Component-based Face Recognition with 3D Morphable Models

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

[1] Center for Biological and Computational Learning, M.I.T., Cambridge, MA, USA
[2] Honda Research Institute US, Boston, MA, USA
[3] Computer Graphics Group, Max-Planck-Institut, Saarbrücken, Germany
jenniferhuang@alum.mit.edu heisele@ai.mit.edu blanz@mpi-sb.mpg.de

Abstract. We present a novel approach to pose and illumination invariant face recognition that combines two recent advances in the computer vision field: component-based recognition and 3D morphable models. First, a 3D morphable model is used to generate 3D face models from three input images from each person in the training database. The 3D models are rendered under varying pose and illumination conditions to build a large set of synthetic images. These images are then used to train a component-based face recognition system. The resulting system achieved 90% accuracy on a database of 1200 real images of six people and significantly outperformed a comparable global face recognition system. The results show the potential of the combination of morphable models and component-based recognition towards pose and illumination invariant face recognition based on only three training images of each subject.

1 Introduction

The need for a robust, accurate, and easily trainable face recognition system becomes more pressing as real world applications such as biometrics, law enforcement, and surveillance continue to develop. However, extrinsic imaging parameters such as pose, illumination and facial expression still cause much difficulty in accurate recognition. Recently, component-based approaches have shown promising results in various object detection and recognition tasks such as face detection [7, 4], person detection [5], and face recognition [2, 8, 6, 3].

In [3], we proposed a Support Vector Machine (SVM) based recognition 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, the component-based system consistently outperformed global 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. These images are often unavailable in real-world applications.

In this paper, the system is further developed through the addition of a 3D morphable face model to the training stage of the classifier. Based on only three images of a person's face, the morphable model allows the computation of 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 briefly explains the generation of 3D head models. Section 3 describes the component-based face detector trained from the synthetic images. Section 4 describes the component-based face recognizer, which was trained from the output of the component-based face detection unit. Section 5 presents the experiments on component-based and global face recognition. Finally, Section 6 summarizes results and outlines future work.

2 Generation of 3D Face Models

We first generate 3D face models based on three training images of each person (frontal, half-profile and profile view). An example of the triplet of training images used in our experiments is shown in the top row of Figure 1. The bottom row shows two synthetic images created by rendering the newly generated 3D face model.

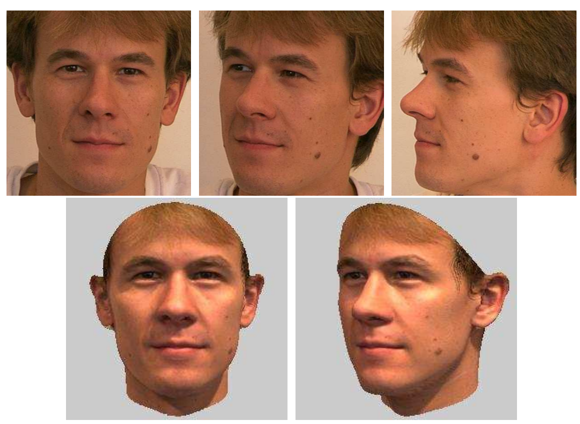

Fig. 1. Generation of the 3D model. The top images were used to compute the 3D model. The bottom images are synthetic images generated from the 3D model.

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 fully automatic way using techniques derived from optical flow computation. Based on these correspondences, a new 3D face model can be generated by morphing between 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. A detailed description of the morphable model approach including the analysis by synthesis algorithm can be found in [1].

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 between the head models, which allows for automatic extraction of facial components.

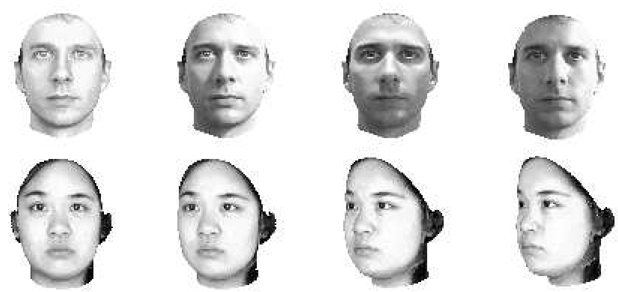

Fig. 2. Synthetic training images. Synthetic face images were generated from the 3D head models under different illuminations (top row) and different poses (bottom row).

