ISSN 1077-3142
Volume 83 Number 3
September 2001

Computer Vision and Image Understanding

ACADEMIC PRESS
A Harcourt Science and Technology Company

Computer Vision and Image Understanding

Volume 83, Number 3, September 2001

1077-3142/01 $35.00

MADE IN THE UNITED STATES OF AMERICA

This journal is printed on acid-free paper. ∞

COMPUTER VISION AND IMAGE UNDERSTANDING

(ISSN 1077-3142)

Published monthly by Academic Press
Editorial and Production Offices: 525 B Street, Suite 1900, San Diego, CA 92101-4495
Accounting and Circulation Offices: 6277 Sea Harbor Drive, Orlando, FL 32887-4900

2001: Volumes 81–84. Price: $899.00 U.S.A. and Canada; $999.00 all other countries
All prices include postage and handling.

Information concerning personal subscription rates may be obtained by writing to the Publishers. All correspondence, permission requests, and subscription orders should be addressed to the office of the Publishers at 6277 Sea Harbor Drive, Orlando, FL 32887-4900 (telephone: 407-345-2000). Send notices of change of address to the office of the Publishers at least 6 to 8 weeks in advance. Please include both old and new addresses. POSTMASTER: Send changes of address to *Computer Vision and Image Understanding*, 6277 Sea Harbor Drive, Orlando, FL 32887-4900.

Periodicals postage paid at Orlando, FL, and at additional mailing offices.

Computer Vision and Image Understanding 83, 236–274 (2001)
doi:10.1006/cviu.2001.0921, available online at http://www.idealibrary.com on IDEAL®

Face Detection: A Survey

Erik Hjelmås[1]

Department of Informatics, University of Oslo, P.O. Box 1080 Blindern, N-0316 Oslo, Norway

and

Boon Kee Low

Department of Meteorology, University of Edinburgh, JCMB, Kings Buildings, Mayfield Road, Edinburgh EH9 3JZ, Scotland, United Kingdom
E-mail: erikh@hig.no; boon@met.ed.ac.uk

Received October 23, 2000; accepted April 17, 2001

In this paper we present a comprehensive and critical survey of face detection algorithms. Face detection is a necessary first-step in face recognition systems, with the purpose of localizing and extracting the face region from the background. It also has several applications in areas such as content-based image retrieval, video coding, video conferencing, crowd surveillance, and intelligent human–computer interfaces. However, it was not until recently that the face detection problem received considerable attention among researchers. The human face is a dynamic object and has a high degree of variability in its apperance, which makes face detection a difficult problem in computer vision. A wide variety of techniques have been proposed, ranging from simple edge-based algorithms to composite high-level approaches utilizing advanced pattern recognition methods. The algorithms presented in this paper are classified as either feature-based or image-based and are discussed in terms of their technical approach and performance. Due to the lack of standardized tests, we do not provide a comprehensive comparative evaluation, but in cases where results are reported on common datasets, comparisons are presented. We also give a presentation of some proposed applications and possible application areas.

Key Words: face detection; face localization; facial feature detection; feature-based approaches; image-based approaches.

1. INTRODUCTION

The current evolution of computer technologies has envisaged an advanced machinery world, where human life is enhanced by artificial intelligence. Indeed, this trend has already

[1] Also with the faculty of technology, Gjøvik University College, Norway.

1077-3142/01 $35.00

FIG. 1. Typical training images for face recognition.

prompted an active development in machine intelligence. Computer vision, for example, aims to duplicate human vision. Traditionally, computer vision systems have been used in specific tasks such as performing tedious and repetitive visual tasks of assembly line inspection. Current development in this area is moving toward more generalized vision applications such as face recognition and video coding techniques.

Many of the current face recognition techniques assume the availability of frontal faces of similar sizes [14, 163]. In reality, this assumption may not hold due to the varied nature of face appearance and environment conditions. Consider the pictures in Fig. 1.[2] These pictures are typical test images used in face classification research. The exclusion of the background in these images is necessary for reliable face classification techniques. However, in realistic application scenarios such as the example in Fig. 2, a face could occur in a complex background and in many different positions. Recognition systems that are based on standard face images are likely to mistake some areas of the background as a face. In order to rectify the problem, a visual front-end processor is needed to localize and extract the face region from the background.

Face detection is one of the visual tasks which humans can do effortlessly. However, in computer vision terms, this task is not easy. A general statement of the problem can be defined as follows: Given a still or video image, detect and localize an unknown number (if any) of faces. The solution to the problem involves segmentation, extraction, and verification of faces and possibly facial features from an uncontrolled background. As a visual front-end processor, a face detection system should also be able to achieve the task regardless of illumination, orientation, and camera distance. This survey aims to provide insight into the contemporary research of face detection in a structural manner. Chellappa *et al.* [14] have conducted a detailed survey on face recognition research. In their survey, several issues, including segmentation and feature extraction, related to face recognition have been reviewed. One of the conclusions from Chellappa *et al.* was that the face detection problem has received surprisingly little attention. This has certainly changed over the past five years as we show in this survey.

The rest of this paper is organized as follows: In Section 2 we briefly present the evolution of research in the area of face detection. Sections 3 and 4 provide a more detailed survey and discussion of the different subareas shown in Fig. 3, while in Section 5 we present some of the possible and implemented applications of face detection technology. Finally, summary and conclusions are in Section 6.

[2] The images are courtesy of the ORL (The Olivetti and Oracle Research Laboratory) face database at http://www.cam-orl.co.uk/facedatabase.html.

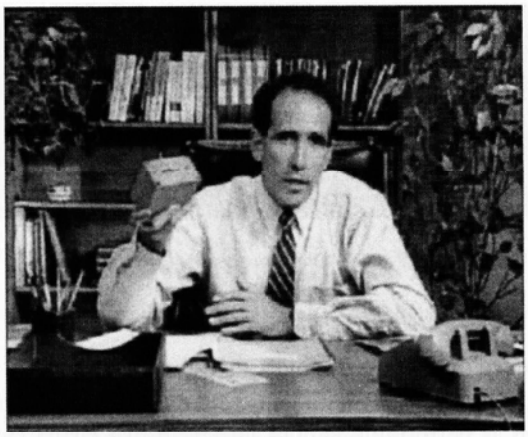

FIG. 2. A realistic face detection scenario.

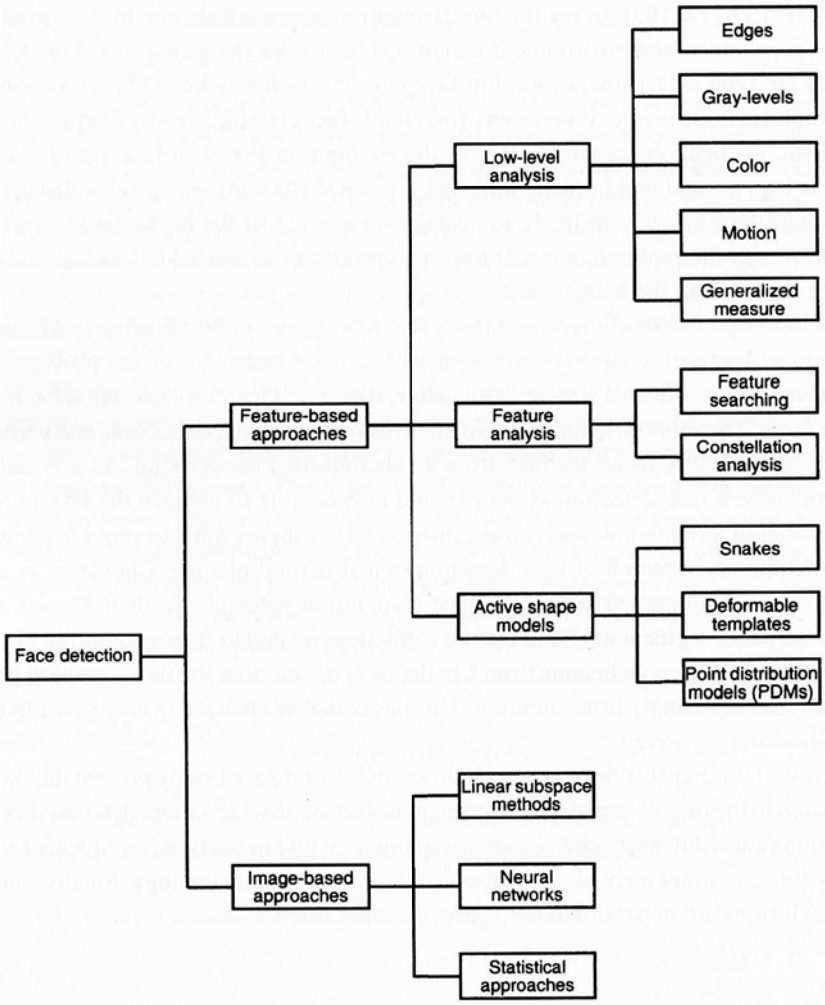

FIG. 3. Face detection divided into approaches.

2. EVOLUTION OF FACE DETECTION RESEARCH

Early efforts in face detection have dated back as early as the beginning of the 1970s, where simple heuristic and anthropometric techniques [162] were used. These techniques are largely rigid due to various assumptions such as plain background, frontal face—a typical passport photograph scenario. To these systems, any change of image conditions would mean fine-tuning, if not a complete redesign. Despite these problems the growth of research interest remained stagnant until the 1990s [14], when practical face recognition and video coding systems started to become a reality. Over the past decade there has been a great deal of research interest spanning several important aspects of face detection. More robust segmentation schemes have been presented, particularly those using motion, color, and generalized information. The use of statistics and neural networks has also enabled faces to be detected from cluttered scenes at different distances from the camera. Additionally, there are numerous advances in the design of feature extractors such as the deformable templates and the active contours which can locate and track facial features accurately.

Because face detection techniques requires *a priori* information of the face, they can be effectively organized into two broad categories distinguished by their different approach to utilizing face knowledge. The techniques in the first category make explicit use of face knowledge and follow the classical detection methodology in which low level features are derived prior to knowledge-based analysis [9, 193]. The apparent properties of the face such as skin color and face geometry are exploited at different system levels. Typically, in these techniques face detection tasks are accomplished by manipulating distance, angles, and area measurements of the visual features derived from the scene. Since features are the main ingredients, these techniques are termed the feature-based approach. These approaches have embodied the majority of interest in face detection research starting as early as the 1970s and therefore account for most of the literature reviewed in this paper. Taking advantage of the current advances in pattern recognition theory, the techniques in the second group address face detection as a general recognition problem. *Image-based* [33] representations of faces, for example in 2D intensity arrays, are directly classified into a face group using training algorithms without feature derivation and analysis. Unlike the feature-based approach, these relatively new techniques incorporate face knowledge implicitly [193] into the system through mapping and training schemes.

3. FEATURE-BASED APPROACH

The development of the feature-based approach can be further divided into three areas. Given a typical face detection problem in locating a face in a cluttered scene, low-level analysis first deals with the segmentation of visual features using pixel properties such as gray-scale and color. Because of the low-level nature, features generated from this analysis are ambiguous. In feature analysis, visual features are organized into a more global concept of face and facial features using information of face geometry. Through feature analysis, feature ambiguities are reduced and locations of the face and facial features are determined. The next group involves the use of active shape models. These models ranging from *snakes*, proposed in the late 1980s, to the more recent point distributed models (PDM) have been developed for the purpose of complex and nonrigid feature extraction such as eye pupil and lip tracking.

3.1. Low-Level Analysis

3.1.1. Edges. As the most primitive feature in computer vision applications, edge representation was applied in the earliest face detection work by Sakai *et al.* [162]. The work was based on analyzing line drawings of the faces from photographs, aiming to locate facial features. Craw *et al.* [26] later proposed a hierarchical framework based on Sakai *et al.*'s work to trace a human head outline. The work includes a line-follower implemented with curvature constraint to prevent it from being distracted by noisy edges. Edge features within the head outline are then subjected to feature analysis using shape and position information of the face. More recent examples of edge-based techniques can be found in [9, 16, 59, 115, 116] for facial feature extraction and in [32, 50, 58, 60, 70, 76, 108, 201, 222] for face detection. Edge-based techniques have also been applied to detecting glasses in facial images [79, 80].

Edge detection is the foremost step in deriving edge representation. So far, many different types of edge operators have been applied. The Sobel operator was the most common filter among the techniques mentioned above [9, 16, 32, 76, 87, 108]. The Marr–Hildreth edge operator [124] is a part of the proposed systems in [50, 70]. A variety of first and second derivatives (Laplacian) of Gaussians have also been used in the other methods. For instance, a Laplacian of large scale was used to obtain a line drawing in [162] and steerable and multi-scale-orientation filters in [59] and [58], respectively. The steerable filtering in [59] consists of three sequential edge detection steps which include detection of edges, determination of the filter orientation of any detected edges, and stepwise tracking of neighboring edges using the orientation information. The algorithm has allowed an accurate extraction of several key points in the eye.

In an edge-detection-based approach to face detection, edges need to be labeled and matched to a face model in order to verify correct detections. Govindaraju [50] accomplishes this by labeling edges as the left side, hairline, or right side of a front view face and matches these edges against a face model by using the golden ratio[3] [40] for an ideal face:

$$\frac{\text{height}}{\text{width}} \equiv \frac{1+\sqrt{5}}{2}. \tag{1}$$

Govindaraju's edge-based feature extraction works in the following steps:

Edge detection: the Marr-Hildreth edge operator
Thinning: a classical thinning algorithm from Pavlidis [141]
Spur removal: each connected component is reduced to its central branch
Filtering: the components with non-face-like properties are removed
Corner detection: the components are split based on corner detection
Labeling: the final components are labeled as belonging to the left side, hairline, or right side of a face

The labeled components are combined to form possible face candidate locations based on a cost function (which uses the golden ratio defined above). On a test set of 60 images with complex backgrounds containing 90 faces, the system correctly detected 76% of the faces with an average of two false alarms per image.

[3] An aesthetically proportioned rectangle used by artists.

3.1.2. Gray information. Besides edge details, the gray information within a face can also be used as features. Facial features such as eyebrows, pupils, and lips appear generally darker than their surrounding facial regions. This property can be exploited to differentiate various facial parts. Several recent facial feature extraction algorithms [5, 53, 100] search for local gray minima within segmented facial regions. In these algorithms, the input images are first enhanced by contrast-stretching and gray-scale morphological routines to improve the quality of local dark patches and thereby make detection easier. The extraction of dark patches is achieved by low-level gray-scale thresholding. On the application side, Wong *et al.* [207] implement a robot that also looks for dark facial regions within face candidates obtained indirectly from color analysis. The algorithm makes use of a weighted human eye template to determine possible locations of an eye pair. In Hoogenboom and Lew [64], local maxima, which are defined by a bright pixel surrounded by eight dark neighbors, are used instead to indicate the bright facial spots such as nose tips. The detection points are then aligned with feature templates for correlation measurements.

Yang and Huang [214], on the other hand, explore the gray-scale behavior of faces in mosaic (pyramid) images. When the resolution of a face image is reduced gradually either by subsampling or averaging, macroscopic features of the face will disappear. At low resolution, face region will become uniform. Based on this observation, Yang proposed a hierarchical face detection framework. Starting at low resolution images, face candidates are established by a set of rules that search for uniform regions. The face candidates are then verified by the existence of prominent facial features using local minima at higher resolutions. The technique of Yang and Huang was recently incorporated into a system for rotation invariant face detection by Lv *et al.* [119] and an extension of the algorithm is presented in Kotropoulos and Pitas [95].

