procedure.

Step1. Fit 88 facial feature points in 2D image. Use 2D face fitting result to estimate the proper 3D shape parameter $\vec{a}$ which could be derived from formula (5). Then, set the derived $\vec{a}$ as the initial value.

Step2. Let k = 100, utilize Levenberg-Marquardt algorithm to optimize the energy error, and deriving the current best shape and texture coefficient $\vec{a}$ and $\vec{b}$

Step3. Change k to a smaller value $k\propto\dfrac{1}{E^2}$ and derive shape and texture coefficient $\vec{a}$ and $\vec{b}$ by formula Eq.(5) until energy error is convergent.

k is changing with overall energy error E. At the beginning of optimization, k is set to a large value. With the energy error degrades, k is changing to small value. This shape constraint means 88 point to point correspondence between 3D and 2D face fitting result. In other words, 2D face fitting provides geometry criterion for 3D reconstruction. The geometry parameter $\vec{a}$ is initially set by 2D fitting result. At the beginning of optimization, 2D fitting result takes a priority.

By proposed optimizing algorithm, shape and texture coefficient $\vec{a}$ and $\vec{b}$ are computed and 3D face of one person's gallery template image is constructed by formula Eq. (1).

5. FACE RECOGNITION WITH VIRTUAL IMAGES

We consider the problem of classifying an input feature vector (e.g. Garbor feature) $X=\left(x_1,\cdots,x_d\right)^T$ produced by one of M variation sources $V=\{v_1,v_2,\cdots,v_M\}$. Each feature vector belongs to one of N classes $C=\{c_1,c_2,\cdots,c_N\}$. So every face feature vector has two classes:

$$X\in\left(C,V\right)=\left(\{c_1,c_2,\cdots,c_N\},\{v_1,v_2,\cdots,v_M\}\right) \tag{6}$$

In reality, the feature distributions are seldom known and have to be estimated from training data, implying the need for sufficient samples for per class. For each sample, it is very easy to generate any views by rotating the constructed 3D model to the different poses. However, the pose variations are continuous. For estimating M, we must sample the continuous variation source. In this paper, we just consider three variation factors. Every face image $X\left(x,y;p\right)$ has two variation sources, $p=\left(\theta_x,\theta_y\right)$, where θ_x,θ_y are the angles of the head with x and y axis. The coordinate is shown in Fig.3. We limit the range of these angles $-\dfrac{\pi}{4}\le\theta_x\le\dfrac{\pi}{4},-\dfrac{\pi}{2}\le\theta_y\le\dfrac{\pi}{2}$. The variation $V=\{v_1,\cdots,v_M\}$ is the impact of pose, so $v_i=\left(\theta_x,\theta_y\right)\ 1\le i\le M$. If θ_x samples h_p points in $\left[-\dfrac{\pi}{4},\dfrac{\pi}{4}\right]$ and θ_y samples h_p points in $\left[-\dfrac{\pi}{2},\dfrac{\pi}{2}\right]$.the total points is

$$M=h_p\times h_p=h_p^{\,2} \tag{7}$$

After selecting the center points of variation, we can confirm the descriptive range of every class variation condition probability density function $p\left(X/c_j,v_m\right)$. Here we give one example when $h_p=3$, the ranges are calculated using Eq.(8) and are shown in Fig. 3 (b):

$$\theta_x\in\left[-\frac{\pi}{4}+i\times\frac{D_X}{h_p},-\frac{\pi}{4}+\left(i+1\right)\times\frac{D_X}{h_p}\right]D_X=\frac{\pi}{2},\ i=0,1,h_p-1$$

$$\theta_y\in\left[-\frac{\pi}{2}+j\times\frac{D_Y}{h_p},-\frac{\pi}{2}+\left(j+1\right)\times\frac{D_Y}{h_p}\right]D_Y=\pi,\ j=0,1,h_p-1 \tag{8}$$

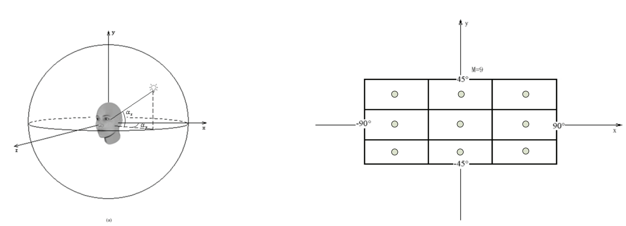

Fig. 3. The order of variation.
(a) Coordinate (b) Example (M=9)

In order to create virtual images, we sample T points in every range. Using these virtual images, the mean vectors are calculated and the discriminant functions are trained. The classifier used in this paper is Minimal Distance Classifier (MDC).

