Making 2D face recognition more robust using AAMs for pose compensation

Peter Huisman*, Ruud van Munster,
Stephanie Moro-Ellenberger
Dept. Imaging Systems, Signal Processing Group
TNO Industry & Science, the Netherlands

Raymond Veldhuis, Asker Bazen
Dept. of EEMCS, Signals and Systems Group
University of Twente, the Netherlands

Abstract

The problem of pose in 2D face recognition is widely acknowledged. Commercial systems are limited to near frontal face images and cannot deal with pose deviations larger than 15 degrees from the frontal view. This is a problem when using face recognition for surveillance applications in which people can move freely. We suggest a preprocessing step to warp faces from a non frontal pose to a near frontal pose. We use view-based active appearance models to fit to a novel face image under a random pose. The model parameters are adjusted to correct for the pose and used to reconstruct the face under a novel pose. This preprocessing makes face recognition more robust with respect to variations in the pose. An improvement in the identification rate of 60% (from 15 % to 75%) is obtained for faces under a pose of 45 degrees.

1. Introduction

This research was initiated by the observation that 2D commercial face recognition is limited to the recognition of faces under near frontal poses. This was also mentioned in a survey by Zhao et al. [14] who state that together with varying illumination, pose is one of the major problems in face recognition. Current commercially available systems are limited to face recognition within 15 degrees deviation from the frontal view. Face recognition performance using frontal faces is considered to be good enough for application in real life situations.

One of the main advantages of face recognition is the non obtrusiveness. This property makes the technique especially attractive for surveillance applications. However in surveillance settings there is often only a small time frame to capture a face with a high probability that the grabbed images do not contain the required frontal face images.

We suggest a preprocessing step in order to warp faces from a non frontal pose to a frontal one in order to perform frontal face recognition with commercial 2D face recognition systems.

First an overview and comparison of a few possible techniques is presented in Section 2. Then an overview of view-based active appearance models (AAMs) used to reconstruct the face images under a novel pose is presented in Section 3. Section 4 gives information on choices that were made with respect to the implementation. The results of two experiments are presented in Section 5. Finally a discussion (Section 6) and recommendations for future work (Section 7) are given.

*Corresponding author. E-mail address: peter.huisman@tno.nl

2. Several methods for reconstruction

We want to construct photo realistic images that can be used in commercial face recognition systems. Three interesting methods are mentioned in the literature that combine face reconstruction with face recognition. The three methods with the main pros and cons per method are given below.

1. 3D Morphable Models (3DMM) [1], [2]
- Pro: Visual image quality.
- Pro: Recognition performance.
- Con: High computational load.
- Con: Manual interaction required in the on-line phase.
- Con: Need for 3D scanners to construct the model.

2. View-based Active Appearance Models (AAM) [7]
- Pro: Fast.
- Con: Lower image quality with respect to the other two.

3. Light-fields (Lf) [4], [9]
- Pro: Image quality.
- Con: Need for a camera array [4] or knowledge of camera intrinsics [9].

AAMs are probably the lesser of the three with respect to visual image quality but the fact that AAMs are fast makes

Proceedings of the 7th International Conference on Automatic Face and Gesture Recognition (FGR'06)
0-7695-2503-2/06 $20.00 © 2006 IEEE
IEEE COMPUTER SOCIETY

the use of these models attractive for use in surveillance settings. The other two methods have a few important disadvantages with respect to AAMs. The large drawbacks of 3DMMs are the high computational load and the need for manual interaction in the process of synthesizing images. The need for camera intrinsics and relative orientation to the camera in Light-fields make it a less attractive candidate than AAMs.

3. Reconstruction of view-based AAM instances

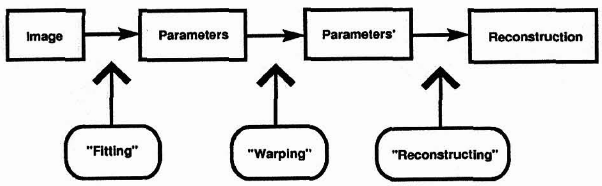

Figure 1. The block diagram illustrates the flow from image to reconstruction. An AAM is "fitted" to a face image which results in a set of parameters. These are "warped" to a novel set of parameters. Finally the face is "reconstructed" which results in a face image under a novel viewing pose

The entire process from original image to reconstruction is illustrated in Figure 1. The first step is fitting an AAM to a face image which results in a set of parameters. These parameters are then warped to a new set of parameters using a rotation model. We use the term warping with respect to changing the model parameters as well as changing the images. The warped model parameters are then used to reconstruct the face image which is now under a novel pose.

