Component-based Face Recognition with 3D Morphable Models

B. Weyrauch[*]
benjamin.weyrauch@vitronic.com
Center for Biological and
Computational Learning, M.I.T.
Cambridge, MA

J. Huang[†]
jenniferhuang@alum.mit.edu
Center for Biological and
Computational Learning, M.I.T.
Cambridge, MA

B. Heisele
bheisele@honda-ri.com
Honda Research Institute USA, Inc.
Boston, MA

V. Blanz
blanz@mpi-sb.mpg.de
Max-Planck-Institute for Computer Science
Saarbrücken, Germany

Abstract

We present a system for pose and illumination invariant face recognition that combines two recent advances in the computer vision field: 3D morphable models and component-based recognition. A 3D morphable model is used to compute 3D face models from three input images of each subject 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 for training a component-based face recognition system. The face recognition module is preceded by a fast hierarchical face detector resulting in a system that can detect and identify faces in video images at about 4 Hz. The system achieved a recognition rate of 88% on a database of 2000 real images of ten people, which is significantly better than a comparable global face recognition system. The results clearly show the potential of the combination of morphable models and component-based recognition towards pose and illumination invariant face recognition.

1. Introduction

The need for a robust, accurate, and easily trainable face recognition system becomes more pressing as real world applications in the areas of law enforcement, surveillance, access control, and human machine interfaces 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 [8, 3, 9], person detection [7], and face recognition [4, 10]. In [4], 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. A system for view and illumination invariant face recognition from single video images has recently been proposed in [1]. By fitting a 3D morphable model to each face image, the system avoids the problems of pose and illumination variations inherent to view-based classification techniques. Matching takes place in the space of morphing parameters, which is invariant to illumination and viewpoint. Disadvantages of the system are its computation time, which is in the order of minutes per image, and the need of a manual initialization of the pose of the 3D morphable model.

In this paper, we combine morphable models and component-based recognition. The morphable model is employed during training only, where slow speed and manual interaction is not as problematic as during classification. Based on three images of a person's face, the morphable model computes a 3D face model using an analysis by synthesis method [2]. Once the 3D face models of all the subjects in the training database are computed, we generate a large number of synthetic face images under varying pose and illumination to train the component-based recognition system. The face recognition module is preceded by

[*] with Vitronic GmbH since 08/03.

[†] with Boston Consulting Group since 10/03.

a hierarchical face detection system similar to the one described in [5], which performs a rough localization of the face in the image. Following the hierarchical system is a component-based face detector [6], which precisely localizes the face and extracts the components for face recognition. The overview of the system is shown in Figure 1.

The outline of the paper is as follows: Section 2 briefly explains the generation of 3D head models. Section 3 is about the hierarchical and component-based face detectors. 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. Examples of the image triplets are shown in Figure 2. Each triplet consisted of a frontal, a half-profile, and a profile high resolution face image.

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. 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 more detailed description of the morphable model approach including the analysis by synthesis algorithm can be found in [2]. The original frontal face images of all ten subjects and the corresponding synthetic images are shown in Figure 3.

3. Face Detection

As shown in Figure 1, the detection of the face is split into two modules. The first module is a fast face detector similar to the one described in [5], consisting of a hierarchy of SVM classifiers which were trained on faces at different resolutions. Low resolution classifiers remove large parts of the background on the bottom of the hierarchy, the most accurate and slowest classifier performs the final detection on the top level. In our experiments we used the following hierarchy of SVM classifiers: 3×3 linear, 11×11 linear, 17×17 linear, and 17×17 second-degree polynomial[1].

