Human and Machine Recognition of Faces: A Survey

RAMA CHELLAPPA, FELLOW, IEEE, CHARLES L. WILSON, SENIOR MEMBER, IEEE, AND SAAD SIROHEY, MEMBER, IEEE

The goal of this paper is to present a critical survey of existing literature on human and machine recognition of faces. Machine recognition of faces has several applications, ranging from static matching of controlled photographs as in mug shots matching and credit card verification to surveillance video images. Such applications have different constraints in terms of complexity of processing requirements and thus present a wide range of different technical challenges. Over the last 20 years researchers in psychophysics, neural sciences and engineering, image processing, analysis and computer vision have investigated a number of issues related to face recognition by humans and machines. Ongoing research activities have been given a renewed emphasis over the last five years. Existing techniques and systems have been tested on different sets of images of varying complexities. But very little synergism exists between studies in psychophysics and the engineering literature. Most importantly, there exist no evaluation or benchmarking studies using large databases with the image quality that arises in commercial and law enforcement applications.

In this paper, we first present different applications of face recognition in commercial and law enforcement sectors. This is followed by a brief overview of the literature on face recognition in the psychophysics community. We then present a detailed overview of more than 20 years of research done in the engineering community. Techniques for segmentation/location of the face, feature extraction and recognition are reviewed. Global transform and feature based methods using statistical, structural and neural classifiers are summarized. A brief summary of recognition using face profiles and range image data is also given. Real-time face recognition from video images acquired in a cluttered scene such as an airport is probably the most challenging problem. We discuss several existing technologies in the image understanding literature that could potentially impact this problem.

Given the numerous theories and techniques that are applicable to face recognition, it is clear that evaluation and benchmarking of these algorithms is crucial. We discuss relevant issues such as data collection, performance metrics, and evaluation of systems and techniques. Finally, summary and conclusions are given.

I. INTRODUCTION

Machine recognition of faces from still and video images is emerging as an active research area spanning several disciplines such as image processing, pattern recognition, computer vision and neural networks. In addition, face recognition technology (FRT) has numerous commercial and law enforcement applications. These applications range from static matching of controlled format photographs such as passports, credit cards, photo ID's, driver's licenses, and mug shots to real-time matching of surveillance video images presenting different constraints in terms of processing requirements. Although humans seem to recognize faces in cluttered scenes with relative ease, machine recognition is a much more daunting task. In this paper we address critical issues involved in understanding how humans perceive faces and follow it with a detailed discussion of several techniques and systems that have been considered in the engineering literature for nearly 25 years. Critical issues such as data collection and performance evaluation are also addressed.

A general statement of the problem can be formulated as follows: Given still or video images of a scene, identify one or more persons in the scene using a stored database of faces. Available collateral information such as race, age and gender may be used in narrowing the search. The solution of the problem involves segmentation of faces from cluttered scenes, extraction of features from the face region, identification, and matching. The generic face recognition task thus posed is a central issue in problems such as electronic line up and browsing through a database of faces.

Over the past 20 years extensive research has been conducted by psychophysicists, neuroscientists and engineers on various aspects of face recognition by humans and machines. Psychophysicists and neuroscientists have been concerned with issues such as: Uniqueness of faces; whether face recognition is done holistically or by local feature analysis; analysis and use of facial expressions for recognition; how infants perceive faces; organization of memory for faces; inability to accurately recognize

Manuscript received July 1, 1994; revised February 14, 1995. Chellappa and Sirohey's work was supported by the Advanced Research Projects Agency under Order Number 8459 and the US Army Topographic Engineering Center under contract DACA76-92-0009. Wilson's work was supported by the FBI through an agreement with NIST.

R. Chellappa and S. Sirohey are with the Department of Electrical Engineering and Center for Automation Research, University of Maryland, College Park, MD 20742 USA.

C. L. Wilson is with the National Institute of Standards and Technology, Gaithersburg, MD 20899 USA.

IEEE Log Number 9410528.

inverted faces; existence of a "grandmother" neuron for face recognition; role of the right hemisphere of the brain in face perception; and inability to recognize faces due to conditions such as prosopagnosia. Some of the theories put forward to explain the observed experimental results are contradictory. Many of the hypotheses and theories put forward by researchers in these disciplines have been based on rather small sets of images. Nevertheless, several of the findings have important consequences for engineers who design algorithms and systems for machine recognition of human faces.

Barring a few exceptions [21], [24], [116], research on machine recognition of faces has developed independent of studies in psychophysics and neurophysiology. During the early and mid-1970's, typical pattern classification techniques, which use measured attributes between features in faces or face profiles, were used. During the 1980's, work on face recognition remained largely dormant. Since the early 1990's, research interest in FRT has grown very significantly. One can attribute this to several reasons: An increase in emphasis on civilian/commercial research projects; the reemergence of neural network classifiers with emphasis on real-time computation and adaptation; the availability of real time hardware; and the increasing need for surveillance-related applications due to drug trafficking, terrorist activities, etc.

Over the last five years, increased activity has been seen in tackling problems such as segmentation and location of a face in a given image, and extraction of features such as eyes, mouth, etc. Also, numerous advances have been made in the design of statistical and neural network classifiers for face recognition. Classical concepts such as Karhunen–Loeve transform based methods [11], [82], [104], [124], [133], singular value decomposition [69] and more recently neural networks [21], [51], have been used. Barring a few exceptions [104], many of the existing approaches have been tested on relatively small datasets, typically less than 100 images.

In addition to recognition using full face images, techniques that use only profiles constructed from a side view are also available. These methods typically use distances between the "fiducial" points in the profile (points such as the nose tip, etc.) as features. Modifications of Fourier descriptors have also been used for characterizing the profiles. Profile based methods are potentially useful for the mug shot problem, due to the availability of side views of the face.

All of the discussion thus far has focused on recognizing faces from still images. The still image problem has several inherent advantages and disadvantages. For applications such as mug shots matching, due to the controlled nature of the image acquisition process, the segmentation problem is rather easy. On the other hand, if only a static picture of an airport scene is available, automatic location and segmentation of a face could pose serious challenges to any segmentation algorithm. However, if a video sequence acquired from a surveillance camera is available, segmentation of a person in motion can be more easily accomplished using motion as a cue. Only a handful of papers on face recognition [117], [121], [133] have addressed the issue of segmenting a face image from the background. However, there is a significant amount of work reported in the image understanding (IU) literature [1], [2] on segmenting a moving object from the background using a sequence. Also, there is a significant amount of work on the analysis of nonrigid moving objects, including faces, in the IU [1], [2] as well as the image compression literature [4]. We briefly discuss those techniques that have potential applications to recovery and reconstruction (in 3D) of faces from a video sequence. The reconstructed image will be useful for recognition tasks when disguises and aging are present.

In addition to the separation of images into static and real-time image sequences several other parameters are important in critically evaluating existing methods. In any pattern recognition problem the accuracy of the solution will be strongly affected by the limitations placed on the problem. To restrict the problem to practical proportions both the image input and the size of the search space must have some limits. The limits on the image might for example include controlled format, backgrounds which simplify segmentation, and controls on image quality. The limits on the database size might include geographic limits and descriptor based limits. Critical issues involving data collection, evaluation and benchmarking of existing algorithms and systems also need to be addressed.

An excellent survey of face recognition research prior to 1991 is in [114]. Still we decided to prepare our survey paper due to the following reasons: The face recognition area has become very active since 1990. Approaches based on Karhunen–Loeve expansion, neural networks and feature matching have all been initiated since the survey paper [114] appeared. Also, [114] did not cover discussions on face recognition from a video, profile, or range imagery nor any aspects of performance evaluation.

The organization of the paper is as follows: In Section II we describe several applications of FRT in still and video images and point out the specific constraints that each set of applications pose. Section III provides a brief summary of issues that are relevant from the psychophysics point of view. In Section IV a detailed review of face recognition techniques, involving still intensity and range images, in the engineering literature is given. Techniques for segmentation of faces from clutter, feature extraction and recognition are detailed. Face recognition using profile images (which has not been pursued with much vigor in recent years, but nevertheless is useful in mug shots matching problem) is discussed in Section V. Section VI presents a discussion on face recognition from video images with special emphasis on how IU techniques could be useful. Some specific examples of face recognition and recall work in law enforcement domains, and commercial applications are briefly discussed in Section VII. Data collection and performance evaluation of face recognition algorithms and architectures are addressed in Section VIII. Finally, summary and conclusions are in Section IX.

Table 1 Applications of Face Recognition Technology

Applications	Advantages	Disadvantages
1a. Credit Card, Driver's License, Passport, and Personal Identification	Controlled image Controlled segmentation Good quality images	No existing database Large potential database Rare search type
1b. Mug shots Matching	Mixed image quality More than one image available	
2. Bank/Store Security	High value Geographically localized search	Uncontrolled segmentation Low image quantity
3. Crowd Surveillance	High value Small file size Availability of video images	Uncontrolled segmentation Low image quality Real-time
4. Expert Identification	High value Enhancement possible	Low image quality Legal certainty required
5. Witness Face Reconstruction	Witness search limits	Unknown similarity
6. Electronic Mug Shots Book	Descriptor search limits	Viewer fatigue
7. Electronic Lineup	Descriptor search limits	Viewer fatigue
8. Reconstruction of Face from Remains	High value	Requires physiological input
9. Computerized Aging	High value	Requires example input

II. Applications

Commercial and law enforcement applications of FRT listed in Table 1 range from static, controlled format photographs to uncontrolled video images, posing a wide range of different technical challenges and requiring an equally wide range of techniques from image processing, analysis, understanding and pattern recognition. One can broadly classify the challenges and techniques into two groups: static (no video) and dynamic (video) matching. Even among these groups, significant differences exist, depending on the specific application. The differences are in terms of image quality, amount of background clutter (posing challenges to segmentation algorithms), the availability of a well defined matching criterion, and the nature, type and amount of input from a human (as in applications 4 and 5). In some applications, such as computerized aging, one is only concerned with defining a set of transformations so that the new images created by the system are similar to what humans expect based on their recollections.

Three different kinds of problems arise in applications listed in Table 1; these are matching, similarity detection, and transformation. Applications 1, 2, and 3 involve matching one face image to another face image. Applications 4–7 involve finding or creating a face image which is similar to the human recollection of a face. Finally, applications 8 and 9 involve generating an image of a face from input data that is useful in other applications by using other information to perform modifications of a face image. Each of these applications imposes different requirements on the recognition process. Matching requires that the candidate matching face image be in some set of face images selected by the system. Similarity detection requires, in addition to matching, that images of faces be found which are similar to a recalled face; this requires that the similarity measure used by the recognition system closely match the similarity measures used by humans. Transformation applications require that new images created by the system be similar to human recollections of a face.

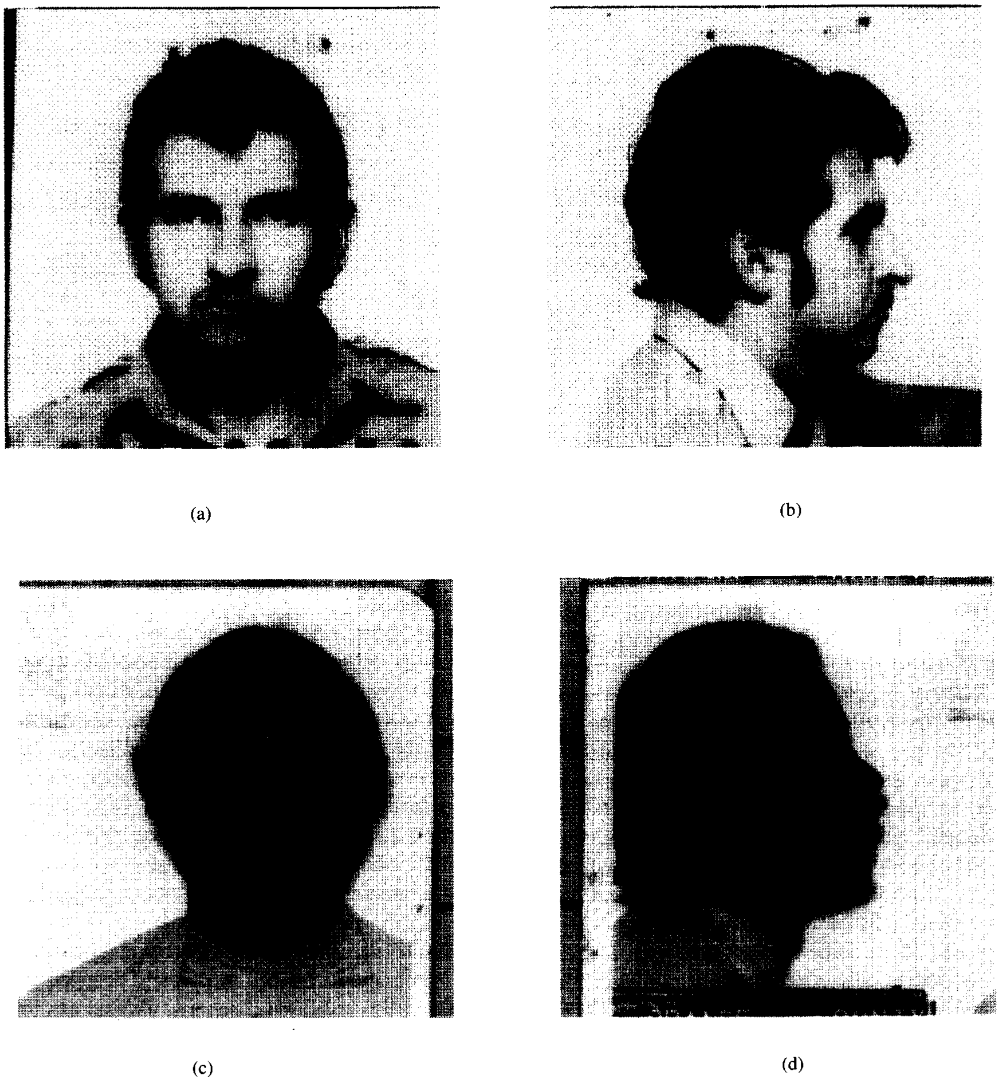
A. Static Matching

Mug shots matching is the most common application in this group. Typically, in mug shots photographs, the illumination is reasonably controlled, and one frontal and one or more side views of a person's face are taken. Although more control can be exercised in image acquisition, no uniform standards exist for use by booking stations across the country. These standards could involve the type of background, illumination, resolution of the camera, and the distance between the camera and the person being photographed. By enforcing such simple controls over the image acquisition process, one can potentially simplify segmentation and matching algorithms. Two examples of typical mug shots images are given in Fig. 1.

Simple versions of the mug shots matching problem are recognition of faces in driver's licenses, credit cards, personal ID cards, and passports. Typical examples of face images in drivers licenses or personal ID cards are shown in Fig. 2. The images in these documents are usually acquired with more control than in mug shots.

Typically, images in mug shots applications are of good quality, consistent with existing law enforcement standards. Given the reasonably controlled imaging conditions, segmentation/location of a face is relatively easy. Potential challenges are in searching through a large dataset and also in matching; though the imaging conditions are controlled, variations in the face due to aging, hair loss, hair growth, etc., have to be accounted for in feature extraction and matching.

Application 2 is more complicated than application 1, largely due to the uncontrolled nature of the image acquisition conditions. As considerable background clutter may be present, segmentation gets harder. Also, the quality of the image tends to be low. An approximate rendition of such an image is shown in Fig. 3. It should be pointed out that application 2 falls between static and dynamic matching. Some of the images that arise in this application are on film while some are acquired from a video camera. As in application 1, variations in face images due to aging and



(a)

(b)

(c)

(d)

Fig. 1. Frontal and profile mug shots images.

disguises must be accounted for in feature extraction and matching. In applications 1 and 2, the matching criterion can be quantified; also, the top few choices can be rank ordered.

Applications 4–7 involve finding or creating a face image which is similar to the human recollection of a face. In application 4, an expert confirms that the face in the given image corresponds to a person in question. It is possible that the face in the image could be disguised, or occluded. Typically, in this application a list of similar looking faces is generated using a face identification algorithm, the expert then performs a careful analysis of the listed faces. In application 5 the witness is asked to compose a picture of a culprit using a library of features such as noses, eyes, lips, etc. For example the library may have examples of noses that are long, short, curved, flat, etc., from which one that is closest to witness's recollection is chosen. In application 6, electronic browsing of photo collection is attempted. Application 7 involves a witness identifying a face from a set of face images which include some false candidates. Typically, in these applications the image quality tends to be low; in addition to matching, it is required to find faces that are similar to a recalled face. The similarity measure is difficult to quantify, as measures supposedly used by humans need to be defined. The problem is complicated further in that when humans search through a mug shots book, they tend to make more recognition errors as the number of mug shots presentations increases. It is difficult to completely quantify the degradation in machine implementation of algorithms developed for applications 4–6. Another issue is the incorporation of mechanisms for recalling faces that humans use in the algorithms. Applications 4–7 need a

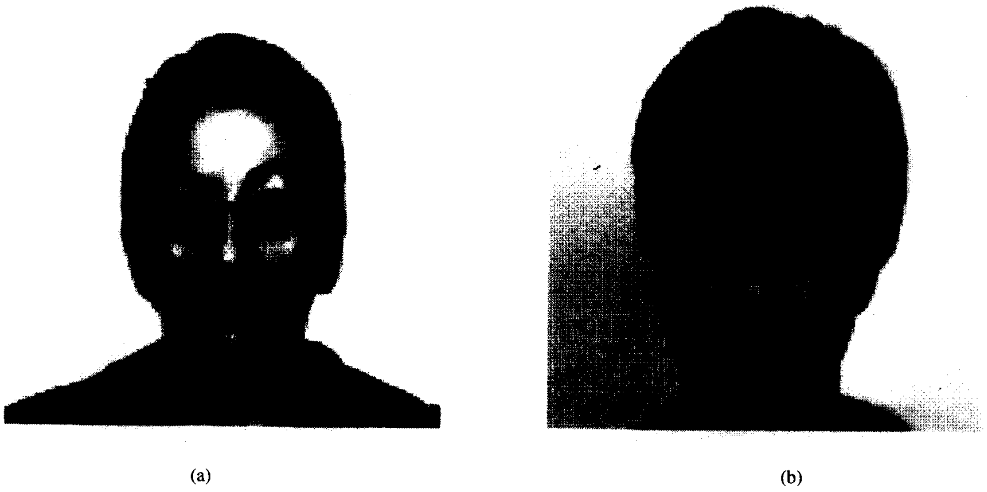


Fig. 2. Face images in a controlled background, as in passport or identification documents.

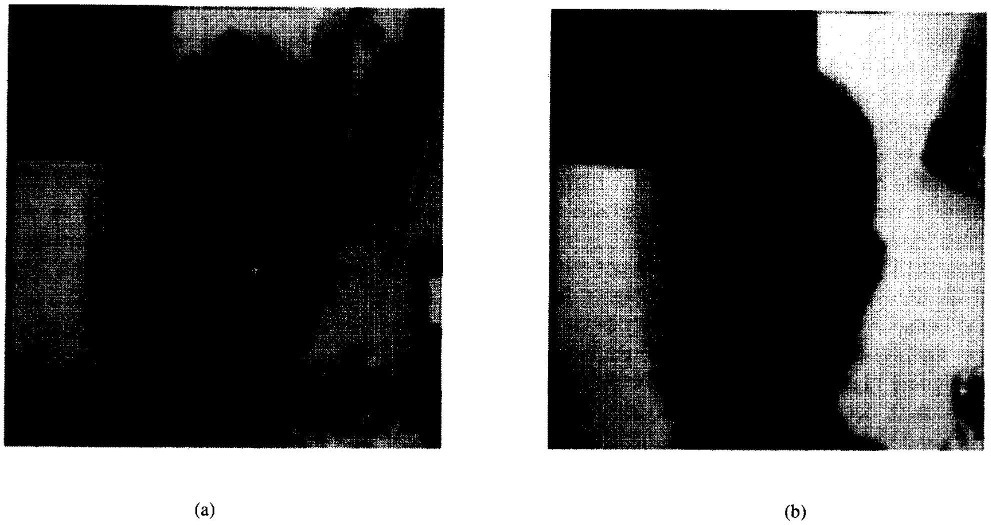


Fig. 3. An approximate illustration of an uncontrolled environment for face images corresponding to application 2.

strong interaction between algorithms and known results in psychophysics and neuroscience studies.

Applications 8 and 9 involve transformations of images from current data to what they could have been (application 8) or to what they will be (application 9). These are even more difficult than applications 4–6, since "smoothing" or "predictive" mechanisms need to be incorporated into the algorithms.

B. Dynamic Matching

We group application 3, and cases of application 2 where a video sequence is available, as dynamic. The images available through a video camera tend to be of low quality. Also, in crowd surveillance applications the background is very cluttered, making the problem of segmenting a face in the crowd difficult. However, since a video sequence is available, one could use motion as a strong cue for segmenting faces of moving persons. One may also be able to do partial reconstruction of the face image using existing models [10], [23], [87] and be able to account for disguises, somewhat better than in static matching problems. One of the strong constraints of this application is the need for real-time recognition. It is expected that several of the existing methodologies in the IU literature [1]–[5] for image sequence based segmentation, structure estimation, nonrigid object motion and recognition will be useful for solving the requirements of application 3.

It should be remarked that the widely varying constraints of the different applications necessitate different methods and scores for evaluating and benchmarking existing algorithms and systems.

III. Psychophysics and Neurophysiological Issues Relevant to Face Recognition

In general, the human recognition system utilizes a broad spectrum of stimuli, obtained from many, if not all, of the senses (visual, auditory, olfactory, tactile, etc.). These stimuli are used in either an individual or collective manner for both the storing and retrieval of face images for the purpose of recognition. There are many instances when contextual knowledge is also applied, i.e., the surroundings play an important role—recognizing faces in relation to where they are supposed to be located. It is futile (impossible with the present technology) to attempt to develop a system which will mimic all these remarkable traits of humans. However, the human brain has its shortcomings in the total number of persons that it can accurately "remember." The benefit of a computer system would be its capacity to handle large datasets of face images. In most of the applications the images are present in single or multiple views of 2D intensity data, which forces the inputs to a computer algorithm to be visual only. It is for this reason that the literature reviewed in this section is related to aspects of human visual perception.

During the course of our literature survey, we have come across several hundred papers that address problems and issues related to human recognition of faces. Many of these studies and their findings have direct relevance to engineers interested in designing algorithms or systems for machine recognition of faces. A detailed review of relevant studies in psychophysics and neuroscience is beyond the scope of this paper. We only summarize findings that are potentially relevant to the design of face recognition systems. For details the reader is referred to the papers cited below and to citations in the supplemental bibliography. In writing this section, we have largely benefited from books [19], [37], [39] and survey papers [14], [20], [41], [65]. Also literature on how animals such as monkeys, dogs and cats recognize faces is not included in our survey. Notable works on experiments with monkeys are [106]–[108].