Cootes et al. [7] introduced view-based AAMs and also showed that the model parameters can be warped and used to reconstruct the face under novel poses. We will show that these reconstructions can be used to make commercial face recognition more robust against variation due to pose.

3.1. Combined AAMs

A combined AAM [6] models the shape and texture variation seen in a training set with one set of model parameters. By using faces with different poses it is possible to construct a model that represents variation due to identity and rotation. Labeled images are needed to construct a training set of shapes and textures. The images are labeled with landmarks which are the coordinates of distinctive points in the face image. These coordinates form the shape vector $\boldsymbol{x}$ of the face. The texture vector $\boldsymbol{g}$ of the face is given by the pixel values inside the convex hull of the landmarks. By using PCA on the training data it is possible to obtain a model in which the model parameters $\boldsymbol{c}$ control both the shape and texture:

$$\begin{aligned} \boldsymbol{x} &= \overline{\boldsymbol{x}} + \boldsymbol{Q}_s \boldsymbol{c} \\ \boldsymbol{g} &= \overline{\boldsymbol{g}} + \boldsymbol{Q}_g \boldsymbol{c} \end{aligned} \tag{1}$$

and where $\overline{\boldsymbol{x}}$ and $\overline{\boldsymbol{g}}$ are respectively the mean shape and mean texture. $\boldsymbol{Q}_s$ and $\boldsymbol{Q}_g$ describe the variation in the training set and come from the eigenvectors of the covariance matrices of the training set.

By fitting the model on a novel face image it is possible to find a set of model parameters that represent the given face. The model parameters are found by determining the smallest possible difference image between the reconstruction and the original in a number of iterative steps [6].

at least three linear models are needed to capture the entire pose range from -90 to 90 degrees around the vertical head axes. Fewer linear models is impossible because landmarks disappear when the head rotates. Gong et al. [8] suggest a non-linear model using KPCA which results in only one model. The issue of rotation (i.e. reconstruction under a novel pose) is however not explicitly addressed.

3.2. Rotation Model

In order to rotate the head it is necessary to have a link between the model parameters and the viewing angle within one model (i.e. intra-model rotation) and over multiple models (i.e. inter-model rotation). Too little annotated training data was available to investigate the inter-model rotation so we only took intra-model rotation into consideration. Cootes [7] suggests that the viewing angle θ can be linked to model parameters $\boldsymbol{c}$ by

$$\boldsymbol{c}(\theta) = \boldsymbol{c}_0 + \boldsymbol{c}_x \cos(\theta) + \boldsymbol{c}_y \sin(\theta) \tag{2}$$

where $\boldsymbol{c}_0$, $\boldsymbol{c}_x$ and $\boldsymbol{c}_y$ are learned from the training data.

The pose can be determined by computing the left pseudo inverse $\boldsymbol{R}_c^{-1}$ of the matrix $(\boldsymbol{c}_x|\boldsymbol{c}_y)$. Now there is a relation between the viewing angle and the model parameters based on Equation 2:

$$(x_a, y_a)^T = \boldsymbol{R}_c^{-1}(\boldsymbol{c}(\theta) - \boldsymbol{c}_0) \tag{3}$$

and

$$\tan\theta = y_a/x_a \tag{4}$$

We verified the rotation model by constructing Figure 2 in which an independent test set was used to estimate θ. The estimate of θ is not entirely correct which is caused by annotation errors, mis fits and inaccuracy in the model. However it is probably be good enough to use in the reconstruction of the face images.

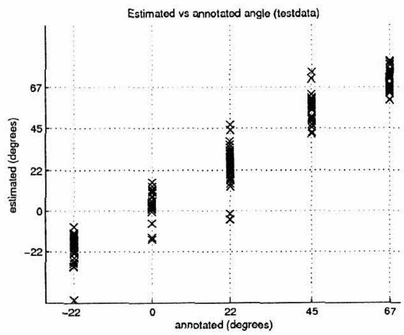

Figure 2. The annotated angles plotted versus the estimated angles using the rotation models for the frontal and the semi profile AAMs.

