A Texture-based Approach to Face Detection

Vidya Manian and Arun Ross
Department of Computer Science and Electrical Engineering, West Virginia University, Morgantown, West Virginia 26506
vidyam@csee.wvu.edu, ross@csee.wvu.edu

1. Introduction

Detection of faces in static or video images is an important but challenging problem in computer vision as faces could occur at different scales, orientations, positions and pose in an unrestrained background. Several methods have been proposed in the literature for detecting faces [1]. Most techniques perform well in constrained environments but perform poorly on noisy images having a cluttered background. Further, techniques based on identifying salient features of the face (such as eyes, skin color, etc.) are sensitive to image quality and the size of constituent face objects. A face image can be viewed as a texture pattern exhibiting symmetry and regularity. In fact, it has been conjectured that the human visual system (HVS) exploits the textural nature of the human face as well as the relationship between component features (viz., eyes, nose and mouth) to detect and recognize faces spontaneously [2]. In this paper, we discuss an algorithm that exploits the textural properties of a face in order to detect faces in diverse scenarios.

2. Texture Features

Let $\{X(i,j)\}, i, j \in I,\ 1 \le i \le N_1,\ 1 \le j \le N_2$, represent a face pattern where N_1 and N_2 are the height and width of the pattern, respectively. The texture features considered in this work are a set of 6 statistical and 3 multiresolution features that capture the gradient, directional variations and the residual energies of a pattern [3]. The texture features include the average (f_1), the standard deviation (f_2), the average deviation of gradient magnitude (f_3), the average residual energy (f_4), and the average deviation of the horizontal (f_5) and vertical directional residuals (f_6) of pixel intensities given by

$$f_1 = \frac{1}{N_1 N_2}\sum_{i=1}^{N_1}\sum_{j=1}^{N_2} X(i,j), \quad f_2 = \sqrt{\sum_{i,j}\left(X(i,j) - f_1\right)^2},$$

$$f_3 = \frac{1}{N_1 N_2}\sum_{i=1}^{N_1-1}\sum_{j=1}^{N_2-1}\left|X(i,j) - X(i+1,j)\right| + \left|X(i,j) - X(i,j+1)\right|, \quad f_4 = \frac{1}{N_1 N_2}\sum_{i=1}^{N_1}\sum_{j=1}^{N_2}\left|X(i,j) - \bar{X}\right|,$$

$$f_5 = \frac{1}{N_1 N_2}\sum_{i=2}^{N_1-1}\sum_{j=1}^{N_2}\left|X(i,j) - \left(X(i-1,j) + X(i+1,j)/2\right)\right|, \quad \text{and} \quad f_6 = \frac{1}{N_1 N_2}\sum_{i=1}^{N_1}\sum_{j=2}^{N_2-1}\left|X(i,j) - \left(X(i,j-1) + X(i,j+1)/2\right)\right|,$$

respectively ($\bar{X}$ is the mean value of the face texture). These features capture the edge information of the texture pattern and are, therefore, useful in characterizing the coarseness and randomness of the texture. Apart from these features, a 3-level wavelet transform of the face image is constructed by convolving the face pattern with the orthonormal Daubechies filter of length 8. This produces 4 sub-images: the approximate, vertical, diagonal and horizontal detail images. The approximate sub-images are further decomposed in a similar manner; specifically, the sub-images at level *r* are half the size of the sub-images at level *r-1*. Wavelet features are computed from the horizontal, vertical and diagonal sub-images at different levels of decomposition. The average energy (f_7) of the 3rd level, the standard deviation (f_8) of the 1st level, and residual energy (f_9) of the 2nd level are computed as:

$$f_7 = \frac{1}{N_1 N_2}\sum_{i=1}^{N1}\sum_{j=1}^{N2}\left|C_{ij}\right|, \quad f_8 = \left[\frac{1}{N_1 N_2}\sum_{i=1}^{N_1}\sum_{j=1}^{N_2}\left|C_{ij} - M\right|^2\right]^{1/2}, \quad \text{and} \quad f_9 = \frac{1}{N_1 N_2}\sum_{i=1}^{N_1}\sum_{j=1}^{N_2}\left|C_{ij} - M\right|,$$

respectively. Here, the C_{ij}'s are the wavelet coefficients of the sub-image while M is the mean of the sub-image. The wavelet energy (f_7) and residual energy (f_9) features measure the regularity of the texture, while the variance feature (f_8) measures the homogeneity of the pattern.

3. Face Detection Algorithm

Given a set of training facial patterns the algorithm randomly partitions it into two sets – TRAIN1 and TRAIN2. The intensity distribution of images in both the sets is normalized to zero-mean-unit-variance in order to reduce the effect of changes in lighting conditions. Every image in TRAIN1 and TRAIN2 is represented by a feature vector comprising of features *f1* to *f9* and a *8x8* image space (resized). The Euclidean distance between every feature vector in TRAIN1 and those in TRAIN2 is computed, and the maximum (T_1) and minimum (T_2) of these distance values are recorded. The feature vectors in TRAIN1 are next subjected to a clustering procedure that emits a pre-determined number of clusters each characterized by a centroid, a maximum radius and a covariance matrix.