The issues that are of potential interest to designers are:

- **Is face recognition a dedicated process? [41]:** Evidence for the existence of a dedicated face processing system comes from three sources. A) Faces are more easily remembered by humans than other objects when presented in an upright orientation. B) Prosopagnosia patients are unable to recognize previously familiar faces, but usually have no other profound agnosia. They recognize people by their voices, hair color, dress, etc. Although they can perceive eyes, nose, mouth, hair, etc., they are unable to put together these features for the purpose of identification. It should be noted that prosopagnosia patients recognize whether the given object is a face or not, but then have difficulty in identifying the face. C) It is argued that infants come into the world pre-wired to be attracted by faces. Neonates seem to prefer to look at moving stimuli that have face-like patterns in comparison to those containing no pattern or with jumbled features.
- **Is face perception the result of holistic or feature analysis?** Both holistic and feature information are crucial for the perception and recognition of faces. Studies suggest the possibility of global descriptions serving as a front end for finer, feature-based perception. If dominant features are present, holistic descriptions may not be used. For example, in face recall studies, humans quickly focus on odd features such as big ears, a crooked nose, a staring eye, etc.
- **Ranking of significance of facial features:** Hair, face outline, eyes and mouth (not necessarily in this order) have been determined to be important for perceiving and remembering faces. Several studies have shown that the nose plays an insignificant role; this may be due to the fact that almost all of these studies have been done using frontal images. In face recognition using profiles (which may be important in mug shots matching applications, where profiles can be extracted from side views), several fiducial points ("features") are around the nose region (see Section V). Another outcome of some of the studies is that both external and internal features are important in the recognition of previously presented but otherwise unfamiliar faces, and internal features are more dominant in the recognition of familiar faces. It has also been found that the upper part of the face is more useful for face recognition than the lower part. The role of aesthetic attributes such as beauty, attractiveness and/or pleasantness has also been studied, with the conclusion that the more attractive the faces are, the better is their recognition rate; the least attractive faces come next, followed by the mid-range faces, in terms of ease of being recognized.
- **Caricatures: [20]** Perkins [105] formally defines a "caricature as a symbol that exaggerates measurements relative to any measure which varies from one person to another." Thus the length of a nose is a measure that varies from person to person, and could be useful as a symbol in caricaturing someone, but not the number of ears. Caricatures do not contain as much information as photographs, but they manage to capture the important characteristics of a face; experiments comparing the usefulness of caricatures and line drawings decidedly favor the former.
- **Distinctiveness:** Studies show that distinctive faces are better retained in recognition memory and are recognized better and faster than typical faces. However, if a decision has to be made as to whether an object is a face or not, it takes longer to recognize an atypical face than a typical face. This may be explained by

different mechanisms being used for detection and for identification.

- **The role of spatial frequency analysis:** Earlier studies [47], [64] concluded that information in low spatial frequency bands play a dominant role in face recognition. Recent studies [119] show that, depending on the specific recognition task, the low, bandpass and high frequency components may play different roles. For example the sex judgment task is successfully accomplished using low frequency components only, while the identification task requires the use of high frequency components. The low frequency components contribute to the global description, while the high frequency components contribute to the finer details required in the identification task.
- **The role of the brain: [40]** The role of the right hemisphere in face perception has been supported by several researchers. In regard to prosopagnosia and the right hemisphere, a retrospective study seems to strongly indicate right hemisphere involvement in face recognition. In other brain damaged victims, those with right hemisphere disease have more impairment in facial recognition then left hemisphere disease. When shown the left half of one face and the right half of another face tachistoscopically, the overwhelming majority of commissurotomy patients selected the face shown to the left vision field (LVF), which arrives initially at the right hemisphere. In other tachistoscopic studies, the LVF has the advantage in both speed and accuracy of response and in long term memory response. Studies have also shown a right hemisphere advantage in reception and/or storage of faces. Some other studies argue against right hemisphere superiority in face perception. Postmortem studies of prosopagnosia victims with known lesions in the right hemisphere have found approximately symmetrical lesions in the left hemisphere. Other cases of bilateral brain damage have been seen or suspected in patients with prosopagnosia. The ways in which the two hemispheres operate may reflect variations in degrees of expertise. It appears that the right hemisphere does possess a slight advantage in aspects of face processing. It is also true that the two hemispheres may simultaneously handle different types of information. The dominance of the right hemisphere in facial processing may be the result of left hemisphere dominance in language. The right hemisphere is also involved in the interpretation of emotions, and this may underlie the slight asymmetry in perceiving and remembering faces.
- **Face recognition by children. [29], [30]** It appears that children under ten years of age code unfamiliar faces using isolated features. Recognition of these faces is done using cues derived from paraphernalia, such as clothes, glasses, hair style, hats, etc. Ten-year-old children exhibit this behavior less frequently, while children older than 12 years rarely exhibit this behavior. It is postulated that around age ten, children seem to change their recognition mechanisms from one of isolated features and paraphernalia to one of holistic analysis. Curiously, when children as young as five years are asked to recognize familiar faces, they do pretty well in ignoring paraphernalia.

  Several other interesting studies related to how children perceive inverted faces are summarized in [29].
- **Facial expression: [19]** Based on neurophysiological studies, it seems that analysis of facial expressions is accomplished in parallel to face recognition. Some prosopagnosic patients, who have difficulties in identifying familiar faces, nevertheless seem to recognize emotional expressions. Patients who suffer from "organic brain syndrome" suffer from poor expression analysis but perform face recognition quite well. Normal humans exhibit parallel capabilities for facial expression analysis and face recognition. Similarly, separation of face recognition and "focused visual processing" (look for someone with a thick mustache) tasks have been claimed.
- **Role of race/gender:** Humans recognize people from their own race better than people from another race. This may be due to the fact that humans may be coding an "average" face with "average" attributes, the characteristic of which may be different for different races, making the recognition of faces from a different race harder. Goldstein [50] gives two possible reasons for the discrepancies: Psychosocial, in which the poor identification results are from the effects of prejudice, unfamiliarity with the class of stimuli, or a variety of other interpersonal reasons; and psychophysical, dealing with loss of facial detail because of different amounts of reflectance from different skin colors, or race-related differences in the variability of facial features. Using tables showing the coefficients of variation for different facial features for different races, it has been concluded that poor identification of other races is not a psychophysical problem but more likely a psychosocial one. Using the same data collected in [50], some studies have been done to quantify the role of gender in face recognition. It has been found [49] that in a Japanese population, a majority of the women's facial features are more heterogeneous than the men's features. It has also been found that white women's faces are slightly more variable than men's, but that the overall variation is small.
- **Image quality:** In [125] the relationship between image quality and recognition of a human face has been explored. The task required of observers is to identify one face from a gallery of 35 faces. The modulation transfer function area (MTFA) was used as a metric to predict an observers performance in a task requiring the extraction of detailed information from both static and dynamic displays. Performance for an observer is measured by two dependent variables—proportion of correct responses and response time. It was found that as the MTFA becomes moderately large, facial recognition performance reaches a ceiling which cannot be exceeded. The MTFA metric

indicates the extent to which a system's response exceeds the minimum contrast requirements, averaged across all spatial frequencies of interest [125].

A. Summary

For engineers interested in designing algorithms and systems for face recognition, numerous studies in psychophysics and neurophysiological literature serve as useful guides. As an example, designers should include both global and local features for representing and recognizing faces. Among the features, some (hairline, eyes, mouth) are more significant or useful than others (nose). This observation is true for frontal images of faces, while for side views and profiles, the nose is an important feature. Studies on distinctiveness and caricatures can help add special features of the face that can be utilized for perceiving and recognizing faces. The role of spatial frequency analysis suggests multiresolution/multiscale algorithms for different problems related to face perception. Issues such as how humans recognize people from their own race better than people from another race, and how infants recognize faces, are very important in the design of systems for expert identification, witness face reconstruction, electronic mug shots books and lineups. Interpreting face recognition using Marr's computational vision paradigm may point to new algorithms and systems; see Chapter 6 of [19]. Other issues, such as organization of face memory, are very pertinent for the design of large databases such as mug shots albums. Usefulness of facial expressions on face recognition needs to be evaluated.

Historically, there has been great interest among computer vision algorithm developers and system designers in learning how our visual system works and in translating these mechanisms into real systems. Marr's paradigm for computational vision [89] is a pioneering example of such an effort. Designers of face recognition algorithms and systems should be aware of relevant psychophysics and neurophysiological studies but should be prudent in using only those that are applicable or relevant from a practical/implementation point of view.

IV. Face Recognition from Still Intensity and Range Images

In this section we survey the state of the art in face recognition in the engineering literature. We have divided the face recognition papers into three groups. Methods for segmentation of faces from a given image are discussed in Section IV-A. Techniques for extraction of statistical features such as Karhunen–Loeve transform and singular value decomposition coefficients, structural features like eyes, nose, lips, and points of high curvature are summarized in Section IV-B. Most papers included in Sections IV-A and IV-B do not report any recognition or matching experiments. Recognition and identification papers that use features described in Section IV-B or other features are surveyed in Section IV-C. The recognition techniques are presented as statistical, neural and feature based. Finally a summary section is included.

A. Segmentation

One of the earliest papers that reported the presence or absence of a face in an image is [113]. An edge map extracted from the input image is matched to a large oval template with possible variations in the position and size of the template. At positions where potential matches are reported, the head hypothesis is confirmed by inspecting the edges produced at expected positions of eyes, mouth, etc. The technique was dependent on the illumination direction.

Kelly [81] introduced a top-down image analysis approach known as PLANNING for automatically extracting the head and body outlines from an image and subsequently the locations of eyes, nose, mouth. As an example, the head extraction algorithm works as follows: Smoothed versions of original images (obtained by local averaging) are first searched for edges that may form the outline of a head; extracted edge locations are then projected back to the original image, and a fine search is locally performed for edges that form the head outline. Several heuristics are used to connect the edges. Once the head outline is obtained, the expected locations for eyes, nose and mouth are searched for locating these features. Several heuristics are again employed in the search process.

The algorithm for extracting the body of a person, subtracts the image of the background without the person from the image that has the person. This difference image is reduced in size by averaging and then thresholded. After applying a connected component algorithm, the extremes of the regions obtained define the region in which the body is located. Details on the feature measurements, dataset etc., used in [81] are given in Section IV-C.

Govindaraju *et al.* [55] consider a computational model for locating the face in a cluttered image. Their technique utilizes a deformable template which is slightly different than that of Yuille *et al.* [145]. Working on the edge image they base their template on the outline of the head. The template is composed of three segments that are obtained from the curvature discontinuities of the head outline. These three segments form the right side-line, the left side-line and the hairline of the head. Each one of these curves is assigned a four-tuple consisting of the length of the curve, the chord in vector form, the area enclosed between the curve and the chord, and the centroid of this area. To determine the presence of the head, all three of these segments should be present in particular orientations. The center of these three segments gives the location of the center of the face. The templates are allowed to translate, scale and rotate according to certain spring-based models. They construct a cost function to determine hypothesized candidates. They have experimented on about ten images, and though they claim to have never failed to miss a face, they do get false alarms.

Craw *et al.* in [34] describe a method for extracting the head area from the image. They use a hierarchical image scale and a template scale. Constraints are imposed on the location of the head in the image. Resolutions of 8 × 8, 16 × 16, 32 × 32, 64 × 64 and full scale 128 × 128 are used in their multiresolution scheme. At the lowest resolution a template is constructed of the head outline. Edge magnitude and direction are calculated from the gray level image using a Sobel mask. A line follower is used to connect the outline of the head. After the head outline has been located a search for lower level features such as eyes, eyebrows, and lips is conducted, guided by the location of the head outline, using a similar line following method. The algorithm for detecting the head outline, performed better than the one searching for the eyes.

Another method of finding the face in an image was defined by Burt [25]. It utilized a coarse to fine approach with a template based match criteria to locate the head. Burt illustrates the usefulness of such techniques by describing a "smart transmission" system. This system could locate and track the human head and then send the information of the head location to an object based compression algorithm.

In [35] Craw, Tock, and Bennet describe a system to recognize and measure facial features. Their work was motivated in part by automated indexing of police mug shots. They endeavor to locate 40 feature points from a grey-scale image; these feature points were chosen according to Shepherd [120], which was also used as a criterion of judgment. The system uses a hierarchical coarse-to-fine search. The template drew upon the principle of polygonal random transformation in Grenander *et al.* [56]. The approximate location, scale and orientation of the head is obtained by iterative deformation of the whole template by random scaling, translation and rotation. A feasibility constraint is imposed so that these transformations do not lead to results that have no resemblance to the human head. Optimization is achieved by simulated annealing [46]. After a rough idea of the location of the head is obtained, refinement is done by transforming individual vectors of the polygon. The authors claim successful segmentation of the head in all 50 images that were tested. In 43 of these images a complete outline of the head was distinguishable; in the remaining ones there was failure in finding the chin. The detailed template of the face included eyes, nose, mouth, etc.; in all, 1462 possible feature points were searched for. The authors claim to be able to identify 1292 of these feature points. The only missing feature was the eyebrow, as they did not have a feature expert for that. They attribute the 6% incorrect identification to be due to presence of beards and mustaches in their database, which caused mistakes in locating the chin and the mouth of the subject. It should be noted that due to its use of optimization and random transformation, the system is inherently computationally intensive.

In [123] the face is segmented from a moderately cluttered background using an approach that involves working with both the intensity image of the face as well as the edge image found using the Canny's edge finder [28]. Preprocessing tasks include locating the intersection points of edges (occlusion of objects), assigning labels to contiguous edge segments and linking of most likely similar edge segments at intersection points. The human face is approximated using the ellipse as the analytical tool. Pairs of labeled edge segments $L_i, L_j^{'}$ are fitted to a linearized equation of the ellipse (1). This linearization is possible under the condition that the semi-major axis a and/or the semi-minor axis b of the ellipse are not 0, which is true for all cases considered.

$$2x_i a_0 - y_i^2 a_1 + 2y_i a_2 - a_3 = x_i^2 \qquad (1)$$

where

$$\begin{aligned} a_0 &= x_0, & a_1 &= \frac{a^2}{b^2} \\ a_2 &= \frac{a^2}{b^2}\, y_0, & a_3 &= x_0^2 + \frac{a^2}{b^2} - a^2. \end{aligned}$$

The resulting parameter set x_0, y_0, a, b is checked against the aspect ratio of the face, and if it is satisfied, is included in the class of parameter sets for final selection. The parameter sets in the class of parameters are reverse fitted with the labeled segments. The parameter set with the most segments (compensated for size) is selected to represent the segmented face. Fig. 4 shows the segmentation process going through its different phases from the input image to its edge representation, then the final grouping of the likely edge segments corresponding to the outline of the face, and finally the output image without background clutter. An accuracy of above 80% was reported when the process was applied to a data set of 48 cluttered images. Fig. 5 shows some of the results of the segmentation algorithm. The image size was 128 × 128 pixels.

The presence or absence of a face using the eigenfaces expansion is reported in [133]. Details on eigenfaces are in Sections IV-B and IV-C.

B. Feature Extraction

Recently, the use of the Karhunen–Loeve (KL) expansion for the representation [82], [124] and recognition [104], [133] of faces has generated renewed interest. The KL expansion has been studied for image compression for more than 30 years [74], [140]; its use in pattern recognition applications has also been documented for quite some time [45]. One of the reasons why KL methods, although optimal, did not find favor with image compression researchers is their computational complexity. As a result, fast transforms such as the discrete sine and cosine transform have been used [74]. In [124], Sirovich and Kirby revisit the problem of KL representation of images (cropped faces). Once the eigenvectors (referred to as "eigenpictures") are obtained, any image in the ensemble can be approximately reconstructed using a weighted combination of eigenpictures. By using an increasing number of eigenpictures, one gets an improved approximation to the given image. The authors also give examples of approximating an arbitrary image (not included in the calculation of eigenvectors) by the eigenpictures. The emphasis in this paper is on the representation of human faces. The weights that characterize

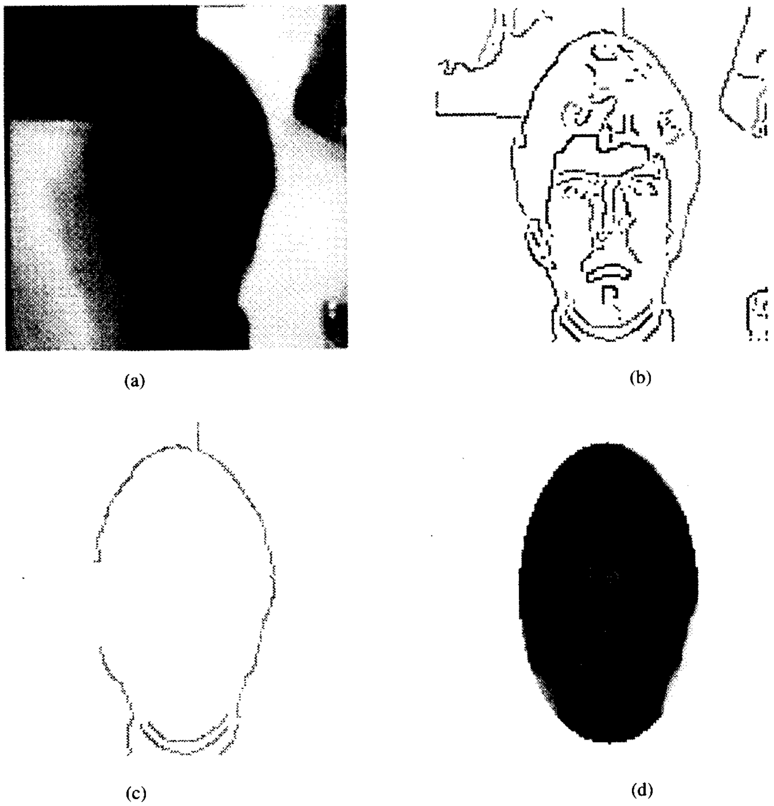


Fig. 4. (a) Input image, (b) edge image, (c) linked segments, and (d) segmented image.

the expansion of the given image in terms of eigenpictures serve the role of features.

In a subsequent extension of their work, Kirby and Sirovich in [82] include the inherent symmetry of faces in the eigenpicture representation of faces, by using an extended ensemble of images consisting of original faces and their mirror images. Since the computations of eigenvalues and eigenvectors can be split into even and odd pictures, there is no overall increase in computational complexity compared to the case in which only the original set of pictures in used. Although the eigenrepresentation for the extended ensemble does not produce dramatic reduction in the error in reconstruction when compared to the unextended ensemble, still the method that accounts for symmetry in the patterns is preferable.

In [11], the KL is combined with two other operations to improve the performance of the extraction technique for the classification of front-view faces. The application of the KL expansion directly to a facial image without standardization does not achieve robustness against variations in image acquisition. [11] uses standardization of the position and size of the face. The center points are the regions corresponding to the eyes and mouth. Each target image is translated, scaled and rotated through affine transformation so that the reference points of the eyes and mouth are in a specific spatial arrangement with a constant distance. An empirically defined standard window encloses the transformed image. The KL expansion applied to the standardized face images is known as the Karhunen–Loeve transform of intensity pattern in affine-transformed target (KL-IPAT) image. The KL-IPAT was extracted from 269 images with 100 eigenfaces. The second step is to apply the Fourier Transform to the standardized image and use the resulting Fourier spectrum instead of the spatial data from the standardized image. The KL expansion applied to the Fourier spectrum is called the Karhunen–Loeve transform of Fourier spectrum in the affine-transformed target (KL-FSAT) image. The robustness of the KL-IPAT and KL-FSAT was checked against geometrical variations using the standard features for 269 face images.

In [69], the image features are divided into four groups: visual features, statistical pixel features, transform coefficient features, and algebraic features, with emphasis on the algebraic features, which represent the intrinsic attributes of an image. The singular value decomposition (SVD) of a matrix is used to extract the features from the pattern. SVD can be viewed as a deterministic counterpart of the KL transform. The singular values (SV's) of an image are very stable and represent the algebraic attributes of the image, being intrinsic but not necessarily visible. [69] proves their stability and invariance to proportional variance of image intensity in the optimal discriminant vector space,

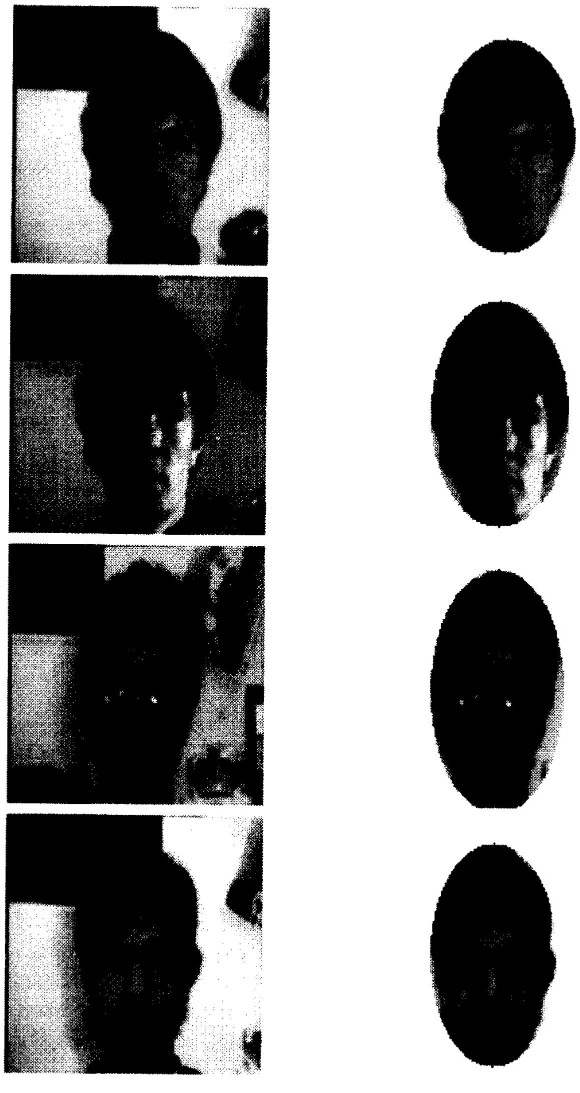


Fig. 5. Results of segmentation. (a) Input image, (b) Extracted image.

to transposition, rotation, translation, and reflection which are important properties of the SV feature vector. The Foley-Sammon transform is used to obtain the optimal set of discriminant vectors spanning the Sammon discriminant plane. For a small set of 45 images of nine persons two of the vectors seem to be adequate for recognition; more discriminant vectors will be needed for recognition with more images. The SVD operation is applied to each image matrix for extracting SV features and the SV vector.

