A Bayesian Discriminating Features Method for Face Detection

Chengjun Liu, *Member*, *IEEE*

Abstract—This paper presents a novel Bayesian Discriminating Features (BDF) method for multiple frontal face detection. The BDF method, which is trained on images from only one database, yet works on test images from diverse sources, displays robust generalization performance. The novelty of this paper comes from the integration of the discriminating feature analysis of the input image, the statistical modeling of face and nonface classes, and the Bayes classifier for multiple frontal face detection. First, feature analysis derives a discriminating feature vector by combining the input image, its 1D Harr wavelet representation, and its amplitude projections. While the Harr wavelets produce an effective representation for object detection, the amplitude projections capture the vertical symmetric distributions and the horizontal characteristics of human face images. Second, statistical modeling estimates the conditional probability density functions, or PDFs, of the face and nonface classes, respectively. While the face class is usually modeled as a multivariate normal distribution, the nonface class is much more difficult to model due to the fact that it includes "the rest of the world." The estimation of such a broad category is, in practice, intractable. However, one can still derive a subset of the nonfaces that lie closest to the face class, and then model this particular subset as a multivariate normal distribution. Finally, the Bayes classifier applies the estimated conditional PDFs to detect multiple frontal faces in an image. Experimental results using 887 images (containing a total of 1,034 faces) from diverse image sources show the feasibility of the BDF method. In particular, the novel BDF method achieves 98.5 percent face detection accuracy with one false detection.

Index Terms—Bayes classifier, Bayesian Discriminating Features (BDF), discriminating feature analysis, face detection, statistical modeling, support nonfaces.

1 Introduction

Among the most challenging tasks for visual form analysis and object recognition are understanding how people process and recognize each other's face, and the development of corresponding computational models for automated face recognition [8], [10]. An automated face recognition system includes several related face processing tasks, such as detection of a pattern as a face, face tracking in a video sequence, face verification, and face recognition. Face detection generally learns the statistical models of the face and nonface images, and then applies a two-class classification rule to discriminate between face and nonface patterns. Face tracking predicts the motion of faces in a sequence of images based on their previous trajectories and estimates the current and future positions of those faces. While face verification is mainly concerned with authenticating a claimed identity posed by a person, face recognition focuses on recognizing the identity of a person from a database of known individuals.

This paper presents a Bayesian Discriminating Features (BDF) method for multiple frontal face detection by integrating feature analysis, modeling, and the Bayes classifier. The main contributions of the paper come from 1) discriminating feature analysis of the input image, 2) statistical modeling of face and nonface classes, and 3) the application of the Bayes classifier for multiple frontal face detection.

• *The author is with the Department of Computer Science, New Jersey Institute of Technology, Newark, NJ 07102. E-mail: liu@cs.njit.edu.*

Manuscript received 24 June 2002; revised 26 Nov. 2002; accepted 10 Feb. 2003.
Recommended for acceptance by J. Goutsias.
For information on obtaining reprints of this article, please send e-mail to: tpami@computer.org, and reference IEEECS Log Number 116830.

First, the discriminating feature analysis combines the input image, its 1D Harr wavelet representation, and its amplitude projections. Recent research has shown that the 2D Harr wavelet representation is effective for human face and pedestrian detection [14]. For efficiency considerations, this paper incorporates the 1D Harr wavelet representation to define the discriminating features. The amplitude projections, namely the column and row projections, capture the vertical symmetric distributions and the horizontal characteristics of human face images. By combining the input image, its 1D Harr wavelet representation, and its amplitude projections, the new feature vector enhances its discriminating power for face detection.

Second, statistical modeling of face and nonface classes essentially estimates the conditional probability density functions, or PDFs, of the two classes. While the face class is usually modeled as a multivariate normal distribution, the nonface class is much more difficult to model due to the fact that it includes "the rest of the world." The estimation of such a broad category is, in practice, intractable. However, one can still derive a subset of the nonfaces that lie closest to the face class, and then model this particular subset of nonfaces as a multivariate normal distribution. The idea of using a subset of nonfaces to design the face detection algorithm is motivated by the recent statistical learning system, support vector machines, or SVMs. In SVM, only the support vectors, the patterns that lie close to the maximal margin hyperplane, are involved in designing the system. Thus, in analogy to SVMs, the BDF method first locates the "support nonfaces," and then models this particular subset of nonfaces as a multivariate normal distribution.

Finally, the BDF method applies the Bayes classifier for multiple frontal face detection. The Bayes classifier yields the minimum error when the underlying PDFs of the face and nonface classes are known. This error, called the Bayes error, is the optimal measure for feature effectiveness when classification is of concern, since it is a measure of class separability [3].

The BDF method is trained using 600 FERET face images (Batch 15) [16] and nine natural images. Experimental results using 887 images (containing a total of 1,034 faces) from diverse image sources show the feasibility of the BDF method. In particular, the novel BDF method achieves 98.5 percent face detection accuracy with one false detection, and compares favorably against the state-of-the-art face detection algorithms, such as the Schneiderman-Kanade method [21], [22].

The novelty of this paper thus comes from:

1. the discriminating feature analysis of the input image, its 1D Harr wavelet representation, and its amplitude projections;
2. statistical modeling of the face class and reducing the dimensionality of the feature vector to a very small number, M, which is 10 in our experiments;
3. nonface class modeling based on the concept of SVM, which models only a small subset of the nonfaces that lie closest to the face class; (note that, in general, the nonface class includes "the rest of the world," which makes the estimation practically intractable. The introduction of the "support nonfaces" in this paper makes the nonface class modeling tractable.);
4. the application of the Bayes classifier with a modified decision rule for multiple frontal face detection; (note that the modified decision rule, (25), introduces a control parameter, θ, which eliminates the nonfaces that are not close to the face class. And, only those subimages that are close enough to the face class are passed to the Bayes decision rule. This modified decision rule thus validates the nonface class modeling, which models only those nonfaces that lie closest to the face class rather than "the rest of the world.");
5. the development of the single response criterion and the early exclusion criterion for computational efficiency;
6. the comprehensive assessments of the BDF method for face detection by applying images from diverse image sources, and the comparative assessments of the BDF method against the state-of-the-art face detection algorithms, such as the Schneiderman-Kanade method [21], [22].

2 Background

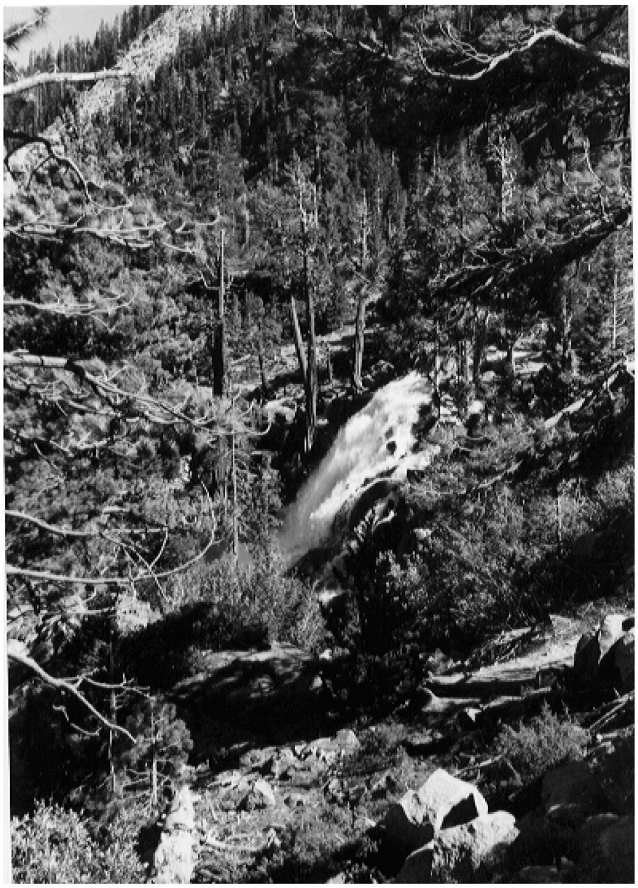
Face detection is the first stage of an automated face recognition system, since a face has to be located before it is recognized. Earlier efforts have been focused on correlation or template matching, matched filtering, subspace methods, deformable templates, etc. [15], [28]. For comprehensive surveys of these early methods, see [23], [1], [20]. Recent approaches emphasize on data-driven learning-based techniques, such as the statistical modeling methods [12], [24], [21], [22], [26], [25], the neural network-based learning methods [18], [19], [24], the statistical learning theory and SVM-based methods [4], [13], [6], the Markov random field-based methods [2], [17], and the color-based face detection [7]. For recent surveys, see [5], [27].

