On the Role of Structural Information for Face Detection

Thomas Serre & Bernd Heisele

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 wish to compare strategies for combining a given set of face-component experts (eyes-, nose-, mouth-detectors, etc) in order to perform robust face detection.

Motivation: The system [3] uses fourteen component experts to build a forty-height dimensional feature vector by concatenating the positions (X, Y) of their best classification results and their corresponding outputs. We wish to study alternate strategies for combining component experts and in particular quantify the role of prior information, structural information and the confidence in the experts’ classification value.

Previous Work: We previously shown that component-based face detection outperforms whole face detection in [3]. Similar results have been proposed in [2] for face recognition and a comparison between features and templates have been proposed in [4].

Approach: 1. Learning automatically discriminative object-parts: Seven 3-D head models only were sufficient to build a robust face detection system. 3-D head models provided us with a) a set of several thousands of 2-D face images generated with standard computer rendering technics and b) their 3-D voxel correspondences for automatic extraction of face parts. Component-experts were built by simultaneously growing image regions (corresponding to face parts such as eyes, mouth, nose, etc) and training SVM classifiers. At each step winning experts that minimize theoretical bounds on the error probability of the SVM were selected. 2. Combining component-experts: Component-experts were combined in a second stage to yield a hierarchical SVM classifier. We here want to compare several strategies to combine those experts including: i. The choice of the features that are used to feed the hierarchical classifier: The position (X, Y) of the experts’ best detection and the experts’ confidence in the detection. ii. The use of prior information (search for components limited to a region defined by the expected position of the component1 vs. whole image).

The novelties of our approach are: 1) a new COMPONENT-BASED algorithm to learn from examples component-based classification experts and their combination, the use of 3D MORPHABLE MODELS to train the system with synthetic faces and automatically extract components using the 3D voxel correspondences and 3) a MAX operation on the output of each component classifier within a search region – which may be relevant for biological models of visual recognition.

Impact: Our study has application in both computer vision and computational neurosciences. First it has potential application in human-computer interfaces and surveillance systems. Second face representation in primates’ cortex has been the center of much debates in the cognitive community. In particular, quantifying the role of con-

1The search region is defined by the mean and std of face components previously calculated from the training set.
Figure 1: Experimental results in face classification show considerable robustness for rotations in depth and suggest performance at significantly better level than other face detection systems.

figurational face encoding is the first step toward inferring the validity of theories arguing that faces are special (see [1]).

Figure 2: An overview of the system (reproduced from [3]).

References: