Virtual Face Image Generation For Illumination And Pose Insensitive Face Recognition

Wen Gao[1,2]*, Shiguang Shan*[1]*, Xiujuan Chai*[2]*, Xiaowei Fu*[2]
[1]Institute of Computing Technology, CAS, P.O.Box 2704, Beijing, China, 100080
[2]Department of Computer Science, Harbin Institute of Technology, Harbin, China, 150001

ABSTRACT

Face recognition has attracted much attention in the past decades for its wide potential applications. Much progress has been made in the past few years. However, specialized evaluation of the state-of-the-art of both academic algorithms and commercial systems illustrates that the performance of most current recognition technologies degrades significantly due to the variations of illumination and/or pose. To solve these problems, providing multiple training samples to the recognition system is a rational choice. However, enough samples are not always available for many practical applications. It is an alternative to augment the training set by generating virtual views from one single face image, that is, relighting the given face images or synthesize novel views of the given face. Based on this strategy, this paper presents some attempts by presenting a ratio-image based face relighting method and a face re-rotating approached based on linear shape prediction and image warp. To evaluate the effect of the additional virtual face images, primary experiments are conducted using our face specific subspace method as face recognition approach, which shows impressive improvement compared with standard benchmark face recognition methods.

1. INTRODUCTION

Face recognition has attracted much attention in the past decades for its wide potential applications in commerce and law enforcement, such as mug-shot database matching, identity authentication, access control, information security, and surveillance. Much progress has been made in the past few years [1,2].

Since the 1990s, appearance based methods have been dominant researches, from which two FRT categories were derived: holistic appearance feature based and analytic local feature based. Popular methods belonging to the former paradigm include Eigenface[3], Fisherface[4]. Local Feature Analysis (LFA)[5] and Elastic Bunch Graph Matching (EBGM)[6] are typical instances of the latter category. In recent years, Eigenface, Fisherface, EBGM, Active Shape Models and Active Appearance Model (ASM/AAM)[7, 22], subspace discrimination analysis[8] and SVM[10] based approaches have attracted much attention. FERET evaluation has provided extensive comparisons of these algorithms [9].

However, face recognition remains a difficult, unsolved problem in general. The performance of almost all current face recognition systems, both best academic systems and most successful commercial systems, is heavily subject to the variations in the imaging conditions. It has been discovered by the FERET and FRVT test that pose and illumination variations are among the several bottlenecks for a practical face recognition system [9]. By far, no revolutionary practical solutions are available for these problems. However, some solutions to pose and illumination problems do have emerged including invariant feature based methods [16], 3D linear illumination subspace [4], linear object class [11], illumination and pose manifold [12], Symmetric Shape-From-Shading [8], photometric alignment [13], Quotient Image [14], illumination cones [15], Lambertian Reflectance and Linear Subspace [17], Eigen light-fields [18] and parametric linear subspace [19].

Generally, we may categorize approaches used to cope with variation in appearance into three kinds: **invariant features, canonical forms, and variation modeling** [20].

The first approach seeks to utilize features that are invariant to the changes in appearance. Examples of such representation considered by early researchers are edge maps, image intensity derivatives, and images convolved with 2D Gabor-like filers. However, Adini's empirical study had shown that "None of the representations considered is sufficient by itself to overcome image variations because of a change in the direction of illumination"[16]. Most recently, the Quotient Image [14] is reported to be invariant to illumination and may be used to recognize faces when lighting conditions change.

The second approach attempts to "normalize" away the variation in appearance, either by image transformations or by synthesizing a new image from the given image in some canonical form. Recognition is then performed using this canonical form. Examples of this approach include [8, 21].

The idea of third approach, variation modeling, is to learn, in some suitable subspace/manifold, the extent of the variation in that space/manifold. Recognition is then conducted by choosing the subspace/manifold closest to the novel image. Currently, this paradigm has been recognized as the dominant one among the three approaches[11, 12, 13, 15, 17, 19, 20].

In this paper, we investigate the possibility to augment the training set for modeling the variations by generating virtual face images when changing lighting conditions or viewpoints. This is especially useful for applications that only limited samples per face are available for training.

The paper is organized as: In section 2 we first described briefly our works on ASM for aligning face images. Section 3 describes the ratio-image based face relighting approach, followed by virtual view prediction based on shape prediction. Our recognition approach based on Face Specific Subspaces (FSS) is presented in section 5. Experiments are set up in the last section.