The statistical methods usually start with the estimation of the distributions of the face and nonface patterns, and then apply a pattern classifier or a face detector to search over a range of scales and locations for possible human faces. The neural network-based methods, however, learn to discriminate the implicit distributions of the face and nonface patterns by means of training samples and the network structure, without involving an explicit estimation procedure. Moghaddam and Pentland [12] applied unsupervised learning to estimate the density in a high-dimensional eigenspace and derived a maximum-likelihood method for single face detection. Rather than using PCA for dimensionality reduction, they implemented the eigenspace decomposition as an integral part of estimating the conditional PDF in the original high-dimensional image space. Face detection is then carried out by computing multiscale saliency maps based on the maximum-likelihood formulation. Sung and Poggio [24] presented an example-based learning method by means of modeling the distributions of face and nonface patterns. To cope with the variability of face images, they empirically chose six Gaussian clusters to model the distributions for face and nonface patterns, respectively. The density functions of the distributions are then fed to a multiple layer perceptron for face detection. Schneiderman and Kanade [21] proposed a face detector based on the estimation of the posterior probability function, which captures the joint statistics of local appearance and position as well as the statistics of local appearance in the visual world. To detect side views of a face, profile images were added to the training set to incorporate such statistics [22]. Rowley et al. [18] developed a neural network-based upright, frontal face detection system, which applies a retinally connected neural network to examine small windows of an image and decide whether each window contains a face. The face detector, which was trained using a large number of face and nonface examples, contains a set of neural network-based filters and an arbitrator which merges detections from individual filters and eliminates overlapping detections. In order to detect faces at any degree of rotation in the image plane, the system was extended to incorporate a separate router network, which determines the orientation of the face pattern. The pattern is then derotated back to the upright position, which can be processed by the early developed system [19].

3 Bayesian Discriminating Features Method for Face Detection

The Bayesian Discriminating Features (BDF) method, which displays robust generalization performance, works by integrating the discriminating feature analysis of the input image, the statistical modeling of face and nonface classes, and the Bayes classifier for multiple frontal face detection. This section details these major components of the BDF method.

(a)



(b)

Fig. 1. Face and natural images. (a) Some examples of the training faces that have been normalized to the standard resolution, 16×16. (b) An example natural image.

3.1 Discriminating Feature Analysis

The discriminating feature analysis derives a new feature vector with enhanced discriminating power for face detection, by combining the input image, its 1D Harr wavelet representation, and its amplitude projections. While the Harr wavelet representation has been shown effective for human face and pedestrian detection [14], the amplitude projections are able to capture the vertical symmetric distributions and the horizontal characteristics of human face images.

Let $I(i,j) \in \mathbb{R}^{m \times n}$ represent an input image (e.g., training images for face and nonface classes, or subimages of test images), and $\mathbf{X} \in \mathbb{R}^{mn}$ be the vector formed by concatenating the rows (or columns) of $I(i,j)$. The 1D Harr representation of $I(i,j)$ yields two images, $I_h(i,j) \in \mathbb{R}^{(m-1)\times n}$ and $I_v(i,j) \in \mathbb{R}^{m\times(n-1)}$, corresponding to the horizontal and vertical difference images, respectively.

(a)

(b)

Fig. 2. Discriminating feature analysis of the mean face and the mean nonface. (a) The first image is the mean face, the second and third images are its 1D Harr wavelet representation, and the last two bar graphs are its amplitude projections. (b) The mean nonface, its 1D Harr wavelet representation, and its amplitude projections. Note that the images and projections in (b) resemble their counterparts in (a), due to the fact that the nonface samples lie close to the face class.

$$I_h(i,j) = I(i+1,j) - I(i,j) \quad 1 \le i < m, 1 \le j \le n \quad (1)$$

$$I_v(i,j) = I(i,j+1) - I(i,j) \quad 1 \le i \le m, 1 \le j < n. \quad (2)$$

Let $\mathbf{X}_h \in \mathbb{R}^{(m-1)n}$ and $\mathbf{X}_v \in \mathbb{R}^{m(n-1)}$ be the vectors formed by concatenating the rows (or columns) of $I_h(i,j)$ and $I_v(i,j)$, respectively.

The amplitude projections of $I(i,j)$ along its rows and columns form the horizontal (row) and vertical (column) projections, $\mathbf{X}_r \in \mathbb{R}^m$ and $\mathbf{X}_c \in \mathbb{R}^n$, respectively.

$$\mathbf{X_r}(i) = \sum_{j=1}^{n} I(i,j) \quad 1 \le i \le m \quad (3)$$

$$\mathbf{X_c}(j) = \sum_{i=1}^{m} I(i,j) \quad 1 \le j \le n. \quad (4)$$

Before forming a new feature vector, the vectors $\mathbf{X}$, $\mathbf{X_h}$, $\mathbf{X_v}$, $\mathbf{X_r}$, and $\mathbf{X_c}$ are normalized by subtracting the means of their components and dividing by their standard deviations, respectively. Let $\hat{\mathbf{X}}$, $\hat{\mathbf{X}}_h$, $\hat{\mathbf{X}}_v$, $\hat{\mathbf{X}}_r$, and $\hat{\mathbf{X}}_c$ be the normalized vectors. A new feature vector $\tilde{\mathbf{Y}} \in \mathbb{R}^N$ is defined as the concatenation of the normalized vectors:

$$\tilde{\mathbf{Y}} = \left(\hat{\mathbf{X}}^t \ \hat{\mathbf{X}}_h^t \ \hat{\mathbf{X}}_v^t \ \hat{\mathbf{X}}_r^t \ \hat{\mathbf{X}}_c^t\right)^t, \quad (5)$$

where t is the transpose operator and $N = 3mn$ is the dimensionality of the feature vector $\tilde{\mathbf{Y}}$. Finally, the normalized vector of $\tilde{\mathbf{Y}}$ defines the discriminating feature vector, $\mathbf{Y} \in \mathbb{R}^N$, which is the feature vector for the multiple frontal face detection system, and which combines the input image, its 1D Harr wavelet representation, and its amplitude projections for enhanced discriminating power:

$$\mathbf{Y} = \frac{\tilde{\mathbf{Y}} - \mu}{\sigma}, \quad (6)$$

where μ and σ are the mean and the standard deviation of the components of $\tilde{\mathbf{Y}}$, respectively.

Fig. 3. Face detection examples from SET1. A square indicates a face region successfully detected. The resolution of the images is 256×384, and the faces are detected at different scales.

3.2 Statistical Modeling of Face and Nonface Classes

The main objective of statistical modeling of face and nonface classes is to estimate the conditional probability density functions, or PDFs, of these two classes, respectively. While the face class contains only faces, the nonface class encompasses all the other objects, i.e., "the rest of the world." Even though it is reasonable to assume that the face class has a multivariate normal distribution, it is pretty awkward to make the same assumption about the nonface class. The BDF method, however, derives a subset of the nonfaces that lie closest to the face class, and then models this particular subset of nonfaces as a multivariate normal distribution. The choosing of a subset of nonfaces for the BDF method resembles the idea of choosing support vectors for the design of support vector machines. In fact, the support vectors are the samples that lie closest to the decision hyperplane of a SVM, and are therefore the most important data for the determination of the optimum location of the decision hyperplane. The same idea is applied here to design the optimal decision surface for face detection by choosing the "support nonfaces" that lie closest to the face class, i.e., closest to the decision surface between the face and nonface classes.