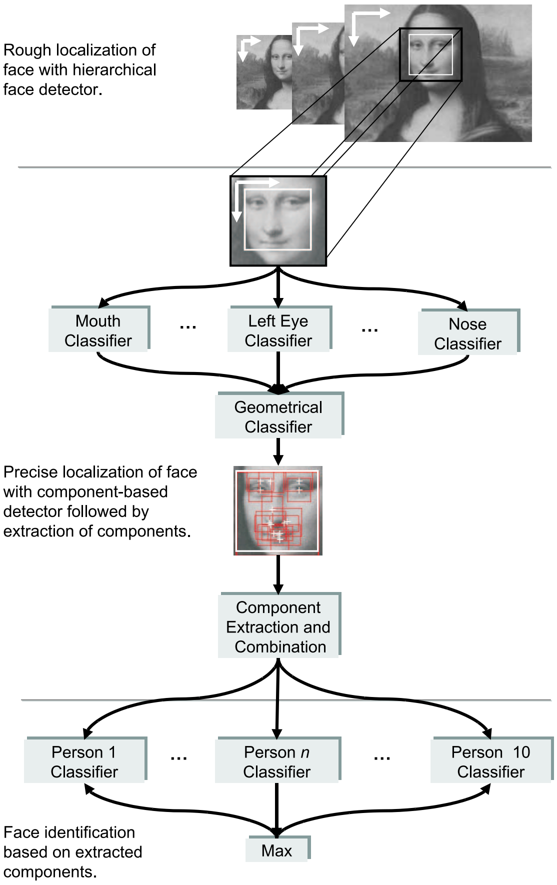
Rough localization of face with hierarchical face detector.
Mouth Classifier
...
Left Eye Classifier
...
Nose Classifier
Geometrical Classifier
Precise localization of face with component-based detector followed by extraction of components.
Component Extraction and Combination
Person 1 Classifier
...
Person *n* Classifier
...
Person 10 Classifier
Face identification based on extracted components.
Max

Figure 1: System overview of the component-based face recognition system. The face is roughly located in the image by a hierarchical face detector. A fine localization of the face and its components is performed with a component-based face detector. The final step is the classification of the face based on the extracted facial components.

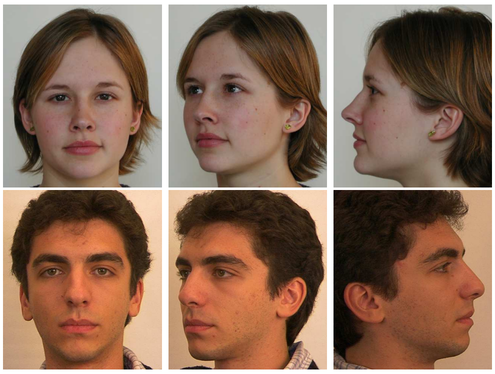

Figure 2: Examples of the image triplets used for generating the 3D models.

[1] Where 17×17 means that the classifier has been trained on face images of size 17×17 pixels.

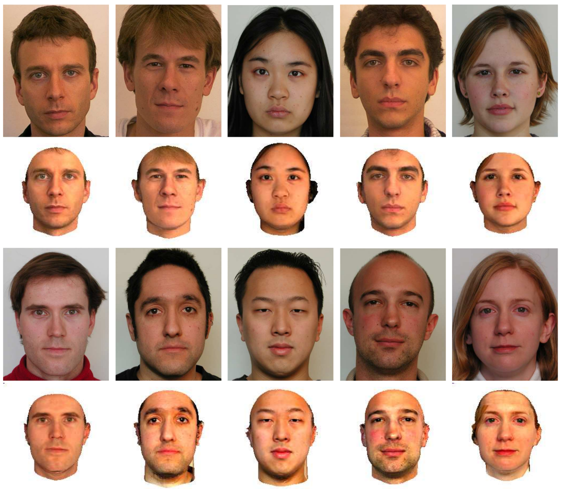

Figure 3: Original images and synthetic images generated from 3D models for all ten subjects in the training database.

The positive training data for all classifiers was generated from 3D head models, with a pose range of ±45° rotation in depth and ±10° rotation in the image plane. The negative training set initially consisted of randomly selected non-face patterns which was enlarged by difficult patterns in several bootstrapping iterations.

Once the face is roughly localized by the hierarchical detector we run the component-based face detector on image part which is slightly bigger than the detection box computed by the hierarchical classifier. The component-based classifier performs a fine search on the given part of the image, detects the face and extracts the facial components. We used the two level component-based face detection system described in [6]. The architecture of the system is schematically shown in Figure 1. The first level consists of 14 independent component classifiers (linear SVMs). Each component classifier was trained on a set of extracted facial components and on a set of randomly selected non-face patterns. The components could be automatically extracted from the synthetic images since the full 3D correspondences between the face models were known. Figure 4 shows examples of the 14 components for two 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 used as inputs to a geometrical classifier (linear SVM), which performed the final detection of the face.