When an input image containing a face object is presented to the system, a windowing process extracts windows (with 50% overlap) from the image and computes a feature set for each intensity-normalized window. A window is deemed to be a facial candidate if the Euclidean distances between its feature vector and the feature vectors in TRAIN1 are contained entirely in the interval $[T_1, T_2]$. Next, the Mahalanobis distance between the feature set of each facial candidate and the centroids of individual face clusters is computed; a candidate window is labeled as a face pattern if its distance from any one of the clusters is within a certain fraction, *p*, of its (i.e., the cluster's) maximum radius.

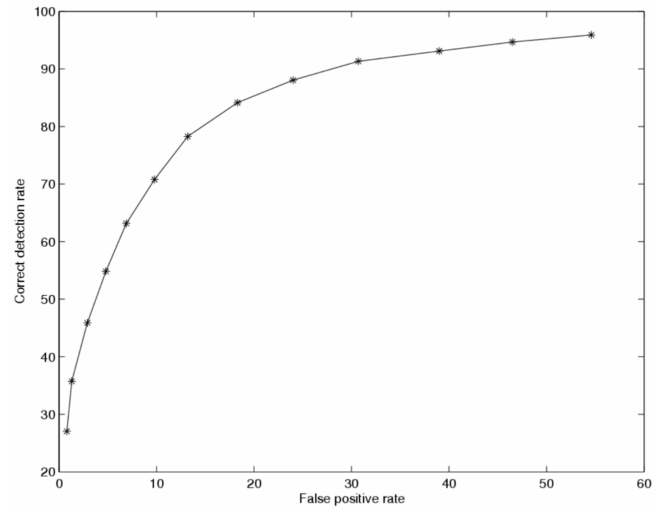

Fig. 1. The ROC curve on the CMU dataset [4] indicating the correct detection and false alarm rates (%). Rates are computed as a function of the number of facial and non-facial images used and not the number of windows examined.

4. Face Detection Results

The proposed algorithm was tested on 3 different scenarios: (a) images containing mug shots of human faces (CMU database [4]); (b) images containing different sized face images in varying background and lighting conditions (training set: AT&T database [5], test set: the BioID database [6]); (c) images possessing multiple frontal and profile facial objects (training set: AT&T [5] and UMIST [6] database, test set: CMU-MIT dataset [8]). In scenario (c) each test image was scanned by windows at three different scales: 92x92, 64x64, and 32x32; a lower scale was used only if no faces were detected in a previous scale. The number of training images used in each of the 3 scenarios varied between 60 and 120. A few results pertaining to the face detection algorithm are presented in Figs. 1, 2 and 3.

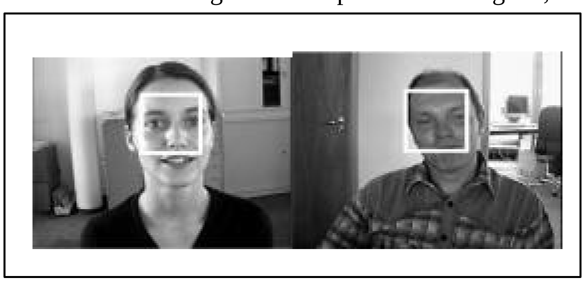

Fig. 2. Face detection on the BioID database [6].

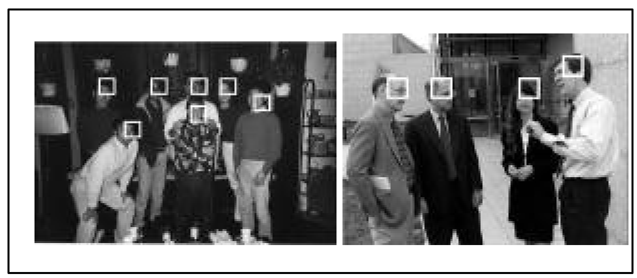

Fig. 3. Experimental results on the CMU-MIT database [8].

5. Summary

A combination of statistical and multi-resolution texture features has been used to design an automatic face detection algorithm. The algorithm is observed to perform well under different lighting and background conditions. By appropriate tuning of the thresholds, the number of false positives can be effectively reduced. The algorithm is computationally less expensive compared to other methods in the literature and is feasible for implementation in real-time face detection systems.

6. References

[1] M. H. Yang, D. J. Kriegman and N. Ahuja, "Detecting faces in images: a survey," IEEE Trans on PAMI, **24**, 34-58 (2002).

[2] E. Hjelmas and B. K. Low, "Face detection: a survey," Computer vision and Image Understanding, **83**, 236-274 (2001).

[3] V. Manian and R. Vasquez, "Approaches to color and texture based image classification," SPIE JOE, **41**, 1480-1490 (2002).

[4] CBCL Face Database #1, MIT Center for Biological and Computation Learning, http://www.ai.mit.edu/projects/cbcl.

[5] F. S. Samaria, "Face recognition using hidden Markov models," PhD thesis, Univ. of Cambridge, 1994.

[6] O. Jesorsky, K. J. Kirchberg, R. W. Frischholz , "Robust face detection using the Hausdorff distance," Proc. 3rd Intl. Conf. Audio and Video based Biometric Person Authentication, Springer, Lecture notes in computer science, (LNCS-2091, Halmstad, Sweden, June 2001), pp. 90-95.

[7] D. B. Graham and N. M. Allinson, "Characterizing Virtual Eigensignatures for General Purpose Face Recognition," Face Recognition: From Theory to Applications, Comp and Sys Sciences, **163**. H. Wechsler et al. (eds), 446-456 (1998).

[8] H. Rowley, S. Baluja and T. Kanade, "Neural network based face detection," IEEE Trans. PAMI, **20**, 23-38 (1998).