In [100] Nixon uses the Hough transform for feature extraction. The transform locates analytically described shapes by using the magnitude of the gradient and the directional information provided by the gradient operator to aid in the recognition process. Two parts of the eye are attractive for recognition of the eye, the iris, and the perimeter of the eye's sclera. The analytic shape representing the iris is a circle with expected gradient directions in each quadrant, given the lighter background of the sclera. An ellipse appears to be the most suitable shape approximating the perimeter of the sclera, but it is unsatisfactory for those parts of the eye furthest from the center of the face. The ellipse is tailored for each eye's face center by using an exponential function. The gradient magnitudes, obtained using a Sobel operator, are thresholded using four brightness levels to represent the direction of the gradient at that point. The directional information is incorporated into the Hough transform technique. The deviation of the position of the iris center from the estimated value has a mean value of 0.33 pixels. The application of the Hough transform to detect the perimeter of the shape of the region below the eyebrows appears on average to yield a spacing 20% larger than the spacing between the irises. Using the Hough transform to find the sclera shows that the spacing differed on average by minus 1.33 pixels. The results show that it is possible to derive a measurement of the spacing by detecting of the position of both the irises, and the shape describing both the perimeter of the sclera and the eyebrows. The measurement by detection of the position of the iris is most accurate. Detection of the perimeter of the sclera is the most sensitive of the methods.

Yuille, Cohen, and Hallinen in [145] extract facial features using deformable templates. These templates are allowed to translate, rotate and deform to fit the best representation of their shape present in the image. Preprocessing is done to the initial intensity image to get representations of peaks and valleys from the intensity image. Morphological filters are used to determine these representations. Their template for the eye has eleven parameters consisting of the upper and lower arcs of the eye; the circle for the iris; the center points; and the angle of inclination of the eye. This template is fit to the image in an energy minimization sense. Energy functions of valley potential, edge potential, image potential, peak potential, and internal potential are determined. Coefficients are selected for each potential and an update rule is employed to determine the best parameter set. In their experiments they found that the starting location of the template is critical for determining the exact location of the eye. When the template was started above the eyebrow, the algorithm failed to distinguish between the eye and the eyebrow. Another drawback to this approach is its computational complexity. Generally speaking, template based approaches to feature extraction are a more logical approach to take. The problem lies in the description of these templates. Whenever analytical approximations are made to the image, the system has to be tolerant to certain discrepancies between the template and the actual image. This tolerance tends to average out the differences that make individual faces unique.

A statistically motivated approach to detecting and recognizing the human eye in an intensity image with the constraint that the face is in a frontal posture is described in [60]. Hallinan [60] uses a template based approach for detecting the eye in an image. The template is depicted as having two regions of uniform intensity. The first is the iris region and the other is the white region of the eye. The approach constructs an "archetypal" eye and models various distributions as variations of it. For the "ideal" eye a uniform intensity for both the iris and whites is chosen. In an actual eye certain discrepancies from the ideal are found which hamper the uniform intensity choice. These discrepancies can be modeled as "noise" components added to the ideal image. For instance, the white region might have speckled (spot) points depending on scale, lighting direction, etc. Likewise the iris can have within it some "white" spots. The author uses an α-trimmed distribution for both the iris and the white. A "blob" detection system is developed to locate the intensity valley caused by the iris enclosed by the white. Using α-trimmed means and variances and a parameter set for the template of the blob, a cost functional is determined for valley detection. A deformable human eye template is constructed around the valley detection scheme. The search for candidates uses a coarse to fine approach. Minimization is achieved using the steepest descent method. After locating the candidate a goodness of fit criteria is used for verification purposes. The inputs used in the experiments were frontal face intensity images. In all three sets of data were used. One consisted of 25 images used as a testing set, another had 107 positive eyes, and the third consisted of images with most probably erroneous locations which could be chosen as candidate templates. For locating the valleys the author reports as many as 60 false alarms for the first data set, 30 for the second and 110 for the third. An increase in hit rate is reported when using the α-trimmed distribution. The overall best hit rate reported was 80%.

Reisfeld and Yeshurun in [112] use a generalized symmetry operator for the purpose of finding the eyes and mouth in a face. Their motivation stems from the almost symmetric nature of the face about a vertical line through the nose. Subsequent symmetries lie within features such as the eyes, nose and mouth. The symmetry operator locates points in the image corresponding to high values of a symmetry measure discussed in detail in [112]. They indicate their procedure's superiority over other correlation based schemes like that of Baron [14] in the sense that their scheme is independent of scale or orientation. However, since no *a priori* knowledge of face location is used, the search for symmetry points is computationally intensive. The authors mention a success rate of 95% on their face image database, with the constraint that the face occupy between 15–60% of the image.

Manjunath *et al.* [88] present a method for the extraction of pertinent feature points from a face image. It employs Gabor wavelet decomposition and local scale interaction to extract features at points of curvature maxima in the image, corresponding to orientation and local neighborhood. These feature points are then stored in a data base and subsequent target face images are matched using a graph matching technique. The 2D Gabor function used and its Fourier transform are:

$$g(x,y:u_0,v_0) = \exp\{-[x^2/2\sigma_x^2 + y^2/2\sigma_y^2] + 2\pi i[u_0x + v_oy]\} \tag{2}$$

$$G(u,v) = \exp\{-2\pi^2[\sigma_x^2(u-u_0)^2 + \sigma_y^2(v-v_0)^2]\} \tag{3}$$

where σ_x and σ_y represent the spatial widths of the Gaussian and (u_0,v_0) is the frequency of the complex sinusoid.

The Gabor functions form a complete though nonorthogonal basis set. Like the Fourier series, a function $g(x,y)$ can easily be expanded using the Gabor function. Consider the following wavelet representation of the Gabor function:

$$\Phi_\lambda(x,y,\theta) = \exp\{[-\lambda^2(x'^2+y'^2)] + i\pi x'\} \tag{4}$$

$$x' = x\cos\theta + y\sin\theta \tag{5}$$

$$y' = -x\sin\theta + y\cos\theta \tag{6}$$

where θ is the preferred spatial orientation and λ is the aspect ratio of the Gaussian. For convenience the subscripts are dropped in further discussions. In the experiments, λ is set to 1, and θ is discretized into four orientations. The resulting family of wavelets is given by

$$\{\Phi[\alpha^j(x-x_0),\alpha^j(y-y_0),\theta_k]\}, \alpha\in\mathbf{R}, \quad j=\{0,-1,-2,\cdots\} \tag{7}$$

where $\theta_k = k\pi/N$, $N=4$, $k=\{0,1,2,3\}$ and $\alpha^j, j\in\mathbf{Z}$.

Feature detection utilizes a simple mechanism to model the behavior of the end-inhibition. It uses interscale interaction to group the responses of cells from different frequency channels. This results in the generation of the end-stop regions. The orientation parameter θ determines the direction of the edges. Hypercomplex cells in animals are sensitive to oriented lines and step edges of short lengths, and their response decreases if the lengths are increased.

$$I_{m,n}(x,y) = \max_\theta g(\| W_m(x,y,\theta) - \gamma W_n(x,y,\theta)\|) \tag{8}$$

and

$$W_j(x,y,\theta) = f\otimes\Phi(\alpha^jx,\alpha^jy,\theta), \quad j=\{0,-1,-2,\cdots\} \tag{9}$$

where f represents the input image, g is a sigmoid nonlinearity, γ is a normalizing factor, and $n > m$. The final step is to actually localize these features, and this is done by looking at the local maximum of these feature responses. A feature point is selected by taking the maxima in a local neighborhood of the pixel location (x,y). Let the neighborhood be N_{xy}:

$$I_{m,n}(x,y) = \max_{(x',y')\in N_{xy}} I_{m,n}(x',y'). \tag{10}$$

The general idea is to use (9) to determine responses at two scales. These scales act as the hypercomplex cells in

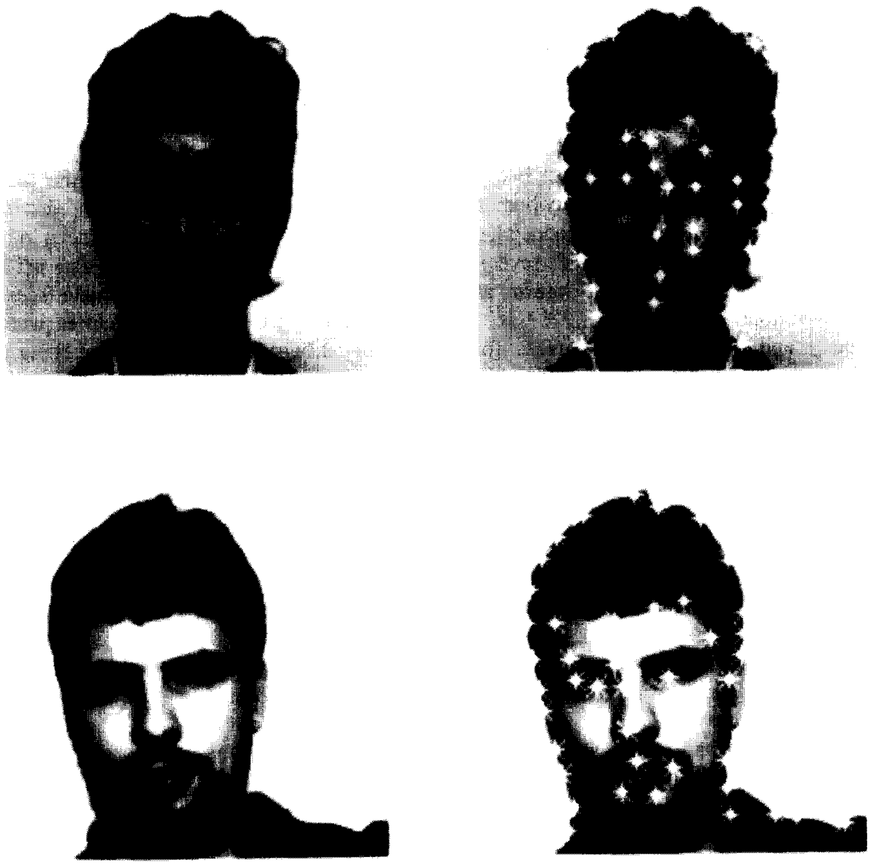


Fig. 6. (a) Input image and (b) feature points extracted.

animals. To determine a high spatial curvature point the response from a larger sized cell is subtracted from the smaller sized cell using (8). A smaller cell will have a higher response for a sharper curvature. This is determined to be a feature point in the image.

Some experimental results for this feature extraction method are shown in Fig. 6. Notice that the background on the image is uniform; this type of image can be seen as representative of passport, driver's license or any identification-type photographs where control over background is easily enforced.

[33] describes a knowledge-based vision system for detection of human faces from hand drawn sketches. The system employs an IF-THEN rule to process its tasks, i.e., "IF: upper mouth line is not found but lower mouth line is found, THEN: look for the upper mouth line in the image area directly above the lower mouth." The template for the face consists of the eyes (both left and right), the nose and the mouth. The processing is done on four different abstraction levels of image information; Line Segment, Component Part, Component, and Face. The line segments are selected as candidates of component parts with probability values associated with them. A component will try to see if a particular area in the image has the necessary component parts (in correct orientations relative to each other) and determine the existence of the component. The Face level will try to determine which geometric layout of the components is best suited to describe a face from the image data. The structure of the system is based on a blackboard architecture; all the tasks have access to (and can write on) to the blackboard. The author reports successful detection of the face using this method with two experiments. The modularity of the system makes it possible to expand it by adding other knowledge sources such as eyebrows, ears, forehead, etc. The usage of sketched images can be extended to the edge map of an intensity image with some processing to get labeled segments, as is done in [123].

C. Recognition

1) Earlier Approaches: One of the earliest works in computer recognition of faces is reported by Bledsoe [18]. In this system, a human operator located the feature points on the face and entered their positions into the computer. Given a set of feature point distances of an unknown person, nearest neighbor or other classification rules were used for identifying the label of the test image. Since feature extraction is manually done, this system could accommodate wide variations in head rotation, tilt, image quality, and contrast.

A landmark work on face recognition is reported in the doctoral dissertation of M. D. Kelly [81]. Kelly's work is similar in framework to that of Bledsoe, but is significantly different in that it does not involve any human intervention. Although we cite this work in connection with face recognition, Kelly's dissertation has made several important contributions to goal directed (also known as top-down) and multiresolution image analysis.

Kelly uses the body and close up head images for recognition. Once the body and head have been outlined as described in Section IV-A, ten measurements are extracted. The body measurements include heights, widths of the head, neck, shoulders, and hips. Measurements from the face include width of the head and distances between eyes, top of head to eyes, between eyes and nose and the distance from eyes to mouth. The nearest neighbor rule was used for identifying the class label of the test image; the leave-one-out [45] strategy was used. The dataset consisted of a total of 72 images, comprised of 24 sets of three images of ten persons. Each set had three images per person; image of the body, image of the background corresponding to the body image and a close-up of the head.

In [80], Kaya *et al.* report a basic study using information theoretic arguments in classifying human faces. They reason from the fact that to represent N different faces a total of $\log_2 N$ bits are required (upper bound on the entropy). They contend that since illumination and background are the same for all face images and the images taken are photographs of front views of human faces, with mouth closed, no beards, and no eyeglasses, therefore the dimensionality of the parameter space can be reduced from the above upper bound. Sixty two photographs were taken with a special apparatus to ensure correct orientation and lighting conditions. An experiment was conducted using 10–40 human subjects to identify prominent geometric features from three different faces. The authors identify nine of these parameters to run statistical experiments on. These parameters form a parameter vector composed of internal biocular breadth, external biocular breadth, nose breadth, mouth breadth, bizygomatic breadth, bigonial breadth, distance between lower lip and chin, distance between upper lip and nose and height of lips. They construct a classifier based on the parameter vector and its estimate, i.e., if $\mathbf{X}$ is the parameter vector then the estimate $\mathbf{Y}$ is given as $\mathbf{Y} = \mathbf{X}+\mathbf{D}$ where $\mathbf{D}$ is the distortion vector. The distortion vector $\mathbf{D}$ has two components $\mathbf{D_m}$, the distortion due to data acquisition and sampling error and $\mathbf{D_i}$ due to inherent variations in facial features. The authors discuss two cases, one in which $\mathbf{D_m}$ is negligible and the other where $\mathbf{D_m}$ is comparable to $\mathbf{D_i}$. For each parameter a threshold is determined from its statistical behavior. Classification is done using the absolute norm between a stored parameter set and the input image parameter values. It should be noted that the parameter values are determined manually. The authors then set a bound on the probability of finding a correct match, using some arbitrary constants, to be about 90% from 15 000 images. However, this is just an extrapolation of the results that they obtained from the sixty two images that were tested and not a result of actual experiments.

One method of characterizing the face is the use of geometrical parameterization, i.e., distances and angles between points such as eye corners, mouth extremities, nostrils, and chin top [78]. The data set used by Kanade consists of 17 male and three female faces without glasses, mustaches, or beards. Two pictures were taken of each individual, with the second picture being taken one month later in a different setting. The face-feature points are located in two stages. The coarse-grain stage simplified the succeeding differential operation and feature-finding algorithms. Once the eyes, nose and mouth are approximately located, more accurate information is extracted by confining the processing to four smaller regions, scanning at higher resolution, and using the "best beam intensity" for the region. The four regions are the left and right eye, nose, and mouth. The beam intensity is based on the local area histogram obtained in the coarse-grain stage. A set of 16 facial parameters which are ratios of distances, areas, and angles to compensate for the varying size of the pictures is extracted. To eliminate scale and dimension differences the components of the resulting vector are normalized. The entire data set of 40 images is processed and one picture of each individual is used in the training set. The remaining 20 pictures are used as a test set. A simple distance measure is used to check for similarity between an image of the test set and the image in the reference set. Matching accuracies range from 45% to 75% correct, depending on the parameters used. Better results are obtained when several of the ineffective parameters are not used [78].

2) Statistical Approach: Turk and Pentland [133] used eigenpictures (also known as "eigenfaces" (see Fig. 7) in [133]) for face detection and identification. Given the eigenfaces, every face in the database can be represented as a vector of weights; the weights are obtained by projecting the image into eigenface components by a simple inner product operation. When a new test image whose identification is required is given, the new image is also represented by its vector of weights. The identification of the test image is done by locating the image in the database whose weights are the closest (in Euclidean distance) to the weights of the test image. By using the observation that the projection of a face image and a nonface image are quite different, a method for detecting the presence of a face in a given image is obtained. Turk and Pentland illustrate their method using a large database of 2500 face images of 16 subjects, digitized at all combinations of three head orientations, three head sizes and three lighting conditions. Several experiments were conducted to test the robustness of the approach to variations in lighting, size, head orientation, and the differences between the training and test conditions. The authors reported 96% correct classification over lighting variations, 85% over orientation

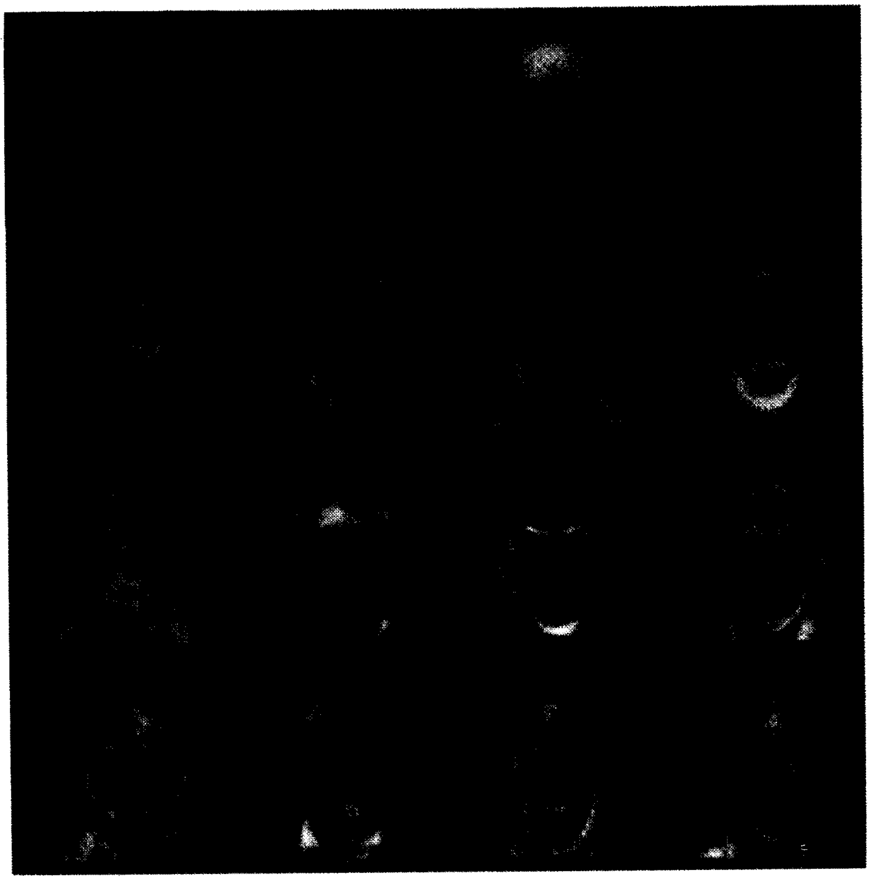


Fig. 7. Eigenfaces [133].

variations and 64% over size variations. It can be seen that the approach is fairly robust to changes in lighting conditions, but degrades quickly as the scale changes. One can explain this by the significant correlation present between images with changes in illumination conditions; the correlation between face images at different scales is rather low. Another way to interpret this is that the approach based on eigenfaces will work well as long as the test image is "similar" to the ensemble of images used in the calculation of eigenfaces. Turk and Pentland also extend their approach to real time recognition of a moving face image in a video sequence. A spatiotemporal filtering step followed by a nonlinear operation is used to identify a moving person. The head portion is then identified using a simple set of rules and handed over to the face recognition module.

In [104], Pentland *et al.* extend the capabilities of their earlier system [133] in several directions. They report extensive tests based on 7562 images of approximately 3000 people, the largest database on which any face recognition study has been reported to date. Twenty eigenvectors were computed using a randomly selected subset of 128 images. In addition to eigenrepresentation, annotated information on sex, race, approximate age and facial expression was included. Unlike mug shots applications, where only one front and one side view of a person's face is kept, in this database several persons have many images with different expressions, head wear, etc.

One of the applications the authors consider is interactive search through the database. When the system is asked to present face images of certain types of people (e.g., white females of age 30 years or younger), images that satisfy this query are presented in groups of 21. When the user chooses one of these images, the system presents faces from the database that look similar to the chosen face in the order of decreasing similarity. In a test involving 200 selected images, about 95% recognition accuracy was obtained—i.e., for 180 images the most similar face was of the same person. To evaluate the recognition accuracy as a function of race, images of white, black and Asian adult males were tested. For white and black males accuracies of 90% and 95% were reported, respectively, while only 80% accuracy was obtained for Asian males. The use of eigenfaces for personnel verification is also illustrated.

In mug shots applications, usually a frontal and a side view of a person are available. In some other applications, more than two views may be available. One can take two approaches to handling images from multiple views. The first approach will pool all the images and construct a set of eigenfaces that represent all the images from all the views. The other approach is to use separate eigenspaces for different views, so that the collection of images taken from each view will have its own eigenspace. The second approach, known as the view-based eigenspace, seems to perform better. For mug shots applications, since two or at most three views are needed, the view-based approach produces two or three sets of eigenspaces.

The concept of eigenfaces can be extended to eigenfeatures, such as eigeneyes, eigenmouth, etc. Just as eigenfaces were used to detect the presence of a face in [133], eigenfeatures are used for the detection of features such as eyes, mouth etc. Detection rates of 94%, 80%, and 56% are reported for the eyes, nose and mouth, respectively, on the large dataset with 7562 images.

Using a limited set of images (45 persons, two views per person, corresponding to different facial expressions such as neutral versus smiling), recognition experiments as a function of number of eigenvectors for eigenfaces only and for the combined representation were performed. The eigenfeatures performed as well as eigenfaces; for lower order spaces, the eigenfeatures fared better; when the combined set was used, marginal improvement was obtained. As summarized in Section III, both holistic and feature-based mechanisms are employed by humans. The feature based mechanisms may be useful when gross variations are present in the input image; the authors' experiments support this.

The effectiveness of standardized KL coefficients such as KL-IPAT and KL-FSAT has been illustrated in [11] using two experiments. In the first experiment, the training and testing samples were acquired under as similar conditions as possible. The test set consisted of five samples from 20 individuals. The KL-IPAT had an accuracy rate of 85% and the KL-FSAT had an accuracy rate of 91%. Both methods misidentified the one example where there is a difference in the wearing and not wearing of glasses between the testing set and the training set. The second experiment checks for feature robustness when there is a variation caused by an error in the positioning of the target window. This is an error usually made during image acquisition due to changing conditions. The test images are created by shifting the reference points in various directions by one pixel. The variances for 4 and 8 pixels are tested. The KL-IPAT having an error rate of 24% for the 4 pixel difference and 81% for the 8 pixel difference. The KL-FSAT had an 4% error rate for the 4 pixel difference and a 44% error rate for the 8 pixel difference. The improvement is due to the shift invariance property in the Fourier spectrum domain. The third experiment used the variations in head positioning. The test samples were taken while the subject was nodding and shaking his head. The KL-FSAT showed high robustness over the KL-IPAT for the different orientations of the head. Good recognition performance was achieved by restricting the image acquisition parameters. Both the KL-IPAT and KL-FSAT have difficulties when the head orientation is varied [11].