Face recognition performance is evaluated on the publicly available CMU-PIE database which includes faces viewed from 11 different poses. c22, c34, c02, c14, c37, c11, c05, c29, c07, c09 are testing sets which have pose changes, c27 is gallery frontal face set.

The recognition results are listed in Table.1. There are five methods to be compared. M-1, A-1 are proposed methods. M-1 utilizes manually annotated 88 facial feature points. A-1 utilizes automatic fitting 88 facial feature points. We realize Blanze 3D reconstruction method and design our classifiers to do face recognition in R-B method. Chai [3] and Hu [4] are published methods. We just copy their results here. Chai [3] only tested on 4 sets, so results on other 6 sets are not listed below.

The mean consuming time per one recognition procedure (T) is also listed. The total time cost is changing with the times of iteration, dimension of coefficients of shape PCA and texture PCA. In this experiment, we choose the dimension of shape PCA and texture PCA by compressing energy to 90% (truncation error = 10%). That is a tradeoff between speed and accuracy.

Table.1. Face Recognition Performance

	M-1	A-1	R-B [1]	Chai [3]	Hu [4]
C22	19.4	14.93	10.45		15.6
C34	10.45	11.94	11.94		9.2
C02	79.1	71.64	64.18		28.6
C14	77.61	80.6	59.7		22.1
C37	98.51	98.51	94.03	88.2	53.3
C11	95.52	98.51	92.54	95.6	48.4
C05	100	100	100	98.5	84.9
C29	100	100	98.51	97.1	68.0
C07	100	100	100		92.1
C09	100	100	100		85.0
T	1.21s	1.31s	7.31s	4s	4s

Note that the recognition rate degrades when utilizing automatic 2D face fitting compared with manually annotation, especially on large pose set.

We can conclude that proposed method is efficient and effective to deal with pose-variant face recognition.

6. CONCLUSION

Experimental results show that proposed 3D assisted pose-invariant face recognition is efficient and effective. Compared with other existing works, proposed algorithm adds shape constraint into optimization energy function. Our method is fully automatic and efficient. However, we only focus on pose changes without considering illumination changes in this paper. Recently, face recognition performance on illumination dataset is still far from satisfactory. Our future work will focus on face recognition algorithm combining pose and illumination influences together. Another future direction is to combine more information into 3D face reconstruction, such as edge, highlights etc.

7. REFERENCES

[1] Volker Blanz and Thomas Vetter, “A morphable model for the synthesis of 3D faces.” In: Proceedings of ACM SIGGRAPH 99 Conference, pp.187-194, 1999

[2] D.G. Jiménez and J. L. Castro, “Toward Pose-Invariant 2-D Face Recognition Through Point Distribution Models and Facial Symmetry.” *IEEE TRANS ON INFORMATION FORENSICS AND SECURITY*, vol. 2(3), pp. 413-429, 2007

[3] X.J. Chai, S.G. Shan, L.Y.Qing, X.L. Chen, W.Gao. “Pose and Illumination Invariant Face Recognition Based on 3D Face Reconstruction.” *Journal of Software*, vol.17(3), pp.525−534, 2006.

[4] Y.X.Hu, D.L.Jiang, S.C Yan, L. Zhang, H.J. Zhang. “Automatic 3D Reconstruction for Face Recognition.” In: Proceedings of IEEE Automatic Face and Gesture Recognition, pp.843 – 848. 2004

[5] A. Franco, D. Maio, and D. Maltoni. “2D face recognition based on supervised subspace learning from 3D models.” *Pattern Recognition* vol. 41, pp.3822 – 3833, 2008

[6] L.T. Wang, X.Q. Ding, C.Fang. “Generic face alignment using an improved Active Shape Model.” In: Proceedings of IEEE Audio, Language and Image Processing, pp.317 – 321, 2008

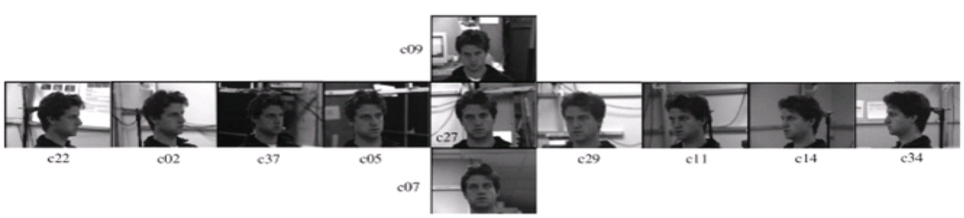

Figure.4. Image Examples in Database