3.3. Reconstruction

The rotation model describes only a part of the model parameters. Therefore the model parameters are divided into a part that describes the variation due to rotation and a part that describes the other variations (e.g. the variation due to identity). This is done by introducing the residual vector $\boldsymbol{r}$:

$$\boldsymbol{r} = \boldsymbol{c}(\theta) - \boldsymbol{c}_0 + \boldsymbol{c}_x \cos(\theta) + \boldsymbol{c}_x \sin(\theta) \qquad (5)$$

This provides a method for rotating faces from the current angle θ to a new angle $\widehat{\theta}$. We estimate the angle θ of the current view with Equation 4 and determine the residual $\boldsymbol{r}$ using Equation 5. Now the model parameters are warped to a novel view by adding $\boldsymbol{r}$ to Equation 2 and substituting $\widehat{\theta}$ for θ.

The reconstruction of the face images is done using thin plate splines [3], mainly because deformations introduced by the warping process are smooth, as opposed to deformations in triangulation.

4. Implementation

Different data sets and tools were used in performing the experiments. Some general remarks on both are given in this section.

4.1. Data

Three image databases are used in the training of the AAMs and the rotation model. The CMU PIE [12] and IMM [10] are publicly available and another one was collected at TNO for previous research on face recognition.

The IMM and TNO data sets were used for finding a relation between the number of training images and the "identity" in the reconstruction. The image set contained three faces of each individual under 3 poses ranging from -40 to +40 degrees divided into a training set and a test set.

Three images (under 22, 45 and 67 degrees) of 68 individuals from the CMU PIE data set were used to build a semi profile view-based AAM.

The images were semi automatically labeled with a set of landmarks defining the shape and texture of the images. In the frontal model 68 landmarks were used and in the semi profile model 63. An AAM was used to find a near fit which was adjusted if necessary (bootstrapping).

In order to enlarge the training set (i.e. introduce more variation) in the frontal model with little effort we extended the training set by adding the mirrored face images and annotations. This is mainly valid because the images of the left and right pose were often taken under different angles.

4.2. Tools

The training and fitting of AAMs is done using the AAM-API by Stegmann et al. [13]. The standard executable is slightly adapted to output the desired variables. The averages and variation matrices in Equation 1 are the output of the training phase. The fitting of the model to a face image results in a set of model parameters. These are then processed using Matlab.

The face recognition software is from Cognitec (which is one of the market leaders on commercial 2D face recognition software). It has the least stringent demands on resolution (32 pixels inter pupil distance) compared to other vendors. The software is robust against pose ($\pm$ 15 degrees deviation from frontal) [5].

5. Experiments and Results

This section presents the experiments and results that provide an answer to the two following questions:
1. Can enough detail (i.e. identity) be captured in the AAM to use the reconstructed faces in 2D face recognition?
2. Can face images under non-frontal poses be warped to (near) frontal poses?

A consequence of using AAMs is the loss of detail since they are based on an average and a model of the variation (Equation 1). The retained detail (i.e. identity) in the reconstructions has to be sufficient to perform face recognition. If this is the case we want to know if we can use the rotation model to adjust the model parameters to compensate for the non-frontal pose. We use the semi profile view-based

AAM to rotate faces from any pose in this model to the near frontal edge of the model. This edge is near the 15 degree pose limit of the face recognition software.

Both questions will be treated separately in the following sections.

5.1. Identity

Four frontal view-based AAMs were trained to establish the relationship between the identity for recognition and the size of an AAM. The models were trained using a different number of face images for each model. By using more face images more variation due to identity is introduced. The training set contained approximately 50% of original images and annotations and 50% of mirrored data. Table 1 shows the number of training images for the 4 models. The training sets were al sub sets of the first larger model. All of the images came from the TNO and IMM data sets (for every IMM image there are three TNO images).

model number	# of training images	# of model parameters
1	411	158
2	391	149
3	321	131
4	250	111

Table 1. Number of training images and model parameters in 4 different AAMs

An independent test set of 243 images and 3 images per individual (from the IMM (54) and TNO (189) data set) is used to evaluate the face recognition performance. Each individual was recorded in a right and left view (within approximately 35 degrees deviation from frontal) and in a frontal view. The 4 AAMs are fitted on the test images and the mis fits are removed after visual inspection. The reconstructions of the frontal face images are the reference images and the reconstructions of the non-frontal face images are the probes.

