Robust Real-Time Face Detection

Article *in* International Journal of Computer Vision · May 2004

DOI: 10.1023/B:VISI.0000013087.49260.fb · Source: DBLP

2 authors:

Paul A. Viola
Microsoft

Michael Jeffrey Jones
Mitsubishi Electric Research Laboratories

Robust Real-time Face Detection

Paul Viola and Michael Jones
Compaq Cambridge Research Laboratory
One Cambridge Center
Cambridge, MA 02142

We have constructed a frontal face detection system which achieves detection and false positive rates which are equivalent to the best published results [7, 5, 6, 4, 1]. This face detection system is most clearly distinguished from previous approaches in its ability to detect faces extremely rapidly. Operating on 384 by 288 pixel images, faces are detected at 15 frames per second on a conventional 700 MHz Intel Pentium III. In other face detection systems, auxiliary information, such as image differences in video sequences, or pixel color in color images, have been used to achieve high frame rates. Our system achieves high frame rates working only with the information present in a single grey scale image. These alternative sources of information can also be integrated with our system to achieve even higher frame rates.

The first contribution of this work is a new image representation called an *integral image* that allows for very fast feature evaluation. Motivated in part by the work of Papageorgiou et al. our detection system does not work directly with image intensities [3]. Like these authors we use a set of features which are reminiscent of Haar Basis functions. In order to compute these features very rapidly at many scales we introduce the integral image representation for images. The integral image can be computed from an image using a few operations per pixel. Once computed, any one of these Harr-like features can be computed at any scale or location in *constant* time.

The second contribution of this work is a method for constructing a classifier by selecting a small number of important features using AdaBoost [2]. Within any image sub-window the total number of Harr-like features is very large, far larger than the number of pixels. In order to ensure fast classification, the learning process must exclude a large majority of the available features, and focus on a small set of critical features. Motivated by the work of Tieu and Viola, feature selection is achieved through a simple modification of the AdaBoost procedure: the weak learner is constrained so that each weak classifier returned can depend on only a single feature [8]. As a result each stage of the boosting process, which selects a new weak classifier, can be viewed as a feature selection process.

The third major contribution of this work is a method for combining successively more complex classifiers in a cascade structure which dramatically increases the speed of the detector by focusing attention on promising regions of the image. More complex processing is reserved only for these promising regions. Those sub-windows which are not rejected by the initial classifier are processed by a sequence of classifiers, each slightly more complex than the last. If any classifier rejects the sub-window, no further processing is performed. The structure of the cascaded detection process is essentially that of a degenerate decision tree, and as such is related to the work of Amit and Geman [1].

The complete face detection cascade has 32 classifiers. Nevertheless the cascade structure results in extremely rapid average detection times. The face detector runs at about 15 frames per second on 384 by 288 pixel images which is about 15 times faster than any previous system. On the MIT+CMU dataset, containing 507 faces and 75 million sub-windows, our detection rate is 90% with 78 false detections (which is 1 false positive in about 961,000 queries).

References

[1] Y. Amit, D. Geman, and K. Wilder. Joint induction of shape features and tree classifiers, 1997.

[2] Y. Freund and R. Schapire. A decision-theoretic generalization of on-line learning and an application to boosting. In *Eurocolt '95*, pages 23–37. Springer-Verlag, 1995.

[3] C. Papageorgiou, M. Oren, and T. Poggio. A general framework for object detection. In *ICCV*, 1998.

[4] D. Roth, M. Yang, and N. Ahuja. A snow-based face detector. In *NIPS 12*, 2000.

[5] H. Rowley, S. Baluja, and T. Kanade. Neural network-based face detection. In *IEEE PAMI*, volume 20, 1998.

[6] H. Schneiderman and T. Kanade. A statistical method for 3D object detection applied to faces and cars. In *ICCV*, 2000.

[7] K. Sung and T. Poggio. Example-based learning for view-based face detection. In *IEEE PAMI*, volume 20, pages 39–51, 1998.

[8] K. Tieu and P. Viola. Boosting image retrieval. In *ICCV*, 2000.

Proceedings of the Eighth IEEE International Conference on Computer Vision (ICCV'01)
0-7695-1143-0/01 $17.00 © 2001 IEEE

0-7695-1143-0/01 $10.00 (C) 2001 IEEE

IEEE COMPUTER SOCIETY