3.1.3. Color. Whilst gray information provides the basic representation for image features, color is a more powerful means of discerning object appearance. Due to the extra dimensions that color has, two shapes of similar gray information might appear very differently in a color space. It was found that different human skin color gives rise to a tight cluster in color spaces even when faces of difference races are considered [75, 125, 215]. This means color composition of human skin differs little across individuals.

One of the most widely used color models is RGB representation in which different colors are defined by combinations of red, green, and blue primary color components. Since the main variation in skin appearance is largely due to luminance change (brightness) [215], normalized RGB colors are generally preferred [27, 53, 75, 88, 92, 165, 181, 196, 202, 207, 213, 215], so that the effect of luminance can be filtered out. The normalized colors can be derived from the original RGB components as follows:

$$r = \frac{R}{R+G+B} \tag{2}$$

$$g = \frac{G}{R+G+B} \tag{3}$$

$$b = \frac{B}{R+G+B}. \tag{4}$$

From Eqs. (2)–(4), it can been seen that $r + g + b = 1$. The normalized colors can be effectively represented using only r and g values as b can be obtained by noting $b = 1 - r - g$. In skin color analysis, a color histogram based on r and g shows that face color

occupies a small cluster in the histogram [215]. By comparing color information of a pixel with respect to the r and g values of the face cluster, the likelihood of the pixel belonging to a flesh region of the face can be deduced.

Besides RGB color models, there are several other alternative models currently being used in the face detection research. In [106] HSI color representation has been shown to have advantages over other models in giving large variance among facial feature color clusters. Hence this model is used to extract facial features such as lips, eyes, and eyebrows. Since the representation strongly relates to human perception of color [106, 178], it is also widely used in face segmentation schemes [51, 83, 118, 125, 175, 178, 187, 220].

The YIQ color model has been applied to face detection in [29, 205]. By converting RGB colors into YIQ representation, it was found that the I-component, which includes color's ranging from orange to cyan, manages to enhance the skin region of Asians [29]. The conversion also effectively suppresses the background of other colors and allows the detection of small faces in a natural environment. Other color models applied to face detection include HSV [48, 60, 81, 216], YES [160], YCrCb [2, 48, 84, 130, 191, 200], YUV [1, 123], CIE-xyz [15], L * a * b * [13, 109], L * u * v * [63], CSN (a modified rg representation) [90, 91] and UCS/Farnsworth (a perceptually uniform color system was proposed by Farnsworth [210]) [208].

Terrilon *et al.* [188] recently presented a comparative study of several widely used color spaces (or more appropriately named chrominance spaces in this context since all spaces seeks luminance-invariance) for face detection. In their study they compare normalized TSL (tint-saturation-luminance [1881]), rg and CIE-xy chrominance spaces, and CIE-DSH, HSV, YIQ, YES, CIE-L * u * v * , and CIE L * a * b * chrominance spaces by modeling skin color distributions with either a single Gaussian or a Gaussian mixture density model in each space. Hu's moments [68] are used as features and a multilayer perceptron neural network is trained to classify the face candidates. In general, they show that skin color in normalized chrominance spaces can be modeled with a single Gaussian and perform very well, while a mixture-model of Gaussians is needed for the unnormalized spaces. In their face detection test, the normalized TSL space provides the best results, but the general conclusion is that the most important criterion for face detection is the degree of overlap between skin and nonskin distributions in a given space (and this is highly dependent on the number of skin and nonskin samples available for training).

Color segmentation can basically be performed using appropriate skin color thresholds where skin color is modeled through histograms or charts [13, 63, 88, 118, 178, 207, 220]. More complex methods make use of statistical measures that model face variation within a wide user spectrum [3, 27, 75, 125, 139, 215]. For instance, Oliver *et al.* [139] and Yang and Waibel [215] employ a Gaussian distribution to represent a skin color cluster of thousands of skin color samples taken from difference races. The Gaussian distribution is characterized by its mean (μ) and covariance matrix (Σ). Pixel color from an input image can be compared with the skin color model by computing the Mahalanobis distance. This distance measure then gives an idea of how close the pixel color resembles the skin color of the model.

An advantage of the statistical color model is that color variation of new users can be adapted into the general model by a learning approach. Using an adaptive method, color detection can be more robust against changes in environment factors such as illumination conditions and camera characteristics. Examples of such a learning approach have been used by Oliver *et al.* and Yang and Waibel according to Eq. (6) which updates the parameters of the Gaussian distribution [139] (a similar approach can be found in the face recognition

system of McKenna *et al.* [127]).

$$\Sigma_{\text{new}} = \left[\Sigma_{\text{general}}^{-1} + \Sigma_{\text{user}}^{-1}\right]^{-1} \tag{5}$$

$$\mu_{\text{new}} = \Sigma_{\text{new}}\left[\Sigma_{\text{general}}^{-1} \times \mu_{\text{general}} + \Sigma_{\text{user}}^{-1} \times \mu_{\text{user}}\right] \tag{6}$$

3.1.4. Motion. If the use of a video sequence is available, motion information is a convenient means of locating moving objects. A straightforward way to achieve motion segmentation is by frame difference analysis. This approach, whilst simple, is able to discern a moving foreground efficiently regardless of the background content. In [5, 53, 152, 192, 218], moving silhouettes that include face and body parts are extracted by thresholding accumulated frame difference. Besides face region, Luthon and Lievin [118], Crowley and Berard [27], and Low [115, 116] also employ frame difference to locate facial features. In [27], the existence of an eye-pair is hypothesized by measuring the horizontal and the vertical displacements between two adjacent candidate regions obtained from frame difference.

Another way of measuring visual motion is through the estimation of moving image contours. Compared to frame difference, results generated from moving contours are always more reliable, especially when the motion is insignificant [126]. A spatio-temporal Gaussian filter has been used by McKenna *et al.* [126] to detect moving boundaries of faces and human bodies. The process involves convolution of gray image $I(x, y)$ with the second order temporal edge operator $m(x, y, t)$ which is defined from the Gaussian filter $G(x, y, t)$ as follows [126],

$$\begin{aligned} G(x, y, t) &= u\left(\frac{a}{\pi}\right)^{\frac{3}{2}} e^{-a(x^2+y^2+u^2t^2)} \\ m(x, y, t) &= -\left(\nabla^2 + \frac{1}{u^2}\frac{\partial^2}{\partial t^2}\right)G(x, y, t), \end{aligned} \tag{7}$$

where u is a time scaling factor, and a is the filter width. The temporal edge operator then convolved with consecutive frames from an image sequence by

$$S(x, y, t) = m(x, y, t) \otimes I(x, y, t). \tag{8}$$

The result of the temporal convolution process $S(x, y, t)$ contains zero-crossings which provide a direct indication of moving edges in $I(x, y, t)$. The locations of the detected zero-crossings are then clustered to finally infer the location of moving objects. More sophisticated motion analysis techniques have also been applied in some of the most recent face detection research. Unlike the methods described above which identify moving edges and regions, these methods rely on the accurate estimation of the apparent brightness velocities called *optical flow*. Because the estimation is based on short-range moving patterns, it is sensitive to fine motion. Lee *et al.* [106] use optical flow to measure face motion. Based on the motion information, a moving face in an image sequence is segmented. The optical flow is modeled by the image flow constraint equation [106]

$$I_x V_x + I_y V_y + I_t = 0, \tag{9}$$

where I_x, I_y, and I_t are the spatio-temporal derivatives of the image intensity and V_x and V_y are the image velocities. By solving the above equation for V_x and V_y, an optical flow

field that contains moving pixel trajectories is obtained. Regions corresponding to different motion trajectories are then classified into motion and nonmotion regions. Since Eq. (9) has two unknowns, additional constraints are required. The choice of additional constraints is a major consideration in optical flow estimation and has been proposed by many researchers in motion analysis. Lee *et al.* proposed a line clustering algorithm which is a modified and faster version of the original algorithm by Schunck [172]. By thresholding the image velocities, moving regions of the face are obtained. Because the extracted regions do not exactly define the complete face area, an ellipse fitting algorithm is employed to complete the face region extraction.

3.1.5. Generalized measures. Visual features such as edges, color, and motion are derived in the early stage of the human visual system, shown by the various visual response patterns in our inner retina [190, 206]. This pre-attentive processing allows visual information to be organized in various bases prior to high-level visual activities in the brain. Based on this observation, Reisfeld *et al.* [153] proposed that machine vision systems should begin with pre-attentive low-level computation of generalized image properties. In their earlier work, Reisfeld and Yeshurun [154] introduced a generalized symmetry operator that is based on edge pixel operation. Since facial features are symmetrical in nature, the operator which does not rely on higher level a priori knowledge of the face effectively produces a representation that gives high responses to facial feature locations. The symmetry measure assigns a magnitude at every pixel location in an image based on the contribution of surrounding pixels. The symmetry magnitude, $M_\sigma(p)$, for pixel p is described as

$$M_\sigma(p) = \sum_{(i,j)\in\Gamma(p)} C(i, j), \tag{10}$$

where $C(i, j)$ is the contribution of the surrounding pixel i and j (of pixel p) in the set of pixels defined by $\Gamma(p)$. Both the contribution and the pixel set are defined as in the following equations,

$$\begin{aligned} \Gamma(p) &= \left[(i, j) \,\middle|\, \frac{p_i + p_j}{2} = p\right] \\ C(i, j) &= D_\sigma(i, j)P(i, j)r_i r_j, \end{aligned} \tag{11}$$

where $D(i, j)$ is a distance weight function, $P(i, j)$ is a phase weight function, and r_i and r_j are defined as below,

$$\begin{aligned} D_\sigma(i, j) &= \tfrac{1}{\sqrt{2\pi}\sigma} e^{-\frac{\|p_i - p_j\|}{2\sigma}} \\ P(i, j) &= (1 - \cos(\theta_i + \theta_j - 2\alpha_{ij}))(1 - \cos(\theta_i - \theta_j)) \\ r_k &= \log\left(1 + \|\nabla_{p_k}\|\right) \\ \theta_k &= \arctan\left[\frac{\frac{\partial}{\partial y} p_k}{\frac{\partial}{\partial x} p_k}\right], \end{aligned} \tag{12}$$

where p_k is any point (x_k, y_k) where $k = 1, \ldots, K$, ∇_{p_k} is the gradient of the intensity at point p_k, and α_{ij} is the counterclockwise angle between the line passing through p_i and p_j and the horizon [154]. Figure 4 shows the example of $M_\sigma(i, j)$ computed from the gradient of a frontal facial image. The symmetry magnitude map clearly shows the locations of facial

FIG. 4. Original image and M_σ of original image Reisfeld *et al.* [153].

features such as the eyes and mouth. Using the magnitude map, Reisfeld *et al.* managed a success rate of 95% in detecting eyes and mouths of various similar faces in a database. The face images in this database also contain various backgrounds and orientations. A more recent work by Reisfeld and Yeshurun is [155].

Later development in generalized symmetry includes [86, 94] and the work by Lin and Lin *et al.* [111]. Lin and Lin proposed a dual-mask operator that exploits the radial symmetry distribution of θ_k (see Eq. (12)) on bright and dark facial parts. Similar to Reisfeld *et al.*'s operator, the dual-mask operator also managed to extract facial features from different backgrounds and under various poses, but the complexity of the latter is reported to be less than the original symmetry operator [111]. A new attentional operator based on smooth convex and concave shapes was presented recently by Tankus *et al.* [185]. Different from previous approaches, they make use of the derivative of gradient orientation, θ_k, with respect to the y-direction which is termed *Y-Phase*. It was shown by Tankus *et al.* that the Y-Phase of concave and convex objects (paraboloids) have a strong response at the negative x-axis. Because facial features generally appear to be parabolic, their Y-Phase response will also resemble that of the paraboloids, giving a strong response at the x-axis. By proposing a theorem and comparing the Y-Phase of $\log(\log(\log(I)))$ and $\exp(\exp(\exp(I)))$ where I is the image, Tankus *et al.* also proved that Y-Phase is invariant under a very large variety of illumination conditions. Further experiments also indicate that the operator is insensitive to strong edges from nonconvex objects and texture backgrounds.

3.2. Feature Analysis

Features generated from low-level analysis are likely to be ambiguous. For instance, in locating facial regions using a skin color model, background objects of similar color can also be detected. This is a classical many to one mapping problem which can be solved by higher level feature analysis. In many face detection techniques, the knowledge of face geometry has been employed to characterize and subsequently verify various features from their ambiguous state. There are two approaches in the application of face geometry among the literature surveyed. The first approach involves sequential *feature searching* strategies based on the relative positioning of individual facial features. The confidence of a feature

existence is enhanced by the detection of nearby features. The techniques in the second approach group features as flexible *constellations* using various face models.

3.2.1. Feature searching. Feature searching techniques begin with the determination of prominent facial features. The detection of the prominent features then allows for the existence of other less prominent features to be hypothesized using anthropometric measurements of face geometry. For instance, a small area on top of a larger area in a head and shoulder sequence implies a "face on top of shoulder" scenario, and a pair of dark regions found in the face area increase the confidence of a face existence. Among the literature survey, a pair of eyes is the most commonly applied reference feature [5, 27, 52, 61, 207, 214] due to its distinct side-by-side appearance. Other features include a main face axis [26, 165], outline (top of the head) [26, 32, 162] and body (below the head) [192, 207].

The facial feature extraction algorithm by De Silva *et al.* [32] is a good example of feature searching. The algorithm starts by hypothesizing the top of a head and then a searching algorithm scans downward to find an eye-plane which appears to have a sudden increase in edge densities (measured by the ratio of black to white along the horizontal planes). The length between the top and the eye-plane is then used as a reference length. Using this reference length, a flexible facial template covering features such as the eyes and the mouth is initialized on the input image. The initial shape of the template is obtained by using anthropometric length with respect to the reference length (given in Table 1 [32]), obtained from the modeling of 42 frontal faces in a database. The flexible template is then adjusted to the final feature positions according to a fine-tuning algorithm that employs an edge-based cost function. The algorithm is reported to have an 82% accuracy (out of 30 images in the same database) in detecting all facial features from quasi-frontal ($<\pm 30°$) head and shoulder faces on a plain background. Although the algorithm manages to detect features of various races since it does not rely on gray and color information, it fails to detect features correctly if the face image contains eyeglasses and hair covering the forehead.

Jeng *et al.* [78] propose a system for face and facial feature detection which is also based on anthropometric measures. In their system, they initially try to establish possible locations of the eyes in binarized pre-processed images. For each possible eye pair the algorithm goes on to search for a nose, a mouth, and eyebrows. Each facial feature has an associated evaluation function, which is used to determine the final most likely face candidate, weighted by their facial importance with manually selected coefficients as shown in Eq. (13). They report a 86% detection rate on a dataset of 114 test images taken under controlled imaging conditions, but with subjects positioned in various directions with a cluttered background.

$$E = 0.5E_{eye} + 0.2E_{mouth} + 0.1E_{Reyebrow} + 0.1E_{Leyebrow} + 0.1E_{nose} \tag{13}$$

TABLE 1
Average Lengths (Times the Reference Length) of Facial Features in De Silva *et al.*'s Algorithm

	Head height	Eye separation	Eye to nose	Eye to mouth
Average length	1.972	0.516	0.303	0.556

An automatic facial features searching algorithm called *GAZE* is proposed by Herpers *et al.* [58] based on the motivation of eye movement strategies in the human visual system (HVS). At the heart of the algorithm is a local attentive mechanism that foveated sequentially on the most prominent feature location. A multi-level saliency representation is first derived using a multi-orientation Gaussian filter. The most prominent feature (with maximum saliency) is extracted using coarse to fine evaluation on the saliency map. The next step involves an enhancement step in which the saliency of the extracted area is reduced while that of the next possible feature is increased for the next iteration. By applying the algorithm iteratively on 50 high resolution frontal face images (no background included), Herpers *et al.* [58] have shown a 98% successful detection rate of eye pairs within the first three foveated steps. Other facial regions like the nose and the mouth are also detected at the later iterations. Because the test images used by the algorithm contain faces of different orientation (some faces are tilted) and slight variation in illumination conditions and scale, the high detection rate indicates that this algorithm is relatively independent of those image variations. Furthermore, unlike the algorithm described previously, it does not depend on specific measurements of facial features.

Eye movement strategies are also the basis of the algorithm proposed by Smeraldi *et al.* [177]. In [177] a description of the search targets (the eyes) is constructed by averaging Gabor responses from a retinal sampling grid centered on the eyes of the subjects in the training set. They use 2-D Gabor functions [31] of six orientations and five different frequencies for feature extraction. The smallest Gabor functions are used at the center of the sampling grid, while the largest are used off-center where there is sparser sampling. For detecting the eyes, a saccadic search algorithm is applied which consists of initially placing the sampling grid at a random position in the image and then moving it to the position where the Euclidian distance between the node of the sampling grid and the node in the search target is the smallest. The grid is moved around until the saccades become smaller than a threshold. If no target can be found (which might be the case if the search is started at a blank area in the image), a new random position is tried. Smeraldi *et al.* report correct detection of the eyes on an entire database of 800 frontal view face images. Gabor responses have also been applied to face and facial feature detection in [18, 56, 66, 146].

Other proposed approaches for feature searching include radial basis functions [71] and genetic algorithms [72, 112, 144].

3.2.2. Constellation analysis. Some of the algorithms mentioned in the last section rely extensively on heuristic information taken from various face images modeled under fixed conditions. If given a more general task such as locating the face(s) of various poses in complex backgrounds, many such algorithms will fail because of their rigid nature. Later efforts in face detection research address this problem by grouping facial features in face-like constellations using more robust modeling methods such as statistical analysis.

Various types of face constellations have been proposed [4, 11, 73, 110, 117, 131, 180, 197, 221]. Burl *et al.* [11, 12] make use of statistical shape theory on the features detected from a multi-scale Gaussian derivative filter. A probabilistic model for the spatial arrangement of facial features enables higher detection flexibility. The algorithm is able to handle missing features and problems due to translation, rotation, and scale to a certain extent. A successful rate of 84% out of 150 images taken from a lab-scene sequence, is obtained. Most detection failures are caused by significant rotation of the subject's head. Huang *et al.* [74] also apply a Gaussian filter for pre-processing in a framework based on image feature analysis. The

pre-processed images are searched with a structure model, a texture model, and a feature model for face-like patterns. In a dataset of 680 images (mostly consisting of single face images), they report only 14 images in which face detection failed.

Probabilistic face models based on multiple face appearance have also been proposed in [180, 204, 221, 224]. In Yow and Cipolla's model [221], faces are classified into several partial facial appearance groups that are common under different viewpoints. These partial facial groups are classified further into feature components. After facial features are obtained from a low-level edge based processing, active grouping then allows various facial groups to be derived hierarchically from the bottom end of the partial face classification. The grouping effectively reduced erroneous features arising from cluttered scenes. A Bayesian network then enables a gross combination of detection confidence of all the partial groups and thereby ensures the hypothesis of true faces to be reinforced with a high confidence level. A 92% detection rate was reported by Yow and Cipolla in 100 lab scene images. The algorithm is able to cope with small variations in scale, orientation, and viewpoint. The presence of eyeglasses and missing features are also handled by the algorithm. Despite some minor differences in the face model and feature detector, the system proposed by Sumi *et al.* [180] also uses similar strategies in which various facial components are processed by concurrent agents in a distributed network. Their system also produces a good rate of 94% in detecting various appearance faces.

In the system of Maio and Maltoni [120], the input images are converted to a directional image using a gradient-type operator over local windows (7×7 pixels) (see also Fröba and Küblebeck [47]). From this directional image they apply a two stage face detection method consisting of a generalized Hough transform and a set of 12 binary templates representing face constellations. The generalized Hough transform is used to generate face candidates by searching for ellipses. The face candidates are then fed to the template matching stage which accepts or rejects the face candidate as shown in Fig. 5. By efficient implementations and

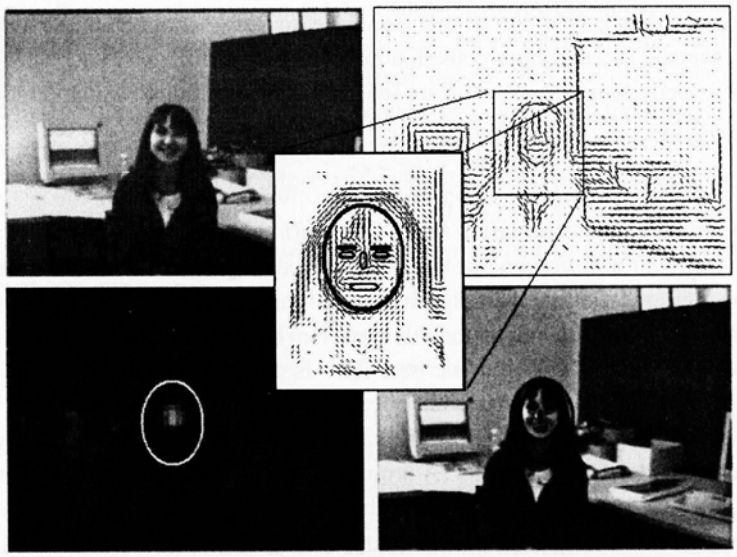

FIG. 5. The face detection system of Maio and Maltoni [120]. Reprinted from Pattern Recognition, 33 Maio, D. and Maltoni, D., Realtime face location on gray-scale static images, 1525–1539, Copyright 2000, with permission from Elsevier Science.

design considerations, the system functions in real-time. Maio and Maltoni report correct detection in 69 out of 70 test images with no false alarms, where the test images consist of single faces of varying sizes with complex backgrounds. The generalized Hough transform has also been used in the system of Schubert [171].

In face recognition systems, one of the most widely used techniques is graph matching. Graphs store local feature information in feature vectors at their nodes and geometrical information in their edges (connecting the nodes). Several recent graph matching systems perform automatic face detection [132, 138], but none have this task as the main goal of the system; thus few extensive quantitative results have been reported on the face detection task specifically. A similar approach to graph matching applied to face detection is the *Potential Net* algorithm of Bessho *et al.* [8].

3.3. Active Shape Models

Unlike the face models described in the previous sections, active shape models depict the actual physical and hence higher-level appearance of features. Once released within a close proximity to a feature, an active shape model will interact with local image features (edges, brightness) and gradually deform to take the shape of the feature. There are generally three types of active shape models in the contemporary facial extraction research. The first type uses a generic active contour called *snakes*, first introduced by Kass *et al.* in 1987 [85]. *Deformable templates* were then introduced by Yuille *et al.* [222] to take into account the a priori of facial features and to better the performance of snakes. Cootes *et al.* [24, 104] later proposed the use of a new generic flexible model which they termed *smart snakes* and PDM to provide an efficient interpretation of the human face. Cootes *et al.*'s model is based on a set of labeled points that are only allowed to vary to certain shapes according to a training procedure.

3.3.1. Snakes. Active contours, or snakes, are commonly used to locate a head boundary [55, 69, 102, 137, 209, 219]. In order to achieve the task, a snake is first initialized at the proximity around a head boundary. It then locks onto nearby edges and subsequently assume the shape of the head. The evolution of a snake is achieved by minimizing an energy function, E_{snake} (analogy with physical systems), denoted as

$$E_{snake} = E_{internal} + E_{external}, \tag{14}$$

where $E_{internal}$, $E_{external}$ are the internal and external energy functions, respectively. The internal energy is the part that depends on the intrinsic properties of the snake and defines its natural evolution. The typical natural evolution in snakes is shrinking or expanding. The external energy counteracts the internal energy and enables the contours to deviate from the natural evolution and eventually assume the shape of nearby features—the head boundary at a state of equilibria.

Two main considerations in implementing a snake are the selection of the appropriate energy terms and the energy minimization technique. Elastic energy [55, 69, 209, 219] is used commonly as internal energy. It is proportional to the distance between the control points on the snake and therefore gives the contour an elastic-band characteristic that causes it to shrink or expand. The external energy consideration is dependent on the type of image features considered. For instance, Gunn and Nixon [55] make the term sensitive to the image gradient so that the contour is convergent toward edge locations. In addition to gradient

information, the external energy term in [209, 219] includes a skin color function which attracts the contour to the face region. Energy minimization can be achieved by optimization techniques such as the steepest gradient descent. Due to the high computational requirement in the minimization process, Huang and Chen [69] and Lam and Yan [101] both employ fast iteration methods (greedy algorithms) for faster convergence.

Even though snakes are generally able to locate feature boundaries, their implementation is still plagued by two problems. Part of the contour often becomes trapped onto false image features. Furthermore, snakes are not efficient in extracting nonconvex features due to their tendency to attain minimum curvature. Gunn and Nixon addressed these problems by introducing a parameterized snake model for face and head boundary extraction. The model consists of dual integrated snakes, one expanding from within the face and the other shrinking from outside the head boundary. Initially, the evolution of both contours is governed by a parameterized model placed between them. The parameterized model biases the contours toward the target shape and thereby allows it to distinguish false image features and not be trapped by them. Once the contours reach equilibrium, the model is removed and the contours are allowed to act individually as a pair of conventional snakes, which leads to the final boundary extraction. Snakes have also been used for eyeglass detection by Saito *et al.* [161].

3.3.2. Deformable template. Locating a facial feature boundary is not an easy task because the local evidence of facial edges is difficult to organize into a sensible global entity using generic contours. The low brightness contrast around some of these features also makes the edge detection process problematic. Yuille *et al.* [222] took the concept of snakes a step further by incorporating global information of the eye to improve the reliability of the extraction process. A deformable eye template based on its salient features is parameterized using 11 parameters. Working according to the same principle as a snake, the template once initialized near an eye feature will deform itself toward optimal feature boundaries. The deformation mechanism involves the steepest gradient descent minimization of a combination of external energy due to valley, edge, peak, and image brightness (E_v, E_e, E_p, E_i) given by

$$E = E_v + E_e + E_p + E_i + E_{internal}. \tag{15}$$

All the energy terms are expressed by an integral using properties of the template such as area and length of the circle and parabolae [222]. The internal energy on the other hand is given according to the template parameters as follows:

$$E_{internal} = \frac{k_1}{2}(x_e - x_c)^2 + \frac{k_2}{2}\left(p_1 + \frac{1}{2}\{r + b\}\right)^2 + \frac{k_2}{2}\left(p_2 + \frac{1}{2}\{r + b\}\right)^2 + \frac{k_3}{2}(b - 2r)^2. \tag{16}$$

The coefficients of energy terms such as $\{k_1, k_2, k_3\}$ control the course and the deformation of the template. Changing the values of these coefficients enables a matching strategy in which the template deforms and translates around image features in different stages. Yuille *et al.* have proposed a six-epochs (per iteration) coefficient changing strategy for eye templates that are initialized below the eye. Their technique has been used as a part of an active-vision-based face authentication system proposed by Tistarelli and Grosso [189].

There are several major considerations in deformable template applications. The evolution of a deformable template is sensitive to its initial position because of the fixed matching strategy. For instance, Yuille *et al.* have shown that if the template is placed above the eye, it could be attracted toward the eyebrow instead. The processing time is also very high due to the sequential implementation of the minimization process. The weights of the energy terms are heuristic and difficult to generalize. Contemporary research in this field [7, 35, 98, 168, 174] has mainly concentrated on issues such as execution time reductions, template modifications, and energy term considerations. Shackleton and Welsh [174] improved the eye template matching accuracy by adding extra parameters and an external energy term that is sensitive to the enhanced white area of eyes. Eighty-four percent (out of 63 eye examples) of successful eye fittings are reported. The lengthy processing time is also reduced by using a simple version of the template by trading off the considerations of some parameters that have lesser effects on the overall template shape [17, 69, 211]. Chow *et al.* [17] employ a two-step approach to the eye extraction task. A circular Hough transform is first used to locate the iris prior to the fitting of a simpler eye template that only models the parabolic eye structure. This simplification has improved the run-time up to about 16 fold compared to Yuille *et al.*'s template. In a more recent development, Lam and Yan [102] make use of eye corner information to estimate the initial parameters of the eye template. The inclusion of this additional information has allowed more reliable template fitting. The time for the template to reach its optimal position has also been reduced up to 40% (compared to the original template). Besides eye templates, the use of mouth templates was also introduced [17, 69] using similar strategies.

3.3.3. Point distributed models. PDM [24] is a compact parameterized description of the shape based upon statistics. The architecture and the fitting process of PDM is different from the other active shape models. The contour of PDM is discretized into a set of labeled points. Variations of these points are first parameterized over a training set that includes objects of different sizes and poses. Using principal component analysis (PCA), variations of the features in a training set are constructed as a linear flexible model. The model comprises the mean of all the features in the sets and the principle modes of variation for each point (for further PCA description, see Section 4.1)

$$x = \bar{x} + Pv, \tag{17}$$

where x represents a point on the PDM, $\bar{x}$ is the mean feature in the training set for that point, $P = [p_1 p_2 \ldots p_t]$ is the matrix of the t most significant variation vectors of the covariance of deviations, and v is the weight vector for each mode.

Face PDM was first developed by Lanitis *et al.* [104] as a flexible model. The model depicts the global appearance of a face which includes all the facial features such as eyebrows, the nose, and eyes. Using 152 manually planted control points (x) and 160 training face images, a face PDM is obtained. Using only 16 principle weights (v of Eq. (17)), the model can approximate up to 95% of the face shapes in the training set [104]. In order to fit a PDM to a face, the mean shape model (model with labeled points $= \bar{x}$) is first placed near the face. Then a local gray-scale search strategy [25] is employed to move each point toward its corresponding boundary point. During the deformation, the shape is only allowed to change in a way which is consistent with the information modeled from the training set.

The advantages of using a face PDM is that it provides a compact parameterized description. In [105], it is implemented as a generic representation for several applications such as

coding and facial expression interpretation. In their subsequent work, Lanitis *et al.* [103] have incorporated a genetic algorithm (GA) and multiresolution approach to address the problem in multiple face candidates. The global characteristic of the model also allows all the features to be located simultaneously and thereby removes the need for feature searching. Furthermore, it has been shown that occlusion of a particular feature does not pose a severe problem since other features in the model can still contribute to a global optimal solution [103]. In [39], Edwards *et al.*'s system is further developed to include person identification in addition to face detection and tracking.

