Face Recognition Robust to Head Pose from One Sample Image

Ting Shan[†], Brian C. Lovell[†], and Shaokang Chen

Intelligent Real-Time Imaging and Sensing Group
EMI, School of ITEE, The University of Queensland
and
[†]National Information and Communications Technology Australia (NICTA)
{shanting, lovell, shaokang}@itee.uq.edu.au

Abstract

Most face recognition systems only work well under quite constrained environments. In particular, the illumination conditions, facial expressions and head pose must be tightly controlled for good recognition performance. In 2004, we proposed a new face recognition algorithm, Adaptive Principal Component Analysis (APCA) [4], which performs well against both lighting variation and expression change. But like other eigenface-derived face recognition algorithms, APCA only performs well with frontal face images. The work presented in this paper is an extension of our previous work to also accommodate variations in head pose. Following the approach of Cootes et al, we develop a face model and a rotation model which can be used to interpret facial features and synthesize realistic frontal face images when given a single novel face image. We use a Viola-Jones based face detector to detect the face in real-time and thus solve the initialization problem for our Active Appearance Model search. Experiments show that our approach can achieve good recognition rates on face images across a wide range of head poses. Indeed recognition rates are improved by up to a factor of 5 compared to standard PCA.

1. Introduction

Variations in head pose, illumination, and facial expression can significantly degrade the performance of a face recognition system. Most of the existing face recognition systems only work well in well-controlled environments. Face recognition in unconstrained environments is still a challenging task. In 2004, we proposed a new face recognition algorithm, Adaptive Principal Component Analysis (APCA) [4], which is insensitive to variations in lighting conditions and facial expression. The work presented in this paper is an extension to our previous work focused on solving the head pose problem.

Existing approaches to pose invariant face recognition can be classified into two categories: 2D based approaches and 3D based approaches.

In 1994, Pentland *et.al* [1] proposed "view-based" eigenspace for use in face recognition under variable pose. Each view-space's eigenvectors are used to compute the "distance-from-face-space" (DFFS), once the proper view space is determined, the input face image is encoded using the eigenvectors of that view-space, and then recognized.

In 2000, Cootes *et.al* proposed "View-based Active Appearance Models" [2] which was based on the idea that a small number of 2D statistical models are sufficient to capture the shape and appearance of a face from any viewpoint. They demonstrated that to deal with pose angle change from left to right profile, only 3 distinct models were needed. They applied this method on face tracking, but did not do any face recognition experiments.

In 2003, Chai *et.al* [8] presented an affine transformation algorithm based on statistical analysis. Their algorithm partitions the face into 3 rectangular regions and the affine transformation parameters associated with different poses are learned from the one-to-one rectangle mapping relations. Experiments on the FERET dataset show that their method can dramatically increase recognition rate compared to the face recognition without pose alignment. But their recognition rate was still quite low, only reaching an average of 58% under pose rotations of 30 degrees. Moreover, they didn't show an automatic way to mark the key points on facial structures.

Also in 2003, Blanz *et.al* [3] developed a 3D morphable model for face recognition. 3D model is learned from a set of textured 3D scans of heads, and

Proceedings of the 18th International Conference on Pattern Recognition (ICPR'06)
0-7695-2521-0/06 $20.00 © 2006 IEEE

this face model is directly matched with single or multiple input images. It achieved promising results for illumination and pose invariant face recognition. But it is computationally too expensive, and couldn't be used in real-time face recognition system.

Motivated by Cootes *et.al*, we propose a similar approach to synthesize frontal view face image from a single novel face image in any pose, and feed this synthesized frontal face image into our face recognition algorithm, APCA [4]. The experimental results show our approach achieves good recognition rates on large head pose angles in near real-time.

2. Background Knowledge

In 2004, we developed a pattern classification algorithm Adaptive Principal Component Analysis (APCA) which inherited merits from both PCA and FLD (Fisher Linear Discriminant) by warping the face subspace according to the within- and between-class covariance. It consisted of four steps:

- Subspace Projection: Apply PCA to project face images into the face subspace to generate the m-dimensional feature vectors
- Whitening Transformation: The subspace is whitened according to the eigenvalues of the subspace with a whitening power p

$$\mathrm{cov} = diag\{\lambda_1^{-2p}, \lambda_2^{-2p}, \ldots \lambda_m^{-2p}\} \qquad (1)$$

- Filtering the Eigenfaces: Eigen-features are weighted according to the identification-to-variation value ITV with a filtering power q.

$$\gamma = diag\ \{ITV_1^{\,q}, ITV_2^{\,q}, \ldots ITV_m^{\,q}\} \qquad (2)$$

- Optimizing the cost function: Minimize the cost function according to the combination of error rate and the ratio of between-class distance and within-class distance:.

$$OPT = \sum_{j=1}^{M} \sum_{k=1}^{K} \sum_{m} \frac{d_{jj,k_0}}{d_{jm,k_0}}, \forall m \in d_{jm,k_0} < d_{jj,k_0}, m \in [1 \cdots m] \qquad (3)$$

More details about APCA can be seen in [4]. Experiments show this technique performs well under changes in lighting conditions and facial expression.