2. Our Works On Feature Correspondence

Both our face relighting method and face rotating method need accurate feature correspondence. Therefore, we first describe our works on feature extraction briefly. Refer to [23], [24], [25] respectively for details of our work on face segmentation, eye localization and face shape extraction. Our face detection method, named Face Center-of-Gravity Template, is based on some observations on the configure relationship between major face organs. The eyes are then localized by growing a region window from the approximate center of the detected face and checking its characteristics. After eyes are located, we attempt to combine the ASM's local texture models and AAM's global appearance models for spare facial feature correspondence. To integrate the local profile and global appearance constraints, the subspace reconstruction residual of the global texture is exploited to evaluate the fitting degree of the current model to the novel image. And, similar to the AAMs, global texture is used to predict and tune the model parameters. Some results of our feature extraction method are shown in Fig.1.

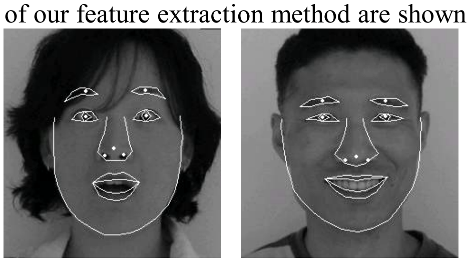


Figure 1. Results of our feature correspondence

3. Ratio-Image Based Face Relighting for Modeling Illumination Variations

In this section, a ratio image based face relighting method is presented. The method is based on the assumptions that any face were a convex surface with a Lambertian function, that is, a face image can be described by the product of the albedo and the cosine angle between a point light source and the surface normal:

$$I(x,y) = \rho(x,y)\vec{n}(x,y)\cdot\vec{s}$$

where $\rho(x,y)$ is the albedo associated with point *x,y* in the image, $\vec{n}(x,y)$ is the surface normal direction associated with point *x,y* in the image, and $\vec{s}$ is the point light source direction and whose magnitude is the light source intensity.

Thus, our problem can be formulated as: Given a face image I_0 under normal light source, $\vec{s}_0$, we need to relight the face under other light sources, e.g. $\vec{s}_i$. To solve this problem, we present a ratio-image based method.

First, we define the ratio-image for the i^{th} face (person) under the k^{th} light source as:

$$\begin{aligned} r_{ik} &= I_{ik} / I_{i0} \\ &= (\rho_i \vec{n}_i \cdot \vec{s}_k)/(\rho_i \vec{n}_i \cdot \vec{s}_0) \\ &= (\vec{n}_i \cdot \vec{s}_k)/(\vec{n}_i \cdot \vec{s}_0) \end{aligned}$$

where ρ_i is the albedo (surface reflectance) associated with the i^{th} face, $\vec{n}_i$ is the corresponding surface normal directions, and $\vec{s}_0$ and $\vec{s}_i$ are the standard and target point light source directions respectively. Thus, we have:

$$I_{ik} = \rho_i \vec{n}_i \cdot \vec{s}_k = (\rho_i \vec{n}_i \cdot \vec{s}_0) \otimes r_{ik} = I_{i0} \otimes r_{ik}\,.$$

where $\otimes$ denotes Cartesian product. This means that, given the ratio-image and the standard face image, we can relight the face to the k^{th} light source.

The ratio-image above defined is almost useless since it is only applicable to the i^{th} face. However, notice that all faces have similar 2D and 3D shapes, so we can try to first warp all faces to the same shape and then compute the ratio-image for relighting the standard face images. It is then easy to reverse warp the relit face image back to its original shape. Currently, we just warp the 2D face image to a predefined mean shape, as defined in ASM. After the warp procedure, all face images are expected to have quite similar 3D shape. Therefore, given a training set, by warp all the face images under different lighting conditions to the same shape, we can define the universal ratio-image for the k^{th} light source as the mean of all the specific ratio-image of each face in the training set:

$$R_k = \frac{1}{N}\sum_{i=1}^{N}\frac{T_{ik}}{T_{i0}} = \tfrac{1}{N}\sum_{i=1}^{N} R_{ik}$$

where N is the total faces in the training set, T_{ik} is the shape-free texture warped from the i^{th} face images lighted under the k^{th} light source, and T_{i0} is its corresponding texture under the standard light source.

By varying the different light source, we can get the ratio-image for each light source and/or combine them to ratio-image for any lighting conditions.

After the ratio-images are computed, any novel face image I_0 with standard lighting conditions can be relit by the following procedure:

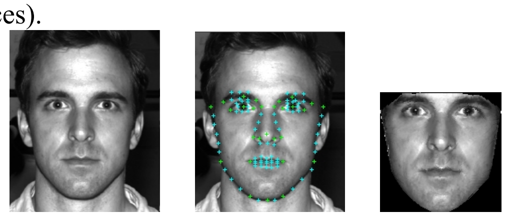
1. Find the face in the picture and extract its 2D face shape using the method descried in Section 2;
2. Warp the face to texture T_0 according to the predefined mean shape;