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 ten subjects by rendering the 3D face models under

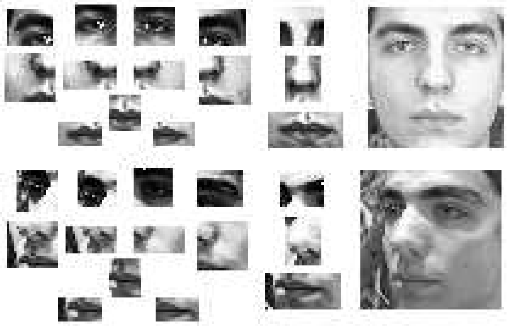

Figure 4: Examples of the 14 components extracted from a frontal view and half profile view of a face.

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. Some example images from the training set are shown in Figure 5.

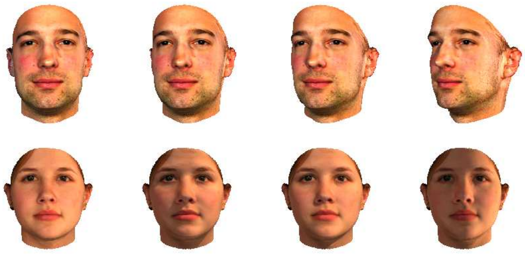

Figure 5: Pose and illumination.

From the originally 14 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). In addition, a global component was added to improve recognition. The location of this component was computed by taking the circumscribing square around the bounding box of the other nine components. After extraction, the squared image patch was normalized to 40×40 pixels. The component-based face detector was applied to each synthetic face image in the training set to extract the ten components. Histogram equalization was then preformed on each component individually, Figure 6 shows the histogram-equalized components for an image from the training data. The gray pixel values of each component were then combined into a single feature vector. 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 continuous outputs of the SVM classifiers. The identity associated with the face classifier with the highest output value was taken to be the identity of the face.

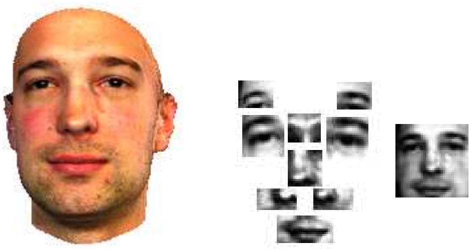

Figure 6: The ten components used for face recognition.

5. Results

A test set was created by taking images of the ten people in the database with a digital video camera. 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 final test set consisted of 200 images of each person under various pose and illumination conditions. Figure 7 shows examples of the images in the test set[2].

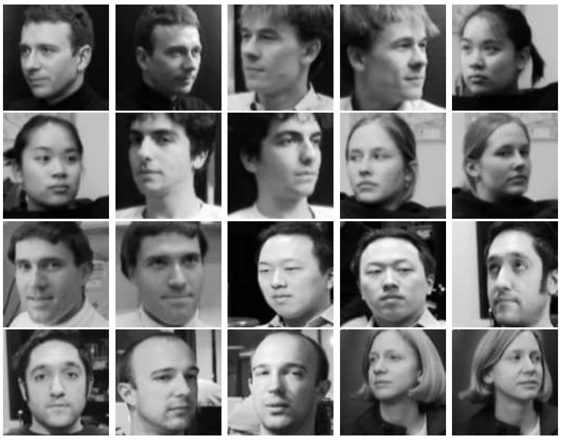

Figure 7: Example images from the real test set. Note the variety of poses and illumination conditions.

The component-based face recognition system was compared to a global face recognition system–both systems were trained and tested on the same images. In contrast to the component-based classifiers, the input vector to the whole face recognizer consisted of the histogram equalized gray values from the entire 40×40 facial region as extracted by the hierarchical face detector. The resulting ROC curves for global and component-based recognition can be seen in Figure 8. Each point on the ROC curve corresponds to a different rejection threshold. A test image was rejected if the maximum output of the ten SVM classifiers was below the given rejection threshold. The rejection threshold is largest at the starting point of an ROC curve, i.e. the recognition and false positive (FP) rates are zero. At the endpoint of an ROC curve the rejection rate is zero, recognition rate and FP rate sum up to 100%. The component-based system achieved a maximal recognition rate of 88%, which is approximately 20% above the recognition rate of the global system. This significant 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, which explains why the component-based recognition is more robust against pose changes. Second, performing histogram equalization on the individual components reduces the in-class variations caused by illumination changes.

