Multiscale Face Detection: A New Approach to Robust Face Detection

Mehdi Aghagolzadeh, Hamid Soltanian-Zadeh, *Senior Member, IEEE,* and Babak N. Araabi

***Abstract*—This paper presents a novel method for detecting multiple frontal faces in still images using multi-scale processing. The main characteristic of this algorithm is its stability in detecting faces with seldom false detections and a high correct detection rate. The novelty of this work comes from the utilization of multiscale detection and using two classifiers to reduce false detections. The algorithm generally has two stages: in the first stage, a face is detected in a unique scale and in the second stage, only the faces that are located in the neighbor scales are accepted as real faces. Consequently, a still image is first resized and scanned block wise, and then each enhanced block is tested for being face. One dimensional Harr wavelet is used for feature extraction, which gives appropriate discriminating features between the face and nonface classes. Detection results at each scale are accumulated in an internal database, so the ultimate detection is prepared based on the mutual detection information between consequent scales. To parameterize both the Bayesian and the simple proposed classifier, 2,643 faces were congregated from famous face databases and more than 10,000 non-face samples were selected from nature images. Experimental results using images gathered from known databases like MIT-CMU show great ability of the proposed algorithm in detecting faces.**

Keywords: Face detection, Bayesian classifier, wavelet transform, multi-scale processing.

I. Introduction

NEW information and multimedia technologies has developed intelligent methods for interaction between human and machines, where keyboard, mouse and monitor play no roles anymore. In recent years, the cost decreasing and performance increasing of computation has developed computer vision systems through our daily life. Human identification and face recognition with applications through surveillance and data security has fascinated a lot of attention. Other applications in relation with human faces are developed, like facial feature detection [1], face authentication [2] and facial expression recognition [3]. In all of these applications and subjects, face detection is the first step.

In face detection, the purpose is to determine whether there is a face in an image or not and locate existing faces. Because of the wide variety of faces and problems transpired to a face, face detection has been an unfathomable dilemma for a while. Some of these difficulties are different face poses, facial expressions and environmental conditions, like light conditions and camera properties.

The preliminary algorithms for face detection have been based on match filters [4] and correlation methods [5] which had poor performance. Recent data driven methods have shown a higher success in detecting faces from still images. Some of these methods use artificial neural networks [6]. Others use Bayesian classifiers [7], SVM [8], FDA [9], and clustering methods [10]. These algorithms use Gabor filter [11], Harr wavelet [12], entropy measurement [13] or image gradient [14] as features to improve correct classification. For detecting faces from the environment, the most distinctive features are extracted.

For detecting faces in a still image, the input image is scanned block wise pixel by pixel for faces. This procedure is done in multiple scales for finding faces with different sizes. Therefore, the computational complexity will increase with a complicated and time consuming classifier. Face candidate finder [15] is a good technique to reduce the scanning region but it decreases the detection rate.

In the proposed algorithm, a combination of two simple classifiers is used and detection results in different scales are used to increase robustness. Although detection in distinctive scales generates different results, mutual information between the results of the neighboring scales is an excellent tool for increasing the stability of the method.

The novelties of this paper are:

1. Suggesting a simple classifier for face localization.
2. Using detection results of two parallel classifiers to boost the classification performance at each scale.
3. Using detection results at multiple scales to increase the stability of the face detection process.

In the next section, the proposed algorithm is presented briefly. The feature extraction method and the two parallel classifiers are described in this section. In Section 3, the method for robust face detection is described. Experiments are presented in Section 4 and conclusions are given in Section 5.

M. Aghagolzadeh is with the Department of Electrical and Computer Engineering, University of Tehran, Tehran 14395-515, Iran, e-mail: m.golzadeh@ece.ut.ac.ir.

H. Soltanian-Zadeh is with the Control and Intelligent Processing Center of Excellence, Electrical and Computer Engineering Department, University of Tehran, Tehran 14395-515, Iran, and also with the Radiology Image Analysis Laboratory, Henry Ford Health System, Detroit, MI 48202 USA, e-mails: hszadeh@ut.ac.ir; hamids@rad.hfh.edu.

B.N. Araabi is with the Control and Intelligent Processing Center of Excellence, Electrical and Computer Engineering Department, University of Tehran, Tehran 14395-515, Iran, e-mails: araabi@ut.ac.ir.

II. PROPOSED FACE DETECTION ALGORITHM

For detecting face in a still image, face patterns must be separated from the "whole world except faces" which is known as the nonface class. Many classifiers are designed

To classify these two groups, but for decreasing computational complexity, simple classifiers like Bayesian classifier and a simple proposed classifier are preferred. Some operations must be done on the input image to generate patterns for classification, like image resizing, image scanning, image enhancement, and feature extraction (Figure 1). Detection results produced by the two combined classifiers at each scale are stored in an internal database, and then the final block composes a robust detection based on the stored data from multiple scales.

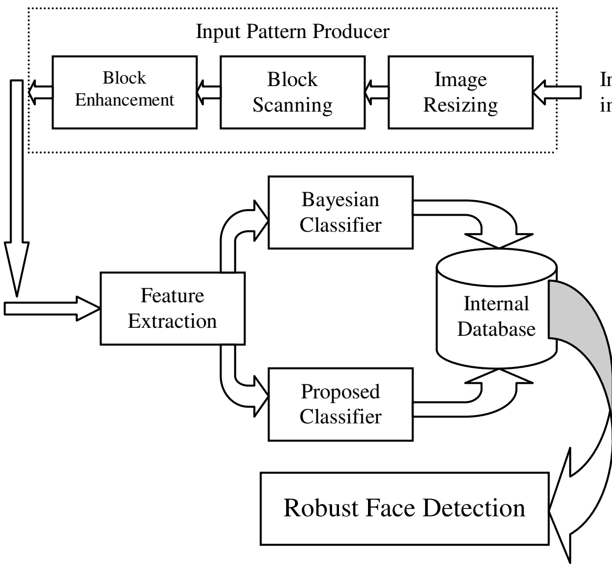

Figure 1: Block diagram of the proposed face detection algorithm.

A. Input Pattern Producer

This block consists of three sub-blocks which are image resizing, block scanning, and block enhancement. In image resizing, the input still image is resized with a factor of α/β. Therefore, first the image is upsampled with the rate of α and then interpolated to estimate inserted zeros and then downsampled by the rate of β. There is a trade off between stability and computational complexity for the scales chosen to resize the input image.

The resized image is scanned from left to right and up to down pointwise and a block is chosen around that pixel with a dimension of 20x20 pixels. Selected bocks naturally have a nonuniform intensity distribution, so histogram equalization is applied to each block to generate a uniform intensity distribution. Before applying histogram equalization, it is necessary to correct possible inconsistent lightning conditions and reduce heavy shadows on faces formed by environmental conditions. Therefore, the mean value of the block is subtracted from the original block and then the block histogram equalized. The block diagram of the enhancement section is shown in Figure 2. Four face blocks that are strongly affected by incongruous lightning conditions are selected to illustrate the affectivity of the input pattern producer (Figure 3).

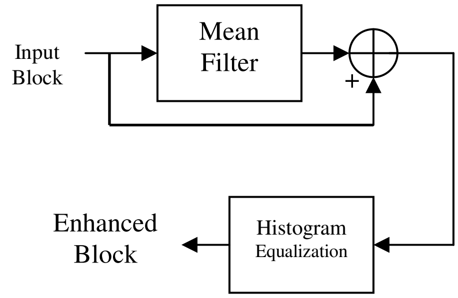

Figure 2: Flowchart of the block enhancement.