4. IMAGE-BASED APPROACH

It has been shown in the previous section that face detection by explicit modeling of facial features has been troubled by the unpredictability of face appearance and environmental conditions. Although some of the recent feature-based attempts have improved the ability to cope with the unpredictability, most are still limited to head and shoulder and quasi-frontal faces (or are included as one of the techniques in a combined system). There is still a need for techniques that can perform in more hostile scenarios such as detecting multiple faces with clutter-intensive backgrounds. This requirement has inspired a new research area in which face detection is treated as a pattern recognition problem. By formulating the problem as one of learning to recognize a face pattern from examples, the specific application of face knowledge is avoided. This eliminates the potential of modeling error due to incomplete or inaccurate face knowledge. The basic approach in recognizing face patterns is via a training procedure which classifies examples into face and non-face prototype classes. Comparison between these classes and a 2D intensity array (hence the name image-based) extracted from an input image allows the decision of face existence to be made. The simplest image-based approaches rely on template matching [62, 114], but these approaches do not perform as well as the more complex techniques presented in the following sections.

Most of the image-based approaches apply a window scanning technique for detecting faces. The window scanning algorithm is in essence just an exhaustive search of the input image for possible face locations at all scales, but there are variations in the implementation of this algorithm for almost all the image-based systems. Typically, the size of the scanning window, the subsampling rate, the step size, and the number of iterations vary depending on the method proposed and the need for a computationally efficient system.

In the following three sections we have roughly divided the image-based approaches into linear subspace methods, neural networks, and statistical approaches. In each section we give a presentation of the characteristics of some of the proposed methods, and in Section 4.4 we attempt to do a comparative evaluation based on results reported on a common dataset. Since we report results in Section 4.4 for most of the systems we describe in Sections 4.1–4.3, we have not presented the same reported results individually for each systems.

4.1. Linear Subspace Methods

Images of human faces lie in a subspace of the overall image space. To represent this subspace, one can use neural approaches (as described in the next section), but there are also several methods more closely related to standard multivariate statistical analysis which can be applied. In this section we describe and present results from some of these techniques,

including principal component analysis (PCA), linear discriminant analysis (LDA), and factor analysis (FA).

In the late 1980s, Sirovich and Kirby [176] developed a technique using PCA to efficiently represent human faces. Given an ensemble of different face images, the technique first finds the principal components of the distribution of faces, expressed in terms of eigenvectors (of the covariance matrix of the distribution). Each individual face in the face set can then be approximated by a linear combination of the largest eigenvectors, more commonly referred to as eigenfaces, using appropriate weights.

Turk and Pentland [192] later developed this technique for face recognition. Their method exploits the distinct nature of the weights of eigenfaces in individual face representation. Since the face reconstruction by its principal components is an approximation, a residual error is defined in the algorithm as a preliminary measure of "faceness." This residual error which they termed "distance-from-face-space" (DFFS) gives a good indication of face existence through the observation of global minima in the distance map.

The basic procedure for computing the face space and DFFS is as follows [192]: We have a dataset of n face images, $\Gamma_1, \Gamma_2, \ldots, \Gamma_n$. The average face is defined by

$$\Psi = \frac{1}{n} \sum_{i=1}^{n} \Gamma_i. \tag{18}$$

The average face is then subtracted from each image and the image is vectorized:

$$\Phi_i = (\Gamma_i - \Psi)^v. \tag{19}$$

Let $D = [\Phi_1 \Phi_2 \cdots \Phi_n]$ and $C = DD^t$. The eigenvectors $\boldsymbol{u}_i$ of C are called the principal components of D, and when converted back to matrices we can view these vectors as the eigenfaces of the dataset. Some examples of these eigenfaces are shown in Fig. 6 (ordered according to eigenvalue). These eigenfaces span the subspace called face space.

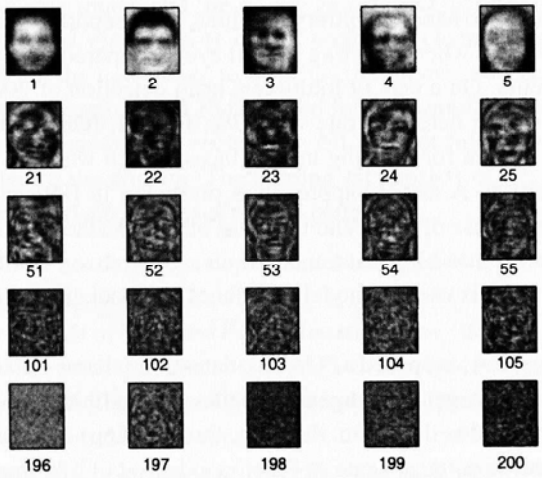

FIG. 6. Some examples of eigenfaces computed from the ORL dataset mentioned earlier (the number below each image indicates the principal component number, ordered according to eigenvalues).

A preprocessed input image Φ can be projected onto face space by

$$\omega_k = \mathbf{u}_k^t \Phi, \qquad k = 1, \ldots, m, \tag{20}$$

where m is the number of principal components selected to span face space. m should be smaller than n, since the principal components with the smallest corresponding eigenvalues do not carry significant information for representation (consider eigenface number 200 in Fig. 6). Φ can be reconstructed by

$$\Phi_r = \sum_{k=1}^{m} \omega_k \mathbf{u}_k. \tag{21}$$

The reconstruction error $\epsilon = \|\Phi - \Phi_r\|^2$ is the DFFS.

Pentland *et al.* [133, 142] later proposed a facial feature detector using DFFS generated from eigenfeatures (eigeneyes, eigennose, eigenmouth) obtained from various facial feature templates in a training set. The feature detector also has a better ability to account for features under different viewing angles since features of different discrete views were used during the training. The performance of the eye locations was reported to be 94% with 6% false positive [142] in a database of 7562 frontal face images on a plain background. One hundred twenty-eight faces are sampled from the database to compute a set of corresponding eigenfeatures. A slightly reduced but still accurate performance for nose and mouth locations was also shown in [142]. The DFFS measure has also been used for facial feature detection in [33] and in combination with Fisher's linear discriminant [43] for face and facial feature detection [173]. A comparison of DFFS with a morphological multiscale fingerprint algorithm is provided in Raducanu and Grana [149].

More recently, Moghaddam and Pentland have further developed this technique within a probabilistic framework [134]. When using PCA for representation, one normally discards the orthogonal complement of face space (as mentioned earlier). Moghaddam and Pentland found that this leads to the assumption that face space has a uniform density, so they developed a maximum likelihood detector which takes into account both face space and its orthogonal complement to handle arbitrary densities. They report a detection rate of 95% on a set of 7000 face images when detecting the left eye. Compared to the DFFS detector, this was significantly better. On a task of multiscale head detection of 2000 face images from the FERET database, the detection rate was 97%. In [77], Jebara and Pentland included this technique in a system for tracking human faces, which was also based on color, 3D, and motion information. A similar approach is presented in [89] where PCA is applied for modeling both the class of faces and the class of pseudo-faces (nonfaces, but face-like patterns), together with matching criteria based on a generalized likelihood ratio. In Meng and Nguyen [128], PCA is used to model both faces and background clutter (an eigenface and an eigenclutter space).

Samal and Iyengar [164] proposed a PCA face detection scheme based on face silhouettes. Instead of eigenfaces, they generate eigensilhouettes and combine this with standard image processing techniques (edge detection, thinning, thresholding) and the generalized Hough transform. They report a detection rate of 92% on a dataset of 129 images (66 face images and 63 general images) where, in the case of face images, the faces occupy most of the image.

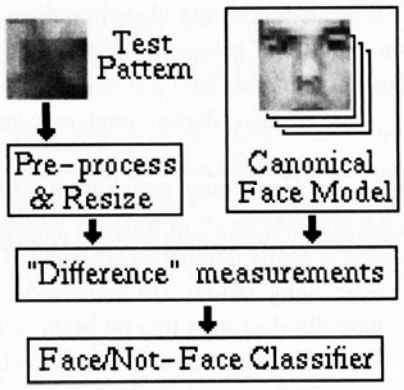

FIG. 7. Classification in Sung and Poggio's [182] system (© 1998/2001 IEEE).

PCA is an intuitive and appropriate way of constructing a subspace for representing an object class in many cases. However, for modeling the manifold of face images, PCA is not necessarily optimal. Face space might be better represented by dividing it into subclasses, and several methods have been proposed for this, of which most are based on some mixture of multidimensional Gaussians. This technique was first applied for face detection by Sung and Poggio [182]. Their method consists mainly of four steps (Fig. 7):

1. The input sub-image is pre-processed by re-scaling to 19×19 pixels, applying a mask for eliminating near-boundary pixels, subtracting a best-fit brightness plane from the unmasked window pixels, and finally applying histogram equalization (Fig. 9).
2. A distribution-based model of canonical face- and nonface-patterns is constructed. The model consists of 12 multi-dimensional Gaussian clusters with a centroid location and a covariance matrix, where six represent face- and six represent no-face pattern prototypes. The clusters are constructed by an elliptical k-means clustering algorithm which uses an adaptively changing normalized Mahalanobis distance metric.
3. A set of image measurements is computed for new images relative to the canonical face model. For each of the clusters, two values are computed. One is a Mahalanobis-like distance between the new image and the prototype centroid, defined within the subspace spanned by the 75 largest eigenvectors of the prototype cluster, while the other is the Euclidean distance from the new image to its projection in the subspace.
4. A multi-layer perceptron (MLP) is trained for face–nonface classification from the 24-dimensional image measurement vector. The MLP is not fully connected, but exploits some prior knowledge of the domain. The training set consists of 47,316 image measurements vectors, where 4150 are examples of face patterns.

When a new image is to bc classified, steps 1 and 3 are applied and the MLP provides the classification.

A similar approach to that of Sung and Poggio based on gray-level features in combination with texture features has been explored in [38] and a similar more computationally efficient method has been proposed by Fouad *et al.* [44]. Gu and Li [54] present a variation of Sung and Poggio's approach where linear discriminant analysis is applied for feature selection before training the neural network classifier. A method following the framework of Sung and Poggio has also been proposed by Rajagopalan *et al.* [150] using higher order statistics to model the face and nonface clusters. They also present a new clustering algorithm using higher order

statistics which replaces a k-means clustering algorithm from Sung and Poggio's work. They report good results on a subset of images from the CMU database using the higher order statistics technique and compare this with a system based on hidden Markov models which does not perform as well. Hidden Markov models have also been applied to face detection in [121, 129].

One issue which arises when training pattern recognition systems for face–nonface classification is how to collect a representable set of training samples for nonface images. The set of positive training samples is easily defined as all kinds of face images, but to define the complementary set is harder. Sung and Poggio suggested a training algorithm, known as "boot-strap training," to partially deal with this problem (a more precise strategy than the one proposed in [10]). The algorithm consists of the following steps:

1. create the initial set of nonface images simply by generating images of random pixels,
2. train the system,
3. run the system on scenes not containing faces and extract the false positives, and
4. pre-process the false positives and add them to the training set of nonfaces; go to step 2.

Yang *et al.* [217] proposed two methods for face detection which also seek to represent the manifold of human faces as a set of subclasses. The first method is based on FA, which is a multivariate statistical technique quite similar to PCA, but in contrast to PCA, FA assumes that the observed data samples come from a well-defined model [122]

$$\mathbf{x} = \Lambda\mathbf{f} + \mathbf{u} + \mu, \tag{22}$$

where Λ is a matrix of constants, $\mathbf{f}$ and $\mathbf{u}$ are random vectors, and μ is the mean. Factor analysis seeks to find Λ and the covariance matrix of $\mathbf{u}$ which best models the covariance structure of $\mathbf{x}$. If the specific variances $\mathbf{u}$ are assumed to be 0, the procedure for FA can be equivalent to PCA. Yang *et al.* use a mixture of factor analyzers in their first method. Given a set of training images, the EM algorithm [34] is used to estimate the parameters in the mixture model of factor analyzers [49]. This mixture model is then applied to subwindows in the input image and outputs the probability of a face being present at the current location.

In the second method, Yang *et al.* use Kohonen's self-organizing map (SOM) [93] to divide the training images into 25 face and 25 nonface classes (the number of classes chosen based on the size of the training set). The training images are then used to compute a LDA, which is a dimensionality reduction technique similar to PCA. Instead of computing a transformation matrix (the $[\boldsymbol{u}_1^t \boldsymbol{u}_2^t \ldots \boldsymbol{u}_m^t]$ from Eq. (20)) aimed for representation of the entire dataset (PCA), LDA seeks to find a transformation matrix which is based on maximizing between class scatter and minimizing within class scatter. The eigenvector in the LDA transformation matrix with the largest eigenvalue is known as Fisher's linear discriminant [43], which by itself has also been used for face detection [179, 203]. PCA is aimed at representation, while LDA aims for discrimination and is therefore appropriate for face detection when the class of faces and nonfaces is divided into subclasses. With this method the input images are projected onto a 49-dimensional space ($25 + 25 - 1$), in which Gaussian distributions are used to model each class conditional density function $P(z \mid X_i)$, where z is the projected image and $i = 1, \ldots, 49$. The mean and covariance matrix for each class are maximum likelihood estimates. A face is detected or not in a subwindow of the input image based on

the maximum likelihood decision rule:

$$X^* = \arg \max_{X_i} P(z \mid X_i). \tag{23}$$

Both of the methods of Yang *et al.* use the window scanning technique with a 20×20 window and scan the input images for 10 iterations with a subsampling factor of 1.2. Other face detection approaches using LDA include [65] and [184], while SOMs have been used in [183]. The likelihood decision rule has also recently been applied in a different subspace face detection system based on wavelet packet decomposition [223].

4.2. Neural Networks

Neural networks have become a popular technique for pattern recognition problems, including face detection. Neural networks today are much more than just the simple MLP. Modular architectures, committee–ensemble classification, complex learning algorithms, autoassociative and compression networks, and networks evolved or pruned with genetic algorithms are all examples of the widespread use of neural networks in pattern recognition. For face recognition, this implies that neural approaches might be applied for all parts of the system, and this had indeed been shown in several papers [14, 113, 163]. An introduction to some basic neural network methods for face detection can be found in Viennet and Fougelman Soulié [195].

The first neural approaches to face detection were based on MLPs [10, 82, 147], where promising results where reported on fairly simple datasets. The first advanced neural approach which reported results on a large, difficult dataset was by Rowley *et al.* [158]. Their system incorporates face knowledge in a retinally connected neural network shown in Fig. 8. The neural network is designed to look at windows of 20×20 pixels (thus 400 input units). There is one hidden layer with 26 units, where 4 units look at 10×10 pixel subregions, 16 look at 5×5 subregions, and 6 look at 20×5 pixels overlapping horizontal stripes. The input window is pre-processed through lighting correction (a best fit linear function is subtracted) and histogram equalization. This pre-processing method was adopted from Sung and Poggio's system mentioned earlier and is shown in Fig. 9. A problem that arises with window scanning techniques is overlapping detections. Rowley *et al.* deals with this problem through two heuristics:

1. Thresholding: the number of detections in a small neighborhood surrounding the current location is counted, and if it is above a certain threshold, a face is present at this location.

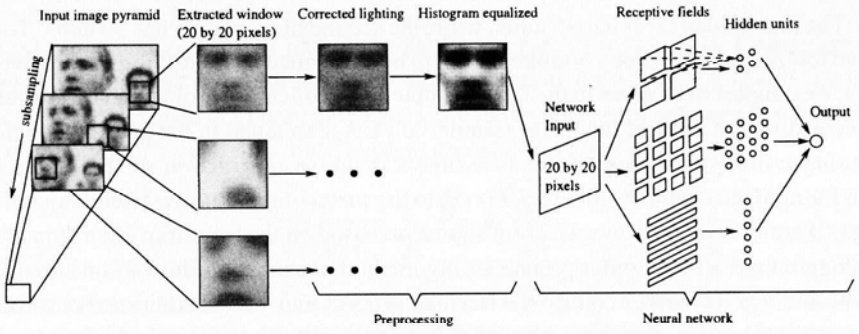

FIG. 8. The system of Rowley *et al.* [158] (© 1998/2001 IEEE).