The ROCs of the 4 AAMs and the original data are presented in Figure 3. It clearly shows that the performance decreases rapidly with a reduced number of training images and hence a reduced number of model parameters. It was impossible to "come closer" to the characteristic of the original data since it was impossible to train a larger model with the computer and software resources that were available. It should be investigated if adding more training data improves the performance even further.

Remarkable is the fact that ROC of model 1 is better than the ROC of the original data (for a FAR > 22%) for which there is no clear explanation.

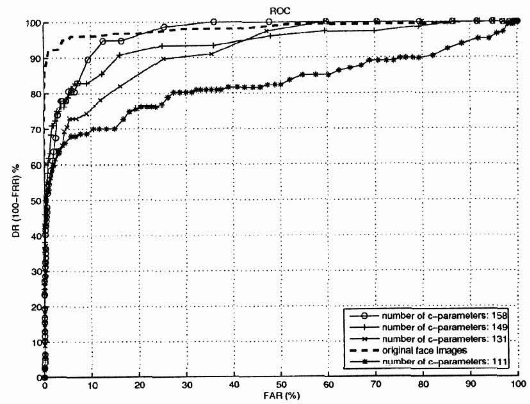

Figure 3. The ROCs of tests with 4 AAMs of different sizes

5.2. Reconstruction under novel poses

In the previous section we showed that it is possible to capture enough variation in the AAM to perform meaningful face recognition in the frontal region. This section will focus on reconstruction under a novel pose in the semi profile region (20 to 70 degrees deviation from the frontal view).

The semi profile AAM and rotation model are trained using the CMU PIE dataset [12]. 170 images of 68 individuals are used to train the models. Each individual was recorded under 22, 45 and 67 degrees. The 67 degrees group contains only 37 individuals since it is impossible to annotate the rest of the group due to disappearing landmarks. These images fall outside the semi profile model and will have to be captured by a profile model.

We construct three probe sets containing:
- the original images;
- the reconstructed images under the original pose and
- the reconstructed images under a novel pose of 22 degrees.

The reference set contained the frontal face images of these individuals. 162 individuals from the IMM and TNO dataset are added to the reference set for the identification tests.

The images in the training set are also used in the testing for lack of more data. Although we realize that independent test and training sets are needed to obtain valid results we choose to perform the tests anyway to get an idea about the possible performance.

Figure 4 shows the original data under two poses, the reconstructed face image under a pose of 67 degrees and the warped reconstruction (under a pose of 22 degrees).

The ROCs and CMCs of the overall performance of ro-

Proceedings of the 7th International Conference on Automatic Face and Gesture Recognition (FGR'06)
0-7695-2503-2/06 $20.00 © 2006 IEEE

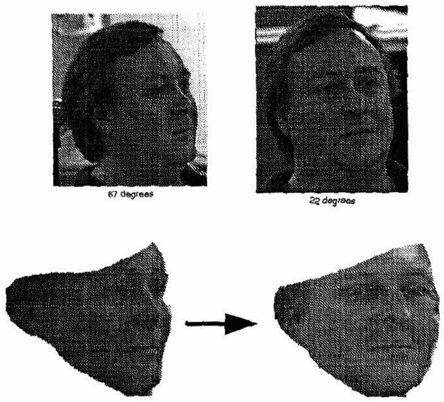

Figure 4. The original face images under 67 and 22 degrees (top). Original reconstruction at 67 degrees and the warped reconstruction to 22 degrees (bottom).

tating the faces in the test set are presented in figure 5. The ROC of the original probes is in fact much lower than shown in Figure 5 since in 28% of the cases the face was not found. These images were therefore not used in the tests. The EER increases with approximately 20% (taking the 28% of missed faces into account).