The effectiveness of SVD for face recognition has been tested in [32], [69]. The optimal discriminant plane and quadratic classifier of the normal pattern is constructed for the 45 SV feature vector samples. The classifier is able to recognize the 45 training samples of the nine subjects. Testing was done using 13 photos which consisted of nine newly sampled photos of the original test subjects with two of one subject and three samples of the subject at different ages. There was a 42.67% error rate which Hong feels was due to the statistical limitations of the small number of training samples [69].

In [32] the SV vector is compressed into a low dimensional space by means of various transforms, the most popular being an optimal discriminant transform based on Fisher's criterion. The Fisher optimal discriminant vector represents the projection of the set of samples on a direction φ, chosen so that the patterns have a minimal scatter within each class and a maximal scatter between classes in the 1D space. Three SV feature vectors are extracted from the training set in [32]. The optimal discriminant transform compresses the high-dimensional SV feature space to a new r-dimensional feature space. The new secondary features are algebraically independent and informational redundancy is reduced. This approach was tested on 64 facial images of eight people (the classes). The images were represented by Goshtasby's shape matrices, which are invariant to translation, rotation, and scaling of the facial images and are obtained by polar quantization of the shape [54]. Three photographs from each class were used to provide a training set of 24 SV feature vectors. The SV feature vectors were treated with the optimal discriminant transform to obtain new feature vectors for the 24 training samples. The class center vectors were obtained using the second feature vectors. The experiment used six optimal discriminant vectors. The separability of training set samples was good with 100% recognition. The remaining 40 facial images were used as the test set, five from each person. Changes were made in the camera position relative to the face, the camera's focus, the camera's aperture setting, the wearing or not wearing of glasses, and blurring. As with the training set, the SV feature vectors were extracted, and the optimal discriminant transform was applied to obtain the transformed feature vector. Again good separability was obtained with an accuracy rate of 100% [32].

Cheng *et al.* [31] develop an algebraic method for face recognition using SVD and thresholding the eigenvalues thus obtained to some value greater than a set threshold value. They use a projective analysis with the training set of images serving as the projection space. A training set in their experiments consists of three instances of face images of the same person. If $A \in \mathcal{R}^{m\times n}$ represents the image, and $A_j^{(i)}$ represents the jth face image of person i, then the average image for person i is given by $(1/N)\ \sum_{j=1}^{N} A_j^{(i)}$. Eigenvalues and eigenvectors are determined for this average image using SVD. The eigenvalues are thresholded to disregard the values close to zero. Average eigenvectors (called feature vectors) for all the average face images are calculated. A test image is then projected onto the space spanned by the eigenvectors. The Frobenius norm is used as a criterion to determine which person the test image belongs to. The authors reported 100% accuracy when working with a database of 64 face images of eight different persons. Each person contributed eight images. Three images from each person were used to determine the feature vector for the face image in question. Eight such feature vectors were determined. They state that the projective distance of the testing image sample was markedly minimum for the correct training set image.

The use of isodensity lines, i.e., curves of constant gray level, for face recognition has been investigated in [98]. Such lines, although they are not directly related to the 3D structure of a face, do provide a relief image of the face. Using images of faces taken with a black background, a Sobel operator and some post-processing steps are used to obtain the boundary of the face region. The gray level histogram (an 8-bin histogram) is then used to trace contour lines on isodensity levels. A template matching procedure is used for face recognition. The method has been illustrated using ten pairs of face images, with three pairs of pictures of men with spectacles, two pairs of pictures of men with thin beards, and two pairs of pictures of women. 100% recognition accuracy was reported on this small data set.

3) Neural Networks Approach: The use of neural networks (NN) in face recognition has addressed several problems: gender classification, face recognition, and classification of facial expressions. One of the earliest demonstrations of NN for face recall applications is reported in Kohonen's associative map [84]. Using a small set of face images, accurate recall was reported even when the input image is very noisy or when portions of the images are missing. This capability was demonstrated using optical hardware by Psaltis's group [6].

A single layer adaptive NN (one for each person in the database) for face recognition, expression analysis and face verification is reported in [128]. Named Wilkie, Aleksander, and Stonham's recognition device (WISARD), the system needs typically 200–400 presentations for training each classifier, the training patterns included translation and variation in facial expressions. Sixteen classifiers were used for the dataset constructed using 16 persons. Classification is achieved by determining the classifier that gives the highest response for the given input image. Extensions to face verification and expression analysis are presented. The sample size is small to make any conclusions on the viability of this approach for large datasets involving a large number of persons.

In [51], Golomb, Lawrence, and Sejnowski present a cascade of two neural networks for gender classification. The first stage is an image compression NN whose hidden nodes serve as inputs to the second NN that performs gender classification. Both networks are fully connected, three-layer networks with two biases and are trained by a standard back-propagation algorithm. The images used for testing and training were acquired such that facial hair, jewelry and makeup were not present. They were then preprocessed so that the eyes are level and the eyes and mouth are positioned similarly. A 30 × 30 cropped block of pixels was extracted for training and testing. The dataset consisted of 45 males and 45 females; 80 were used for training, with 10 serving as testing examples. The compression network indirectly serves as a feature extractor; in that the activities of 40 hidden nodes (in a 900 × 40 × 900 network) serve as features for the second network, that performs gender classification. The hope is that due to the nonlinearities in the network, the feature extraction step may be more efficient than the linear KL methods. The gender classification network is a $40 \times n \times 1$ network, where the number n of hidden nodes has been 2, 5, 10, 20, or 40. Experiments with 80 training images and 10 testing images have shown the feasibility of this approach. This method has also been extended to classifying facial expressions into eight types.

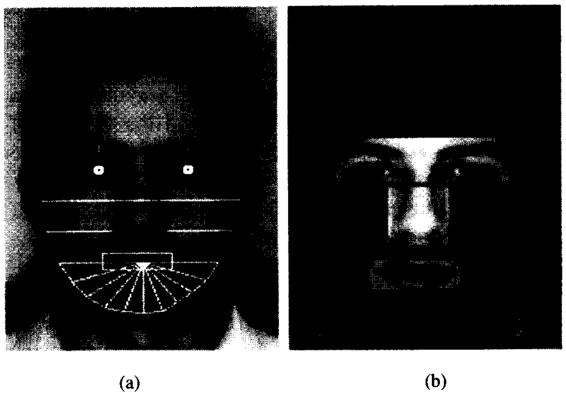


Fig. 8. Radius vectors and other feature points [22].

Using a vector of 16 numerical attributes (Fig. 8) such as eyebrow thickness, widths of nose and mouth, six chin radii, etc., Brunelli and Poggio [21] also develop a NN approach for gender classification. They train two HyperBF networks [109], one for each gender. The input images are normalized with respect to scale and rotation by using the positions of the eyes which are detected automatically. The 16D feature vector is also automatically extracted. The outputs of the two HyperBF networks are compared, the gender label for the test image being decided by the network with greater output. In the actual classification experiments only a subset of the 16D feature vector is used. The database consists of 21 males and 21 females. The leave-one-out strategy [45] was employed for classification. When the feature vector from the training set was used as the test vector, 92.5% correct recognition accuracy was reported; for faces not in the training set, the accuracy further dropped to 87.5%. Some validation of the automatic classification results has been reported using humans.

By using an expanded 35D feature vector, and one HyperBF per person, the gender classification approach has been extended to face recognition. The motivation for the underlying structure is the concept of a grandmother neuron: a single neuron (the Gaussian function in the HyperBF network) for each person. As there were relatively few training images per person, a synthetic data base was generated by perturbing around the average of the feature vectors of available persons and the available persons were used as testing samples. For different sets of tuning parameters (coefficients, centers and metrics of the HyperBF's) classification results are reported. Some corroboration of the caricatural behavior of the HyperBF networks, by psychophysical studies, is also presented.

In [22], Brunelli and Poggio compare the merits of both feature based and template based approaches. Their feature based approach is motivated by [78] and [80]. They determine 35 features which are also used in [109] [see also Fig. 8(a)]. They mention the use of various classifiers to accomplish the task of matching these features, namely Bayes, nearest neighbor, or the HyperBF. For the template based approach they have selected various regions of the face as templates and used a correlation based matching technique [see Fig. 8(b)]. From their experiments they concluded that the template based approach, though computationally complex, was superior on their database over the feature based approach. An accuracy of 100% for the template based approach compared to 90% for the feature based one.

The use of HyperBF networks for face recognition is also reported in [21]. To remove variations due to changing viewpoint, the images are first transformed using 2D affine transforms. The transformation parameters are obtained by using the detected positions of the eyes and mouth in the given image and the desired positions of these features. The transformed image is then subjected to a directional derivative operator to reduce the effects of illumination. The resulting image is multiplied by a Gaussian function and integrated over the receptive field to achieve dimensionality reduction. The MIT Media Lab database of 27 images, of each of 16 different persons was used, with the images of 17 persons being used for training, while the rest were used as testing samples. A HyperBF was trained for each person. An average accuracy of 79% was reported compared with 90% accuracy when tested with human subjects. By feeding the outputs of 16 HyperBF's to another HyperBF, significant reductions in error rates were reported.

[111] presents the results of work using a connectionist model of facial expression. The model uses the pyramid structure to represent image data. Each level of the pyramid is represented by a network consisting of one input, one hidden, and one output layer. The input layers of the middle levels of the pyramid are the outputs of the previous level's hidden units when training is complete. Network training at the lowest level is carried out conventionally. Each network is trained using a fast variation of the back propagation learning algorithm. The training pattern set for the subsequent levels is obtained by combining and partitioning the hidden units' outputs of the preceding level. The original images of the training set are partitioned into blocks of overlapping squares. The overlapping blocks simulate the local receptive fields of the human visual system. Each block consists of the set of block patterns partitioned in the same positions over the image pattern set. The data set for training consists of six hand drawn faces with six different expressions: happiness, surprise, sadness, anger, fear, and normal. The outer features of each face are its shape and the ears. The inner features are the eyebrows, the eyes, the nose, and the mouth. Each face is drawn to be as dissimilar as possible from the others. The testing set consists of the six training faces and the images from the training set masked with a horizontal bar across the upper, middle, and lower portions of the face covering approximately 20% of the total image. The horizontal bar is used to demonstrate the network's associative memory capability. The network has four levels. Levels 1–3 consist of 25 input units, six hidden units, and 25 output units. The fourth level has 18 input units, eight hidden units, and 25 output units. The network training process at each level results in a different representation of the original image data. The last level of the pyramid has the leanest and most abstract representation. The representation is viewed as a unique identification of the face and the information it conveys. The network is able to successfully recognize the members of the training set when tested on them. The network poorly recognizes (50%) the various masked, blurred, or distorted facial expressions. It is unable to recognize the various masks of the happy face. The error rate is the result of obtaining a totally different abstract representation which the network has not learned. On analysis of the hidden units, patches are found. The patches block off some of the features of the faces and appear unimportant to the hidden node. The hidden units' internal representations show that many of them are in the form of eigenfeatures where the features of the faces are combined in an overlaying manner on top of each other. The eigenfeatures are only a portion of all the features. In the happy face the blocked patches of the hidden units are mainly outside of the face while the others are inside the face. This may be explained by the fact that the happy face does not have many facial features in common with the other faces in the training set. It appears that the network developed a holistic representation of the happy face so that it could be recognized. The leaner representations of the face are automatically generated and are a unique identification of the learned object. The unique representation may be associated with the original object in the form of one-to-many. The model is able to successfully identify the same face but not the masked faces of the same type. The masking of areas shows where the network's learning is focused. It appears that the middle portion of the face image is not as important as the upper and lower portions and may be used to develop a focus of attention [111].

The systems presented in [24] and [85] are based on the dynamic link architecture (DLA). DLA's attempt to solve some of the conceptual problems of conventional artificial neural networks, the most prominent problem being the expression of syntactical relationships in neural

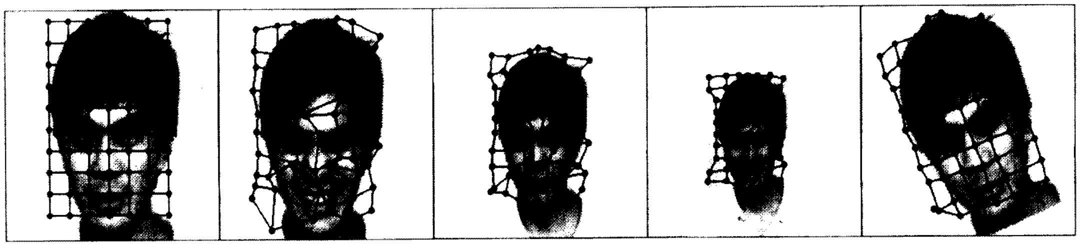


Fig. 9. System for DLA.

networks. DLA's use synaptic plasticity and are able to instantly form sets of neurons grouped into structured graphs and maintain the advantages of neural systems. A DLA permits pattern discrimination with the help of an object-independent standard set of feature detectors, automatic generalization over large groups of symmetry operations, and the acquisition of new objects by one-shot learning, reducing the time-consuming learning steps. Invariant object recognition is achieved with respect to background, translation, distortion and size by choosing a set of primitive features which is maximally robust with respect to such variations. Both [24] and [85] use Gabor based wavelets for the features. The wavelets are used as feature detectors, characterized by their frequency, position, and orientation. Two nonlinear transforms are used to help during the matching process. A minimum of two levels, the image domain and the model domain, are needed for a DLA. The image domain corresponds to primary visual cortical areas and the model domain to the intertemporal cortex in biological vision. The image domain consists of a 2D array of nodes $A^I_x = \{(x, \alpha)$, where $\alpha = 1, \cdots, F\}$. Each node at position x consists of F different feature detector neurons (x, α) that provide local descriptors of the image. The label α is used to distinguish different feature types. The amount of feature type excitation is determined for a given node by convolving the image with a subset of the wavelet functions for that location. Neighboring nodes are connected by links, encoding information about the local topology. Images are represented as attributed graphs. Attributes attached to the graph's nodes are activity vectors of local feature detectors. An object in the image is represented by a subgraph of the image domain. The model domain is an assemblage of all the attributed graphs, being idealized copies of subgraphs in the image domain. Excitatory connections are between the two domains and are feature preserving. The connection between domains occurs if and only if the features belong to corresponding feature types. The DLA machinery is based on a data format which is able to encode information on attributes and links in the image domain and to transport that information to the model domain without sending the image domain position. The structure of the signal is determined by three factors: the input image, random spontaneous excitation of the neurons, and interaction with the cells of the same or neighboring nodes in the image domain. Binding between neurons is encoded in the form of temporal correlations and is induced by the excitatory connections within the image. Four types of bindings are relevant to object recognition and representation: Binding all the nodes and cells together that belong to the same object, expressing neighborhood relationships with the image of the object, bundling individual feature cells between features present in different locations, and binding corresponding points in the image graph and model graph to each other. DLA's basic mechanism, in addition to the connection parameter between two neurons, is a dynamic variable (J) between two neurons (i, j). J-variables play the role of synaptic weights for signal transmission. The connection parameters merely act to constrain the J-variables. The connection parameters may be changed slowly by long-term synaptic plasticity. The connection weights J_{ij} are subject to a process of rapid modification. J_{ij} weights are controlled by the signal correlations between neurons i and j. Negative signal correlations lead to a decrease and positive signal correlations lead to an increase in J_{ij}. In the absence of any correlation, J_{ij} slowly returns to a resting state. Rapid network self-organization is crucial to the DLA. Each stored image is formed by picking a rectangular grid of points as graph nodes. A locally determined jet for each of these nodes is stored and used as the pattern class. New image recognition takes place by transforming the image into the grid of jets, and all stored model graphs are tentatively matched to the image. Conformation of the DLA is done by establishing and dynamically modifying links between vertices in the model domain. During the recognition process an object is selected from the model domain. A copy of the model graph is positioned in a central position in the image domain. Each vertex in the model graph is connected to the corresponding vertex in the image graph. The match quality is evaluated using a cost function. The image graph is scaled by a factor while keeping the center fixed. If the total cost is reduced the new value is accepted. This is repeated until the optimum cost is reached. The diffusion and size estimation are repeated for increasing resolution levels and more of the image structure is taken into account. Recognition takes place after the optimal total cost is determined for each object. The object with the best match to the image is determined. Identification is a process of elastic graph matching. In the case of faces, if one face model matches significantly better than all competitor models, the face in the image is considered as recognized. The system identifies a person's face by comparing an extracted graph with a set of stored graphs. In [24] the experiment consists of a gallery of over 40 different face images. With little effort to standardize the images, the system's recognition success is remarkably consistent. The system shows that a neural system gains power when provided with a mechanism for grouping. The system used in [85] has a larger gallery of faces and recognizes them under different types of distortion and rotation in depth, achieving less than 5% false assignments. Lades *et al.* state that when a clear criterion for the significance for the recognition process is determined, all false assignments are rejected and no image is accepted if its corresponding model is temporarily removed from the gallery. This means that the capacity of the gallery to store distinguishable objects is certainly larger than its present size. No limits to this capacity other than a linear increase in computation time have been encountered so far. Most of the time is spent on image transformation and on optimizing the map between the image and individual stored models [24], [85].

4) Feature Matching Approach: Manjunath *et al.* [88] store feature points detected using the Gabor wavelet decomposition into data files for each image. This greatly reduces the storage requirements for the database. Typically 35–45 points per face image are generated and stored. The identification process utilizes the information present in a topological graphic representation of the feature points. After compensating for differing centroid locations, two cost values are evaluated. One is the topological cost and the other a similarity cost.

The identification process utilizes the information present in a topological graphic representation of the feature points. The feature points are represented by nodes V_i where $i = \{1, 2, 3, \cdots\}$, a consistent numbering technique. The information about a feature point is contained in $\{S, \mathbf{q}\}$, where S represents the spatial location and q is the feature vector defined by

$$\mathbf{q_i} = [Q_i(x, y, \theta_1), \cdots, Q_i(x, y, \theta_N)] \qquad (11)$$

corresponding to the ith feature point. The vector q_i is a set of spatial and angular distances from feature point i to its N nearest neighbors denoted by $Q_i(x, y, \theta_j)$, where j is the jth of the N neighbors. N_i represents a set of neighbors which are of consequence for the feature point in question. The neighbors satisfying both maximum number N and minimum Euclidean distance d_{ij} between two points V_i and V_j are said to be of consequence for the ith feature point.

To identify an input graph with a stored one which is different, either in total number of feature points or in the location of the respective faces, we proceed in a stepwise manner. If i, j refer to nodes in the input graph $\mathcal{I}$ and x', y', m', n' refer to nodes in the stored graph $\mathcal{O}$ then the two graphs are matched as follows:

1) The centroids of the feature points of $\mathcal{I}$ and $\mathcal{O}$ are aligned.

2) Let N_i be the ith feature point $\{V_i\}$ of $\mathcal{I}$. Search for the best feature point $\{V_{i'}\}$ in $\mathcal{O}$ using the criterion

$$S_{ii'} = 1 - \frac{\mathbf{q}_i \cdot \mathbf{q}_{i'}}{||\mathbf{q}_i||\,||\mathbf{q}_{i'}||} = \min_{m' \in N_i} S_{im'}. \qquad (12)$$

3) After matching, the total cost is computed taking into account the topology of the graphs. Let nodes i and j of the input graph match nodes i' and j' of the stored graph and let $j \in N_i$ (i.e., V_j is a neighbor of V_i). Let $\rho_{ii'jj'} = \min\{d_{ij}/d_{i'j'}, d_{i'j'}/d_{ij}\}$. The topology cost is given by

$$T_{ii'jj'} = 1 - \rho_{ii'jj'}. \qquad (13)$$

4) The total cost is computed as

$$C_1(\mathcal{I}, \mathcal{O}) = \sum_i S_{ii'} + \lambda_t \sum_i \sum_{j \in N_i} T_{ii'jj'} \qquad (14)$$

where λ_t is a scaling parameter assigning relative importance to the two cost functions.

5) The total cost is scaled appropriately to reflect the possible difference in the total number of the feature points between the input and stored graph. If $n_\mathcal{I}, n_\mathcal{O}$ are the numbers of feature points in the input and stored graph respectively, then the scaling factor $s_f = \max\{n_\mathcal{I}/n_\mathcal{O}, n_\mathcal{O}/n_\mathcal{I}\}$ and the scaled cost is $C(\mathcal{I}, \mathcal{O}) = s_f C_1(\mathcal{I}, \mathcal{O})$.

6) The best candidate is the one with the least cost, i.e., it satisfies

$$C(\mathcal{I}, \mathcal{O}^*) = \min_{\mathcal{O}'} C(\mathcal{I}, \mathcal{O}'). \qquad (15)$$

The recognized face is the one that has the minimum of the combined cost value. An accuracy of 94% is reported. The method shows a dependency on the illumination direction and works on controlled background images like passport and drivers license pictures. Fig. 10 shows a set of input and identified images for this method.

Seibert and Waxman [116] have proposed a system for recognizing faces from their parts using a neural network. The system is similar to a modular system they have developed for recognizing 3D objects [117] by combining 2D views from different vantage points; in the case of faces, arrangement of features such as eyes and nose play the role of the 2D views. The processing steps involved are segmentation of a face region using interframe change detection techniques, extraction of features such as eyes, mouth, etc., using symmetry detection, grouping and log-polar mapping of the features and their attributes such as centroids, encoding of feature arrangements, clustering of feature vectors into view categories using ART 2, and integration of accumulated evidence using an aspect network.