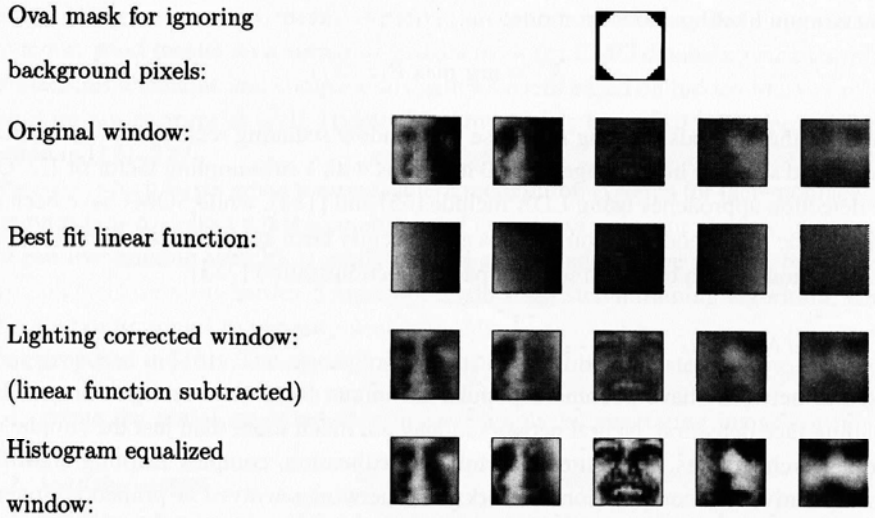

FIG. 9. The preprocessing method applied by Rowley *et al.* [158] (© 1998/2001 IEEE) and Sung and Poggio. A linear function is fit to the intensity values in the window (inside the oval mask) and then subtracted out (from the entire window). Then histogram equalization is applied to improve contrast.

2. Overlap elimination: when a region is classified as a face according to thresholding, then overlapping detections are likely to be false positives and thus are rejected.

To further improve performance, they train multiple neural networks and combine the output with an arbitration strategy (ANDing, ORing, voting, or a separate arbitration neural network). This algorithm was applied in a person tracking system in [30] and [166] and for initial face detection in the head tracking system of La Cascia *et al.* [99]. A similar system was recently proposed in [57].

Recently, Rowley *et al.* [159] combined this system with a router neural network to detect faces at all angles in the image plane. They use a fully connected MLP with one hidden layer and 36 output units (one unit for each 10° angle) to decide the angle of the face. The system detects 79.6% of the faces in two large datasets with a small number of false positives.

In Feraud *et al.* [41] a different neural approach is suggested, based on a constrained generative model (CGM). Their CGM is an autoassociative fully connected MLP with three layers of weights, with 300 (15×20) input and output units (corresponding to the size of the image). The first hidden layer has 35 units, while the second hidden layer has 50 units. The idea behind this model is to force a nonlinear PCA to be performed by modifying the projection of nonface examples to be close to the face examples. Classification is obtained by considering the reconstruction error of the CGM (similar to PCA, explained in the previous section).

During training, the target for a face-image is the reconstruction of the image itself, while for nonface examples, the target is set to the mean of the n nearest neighbors of face-images. Feraud *et al.* employ a training algorithm based on the bootstrap algorithm of Sung and Poggio (and also a similar preprocessing method consisting of histogram equalization and smoothing). To further control the learning process they use an additional cost function based on the minimum description length (MDL) principle. The system is further developed

in [42] to include color information and multiple views and applied to the problem of finding face images on the Web.

In Lin *et al.* [113], a fully automatic face recognition system is proposed based on probabilistic decision-based neural networks (PDBNN). A PDBNN is a classification neural network with a hierarchical modular structure. The network is similar to the DBNN [97], but it has an additional probabilistic constraint. The network consists of one subnet for each object class, combined with a winner-take-all strategy. For the case of face detection, the network has only one subnet representing the face class. Training is performed with DBNN learning rules, which means that the teacher only tells the correctness of the classification (no exact target values) and LUGS (locally unsupervised globally supervised) learning is applied. With LUGS, each subnet is trained individually with an unsupervised training algorithm (K-mean and vector quantization or the EM algorithm). The global training is performed to fine-tune decision boundaries by employing reinforced or antireinforced learning when a pattern in the training set is misclassified. The input images are originally 320×240 (from the MIT dataset [182]), but are scaled down to approximately 46×35, and a 12×12 window is used to scan this image (search step 1 pixel).

A new learning architecture called SNoW (sparse network of winnows) [156] is applied to face detection in Roth *et al.* [157]. SNoW for face detection is a neural network consisting of two linear threshold units (LTU) (representing the classes of faces and nonfaces). The two LTUs operate on an input space of Boolean features. The best performing system derives features from 20×20 subwindows in the following way: for 1×1, 2×2, 4×4, and 10×10 subsubwindows, compute {`position` $\times$ `intensity mean` $\times$ `intensity variance`}. This gives Boolean features in a 135,424-dimensional feature space, since the mean and variance have been discretized into a predefined number of classes. The LTUs are separate from each other and are sparsely connected over the feature space. The system is trained with a simple learning rule which promotes and demotes weights in cases of misclassification. Similar to the previously mentioned methods, Roth *et al.* use the bootstrap method of Sung and Poggio for generating training samples and preprocess all images with histogram equalization.

Apart from face classification, neural networks have also been applied for facial features classification [36, 125, 151, 196] and a method for improving detection time for MLPs is presented in [6].

4.3. Statistical Approaches

Apart from linear subspace methods and neural networks, there are several other statistical approaches to face detection. Systems based on information theory, a support vector machine and Bayes' decision rule are presented in this section.

Based on an earlier work of maximum likelihood face detection [21], Colmenarez and Huang [22] proposed a system based on Kullback relative information (Kullback divergence). This divergence is a nonnegative measure of the difference between two probability density functions P_{X^n} and M_{X^n} for a random process X^n:

$$H_{P\|M} = \sum_{X^n} P_{X^n} \ln \frac{P_{X^n}}{M_{X^n}}. \tag{24}$$

During training, for each pair of pixels in the training set, a join-histogram is used to create probability functions for the classes of faces and nonfaces. Since pixel values are

highly dependent on neighboring pixel values, X^n is treated as a first order Markov process and the pixel values in the gray-level images are requantized to four levels. Colmenarez and Huang use a large set of 11×11 images of faces and nonfaces for training, and the training procedure results in a set of look-up tables with likelihood ratios. To further improve performance and reduce computational requirements, pairs of pixels which contribute poorly to the overall divergency are dropped from the look-up tables and not used in the face detection system. In [23], Colmenarez and Huang further improved on this technique by including error bootstrapping (described in Section 4.2), and in [20] the technique was incorporated in a real-time face tracking system. A similar system has been developed by Lew and Huijsmans [107].

In Osuna *et al.* [140], a support vector machine (SVM) [194] is applied to face detection. The proposed system follows the same framework as the one developed by Sung and Poggio [182], described in Section 4.1 (scanning input images with a 19×19 window). A SVM with a 2nd-degree polynomial as a kernel function is trained with a decomposition algorithm which guarantees global optimality. Training is performed with the boot-strap learning algorithm, and the images are pre-processed with the procedure shown in Fig. 9. Kumar and Poggio [96] recently incorporated Osuna *et al.*'s SVM algorithm in a system for real-time tracking and analysis of faces. They apply the SVM algorithm on segmented skin regions in the input images to avoid exhaustive scanning. SVMs have also been used for multiview face detection by constructing separate SVMs for different parts of the view sphere [136]. In Terrillon *et al.* [186], SVMs improved the performance of the face detector compared to the earlier use of a multi-layer perceptron.

Schneiderman and Kanade [169, 170] describe two face detectors based on Bayes' decision rule (presented as a likelihood ratio test in Eq. (25), where italics indicate a random variable).

$$\frac{P(image \mid \text{object})}{P(image \mid \text{non-object})} > \frac{P(\text{non-object})}{P(\text{object})} \tag{25}$$

If the likelihood ratio (left side) of Eq. (25) is greater than the right side, then it is decided that an object (a face) is present at the current location. The advantage with this approach is that if the representations for $P(image \mid \text{object})$ and $P(image \mid \text{non-object})$ are accurate, the Bayes decision rule is proven to be optimal [37]. In the first proposed face detection system of Schneiderman and Kanade [169], the posterior probability function is derived based on a set of modifications and simplifications (some of which are mentioned here):

- the resolution of a face image is normalized to 64×64 pixels
- the face images are decomposed into 16×16 subregions and there is no modeling of statistical dependency among the subregions
- the subregions are projected onto a 12-dimensional subspace (constructed by PCA)
- the entire face region is normalized to have zero mean and unit variance.

In the second proposed system [170], the visual attributes of the image are not represented by local eigenvector coefficients (as in the first approach), but instead by a locally sampled wavelet transform. A wavelet transform can capture information regarding visual attributes in space, frequency, and orientation and thus should be well suited for describing the characteristics of the human face. The wavelet transform applied in [170] is a three-level decomposition using a 5/3 linear phase filter-bank. This transform decomposes the image

FIG. 10. Face detection examples from Schneiderman and Kanade [170] (© 2000/2001 IEEE).

into 10 subbands. From these subbands 17 visual attributes (each consisting of 8 coefficients) are extracted and treated as statistical independent random variables. The coefficients are requantized to three levels and the visual attributes are represented using histograms. With this approach, a view-based detector is developed with a frontal view detector and a right profile detector (to detect left profile images, the right profile detector is applied to mirror reversed images). The best results from the two systems described here are obtained with the eigenvector system, but this is due to the dataset consisting of mostly frontal view faces. In a separate experiment on a dataset consisting mostly of profile view faces, the wavelet detector outperformed the eigenvector detected (which of course had been modified to detect profile views also). Some examples of processed images with the wavelet system are shown in Fig. 10. Bayes' decision rule has also been applied for face detection by Qian and Huang [148].

4.4. Comparative Evaluation

Since some of the image-based systems report results on the same datasets, we have compared performance in Table 2. The datasets have been collected at CMU by Rowley *et al.* [158] and at MIT by Sung and Poggio [182]. The CMU dataset also includes the MIT dataset. Since some systems report results when excluding some of the images from the datasets, we had to consider each set in two separate columns in Table 2. Thus, we ended up with the following four datasets:

CMU-130: The entire CMU dataset with 130 images with 507 labeled faces (which includes the MIT dataset). The images are grayscale and of varying size and quality.

CMU-125: The CMU dataset excluding hand-drawn and cartoon faces for a total of 125 images with 483 labeled faces. There are additional faces in this dataset (for a total of over 495 faces), but the ground truth of 483 has been established based on excluding some of the occluded faces and nonhuman faces. However, the ground truth of 483 established by Rowley *et al.* does not necessarily refer to the same 483 faces in all the reported results since at least one of the papers [157] indicate that they have labeled the set of faces themselves.

MIT-23: The entire MIT dataset (also known as test set B of the CMU dataset) with 23 images. The number of labeled faces was originally 149 by Sung and Poggio (which is the number used in the reported results in [182] and [107], but it was changed to 155 (which

TABLE 2
Results Reported in Terms of Percentage Correct Detection (CD) and Number of False Positives (FP), CD/FP, on the CMU and MIT Datasets

Face detection system	CMU-130	CMU-125	MIT-23	MIT-20
Schneiderman & Kanade—E[a] [170]		94.4%/65		
Schneiderman & Kanade—W[b] [170]		90.2%/110		
Yang *et al.*—FA [217]		92.3%/82		89.4%/3
Yang *et al.*—LDA [217]		93.6%/74		91.5%/1
Roth *et al.* [157]		94.8%/78		94.1%/3
Rowley *et al.* [158]	86.2%/23		84.5%/8	
Feraud *et al.* [42]	86%/8			
Colmenarez & Huang [22]	93.9%/8122			
Sung & Poggio [182]			79.9%/5	
Lew & Huijsmans [107]			94.1%/64	
Osuna *et al.* [140]			74.2%/20	
Lin *et al.* [113]			72.3%/6	
Gu and Li [54]			87.1%/0	

[a] Eigenvector coefficients.
[b] Wavelet coefficients.

is the number used in the reported results in [158], [140], and [113]) when included in the CMU dataset.

MIT-20: The MIT dataset excluding hand-drawn and cartoon faces for a total of 20 images with 136 labeled faces.

It is hard to draw any conclusions from Table 2 due to inaccurate or incomplete information in the respective publications. There are several problems which need to be addressed when detecting faces with the window scanning technique on such complex data as the CMU and MIT datasets. We have tried to summarize some of these problems we have encountered:

How does one count correct detections and false positives? Since a detection is a placement of a window at a location in the input image, we need to decide how accurate this placement needs to be. Yang *et al.* [217] make this decision based on the rule "... a detected face is a successful detect if the subimage contains eyes and mouth." For all systems using the window scanning technique, a face might be detected at several scales and at several positions close to each other. Rowley *et al.* and Gu and Li address this problem by using two merging heuristics (mentioned in Section 4.2), while few others seem to care about this. This is probably due to the fact that correct detections and false positives are manually counted in the tests.

What is the system's ROC curve? It is well known that some systems have high detection rates while others have a low number of false positives. Most systems can adjust their threshold depending on how conservative one needs the system to be. This can be reported in terms of a ROC curve to show the correct detections–false positives trade-off.

What is the size of the training set and how is training implemented? Some systems use a large training set and generate additional training samples by rotating, mirroring, and adding noise, while others have a smaller training set. Also, the proposed bootstrap training algorithm by Sung and Poggio is implemented in some of the systems.

What is a face? Since the CMU dataset contain a large number of faces, there seems to be some disagreement of how many faces the dataset actually contains. This is due to the fact there are human faces, animal faces, cartoon faces, and line-drawn faces present in the dataset.

Since many of these questions are left unanswered in most of the papers, this introduces a degree of uncertainty in the systems performance. Without ROC curves it is hard to tell how the systems are affected by parameter adjustments. Also the number of false positives is dependent upon the number of subwindows tested with the window scanning technique, which makes it hard to evaluate the false detection rate.

However, all the results presented in Table 2 are quite impressive considering the complexity of the CMU and MIT datasets.

5. APPLICATIONS

Face detection technology can be useful and necessary in a wide range of applications, including biometric identification, video conferencing, indexing of image and video databases, and intelligent human–computer interfaces. In this section we give a brief presentation of some of these applications.

As mentioned earlier, face detection is most widely used as a preprocessor in face recognition systems. Face recognition is one of many possible approaches to biometric identification; thus many biometric systems are based on face recognition in combination with other biometric features such as voice or fingerprints. In BioID [46], a model-based face detection method based on edge information [28] is used as a preprocessor in a biometric systems which combines face, voice, and lip movement recognition. Other biometric systems using face detection include template-based methods in the CSIRO PC-Check system [145] and eigenface methods [142, 192] in FaceID from Viisage Technologies.