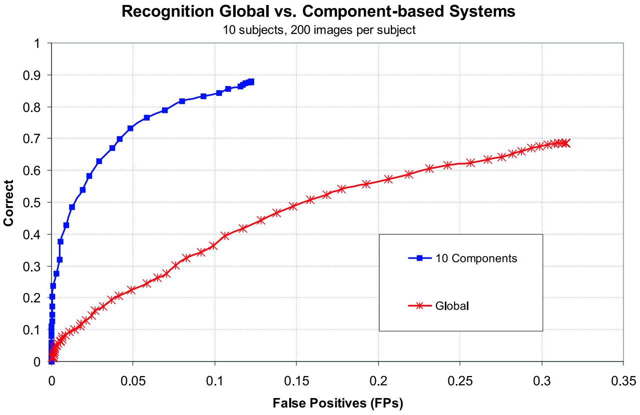

Figure 8: ROC curves for the component-based and the global face recognition system. Both systems were trained and tested on the same data.

The error distribution among the ten subjects was highly unbalanced. While nine out of the ten people could be recognized with about 92% accuracy, the recognition rate for the tenth subject (person on the bottom right in Figure 3) was as low as 49%. This might be explained by an inaccurate 3D head model or by the fact that for this subject training and test data were recorded six months apart from each other. Upon visual inspection of the misclassified faces about 50% of the errors could be attributed to pose, facial

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

expression, illumination, and failures in the component detection stage. Figure 9 shows some of these images. The remaining 50% of the errors could not be explained by visual inspection.

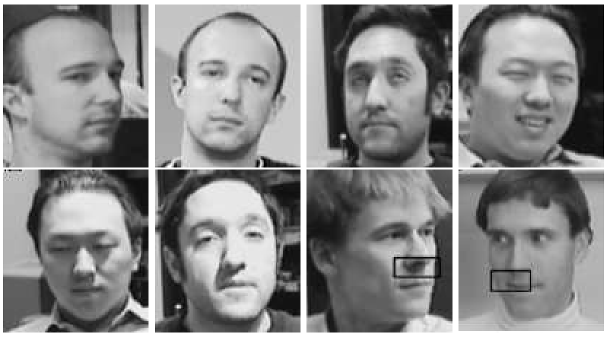

Figure 9: Examples of misclassified faces in the test set. From top left to bottom right the reasons for misclassification are: Pose, expression, illumination, and failure in detecting the mouth component.

The speed measurements of the system were conducted on a different test set of 100 images of size 640×480. Each image included a single face at a resolution between 80×80 to 120×120 pixels. The overall speed was 4 Hz, the hierarchical detector took about 81%, the component detection 10%, and the recognition 6.4% of the overall computation time.

6. Conclusion

This paper presented a new development in component-based face recognition by the incorporation a 3D morphable model into the training process. From only three images per subject, 3D face models were computed and subsequently rendered under varying poses and lighting conditions to build a large number of synthetic images. These synthetic images were then used to train a component-based face recognizer. The face recognition module was combined with a hierarchical face detector, resulting in a system that could detect and identify faces in video images at about 4 Hz. Results on 2000 real images of ten subjects show that the component-based recognition system clearly outperforms a comparable global face recognition system. Component-based recognition was at 88% for faces rotated up to approximately half profile in depth.

Future work includes using a different set of learned components, which is optimized for face recognition, and expanding the morphable model to generate facial expressions.

References

[1] V. Blanz, S. Romdhani, and T. Vetter. Face identification across different poses and illuminations with a 3d morphable model. In *Proc. of the 5th Int. Conference on Automatic Face and Gesture Recognition (FG)*, pages 202–207, 2002.

[2] V. Blanz and T. Vetter. A morphable model for synthesis of 3D faces. In *Computer Graphics Proceedings SIGGRAPH*, pages 187–194, Los Angeles, 1999.

[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, 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, S. Mukherjee, and T. Poggio. Feature reduction and hierarchy of classifiers for fast object detection in video image. In *Proc. IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, volume 2, pages 18–24, Kauai, 2001.

[6] 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.

[7] 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.

[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] S. Ullman and E. Sali. Object classification using a fragment-based representation. In *Biologically Motivated Computer Vision (eds. S.-W. Lee, H. Bulthoff and T. Poggio)*, pages 73–87 (Springer, New York), 2000.

[10] L. Wiskott, J.-M. Fellous, N. Krüger, and C. von der Malsburg. Face recognition by elastic bunch graph matching. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 19(7):775 –779, 1997.