3. Facial Feature Interpretation

3.1. Active Appearance Models

First introduced by Cootes and Taylor [5] in 2001, Active Appearance Models are a powerful tool to describe deformable object images. Given a collection of training images for a certain object class where the feature points have been manually marked, a shape and texture can be represented by applying PCA to the sample shape and texture distributions as:

$$x = \bar{x} + P_s c \qquad (4)$$

and

$$g = \bar{g} + P_g c \qquad (5)$$

where $\bar{x}$ is the mean shape, $\bar{g}$ is the mean texture and P_s, P_g are matrices describing the respective shape and texture variations learned from the training sets. The parameters, **c** are used to control the shape and texture change.

3.2 Combination of Cascade Face Detector with Active Appearance Model Search

The initialization of the Active Appearance Model search is a critical problem since the original AAM search is a local gradient ascent. Some failed AAM searches due to the poor initialization can be seen in Figure 1.

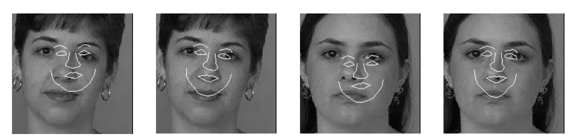

Figure 1. Failed AAM search due to poor initialization

We solve the initialization position problem by applying the Cascade face detector [6] to provide the initialization position to subsequent AAM searches.

3.2.1. Adaboost-based Cascade Face Detection. In 2001, Viola and Jones [6] proposed an image-based face detection system which can achieve remarkably good performance in real-time. The main idea of their method is to combine weak classifiers based on simple binary features which can be computed extremely fast. Simple rectangular Haar-like features are extracted; face and non-face classification is done using a cascade of successively more complex classifiers which discards non-face regions and only sends face-like candidates to the next layer's classifier. Each layer's classifier is trained by the AdaBoost learning algorithm. Our face detector is based on the Viola-Jones approach using our own training sets.

The cascade face detector finds the location of a human face in an input image and provides a good starting point for the subsequent AAM search which then precisely marks the major facial features.

4. Frontal-View Synthesis

4.1 Rotation Model and Pose Estimation

Here we follow the method of Cootes *et al* [2]. They assume that the model parameter **c** is related to the viewing angle, θ, approximately by:

$$c = c_0 + c_c \cos(\theta) + c_s \sin(\theta) \qquad (6)$$

where c_0, c_c and c_s are vectors which are learned from the training data. (Here we consider only head turning. Head nodding can be dealt with in a similar way).

Given a new face image with parameters **c**, we can estimate orientation as follows. We first transform equation (6) to:

$$c - c_0 = (c_c\, c_x)\begin{pmatrix} \cos\theta \\ \sin\theta \end{pmatrix} \qquad (7)$$

let R_c^{-1} be the left pseudo-inverse of the matrix $(c_c \mid c_s)$, then (7) becomes

$$R_c^{-1}(c - c_0) = \begin{pmatrix} \cos\theta \\ \sin\theta \end{pmatrix}. \qquad (8)$$

Let $(x_\alpha, y_\alpha)' = R_c^{-1}(c - c_0)$, then the best estimate of the orientation is $\tan^{-1}(y_\alpha / x_\alpha)$.

4.2 Frontal-View Synthesis

After we estimate the angle θ, we can use the model to synthesize new views. Here we will synthesize a frontal view face image, which will be used for face recognition.

Let c_{res} be the residual vector which is not explained by the rotation model,

$$c_{res} = c - (c_0 + c_s \cos(\theta) + c_s \sin(\theta)). \qquad (9)$$

To reconstruct at a new angle, α, we simply use the parameters:

$$c(\alpha) = c_0 + c_c \cos(\alpha) + c_s \sin(\alpha) + c_{res}\,. \qquad (10)$$

here α is 0, so this becomes:

$$c(0) = c_0 + c_c + c_{res} \qquad (11)$$

The shape and texture at angle 0°can be calculated by:

$$x(0) = \bar{x} + P_s c(0) \qquad (12)$$

and

$$g(0) = \bar{g} + P_g c(0) \qquad (13)$$

The new frontal face image then can be reconstructed.

5. Experimental Results

5.1 Training of Cascade Face Detector

The training database includes 4916 hand labeled faces (Figure 2), and the negative training data were randomly collected from the internet. After training, our final face detector has 24 layers with a total of 2913 features.

Figure 2. Example face images used for training

5.2 Training of Face Model and Rotation Model

In our trials, we collect face image samples from 40 individuals for the training set. For each person we have 3 images in 3 pose views (left 15°, frontal, right 15°) extracted from the Feret face database [7]. We label each of these 120 face images with 58 points around the main features (Figure 3).

