Face Recognition with Support Vector Machines and 3D Head Models

Jennifer Huang[1], Volker Blanz[2], and Bernd Heisele[3]

[1] Center for Biological and Computer Learning, M.I.T, Cambridge, MA, USA
[2] Computer Graphics Research Group, University of Freiburg, Freiburg, Germany
[3] Honda R&D Americas, Inc., Boston, MA, USA
j_huang@mit.edu volker@informatik.uni-freiburg.de heisele@ai.mit.edu

Abstract. We present a novel approach to view and pose invariant face recognition that combines two recent advances in the computer vision field: component-based recognition and 3D morphable models. In a first step a 3D morphable model is used to generate 3D face models from only two input images from each person in the training database. By rendering the 3D models under varying pose and illumination conditions we then create a vast number of synthetic face images which are used to train a component-based face recognition system. In preliminary experiments we show the potential of our approach regarding pose and illumination invariance.

1 Introduction

As real-world applications for face recognition systems continue to increase, the need for an accurate, easily trainable recognition system becomes more pressing. Current systems (for a survey see e.g. [3]) have advanced to be fairly accurate in recognition under constrained scenarios, but extrinsic imaging parameters such as pose, illumination, and facial expression still cause much difficulty in correct recognition.

Recently, component-based approaches have shown promising results in various object detection and recognition tasks such as face detection [8, 5], person detection [6], and face recognition [2, 9, 7, 4]. In [4] we proposed an SVM based recognitions system which decomposes the face into a set of components that are interconnected by a flexible geometrical model. Changes in the head pose mainly lead to changes in the position of the facial components which could be accounted for by the flexibility of the geometrical model. In our experiments, we have shown that the component-based system consistently outperformed whole face recognition systems in which classification was based on the whole face pattern. A major drawback of the system was the need of a large number of training images taken from different viewpoints and under different lighting conditions.

In this paper we further develop this system by adding a 3D morphable face model to the training stage of the classifier. Based on only two images of a person's face and the morphable model we can compute a 3D face model using an analysis by synthesis method [1]. Once the 3D face models of all the subjects in

the training database are computed we generate arbitrary synthetic face images under varying pose and illumination to train the component-based recognition system.

The outline of the paper is as follows: Section 2 explains the generation of 3D head models and synthetic images from two input images. Section 3 describes the training of the component-based face detector from the synthetic images. Section 4 incorporates the synthetic images and face detector from the previous sections into a face recognition system. Section 5 briefly presents preliminary experimental results. Finally, Section 6 summarizes results and outlines future work.

2 Generation of 3D Face Models

The first step in the process of training a component-based recognizer is the generation of 3D face models based on two images of each person in the training database. Examples of the pairs of training images are shown in the top row of Figure 1. The bottom row shows the corresponding synthetic images created by rendering the 3D face models.

In the following we give a brief overview of the morphable model approach, a detailed description can be found in [1]. The main idea behind the morphable model approach is that given a sufficiently large database of 3D face models any arbitrary face can be generated by morphing the ones in the database. An initial database of 3D models was built by recording the faces of 200 subjects with a 3D laser scanner. Then 3D correspondences between the head models were established in a semi-automatic way using techniques derived from optical flow computation. Using these correspondences, a new 3D face model can be generated by morphing the existing models in the database. To create a 3D face model from a set of 2D face images, an analysis by synthesis loop is used to find the morphing parameters such that the rendered images of the 3D model are as close as possible to the input images.

Using the 3D models, synthetic images such as the ones in Figure 2 can easily be created by rendering the models. The 3D morphable model also provides the full 3D correspondence information which allows for automatic extraction of facial components. Prior to using 3D face models, countless images under different pose and illumination conditions had to be recorded for each subject and the components had be extracted in time consuming computations [4].

3 Component based Face Detection

The component-based detector performs two tasks: the detection of the face in a given input image and the extraction the facial components which are later needed to recognize the face. In the following we describe how we trained the component-based face detection system using synthetic face images.