In a subsequent paper Seibert and Waxman [118] exploit the role of caricatures and distinctiveness (summarized in Section III of the report) in human face recognition to

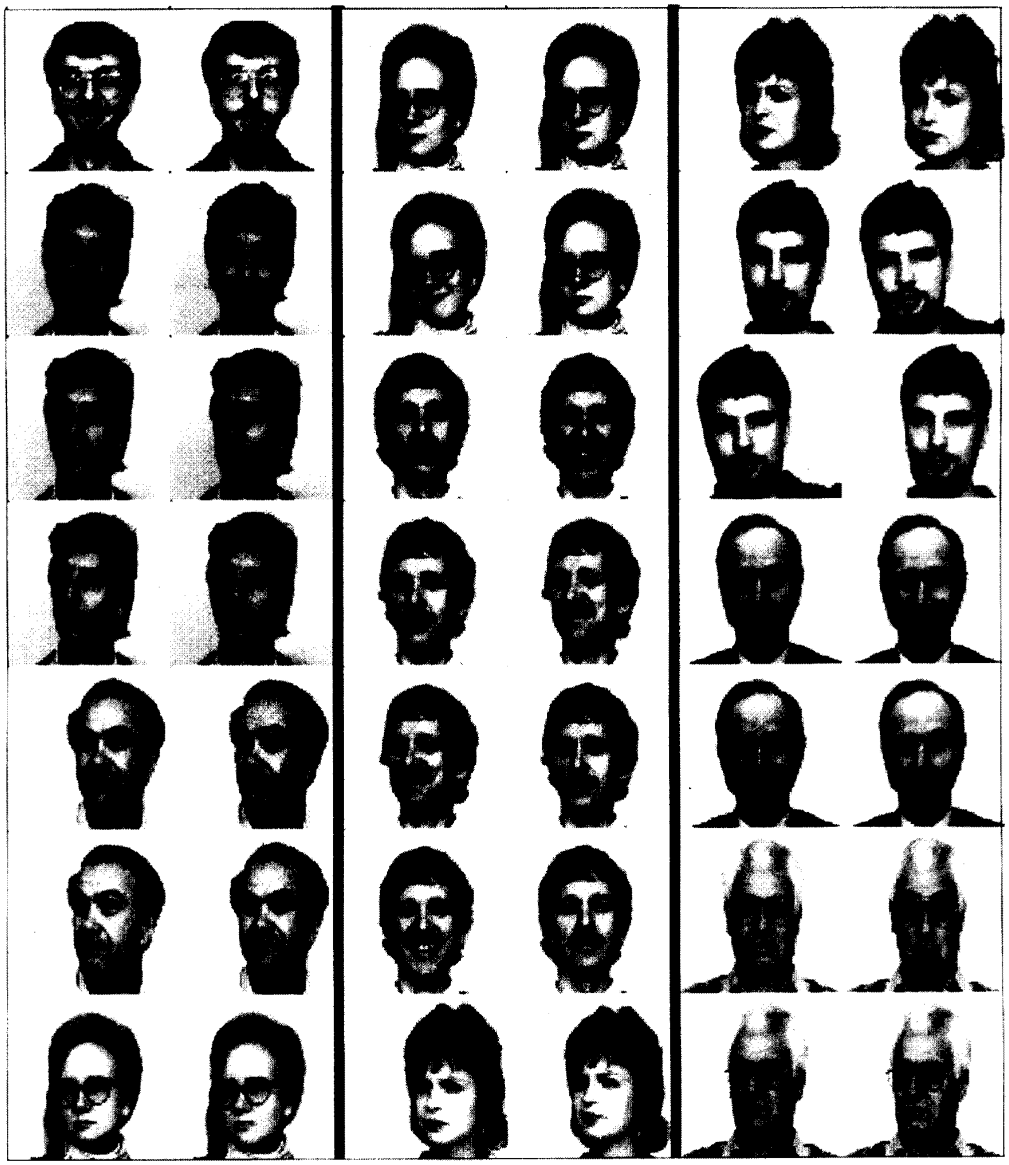


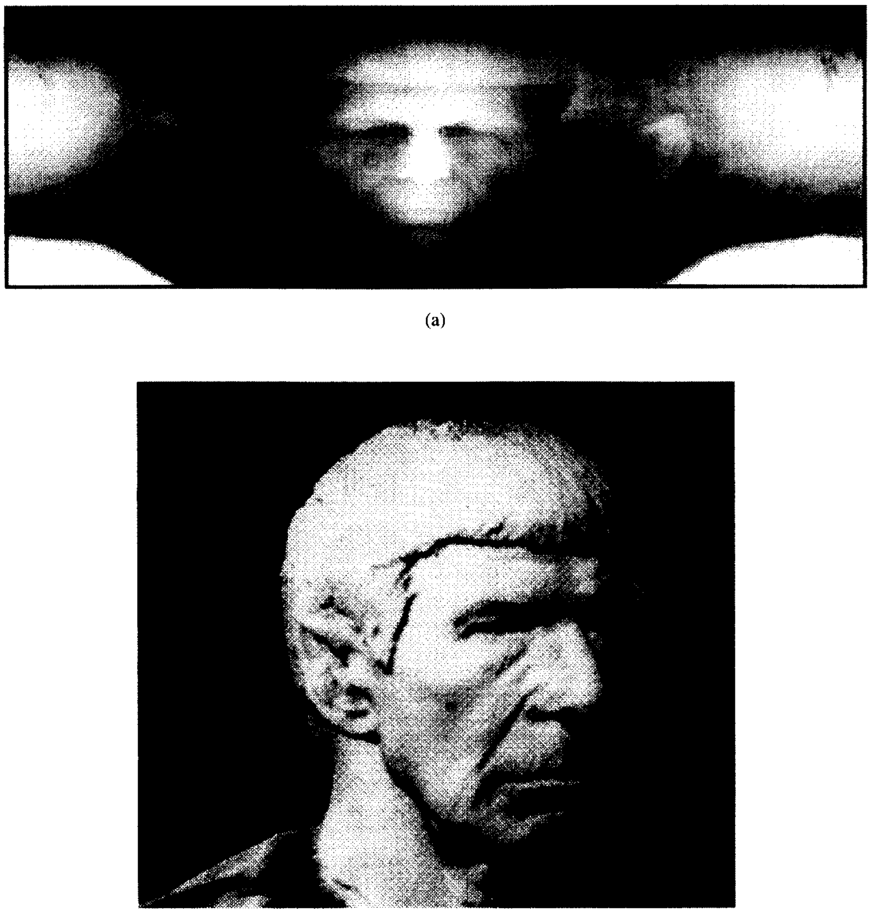
Fig. 10. Some results of the recognition system in Manjunath *et al.*

enhance the capabilities of their previously reported face recognition system. Both of their papers are preliminary reports and lack experimental validation using a large dataset of faces.

Huang *et al.* have been working on a system for detection and recognition of human faces in still monochromatic images. A rule-based algorithm is first used to locate faces in the image [144]. Then each face is recognized by a neural network like structure called Cresceptron [139]. The Cresceptron has a multi-resolution pyramid structure. It is on the surface similar to Fukushima's Neocognitron; however, the learning in cresceptron is completely automatic, and incremental. In a small-scale experiment involving 50 persons, the Cresceptron performs well.

D. Range Images

The discussion so far has revolved around face recognition methods and systems which use data obtained from a 2D intensity image. Another topic being studied by researchers is face recognition from range image data. A range image contains the depth structure of the object in question. Although such data is not available in most applications, it is still important to note the benefit of the added information present in range data in terms of accuracy of the face recognition system. For further study, the reader is encouraged to read some of the selected papers presented in the bibliography at the end of this report.



(a)

(b)

Fig. 11. (a) Depth of face parameterized as $f(\theta, y)$ (Leonard Nimoy as Spock), (b) rendered polygonal model of face composed from coarse sampling of depth data [52].

Gordon in [52] describes a template based recognition system involving descriptors determined from curvature calculations of range image data. The data is obtained from a rotating laser scanner system with resolution of better then 0.4 mm. Segmentation is done by classifying surfaces into planar regions, spherical regions and surfaces of revolution. The image data is stored in a cylindrical coordinate system as $f(\theta, y)$. An example of such data is shown in Fig. 11. At each point on the surface of the curve the magnitude and direction of the minimum and maximum normal curvatures are calculated. Since the calculations involve second order derivatives, smoothing is required to remove the affect of noise in the image. This smoothing is achieved by a Gaussian smoothing filter.

The segmentation produces four surface regions: one convex, one concave and two saddle regions. Ridges and valley lines are determined by obtaining the maxima and minima of the curvatures. These as a whole represent the information pertinent to feature location for an individual. Next comes a comparison strategy as applied to face recognition.

- The nose is located.
- Locating the nose facilitates the search for the eyes and mouth.
- Other features such as forehead, neck, cheeks, etc., are determined by their surface smoothness (unlike hair or eye regions).
- This information is then used for depth template comparison. Using the location of the eyes, nose and mouth the faces are normalized into a standard position. This standard position is reinterpolated to a regular cylindrical grid and the volume of space between two normalized surfaces is used as the criterion for a match.

The system was tested on a dataset of 24 images of eight persons with three views of each. The data represented four male and four female faces. Sufficient feature detection was achieved for 100% of these faces. For recognition 97% accuracy is reported for individual features rather than the whole face (which yielded 100% accuracy). In a related work [53], the process of finding the features is formalized for recognition purposes.

E. Summary

Significant progress has been achieved in segmentation, feature extraction and recognition of faces in intensity images. As range images or stereo pairs may not be available in most of the commercial and law enforcement applications, the face recognition problem, by and large, may be viewed as a 2D image matching and recognition problem with provisions for two or three views of a person's face. Given that in mug shots, drivers' licenses, and personal ID cards, the backgrounds are relatively uncluttered, and only the face is present, the segmentation of the face image, for subsequent processing, could be reasonably handled by any one of the methods in Section IV-A. Segmentation becomes more difficult when beards, baldness are present; also if a face has to be segmented in a cluttered scene where other objects are present, the techniques presented in [55] and [123] may be applicable. It should be pointed out that a strict bottom up procedure may use segmentation as the first step. Techniques based on KL transforms often sidestep this issue, treating the entire image including the background as a pattern. This strategy may be appropriate when similar homogenous backgrounds are present, but if widely varying backgrounds are present, more KL features will be required.

As discussed in Section III, both holistic and independent features contribute to face recognition. The notion of eigenfaces, and eigenfeatures such as eigenmouths, eigeneyes, etc., captures both holistic and independent features in a unified way. Methods based on deformable templates, and their variants for the extraction of eyes and mouth, suffer from dependence on a large number of parameters and also depend on good initial placement of the different templates. One of the more serious concerns with the deformable templates approach is its computational complexity. The eigenfeature approach looks more promising as one can roughly construct a region around an eye or both eyes and mouth and perform eigenanalysis. From an aesthetic point of view, the eigenapproach is not appealing as structure information is coded purely in terms of numbers, but has advantages of being rather straightforward from a computational point of view. The features extracted using Gabor wavelets capture some types of holistic attributes in that they represent the face as a cluster of points; location of these points in and around eyes, mouth, etc., could be quite accidental, although one can use masks around these regions to highlight point features. The numerical features extracted from eyes, nose, mouth, and chin regions concentrate more on the lower parts of the face region, which may be adequate for gender classification. It appears that in face recognition, the upper parts of the face play a more important role.

There is not much of a consensus as to what should be coded to represent a face. The studies alluded to in Section III point out the significance of different internal features, but do not say much about how these features are coded numerically or in subjective terms such as thick eyebrow, wide nose, narrow mouth, etc. Many systems that do face reconstruction for witness recall utilize such subjective descriptions.

In the face recognition and identification area the eigenfaces and eigenfeatures approach of Pentland and his students seems to be the most tested system, using several thousands of images. Their eigenfunction approach has been shown to be useful in personal identification search through a database, recognition from multiple views, etc. The interesting aspect of this approach is that one can develop multiple eigenrepresentations corresponding to not only different views but also corresponding to different races, age groups, gender, etc. It is almost always true that in mug shot and other similar applications such information is available. This enables efficient representation of a potentially very large number of people. It is our view that the eigenface and feature point based approaches are the most developed and tested ones yet and deserve very serious consideration for evaluation in real applications involving hundreds of thousands of examples.

Techniques based on NN also deserve more study. It is our view that NN-based methods can potentially incorporate both numerical and structural information relevant to face recognition; all of the existing work on NN approaches to FRT have demonstrated this on limited sets of images. In addition, the ability to generalize and recognize using incomplete information gives NN classifiers significant advantages over the simple minimum distance classifiers used by Pentland's group. By appropriately combining the eigenfaces and eigenfeatures with NN classifiers, it will be possible to improve the performance of Pentland's system. In any case, the usefulness of NN classifiers needs to be evaluated on significantly larger datasets than reported in the literature.

In the final stages of the mug shot matching problem, finer attributes of facial features are usually matched for identification purposes. The point features used in [88] which correspond to points of high curvature, may serve the role of "minutiae" in the fingerprint matching problem. The usefulness of point features and appropriate recognizers and identifiers that use them should be studied and evaluated on large datasets.

Owing to the cost of the equipment and the need for easy maintenance, intensity based systems are preferred in law enforcement applications. Although range information is richer than the 2D intensity array, we feel that cost considerations will make range image based techniques less attractive for field use.

V. Face Recognition from Profiles

Another area for recognition of faces involving intensity images is that of profile images. Research in this area is basically motivated from requirements of law enforcement agencies with their mug shot databases. Profile images provide a detailed structure of the face that is not seen in frontal images. In particular, the size and orientation of the nose is delineated. Face recognition from profiles concentrates on locating points of interest, called fiducial

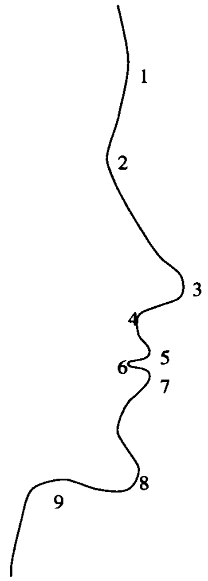


Fig. 12. The nine fiducial points of interest for face recognition using profile images (similar to figure in [61]).

points (Fig. 12). Recognition involves the determination of relationships among these fiducial points.

Kaufman and Breeding [79] developed a face recognition system using profile silhouettes. The image acquired by a black and white TV camera is thresholded to produce a binary, black and white image, the black corresponding to the face region. A preprocessing step then extracts the front portion of the silhouette that bounds the face image. This is to ensure variations in the profile due to changes in hairline. A set of normalized autocorrelations expressed in polar coordinates is used as a feature vector. Normalization and polar representation steps insure invariance to translation and rotation. A distance weighted k-nearest neighbor rule is used for classification. Experiments were performed on a total of 120 profiles of ten persons, half of which were used for training. A set of 12 autocorrelation features was used as a feature vector. Three sets of experiments were done. In the first two, 60 randomly chosen training samples were used, while in the third experiment 90 samples were used in the training set. Experiments with varying dimensionality of the training samples are also reported. The best performance (90% accuracy) was achieved when 90 samples were stored in the training set and the dimensionality of the training feature vector was four. Comparisons with features derived from moment invariants [38] show that the circular autocorrelations performed better.

Harmon and Hunt [61] presented a semi-automatic recognition system for profile-posed face recognition by treating the problem as a "waveform" matching problem. The profile photos of 256 males were manually reduced to outline curves by an artist. From these curves, a set of nine fiducial marks (see Fig. 12) such as nose tip, chin, forehead, bridge, nose bottom, throat, upper lip, mouth and lower lip were automatically identified. The details of how each of these fiducial marks was identified are given in [61]. From these fiducial marks, a set of six feature characteristics were derived. These were protrusion of nose, area right of base line, base angle of profile triangle, wiggle, distances and angles between fiducials. A total of eleven numerical features were extracted from the characteristics mentioned above. After aligning the profiles by using two selected fiducial marks, an Euclidean distance measure was used for measuring the similarity of the feature vectors derived from the outline profiles. A ranking of most similar faces was obtained by ordering the Euclidean norms. In subsequent work, Harmon *et al.* [63] added images of female subjects and experimented with the same feature vector. By noting that the values of the features of a face do not change very much in different images and that the faces corresponding to feature vectors with a large Euclidean distance between them will be different, a partitioning step is included to improve computational efficiency.

[63] used the feature extraction methods developed in [61] to create 11 feature vector components. The 11 features were reduced to 10, because nose protrusion is highly correlated with two other features. The 10D feature vector was found to provide a high rate of recognition. Classification was done based on both Euclidean distances and set partitioning. Set partitioning was used to reduce the number of candidates for inclusion in the Euclidean distance measure and thus increase performance and diminish computation time. Reference [62] is a continuation of the research done in [61] and [63]. The aim is basic understanding of how to achieve automatic identification of human face profiles, to develop robust and economical procedures to use in real-time systems, and to provide the technological framework for further research. The work defines 17 fiducial points which appear to the best combination for face recognition. The method uses the minimum Euclidean distance between the unknown and the reference file to determine the correct identification of a profile, and uses thresholding windows for population reduction during the search for the reference file. The thresholding window size is based on the average vector obtained from multiple samples of an individual's profile. In [62], the profiles are obtained from high contrast photography from which transparencies are made, scanned, and digitized. The test set consists of profiles of the same individuals taken at a different setting. The resulting 96% rate of correctness occurs both with and without population reduction [62].

Wu and Huang [142] also report a profile-based recognition system using an approach similar to that of Harmon and his group [61], but significantly different in detail. First of all, the profile outlines are obtained automatically. B-splines are used to extract six interest points. These are the nose peak, nose bottom, mouth point, chin point, forehead point, and eye point. A feature vector of dimension 24 is constructed by computing distances between two neighboring points, length, angle between curvature segments joining two adjacent points, etc. Recognition is done by comparing the feature vector extracted from the test image with stored vectors using a sequential search method and an absolute norm. The stored features are obtained from three instances of a person's profile; in all, 18 persons were used for the training phase. The testing dataset was generated from the same set of persons used in training, but from different images. In the first attempt 17 of the 18 test images were correctly recognized. The face image corresponding to the failed case was relearned (by including the failed image feature vector in the training set). Another instance of this person was then correctly recognized using the expanded dataset.

Traditional approaches such as Fourier descriptors (FD) have been used [146] for recognizing closed contours. Using p-type FD's that can describe open as well as closed curves, Aibara *et al.* [9] describe a technique for profile-based face recognition. The p-type FD's are derived by discrete Fourier transforming normalized line segments from profiles, are invariant to parallel translation or enlargement/reduction, and satisfy a simple relation between the original and rotated curves. The training set was generated from three sittings of 90 persons (all males) with the fourth sitting used as the testing data. The p-type FD's (ten coefficients) obtained from three sittings were averaged and used as prototypes. Using four coefficients, 65 persons were recognized perfectly. For the full set of 90 test samples, close to 98% accuracy was obtained using ten coefficients.

A. Summary

Profile based recognition has not been pursed with as much vigor as frontal face recognition. Given that mug shots have at least one side view, one could pursue a combination of the eigenapproach for side views of the face, as done in Pentland *et al.* [104] and a similar approach for the profiles. The p-type FD's, used in [9] for profiles, belong to a class of methods similar to eigenanalysis of waveforms. The profile-based approaches, reported in the literature, have not been tested extensively on large datasets. An evaluation of the eigenapproach for sideview and profiles deserves serious investigation on large mug shots datasets.

VI. Face Recognition from an Image Sequence

In surveillance applications, face recognition and identification from a video sequence is an important problem. Although over 20 years of active research on image sequence analysis is available [1], [2], [4], [71], [134], very little of that research has been applied to the face recognition problem. We have identified at least four important areas relevant to FRT where techniques from image sequence analysis are useful:

1) Segmentation of moving objects (humans) from a video sequence
2) Structure estimation
3) 3D models for faces
4) Nonrigid motion analysis.

Current attempts [117], [133] at segmenting moving faces from an image sequence have used pixel based, simple change detection procedures based on difference images. These techniques may run into difficulties when multiple moving objects and occlusion are present. Flow field based methods for segmenting humans in motion is reported in [121]. If there are situations where both camera and objects are moving, more sophisticated segmentation procedures may be required.

There is a large body of existing literature in image sequence analysis on segmenting/detecting moving objects from a stationary or moving platform. Methods based on analysis of difference images, discontinuities in flow fields using clustering, line processes or Markov random field models are available. Some of these techniques have been extended to the case when both the camera and objects are moving.

The problem of structure from motion is to estimate 3D depth of points on objects from a sequence of images. Since in most cases involving surveillance applications, it is nearly impossible to move the camera along a known baseline, thus techniques such as motion stereo may not be useful. The structure from motion problem has been approached in two ways. In the differential method, one computes some form of a flow field (optical, image or normal) and from the computed flow field estimates structure or depth of visible points on the object. The bottleneck in this approach is the reliable computation of the flow field. In the discrete approach, a set of features such as points, edges, corners, lines and contours are tracked over a sequence of frames, and the structure of these features is computed. The bottleneck here is the correspondence problem—the task of extracting and matching features over a sequence of frames. It should be pointed out that in both differential and discrete approaches, the parameters that characterize the motion of the object jointly appear along with the structure parameters of the object. In FRT, the motion parameters may be useful in predicting where the object will appear in subsequent frames, making the segmentation task somewhat easier. The usefulness of structure information is in building 3D models for faces and possibly using the models for face recognition in the presence of occlusion, disguises and face reconstruction. It should be pointed out that if only a monocular image sequence is available, the depth information is available only up to a scaling constant; if binocular image sequences are available, one can get absolute depth values using stereo triangulation. Given that laser range finders may not be practical, for surveillance applications, binocular image sequences may be the best way to get depth information. Another point worth observing is that when discrete approaches are used, the depth values are available only at sparse points requiring interpolation techniques; when flow based methods are used, dense depth maps can be constructed.

Over the last 20 years, hundreds of papers dealing with structure from motion have appeared in the literature. It is beyond the scope of this paper to even include a brief summary of major techniques. We simply list books [68],

[71], [92], [134], [138], [147] and review papers [8], [91], [95]–[97] here; papers that describe major approaches are listed in the additional bibliography.

3D models of faces have been employed [10], [23], [87] in the model based compression literature by several research groups. Such models are very useful for applications such as witness face reconstruction, reconstruction of faces from partial information and computerized aging. 3D models of faces could also be useful for face recognition in the presence of disguises.

Another area of relevance to FRT is the motion analysis of nonrigid objects [72], [93], [103], [130]. There is emerging work in image sequence analysis dealing with nonrigid objects with emphasis on medical applications. Some of the ongoing work on nonrigid motion analysis will be useful in face recognition. An application of nonrigid motion to face recognition is reported in Yacoob and Davis's [143] approach for recognizing facial expressions and actions from image sequences. Their work focuses on six universal emotion expressions (i.e., anger, disgust, fear, happiness, sadness, and surprise), and detection of eye blinking. The approach consists of: spatial tracking of face features, optical flow computation of these features, and psychologically motivated analysis of these spatio-temporal results. The system has been successfully employed to classify the expressions of 30 subjects with a total of 120 instances of the above six emotions.

In sum, we feel that segmentation of moving persons from a video is the most important area in image sequence analysis with direct applications to face recognition. Structure from motion, 3D modeling of faces and nonrigid motion analysis potentially offer new solutions to various aspects of the face recognition and reconstruction problem. In the next subsection, we will briefly summarize the relevant literature on the problem of segmenting moving objects from an image sequence.

A. Segmentation

In [17], a template based approach is used to locate and track a moving head in a video sequence. This approach derives its motivation from the view based approach of [104]. It utilizes the minimum number of different templates of a face determined by analysis of geometrical transformations of a face over parameters such as translation, large rotation, scaling etc. The minimum set is called the face set topology. Restrictions on the transformation parameters lead to monotone regions within which variations of the parameter values monotonically ascend to the exact match. A coarse to fine approach is used to zero in on the most likely match viz-a-viz the parameter in question. On the coarsest scale a rough estimate of the probable parameter values is determined through a correlation match. This rough estimate is fine tuned in successive finer scales by changing the acceptance threshold. They authors describe a system which locates and then tracks a moving head in a video sequence using the above approach. It was used to test both the translation as well as the rotational accuracy of the algorithm. In the case of translation the authors reported error with standard deviation in horizontal direction of 4.1 pixels and in the vertical direction of 2.0 pixels. For the orientational accuracy on a test set of 189 images with nine views of 21 people evenly spaced from −90° to 90° along the horizontal plane, they reported an error with the average standard deviation of 15°.