3.2.1 Face Class Modeling

The conditional density function of the face class, ω_f, is modeled as a multivariate normal distribution:

$$p(\mathbf{Y}|\omega_f) = \frac{1}{(2\pi)^{N/2}|\mathbf{\Sigma}_f|^{1/2}} exp\left\{-\frac{1}{2}(\mathbf{Y}-\mathbf{M}_f)^t\mathbf{\Sigma}_f^{-1}(\mathbf{Y}-\mathbf{M}_f)\right\}, \tag{7}$$

where $\mathbf{M}_f \in \mathbb{R}^N$ and $\mathbf{\Sigma}_f \in \mathbb{R}^{N\times N}$ are the mean and the covariance matrix of face class ω_f, respectively. Take the natural logarithm on both sides, we have

$$ln\big[p(\mathbf{Y}|\omega_f)\big] = -\frac{1}{2}\big\{(\mathbf{Y}-\mathbf{M}_f)^t\mathbf{\Sigma}_f^{-1}(\mathbf{Y}-\mathbf{M}_f) + Nln(2\pi) + ln|\mathbf{\Sigma}_f|\big\}. \tag{8}$$

The covariance matrix, $\mathbf{\Sigma}_f$, can be factorized into the following form using the principal component analysis, or PCA [9]:

Fig. 4. Face detection examples from SET2. Some images contain faces with glasses having bright reflections.

$$\begin{aligned}\boldsymbol{\Sigma}_f &= \boldsymbol{\Phi}_f \boldsymbol{\Lambda}_f \boldsymbol{\Phi}_f^t \;\; with \;\; \boldsymbol{\Phi}_f \boldsymbol{\Phi}_f^t = \boldsymbol{\Phi}_f^t \boldsymbol{\Phi}_f = \mathbf{I}_N, \boldsymbol{\Lambda}_f \\ &= diag\{\lambda_1, \lambda_2, \ldots, \lambda_N\},\end{aligned} \tag{9}$$

where $\boldsymbol{\Phi}_f \in \mathbb{R}^{N\times N}$ is an orthogonal eigenvector matrix, $\boldsymbol{\Lambda}_f \in \mathbb{R}^{N\times N}$ a diagonal eigenvalue matrix with diagonal elements (the eigenvalues) in decreasing order ($\lambda_1 \geq \lambda_2 \geq \cdots \geq \lambda_N$), and $\mathbf{I}_N \in \mathbb{R}^{N\times N}$ an identity matrix. An important property of PCA is its optimal signal reconstruction in the sense of minimum mean-square error when only a subset of principal components is used to represent the original signal [11]. The principal components are defined by the following vector, $\mathbf{Z} \in \mathbb{R}^N$:

$$\mathbf{Z} = \boldsymbol{\Phi}_f^t (\mathbf{Y} - \mathbf{M}_f). \tag{10}$$

It then follows from (8), (9), and (10) that

$$ln\big[p(\mathbf{Y}|\omega_f)\big] = -\frac{1}{2}\Big\{\mathbf{Z}^t \boldsymbol{\Lambda}_f^{-1} \mathbf{Z} + N ln(2\pi) + ln|\boldsymbol{\Lambda}_f|\Big\}. \tag{11}$$

Note that the components of $\mathbf{Z}$ are the principal components. Applying the optimal signal reconstruction property of PCA, we use only the first M ($M \ll N$) principal components to estimate the conditional density function. We further adopt a model by Moghaddam and Pentland [12] that estimates the remaining $N - M$ eigenvalues, $\lambda_{M+1}, \lambda_{M+2}, \cdots, \lambda_N$, by the average of those values:

$$\rho = \frac{1}{N-M} \sum_{k=M+1}^{N} \lambda_k. \tag{12}$$

Note that, from (10), we have $\|\mathbf{Z}\|^2 = \|\mathbf{Y} - \mathbf{M}_f\|^2$, where $\|\cdot\|$ denotes the norm operator. This result shows that the PCA transformation, which is an orthonormal transformation, does not change norm. Now, it follows from (11) and (12) that

$$\begin{aligned}ln\big[p(\mathbf{Y}|\omega_f)\big] = -\frac{1}{2}\Bigg\{ &\sum_{i=1}^{M} \frac{z_i^2}{\lambda_i} + \frac{\|\mathbf{Y} - \mathbf{M}_f\|^2 - \sum_{i=1}^{M} {z_i}^2}{\rho} + \\ &ln\left(\prod_{i=1}^{M} \lambda_i\right) + (N-M) ln\rho + N ln(2\pi)\Bigg\},\end{aligned} \tag{13}$$

where z_is are the components of $\mathbf{Z}$ defined by (10). Equation (13) states that the conditional density function of face class can be estimated using the first M principal

TABLE 1
Testing Data Sets: SET1, SET2, and SET3, and Testing Performance

data	sources	images	faces	detected	false detections
SET1	FERET Batches 12, 13, and 14	511	511	507	0
SET2	FERET Batch 2	296	296	290	0
SET3	MIT-CMU Test Sets	80	227	221	1
Total	—	887	1,034	1,018	1

components, the input image, the mean face, and the eigenvalues of the face class.

3.2.2 Nonface Class Modeling

The nonface class modeling starts with the generation of nonface samples by applying (13) to natural images that do not contain any human faces at all. Those subimages of the natural scene that lie closest to the face class are chosen as training samples for the estimation of the conditional density function of the nonface class, ω_n, which is also modeled as a multivariate normal distribution:

$$p(\mathbf{Y}|\omega_n) = \frac{1}{(2\pi)^{N/2}|\mathbf{\Sigma}_n|^{1/2}} exp\left\{-\frac{1}{2}(\mathbf{Y}-\mathbf{M}_n)^t\mathbf{\Sigma}_n^{-1}(\mathbf{Y}-\mathbf{M}_n)\right\}, \tag{14}$$

where $\mathbf{M}_n \in \mathbb{R}^N$ and $\mathbf{\Sigma}_n \in \mathbb{R}^{N\times N}$ are the mean and the covariance matrix of nonface class ω_n, respectively.

Factorize the covariance matrix, $\mathbf{\Sigma}_n$, using PCA [9]:

$$\begin{aligned}\mathbf{\Sigma}_n &= \mathbf{\Phi}_n\mathbf{\Lambda}_n\mathbf{\Phi}_n^t \;\; with \;\; \mathbf{\Phi}_n\mathbf{\Phi}_n^t = \mathbf{\Phi}_n^t\mathbf{\Phi}_n = \mathbf{I}_N, \mathbf{\Lambda}_n \\ &= diag\left\{\lambda_1^{(n)}, \lambda_2^{(n)}, \ldots, \lambda_N^{(n)}\right\},\end{aligned} \tag{15}$$

where $\mathbf{\Phi}_n \in \mathbb{R}^{N\times N}$ is an orthogonal eigenvector matrix, $\mathbf{\Lambda}_n \in \mathbb{R}^{N\times N}$ a diagonal eigenvalue matrix with diagonal elements (the eigenvalues) in decreasing order ($\lambda_1^{(n)} \geq \lambda_2^{(n)} \geq \cdots \geq \lambda_N^{(n)}$), and $\mathbf{I}_N \in \mathbb{R}^{N\times N}$ an identity matrix. The principal component vector, $\mathbf{U} \in \mathbb{R}^N$, is defined as follows:

$$\mathbf{U} = \mathbf{\Phi}_n^t(\mathbf{Y}-\mathbf{M}_n). \tag{16}$$

Estimate the remaining $N-M$ eigenvalues, $\lambda_{M+1}^{(n)}, \lambda_{M+2}^{(n)}, \cdots, \lambda_N^{(n)}$, by the average of those values:

$$\varepsilon = \frac{1}{N-M}\sum_{k=M+1}^{N}\lambda_k^{(n)}. \tag{17}$$

The conditional density function of the nonface class can be estimated as follows:

$$\begin{aligned}ln[p(\mathbf{Y}|\omega_n)] = -\frac{1}{2}\Bigg\{&\sum_{i=1}^{M}\frac{u_i^2}{\lambda_i^{(n)}} + \frac{\|\mathbf{Y}-\mathbf{M}_n\|^2 - \sum_{i=1}^{M}u_i^2}{\varepsilon} + \\ &ln\left(\prod_{i=1}^{M}\lambda_i^{(n)}\right) + (N-M)ln\varepsilon + Nln(2\pi)\Bigg\},\end{aligned} \tag{18}$$

where u_is are the components of $\mathbf{U}$ defined by (16). Equation (18) states that the conditional density function of nonface class can be estimated using the first M principal components, the input image, the mean nonface, and the eigenvalues of the nonface class.

3.3 The Bayesian Classifier for Multiple Frontal Face Detection

After modeling the conditional PDFs of the face and nonface classes, the BDF method applies the Bayes classifier for multiple frontal face detection, since the Bayes classifier yields the minimum error when the underlying PDFs are known. This error, called the Bayes error, is the optimal measure for feature effectiveness when classification is of concern, since it is a measure of class separability [3].