The correct identification rate at rank one increases with 25%. We also evaluate the performance of the three angles seperately. The CMC for 45 and 67 degrees are given in Figure 6. The CMC at 22 degrees of the warped images shows a decrease in the performance (between 20% and 2%) with respect to the original data. In this case the detail lost in the reconstruction is higher than the benefit gained from the warping. This seems logical because we warp the images to an angle that is close to the original view. The CMCs of the other two poses show significant increases in performance. Warping face images from 45 degrees to 22 degrees improves the performance with 60% (from 15% to 75%).

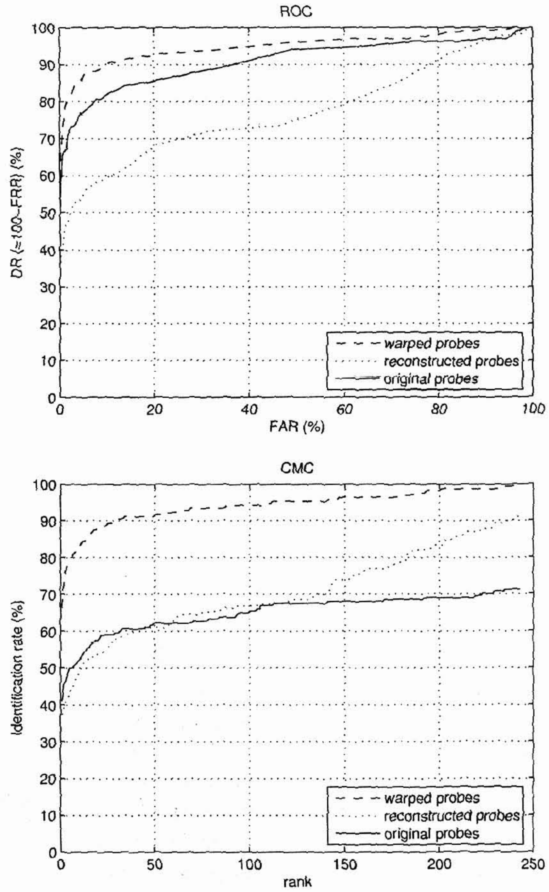

Figure 5. ROC (top) and CMC (bottom) of tests with a polluted reference set.

6. Discussion

We showed that it is possible to capture a large part of the variation that defines the identity in AAMs needed for face recognition. Although it was not possible to reach the performance obtained with the original data we showed that we can capture enough detail to gain more from the rotation model than that we loose detail by using a limited model.

The results of warping face images show a large improvement in the performance and the suggested method makes face recognition under pose more robust for faces under large pose deviations from frontal (45 and 67 degrees). The performance can be increased by 25% for identification at rank one for faces under poses ranging from 22 to 67 degrees. For faces under 45 degrees the identification rate increases by 60% (from 15% to 75%) at rank one. These results were obtained using the same set for testing and training which means that the exact increase in performance has to be investigated in future research. The results however are very encouraging.

Blanz and Vetter [2] used their method (3DMM) in the face recognition vendor test 2002 (FRVT2002) [11]. They obtained an improvement of up to 40% in the verification at a fixed FAR of 1% for reconstructions from 45 degrees to frontal. The results for Cognitec were similar to that obtained in this research. The manual interaction and the slow procedure are however large disadvantages of using 3DMMs.

It was possible to locate the face in all but one of the

0-7695-2503-2/06 $20.00 © 2006 IEEE
IEEE COMPUTER SOCIETY

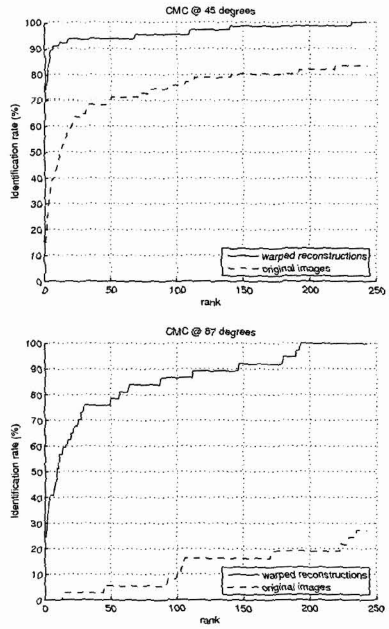

Figure 6. CMCs at 45 (top) and 67 (bottom) degrees.

warped reconstructions. The method is in essence automatic since no manual interaction is needed. Due to the reasonably large images of 140 by 140 pixels the method becomes somewhat sluggish. It takes approximately half a minute to find a correct fit and another few seconds to reconstruct the face image (in Matlab) in the current implementation. A significant gain in speed is expected by implementing the entire procedure more efficiently in for instance C++.