3. Relight the face image under k-th lighting condition according to the k-th ratio-image by:

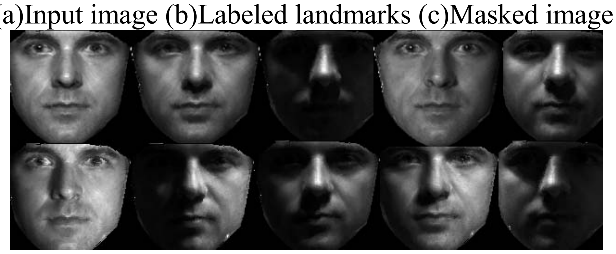
$$T_k = T_0 \otimes R_k ;$$

4. Reverse-warp the texture T_k to its original shape to get the relit image I_k.;

Fig. 2 illustrates some relighting effect of our method on the Yale face Database B (The other 9 person's faces are used to compute the 63 ratio-images for 63 light sources).



(a)Input image (b)Labeled landmarks (c)Masked image



(d) Relit images

Figure2. Ratio-image based face relighting

4. Virtual View Generation For Different Poses

Similar to face relighting, the virtual view generation problem is defined as: given a frontal view of an unknown face, generating its view under other poses. Let P_0 denote the frontal pose, P_1 be another pose (e.g. right rotating 30^o out of the image plane), and I_0 be the image under pose P_0. Our goal is generating its view I_1 under pose P_1.

A linear regression method is exploited to solve this problem: a learning set containing pairs of the shapes of the two views under P_0 and P_1 is collected, a linear mapping between them are learned and applied to any given novel frontal image to predict its shape under pose P_1. Let $\Re = \{(I_1^0, I_1^1), (I_2^0, I_2^1), \cdots, (I_m^0, I_m^1))\}$ be an image set containing pairs of the two views under P_0 and P_1, and $\Im = \{(S_1^0, S_1^1), (S_2^0, S_2^1), \cdots, (S_m^0, S_m^1))\}$ be the corresponding shape set containing pairs of the shapes for the two views in the learning set $\Re$. A linear mapping P can be learnt easily from $\Im$. So, for a given novel image I_0 under pose, P_0, its shape vector S^0 is first extracted using method in Section 2. Then, its face shape S^1 viewed under pose P_1 is predicted by:

$$S^1 = PS^0 .$$

Then we can generate the virtual view by an image warping procedure based on S^0 and S^1. Fig.3 shows two examples of the generation results, in which the first row are the original frontal views with landmarks overlapped; the second row is the generation results.

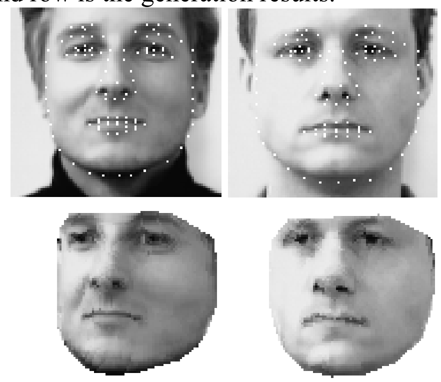


Figure 3. Virtual view generation

5. FSS-based Face Recognition

In our previous work, we have proposed a Face-Specific Subspace (FSS) based face recognition method [26]. This method is motivated, but essentially different from the traditional Eigenface. In Eigenface, each face image is represented as a point in a low dimensional face subspace shared by ***all*** faces; however, we experimentally show that one of the demerits of such a strategy is that the most discriminant features of a specific face are not accurately represented. Therefore, we propose to model each face by one individual face subspace, named Face-Specific Subspace. Distance from the face-specific subspace, that is, the reconstruction error, is then exploited as the similarity measurement for identification.

Each FSS is learnt from the training images of the specific face and represented as a *4*-tuple by:

$$\Re_k = (U_k, \Psi_k, \Lambda_k, d_k),$$

where U_k is the eigenvector matrix, Λ_k is the eigenvalues, Ψ_k is the mean of the k^{th} face, and d_k is the dimension of the FSS.

Similar to DFFS in Eigenface method, the similarity of any image to a face can be measured by using the Distance From FSS (DFFSS): less DFFSS means more probability that the image belongs to the corresponding face. It can be formulated as follows: Let Γ be any input image. It can be projected to the k^{th} FSS by: $W^{(k)} = U_k{}^T \Phi^{(k)}$, where $\Phi^{(k)} = \Gamma - \Psi_k$. Then $\Phi^{(k)}$ can be reconstructed by: $\Phi_r^{(k)} = U_k W^{(k)}$. So, Γ's distance from k^{th} FSS (DFFSS) is computed as the following reconstruction error:

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

The DFFSS can be regarded as the similarity of the input pattern Γ to the face corresponding to the k^{th} FSS. Therefore, the following minimal distance classifier can be naturally formulated:

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

6. Experiments on Yale Face Database B

To evaluate the effect of augmenting training set for face recognition, we conduct experiments on the Yale Face Database B (Refer [15] for detailed information on this face database). We choose just the frontal set in this DB, containing 640 images from 10 persons, each person has 64 frontal images under 64 different lighting conditions. To test different face recognition methods, we choose the frontal face image under the standard lighting of each person as training images, other 63 images lighted under different for testing.