Let $\mathbf{Y} \in \mathbb{R}^N$ be the discriminating feature vector constructed from an input pattern, i.e., a subimage of some test image (see Section 3.1). Let the a posteriori probabilities of face class (ω_f) and nonface class (ω_n) given $\mathbf{Y}$ be $P(\omega_f|\mathbf{Y})$ and $P(\omega_n|\mathbf{Y})$, respectively. The pattern is classified to the face class or the nonface class according to the Bayes decision rule for minimum error [3]:

$$\mathbf{Y} \in \begin{cases}\omega_f & \text{if } P(\omega_f|\mathbf{Y}) > P(\omega_n|\mathbf{Y}).\\ \omega_n & \text{otherwise.}\end{cases} \tag{19}$$

Note that the Bayes decision rule optimizes the class separability in the sense of the Bayes error, hence, should yield the best performance on face detection.

The a posteriori probabilities, $P(\omega_f|\mathbf{Y})$ and $P(\omega_n|\mathbf{Y})$, can be computed from the conditional PDFs as defined in Sections 3.2.1 and 3.2.2 using the Bayes theorem:

$$P(\omega_f|\mathbf{Y}) = \frac{P(\omega_f)p(\mathbf{Y}|\omega_f)}{p(\mathbf{Y})}, \quad P(\omega_n|\mathbf{Y}) = \frac{P(\omega_n)p(\mathbf{Y}|\omega_n)}{p(\mathbf{Y})}, \tag{20}$$

where $P(\omega_f)$ and $P(\omega_n)$ are the a priori probabilities of face class ω_f and nonface class ω_n, respectively, and $p(\mathbf{Y})$ is the mixture density function.

From (13), (18), and (20), the Bayes decision rule for face detection is then defined as follows:

$$\mathbf{Y} \in \begin{cases}\omega_f & \text{if } \delta_f + \tau < \delta_n\\ \omega_n & \text{otherwise,}\end{cases} \tag{21}$$

where δ_f, δ_n, and τ are as follows:

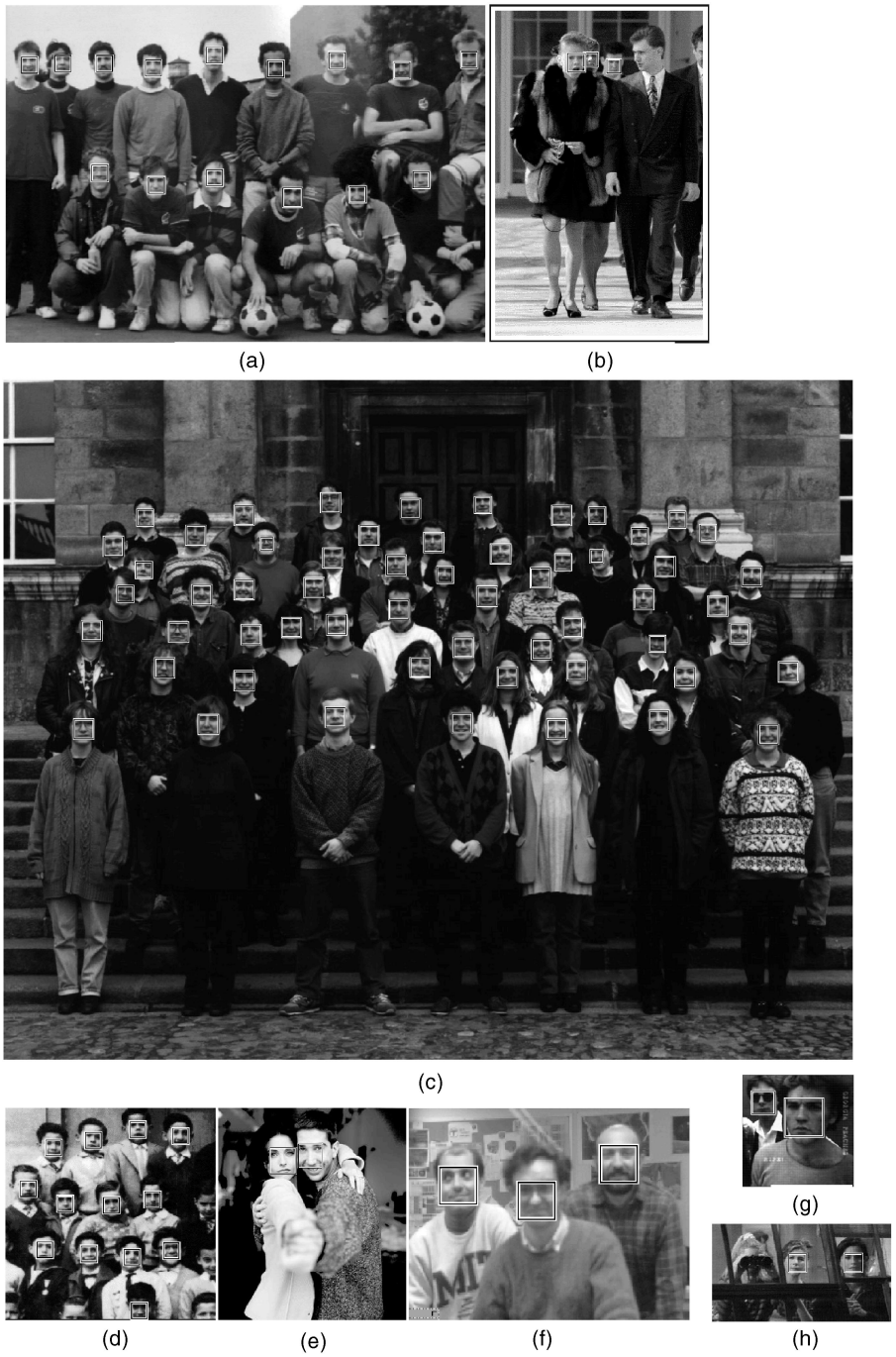


Fig. 5. Detection of multiple frontal faces.

$$\delta_f = \sum_{i=1}^{M} \frac{z_i^2}{\lambda_i} + \frac{\|\mathbf{Y} - \mathbf{M}_f\|^2 - \sum_{i=1}^{M} z_i^2}{\rho} + ln\left(\prod_{i=1}^{M} \lambda_i\right) + (N - M) ln\rho \tag{22}$$

$$\delta_n = \sum_{i=1}^{M} \frac{u_i^2}{\lambda_i^{(n)}} + \frac{\|\mathbf{Y} - \mathbf{M}_n\|^2 - \sum_{i=1}^{M} u_i^2}{\varepsilon} + ln\left(\prod_{i=1}^{M} \lambda_i^{(n)}\right) + (N - M) ln\varepsilon \tag{23}$$

$$\tau = 2ln\left[\frac{P(\omega_n)}{P(\omega_f)}\right] \tag{24}$$

δ_f and δ_n can be calculated from the input pattern $\mathbf{Y}$, the face class parameters (the mean face, the first M eigenvectors, and the eigenvalues), and the nonface class parameters (the mean nonface, the first M eigenvectors, and the eigenvalues). τ is a constant, which functions as a control parameter—the larger the value is the fewer the false detections are. To further control the false detection rate, the BDF method introduces another control parameter, θ, to the face detection system, such that

$$\mathbf{Y} \in \begin{cases} \omega_f & \text{if } (\delta_f < \theta) \text{ and } (\delta_f + \tau < \delta_n). \\ \omega_n & \text{otherwise.} \end{cases} \tag{25}$$

The control parameters, τ and θ, are empirically chosen for the face detection system.

Fig. 6. Detection of multiple frontal faces with rotations.

4 Experiments

The Bayesian Discriminating Features (BDF) method integrates feature analysis, statistical modeling, and the Bayes classifier for multiple frontal face detection. The training data for the BDF method consist of 600 FERET frontal face images from Batch 15 [16], and nine natural images. The face class thus contains 1,200 face samples for training after including the mirror images of the FERET data, and the nonface class consists of 4,500 nonface samples, which are generated by choosing the subimages that lie closest to the face class from the nine natural images.