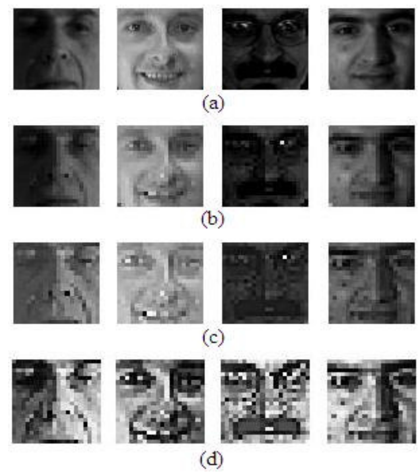

Figure 3: (a) Original Block, (b) Resampled Block to 20x20, (c) Intensity Corrected Block, (d) Histogram Equalized Block.

B. Feature Extraction

Face detection algorithms use different tools for feature extraction like Gabor filters, Harr wavelet, PCA [16], and ICA [17]. Between these methods, the Harr wavelet seems to give distinctive features between face and nonface classes. Therefore, one-dimensional Harr wavelet is applied on the input block in both horizontal and vertical directions. The horizontal and vertical one-dimensional Harr wavelet features are extracted from the input block by:

$$H_{hor}(i, j) = B(i, j+1) - B(i, j) \tag{1}$$

$$H_{ver}(i, j) = B(i+1, j) - B(i, j) \tag{2}$$

If the input block vector is denoted by I_b and the vectors of the horizontal and vertical Harr wavelet are shown by I_{hor} and I_{ver}, respectively, the feature vector is formed by:

$$I_{feature} = (I_b{}^T, I_{hor}{}^T, I_{ver}{}^T)^T \tag{3}$$

Since the block size is 20x20 pixels, then the length of I_{hor} is one pixel less in horizontal dimension and I_{ver} is one pixel less in vertical dimension, so the blocks sizes are 20x19 and 19x20 pixels respectively and the total length of the feature vector is 1,160. Feature extraction is applied on a face pattern (Figure 4-1) and a nonface pattern (Figure 4-2). Figures 4-3 and 4-4 show the mean of the blocks and their Harr wavelet transforms in the horizontal and vertical directions, respectively.

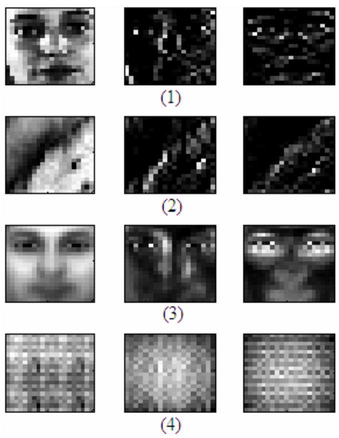

Figure4: The first column shows the enhanced input block, the second column is its horizontal one-dimensional Harr wavelet feature and the third column is its vertical one-dimensional Harr wavelet feature.

C. Classifiers

Two combined classifiers are used to classify input patterns as face or nonface. To design any classifier, a training process is done on the gathered training samples. Therefore, a plenty of data samples must be gathered as training samples for both the face and nonface class to define the parameters of any classifier.

For the face class, face samples are selected manually from well-known face databases like BioID, Yale, ORL, and World Wide Web images. In this way, a face block is selected by locating the position of the left and right eyes in the face and cluttering a size of the face based upon the distance between the two eyes shown by M (figure 5).

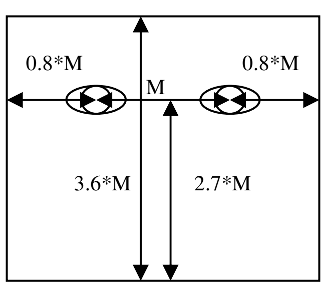

Figure5: Frontal face pattern selected from face databases.

Manually selected frontal faces have different face rotations (between 0 degree to 15 degree with directions to right or left) with different facial expressions (happy, angry, sad, etc.) with differences in wearing glasses or having mustache or beard. There are 2,643 frontal faces gathered.

Nonface samples, any sample that is not a face, are selected from images taken from nature for the nonface class, where it has different patterns resembling as nonface. More than 10,500 nonface patterns are selected from 7 real images of nature each in the size of 20x20 pixels. To each of the face and nonface blocks, the block enhancement and feature extraction methods are applied. Consequently, face and nonface vectors are gathered in the face and nonface training classes, respectively.

The Bayesian classifier is applied onto the two training classes to find the discriminating function that produces the minimum classification error. For this purpose, both the face and nonface classes are modeled by Gaussian distributions. The conditional probability density function (pdf) of the face class (ω_f) is modeled as:

$$P(Y \mid \omega_f) = \frac{1}{(2\pi)^{N/2}\left|\Sigma_f\right|^{1/2}} \exp\left\{-\frac{1}{2}\left(Y - M_f\right)^t \Sigma_f^{-1}\left(Y - M_f\right)\right\} \quad (4)$$

where M_f and Σ_f are the mean vector and covariance matrix of the face class, respectively. Similarly, the conditional pdf of the nonface class is modeled as:

$$P(Y \mid \omega_n) = \frac{1}{(2\pi)^{N/2}\left|\Sigma_n\right|^{1/2}} \exp\left\{-\frac{1}{2}\left(Y - M_n\right)^t \Sigma_n^{-1}\left(Y - M_n\right)\right\} \quad (5)$$

where M_n and Σ_n are the mean vector and covariance matrix of the nonface class. If Y is the input feature vector, the Bayesian method classifies it as face or nonface by the following decision rule:

$$Y \in \begin{cases} \omega_f & \text{if } P(\omega_f \mid Y) \;>\; P(\omega_n \mid Y) \\ \omega_n & \text{otherwise} \end{cases}$$

In the above decision rule, a posteriori probabilities are used. These probabilities can be directly extracted from the conditional pdf's by the following:

$$P(\omega_f \mid Y) = \frac{P(Y \mid \omega_f) P(\omega_f)}{P(Y)} \quad (6)$$

$$P(\omega_n \mid Y) = \frac{P(Y \mid \omega_n) P(\omega_n)}{P(Y)} \quad (7)$$

Using the above equations in the decision rule, a simple decision rule is achieved for face and nonface classification:

$$Y \in \begin{cases} \omega_f & \text{if } \varphi_f \;>\; \kappa\varphi_n \\ \omega_n & \text{otherwise} \end{cases}$$

where φ_f , φ_n and κ are:

$$\varphi_f = \exp\left\{-\frac{1}{2}\left(Y - M_f\right)^t \Sigma_f^{-1}\left(Y - M_f\right)\right\} \quad (8)$$

$$\varphi_n = \exp\left\{-\frac{1}{2}\left(Y - M_n\right)^t \Sigma_n^{-1}\left(Y - M_n\right)\right\} \quad (9)$$

$$\kappa = \frac{P(\omega_n)}{P(\omega_f)}\left(\frac{\left|\Sigma_f\right|}{\left|\Sigma_n\right|}\right)^{1/2} \quad (10)$$

In the above equations, φ_f and φ_n are directly calculated from the input pattern Y and κ is a constant value that can be learned from the training data. If κ is set to a large value, both of the correct and false face detection rates will decrease. If κ is set to a small value, both of the rates will increase. Therefore, there is a trade off in choosing κ. In the proposed algorithm, κ is chosen so that the correct detection rate is 100% but with a high number of false detections. The proposed algorithm decreases the false detection using the Bayesian classifier in combination with a second classifier.

The second classifier compares the input pattern with the mean of face and nonface training samples and classifies it to a face or nonface class. This classifier uses the parameters computed for the Bayesian classifier and thus no new training process is needed. In this classifier, the following conditional pdf's are computed.

$$P(\omega_f \mid Y) = \sum_{i=1}^{1160} \frac{I_{feature}(i)}{\left(M_{face}(i)\right)^p} \tag{11}$$

$$P(\omega_n \mid Y) = \sum_{i=1}^{1160} \frac{I_{feature}(i)}{\left(M_{nonface}(i)\right)^p} \tag{12}$$

where M_{face} and $M_{nonface}$ are the mean of face and nonface training samples, respectively. The decision rule using the above conditional pdf's is the same as that for the Bayesian classifier.

$$Y \in \begin{cases} \omega_f & if\ P(\omega_f \mid Y) \ < \ \eta \times P(\omega_n \mid Y) \\ \omega_n & otherwise \end{cases}$$

Experiments show that for $p = 3$, the maximum discrimination is achieved. To learn κ and η for the classifiers, the likelihood ratio (ratio of conditional face probability to conditional nonface probability) for the classifiers are drawn in Figure 6.