3 Component-based Face Detection

The component-based detector detects the face in a given input image and extracts the facial components which are later used to recognize the face. We used the two level component-based face detection system described in [3]. 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 a set of extracted facial components[4] and on a set of randomly selected non-face patterns. As mentioned in the previous Section, the components could be automatically extracted from the synthetic images since the full 3D correspondences between the face models were known. Figure 3 shows examples of the fourteen components for three training images. On the second level, the maximum continuous outputs of the component classifiers within rectangular search regions around the expected positions of the components were

[4] The shape of the components was learned by an algorithm described in [4] to achieve optimal detection results.

used as inputs to a combination classifier (linear SVM), which performed the final detection of the face.

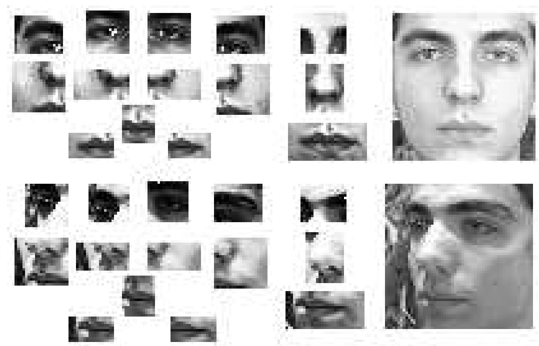

Fig. 3. Examples of the fourteen components extracted from a frontal view and half-profile view of a face.

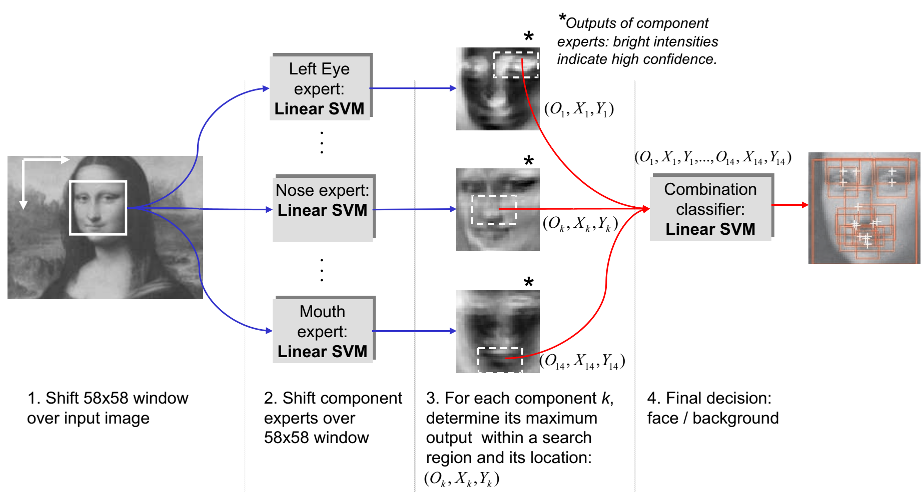

Fig. 4. System overview of the component-based face detector.

4 Component-based Face Recognition

The component-based face recognizer uses the output of the face detector in the form of extracted components. First, synthetic faces were generated at a resolution of 58×58 for the six 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 and rendered with two illumination models at each pose. The first model consisted of ambient light alone. The second model included ambient light and a directed light source, which 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.

From the fourteen components extracted by the face detector, only nine components were used for face recognition. Five components were eliminated because they strongly overlapped with other components or contained few gray value structure (e.g. cheeks). Figure 5 shows the composite of the nine extracted components used for face recognition for some example images. In addition, a histogram-equalized inner face region was added to improve recognition[5]. Some examples of this inner face region are shown in Figure 6.

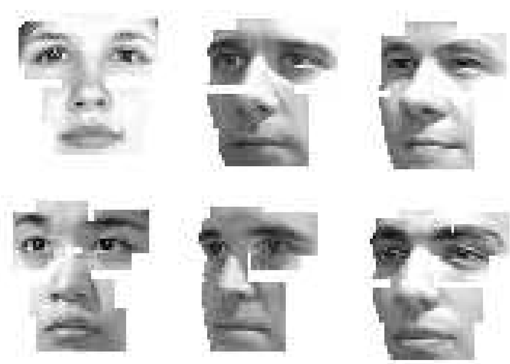

Fig. 5. Composite of the nine components retained for face recognition.