Three testing data sets, SET1, SET2, and SET3, are created to evaluate the face detection performance of the BDF method. SET1, consisting of all the frontal face images of the Batches 12, 13, and 14 from the FERET database [16], contains mainly head or head and shoulder pictures as shown in Fig. 3. SET2, consisting of all those frontal face images from the FERET Batch 2, contains upper body pictures as shown in Fig. 4. And, SET3 consists of images chosen from the MIT-CMU test sets [18] that contain frontal faces. Table 1 shows the configurations of these test sets. In particular, SET1 and SET2 consist of 511 and 296 images,

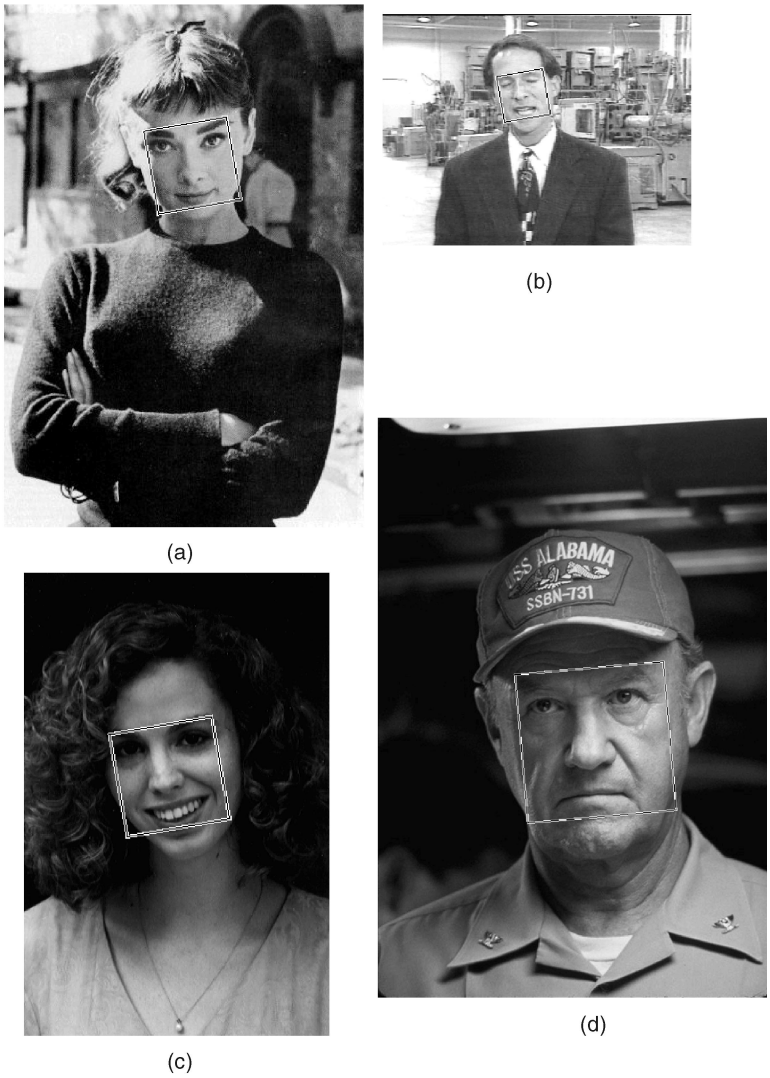


Fig. 7. Detection of rotated faces.

respectively. Note that each image in SET1 and SET2 has only one face in it.

SET3,[1] chosen from the MIT-CMU test sets [18], consists of 80 images containing a total of 227 faces. As the BDF method addresses detection of frontal and real human faces, the images that contain large pose-angled face, line-drawn face, poker face, masked face, or Cartoon face, are not included in this set. Note that the testing data are more diverse than the training data, which consist of images from only one database. Experimental results, however, show that the BDF method, which is trained on a simple image set, yet works on much more complex images, displays robust generalization performance.

4.1 Statistical Learning of the BDF Method

The statistical modeling of the face and the nonface classes requires the estimation of the parameters of these two classes from the training images. The face class parameters are computed as follows:

1. Normalize the 600 FERET images to a spatial resolution of 16×16, which is the standard resolution used in this paper for multiple frontal face detection. Fig. 1a shows some examples of the training faces that have been normalized to the standard resolution. Note that the normalization is based on the fixed eye locations and interocular distance.
2. Add the mirror images of the 600 FERET faces to the face training set and increase the number of the training samples to 1,200.

1. The 80 images (from Website: http://vasc.ri.cmu.edu/IUS/eyes_usr17/har/har1/usr0/har/faces/test/images.html) are listed as follows: albert.gif, Argentina.gif, audrey1.gif, audrey2.gif, audrybt1.gif, baseball.gif, bksomels.gif, brian.gif, bwolen.gif, cfb.gif, churchill-downs.gif, class57.gif, cluttered-tahoe.gif, cnn1085.gif, cnn1160.gif, cnn1260.gif, cnn1714.gif, cnn2020.gif, cnn2221.gif, cnn2600.gif, crimson.gif, ds9.gif, ew-courtney-david.gif, ew-friends.gif, fleetwood-mac.gif, frisbee.gif, Germany.gif, giant-panda.gif, gigi.gif, gpripe.gif, harvard.gif, hendrix2.gif, henry.gif, john.coltrane.gif, kaari-stef.gif, kaari1.gif, kaari2.gif, karen-and-rob.gif, knex0.gif, kymberly.gif, lacrosse.gif, larroquette.gif, madabout-you.gif, married.gif, me.gif, mom-baby.gif, mona-lisa.gif, natalie1.gif, nens.gif, oksana1.gif, pittsburgh-park.gif, police.gif, sarah4.gif, sarah_live_2.gif, seinfeld.gif, shumeet.gif, soccer.gif, speed.gif, tahoe-and-rich.gif, tammy.gif, tommyrw.gif, tori-crucify.gif, tori-entweekly.gif, tori-live3.gif, torrance.gif, tp-reza-girosi.gif, tp.gif, tree-roots.gif, trek-trio.gif, trekcolr.-gif, tress-photo-2.gif, tress-photo.gif, u2-cover.gif, uprooted-tree.gif, voyager2.gif, wall.gif, window.gif, wxm.gif, yellow-pages.gif, and ysato.gif.

Fig. 8. Detection of large frontal faces.

3. Incorporate the 1D Harr wavelet representation and the amplitude projections into the face images and derive the discriminating feature vectors as detailed in Section 3.1.
4. Derive the face class parameters: the mean face, the face class eigenvectors, and eigenvalues. Fig. 2a shows the mean face, its 1D Harr wavelet representation, and its amplitude projections. The first image is the mean face. The second and the third images are the horizontal and vertical difference images of the mean face, respectively, which correspond to the 1D Harr wavelet representation. The last two bar graphs draw the horizontal (row) and vertical (column) projections of the mean face, which correspond to the amplitude projections. From the figure, one can see that the amplitude projections are able to capture the vertical symmetric distributions and the horizontal characteristics of human face images.
5. The face class parameters also include M, the number of principal components used to model the conditional PDF of face class. A good choice of M should balance both the face detection performance and the computational complexity. M is empirically chosen to equal 10 for the experiments in this paper.

The learning of the nonface class parameters starts with the generation of nonface samples from the nine natural images that do not contain any human faces at all. Fig. 1b shows an example natural image, which is a natural scene image. The nonface images, chosen from the subimages of these nine natural images, have the standard spatial resolution of 16×16 and lie closest to the face class whose parameters have just been computed and whose conditional PDF is specified by (13). In particular, 4,500 nonface samples are generated from the nine natural images. Fig. 2b shows the mean nonface, its 1D Harr wavelet representation, and its amplitude projections. Note that the images and projections in Fig. 2b resemble their counterparts in Fig. 2a, due to the fact that the nonface samples lie close to the face class. After the generation of the nonface samples, the nonface class parameters can be calculated in the same way the face class parameters are computed.