With the increasing amount of digital images available on the Internet and the use of digital video in databases, face detection has become an important part of many content-based image retrieval (CBIR) systems. The neural network-based face detection system of Rowley *et al.* [158] is used as a part of an image search engine for the World Wide Web in WebSeer [45]. The idea behind CBIR systems using face detection is that faces represent an important cue for image content; thus digital video libraries consisting of terabytes of video and audio information have also percieved the importance of face detection technology. One such example is the Informedia project [198] which provides search and retrieval of TV news and documentary broadcasts. Name-It [167] also processes news videos, but is focused on associating names with faces. Both systems use the face detection system of Rowley *et al.* [158].

In video conferencing systems, there is a need to automatically control the camera in such a way that the current speaker always has the focus. One simple approach to this is to guide the camera based on sound or simple cues such as motion and skin color. A more complex approach is taken by Wang *et al.* [199], who propose an automatic video conferencing system consisting of multiple cameras, where decisions involving which camera to use are based on an estimate of the head's gazing angle. Human faces are detected using features derived from motion, contour geometry, color, and facial analysis. The gazing angle is computed based on the hairline (the border between an individuals hair and skin). Approaches which focus on single camera control include LISTEN [19],

SAVI [60], LAFTER [139], Rits Eye [212], and the system of Morimoto and Flickner [135].

6. SUMMARY AND CONCLUSIONS

We have presented an extensive survey of feature-based and image-based algorithms for face detection, together with a brief presentation of some of the application areas. The following is a concise summary with conclusions representing the main topics from this paper.

- Face detection is currently a very active research area and the technology has come a long way since the survey of Chellappa *et al.* [14]. The last couple of years have shown great advances in algorithms dealing with complex environments such as low quality gray-scale images and cluttered backgrounds. Some of the best algorithms are still too computationally expensive to be applicable for real-time processing, but this is likely to change with coming improvements in computer hardware.
- Feature-based methods are applicable for real-time systems where color and motion is available. Since an exhaustive multiresolution window scanning is not always preferable, feature-based methods can provide visual cues to focus attention. In these situations, the most widely used technique is skin color detection based on one of the color models mentioned in Section 3.1.3. Out of the feature-based approaches which perform on gray-scale static images, Maio and Maltoni's [120] algorithm seems very promising, showing good detection results while still being computationally efficient.
- Image-based approaches are the most robust techniques for processing gray-scale static images. Sung and Poggio [182] and Rowley *et al.* [158] set the standards for research on this topic, and the performances of their algorithms are still comparable to more recent image-based approaches. All these algorithms are based on multiresolution window scanning to detect faces at all scales, making them computationally expensive. Multiresolution window scanning can be avoided by combining the image-based approach with a feature-based method as a preprocessor with the purpose of guiding the search based on visual clues such as skin color.
- The most important application for face detection is still as a preprocessor in face recognition systems. For offline processing, face detection technology has reached a point where the detection of a single face in an image with fair resolution (typical for a face recognition system) is close to being a solved problem. However, accurate detection of facial features such as the corners of the eyes and mouth is more difficult, and this is still a hard problem to solve. Face detection has also found its way to CBIR systems such as Web search engines and digital video indexing.
- As shown in Section 4.4, it is not easy to evaluate and compare current algorithms. The MIT and CMU datasets provide some basis for comparison, but since there are no standard evaluation procedures or agreements on the number of faces in the dataset, it is hard to draw any conclusions. There is a need for an evaluation procedure for face detection algorithms similar to the FERET [143] test for face recognition.
- The human face is a dynamic object but with a standard configuration of facial features which can vary within a limited range. It is a difficult problem to detect such dynamic objects and considering the changes in faces over time (facial hair, glasses, wrinkles, skin color, bruises) together with variations in pose, developing a robust face detection algorithm is still a hard problem to solve in computer vision systems.

ACKNOWLEDGMENTS

The authors are grateful to Professor H. Yeshurun for providing Fig. 4, Dr. D. Maltoni for providing Fig. 5, Dr. K.-K. Sung for providing Fig. 7, Dr. H. Rowley for providing Figs. 8 and 9, and Dr. H. Schneiderman for providing Fig. 10. The authors thank Professor F. Albregtsen, Dr. J. Wroldsen, and Dr. O. Lund for useful discussions during the preparation of this manuscript.

REFERENCES

1. M. Abdel-Mottaleb and A. Elgammal, Face detection in complex environments, in *Proceedings International Conference on Image Processing, 1999.*
2. A. Albiol, A simple and efficient face detection algorithm for video database applications, in *Proceedings of the 2000 International Conference on Image Processing, 2000,* p. TA07.09.
3. A. Albiol, C. A. Bouman, and E. J. Delp, Face detection for pseudosemantic labeling in video databases, in *Proceedings International Conference on Image Processing, 1999.*
4. Y. Amit, D. Geman, and B. Jedynak, Efficient focusing and face detection, in *Face Recognition: From Theory to Application,* Springer-Verlag, Berlin/New York, 1998.
5. P. J. L. Van Beek, M. J. T. Reinders, B. Sankur, and J. C. A. Van Der Lubbe, Semantic segmentation of videophone image sequences, in *Proc. of SPIE Int. Conf. on Visual Communications and Image Processing, 1992,* pp. 1182–1193.
6. S. Ben-Yacoub, B. Fasel, and J. Lüttin, Fast face detection using MLP and FFT, in *Proceedings Second International Conference on Audio- and Video-based Biometric Person Authentication (AVBPA), 1999.*
7. D. E. Benn, M. S. Nixon, and J. N. Carter, Extending concentricity analysis by deformable templates for improved eye extraction, in *Proceedings Second International Conference on Audio- and Video-based Biometric Person Authentication (AVBPA), 1999.*
8. H. Bessho, Y. Iwai, and M. Yachida, Detecting human face and recognizing facial expressions using potential net, in *Proceedings of the 15th International Conference on Pattern Recognition, 2000,* Vol. I, p. 3A.
9. R. Brunelli and T. Poggio, Face recognition: Feature versus templates, *IEEE Trans. Pattern Anal. Mach. Intell.* **15**, 1993, 1042–1052.
10. G. Burel and D. Carel, Detection and localization of faces on digital images, *Pattern Recog. Lett.* **15**, 1994, 963–967.
11. M. C. Burl, T. K. Leung, and P. Perona, Face localization via shape statistics, in *Int. Workshop on Automatic Face and Gesture Recognition, Zurich, Switzerland, June 1995.*
12. M. C. Burl and P. Perona, Recognition of planar object classes, in *IEEE Proc. of Int. Conf. on Computer Vision and Pattern Recognition, 6, 1996.*
13. J. Cai and A. Goshtasby, Detecting human faces in color images, *Image Vision Comput.* **18**, 1999, 63–75.
14. R. Chellappa, C. L. Wilson, and S. Sirohey. Human and machine recognition of faces: A survey, *Proc. IEEE* **83**, 5, 1995.
15. C. Chen and S. P. Chiang, Detection of human faces in color images, *IEE Proc. Vision Image Signal Process.* **144**, 1997, 384–388.
16. J. Choi, S. Kim, and P. Rhee, Facial components segmentation for extracting facial feature, in *Proceedings Second International Conference on Audio- and Video-based Biometric Person Authentication (AVBPA), 1999.*
17. G. Chow and X. Li, Towards a system for automatic facial feature detection, *Pattern Recog.* **26**, 1993, 1739–1755.
18. S. Clippingdale and T. Ito, A unified approach to video face detection—tracking and recognition, in *Proceedings International Conference on Image Processing, 1999.*
19. M. Collobert, R. Feraud, G. Le Tourneur, O. Bernier, J. E. Viallet, Y. Mahieux, and D. Collobert, LISTEN: A system for locating and tracking individual speaker, in *IEEE Proc. of 2nd Int. Conf. on Automatic Face and Gesture Recognition, Vermont, 1996.*

20. A. J. Colmenarez, B. Frey, and T. S. Huang, Detection and tracking of faces and facial features, in *Proceedings International Conference on Image Processing, 1999.*

21. A. J. Colmenarez and T. S. Huang, Maximum likelihood face detection, in *IEEE Proc. of 2nd Int. Conf. on Automatic Face and Gesture Recognition, Vermont, 1996*, pp. 222–224.

22. A. J. Colmenarez and T. S. Huang, Face detection with information-based maximum discrimination, in *IEEE Proc. of Int. Conf. on Computer Vision and Pattern Recognition, 6 1997.*

23. A. J. Colmenarez and T. S. Huang, Pattern detection with information-based maximum discrimination and error bootstrapping, in *Proc. of International Conference on Pattern Recognition, 1998.*

24. T. F. Cootes and C. J. Taylor, Active shape models—'smart snakes,' in *Proc. of British Machine Vision Conference, 1992*, pp. 266–275.

25. T. F. Cootes, C. J. Taylor, and A. Lanitis, Active shape models: Evaluation of a multi-resolution method for improving image search, in *Proc. of British Machine Vision Conference, 1994*, pp. 327–336.

26. I. Craw, H. Ellis, and J. R. Lishman, Automatic extraction of face-feature, *Pattern Recog. Lett.* Feb. 1987, 183–187.

27. J. L. Crowley and F. Berard, Multi-model tracking of faces for video communications, in *IEEE Proc. of Int, Conf. on Computer Vision and Pattern Recognition, Puerto Rico, Jun. 1997.*

28. G. A. Klanderman, D. P. Huttenlocher, and W. J. Rucklidge, Comparing images using the Hausdorff distance, *IEEE Trans. Pattern Anal. Mach. Intell.* 1993, 850–863.

29. Y. Dai and Y. Nakano, Face-texture model based on sgld and its application, *Pattern Recog.* **29**, 1996, 1007–1017.

30. T. Darrell, G. Gordon, M. Harville, and J. Woodfill, Integrated person tracking using stereo, color, and pattern detection, in *IEEE Proc. of Int. Conf. on Computer Vision and Pattern Recognition, 1998.*

31. J. G. Daugman, Complete discrete 2-D gabor transforms by neural networks for image analysis and compression, *IEEE Trans. Acoustics, Speech Signal Process.* **36**, 1988, 1169–1179.

32. L. C. De Silva, K. Aizawa, and M. Hatori, Detection and tracking of facial features by using a facial feature model and deformable circular template, *IEICE Trans. Inform. Systems* **E78–D(9)**, 1995, 1195–1207.

33. H. Demirel, T. J. Clarke, and P. J. K. Cheung, Adaptive automatic facial feature segmentation, in *IEEE Proc. of 2nd Int. Conf. on Automatic Face and Gesture Recognition, Vermont, Oct. 1996*, pp. 277–282.

34. A. P. Dempster, N. M. Laird, and D. B. Rubin, Maximum likelihood from incomplete data via the EM algorithm, *J. Roy. Statist. Soc.* **39**, 1977, 1–39.

35. J.-Y. Deng and F. Lai, Recognition-based template deformation and masking for eye-feature extraction and description, *Pattern Recog.* **30**, 1997, 403–419.

36. P. Duchnowski, M. Hunke, D. Busching, U. Meier, and A. Waibel, Toward movement invariant automatic lip-reading and speech recognition, in *Proc. Int. Conf. on Acoustic, Speech and Signal Processing, 1995.*

37. R. O. Duda and P. E. Hart, *Pattern Classification and Scene Analysis*, Wiley, New York, 1973.

38. N. Duta and A. K. Jain, Learning the human face concept from black and white images, in *Proc. of International Conference on Pattern Recognition, 1998.*

39. G. J. Edwards, C. J. Taylor, and T. F. Cootes, Learning to identify and track faces in image sequences, in *IEEE Proc. of 3rd Int. Conf. on Automatic Face and Gesture Recognition, 1998.*

40. L. G. Frakas and I. R. Munro, *Anthropometric Facial Proportions in Medicine*. Charles C. Thomas, Springfield, IL, 1987.

41. R. Feraud, O. Bernier, and D. Collobert, A constrained generative model applied to face detection, *Neural Process. Lett.* **5**, 1997, 73–81.

42. R. Feraud, O. Bernier, J.-E. Viallet, and M. Collobert, A fast and accurate face detector for indexation of face images, in *Proceedings Fourth IEEE International Conference on Automatic Face and Gesture Recognition, 2000.*

43. R. A. Fisher, The use of multiple measurements in taxonomic problems, *Ann. Eugenics* **7**, 1936, 179–188.

44. M. Fouad, A. Darwish, S. Shaheen, and F. Bayoumi, Mode-based human face detection in unconstrained scenes, in *Proceedings of the 2000 International Conference on Image Processing, 2000*, p. TA07.06.

45. C. Frankel, M. J. Swain, and V. Athitsos, *WebSeer: An Image Search Engine for the World Wide Web*, Technical Report TR 96-14, Computer Science Department, Univ. of Chicago, 1996.

46. R. W. Frischolz and U. Dieckmann, BioID: A multimodal biometric identification system, *IEEE Comput.* **33**(2), 2000.

47. B. Fröba and C. Küblbeck, Orientation template matching for face localization in complex visual scenes, in *Proceedings of the 2000 International Conference on Image Processing, 2000*, p. TA07.12.

48. C. Garcia and G. Tziritas, Face detection using quantized skin color regions, merging and wavelet packet analysis, *IEEE Trans. Multimedia* **1**, 1999, 264–277.

49. Z. Ghahramani and G. Hinton, *The EM Algorithm for Mixtures of Factor Analyzers*, Technical Report CRG-TR-96-1, Dept. of Computer Science, University of Toronto, 1996.

50. V. Govindaraju, Locating human faces in photographs, *Int. J. Comput. Vision* **19**, 1996.

51. H. Graf, E. Cosatto, and T. Ezzat, Face analysis for the synthesis of photo-realistic talking heads, in *Proceedings Fourth IEEE International Conference on Automatic Face and Gesture Recognition, 2000*.

52. H. P. Graf, T. Chen, E. Petajan, and E. Cosatto, Locating faces and facial parts, in *IEEE Proc. of Int. Workshop on Automatic Face-and Gesture-Recognition, Zurich, Switzerland, Jun. 1995*, pp. 41–45.

53. H. P. Graf, E. Cosatto, D. Gibson, E. Petajan, and M. Kocheisen, Multi-modal system for locating heads and faces, in *IEEE Proc. of 2nd Int. Conf. on Automatic Face and Gesture Recognition, Vermont*, Oct. 1996, pp. 277–282.

54. Q. Gu and S. Z. Li, Combining feature optimization into neural network based face detection, in *Proceedings of the 15th International Conference on Pattern Recognition, 2000*, Vol. II, p. 4A.

55. S. R. Gunn and M. S. Nixon, A dual active contour for head and boundary extraction, in *IEE Colloquium on Image Processing for Biometric Measurement, London, Apr. 1994*, pp. 6/1.

56. T. Hamada, K. Kato, and K. Kawakami, Extracting facial features as in infants, *Pattern Recog. Lett.* **21**, 2000, 407–412.

57. C. C. Han, H. Y. M. Liao, G. J. Yu, and L. H. Chen, Fast face detection via morphology-based pre-processing, *Pattern Recog.* **33**, 2000.

58. R. Herpers, H. Kattner, H. Rodax, and G. Sommer, Gaze: An attentive processing strategy to detect and analyze the prominent facial regions, in *IEEE Proc. of Int. Workshop on Automatic Face- and Gesture-Recognition, Zurich, Switzerland, Jun. 1995*, pp. 214–220.

59. R. Herpers, M. Michaelis, K.-H. Lichtenauer, and G. Sommer, Edge and keypoint detection in facial regions, in *IEEE Proc. of 2nd Int. Conf. on Automatic Face and Gesture Recognition*, Vermont, Oct. 1996, pp. 212–217.