As noted earlier, thresholding of frame differences is one of the simplest methods for detecting moving objects. Several of the earlier papers [75]–[77], [96] have analyzed difference images to draw inferences as to whether an object is approaching, receding, translating, etc. An example of face location from a video sequence using simple change detection algorithms based on difference images is shown in Fig. 13. Analysis of difference images becomes complicated when occlusion and illumination changes are present or when the camera is moving. More sophisticated techniques for the segmentation of moving objects rely on analyzing optical flow field or its variations. Optical flow is the distribution of apparent velocities of brightness patterns in an image [70] and may arise from relative motion of objects and the viewer. Analysis of optical flow field is useful for estimating the egomotion of the observer, segmentation/detection of moving objects, image stabilization and estimation of depth for scene reconstruction and collision avoidance. Although computation of optical flow has been studied in the image sequence coding literature for nearly 25 years, it has received significant attention in the computer vision literature only over the last 15 years. Since computation of the optical flow field involves two unknowns (velocity components along the x and y directions) but only one measurement (intensity) at each pixel, additional constraints such as smoothness of flow are enforced to find a solution to what is essentially an ill-posed problem. Such ill-posed problems are handled using the regularization approach [110], [132]. Horn and Schunck [70] developed an iterative method for computing optical flow field using the regularization approach. Subsequently Anandan [13] has presented hierarchical approaches to the computation of the optical flow field. The literature on computation of optical flow is quite extensive. We refer the reader to other significant papers by Enklemann [42], Glazer [48], Heeger [66], Hildreth [68], and Fleet and Jepson [44]. Accurate computation of optical flow is still an unsolved problem leading researchers into computation of other flow fields such as image flow [122], [129], [136] and normal flow [12]. Detailed discussion of the relative merits of the computation and interpretation of different types of flow field patterns is beyond the scope of this report. A systematic evaluation of methods that compute optical flow may be found in [15].

Given that an accurate flow field is available, several techniques are available for detecting motion boundaries or clustering of flow fields. Adiv [7] first partitions the computed flow field into connected segments of flow vectors, where each segment is generated by rigid motion of a planar surface. Subsequently, segments coming from a rigid object are grouped. Grouping is done by using the motion coherency of the planar surfaces. From the grouped

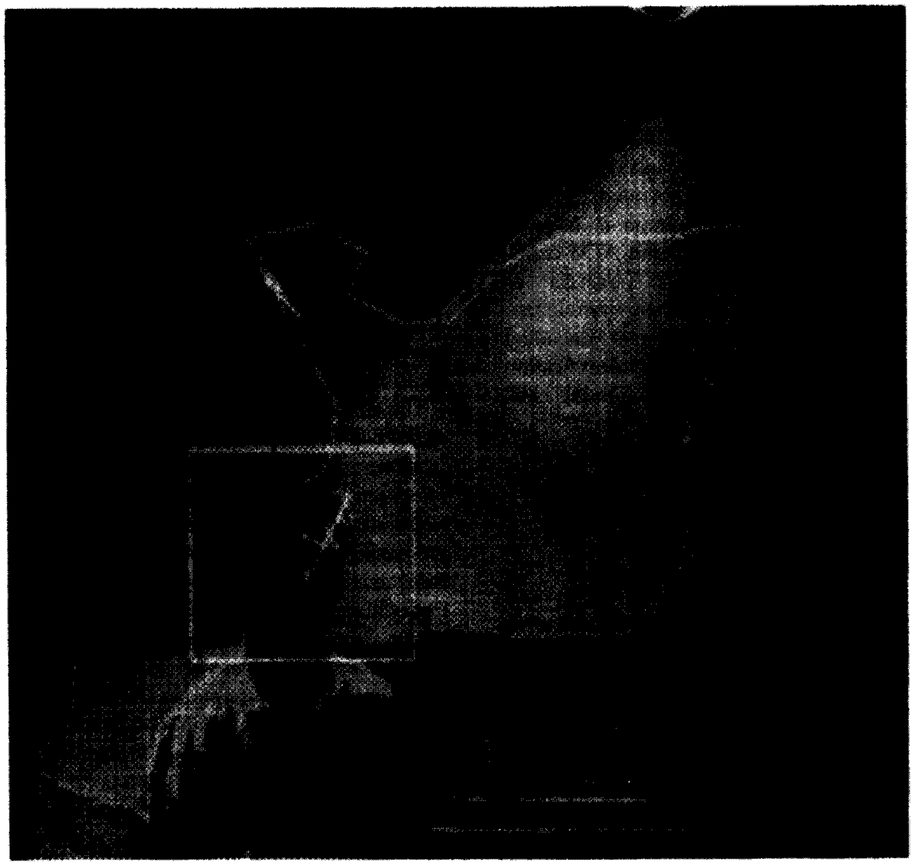


Fig. 13. Locating the head from a video sequence applying the method of Pentland *et al.* [133].

flow fields, motion and structure parameters are computed under the assumptions the field of view (FOV) is narrow and the images of moving objects are small with respect to the FOV and that the optical flow field is computed from monocular, noisy imagery. Thompson and Pong [131] present algorithms for the detection of moving objects from a moving platform. Under various assumptions about the camera motion (the complete camera motion is known, only the rotation or translation is known, etc.), several versions of motion detection algorithms are presented with examples drawn from indoor scenes.

Analysis of optical flow for detecting motion boundaries and subsequently for motion detection requires the availability of accurate estimates of optical flow. But to obtain these accurate estimates, we need to account for or model the motion discontinuities in the flow field due to the presence of moving objects. Simultaneous computation of optical flow and modeling of discontinuities has been addressed by several research groups [67], [73], [94]. The central theme of this approach is to model the discontinuities using the "line processes" of Geman and Geman [46], pose the computation of optical flow in a Bayesian framework, and derive iterative techniques from the application of an optimization procedure such as simulated annealing [46], maximum posterior marginal [90], and iterated conditional mode [16]. Implementation of these algorithms using analog VLSI hardware is addressed in [73], [83]. A recent paper [83] presents a multiscale approach with supporting physiological theory for the computation of optical flow. Other significant papers that deal with segmentation, motion detection from optical flow and normal flow may be found in [26], [36], [102], [115], [126], [135], [141], [148].

Two algorithms that use models of human motion for the purpose of segmentation are described in [101], [121]. An algorithm for the detection of moving persons from normal flows is also described in [99].

VII. Applications

Current applications of FRT include both commercial and law enforcement agencies. Although we have not been able to find many publications detailing commercial applications nevertheless a brief description of possible application areas is given. For the case of law enforcement agencies the emphasis has been on face identification, witness recall, storage and retrieval of mug shots and user interactive systems. Papers describing these applications are summarized.

A. Commercial Applications

Commercial application of face recognition technology may soon become more important economically than law enforcement applications. This has the potential for drastically reducing the cost of face recognition in law enforcement. Bank cards, both credit cards and ATM cards, need a better means of user identification. Insurance for ATM transactions costs about 1.5 cents per transaction; this amounts to several billion dollars per year. Encoding each card with facial features could provide an identification method which would be very effective in improving bank card fraud statistics.

The advantages of facial identification over alternative methods, such as fingerprint identification, are based primarily on users convenience and cost. Facial recognition can be corrected in uncertain cases by humans without extensive training. The development of multi-media computing promises to make low cost video input devices available for PC's. This should allow a facial image to be acquired by any PC based cash register systems. For systems of this kind to be widely accepted, standard PC compatible methods for facial recognition must be available. If a low cost method for acquiring and encoding facial images is developed, then this technology can also be applied to provide a low cost booking station technology.

B. Law Enforcement

The basic approach to the mug shots problem is for the system to compare features from the target with those stored in the database. The nature of the target image relative to the images in the database is crucial and determines the difficulty of the overall procedure. The target may be a mug-shot or from another photographic source and may need to be rotated before the features can be extracted and compared to the mug-file images. In [80] ten feature distances are measured to code nine features with each feature being a distance divided by the nose length. It was found that 92% of the variation in the normalized data could be explained by five components. Similarity measures are used in sequencing algorithms with geometric coding of features. The objective of this approach is to sequence the photographs in the mug shots album on the basis of similarity to the target. The algorithm's design must consider the precision and accuracy of the measurements and develop appropriate scales for the components of the feature vector. Algorithm design is concerned with the sequence in which selections are made and how to minimize errors in the face description. The matching algorithm uses a window for each feature and selects all the images that fall within the window. A larger window permits more of the database's population to fit, while a smaller window increases the probability that an error in coding a feature will cause the correct image to be missed during the search. Harmon *et al.* [62], using geometric feature codes of images, present a matching algorithm to eliminate the mismatches, and a sequencing algorithm on the remaining subset achieved "nearly perfect" identification for a population of over 100 faces.

In [86] Laughery *et al.* deal with the storage and retrieval of mug shots. Three distinctions are made in the prototype development efforts, the nature of the target image that serves as input to the search, the method of coding the faces, and the pose position of the faces in the mug shots album. Three prototype systems characterized by varying methods of interaction and input are reviewed in this paper. Laughery *et al.* also review the different interactive and automatic measurement algorithms for locating and measuring facial features.

The initial forensic evidence is often a witness' recall of the culprit's appearance. Verbal descriptions of people's faces very often lack detail. In [120] two factors that might affect retrieval—distinctiveness of target face and position in the album—were independently varied. Distinctive faces are easier to remember than nondistinctive faces. Target faces occurring later in the album are believed to be more difficult to detect than those encountered earlier in the search. The faces were rated on a set of five-point descriptive scales. The scales were derived from the analysis of free descriptions of a different set of faces. The physical measurements corresponding to the features which had been rated were taken from the faces, and these were converted to values on five-point scales using linear regression techniques. The complete record for each face comprises 47 face parameters plus the age. Thirty-eight of the parameters are five-point scale parameters (breadth of face, length of hair, eye color), and nine are dichotomous parameters which code for the presence or absence of facial hair, peculiarities, and accessories. Age is coded on five-point scale. The database consists of 1000 photographs of males between 18–70 taken under controlled conditions. Three photographs were taken, frontal view, profile, and $\frac{3}{4}$ view. Four nondistinctive and four distinctive faces were chosen from the set. Eight paid subjects were shown one of the $\frac{3}{4}$ view test photographs for 10 s, provided a detailed description of the photograph, and were assigned randomly to either the computer or album search group. There were four albums in which there were four photographs per page and 250 pages. Each photograph appeared four times within an album. The computer search was performed using the subjects' descriptions and ratings and could be changed and repeated up to four times. The computer search retrieval rate for the distinctive faces was 75% and for the nondistinctive faces 69%. The rate for the album searches was 78% for the distinctive faces and 44% for the nondistinctive faces [120].

In [59] Haig presents a system that can insert new targets into storage, control and change the intensities both locally and globally, move the targets around, change their size and orientation, present them for a wide range of fixed time intervals, run experiments automatically, collect data, and analyze the results. Haig's database consists of 100 target faces, taken under reasonably standardized conditions from the direct frontal aspect. The pictures are registered such that the inter-pupilary distance is 30 pixels. The goal of face distortion experiments is to measure the sensitivity of adult observers to slight positional changes of prominent facial features. Each target face is subjected to the same operations, in which certain features are moved by defined amounts. Greatest sensitivity is to the movement of the mouth upward by 1.2 pixels, close to the visual acuity limit. In other experiments, features are interchanged among four different faces. The head outline is the major focus of attention when four features are interchanged. A changed head outline, while maintaining the inner features, very strongly influences the observer's responses. Fixing both the head and the eyes shows a dominance in the mouth over the nose. These experiments establish a clear hierarchy of feature saliency in which the head's outline plays a major role. In the distributed aperture experiments, Haig attempts to find what constitutes a facial recognition feature. The technique used implies that all parts of the face are equally likely to be masked or unmasked in any combination. The program selects one of the four target faces at random and presents the target to a random number of apertures with their actual addresses selected at random from the 38 possible addresses. An analysis of results revealed a very high proportion of correct responses across the eyes-eyebrows and across the hairline at the forehead. Few correct responses may be seen around the side of the temples and at the mouth, and the lower chin area is clearly not a strong recognition feature.

Starkey [127] presents an overview of work done using 96 police photographs. The images of the faces are normalized by fixing the pupil distance at 80 pixels. A target face is chosen at random and a neural network is trained to recognize it. The correct face is identified from among the 96 photos 100% of the time. With the addition of noise levels of 5% and 10% to the target image, the correct face is found 62.5% and 36.5% of the time. In an experiment using 100 faces, 43 noses are used to train the neural network to find features. The net is able to find the nose feature within 2 pixels using a Euclidean metric. In profile analysis, a set of 36 profiles is prepared using Fourier descriptors. Cluster analysis is used to group them by similarity. The descriptors from a profile can be displayed with other data such as height, age, sex, etc. in the form of a histogram or bar-code and may be used to increase search accuracy.

VIII. Evaluation of a Face Recognition System

In addition to the material contained in this paper, previous work on the evaluation of OCR and fingerprint classification systems [27], [58] provides insights into how the evaluation of recognition algorithms and systems is most efficiently performed. One of the most important facts learned in previous evaluations is that large sets of test images are essential for adequate evaluation. It is also extremely important that the sample be statistically as similar as possible to the images that arise in the application being considered. Scoring should be done in a way which reflects the costs or other system requirement changes that result from errors in recognition. System reject-error behavior should be studied, not just forced recognition.

In planning these evaluations, it is also important to keep in mind that pattern recognition is not governed by a formal theory, as physics or mathematics is, which allows clearly applicable governing principles to determine the extent to which results derived form one set of applications can be applied to other applications. The operation of these systems is statistical, with measurable distributions of success and failure. The specific values in these distributions are very application-dependent and no existing theory seems to allow their prediction for a new application. This strongly suggests that the most useful form of evaluation is one based as closely as possible on a specific application.

A. Evaluation Requirements

1) Accuracy Requirements: The face recognition applications to be evaluated here will usually take the form of a computer search of a large set of face images which generates a list of possible match candidates which are evaluated by the users of the system. In this kind of application accuracy requirements for the face recognition system are bounded by two factors: 1) the acceptable probability of missed recognition (the computer never finds the right face) and 2) the ability of humans to distinguish similar faces from a candidate list generated by the recognition system without unacceptable confusion or fatigue (the human can't find the face in the set generated by the computer). This tradeoff is substantially different from the trade-off inherent in the reject-error characteristics found in OCR and fingerprint classification. In these two applications, if an image is rejected human correction can always lower the classification error to practical levels since humans can do the job without computer assistance. With face recognition, some faces are known to look alike and when humans are presented with large sets of face images their ability to distinguish similar faces drops due to confusion and fatigue. As the candidate list of similar faces produced by the recognition system increases, the probability that a matching face exits to the candidate face increases. At the same time, as the candidate list increases, the probability of confusion in human selection causes the probability of selecting the matching face to decrease. At some point the highest probability of a correct match will be achieved. This will not be a serious problem if the face of interest is in the first few faces; the probability of finding the correct match then increases sharply as candidate faces are added. If these lists contain hundreds of faces to obtain adequate probability of selection, then the limiting performance factor will be human ability to select faces from the candidate list.

2) Constraints on Samples: In both OCR and fingerprint classification work, use of images with atypical image quality or overly simplified segmentation characteristics has led to misleading conclusions about system requirements. In OCR two factors were found to be very important. First, algorithms should be tested using images from sources of comparable image complexity to the target application.

Second, since many early OCR studies were done on isolated characters or characters with moderate segmentation problems, the need for robust generalization was underestimated. This caused too much effort to be expended on systems that did not address the types of recognition problems that arise in real Based on this experience, we feel that initial studies using realistic images from some specific commercial/law enforcement application should be carried out. An initial image set based on mug shots seems appropriate since these images span a wide range of the possible applications shown in Table 1, will have realistic image segmentation problems, and have realistic image quality parameters. This should provide commercial/law enforcement agencies such as credit card companies or the FBI with a more realistic estimate of the utility of FRT than studies done on idealized datasets or datasets which are unrelated to specific applications.

3) Speed and Hardware Requirements: We recommend that where possible all algorithms under test be evaluated on several types of parallel computer hardware as well as standard engineering workstations. In high volume applications speed will be an important factor in evaluating applicability. In many potential applications, parallel computers may be too costly but developments of effective high speed methods on parallel computers should allow special purpose hardware to be developed to reduce costs.

4) Human Interface: The utility of face recognition systems will be strongly affected by the type of human interface that is used in conjunction with this technology. The human factors which will affect this interface are dealt with in Section III. The literature on human perception and recognition of faces will be important in designing human interfaces which allow users to make efficient use of the results of machine-based face recognition.

B. Evaluation Methods

1) Database Size and Uniformity: For law enforcement applications, as an initial evaluation sample, a collection of a minimum of 5000 and a maximum of 50 000 mug shot images may be appropriate. A testing sample containing 500 to 5000 different mug shots of faces in the original training set and 500 to 5000 different mug shots of faces not in the original training set should be collected to allow testing of machine face matching. Similar samples for commercial applications are also suggested.[1] The minimum sample sizes for the test sets is based on the need to obtain accurate matching and false matching statistics. The 10:1 ratio of the evaluation set size to the testing set size is designed to minimize false match statistics due to random matches and provide statistical accuracy in probability of match versus candidate list size statistics.

2) Sample Size Issues—Feasibility of Resampling: We suggest that images be collected at relatively high resolution, 512 × 512, and using 8 b of gray or intensity. If color images are used, matching will initially use only intensity data. With this level of image resolution, down-sampling of the images and digital filtering to provide lower resolution and image quality can be done with a single set of master images. Images can also be cropped after segmentation to provide more usable image areas containing the face image. Resampling the image to provide a greater area of background and less active image area may also be possible, but may introduce artifacts that change the difficulty of the segmentation problem.

3) Test Methods for Algorithm Accuracy and Probability of Match: The scoring of face matching systems should be based on the probability of matching a candidate face in the first n faces on a candidate list. Two sets of probabilities of this type can be defined, one for faces in the database and one for faces not in the database. The first will generate true positives and the second will generate false positives. The comparison of true and false recognition probabilities assumes that each recognition produces a confidence number which is higher for faces with greater similarity. For each specified level of confidence, the number of faces matching true and false faces can be generated. The simplest accuracy measure of each type of recognition is the cumulative probability of a match for various values of n, and at the same confidence level, the probability of a false match. It seems likely that in addition to the raw cumulative probability curve, some simple models of the shape of the curve, such as a linear model, may be of interest in comparing different algorithms. In many applications it will be as important that the face recognition system avoid false positives as that it produce good true positive candidate lists.

Many of the face recognition systems discussed in this paper reduce the face to a set of features and measure the similarity of faces by comparing the distance between faces in this feature space. For all of the test faces the distance between each test face and all other faces in the database is calculated. The probability, over the entire test sample, and the average confidence of the first n near neighbors is then calculated. A similar calculation is made using faces not in the database and the average confidence of the first n candidates evaluated. At each confidence level for these faces a probability of finding a false match can be calculated as the ratio of false candidates to true candidates at comparable confidence. If the recognition process is to be successful the probability of detection of a face in the database should always exceed the probability of false detection of a face not included in the database.

4) Similarity Measures: The example calculation discussed above requires that the recognition system produce a measure of confidence of recognition and of similarity to other possible recognitions. Similarity differs from confidence in that similarity is measured between any two points in the feature space of the database while confidence is a measure between a test image and a trial match. In the example a reasonable measure might be $1/(1+kd_{ij}^2)$. Using this measure the similarity of two faces is 1.0 if their features are identical and approaches 0.0 as the features are displaced to infinity. This type of similarity measure

[1] NIST has recently made available a mug shot identification database containing a total of 3248 images. For details the reader may contact: craig@magi.ncsl.nist.gov.

only works if all of the feature are normalized to a similar scale; otherwise, a single large feature can control the entire process.

5) Rank Statistic Comparison: The ability to address applications where lists of candidate faces are to be used for human ranking requires that a method for comparison of human ranking of similar faces and machine ranking of faces be developed. This problem can be addressed by treating each ranked list of faces as a symbol string and using the string comparison methods which have been applied to OCR. These lists will contain insertions, deletions, and substitutions just as symbol strings would. The comparison of strings can then be effected using Levenstein distance as a measure of cost of sequence alignment. The problem differs from the OCR problem in that in OCR the use of confidence measures allows unknown symbols to be included in sequences. In the face recognition problem these sequences are completely determined by similarity measures and will only decrease in similarity with sequence length.

6) Summary: All of the evaluation methods discussed here are directed toward the evaluation of machine base similarity metrics for specific applications. The two which should be addressed first are: Can the methods find similar faces in a large database with an acceptable false detection rate? and: Will machine base rank ordering of similarity be sufficiently similar to human ranking to provide useful input for human list correction?

IX. Summary and Conclusions

In this paper, we have presented an extensive survey of human and machine recognition of faces with applications to law enforcement and other commercial sectors. We have focused on segmentation, feature extraction and recognition aspects of the face recognition problem, using information drawn from intensity and range images of faces and profiles. A brief summary of relevant psychophysics and neurosciences literature has also been included.

The survey presented here is relevant to applications 1–7. While many of the methods described here are of interest in applications 8 and 9, a detailed discussion of them is beyond the scope of this paper.

We give below a concise summary followed by conclusions in the same order as the topics appear in the paper.

- Face recognition technology is an essential tool for law enforcement agencies' efforts tc combat crime. Given that crime is seen as the most important problem facing the country, even ahead of job security, health care, and the economy, the use of high technology to effectively fight crime will receive support from the people and from their elected representatives. Furthermore, commercial applications of this technology has received a growing interest, most importantly in the case of credit card companies and their desire to reduce fraudulent usage of credit cards. Face recognition, in addition to fingerprint recognition, will remain a critical high-technology strategic research area with significant potential impact on reducing crime.
- Over 30 years of research in psychophysics and neurosciences on human recognition of faces is documented in the literature. Although we do not feel that machine recognition of faces should strictly follow what is known about human recognition of faces, it is beneficial for engineers designing face recognition systems to be aware of the relevant findings. This is especially crucial for applications where human interaction is involved, such as in expert identification, electronic mug shots books, and lineups. Even for applications 1–3, what is known about human recognition of faces obviously impact feature selection and recognition strategies.
- Segmentation of a face region from a still image or a video is one of the key first steps in face recognition. Surprisingly, this problem has not received much attention. Although the segmentation problem is easier for mug shots, drivers' licenses, personal ID's, and passport pictures, as in applications 2 and 3, segmentation in general is a nontrivial task. More effort needs to be directed in addressing the segmentation problem.
- Both global and local feature descriptions are useful. Popular global descriptions are based on the KL expansion. The local descriptors are derived from regions that contain the eyes, mouth, nose, etc., using approaches such as deformable templates or eigenexpansion of regions surrounding the eyes, mouth, etc. Minutiae-type point features have also been extracted. It appears that feature extraction is better understood and developed in connection with recognition. It may be worthwhile to investigate the use of other possible global and local transforms, and better methods for detection, localization and description of features.
- A multitude of techniques are available in IU literature [1], [3]–[5], [43], [57], [137] for recognition. The eigenapproach of Pentland's group has been tested on a large number of images of 3000 persons. Other promising approaches based on neural networks and graph matching have not been tested on such large datasets. All of these approaches should be tested on the same dataset derived from a practical application. Although a complete algorithm that can solve even the simplest of the applications in Table 1 is not yet available, one can begin the task of evaluating the existing methods on a dataset that truly represents the data available in real applications. Methods that may not scale with the size of the dataset can be identified in this way and discouraged from further development. For commercial/law enforcement applications, the use of range data is not feasible due to the cost of the data acquisition process. Consequently, only intensity-based approaches may be pursued. Profile-based recognition schemes should be evaluated on a large dataset

and methods for integrating profile and face based methods should be developed.