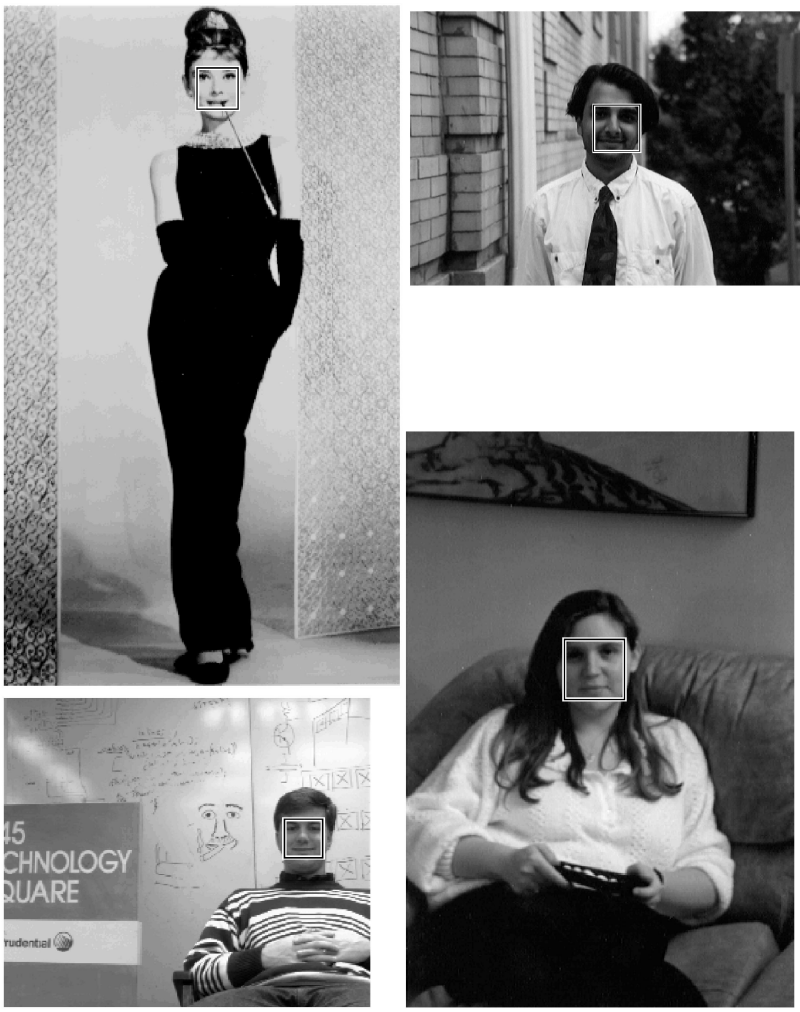


Fig. 9. Detection of small frontal faces.

Finally, the BDF method has to set the values for the two control parameters, τ and θ, whose function is to control the false detection rate. These two control parameters are empirically chosen for the BDF face detection system, and are set to equal 300 for τ and 500 for θ, respectively.

4.2 Testing Performance of the BDF Method

The BDF method is applied to detect frontal faces from three testing data sets: SET1, SET2, and SET3. SET1 and SET2 are from the FERET database [16], and the major differences between these two sets come from 1) SET1 contains mainly head or head and shoulder pictures, while SET2 contains upper body pictures, and 2) SET2 consists of more face images with glasses, and some glasses have bright reflections. Table 1 shows the detection performance of the BDF method when applied to SET1 and SET2. In particular, the BDF method successfully detects 507 faces from the 511 images (each image contains only one face) without any false detection. Fig. 3 shows examples of the detected faces from SET1, where a square indicates a face region successfully detected. Note that the resolution of the images is 256×384, and the faces are detected at different scales. The BDF method again successfully detects 290 faces from the 296 images with no false detection when it is applied to SET2. Fig. 4 shows examples of the detected faces from SET2, which contains face images with glasses having bright reflections.

Up to now, the training and testing face data are from the FERET database: The training face data are from the Batch 15, and the testing data are from the Batches 12, 13, and 14 for SET1, and the Batch 2 for SET2, respectively. To test the generalization performance of the BDF method, a third testing data set, SET3, is created from the MIT-CMU test sets [18]. SET3 consists of 80 images that contain a total of 227 faces. Note that the training face images for the BDF method are from only one database, but the test images in SET3 are from diverse sources: Some of the images are from the World Wide Web, some are scanned from photographs and newspaper pictures, and some are digitized from broadcast television [18]. Some images contain many different sized faces (Fig. 5); some include rotated faces (Figs. 6 and 7); some have very large faces (Fig. 8) or very small faces (Fig. 9); yet others involve low quality face images (Fig. 10), partially occluded faces, or slightly pose-angled faces (Fig. 11). Experiments based on such a simple training set and such a diverse testing set should be able to test the generalization performance of the BDF method.

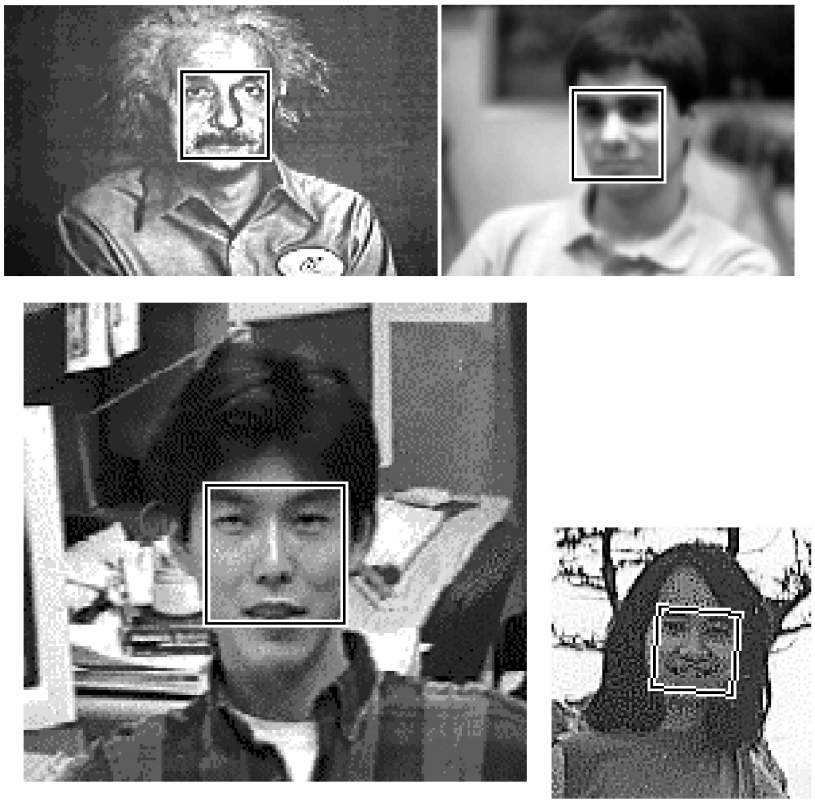


Fig. 10. Face detection in low quality images.

Fig. 5 shows the results of detection of multiple frontal faces. The BDF method successfully detects all the face images except a face with a large pose in Fig. 5b and a (downward) pose-angled face in image (d). In particular, all the 15 faces in Fig. 5a are detected at the scales 22, 26, and 27, respectively. Note that the scale 22 means that the original image is resized by a ratio, $\frac{16}{22}$. Three faces in Fig. 5b are successfully detected at the scales 20 and 26, respectively. Note that one face with a large pose is not detected, since the BDF method is trained to detect multiple frontal faces. Fig. 5c contains 57 faces and they are detected at the scales 20, 25, and 30, respectively. Fig. 5d contains 14 faces and 13 complete faces are detected at the scale 20, with a (downward) pose-angled face missed. The two faces in Fig. 5e are detected at the scales 40 and 44, respectively, and the three faces in Fig. 5f are detected at a scale of 30. Fig. 5g contains two faces, which are detected at scales 33 and 55, respectively, while Fig. 5h has two faces, which are detected at a scale of 30. The reason that the system misses one (downward) pose-angled face in Fig. 5d and one face with a large pose in Fig. 5b is that the system is trained to detect frontal faces, and the training images do not contain any pose-angled faces.

The BDF method, trained only on the upright frontal faces, can also detect rotated faces in test images by means of rotating the test images to a number of predefined degrees, such as $\pm 5^\circ$, $\pm 10^\circ$, $\pm 15^\circ$, and $\pm 20^\circ$. Fig. 6 shows the results of detection of multiple frontal faces with rotations. The BDF method successfully detects all the faces in the six test images. In particular, Fig. 6a requires two scales (26, 29) and two rotations (-10°, -15°) for the detection of all the faces, Fig. 6b requires two scales (36, 38) and one rotation (-10°), Fig. 6c requires two scales (20, 23) and two rotations (-10°, 20°), Fig. 6d requires two scales (30, 38) and one rotation (-20°), Fig. 6e requires one scale (25) and one rotation (-10°), and Fig. 6f requires two scales (35, 36) and one rotation (-20°). Fig. 7 shows some additional examples of rotated face detection using the BDF method. Figs. 7a, 7b,

Fig. 11. Detection of partially occluded or slightly pose-angled faces.

Fig. 12. Examples of missed faces and false detection. A face in (a), a baby face in (b), and a slightly pose-angled face in (c) are not detected. A false detection occurs in (c).

and 7c are rotated -10° for the detection of the faces, while image (d) is rotated -5° for the detection of the face.

The BDF method is also tested on images that contain very large faces or very small faces. Figs. 8 and 9 show the face detection performance on these test images, respectively. All the faces in Figs. 8 and 9 are successfully detected. Since the BDF method is trained on real face images, it does not detect a hand drawn face in Fig. 9, which shows the robustness of the BDF method in detection of real faces.