We then apply our combined Adaboost-based cascade face detector and AAM search on the rest of the Feret b-series dataset where each person has 7 pose angles, ranging from left 40°, 25°, 15°, 0° to right 15°, 25°, 40°. Despite Cootes et al's claim that a near frontal face model can deal with pose change from left 45° to right 45°, we found that AAM search can't locate the facial features precisely on most of face images with high (40°) pose rotation, even given a very good initialization position —especially for face images with a large nose. AAM search on the remaining face images with smaller pose angle change can achieve 95% search accuracy rate. We discarded those face images with 40° pose rotation and will leave these to future work. Parameters c_0, c_c and c_s are learned from the successful AAM search samples.

Proceedings of the 18th International Conference on Pattern Recognition (ICPR'06)
0-7695-2521-0/06 $20.00 © 2006 IEEE

Figure 3. Sample training face images with 58 points labeled on the main facial features

Figure 4. (L to R) Frontal view, turned view, synthesized frontal view from turned view using proposed methods

5.3 High Pose Angle Face Recognition Results

We trained our APCA face recognition algorithm using the Asian Face Database as in [4]. We then choose face images from 46 persons with good AAM search results. We applied both PCA and APCA on original face images and synthesized frontal images respectively for testing. We only register the frontal view images into the gallery and use the high pose angle images for testing. The recognition results are shown in figure 5.

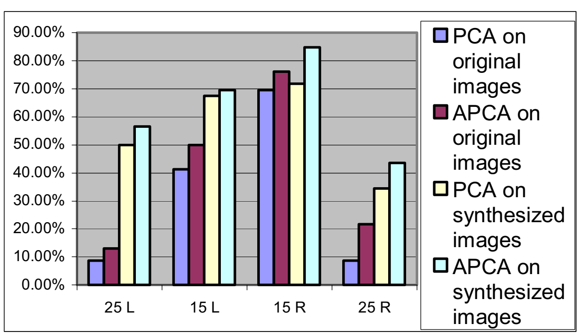

Figure 5. Recognition rate for PCA and APCA on original and synthesized images

It can be seen from figure 5 that the recognition rates of PCA and APCA on synthesized images is much higher than that of the original high pose angle images. The recognition rate increases by up to 500% for images with view angle of 25º. Yet even for smaller rotation angles less than 15º, the accuracy increases by up to 20%. Note that the recognition performance of APCA is always significantly higher than PCA, which is consistent with the results in [4].

6. Conclusions and Future Work

In this paper we proposed a framework for face recognition robust to changes in head pose using only one image per person in the gallery. We developed a Face Model to describe face shape and texture change, a Rotation Model to estimate pose angle and synthesize the frontal face image. We also trained a cascade face detector to detect the approximate location of the face to provide the initialization position for AAM search. Finally we use Adaptive Principal Component Analysis (APCA) which is insensitive to illumination and expression change for face recognition. The experimental results show that after frontal pose synthesis, the recognition rate increases significantly especially for larger rotation angles.

Currently we only have one frontal face model, and only can handle pose angle change from left 25° to right 25°. In our continuing work, we will apply similar techniques to implement quarter-left and quarter-right face models and rotation models to address the larger pose rotation problem.

7. References

[1] A. Pentland, B.Moghaddam and T.Starner, “View-based and Modular Eigenspaces for Face Recognition”, Proc. of ITEE Conference on CVPR 94

[2] T.F.Cootes, K.Walker and C.J.Taylor, “View-based Active Appearance Models”, Proc. of The Fourth ITEE International Conference on AFGR 2000

[3] V.Blanz and T.Vetter, “Face Recognition Based on Fitting a 3D Morphable Model”, ITEE Transactions of Pattern Analysis and Machine Intelligence, Volume 25, Issue 9, Pages: 1063 – 1074, 2003

[4] Shaokang Chen and Brian C. Lovell, “Illumination and Expression Invariant Face Recognition with One Sample Image”, ICPR, pp.300-303, 17th International Conference on Pattern Recognition (ICPR’ 04) –Volume 1, 2004

[5] T.F.Cootes and C.J.Taylor, “Active Appearance Models”, IEEE PAMI, Vol.23, No.6, pp. 681-685, 2001

[6] Paul Viola and Michael Jones, “Rapid Object Detection using a Boosted Cascade of Simple Features”, In Proc on CVPR, pp. 511-518, 2001

[7] http://www.itl.nist.gov/iad/humanid/feret/ [last visited 20-Dec-2005]

[8] Xiujuan Chai, Shiguang Shan and Wen Gao, “Pose normalization for robust face recognition based on statistical affine transformation”, Information, Communications and Signal Processing, and the Fourth Pacific Rim Conference on Multimedia, 2003