Fig. 1. The upper row consists of the two pictures per person used to generate a 3D model. The lower row consists of the synthetic pictures generated from the model. Notice the similarity between the original and synthetic images.
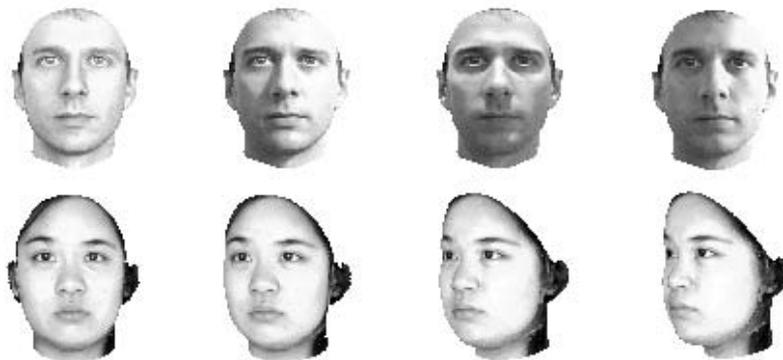

Fig. 2. Synthetic face images generated from the 3D head models under different illuminations (top row) and different poses (bottom row). Synthetic images are used for training the face detection and recognition system.

3.1 Training Set

Approximately 7700 synthetic faces were generated at a resolution of 58×58 for the 6 subjects by rendering the 3D face models under varying pose and illumination. Specifically, the faces were rotated in depth from $0°$ to $34°$ in $2°$ increments. We rendered the faces under for two illumination models. One model consisted of ambient light alone, while the other model was composed of directed light in addition to ambient light, both at equal intensities. The directed light was pointed at the center of the face and positioned between $-90°$ and $+90°$ in azimuth and $0°$ and $75°$ in elevation. The angular position of directed light was incremented by $15°$ in both directions. To build the negative training set we randomly extracted 13655 patterns of size 58×58 from a database of non-face images.

3.2 Extraction of Components

Fourteen components were extracted from every face image based on the correspondence information given by the morphable model. The shape of the components was learned by an algorithm described in [5] to achieve optimal detection results. Figure 3 shows examples of the fourteen components. The components included the left eyebrow, right eyebrow, left eye, right eye, area between eyebrows, bridge of nose, right lip, left lip, right cheek, left cheek, center of mouth, entire mouth, right side of nose, and left side of the nose. Negative training images for the component classifiers were extracted from the 58×58 non-face images.
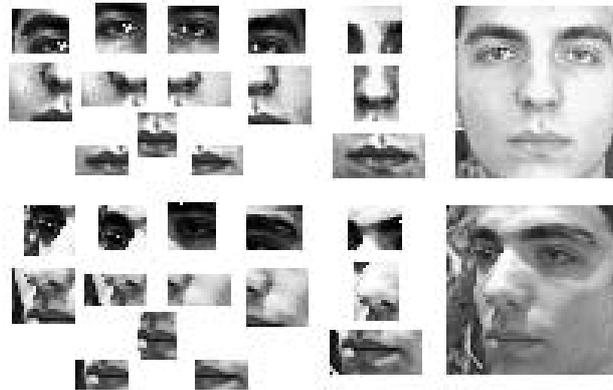

Fig. 3. Examples of the fourteen components extracted from a frontal view and half profile view of a face.

3.3 Architecture of the Face Detector

We used the two level component-based face detection system described in [4]. The architecture of the system is schematically shown in Figure 4. The first level consists of fourteen independent component classifiers (linear SVMs). Each component classifier was trained on the above described set of extracted facial components and on a set of randomly selected non-face patterns. On the second level, the maximum continuous outputs of the component classifiers within rectangular search regions around the expected positions of the components were used as inputs to a geometrical classifier (linear SVM) which performed the final detection of the face. The rectangular search regions were determined from statistical information about the location of the components in the training images.
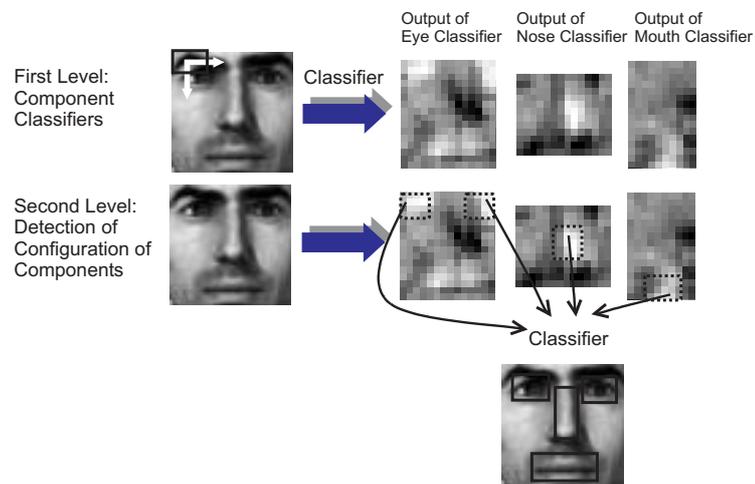

Fig. 4. System overview of component-based face detector. On the first level, windows (lined boxes) of component size are shifted over the face image and classified by the component classifiers. On the second level, the maximum outputs of the component classifiers within the predefined search regions (dotted boxes) and the positions of the detected components are fed to the geometrical classifier.