The generalization performance of the BDF method is further tested using low quality face images, partially occluded faces, and slightly pose-angled faces. Figs. 10 and 11 show the face detection performance of the BDF method for the detection of these three categories of faces, respectively. All the faces in Figs. 10 and 11 are successfully detected, which again shows the robustness of the BDF method in real face detection. Note that the last image in Fig. 10 is rotated 5° for the detection of the face.

For SET3, there are six faces that are not detected by the BDF method: Three faces are posed-angled, one is a baby face, one is a masked face, and one is in a low quality image. In particular, one large pose-angled face in Fig. 5b and one (downward) pose-angled face in Fig. 5d are not detected by the BDF method. Fig. 12 shows some other examples of missed faces and false detection: A low resolution face in Fig. 12a, a baby face in Fig. 12b, and a slightly pose-angled face in Fig. 12c are not detected. Also, a false detection

TABLE 2
Comparative Face Detection Performance of the Schneiderman-Kanade Method and the BDF Method on Testing Data Set, SET3, which Contains 80 Images and 227 Faces

method	faces detected	false detections	detection rate
Schneiderman-Kanade (1.0, 1.0)	218	41	96.0%
Schneiderman-Kanade (2.0, 2.0)	214	5	94.3%
Schneiderman-Kanade (3.0, 3.0)	208	1	91.6%
the BDF method	221	1	97.4%

Note that the two numbers in the parentheses of the Schneiderman-Kanade method correspond to the frontal detection threshold and the profile detection threshold, respectively, which control the number of faces detected and the number of false detections.

occurs in Fig. 12c. The experimental results using 80 test images (containing in total 227 faces) from the MIT-CMU test sets show that the BDF method detects 221 out of the 227 faces in these images with one false detection.

Table 1 summarizes the detection performance of the BDF method for the testing data sets: SET1, SET2, and SET3. The overall face detection performance of the BDF method using the 887 images containing a total of 1,034 faces is 98.5 percent correct face detection rate with one false detection.

4.3 Comparative Face Detection Performance

Among the state-of-the-art face detection algorithms, the Schneiderman-Kanade method [21], [22] is publicly available at http://www.vasc.ri.cmu.edu/cgi-bin/demos/findface.cgi. This method has two thresholds, the frontal detection threshold and the profile detection threshold, which control the number of faces detected and the number of false detections—increasing these thresholds decreases both numbers. Table 2 shows the comparative face detection performance of the Schneiderman-Kanade method and the BDF method on the testing data set, SET3, which contains 80 images and 227 faces. Note that the two numbers in the parentheses correspond to the frontal detection threshold and the profile detection threshold, respectively.

Experimental results[2] show that the Schneiderman-Kanade method achieves 96.0 percent detection rate with 41 false detections when the thresholds are set to be 1.0. The detection rate decreases when the thresholds get larger: 94.3 percent detection rate with five false detections when the thresholds are 2.0, and 91.6 percent detection rate with one false detection when the thresholds are 3.0. The BDF method, achieving 97.4 percent face detection accuracy with one false detection, thus compares favorably against the state-of-the-art face detection algorithms, such as the Schneiderman-Kanade method [21], [22].

2. The experimental results of the Schneiderman-Kanade method are derived by submitting the images in SET3 to the face detector at http://www.vasc.ri.cmu.edu/cgi-bin/demos/findface.cgi. The complete face detection results for SET3 using the BDF method are available at http://www.cs.njit.edu/~liu/RESEARCH/fd/fd.html. Note that the BDF method is a frontal face detection method, and it cannot detect large pose-angled faces in an image. The Schneiderman-Kanade method, however, is capable of detecting both frontal and profile faces.

4.4 Computational Efficiency of the BDF Method

The computational efficiency of the BDF method is mainly due to two criteria, namely, the single response criterion and the early exclusion criterion. The single response criterion circumvents the possibility of multiple responses to a singe face, while the early exclusion criterion uses a heuristic procedure to eliminate subimages that could not be faces.

Fig. 13a shows the idea of the single response criterion. Let the searching order of the subimages be from top to bottom, and then from left to right. Suppose a face is detected and a point p, the first pixel (the upper left pixel) of the subimage, is used to represent this face. For simplicity, we use the upper left pixel to represent a 16×16 subimage in the following discussion. Now, we want to search a small neighborhood of p, say, 7×7, in order to find among these 49 face candidates the one that lies closest to the face class. Note that 49 face candidates do not mean 49 faces, since some of the candidates may not be classified as face. But, at least we know p corresponds to a face, hence, there should be one face defined by one of these 49 pixels, and our purpose is to find the one that lies closest to the face class. Due to the predefined searching order, half of these neighbors have already been searched, the remaining unsearched neighbors are the pixels inside the area A. Note that each of these 24 neighbors defines a 16×16 subimage, which could be a face. Suppose q defines a face that lies closest to the face class, then all the pixels inside area B, which defines the half 16×16 neighborhood of the pixel q, should not be search again due to the nonoverlapping assumption of human faces. Note that the nonoverlapping assumption really means that we are interested in detecting complete faces. Fig. 13a shows that once a face is detected, 470 subimages are excluded from further processing, and the area B is an eliminated area. As a result, the single response criterion improves the speed of face detection by excluding subimages in eliminated areas from further processing. Note that, when carrying the eliminated area from one scale to another scale, one should shrink the size of the neighborhood by one or two pixels in order to detect closely adjacent or partially overlapping faces as those shown in Fig. 5e.

To further improve the computational efficiency, we define a heuristic procedure that excludes subimages which could not be faces at all, such as some homogeneous background. As the major computation takes place in discriminate feature analysis and evaluation of the Bayes decision rule, an early exclusion of those nonface subimages

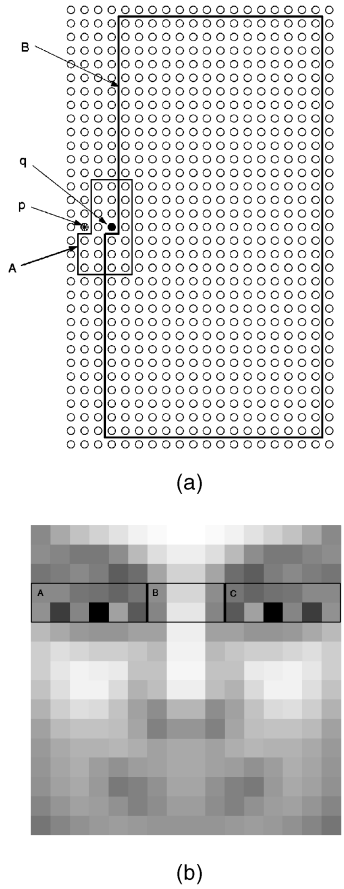


Fig. 13. The single response criterion and the early exclusion criterion. (a) The single response, q, and the eliminated area, B. (b) A 16×16 subimage with three labeled regions corresponding to the left eye area (A), the nose bridge area (B), and the right eye area (C), respectively.

would greatly improve the computational efficiency of the BDF face detection system. Fig. 13b shows a 16×16 subimage with three labeled regions corresponding to the left eye area (A), the nose bridge area (B), and the right eye area (C), respectively. The idea of the exclusion criterion is based on some simple statistics. First, calculate the mean values, μ_A, μ_B, and μ_C, of regions A, B, and C, respectively. Then, compute the average values, m_A and m_C, of the pixels whose intensity values are above the mean values of regions A and C, respectively, and compute the average value, m_B, of the pixels whose intensity values are below the mean value of regions B. Note that, if no pixels in a region are above (or below) the mean value, then assign the mean value to the average value. Actually, such a region is a homogenous region, i.e., all the pixels have the same intensity value. Finally, the exclusion criterion states that a subimage is excluded from further processing if $m_A \leq \kappa m_B$ or $m_C \leq \kappa m_B$, where κ, $0 < \kappa < 1$, is a control factor.

The main factor of the running time of the BDF face detection system is the number of subimages the system has to process. Currently, it takes the system an average of one second to process a 320×240 image without any scaling (containing $68,625$ subimages in total) on a 900 MHz Sun Blade 1000 workstation. Note that different image complexity requires different processing time.