60. R. Herpers, G. Verghese, K. Derpanis, R. McCready, J. MacLean, A. Jepson, and J. K. Tsotsos, Detection and tracking of faces in real environments, in *Proc. IEEE International Workshop on Recognition, Analysis, and Tracking of Faces and Gestures in Real-Time Systems, 1999*.

61. E. Hjelmås and J. Wroldsen, Recognizing faces from the eyes only, in *Proceedings of the 11th Scandinavian Conference on Image Analysis, 1999*.

62. G. Holst, Face detection by facets: Combined bottom-up and top-down search using compound templates, in *Proceedings of the 2000 International Conference on Image Processing, 2000*, p. TA07.08.

63. H. Hongo, M. Ohya, M. Yasumoto, Y. Niwa, and K. Yamamoto, Focus of attention for face and hand gesture recognition using multiple cameras, in *Proceedings Fourth IEEE International Conference on Automatic Face and Gesture Recognition, 2000*.

64. R. Hoogenboom and M. Lew, Face detection using local maxima, in *IEEE Proc. of 2nd Int. Conf. on Automatic Face and Gesture Recognition, Vermont, Oct. 1996*, pp. 334–339.

65. K. Hotta, T. Kurita, and T. Mishima, Scale invariant face detection method using higher-order local autocorrelation features extracted from log-polar image, in *IEEE Proc. of 3rd Int. Conf. on Automatic Face and Gesture Recognition, 1998*.

66. 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*.

67. J. Hu, H. Yan, and M. Sakalli, Locating head and face boundaries for headshoulder images, *Pattern Recog.* **32**, 1999, 1317–1333.

68. M. K. Hu, Visual pattern recognition by moment invariants, *IRE Transactions on Information Theory* **8**, 1962, 179–187.

69. C. L. Huang and C. W. Chen, Human facial feature extraction for face interpretation and recognition, *Pattern Recog.* **25**, 1992, 1435–1444.

70. J. Huang, S. Gutta, and H. Wechsler, Detection of human faces using decision trees, in *IEEE Proc. of 2nd Int. Conf. on Automatic Face and Gesture Recognition, Vermont, 1996.*

71. J. Huang and H. Wechsler, Eye detection using optimal wavelet packets and radial basis functions (rbfs), *Int. J. Pattern Recog. Artificial Intell.* **13**, 1999.

72. J. Huang and H. Wechsler, Visual routines for eye location using learning and evolution, in *IEEE Transactions on Evolutionary Computation, 1999.*

73. W. Huang and R. Mariani, Face detection and precise eyes location, in *Proceedings of the 15th International Conference on Pattern Recognition, 2000*, Vol. IV, p. 3B.

74. W. Huang, Q. Sun, C. P. Lam, and J. K. Wu, A robust approach to face and eyes detection from images with cluttered background, in *Proc. of International Conference on Pattern Recognition, 1998.*

75. M. Hunke and A. Waibel, Face locating and tracking for human-computer interaction, in *28th Asilomar Conference on Signals, Systems and Computers, Monterey, CA, 1994.*

76. A. Jacquin and A. Eleftheriadis, Automatic location tracking of faces and facial features in video sequences, in *IEEE Proc. of Int. Workshop on Automatic Face- and Gesture-Recognition, Zurich, Switzerland, Jun. 1995.*

77. T. S. Jebara and A. Pentland, Parametrized structure from motion for 3D adaptive feedback tracking of faces, in *IEEE Proc. of Int. Conf. on Computer Vision and Pattern Recognition, Puerto Rico, 1997.*

78. S. H. Jeng, H. Y. M. Liao, C. C. Han, M. Y. Chern, and Y. T. Liu, Facial feature detection using geometrical face model: An efficient approach, *Pattern Recog.* **31**, 1998.

79. X. Jiang, M. Binkert, B. Achermann, and H. Bunke, Towards detection of glasses in facial images, *Pattern Anal. Appl.* **3**, 2000, 9–18.

80. Z. Jing and R. Mariani, Glasses detection and extraction by deformable contour, in *Proceedings of the 15th International Conference on Pattern Recognition, 2000*, Vol. II, p. 4B.

81. L. Jordão, M. Perrone, and J. P. Costeira, Active face and feature tracking, in *Proceedings of the 10th International Conference on Image Analysis and Processing, 1999.*

82. P. Juell and R. Marsh, A hierarchical neural network for human face detection, *Pattern Recog.* **29**, 1996, 781–787.

83. M. Kapfer and J. Benois-Pineau, Detection of human faces in color image sequences with arbitrary motions for very low bit-rate videophone coding, *Pattern Recog. Lett.* **18**, 1997.

84. J. Karlekar and U. B. Desai, Finding faces in color images using wavelet transform, in *Proceedings of the 10th International Conference on Image Analysis and Processing, 1999.*

85. M. Kass, A. Witkin, and D. Terzopoulos, Snakes: active contour models, in *Proc. of 1st Int Conf. on Computer Vision, London, 1987.*

86. S. Katahara and M. Aoki, Face parts extraction window based on bilateral symmetry of gradient information, in *Proceedings 8th International Conference on Computer Analysis of Images and Patterns (CAIP), 1999.*

87. T. Kawaguchi, D. Hidaka, and M. Rizon, Robust extraction of eyes from face, in *Proceedings of the 15th International Conference on Pattern Recognition, 2000*, Vol. I, p. 3A.

88. S. Kawato and J. Ohya, Real-time detection of nodding and head-shaking by directly detecting and tracking the "between-eyes," in *Proceedings Fourth IEEE International Conference on Automatic Face and Gesture Recognition, 2000.*

89. C. Kervrann, F. Davoine, P. Pérez, R. Forchheimer, and C. Labit, Generalized likelihood ratio-based face detection and extraction of mouth features, *Pattern Recog. Lett.* **18**, 1997.

90. S.-H. Kim and H.-G. Kim, Face detection using multi-modal information, in *Proceedings Fourth IEEE International Conference on Automatic Face and Gesture Recognition, 2000.*

91. S.-H. Kim, N.-K. Kim, S. C. Ahn, and H.-G. Kim, Object oriented face detection using range and color information, in *IEEE Proc. of 3rd Int. Conf. on Automatic Face and Gesture Recognition, 1998.*

92. L. H. Koh, S. Ranganath, M. W. Lee, and Y. V. Venkatesth, An integrated face detection and recognition system, in *Proceedings of the 10th International Conference on Image Analysis and Processing, 1999*.

93. T. Kohonen, *Self-Organizing Maps*, Springer-Verlog, Berlin, 1995.

94. T. Kondo and H. Yan, Automatic human face detection and recognition under nonuniform illumination, *Pattern Recog.* **32**, 1999, 1707–1718.

95. C. Kotropoulos and I. Pitas, Rule-based face detection in frontal views, in *Proc. Int. Conf. on Acoustic, Speech and Signal Processing, 1997*.

96. V. Kumar and T. Poggio, Learning-based approach to real time tracking and analysis of faces, in *Proceedings Fourth IEEE International Conference on Automatic Face and Gesture Recognition, 2000*.

97. S. Y. Kung and J. S. Taur, Decision-based neural networks with signal/image classification applications, *IEEE Trans. Neural Networks* **6**, 1995, 170–181.

98. Y. H. Kwon and N. da Vitoria Lobo, Face detection using templates, in *Proceedings of the 12th International Conference on Pattern Recognition* **1994**, pp. A:764–767.

99. M. LaCascia, S. Sclaroff, and V. Athitsos, Fast, reliable head tracking under varying illumination: An approach based on registration of textured-mapped 3D models, *IEEE Trans. Pattern Anal. Mach. Intell.* **22**, 2000, 322–336.

100. K. M. Lam and H. Yan, Facial feature location and extraction for computerised human face recognition, in *Int. Symposium on information Theory and Its Applications, Sydney, Australia, Nov. 1994*.

101. K. M. Lam and H. Yan, Fast greedy algorithm for locating head boundaries, *Electron. Lett.* **30**, 1994, 21–22.

102. K. M. Lam and H. Yan, Locating and extracting the eye in human face images, *Pattern Recog.* **29**, 1996, 771–779.

103. A. Lanitis, A. Hill, T. Cootes, and C. Taylor, Locating facial features using genetics algorithms, in *Proc. of Int. Conf. on Digital Signal Processing, Limassol, Cyrus, 1995*, pp. 520–525.

104. A. Lanitis, C. J. Taylor, and T. F. Cootes, Automatic tracking, coding and reconstruction of human faces, using flexible appearance models, *IEEE Electron. Lett.* **30**, 1994, 1578–1579.

105. A. Lanitis, C. J. Taylor, and T. F. Cootes, Automatic interpretation and coding of face images using flexible models, *IEEE Trans. Pattern Anal. Mach. Intell.* **19**, 1997.

106. C. H. Lee, J. S. Kim, and K. H. Park, Automatic human face location in a complex background, *Pattern Recog.* **29**, 1996, 1877–1889.

107. M. S. Lew and N. Huijsmans, Information theory and face detection, in *Proc. of International Conference on Pattern Recognition, 1996*.

108. X. Li and N. Roeder, Face contour extraction from front-view images, *Pattern Recog.* **28**, 1995.

109. Y. Li, A. Goshtasby, and O. Garcia, Detecting and tracking human faces in videos, in *Proceedings of the 15th International Conference on Pattern Recognition, 2000*, Vol. I, p. 2A.

110. C. Lin and K. Fan, Human face detection using geometric triangle relationship, in *Proceedings of the 15th International Conference on Pattern Recognition, 2000*, Vol. II, p. 4B.

111. C. C. Lin and W. C. Lin, Extracting facial features by an inhibitory mechanism based on gradient distributions, *Pattern Recog.* **29**, 1996, 2079–2101.

112. C. H. Lin and J. L. Wu, Automatic facial feature extraction by genetic algorithms, *IEEE Trans. Image Process.* **8**, 1999, 834–845.

113. S.-H. Lin, S.-Y. Kung, and L.-J. Lin, Face recognition/detection by probabilistic decision-based neural network, *IEEE Trans. Neural Networks* **8**, 1997, 114–132.

114. Z. Liu and Y. Wang, Face detection and tracking in video using dynamic programming, in *Proceedings of the 2000 International Conference on Image Processing, 2000*, pp. MA02.08.

115. B. K. Low, *Computer Extraction of Human Faces*, PhD thesis, Dept. of Electronic and Electrical Engineering, De Montfort University, 1998.

116. B. K. Low and M. K. Ibrahim, A fast and accurate algorithm for facial feature segmentation, in *Proceedings International Conference on Image Processing, 1997*.

117. J. Luo, C. W. Chen, and K. J. Parker, Face location in wavelet-based video compression for high perceptual quality videoconferencing, *IEEE Trans. Circuits Systems Video Technol.* **6**(4), 1996.

118. F. Luthon and M. Lievin, Lip motion automatic detection, in *Scandinavian Conference on Image Analysis, Lappeenranta, Finland, 1997.*

119. X.-G. Lv, J. Zhou, and C.-S. Zhang, A novel algorithm for rotated human face detection, in *IEEE Conference on Computer Vision and Pattern Recognition, 2000.*

120. D. Maio and D. Maltoni, Real-time face location on gray-scale static images, *Pattern Recog.* **33**, 2000, 1525–1539.

121. S. Marchand-Maillet and B. Mérialdo, Pseudo two-dimensional hidden markov models for face detection in colour images, in *Proceedings Second International Conference on Audio- and Video-based Biometric Person Authentication (AVBPA), 1999.*

122. K. V. Mardia, J. T. Kent, and J. M. Bibby, *Multivariate Analysis*, Academic Press, San Diego, 1979.

123. F. Marqués and V. Vilaplana, A morphological approach for segmentation and tracking of human faces, in *Proceedings of the 15th International Conference on Pattern Recognition, 2000*, Vol. I, P. 3A.

124. D. Marr and E. Hildreth, Theory of edge detection, in *Proc. of the Royal Society of London, 1980.*

125. S. McKenna, S. Gong, and J. J. Collins, Face tracking and pose representation, in *British Machine Vision Conference, Edinburgh, Scotland, Sept. 1996.*

126. S. McKenna, S. Gong, and H. Liddell, Real-time tracking for an integrated face recognition system, in *2nd Workshop on Parallel Modelling of Neural Operators, Faro, Portugal, Nov. 1995.*

127. S. J. McKenna, S. Gong, and Y. Raja, Modelling facial colour and identity with Gaussian mixtures, *Pattern Recog.* **31**, 1998.

128. L. Meng and T. Nguyen, Two subspace methods to discriminate faces and clutters, in *Proceedings of the 2000 International Conference on Image Processing, 2000*, p. TA07.03.

129. L. Meng, T. Nguyen, and D. A. Castañon, An image-based bayesian framework for face detection, in *IEEE Conference on Computer Vision and Pattern Recognition, 2000.*

130. B. Menser and M. Brunig, Segmentation of human faces in color images using connected operators, in *Proceedings International Conference on Image Processing, 1999.*

131. J. Miao, B. Yin, K. Wang, L. Shen, and X. Chen, A hierarchical multiscale and multiangle system for human face detection in a complex background using gravitycenter template, *Pattern Recog.* **32**, 1999, 1237–1248.

132. A. R. Mirhosseini, H. Yan, K.-M. Lam, and T. Pham, Human face image recognition: An evidence aggregation approach, *Computer Vision and Image Understanding* **71**, 1998, doi:10.1006/cviu.1998.0710.

133. B. Moghaddam and A. Pentland, Face recognition using view-based and modular eigenspaces, in *Automatic Systems for the Identification of Humans, SPIE, 1994*, Vol. 2277.

134. B. Moghaddam and A. Pentland, Probabilistic visual learning for object representation, *IEEE Trans. Pattern Anal. Mach. Intell.* **19**(1), 1997.

135. C. Morimoto and M. Flickner, Real-time multiple face detection using active illumination, in *Proceedings Fourth IEEE International Conference on Automatic Face and Gesture Recognition, 2000.*

136. J. Ng and S. Gong, Performing multi-view face detection and pose estimation using a composite support vector machine across the view sphere, in *Proc. IEEE International Workshop on Recognition, Analysis, and Tracking of Faces and Gestures in Real-Time Systems, 1999.*

137. A. Nikolaidis and I. Pitas, Facial feature extraction and pose determination, *Pattern Recog.* **33**, 2000, 1783–1791.

138. K. Okada, J. Steffens, T. Maurer, H. Hong, E. Elagin, H. Neven, and C. v. d. Malsburg, The bochum/USC face recognition system and how it fared in the FERET phase III test, in *Face Recognition: From Theory to Application*, Springer-Verlag, Berlin/New York, 1998.

139. N. Oliver, A. Pentland, and F. Bérard, LAFTER: A real-time face and lips tracker with facial expression recognition, *Pattern Recog.* **33**, 2000, 1369–1382.

140. E. Osuna, R. Freund, and F. Girosi, Training support vector machines: An application to face detection, in *IEEE Proc. of Int. Conf. on Computer Vision and Pattern Recognition, 6, 1997.*

141. T. Pavlidis, *Graphics and Image Processing*, Computer Science Press, Rockville, MD, 1982.

142. A Pentland, B. Moghaddam, and T. Strarner, View-based and modular eigenspaces for face recognition, in *IEEE Proc. of Int. Conf. on Computer Vision and Pattern Recognition, 1994*.

143. J. P. Phillips, H. Wechsler, J. Huang, and P. Rauss, The FERET database and evaluation procedure for face-recognition algorithms, *Image Vision Comput.* **16**(5), 1998.