Our research focuses on face recognition in surveillance. The suggested method makes 2D face recognition more robust but we experienced quite some problems with the fitting of the AAM when the test and training set contained images from different data sets. Great care should therefore be taken in the training of the AAM.

7. Future Work

AAMs need to be more robust. Improvements can be made to the initialization and fitting of the AAMs to novel face images. Also the speed of the implementation can be improved by making it more efficient.

More insight has to be gained on how different variations manifest themselves in generic models. Is it possible to train one generic model capturing all the variation in the test set (without an exploding number of model parameters) or are other strategies needed (e.g. a set of generic models)?

The actual improvement using pose correction has to be determined with independent test and training sets.

Our main objective is to make face recognition more robust against different variations. This means we will also look at other possible methods that can be used as a preprocessing step and operate real-time in order to make 2D face recognition in surveillance settings more successful.

References

[1] V. Blanz and T. Vetter. A morphable model for the synthesis of 3d faces. *Conf. Proc. Computer Graphics SIGGRAPH '99*, 1999.

[2] V. Blanz and T. Vetter. Face recognition based on fitting a 3d morphable model. *IEEE trans. On Pattern Analysis and Machine Intelligence*, 25(9), September 2003.

[3] F. Bookstein. Principal warps: Thin-plate splines and the decomposition of deformations. *IEEE trans. On Pattern Analysis and Machine Intelligence*, 11(6), June 1989.

[4] C. Christoudias, L. Morency, and T. Darrell. Light field appearance manifolds. *Proc. of the European Conference on Computer Vision*, (4):481–493, 2004.

[5] Cognitec. Technical specification facevacs-sdk version 2.4. may 2004.

[6] T. Cootes, G. Edwards, and C. Taylor. Active appearance models. *IEEE Trans. on Pattern Analysis and Machine Intelligince*, 23(6):681–685, june 2001.

[7] T. Cootes, G. Wheeler, K. Walker, and C. Taylor. View-based active appearance models. *Image and Vision Computing 20*, pages 657–664, 2002.

[8] S. Gong, A. Psarrou, and S. Romdhani. Corresponding dynamic appearances. *Image and Vision Computing*, 20:307–318, 2002.

[9] R. Gross, I. Matthews, and S. Baker. Eigen light-fields and face recognition across pose. *International Conf. On Automatic Face and Gesture Recognition*, 2002.

[10] M. M. Nordstrøm, M. Larsen, J. Sierakowski, and M. B. Stegmann. The IMM face database - an annotated dataset of 240 face images. Technical report, Informatics and Mathematical Modelling, Technical University of Denmark, DTU, Richard Petersens Plads, Building 321, DK-2800 Kgs. Lyngby, may 2004.

[11] P. Philips, P. Grother, R. Micheals, D. Blackburn, E. Tabassie, and M. Bone. Face recognition vendor test 2002 evaluation report. National Institute of Standards and Technology (www.frvt.org), 2003.

[12] T. Sim, S. Baker, and M. Bsat. The cmu pose, illumination, and expression database. *IEEE Trans. on Pattern Analysis and Machine Intelligence*, 25(12):1615 – 1618, December 2003.

[13] M. B. Stegmann, B. K. Ersbøll, and R. Larsen. FAME - a flexible appearance modelling environment. *IEEE Trans. on Medical Imaging*, 22(10):1319–1331, 2003.

[14] W. Zhao, R. Chellappa, P. Phillips, and A. Rosenfeld. Face recognition: A literature survey. *ACM Computing Surveys*, 35(4):399–458, December 2003.

0-7695-2503-2/06 $20.00 © 2006 IEEE
IEEE COMPUTER SOCIETY