5 Discussion

This paper presents a novel Bayesian Discriminating Features (BDF) method for multiple frontal face detection. The BDF method, which is trained on images from only one database, yet works on test images from diverse sources, displays robust generalization performance. The novelty of this paper comes from the integration of the discriminating feature analysis of the input image, the statistical modeling of face and nonface classes, and the Bayes classifier for multiple frontal face detection. First, feature analysis derives a discriminating feature vector by combining the input image, its 1D Harr wavelet representation, and its amplitude projections. Second, statistical modeling estimates the conditional probability density functions, or PDFs, of the face and nonface classes. Finally, the Bayes classifier applies the estimated conditional PDFs to detect multiple frontal faces in an image. The BDF method is trained using 600 FERET face images and nine natural images. Experimental results using 887 images (containing a total of 1,034 faces) from diverse image sources show the feasibility of the BDF method. In particular, the novel BDF method achieves 98.5 percent face detection accuracy with one false detection.

Closely related to the BDF method is the maximum-likelihood method developed by Moghaddam and Pentland [12] for single face detection. In comparison, the BDF method differs from this maximum-likelihood method in the following aspects:

1. the discriminating feature analysis, which integrates the input image, its 1D Harr wavelet representation, and its amplitude projections,
2. the statistical modeling of the nonface class,
3. the application of the Bayes classifier for multiple frontal face detection,
4. the computational efficiency due to the design of the single response criterion and the early exclusion criterion, and
5. multiple frontal face detection. Note that the maximum-likelihood method [12] does not contain nonface modeling. In analogy to support vector machines, the BDF method first locates the support nonfaces and then models this particular subset of nonfaces as a multivariate normal distribution.

Future research will consider pose-angled face detection and detecting faces in video. One possibility is to discretize the pose space and design algorithms for face detection for each possible pose. The algorithms should consider among others feature analysis and statistical modeling of the different pose classes. Regarding detecting faces in video, one possibility is to use motion information to detect quickly region of interest, or ROI, from video, and then apply the detection algorithms, such as the BDF method introduced in this paper, to the ROI areas and locate faces.

Acknowledgments

The author would like to thank the anonymous reviewers for their critical and constructive comments and suggestions.

REFERENCES

[1] R. Chellappa, C.L. Wilson, and S. Sirohey, "Human and Machine Recognition of Faces: A Survey," *Proc. IEEE,* vol. 83, no. 5, pp. 705-740, 1995.

[2] S.C. Dass and A.K. Jain, "Markov Face Models," *Proc. Eighth IEEE Int'l Conf. Computer Vision,* pp. 680-687, 2001.

[3] K. Fukunaga, *Introduction to Statistical Pattern Recognition.* second ed. Academic Press, 1991.

[4] B. Heisele, T. Poggio, and M. Pontil, "Face Detection in Still Gray Images," A.I. memo AIM-1687, Artificial Intelligence Laboratory, MIT, 2000.

[5] E. Hjelmas and B.K. Low, "Face Detection: A Survey," *Computer Vision and Image Understanding,* vol. 83, pp. 236-274, 2001.

[6] P. Ho, "Rotation Invariant Real-Time Face Detection and Recognition System," A.I. memo AIM-2001-010, Artificial Intelligence Laboratory, MIT, 2001.

[7] R.L. Hsu, M. Abdel-Mottaleb, and A.K. Jain, "Face Detection in Color Images," *Proc. Int'l Conf. Image Processing,* pp. 1046-1049, 2001.

[8] C. Liu and H. Wechsler, "Evolutionary Pursuit and Its Application to Face Recognition," *IEEE Trans. Pattern Analysis and Machine Intelligence,* vol. 22, no. 6, pp. 570-582, June 2000.

[9] C. Liu and H. Wechsler, "Robust Coding Schemes for Indexing and Retrieval from Large Face Databases," *IEEE Trans. Image Processing,* vol. 9, no. 1, pp. 132-137, 2000.

[10] C. Liu and H. Wechsler, "A Shape and Texture Based Enhanced Fisher Classifier for Face Recognition," *IEEE Trans. Image Processing,* vol. 10, no. 4, pp. 598-608, 2001.

[11] C. Liu and H. Wechsler, "Gabor Feature Based Classification Using the Enhanced Fisher Linear Discriminant Model for Face Recognition," *IEEE Trans. Image Processing,* vol. 11, no. 4, pp. 467-476, 2002.

[12] B. Moghaddam and A. Pentland, "Probabilistic Visual Learning for Object Representation," *IEEE Trans. Pattern Analysis and Machine Intelligence,* vol. 19, no. 7, pp. 696-710, July 1997.

[13] A. Mohan, C. Papageorgiou, and T. Poggio, "Example-Based Object Detection in Images by Components," *IEEE Trans. Pattern Analysis and Machine Intelligence,* vol. 23, no. 4, pp. 349-361, Apr. 2001.

[14] C.P. Papageorgiou, M. Oren, and T. Poggio, "A General Framework for Object Detection," *Proc. Int'l Conf. Computer Vision,* pp. 555-562, Jan. 1998.

[15] A. Pentland, B. Moghaddam, and T. Starner, "View-Based and Modular Eigenspaces for Face Recognition," *Proc. Conf. Computer Vision and Pattern Recognition,* pp. 84-91, 1994.

[16] P.J. Phillips, H. Wechsler, J. Huang, and P. Rauss, "The FERET Database and Evaluation Procedure for Face-Recognition Algorithms," *Image and Vision Computing,* vol. 16, pp. 295-306, 1998.

[17] R.J. Qian and T.S. Huang, "Object Detection Using Hierarchical MRF and MAP Estimation," *Proc. Conf. Computer Vision and Pattern Recognition,* pp. 186-192, 1997.

[18] H.A. Rowley, S. Baluja, and T. Kanade, "Neural Network-Based Face Detection," *IEEE Trans. Pattern Analysis and Machine Intelligence,* vol. 20, no. 1, pp. 23-38, Jan. 1998.

[19] H.A. Rowley, S. Baluja, and T. Kanade, "Rotation Invariant Neural Network-Based Face Detection," *Proc. Conf. Computer Vision and Pattern Recognition,* pp. 38-44, June 1998.

[20] A. Samal and P.A. Iyengar, "Automatic Recognition and Analysis of Human Faces and Facial Expression: A Survey," *Pattern Recognition,* vol. 25, no. 1, pp. 65-77, 1992.

[21] H. Schneiderman and T. Kanade, "Probabilistic Modeling of Local Appearance and Spatial Relationships for Object Recognition," *Proc. Conf. Computer Vision and Pattern Recognition,* pp. 45-51, June 1998.

[22] H. Schneiderman and T. Kanade, "A Statistical Method for 3D Object Detection Applied to Faces and Cars," *Proc. Conf. Computer Vision and Pattern Recognition,* pp. 746-751, 2000.

[23] K.K. Sung, "Learning and Example Selection for Object and Pattern Detection," PhD thesis, AI Lab, MIT, 1996.

[24] K.K. Sung and T. Poggio, "Example-Based Learning for View-Based Human Face Detection," *IEEE Trans. Pattern Analysis and Machine Intelligence,* vol. 20, no. 1, pp. 39-51, Jan. 1998.

[25] P. Viola and M. Jones, "Rapid Object Detection Using a Boosted Cascade of Simple Features," *Proc. Conf. Computer Vision and Pattern Recognition,* pp. 511-518, 2001.

[26] M.H. Yang, N. Ahuja, and D. Kriegman, "Face Detection Using Mixtures of Linear Subspaces," *Proc. Fifth Int'l Conf. Automatic Face and Gesture Recognition,* pp. 70-76, 2000.

[27] M.H. Yang, D. Kriegman, and N. Ahuja, "Detecting Faces in Images: A Survey," *IEEE Trans. Pattern Analysis and Machine Intelligence,* vol. 24, no. 1, pp. 34-58, Jan. 2002.

[28] A.L. Yuille, "Deformable Templates from Face Recognition," *J. Cognitive Neuroscience,* vol. 3, no. 1, pp. 59-70, 1991.

Chengjun Liu received the PhD degree from George Mason University in 1999, and he is presently an assistant professor of computer science at New Jersey Institute of Technology. His research interests are in computer vision, pattern recognition, image processing, evolutionary computation, and neural computation. His recent research has been concerned with the development of novel and robust methods for image/video retrieval and object detection, tracking, and recognition based upon statistical and machine learning concepts. The class of new methods includes the Probabilistic Reasoning Models (PRM), the Enhanced Fisher Models (EFM), the Enhanced Independent Component Analysis (EICA), the Shape and Texture-Based Fisher method (STF), the Gabor-Fisher Classifier (GFC), and the Independent Gabor Features (IGF) method. He has also pursued the development of novel evolutionary methods leading to the development of the Evolutionary Pursuit (EP) method for pattern recognition in general, and face recognition in particular. He is a member of the IEEE and the IEEE Computer Society.