- Face recognition from a video sequence is probably the most challenging problem in face recognition. Up to now, fairly simple thresholding of difference images has been used for locating a moving person's face, and has been followed by a 2D recognition algorithm. In our opinion, recognition from an image sequence offers excellent opportunities for applying several concepts from the IU literature; specifically, the usefulness of flow fields for the segmentation of the face region, and the reconstruction and refinement of 3D structure from 2D frames, must be investigated.
- The most important step in face recognition is the ability to evaluate existing methods and provide new directions on the basis of these evaluations. The images used in the evaluation should be derived from operational situations, similar to those in which the recognition system is expected to be installed. An important subproblem is the definition of a similarity measure that can be used in matching two face images. In witness and electronic mug shots matching problems, a face recognition system is expected to rank the chosen images in the same way that humans do, in terms of how similar the test and stored images are. The similarity measure used in a face recognition system should be designed so that humans' ability to perform face recognition and recall are imitated as closely as possible by the machine. As of this writing, no such evaluation of a face recognition system has been reported in the literature.

ACKNOWLEDGMENT

The authors are grateful to Prof. A. Pentland for providing Figs. 3, 5(a), 7, and 13, Prof. T. Poggio for providing Fig. 8, Prof. C. V. d. Malsburg for providing Figs. 2, 6(a), 9, and 10, Dr. G. Gordon for providing Fig. 11, Prof. B. S. Manjunath for providing the feature extraction and face identification algorithms (results of which are shown in Fig. 6(b) and Fig. 10), and FBI for providing Fig. 1. The authors would like to thank Ms. C. S. Barnes for helpful discussions in the preparation of this manuscript. Finally, the authors are grateful to Prof. A. Rosenfeld for carefully reading the manuscript and suggesting useful improvements.

REFERENCES

[1] Proc. DARPA Image Understanding Workshop, 1984, —.

[2] Proc. IEEE Workshops on Motion, 1986, 1989, 1991.

[3] *Computer Vision, Graphics and Image Process.*, 1972, —.

[4] *IEEE Trans. Patt. Anal. and Mach. Intell.*, 1979, —.

[5] *Int. J. Compu. Vision*, 1988, —.

[6] Y. S. Abu-Mostafa and D. Psaltis, "Optical neural computers," *Scientific American*, vol. 256, pp. 88–95, 1987.

[7] G. Adiv, "Determining three-dimensional motion and structure from optical flow generated by several moving objects," *IEEE Trans. Patt. Anal. and Mach. Intell.*, vol. 7, pp. 525–542, 1985.

[8] J. K. Aggarwal and N. Nandhakumar, "On the computation of motion from sequences of images," *Proc. IEEE*, vol. 76, pp. 917–935, 1988.

[9] T. Aibara, K. Ohue, and Y. Matsuoka, "Human face recognition of P-type Fourier descriptors," in *SPIE Proc., Vol. 1606: Visual Commun. and Image Process.*, 1991, pp. 198–203.

[10] K. Aizawa *et al.*, "Human facial motion analysis and synthesis with application to model-based coding," in *Motion Analysis and Image Sequence Processing*, M. I. Sezan and R. L. Lagendijk, Eds. Boston, MA: Kluwer, 1993, pp. 317–348.

[11] S. Akamatsu, T. Sasaki, H. Fukamachi, and Y. Suenaga, "A robust face identification scheme—KL expansion of an invariant feature space," in *SPIE Proc.: Intell. Robots and Computer Vision X: Algorithms and Techn.*, vol. 1607, 1991, pp. 71–84.

[12] Y. Aloimonos *et al.*, "Behavioral visual motion analysis," in *Proc., DARPA Image Understanding Workshop*, 1992, pp. 521–541.

[13] P. Anandan, "A computational framework and an algorithm for the measurement of visual motion," *International Journal of Computer Vision*, vol. 2, pp. 283–310, 1989.

[14] R. Baron, "Mechanisms of human facial recognition," *Int. J. Man-Machine Studies*, vol. 15, pp. 137–178, 1981.

[15] J. L. Barron, D. J. Fleet, and S. S. Beauchermin, "Performance of optical flow techniques," *Int. J. on Computer Vision*, vol. 12, pp. 43–77, 1994.

[16] J. Besag, "On the statistical analysis of dirty pictures," *J. Royal Statistical Soc.*, vol. 48, pp. 259–302, 1986.

[17] M. Bichsel and A. Pentland, "Human face recognition and the face image set's topology," *Computer Vision, Graphics, and Image Process.: Image Understanding*, vol. 59, pp. 254–261, 1994.

[18] W. W. Bledsoe, "The model method in facial recognition," Panoramic Research Inc., Tech. Rep. PRI:15, Palo Alto, CA, 1964.

[19] V. Bruce, *Recognizing Faces.* London: Erlbaum, 1988.

[20] ——, "Perceiving and recognizing faces," in *Mind and Language*, 1990, pp. 342–364.

[21] R. Brunelli and T. Poggio, "HyperBF networks for gender classification," in *Proc., DARPA Image Understanding Workshop*, 1992, pp. 311–314.

[22] ——, "Face recognition: Features versus templates," *IEEE Trans. Patt. Anal. and Mach. Intell.*, vol. 15, pp. 1042–1052, 1993.

[23] M. Buck and N. Diehl, "Model-based image sequence coding," in *Motion Analysis and Image Sequence Processing*, M. I. Sezan and R. L. Lagendijk, Eds. Boston, MA: Kluwer, 1993, pp. 285–315.

[24] J. Buhmann, M. Lades, and C. v. d. Malsburg, "Size and distortion invariant object recognition by hierarchical graph matching," in *Proc., Int. Joint Conf. on Neural Networks*, pp. 411–416, 1990.

[25] P. J. Burt, "Multiresolution techniques for image representation, analysis, and 'smart' transmission," in *SPIE Proc.: Visual Commun. and Image Process. IV*, vol. 1199, pp. 2–15, 1989.

[26] P. J. Burt, R. Hingorani, and R. J. Kolozunski, "Mechanisms for isolating component patterns in the sequential analysis of multiple motion," in *Proceedings, IEEE Workshop on Visual Motion*, pp. 187–193, 1991.

[27] G. T. Candela and R. Chellappa, "Comparative performance of classification methods for fingerprints," National Institute of Standards and Technology, Gaithersburg, MD, Tech. Rep. NISTIR-5163, 1993.

[28] J. Canny, "A computational approach to edge detection," *IEEE Trans. Patt. Anal. and Mach. Intell.*, vol. 8, pp. 679–689, 1986.

[29] S. Carey, "A case study: Face recognition," in *Explorations in the Biological Language*, E. Walker, Ed. New York: Bradford, 1987, pp. 175–201.

[30] S. Carey, R. Diamond, and B. Woods, "The development of face recognition—A maturational component?" *Develop. Psych.*, vol. 16, pp. 257–269, 1980.

[31] Y. Cheng, K. Liu, J. Yang, and H. Wang, "A robust algebraic method for human face recognition," in *Proc. 11th Int. Conf. on Patt. Recog.*, 1992, pp. 221–224.

[32] Y. Cheng, K. Liu, J. Yang, Y. Zhuang, and N. Gu, "Human face recognition method based on the statistical model of small sample size," in *SPIE Proc.: Intell. Robots and Compu. Vision X: Algorithms and Techn.*, vol. 1607, 1991, pp. 85–95.

[33] M. J. Conlin, "A ruled based high level vision system," in *SPIE Proc.: Intell. Robots and Compu. Vision*, vol. 726, 1986, pp. 314–320.

[34] I. Craw, H. Ellis, and J. Lishman, "Automatic extraction of face features," *Patt. Recog. Lett.*, vol. 5, pp. 183–187, 1987.

[35] I. Craw, D. Tock, and A. Bennett, "Finding face features," in

Proc. 2nd Europe. Conf. on Compu. Vision, 1992, pp. 92–96.

[36] T. Darrell and A. Pentland, "Robust estimation of a multilayered motion representation," in *Proc. IEEE Conf. on Visual Motion,* 1991, pp. 173–178.

[37] G. Davies, H. Ellis, and E. J. Shepherd, *Perceiving and Remembering Faces.* New York: Academic, 1981.

[38] S. Dudani, K. Breeding, and R. McGhee, "Aircraft identification by moment invariants," *IEEE Trans. Computers,* vol. 26, pp. 39–45, 1977.

[39] H. Ellis, M. Jeeves, F. Newcombe, and A. Young, *Aspects of Face Processing.* Dordrecht: Nijhoff, 1986.

[40] H. D. Ellis, "The role of the right hemisphere in face perception," in *Function of the Right Cerebral Hemisphere,* A. W. Young, Ed. London: Academic, 1983, pp. 33–64.

[41] ——, "Introduction to aspects of face processing: Ten questions in need of answers," in *Aspects of Face Processing,* H. Ellis, M. Jeeves, F. Newcombe, and A. Young, Eds. Dordrecht: Nijhoff, 1986, pp. 3–13.

[42] W. Enklemann, "Investigations of multigrid algorithms for the estimation of optical flow fields in image sequences," *Computer Vision, Graphics and Image Process.,* vol. 43, pp. 150–177, 1988.

[43] O. Faugeras, *Three-Dimensional Computer Vision, A Geometric Viewpoint.* Cambridge, MA: MIT Press, 1990.

[44] D. J. Fleet and A. D. Jepson, "Stability of phase information," *IEEE Trans. Patt. Anal. and Mach. Intell.,* vol. 15, pp. 1253–1269, 1993.

[45] K. Fukunaga, *Statistical Pattern Recognition.* New York: Academic, 1989.

[46] S. Geman and D. Geman, "Stochastic relaxation, Gibbs distribution and Bayesian restoration of images," *IEEE Trans. Patt. Anal. and Mach. Intell.,* vol. 6, pp. 721–741, 1984.

[47] A. P. Ginsburg, "Visual information processing based on spatial filters constrained by biological data," AMRL Tech. Rep., pp. 78–129, 1978.

[48] F. Glazer, "Hierarchical motion detection," Ph.D. dissertation, Univ. Massachusetts, Amherst, MA, 1987.

[49] A. G. Goldstein, "Facial feature variation: Anthropometric data II," *Bull. Psychonomic Soc.,* vol. 13, pp. 191–193, 1979.

[50] ——, "Race-related variation of facial features: Anthropometric Data I," *Bull. Psychonomic Soc.,* vol. 13, pp. 187–190, 1979.

[51] B. A. Golomb and T. J. Sejnowski, "SEXNET: A neural network identifies sex from human faces," in *Advances in Neural Information Processing Systems 3,* D. S. Touretzky and R. Lipmann, Eds. San Mateo, CA: Morgan Kaufmann, 1991, pp. 572–577.

[52] G. Gordon, "Face recognition based on depth maps and surface curvature," in *SPIE Proc.: Geometric Methods in Computer Vision,* vol. 1570, 1991, pp. 234–247.

[53] G. G. Gordon and L. Vincent, "Application of morphology to feature extraction for face recognition," in *SPIE Proc.: Nonlinear Image Process.,* vol. 1658, 1992.

[54] A. Goshtasby, "Description and discrimination of planar shapes using shape matrices," *IEEE Trans. Patt. Anal. and Mach. Intell.,* vol. 7, pp. 738–743, 1985.

[55] V. Govindaraju, S. N. Srihari, and D. B. Sher, "A computational model for face location," in *Proc. 3rd Int. Conf. on Computer Vision,* 1990, pp. 718–721.

[56] U. Grenander, Y. Chow, and D. Keenan, *Hands: A Pattern Theoretic Study of Biological Shapes.* New York: Springer-Verlag, 1991.

[57] W. E. L. Grimson, *Object Recognition by Computer: The Role of Geometric Constraints.* Cambridge, MA: MIT Press, 1990.

[58] P. J. Grother and G. T. Candela, "Comparison of handprinted digit classifiers," Tech. Rep. NISTIR, National Institute of Standards and Technology, Gaithersburg, MD, 1993.

[59] N. D. Haig, "Investigating face recognition with an image processing computer," in *Aspects of Face Processing,* H. D. Ellis, M. Jeeves, F. Newcombe, and A. Young, Eds. Dordrecht: Nijhoff, 1985, pp. 410–425.

[60] P. W. Hallinan, "Recognizing human eyes," in *SPIE Proc.: Geometric Methods in Compu. Vision,* vol. 1570, 1991, pp. 214–226.

[61] L. Harmon and W. Hunt, "Automatic recognition of human face profiles," *Computer Graphic and Image Process.,* vol. 6, pp. 135–156, 1977.

[62] L. Harmon, M. Khan, R. Lasch, and P. Ramig, "Machine identification of human faces," *Patt. Recog.,* vol. 13, pp. 97–110, 1981.

[63] L. Harmon, S. Kuo, P. Ramig, and U. Raudkivi, "Identification of human face profiles by computer," *Patt. Recog.,* vol. 10, pp. 301–312, 1978.

[64] L. D. Harmon, "The recognition of faces," *Scientific American,* vol. 229, pp. 71–82, 1973.

[65] D. C. Hay and A. W. Young, "The human face," in *Normality and Pathology in Cognitive Function,* A. W. Ellis, Ed. London: Academic, 1982, pp. 173–202.

[66] D. Heeger, "Optical flow from spatio-temporal filters," *Int. J. Computer Vision,* vol. 1, pp. 279–302, 1988.

[67] F. Heitz and P. Bouthemy, "Multimodal motion estimation and segmentation using Markov random fields," in *Proc. Int. Conf. on Patt. Recog.,* 1990, pp. 378–383.

[68] E. C. Hildreth, *The Measurement of Visual Motion.* Cambridge, MA: MIT Press, 1984.

[69] Z. Hong, "Algebraic feature extraction of image for recognition," *Patt. Recog.,* vol. 24, pp. 211–219, 1991.

[70] B. K. P. Horn and B. G. Schunck, "Determining optical flow," *Artif. Intell.,* vol. 17, pp. 185–203, 1981.

[71] T. S. Huang, *Image Sequence Analysis.* New York: Springer-Verlag, 1981.

[72] ——, "Modeling, analysis, and visualization of nonrigid object motion," in *Proc. Int. Conf. Patt. Recog.,* pp. 361–364, 1990.

[73] J. Hutchinson, C. Koch, J. Luo, and C. Mead, "Computing motion using analog and binary resistive networks," *Computer,* vol. 21, pp. 52–63, 1988.

[74] A. K. Jain, *Fundamentals of Digital Image Processing.* Englewood Cliffs, NJ: Prentice Hall, 1989.

[75] R. Jain, "Extraction of motion information from peripheral processes," *IEEE Trans. Patt. Anal. and Mach. Intell.,* vol. 3, pp. 489–503, 1981.

[76] R. Jain, W. N. Martin, and J. K. Aggarwal, "Segmentation through the detection of changes due to motion," *Computer Graphics and Image Process.,* vol. 11, pp. 13–34, 1979.

[77] R. Jain and H. H. Nagel, "On the analysis of accumulative difference pictures from image sequence of real world scenes," *IEEE Trans. Patt. Anal. and Mach. Intell.,* vol. 1, pp. 206–214, 1979.

[78] T. Kanade, *Computer Recognition of Human Faces.* Basel and Stuttgart: Birkhauser, 1977.

[79] G. J. Kaufman, Jr., and K. J. Breeding, "The automatic recognition of human faces from profile silhouettes," *IEEE Trans. Syst., Man, and Cybern.,* vol. SMC-6, pp. 113–121, 1976.

[80] Y. Kaya and K. Kobayashi, "A basic study on human face recognition," in *Frontiers of Pattern Recognition,* S. Watanabe, Ed. New York: Academic, 1972, pp. 265–289.

[81] M. D. Kelly, "Visual identification of people by computer," Tech. Rep. AI-130, Stanford AI Proj., Stanford, CA, 1970.

[82] M. Kirby and L. Sirovich, "Application of the Karhunen-Loeve procedure for the characterization of human faces," *IEEE Trans. Patt. Anal. and Mach. Intell.,* vol. 12, pp. 103–108, 1990.

[83] C. Koch, H. T. Wang, R. Battiti, B. Mathur, and C. Ziomkowski, "An adaptive multiscale approach for estimating optical flow: Computational theory and physiological implementation," in *Proc. IEEE Workshop on Visual Motion,* 1991, pp. 111–122.

[84] T. Kohonen, *Self-Organization and Associative Memory.* Berlin: Springer, 1988.

[85] M. Lades, J. Vorbruggen, J. Buhmann, J. Lange, C. v. d. Malsburg, and R. Wurtz, "Distortion invariant object recognition in the dynamic link architecture," *IEEE Trans. Computers,* vol. 42, pp. 300–311, 1993.

[86] K. R. Laughery, J. B. T. Rhodes, and J. G. W. Batten, "Computer-guided recognition and retrieval of facial images," in *Perceiving and Remembering Faces,* G. Davis, H. Ellis, and J. Shepherd, Eds. New York: Academic, 1981, pp. 250–269.

[87] H. Li, P. Roivainen, and R. Forchheimer, "3-D motion estimation in model-based facial image coding," *IEEE Trans. Patt. Anal. and Mach. Intell.,* vol. 15, pp. 545–555, 1993.

[88] B. S. Manjunath, R. Chellappa, and C. v. d. Malsburg, "A feature based approach to face recognition," in *Proc. IEEE Computer Soc. Conf. on Computer Vision and Patt. Recog.,* 1992, pp. 373–378.

[89] D. Marr, *Vision.* San Francisco, CA: Freeman, 1982.

[90] J. Marroquin, S. Mitter, and T. Poggio, "Probabilistic solutions for ill-posed problems in computational vision," *J. Amer. Statistical Assoc.,* vol. 82, pp. 76–89, 1987.

[91] W. N. Martin and J. K. Aggarwal, "Dynamic scene analysis: A survey," *Computer Vision, Graphics and Image Process.,* vol. 7, pp. 356–374, 1978.

[92] ——, *Motion Understanding, Robot and Human Vision.* Boston, MA: Kluwer, 1988.

[93] D. Metaxas and D. Terzopoulus, "Recursive estimation of shape and nonrigid motion," in *Proc. IEEE Workshop on Visual Motion,* 1991, pp. 296–311.

[94] D. W. Murray and B. F. Buxton, "Scene segmentation from visual motion using global optimization," *IEEE Trans. Patt. Anal. and Mach. Intell.,* vol. 9, pp. 220–228, 1987.

[95] H. H. Nagel, "Analysis techniques for image sequences," in *Proc. Int. Conf. on Patt. Recog.,* 1978, pp. 186–211.

[96] ——, "Overview on image sequence analysis," in *Image Sequence Analysis,* T. S. Huang, Ed. New York: Springer-Verlag, 1981, pp. 19–228.

[97] ——, "Image sequences—ten (octal) years—from phenomenology toward a theoretical foundation," in *Proc. Int. Conf. on Patt. Recog.,* 1986, pp. 1174–1185.

[98] O. Nakamura, S. Mathur, and T. Minami, "Identification of human faces based on isodensity maps," *Patt. Recog.,* vol. 24, pp. 263–272, 1991.

[99] R. Nelson, "Qualitative detection of motion by a moving observer," in *Proc. DARPA Image Understanding Workshop,* 1990, pp. 329–338.

[100] M. Nixon, "Eye spacing measurement for facial recognition," in *SPIE Proc.,* 1985, vol. 575, pp. 279–285.

[101] O'Rourke and N. L. Badler, "Model-based image analysis of human motion using constraint propagation," *IEEE Trans. Patt. Anal. and Mach. Intell.,* vol. 2, pp. 522–536, 1980.

[102] S. Peleg and H. Rom, "Motion based segmentation," in *Proc., Int. Conf. on Patt. Recog.,* 1990, pp. 109–113.

[103] A. Pentland and B. Horowitz, "Recovery of nonrigid motion and structure," *IEEE Trans. Patt. Anal. and Mach. Intell.,* vol. 13, pp. 730–742, 1991.

[104] A. Pentland, B. Moghaddam, T. Starner, and M. Turk, "View-based and modular eigenspaces for face recognition," in *Proc. IEEE Computer Soc. Conf. on Computer Vision and Patt. Recog.,* 1994, pp. 84–91.

[105] D. Perkins, "A definition of caricature and recognition," *Studies in the Anthropology of Visual Commun.,* vol. 2, pp. 1–24, 1975.

[106] D. I. Perrett, A. J. Mistlin, and A. J. Chitty, "Visual neurons responsive to faces," *Trends in Neuroscience,* vol. 10, pp. 358–363, 1987.

[107] D. I. Perret, A. J. Mistlin, A. J. Chitty, P. A. Smith, D. D. Potter, R. Broennimann, and M. H. Harries, "Specialized face processing and hemispheric asymmetry in man and monkey: Evidence from single unit and reaction time studies," *Behav. Brain Res.,* vol. 29, pp. 245–258, 1988.

[108] D. I. Perret, P. A. Smith, D. D. Potter, A. J. Mistlin, A. S. Head, A. D. Milner, and M. A. Jeeves, "Visual cells in temporal cortex sensitive to face view and gaze direction," in *Proc. Royal Soc. of London, Series B,* 1985, vol. 223, pp. 293–317.

[109] T. Poggio and F. Girosi, "Networks for approximation and learning," *Proc. IEEE,* vol. 78, pp. 1481–1497, 1990.

[110] T. Poggio and V. Torre, "Ill-posed problems and regularization analysis in early vision," MIT AI Lab, Tech. Rep. AI Memo 773, 1984.

[111] A. Rahardja, A. Sowmya, and W. Wilson, "A neural network approach to component versus holistic recognition of facial expressions in images," in *SPIE Proc.: Intell. Robots and Computer Vision X: Algorithms and Techn.,* vol. 1607, 1991, pp. 62–70.

[112] D. Reisfeld and Y. Yeshuran, "Robust detection of facial features by generalized symmetry," in *Proc. 11th Int. Conf. on Patt. Recog.,* 1992, pp. 117–120.

[113] T. Sakai, M. Nagao, and S. Fujibayashi, "Line extraction and pattern recognition in a photograph," *Patt. Recog.,* vol. 1, pp. 233–248, 1969.

[114] A. Samal and P. Iyengar, "Automatic recognition and analysis of human faces and facial expressions: A survey," *Patt. Recog.,* vol. 25, pp. 65–77, 1992.