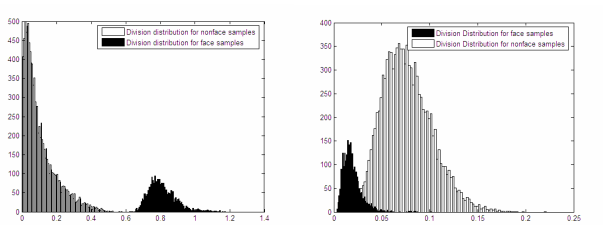

Figure 6: (a) Division distribution for Bayesian classifier, (b) Division distribution for the proposed classifier.

As it can be seen, the Bayesian classifier gives a better discrimination than the proposed second classifier, but the advantage of the second classifier is that it does not repeat the false detections of the Bayesian classifier, reducing the overall false detection rate.

III. ROBUST FACE DETECTION

As mentioned before, detection at multiple scales by the two parallel classifiers are collected in an internal database. Robust face detection is obtained using the stored data in two stages. In the first stage, a face is located in a unique scale based on the two classifier's detection results at that scale, with the definition of centralization. In the second stage, final decision for face detection is made upon the detection results at sequent scales with the definition of intersection.

A. Detection in a Unique Scale

Since face detection algorithms use the scanning method for finding faces, instead of one pixel, a group of neighboring pixels resemble a face pattern, but this differs for false detections. In a false detection, a nonface pattern is wrongly classified as a face pattern, but its neighboring pixel is not detected as face patterns. Therefore, an alone detected pixel is assumed to be a false detection. A set of spread detected pixels is also distinguished as false detections.

Consequently, a face pattern is recognized when a number of detected pixels have a limited variance in that scale. So a classifier locates a face when a group of detected pixels satisfies the conditions of the following decision rules.

$$Bayesian \begin{cases} face & if\ Var(Detection) < \lambda_{Bayesian} \ \ \& \ \ P_{Detected} > N \\ nonface & otherwise \end{cases}$$

$$\Pr oposed \begin{cases} face & if\ Var(Detection) < \lambda_{\Pr oposed} \ \ \& \ \ P_{Detected} > N \\ nonface & otherwise \end{cases}$$

In the above decision rules, $\lambda_{Bayesian} \approx 50$ and $\lambda_{\Pr oposed} \approx 120$. The number of detected pixels resembling a face pattern must be more than 10 pixels at a scale (N=10).

A face is located when both of the classifiers detect at the same location a face pattern. If only one classifier detects a face pattern, the detection is considered as a false detection. To show the performance of this method, two regions are cropped from an image of the MIT-CMU face database. One image contains two faces, shown at lower left part of Figure 7. The detection results of the Bayesian and proposed classifiers are shown. The second image is selected so that it produces false detections in both of the classifiers; a group of spread detected pixels by the Bayesian classifier and separated detected pixels by the proposed classifier. Since the detection location differs in the two classifiers and the detected pixels do not satisfy the decision rules, the locations are not detected as face patterns.

B. Final Detection in Sequent Scales

After detecting faces at each scale, only if the same location is detected as a face pattern in the neighboring scales, the detected locations are considered as the ultimate face locations. Therefore, the following equation is applied to detect the ultimate faces at a desired scale.

$$Ultimate\_Face(S_i) \equiv Face(S_{i-1}) \cap Face(S_i) \cap Face(S_{i+1}) \tag{13}$$

Figure 8 shows an image from the MIT-CMU face database and detection results at some sequent scales. Using equation (13), a face is detected at scales 0.8 and 0.75. It must be noted that detection results for sequent scales with small disparities should be used. Otherwise, the proposed method may not detect faces with a high stability.