The component-based face detector was applied to each synthetic face image in the training set to detect the components and thereby the facial region. Histogram equalization was then preformed on the bounding box around the components. The gray pixel values of each component were then taken from the histogram equalized image and combined into a single feature vector. Feature vectors were constructed for each person, and corresponding classifiers were trained.

[5] The location of the face region was computed by taking the bounding box around the other nine detected components and then subtracting from the larger edge to form a square component. This square was then normalized to 40×40 and histogram-equalized.

Fig. 6. Histogram-equalized inner face region, added as a tenth component for face recognition.

A face recognition system consisting of second-degree polynomial SVM classifiers was trained on these feature vectors in a one vs. all approach. 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 in the feature space. The identity associated with the face classifier with the highest normalized output was taken to be the identity of the face.

5 Results

A test set was created by taking images of the six people in the database. The subjects were asked to rotate their faces in depth and the lighting conditions were changed by moving a light source around the subject. The test set consisted of 200 images of each person under various pose and illumination conditions. Figure 7 contains examples of the images in the test set[6].

Fig. 7. Examples of the real test set. Note the variety of poses and illumination conditions.

The component-based face recognition system was compared to a global face recognition system; both systems were trained and tested on the same images.

[6] Training and test set will be made available on our website upon publication.

In contrast to the component-based classifiers, the input vector to the whole face detector and recognizer consisted of the histogram-equalized gray values from the entire 58×58 facial region[7]. The resulting ROC curves of global and component-based recognition on the test set can be seen in Figure 8. The component-based system achieved a recognition of 90%, which is approximately 50% above the recognition rate of the global system.

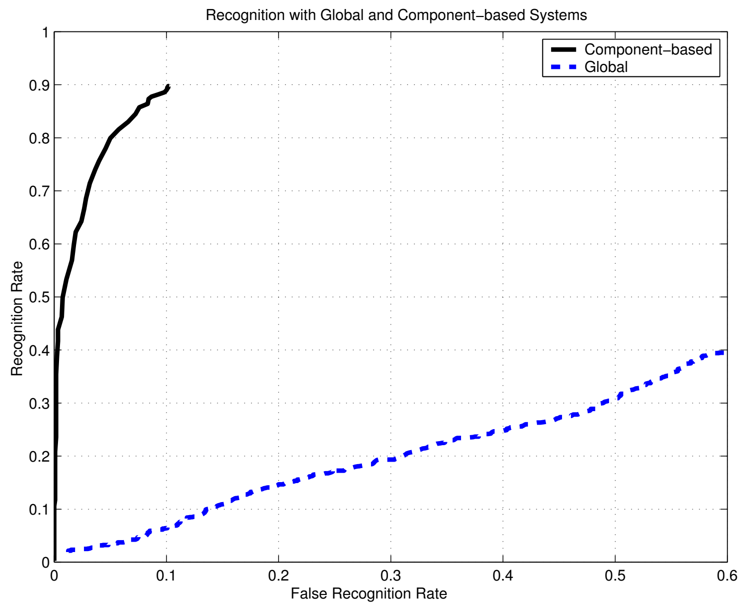

Fig. 8. ROC curves for the component-based and the global face recognition systems. Both systems were trained and tested on the same data.

This large discrepancy in results can be attributed to two main factors: First, the components of a face vary less under rotation than the whole face pattern, explaining why the component-based recognition is more robust against pose changes. Second, in contrast to the training data, the backgrounds in the test images were non-uniform. Component-based recognition only used face parts as input features for the classifier while the input features to the global system occasionally contained distracting background parts.

6 Conclusion and Future Work

This paper presented a new development in component-based face recognition by the incorporation of a 3D morphable model into the training process. This

[7] For a detailed description of the whole face system see [3].

combination allowed the training of a face recognition system which required only three face images of each person. From these three images, 3D face models were computed and then used to render a large number of synthetic images under varying poses and lighting conditions. The synthetic images were then used to train a component-based face detection and recognition system. A global face detection and recognition system was also trained for comparison.

Results on 1200 real images of six subjects show that the component-based recognition system clearly outperforms a comparable global face recognition system. Component-based recognition was at around 90% for faces rotated up to approximately half-profile in depth. Future work includes increasing the size of the database and the range of the pose and illumination conditions which the system can handle.

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. 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.
4. 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.
5. 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.
6. 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.
7. 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.
8. L. Wiskott. *Labeled Graphs and Dynamic Link Matching for Face Recognition and Scene Analysis*. PhD thesis, Ruhr-Universität Bochum, Bochum, Germany, 1995.