Leave-one-out strategy is exploited to generate the ratio-image for each lighting configure. Then, 63 additional face images from each standard lighting face images are generated to augment the training set for the FSS method. The testing results on the 5 subsets are shown in Fig. 4.(Note: other methods tested are correlation, PCA and FaceIt3.0 system. They do not using the additional virtual images for train).

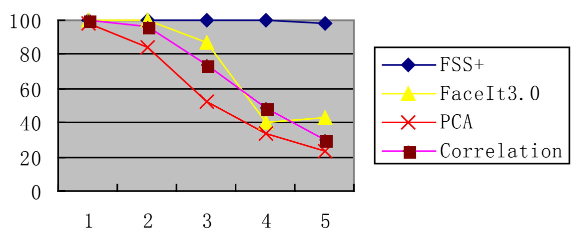


Figure 4. Performance comparison on the 5 subsets in the Yale Face Database B

From Fig.4, a significant performance improved can be observed, which obviously profits from the augment of the training set.

Acknowledgements

This research is sponsored partly by NSF of China (No.69789301), National Hi-Tech Program of China (No.2001AA114190), and YinChen Net. Co.

REFERENCES

[1] R. Brunelli and T. Poggio, "Face Recognition: Features versus Template", *TPAMI*, 15(10), pp1042-1052, 1993

[2] R.Chellappa, C.L.Wilson ect. "Human and Machine Recognition of faces: A survey", *Proc. of the IEEE*, 83(5), pp705-740, 1995.5

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

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

[5] 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

[6] 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

[7] T.F.Cootes, G.J.Edwards, C.J.Taylor, "Active Appearance Models", *ECCV,* vol.2, pp484-498, 1998.

[8] W.Zhao and R.Chellappa, "Robust Image-Based 3D Face Recognition", *CAR-TR-932, N00014-95-1-0521, CS-TR-4091*, Center for Auto Research, UMD, 2000.1

[9] P.J.Phillips, H.Moon, etc. "The FERET Evaluation Methodology for Face-Recognition Algorithms", *IEEE TPAMI*, Vol.22, No.10, pp1090-1104, 2000

[10] G.Guo, S.Z.Li and K.Chan, "Face Recognition by Support Vector Machines", FG'02, pp196-201, Grenoble, 2000.3

[11] 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

[12] H.Murase, S.Nayar, Visual Learning and recognition of 3D object from appearance, *IJCV*, 14:5-24, 1995

[13] A.Shashua, On Photometric Issues in 3D visual recognition from a single 2D Image, *International Journal of Computer Vision*, 21(1/2), 99-122, 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] 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*

[16] Y.Adini, Y. Moses, S.Ullman, Face Recognition: The Problem of Compensting for changes in illumination Direction, *IEEE TPAMI,* Vol.19, No.7, pp721-732, 1997

[17] R.Basri, D. Jacobs, Lambertian Reflectance and Linear Subspaces, ICCV2001, Beckman Institute. Vol.2, p383-390,

[18] R.Gross, I.Matthews, S.Baker, Eigen Light-Fields and Face Recognition Across Pose, Proc. of FG02

[19] K.Okada, C.Malsburg, Pose-Invariant Face Recognition with Parametric Linear Subspaces, Proc. of FG02

[20] T.Sim, T.Kanade, Combining Models and Exemplars for Face Recognition: An Illuminating Example,In Proceedings of Workshop on Models versus Exemplars in Computer Vision, CVPR 2001.

[21] P.J.Phillips, Y.Vardi, Efficient Illumination Normalization of Facial Images", PRL, 17(1996), 921-927

[22] T.F.Cootes, C.J.Taylor, D.Cooper, and J. Graham. Active shape models--their training and application. Computer vision and image understanding, 61(1): pp38-59, 1995.

[23] J.Miao, H.Zhang, W.Gao, et al., FaceTracker: A Human Face Tracking and Facial Organ Localizing System, the Proc. of the 8th ICCV, Cananda July.2001 pp.743-743,

[24] B.Cao, S.Shan, W.Gao, D.Zhao. Localizing the iris center by region growing search. Proc. of the IEEE ICME 2002.

[25] W.Wang, S.Shan, W.Gao, B.Yin, Combining Active Shape Models and Active Appearance Models For Accurate Image Interpretation, Submit to ICASSP 2003.

[26] S.Shan, W.Gao, D.Zhao, Face Identification From A Single Example Image Based On Face-Specific Subspace (FSS), Proceedings of the IEEE ICASSP2002