4 Component based Face Recognition

From the fourteen components extracted by the face detector we used only ten components for face recognition. The four components that were eliminated either strongly overlapped with other components or contained few grey value structures (e.g. cheeks). The face detection system was applied to each synthetic training to detect the facial region and extract the components. Figure 5 shows the composite of the ten extracted components for some example images. For

each face, the pixel values of the extracted components were combined into a single feature vector. A face recognition system consisting of SVM classifiers was trained on these feature vectors in a one vs. all strategy. In other words, an SVM was trained for each subject in the database to separate her/him from all the other subjects. To determine the identity of a person at runtime, we compared the normalized outputs of the SVM classifiers, i.e. the distances to the hyperplanes. The identity associated with the face classifier with the highest normalized output was taken to be the identity of the face.

Fig. 5. Composite of the ten face components used for face recognition.

5 Experimental Results

The component based face recognition system was compared to a whole face recognition system, both systems were trained and tested on the same images. In contrast to the component based classifiers, the input vector to the whole face detector and recognizer consisted of the pixel values from the entire 58×58 facial region. For a more detailed description of the whole face system see [4].

The whole face and the component based face detectors were trained with linear SVMs. For the recognition systems, we trained both linear and 2nd degree polynomial SVMs. This resulted in four different face classification systems which were compared on two test sets. In the following, the classifiers will be referred to as whole linear, whole polynomial, component linear, and component polynomial, respectively.

The two test sets of novel images were rendered from the 3D models described in Section 2. The first test set consisted of synthetic faces of the six subjects in the database rotated between $-36°$ and $+36°$ in $6°$ steps. The point light source was positioned in elevation between $-22.5°$ and $67.5°$ in $30°$ steps and in azimuth from $-112.5°$ to $97.5°$ in $30°$ steps. The second test set had the same parameters as the first test set except that the faces were additionally rotated in the image plane by $4°$. The former test set will be referred to as regular test set and the latter as rotated test set.

The resulting ROC curves for the four classifiers can be seen in Figure 6. For the whole face system the accuracy of the polynomial classifier exceeds that of its linear counterpart while for the component based system the polynomial and linear classifiers are roughly equal. On both test sets the component based system clearly outperforms the whole face system. The large discrepancy on the the rotated test is explained by the sensitivity of the whole face recognition system to rotations.

6   Conclusion and Future Work

This paper presented a new development in component based face recognition by incorporation a 3D morphable model into the training process. Based on two face images of a person and a 3D morphable model we computed the 3D face model of each person in the database. By rendering the 3D models under varying poses and lighting conditions we automatically generated a large number of synthetic face images to train the component based recognition system. Preliminary results on synthetic test images show that the component based recognition system clearly outperforms a comparable whole face recognition system. We achieved component based recognition rates around 98% for faces rotated up to $\pm36°$ in depth.