Figure 7: Face detection at a unique scale, two faces are detected in the lower left image but no face is detected in the lower right image.

Figure 8: Final face detection based on detection results of sequent scales. A face pattern is detected in scale=0.8 and scale=0.75.

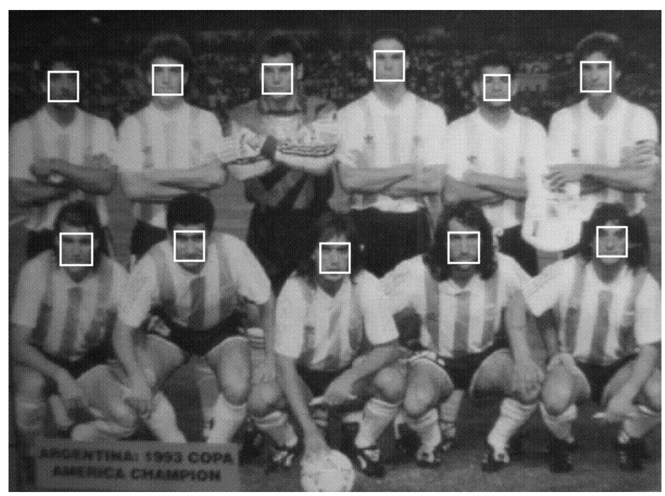

Figure 9: Face detection in low quality images.

IV. EXPERIMENTS AND RESULTS

The proposed algorithm was tested on two sets of images, which are totally different with those used for the training process. Test set 1 consists of 23 MIT-CMU images, mostly with a low quality and small faces (challenging problem in face detection) with cluttered background. Test 1 has a total of 130 faces. Test set 2 consists of 52 images, each containing only one face. Some of the images have cluttered background and others have a simple background, gathered from several websites.

Experiments show that the proposed algorithm for face detection has a very good performance in detecting low quality faces and faces affected by environmental lightning conditions. However, it is sensitive to pose rotated faces and faces corrupted by other objects.

The detection rates and false positive rates on test sets 1-2 are listed in Table 1. Note that the false detection, especially for test set 2, is very low, while the detection rate is acceptable. Two examples from the MIT-CMU face database are shown in Figures 9-10. Also, three examples from the test set 2 with cluttered and simple backgrounds are shown in Figure 11.

Figure 10: Example of missed faces (9 missed faces) in low quality images.

Table: Face detection results on test sets 1-2.

	Test set 1 (MIT-CMU) 23 images with 130 faces	Test set 2 (gathered) 52 images with 52 faces
Detection rate	93.1%	100%
False detection	1	0

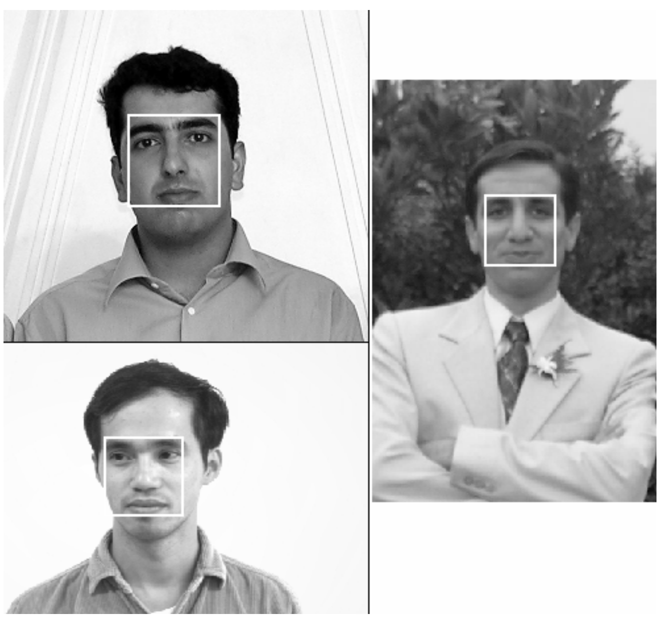

Figure 11: Detection of large frontal faces with different backgrounds.