144. R. Pinto-Elias and J. H. Sossa-Azuela, Automatic facial feature detection and location, in *Proc. of International Conference on Pattern Recognition, 1998*.

145. G. E. Poulton, N. A. Oakes, D. G. Geers, and R.-Y. Qiao, The CSIRO PC-check system, in *Proceedings Second International Conference on Audio- and Video-based Biometric Person Authentication (AVBPA), 1999*.

146. D. Pramadihanto, Y. Iwai, and M. Yachida, A flexible feature matching for automatic face and facial feature points detection, in *Proc. of International Conference on Pattern Recognition, 1998*.

147. M. Propp and A. Samal, Artificial neural network architecture for human face detection, *Intell. Eng. Systems Artificial Neural Networks* **2**, 1992, 535–540.

148. R. Qian and T. Huang, Object detection using hierarchical mrf and map estimation, in *IEEE Proc. of Int. Conf. on Computer Vision and Pattern Recognition, 1997*.

149. B. Raducanu and M. Grana, Face localization based on the morphological multiscale fingerprint, in *Proceedings of the 15th International Conference on Pattern Recognition, 2000*, Vol. II, p. 4B.

150. A. N. Rajagopalan, K. S. Kumar, J. Karlekar, R. Manivasakan, M. M. Patil, U. B. Desai, P. G. Poonacha, and S. Chaudhuri, Finding faces in photographs, in *Proceedings of International Conference on Computer Vision, 1998*.

151. M. J. T. Reinders, R. W. C. Koch, and J. J. Gerbrands, Tracking facial features in image sequences using neural networks, in *IEEE Proc. of 2nd Int. Conf. on Automatic Face and Gesture Recognition, Vermont, Oct. 1996*, pp. 230–235.

152. M. J. T. Reinders, P. J. L. van Beek, B. Sankur, and J. C. A. van der Lubbe, Facial feature localization and adaptation of a generic face model for model-based coding, in *Signal Processing: Image Communication*, 1995, pp. 57–74.

153. D. Reisfeld, H. Wolfson, and Y. Yeshurun, Context-free attentional operators: The generalized symmetry transform, *Int. J. Comput. Vision* **14**, 1995, 119–130.

154. D. Reisfeld and Y. Yeshurun, Robust detection of facial features by generalised symmetry, in *Proc. of 11th Int. Conf. on Pattern Recognition, The Hague, The Netherlands, August 1992*, pp. A117–120.

155. D. Reisfeld and Y. Yeshurun, Preprocessing of face images: Detection of features and pose normalization, *Comput. Vision Image Understanding* **71**, 1998. doi:10.1006/cviu.1997.0640.

156. D. Roth, "The SNoW Learning Architecture," Technical Report UIUCDCS-R-99-2102, UIUC Computer Science Department, 1999.

157. D. Roth, M.-H. Yang, and N. Ahuja, A SNoW-based face detector, in *Advances in Neural Information Processing Systems 12 (NIPS 12)*, MIT Press, Cambridge, MA, 2000.

158. H. A. Rowley, S. Baluja, and T. Kanade, Neural network-based face detection, *IEEE Trans. Pattern Anal. Mach. Intell.* **20**, January 1998, 23–38.

159. H. A. Rowley, S. Baluja, and T. Kanade, Rotation invariant neural network-based face detection, in *Proc. IEEE Intl. Conf. on Computer Vision and Pattern Recognition, 1998*, pp. 38–44.

160. E. Saber and A. M. Tekalp, Frontal-view face detection and facial feature extraction using color, shape and symmetry based cost functions, *Pattern Recog. Lett.* **19**, 1998.

161. Y. Saito, Y. Kenmochi, and K. Kotani, Extraction of a symmetric object for eyeglass face analysis using active contour model, in *Proceedings of the 2000 International Conference on Image Processing, 2000*, p. TA07.07.

162. T. Sakai, M. Nagao, and T. Kanade, Computer analysis and classification of photographs of human faces, in *Proc. First USA—Japan Computer Conference, 1972*, p. 2.7.

163. A Samal and P. A. Iyengar, Automatic recognition and analysis of human faces and facial expressions: a survey, *Pattern Recog.* **25**, 1992, 65–77.

164. A. Samal and P. A. Iyengar, Human face detection using silhouttes, *Int. J. Pattern Recog. Artificial Intell.* **9**(6), 1995.

165. M. U. Ramos Sanchez, J. Matas, and J. Kittler, Statistical chromaticity models for lip tracking with b-splines, in *Int. Conf. on Audio- and Video-Based Biometric Person Authentication, Crans Montana, Switzerland, 1997.*

166. S. Satoh, Comparative evaluation of face sequence matching for content-based video access, in *Proceedings Fourth IEEE International Conference on Automatic Face and Gesture Recognition, 2000.*

167. S. Satoh, Y. Nakamura, and T. Kanade, Name-It: Naming and detecting faces in news videos, *IEEE Multimedia* **6**, 1999, 22–35.

168. B. Scasselatti, Eye finding via face detection for a foveated, active vision system, in *Proceedings AAAI Conference, 1998.*

169. H. Schneiderman and T. Kanade, Probabilistic modeling of local appearance and spatial relationships for object recognition, in *IEEE Conference on Computer Vision and Pattern Recognition, 6, 1998.*

170. H. Schneiderman and T. Kanade, A statistical model for 3D object detection applied to faces and cars, in *IEEE Conference on Computer Vision and Pattern Recognition, 2000.*

171. A. Schubert, Detection and tracking of facial features in real time using a synergistic approach of spatio-temporal models and generalized hough-transform techniques, in *Proceedings Fourth IEEE International Conference on Automatic Face and Gesture Recognition, 2000.*

172. B. G. Schunck, Image flow segmentation and estimation by constraint line clustering, *IEEE Trans. Pattern Anal. Mach. Intell.* **11**, 1989, 1010–1027.

173. A. W. Senior, Face and feature finding for a face recognition system, in *Proceedings Second International Conference on Audio- and Video-based Biometric Person Authentication (AVBPA), 1999.*

174. A. Shackleton and W. J. Welsh, Classification of facial features for recognition, in *IEEE Proc. of Int. Conf.* on Computer Vision and Pattern Recognition, *Hawaii, 1991*, pp. 573–579.

175. B. E. Shpungin and J. R. Movellan, *A Multi-threaded Approach to Real Time Face Tracking,* Technical Report TR 2000.02, UCSD MPLab, 2000.

176. L. Sirovich and M. Kirby, Low-dimensional procedure for the characterization of human faces, *J. Opt. Soc. Amer.* **4**, 1987, 519–524.

177. F. Smeraldi, O. Carmona, and J. Bigün, Saccadic search with Gabor features applied to eye detection and real-time head tracking, *Image Vision Comput.* **18**, 2000, 323–329.

178. K. Sobottka and I. Pitas, Extraction of facial regions and features using color and shape information, in *Proc. of Int. Conf. on Pattern Recognition, 1996.*

179. Q. Song and J. Robinson, A feature space for face image processing, in *Proceedings of the 15th International Conference on Pattern Recognition, 2000*, Vol II, p. 2B.

180. Y. Sumi and Y. Ohta, Detection of face orientation and facial components using distributed appearance modelling, in *IEEE Proc. of Int. Workshop on Automatic Face- and Gesture-Recognition, Zurich, Switzerland, Jun. 1995*, pp. 254–255.

181. Q. B. Sun, W. M. Huang, and J. K. Wu, Face detection based on color and local symmetry information, in *IEEE Proc. of 3rd Int. Conf. on Automatic Face and Gesture Recognition, 1998.*

182. K.-K. Sung and T. Poggio, Example-based learning for view-based human face detection, *IEEE Trans. Pattern Anal. Mach. Intelligence* **20**, 1998, 39–51.

183. B. Takacs and H. Wechsler, Detection of faces and facial landmarks using iconic filter banks, *Pattern Recog.* **30**, 1997.

184. M. Tanaka, K. Hotta, T. Kurita, and T. Mishima, Dynamic attention map by ising model for human face detection, in *Proc. of International Conference on Pattern Recognition, 1998.*

185. A. Tankus, H. Yeshurun, and N. Intrator, Face detection by direct convexity estimation, in *Proc. of the 1st Int. Conf. on Audio- and Video-based Biometric Person Authentication, Crans-Montana, Switzerland, 1997.*

186. J. Terrillon, M. Shirazi, M. Sadek, H. Fukamachi, and S. Akamatsu, Invariant face detection with support vector machines, in *Proceedings of the 15th International Conference on Pattern Recognition, 2000*, Vol. IV, p. 4B.

187. J.-C. Terrillon, M. David, and S. Akamatsu, Automatic detection of human faces in natural scene images by use of a skin color model and of invariant moments, in *IEEE Proc. of 3rd Int. Conf. on Automatic Face and Gesture Recognition, 1998.*

188. J.-C. Terrillon, M. Shirazi, H. Fukamachi, and S. Akamatsu, Comparative performance of different skin chrominance models and chrominance spaces for the automatic detection of human faces in color images, in *Proceedings Fourth IEEE International Conference on Automatic Face and Gesture Recognition, 2000*.

189. M. Tistarelli and E. Grosso, Active vision-based face authentication, *Image Vision Comput.* **18**, 2000, 299–314.

190. N. Treisman, Preattentive processing in vision, *Comput. Vision, Graphics Image Process.* **31**, 1985, 156–177.

191. N. Tsapatsoulis, Y. Avrithis, and S. Kollias, Efficient face detection for multimedia applications, in *Proceedings of the 2000 International Conference on Image Processing, 2000*, p. TA07.11.

192. M. Turk and A. Pentland, Eigenfaces for recognition, *J. Cog. Neurosci.* **3**, 1991, 71–86.

193. D. Valentin, H. Abdi, A. J. O'Toole, and G. Cottrell, Connectionist models of face processing: A survey, *Pattern Recog.* **27**, 1994, 1209–1230.

194. V. Vapnik, *The Nature of Statistical Learning Theory*. Springer-Verlag, New York, 1995.

195. E. Viennet and F. Fogelman Soulié, Connectionist methods for human face processing, in *Face Recognition: From Theory to Application*. Springer-Verlag, Berlin/New York, 1998.

196. J. M. Vincent, J. B. Waite, and D. J. Myers, Automatic location of visual features by a system of multilayered perceptrons, in *IEE Proceedings-F, Dec. 1992*, Vol. 139.

197. V. Vogelhuber and C. Schmid, Face detection based on generic local descriptors and spatial constraints, in *Proceedings of the 15th International Conference on Pattern Recognition, 2000*, Vol. I, p. 3A.

198. H. D. Wactlar, T. Kanade, M. A. Smith, and S. M. Stevens, Intelligent access to digital video: Informedia project, *IEEE Comput.* **29**(5), 1996, 46–52.

199. C. Wang, S. M. Griebel, and M. S. Brandstein, Robust automatic video-conferencing with multiple cameras and microphones, in *Proc. IEEE International Conference on Multimedia and Expo, 2000*.

200. H. Wang and S.-F. Chang, A highly efficient system for automatic face region detection in MPEG video, *IEEE Trans. Circuits Systems Video Technol.* **7**, 1997, 615–628.

201. J. Wang and T. Tan, A new face detection method based on shape information, *Pattern Recog. Lett.* **21**, 2000, 463–471.

202. J. G. Wang and E. Sung, Frontalview face detection and facial feature extraction using color and morphological operations, *Pattern Recog. Lett.* **20**, 1999, 1053–1068.

203. F. Weber and A. Herrera Hernandez, Face location by template matching with a quadratic discriminant function, in *Proc. IEEE International Workshop on Recognition, Analysis, and Tracking of Faces and Gestures in Real-Time Systems, 1999*.

204. M. Weber, W. Einhäuser, M. Welling, and P. Perona, Viewpoint-invariant learning and detection of human heads, in *Proceedings Fourth IEEE International Conference on Automatic Face and Gesture Recognition, 2000*.

205. G. Wei and I. K. Sethi, Face detection for image annotation, *Pattern Recog. Lett.* **20**, 1999, 1313–1321.

206. F. Werblin, A. Jacobs, and J. Teeters, The computational eye, in *IEEE Spectrum: Toward an Artificial Eye*, May 1996, pp. 30–37.

207. C. Wong, D. Kortenkamp, and M. Speich, A mobile robot that recognises people, in *IEEE Int. Conf. on Tools with Artificial Intelligence, 1995*.

208. H. Wu, Q. Chen, and M. Yachida, Face detection from color images using a fuzzy pattern matching method, *IEEE Trans. Pattern Anal. Mach. Intell.* **21**, 2000, 557–563.

209. H. Wu, T. Yokoyama, D. Pramadihanto, and M. Yachida, Face and facial feature extraction from colour image, in *IEEE Proc. of 2nd Int. Conf. on Automatic Face and Gesture Recognition, Vermont, Oct. 1996*, pp. 345–349.

210. G. Wyszecki and W. S. Stiles, *Color Science*, Wiley, New York, 1967.

211. X. Xie, R. Sudhhakar, and H. Zhuang, On improving eye feature extraction using deformable templates, *Pattern Recog.* **27**, 1994, 791–799.

212. G. Xu and T. Sugimoto, Rits eye: A software-based system for realtime face detection and tracking using pan-tilt-zoom controllable camera, in *Proc. of International Conference on Pattern Recognition, 1998*.

213. K. Yachi, T. Wada, and T. Matsuyama, Human head tracking using adaptive appearance models with a fixed-viewpoint pan-tilt-zoom camera, in *Proceedings Fourth IEEE International Conference on Automatic Face and Gesture Recognition, 2000.*

214. G. Yang and T. S. Huang, Human face detection in a complex background, *Pattern Recog.* **27**, 1994, 53–63.

215. J. Yang and A. Waibel, A real-time face tracker, in *IEEE Proc. of the 3rd Workshop on Applications of Computer Vision, Florida, 1996.*

216. L. Yang, Multiple-face tracking system for general region-of-interest video coding, in *Proceedings of the 2000 International Conference on Image Processing, 2000*, p. MA09.13.

217. M.-H. Yang, N. Ahuja, and D. Kriegman, Face detection using mixtures of linear subspaces, in *Proceedings Fourth IEEE International Conference on Automatic Face and Gesture Recognition, 2000.*

218. L. Yin and A. Basu, Integrating active face tracking with model based coding, *Pattern Recog. Lett.* **20**, 1999, 651–657.

219. T. Yokoyama, Y. Yagi, and M. Yachida, Facial contour extraction model, in *IEEE Proc. of 3rd Int. Conf. on Automatic Face and Gesture Recognition, 1998.*

220. T. W. Yoo and I. S. Oh, A fast algorithm for tracking human faces based on chromatic histograms, *Pattern Recog. Lett.* **20**, 1999, 967–978.

221. K. C. Yow and R. Cipolla, Feature-based human face detection, *Image Vision Comput.* **15**(9), 1997.

222. A. L. Yuille, P. W. Hallinan, and D. S. Cohen, Feature extraction from faces using deformable templates, *Int. J. Comput. Vision* **8**, 1992, 99–111.

223. Y. Zhu, S. Schwartz, and M. Orchard, Fast face detection using subspace discriminant wavelet features, in *IEEE Conference on Computer Vision and Pattern Recognition, 2000.*

224. M. Zobel, A. Gebhard, D. Paulus, J. Denzler, and H. Niemann, Robust facial feature localization by coupled features, in *Proceedings Fourth IEEE International Conference on Automatic Face and Gesture Recognition, 2000.*