[115] B. G. Schunck, "Image flow segmentation and estimation by constraint line clustering," *IEEE Trans. Patt. Anal. and Mach. Intell.,* vol. 11, pp. 1010–1027, 1989.

[116] M. Seibert and A. Waxman, "Recognizing faces from their parts," in *SPIE Proc.: Sensor Fusion IV: Control Paradigms and Data Structures,* vol. 1611, 1991, pp. 129–140.

[117] ——, "Combining evidence from multiple views of 3-D objects," in *SPIE Proc.: Sensor Fusion IV: Control Paradigms and Data Structures,* vol. 1611, 1991.

[118] ——, "An approach to face recognition using saliency maps and caricatures," in *Proc. World Cong. on Neural Networks,* 1993, pp. 661–664.

[119] J. Sergent, "Microgenesis of face perception," in *Aspects of Face Processing,* H. D. Ellis, M. A. Jeeves, F. Newcombe, and A. Young, Eds. Dordrecht: Nijhoff, 1986.

[120] J. W. Shepherd, "An interactive computer system for retrieving faces," in *Aspects of Face Processing,* H. D. Ellis, M. A. Jeeves, F. Newcombe, and A. Young, Eds. Dordrecht: Nijhoff, 1985, pp. 398–409.

[121] A. Shio and J. Sklansky, "Segmentation of people in motion," in *Proc. IEEE Workshop on Visual Motion,* 1991, pp. 325–332.

[122] A. Singh, "An estimation—Theoretic framework for image flow analysis," in *Proc. Int. Conf. on Computer Vision,* 1990, pp. 167–177.

[123] S. A. Sirohey, "Human face segmentation and identification," Tech. Rep. CAR-TR-695, Center for Autom. Res., Univ. Maryland, College Park, MD, 1993.

[124] L. Sirovich and M. Kirby, "Low-dimensional procedure for the characterization of human face," *J. Opt. Soc. Amer.,* vol. 4, pp. 519–524, 1987.

[125] H. L. Snyder, "Image quality and face recognition on a television display," *Human Factors,* vol. 16, pp. 300–307, 1974.

[126] A. Spoerri and S. Ullman, "The early detection of motion boundaries," in *Proc. Int. Conf. on Computer Vision,* 1987, pp. 209–218.

[127] R. B. Starkey and I. Aleksander, "Facial recognition for police purpose using computer graphics and neural networks," in *Proc. Colloquium on Electron. Images and Image Proc. in Security and Forensic Science,* dig. no. 0871990, pp. 2/1–2.

[128] T. J. Stonham, "Practical face recognition and verification with WISARD," in *Aspects of Face Processing,* H. D. Ellis, M. A. Jeeves, F. Newcombe, and A. Young, Eds. Dordrecht: Nijhoff, 1984, pp. 426–441.

[129] M. Subbarao, "Interpretation of image flow: A spatio-temporal approach," *IEEE Trans. Patt. Anal. and Mach. Intell.,* vol. 11, pp. 266–278, 1989.

[130] D. Terzopoulus, A. Watkin, and M. Kass, "Constraints on deformable models: 3-D shape on nonrigid motion," *Artif. Intell.,* vol. 36, pp. 91–123, 1988.

[131] W. B. Thompson and T. C. Pong, "Detecting moving objects," in *Proc. 1st Int. Conf. on Computer Vision,* 1987, pp. 201–208.

[132] A. N. Tikhonov and V. Y. Arsenin, *Solution of Ill-posed Problems.* Washington, DC: Winston and Wiley, 1977.

[133] M. A. Turk and A. P. Pentland, "Face recognition using eigenfaces," in *Proc. Int. Conf. on Patt. Recog.,* 1991, pp. 586–591.

[134] S. Ullman, *The Interpretation of Visual Motion.* Cambridge, MA: MIT Press, 1979.

[135] J. Y. A. Wang and E. H. Adelson, "Layered representation for motion analysis," in *Proc. IEEE Computer Soc. Conf. Computer Vision and Patt. Recog.,* 1993, pp. 361–366.

[136] A. M. Waxman, B. Kamgar-Parsi, and M. Subbarao, "Closed-form solutions to image flow equations for 3-D structure and motion," *Int. J. Computer Vision,* vol. 1, pp. 239–258, 1987.

[137] H. Wechsler, *Computational Vision.* Boston: Academic, 1990.

[138] J. Weng, T. S. Huang, and N. Ahuja, *Motion and Structure for Image Sequences,* T. S. Huang, Ed. New York: Springer-Verlag, 1993.

[139] ——, "Learning recognition and segmentation of 3D objects from 2D images," in *Proc. IEEE Int. Conf. on Computer Vision,* 1993, pp. 121–128.

[140] P. A. Wintz, "Transform picture coding," *Proc. IEEE,* vol. 60, pp. 809–820, 1972.

[141] K. Wohn and A. M. Waxman, "The analytic structure of image flows: Deformation and segmentation," *Computer Vision, Graphics and Image Process.,* vol. 49, pp. 127–151, 1990.

[142] C. Wu and J. Huang, "Human face profile recognition by computer," *Patt. Recog.,* vol. 23, pp. 255–259, 1990.

[143] Y. Yacoob and L. S. Davis, "Computing spatio-temporal representations of human faces," in *Proc. IEEE Computer Soc. Conf. on Computer Vision and Patt. Recog.,* 1994, pp. 70–75.

[144] G. Yang and T. S. Huang, "Human face detection in a scene," in *Proc. IEEE Conf. on Computer Vision and Patt. Recog.,* 1993, pp. 453–458.

[145] A. Yuille, D. Cohen, and P. Hallinan, "Feature extraction from faces using deformable templates," in *Proc. IEEE Computer Soc. Conf. on Computer Vision and Patt. Recog.,* 1989, pp. 104–109.

[146] C. T. Zahn and R. S. Roskies, "Fourier descriptors for plane closed curves," *IEEE Trans. Computers,* vol. COM–21, pp. 269–281, 1972.

[147] Z. Zhang and O. Faugeras, "3D dynamic scene analysis," in *Springer Series in Information Sciences, Vol. 27,* T. S. Huang, Ed. New York: Springer-Verlag, 1992.

[148] Y. T. Zhou and R. Chellappa, "A network for motion perception," in *Proc. Int. Conf. on Neural Networks,* pp. 875–884, 1990.

Additional References

[1] M. K. Fleming and G. W. Cottrell, "Categorization of faces using unsupervised feature extraction," in *Proc. Int. Joint Conf. on Neural Networks,* pp. 65–70, 1990.

[2] N. M. Marinovic and G. Eichmann, "Feature extraction and pattern classification in space—Spatial frequency domain," in *SPIE Proc.: Intell. Robots and Computer Vision,* vol. 579, 1985, pp. 19–26.

[3] L. de Floriani. "Feature extraction from boundary models of three-dimensional objects," *IEEE Trans. Patt. Anal. and Mach. Intell.,* vol. 11, pp. 785–798, 1989.

[4] T. Abe, "Automatic identification of human faces by 3-D shape of surface—Using vertices of B-spline surface," in *Systems and Computers in Japan,* 1991.

[5] E. R. Brocklehurst, "Computer methods of signature verification," *IEE Colloq. Dig. (80): Colloq. on MMI in Computer Security,* 1986, pp. 3/1–5.

[6] L. D. Harmon and K. C. Knowlton, "Picture processing by computer," *Science,* vol. 164, pp. 19–29, 1969.

[7] P. L. Hawkes, "The commercialization of biometric techniques for automatic personal identification," *IEE Colloq. Dig. (80): Colloq. on MMI in Computer Security,* 1986, pp. 1/1–2.

[8] H. Mannaert and A. Oosterlinck, "Self-organizing system for analysis and identification of human face," in *SPIE Proc. Appl. of Digital Image Process. XIII,* vol. 1349, 1990, pp. 227–232.

[9] K. Flaton, "2D object recognition by adaptive feature extraction and dynamical link graph matching," Tech. Rep., Univ. S. Calif., 1992.

[10] Midorikawa, "Face pattern identification by back propagation learning procedure," *Neural Networks,* vol. 1, p. 515, 1988.

[11] M. Nixon, "Automated facial recognition and its potential for security," in *IEE Colloq. Dig. (80): Colloq. on MMI in Computer Security,* 1986, pp. 5/1–4.

[12] A. J. O'Toole, R. B. Millward, and J. A. Anderson, "A physical system approach to recognition memory for spatially transformed faces," *Neural Networks,* vol. 1, pp. 179–199, 1988.

[13] R. Brunelli and T. Poggio, "Face recognition through geometrical features," in *Proc. Europe. Conf. on Computer Vision,* 1992, pp. 792–800.

[14] D. Forsyth and A. Zisserman, "Invariant descriptors for 3-D object recognition and pose," *IEEE Trans. Patt. Anal. and Mach. Intell.,* vol. 13, pp. 971–991, 1991.

[15] T. Sakaguchi, O. Nakamura, and T. Minami, "Personal identification through facial images using isodensity lines," in *SPIE Proc.: Visual Commun. and Image Process. IV,* vol. 1199, 1989, pp. 643–654.

[16] I. Aleksander, W. Thomas, and P. Bowden, "A step forward in image processing," *Sensor Rev.,* pp. 120–124, 1984.

[17] S. Sclaroff and A. Pentland, "Closed-form solutions for physically based shape modeling and recognition," in *Proc. IEEE Computer Soc. Conf. on Computer Vision and Patt. Recog.,* 1991, pp. 238–243.

[18] A. G. Goldstein *et al.,* "Recognition of human faces from isolated facial features: A developmental study," *Psychonomic Sci.,* vol. 6, pp. 149–150, 1966.

[19] K. Takahashi, T. Sakaguchi, T. Minami, and O. Nakamura, "Description and matching of density variation for personal identification through facial images," in *SPIE Proc.: Visual Commun. and Image Process.,* vol. 1360, 1990, pp. 1694–1704.

[20] K. H. Wong, H. H. M. Law, and P. W. M. Tsang, "A system for recognizing human faces," in *Proc. Int. Conf. on Acoust., Speech, and Signal Process.,* 1989, pp. 1638–1642.

[21] S. Akamatsu, T. Sasaki, N. Masui, H. Fukamachi, and Y. Suenage, "A new method for designing face image classifiers using 3-D CG models," in *SPIE Proc.: Visual Commun. and Image Process.,* vol. 1606, 1991, pp. 204–216.

[22] S. Lele and J. T. Richtsmeier, "Euclidean distance matrix analysis: A coordinate-free approach for comparing biological shapes using landmark data," *Amer. J. Phys. Anthrop.,* vol. 86, pp. 415–427, 1991.

[23] G. Rhodes, "Lateralized process in face recognition," *British J. Psych.,* vol. 76, pp. 249–271, 1985.

[24] H. D. Ellis, "Processes underlying face recognition: Introduction," in *The Neurophysiology of Face Perception and Facial Expression,* R. Bruyer, Ed. Hillsdale, NJ: Erlbaum, 1986, pp. 1–27.

[25] J. Hochberg and R. E. Galper, "Recognition of faces: I. An exploratory study," *Psychonomic Science,* vol. 9, pp. 619–620, 1967.

[26] D. C. Hay and A. W. Young, "The human face," in *Normality and Pathology in Cognitive Function,* A. W. Ellis, Ed. London: Academic, 1982, pp. 173–202.

[27] G. Davies, H. Ellis, and J. Shepherd, "Perceiving and remembering faces," *Perceptual and Motor Skills,* 1965.

[28] S. Deutsch, "Conjectures on mammalian neuron networks for visual pattern recognition," *IEEE Trans. Syst. Sci. and Cybern.,* vol. 2, pp. 81–85, 1966.

[29] P. D. McCormack and S. P. Colletta, "Recognition memory for items from unilingual and bilingual lists," *Bull. Psychonomic Soc.,* vol. 6, pp. 149–151, 1975.

[30] H. Ellis, J. Sheppard, and G. Davies, "An investigation into the use of the photofit technique for recalling faces," *British J. Psych.,* vol. 66, pp. 29–37, 1975.

[31] H. Ellis, "Recognizing faces," *British J. Psych.,* vol. 66, pp. 409–426, 1975.

[32] M. S. Wogalter and D. B. Marwitz, "Face composite construction: In-view and from memory quality and improvement with practice," *Ergonomics,* 1991.

[33] J. F. Fagan, III, "Infants' recognition of invariant features of faces," *Child Develop.,* vol. 47, pp. 627–638, 1976.

[34] M. P. Young and S. Yamane, "Sparse population coding of faces in the inferotemporal cortex," *Sci.,* vol. 256, pp. 1327–1331, 1992.

[35] R. K. Yin, "Looking at upside-down faces," *J. Experimental Psych.,* vol. 81, pp. 141–145, 1969.

[36] J. Y. Cartoux, J. T. Lapreste, and M. Richetin, "Face authentification or recognition by profile extraction from range images," in *Proc. Workshop on Interp. of 3D Scenes,* 1990, pp. 194–199.

[37] P. Maragos, "Tutorial on advances in morphological image processing and analysis," *Opt. Eng.,* vol. 26, pp. 623–632, 1987.

[38] J. C. Lee and E. Milios, "Matching range images of human faces," in *Proc. 3rd Int. Conf. on Computer Vision,* 1990, pp. 722–726.

[39] Y. Yacoob and L. S. Davis, "Labeling of human face components from range data," in *Proc. IEEE Computer Soc. Conf. on Computer Vision and Patt. Recog.,* 1993, pp. 592–593.

[40] T. Minami, "Present state of the art on identification of human face," *J. Soc. Instrum. and Control Eng. (Japan),* vol. 25, pp. 707–713, 1986.

[41] M. A. Fischler and R. A. Elschlager, "The representation and matching of pictorial structures," *IEEE Trans. Computers,* vol. COM–22, pp. 67–92, 1973.

[42] S. Carey and R. Diamond, "From piecemeal to configurational representation of faces," *Sci.,* vol. 195, pp. 312–314, 1977.

[43] S. S. Culbert, "Object recognition as a function of number of different views during training," *Perceptual and Motor Skills,* 1965.

[44] M. A. Jeeves, "Plenary Session. An overview. Complementary approaches to common problems in face recognition," in *Aspects of Face Processing,* H. D. Ellis, M. A. Jeeves, F. Newcombe and A. Young, Eds. Dordrecht: Nijhoff, 1986, pp. 445–452.

[45] J. Carter and M. Nixon, "An integrated biometric database," in *Colloq. on Electron. Images and Image Process. in Security and Forensic Sci.,* dig. no. 87, 1986, pp. 8/1–6.

[46] T. Y. Tu, C. Ma, and Y. L. Ma, "Image recognition by the Kolmogorov complexity program," in *SPIE Proc.: Int. Symp. on Patt. Recog. and Acoust. Imaging,* vol. 768, 1987, pp. 343–350.

[47] P. O'Higgins and N. W. Williams, "An investigation into the use of Fourier coefficients in characterizing cranial shapes in primates," *J. Zoology, London,* vol. 211, pp. 409–430, 1987.

[48] Z. De-Chen and Shen Xuan-Jing, "Fast texture image segmentation," in *SPIE Proc.: Appli. of Digital Image Process. XIII,* vol. 1349, 1990, pp. 277–282.

[49] C. A. Rothwell, A. Zisserman, C. I. Marinos, D. A. Forsyth, and J. L. Mundy, "Relative motion and pose from arbitrary plane curves," *Image and Vision Computing,* vol. 10, pp. 250–262, 1992.

[50] R. M. Haralick, "A facet model for image data," *Computer Graphics and Image Process.,* vol. 15, pp. 113–129, 1981.

[51] A. Pentland and B. Horowitz, "Recovery of nonrigid motion and structure," *IEEE Trans. Patt. Anal. and Mach. Intell.,* vol. 13, pp. 730–742, 1991.

[52] M. Musen and J. Van˜Der Lei, "Of brittleness and bottlenecks: Challenges in the creation of pattern recognition and expert system models," in *Pattern Recognition and Artificial Intelligence,* E. S. Gelsema and L. N. Kanal, Eds. Amsterdam: North Holland, 1988.

[53] R. M. Haralick, "Ridges and valleys on digital images," *Computer Graphics and Image Process.,* vol. 22, pp. 28–38, 1983.

[54] Y. Q. Cheng, Y. M. Zhuang, and J. Y. Yang, "Optimal Fisher discriminant analysis using the rank decomposition," *Patt. Recog.,* vol. 25, pp. 101–111, 1992.

[55] B. P. Yuchas, Jr., M. H. Goldstein, T. J. Sejnowski, and R. E. Jenkins, "Neural network models of sensory integration for improved vowel recognition," *Proc. IEEE,* vol. 78, pp. 1658–1668, 1990.

[56] J. Q. Fang and T. S. Huang, "Some experiments on estimating the 3-D motion parameters of a rigid body from two consecutive image frames," *IEEE Trans. Patt. Anal. and Mach. Intell.,* vol. 6, pp. 547–554, 1984.

[57] J. W. Roach and J. K. Aggarwal, "Determining the movements of objects from a sequence of images," *IEEE Trans. Patt. Anal. and Mach. Intell.,* vol. 2, pp. 554–562, 1980.

[58] R. Y. Tsai and T. S. Huang, "Uniqueness and estimation of 3-D motion parameters of rigid bodies with curved surfaces," *IEEE Trans. Patt. Anal. and Mach. Intell.,* vol. 6, pp. 13–27, 1984.

[59] J. Weng, T. S. Huang, and N. Ahuja, "Motion and structure from two prospective views: Algorithms, error analysis and error estimation," *IEEE Trans. Patt. Anal. and Mach. Intell.,* vol. 11, pp. 451–476, 1989.

[60] T. J. Broida and R. Chellappa, "Estimation of object motion parameters from noisy images," *IEEE Trans. Patt. Anal. and Mach. Intell.,* vol. 8, pp. 90–99, 1986.

[61] J. Aisbett, "An iterated estimation of the motion parameters of a rigid body from noisy displacement vectors," *IEEE Trans. Patt. Anal. and Mach. Intell.,* vol. 12, pp. 1092–1098, 1990.

[62] H. C. Longuet-Higgins, "A computer program for reconstructing a scene from two projections," *Nature,* vol. 293, pp. 133–135, 1981.

[63] Y. Yasumoto and G. Medioni, "Robust estimation of three-dimensional motion parameters from sequence of image frames using regularization," *IEEE Trans. Patt. Anal. and Mach. Intell.,* vol. 8, pp. 464–471, 1986.

[64] C. L. Fennema and W. R. Thompson, "Velocity determination in scenes containing several moving objects," *Computer Graphics and Image Process.,* vol. 9, pp. 301–315, 1979.

[65] X. Zhuang, T. S. Huang, and R. M. Haralick, "Two-view motion analysis: A unified algorithm," *J. Opt. Soc. Amer.,* vol. 3, pp. 1492–1500, 1986.

[66] G. S. Young and R. Chellappa, "3-D motion estimation using a sequence of noisy stereo images: Models, estimation, and uniqueness results," *IEEE Trans. Patt. Anal. and Mach. Intell.,* vol. 12, pp. 735–759, 1990.

[67] J. A. Webb and J. K. Aggarwal, "Structure from motion of rigid and jointed objects," *Artif. Intell.,* vol. 19, pp. 107–130, 1982.

[68] M. Spetasakis and Y. Aloimonus, "A multi-frame approach to visual motion perception," *Int. J. Computer Vision,* vol. 6, pp. 245–255, 1991.

[69] Y. Kim and J. Aggarwal, "Determining object motion in a sequence of stereo images," *IEEE J. Robotics and Autom.,* vol. RA-3, pp. 599–614, 1987.

[70] S. Liou and R. Jain, "Motion detection in spatio-temporal space," *Computer Vision, Graphics and Image Process.,* vol. 45, pp. 227–250, 1989.

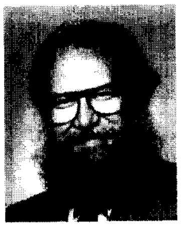
[71] Z. Zhang, O. Faugeras, and N. Ayache, "Analysis of a sequence of stereo scenes containing multiple moving objects using rigidity constraints," in *Proc. Int. Conf. on Computer Vision,* 1988, pp. 177–186.

[72] H. Chen and T. Huang, "Matching 3-D line segments with applications to multiple-object motion estimation," *IEEE Trans. Patt. Anal. and Mach. Intell.,* vol. 12, pp. 1002–1008, 1990.

[73] I. Sethi and R. Jain, "Finding trajectories of feature points in a monocular image sequence," *IEEE Trans. Patt. Anal. and Mach. Intell.,* vol. 9, pp. 56–73, 1987.

Rama Chellappa (Fellow, IEEE) is a Professor in the Department of Electrical Engineering at the University of Maryland, where he is also affiliated with the Institute for Advanced Computer Studies, the Center for Automation Research and the Computer Science Department. He is an Editor of *Collected Papers on Digital Image Processing* (IEEE Computer Society Press, 1992). He coauthored *Artificial Neural Networks for Computer Vision* (Springer Verlag, 1992) and coedited *Markov Random Fields: Theory and Applications* (Academic Press, 1993). He was an Associate Editor for *IEEE Transactions on Acoustics, Speech, and Signal Processing* and *IEEE Transactions on Neural Networks.* He is presently Coeditor-in-Chief of *Computer Vision, Graphics, and Image Processing: Graphic Models and Image Processing* and *IEEE Transactions on Image Processing.* He has authored 20 book chapters and over 150 peer-reviewed journal and conference papers. He was the General Chairman of the IEEE Computer Society Conference on Computer Vision and Pattern Recognition and of the IEEE Computer Society Workshop on Artificial Intelligence for Computer Vision (1989). He was the Program Chairman of the IEEE Signal Processing Workshop on Neural Networks for Signal Processing, and is the Program Chairman for the 2nd International Conference on Image Processing.

Dr. Chellappa received the 1985 National Science Foundation Presidential Young Investigator Award and the 1985 IBM Faculty Development Award. In 1990 he received the Excellence in Teaching Award from the School of Engineering at the University of Southern California. He is a corecipient of four NASA certificates for his work on synthetic aperture radar image segmentation.



Charles L. Wilson (Senior Member, IEEE) has been with the National Institute of Standards and Technology, Gaithersburg, MD, for the past 15 years. He is presently Manager of the Visual Image Processing Group of the Advanced Systems Division. His was with Los Alamos National Laboratory and AT&T Bell Laboratories. His current research interests are in application of statistical pattern recognition, neural network methods, and dynamic training methods for image recognition, image compression, and in standards used to evaluate recognition systems.

Dr. Wilson received a DOC Gold Medal in 1983 for his work in semiconductor device simulation.

Saad A. Sirohey (Member, IEEE) received the B.Sc. (highest honors) in electrical enginering from King Fahd University of Petroleum and Minerals, Dhahran, Saudi Arabia, and the M.S. degree in electrical engineering from the University of Maryland at College Park, in 1990 and 1993, respectively. He is working toward the Ph.D. in electrical engineering at the University of Maryland at College Park.

He is a Research Assistant at the Center for Automation Research at the University of Maryland. His current research interests include signal/image processing and computer vision, specifically automated face recognition.