V. CONCLUSION

A novel and robust face detection algorithm is proposed utilizing two independent classifiers at multiple sequent scales. The proposed algorithm has reduced false detections and enhanced detection stability. Face localization is done in two stages. In the first stage, candidate faces are detected based on the results of the parallel classifiers in different scales. In the second stage, robust face detection is achieved using detection results of the neighboring scales. Experiments using the MIT-CMU face database show great ability of the proposed algorithm in detecting faces, especially detecting faces in low quality images, without giving too much false detections. The reason that the proposed algorithm has reduced false detections is that a false detection occurs when both of the classifiers detect a nonface pattern as a face at the same location in three sequent scales; this is almost impossible. By setting the parameters of the two classifiers such that almost all frontal faces are detected, a high correct face detection rate is achieved.

REFERENCES

[1] Craw, D. Tock, and A. Bennett, "Finding Face Features," Proc. Second European Conf. Computer Vision, pp. 92-96, 1992.

[2] C. Kotropoulos, A. Tefas, and I. Pitas, "Frontal Face Authentication Using Variants of Dynamic Link Matching Based on Mathematical Morphology," Proc. IEEE Int'l Conf. Image Processing, pp. 122-126, 1998.

[3] G. Donato, M.S. Bartlett, J.C. Hager, P. Ekman, and T.J. Sejnowski, "Classifying Facial Actions," IEEE Trans. Pattern Analysis and Machine Intelligence, vol. 21, no. 10, pp. 974-989, Oct. 2000.

[4] T.K. Leung, M.C. Burl, and P. Perona, "Finding Faces in Cluttered Scenes Using Random Labeled Graph Matching," Proc. Fifth IEEE Int'l Conf. Computer Vision, pp. 637-644, 1995.

[5] T. Sakai, M. Nagao, and S. Fujibayashi, "Line Extraction and Pattern Detection in a Photograph," Pattern Recognition, vol. 1, pp. 233-248, 1969.

[6] H. 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.

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

[8] E. Osuna, R. Freund, and F. Girosi, "Training Support Vector Machines: An Application to Face Detection," Proc. IEEE Conf. Computer Vision and Pattern Recognition, pp. 130-136, 1997.

[9] A. Martinez and A. Kak, "PCA versus LDA," IEEE Trans. Pattern Analysis and Machine Intelligence, vol. 23, no. 2, pp. 228-233, Feb. 2001.

[10] T. Rikert, M. Jones, and P. Viola, "A Cluster-Based Statistical Model for Object Detection," Proc. Seventh IEEE Int'l Conf. Computer Vision, vol. 2, pp. 1046-1053, 1999.

[11] K. Hotta, T. Mishima, T. Kurita, and S. Umeyama, "Face matching through information theoretical attention points and its applications to face detection and classification," in Proceedings Fourth IEEE International Conference on Automatic Face and Gesture Recognition, 2000.

[12] 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.

[13] J. Huang, S. Gutta, and H. Wechsler, "Detection of Human Faces Using Decision Trees," Proc. Second Int'l Conf. Automatic Face and Gesture Recognition, pp. 248-252, 1996.

[14] L. L. Huang, A. Shimizu, Y. Hagihara and H. Kobatake, "Gradient feature extraction for classification-based face detection," Pattern Recognition 36, 2003, 2501-2511.

[15] J. Wu and Z. H. Zhou, "Efficient face candidate selector for face detection," Pattern Recognition 36, 2003, 1175-1186.

[16] C. Liu, "A Bayesian Discriminating Features Method for Face Detection," IEEE Trans. Pattern Analysis and Machine Intelligence, Vol. 25, No. 6, June 2003.

[17] B. A. Draper, K. Baek, M. S. Bartlett and R. Beveridge, "Recognizing faces with PCA and ICA," Computer Vision and Image Understanding 91, 2003, 115-137.