Component-Based Approach to Face Detection

Bernd Heisele, Thomas Serre & Stanley Bileschi

Artificial Intelligence Laboratory and
The Center for Biological and Computational Learning
Massachusetts Institute of Technology
Cambridge, Massachusetts 02139

http://www.ai.mit.edu

The Problem: We want to develop a trainable system for face detection that is able to handle faces rotated in depth and partially occluded in still gray images.

Motivation: Faces form a class of visually similar objects that simplifies the generally difficult task of object detection. However, the problem of pose invariance is still unsolved. There are three basic ideas behind part- or component-based detection of objects. First, some objects classes can be described well by a few characteristic object parts and their geometrical relation. Second, the patterns of some object parts can might less under pose changes than the pattern belonging to the whole object. Third, a component-based approach might be more robust against partial occlusions than a global approach. Fourth, recent psycho-physiological experiments show that building a hierarchy of simple detector would be close to the way the brain works.

Previous Work: Most of the previous work dealt with frontal faces. A technique that uses clustering to generate face and non-face prototypes is presented in [6]. For each test pattern the distances between the pattern and the prototypes are calculated. These distances form the input to a multi-layer perceptron that classifies the pattern into a face and non-face class. In [2] a Support Vector Machine (SVM) with a 2nd degree polynomial kernel is used to classify normalized gray value patterns. A neural network based system which is able to deal with rotations in the image plane is proposed in [3] and [4]. An approach based on the probabilities of the occurrence of small intensity patterns (16×16 pixels) in the image of the whole face (64×64 pixels) is described in [5]. In [1] we propose a component-based technique for detecting frontal and near frontal faces in gray images using SVMs. The system was trained on frontal faces and has been compared with a whole face detection system similar to the one proposed in [2]. With increasing rotation the component-based classifier clearly outperformed the whole face classifier (about 50% less FP at ± 30 deg. rotation in depth and 50% less FP at ± 10 deg. rotation in the image plane).

Approach: The basic idea is to train a set of classifiers on components of a face instead of training one classifier on the whole face image. We propose a component-based, trainable system for detecting frontal and near-frontal views of faces in still gray images. The system consists of a two-level hierarchy of Support Vector Machine (SVM) classifiers. On the first level, component classifiers independently detect components of a face. On the second level, a single classifier checks if the geometrical configuration of the detected components in the image matches a geometrical model of a face. A fundamental problem of the component-based approach is how to determine discriminative components for a given class of objects. We consider two approaches to solve this problem. In [1] we propose a method for automatically learning components by using 3-D head models where we determine components by growing regions around seed points located on 3-D head models. The region growing is controlled by the discriminative power of the components and their robustness against pose changes. This approach has the advantage that no manual interaction is required for choosing and extracting components. Experiments show that the component-based system is significantly more robust against rotations in depth than a comparable system trained on whole face patterns. Current work focuses on extracting a large set of random components (random size and position) and on combining then based on a criterion derived from the SVM theory. We also consider correlation between classifiers when combining them.

Difficulty: The detection system must be robust against changes in the appearance of faces due to rotations in depth and against partial occlusions.

Impact: Face detection is the first step of an autonomous face recognition system. It also has potential application in human-computer interfaces and surveillance systems.

Future Work: The component-based system outperforms the whole face system. Similar results have been obtained comparing a component-based and a whole face approach for face identification. Further work has to be done on the combination of component classifiers. We also plan to apply SVM regression to estimating the pose of a face using the locations of the detected components.

Research Support: Research at CBCL is supported by ONR, Darpa, NSF, Kodak, Siemens, DaimlerChrysler, ATR, ATT, Compaq, Honda, CRIEPI.

References:

[1] B. Heisele, T. Poggio, and M. Pontil. Face detection in still gray images. A.I. memo 1687, Center for Biological and Computational Learning, MIT, Cambridge, MA, 2000.

[2] E. Osuna. *Support Vector Machines: Training and Applications*. PhD thesis, MIT, Department of Electrical Engineering and Computer Science, Cambridge, MA, 1998.

[3] H. A. Rowley, S. Baluja, and T. Kanade. Rotation invariant neural network-based face detection. Computer Scienct Technical Report CMU-CS-97-201, CMU, Pittsburgh, 1997.

[4] H. A. Rowley, S. Baluja, and T. Kanade. Neural network-based face detection. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 20(1):23–38, 1998.

[5] H. Schneiderman and T. Kanade. Probabilistic modeling of local appearance and spatial relationships for object recognition. In *Proc. IEEE Conference on Computer Vision and Pattern Recognition*, pages 45–51, Santa Barbara, 1998.

[6] K.-K. Sung. *Learning and Example Selection for Object and Pattern Recognition*. PhD thesis, MIT, Artificial Intelligence Laboratory and Center for Biological and Computational Learning, Cambridge, MA, 1996.

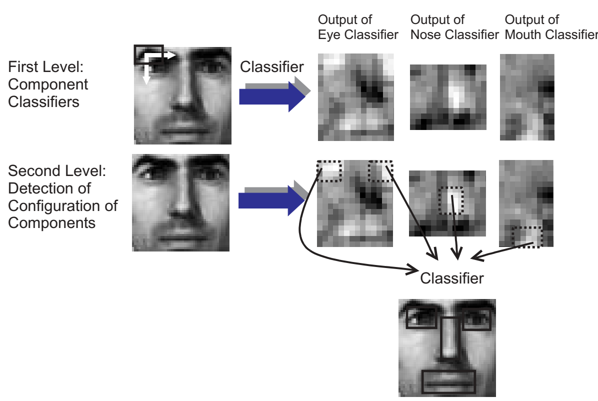

Figure 1: System overview of the component-based classifier. On the first level, windows of the size of the components (solid lined boxes) are shifted over the face image and classified by the component classifiers. On the second level, the maximum outputs of the component classifiers within predefined search regions (dotted lined boxes) are fed into the geometrical configuration classifier.

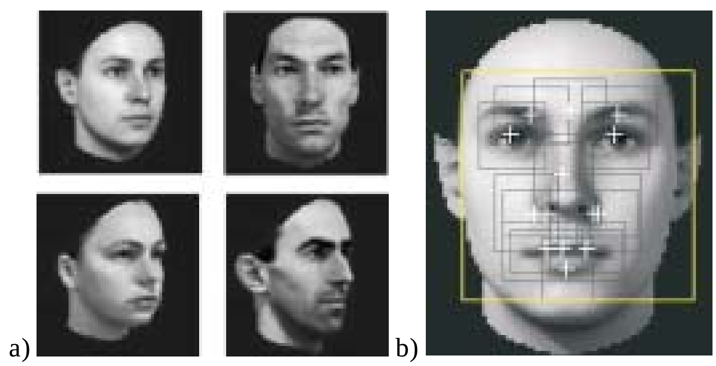

Figure 2: a) Examples of synthetic faces from the training set. b) Components that have been automatically learned from 3-D head models.