Future work includes the application of the system to a test set of real face images and extending the pose invariance by training on synthetic images over a larger range of views.
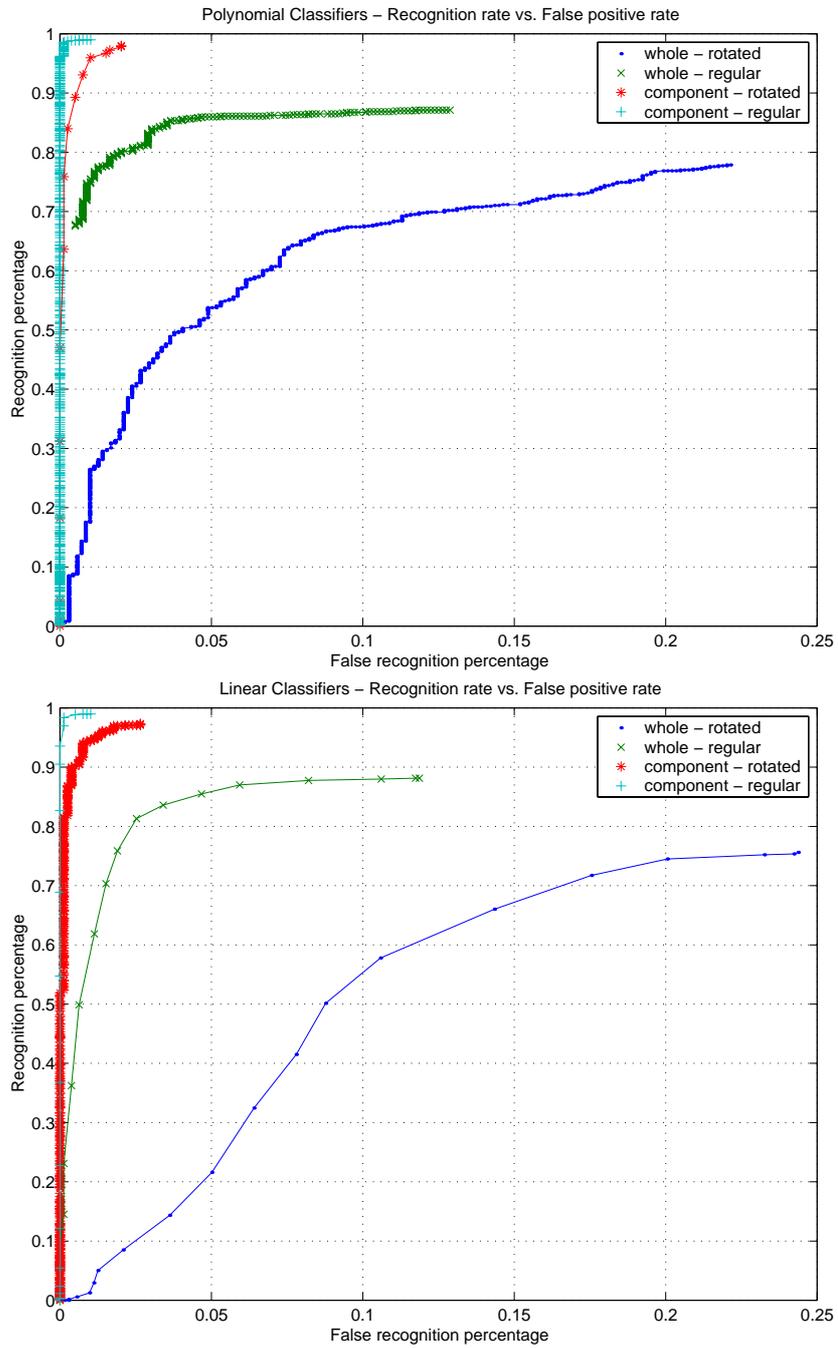
Fig. 6. The left diagram shows the ROC curve of the linear and polynomial component based face recognition system on two different test sets. The right diagram shows the ROC curves of the linear and polynomial whole face recognition system.

References

1. V. Blanz and T. Vetter. A morphable model for synthesis of 3D faces. In *Computer Graphics Proceedings SIGGRAPH*, pages 187–194, Los Angeles, 1999.
2. R. Brunelli and T. Poggio. Face recognition: Features versus templates. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 15(10):1042–1052, 1993.
3. R. Chellapa, C. Wilson, and S. Sirohey. Human and machine recognition of faces: a survey. *Proceedings of the IEEE*, 83(5):705–741, 1995.
4. B. Heisele, P. Ho, and T. Poggio. Face recognition with support vector machines: global versus component-based approach. In *Proc. 8th International Conference on Computer Vision*, volume 2, pages 688–694, Vancouver, 2001.
5. B. Heisele, T. Serre, M. Pontil, and T. Poggio. Component-based face detection. In *Proc. IEEE Conference on Computer Vision and Pattern Recognition*, volume 1, pages 657–662, Hawaii, 2001.
6. A. Mohan, C. Papageorgiou, and T. Poggio. Example-based object detection in images by components. In *IEEE Transactions on Pattern Analysis and Machine Intelligence*, volume 23, pages 349–361, April 2001.
7. A.V Nefian and M.H Hayes. An embedded HMM-based approach for face detection and recognition. In *Proc. IEEE International Conference on Acoustics, Speech, and Signal Processing*, volume 6, pages 3553–3556, 1999.
8. H. Schneiderman and T. Kanade. A statistical method for 3d object detection applied to faces and cars. In *Proc. IEEE Conference on Computer Vision and Pattern Recognition*, pages 746–751, 2000.
9. L. Wiskott. *Labeled Graphs and Dynamic Link Matching for Face Recognition and Scene Analysis*. PhD thesis, Ruhr-Universität Bochum, Bochum, Germany, 1995.