NISTIR 5465

Face Recognition Technology for Law Enforcement Applications

C. L. Wilson
C. S. Barnes
R. Chellappa
S. A. Sirohey

U.S. DEPARTMENT OF COMMERCE
Technology Administration
National Institute of Standards
and Technology
Computer Systems Laboratory
Advanced Systems Division
Gaithersburg, MD 20899

NIST

NISTIR 5465

Face Recognition Technology for Law Enforcement Applications

C. L. Wilson
C. S. Barnes
R. Chellappa
S. A. Sirohey

U.S. DEPARTMENT OF COMMERCE
Technology Administration
National Institute of Standards
and Technology
Computer Systems Laboratory
Advanced Systems Division
Gaithersburg, MD 20899

July 1994

U.S. DEPARTMENT OF COMMERCE
Ronald H. Brown, Secretary

TECHNOLOGY ADMINISTRATION
Mary L. Good, Under Secretary for Technology

NATIONAL INSTITUTE OF STANDARDS AND TECHNOLOGY
Arati Prabhakar, Director

Face Recognition Technology for Law Enforcement Applications

C. L. Wilson
C. S. Barnes
National Institute of
Standards and Technology
Gaithersburg, MD 20899

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

Abstract

The goal of this report is to relate existing face recognition technology to law enforcement applications. These applications range from static matching of controlled photographs as in mugshot matching to surveillance video images and 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. The ongoing research activities have been given a renewed emphasis over the last five years. The existing techniques and systems have been tested on different sets of images of varying complexities. Also, 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 law enforcement applications.

In this report, we first present different possible scenarios that arise in law enforcement applications. Special constraints that are present in these applications are pointed out. 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 twenty 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. Examples of face recognition and matching techniques applied to specific law enforcement problems are pointed out.

Real-time recognition from video images acquired in a cluttered scene such as an airport is probably the most challenging problem of face recognition. As not much has been reported on this 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 several issues such as data collection, performance metrics and evaluation of systems and techniques that are relevant to law enforcement. Finally, a summary and conclusions are given.

Acknowledgements

The authors are grateful to Professor A. Pentland for providing Figures 3, 5(a), 7, 13, Professor T. Poggio for providing Figure 8, Professor C. V. d. Malsburg for providing Figures 2, 6(a), 9 and 10, Dr. G. Gordon for providing Figure 11, Professor B. S. Manjunath for providing the feature extraction and face identification algorithms (results of which are shown in Figure 6(b) and Figure 10), and Ms. M. Lepley for providing Figure 1.

Contents

1 Introduction

The goal of this report is to relate the extensive Face Recognition Technology (FRT) which is available in the literature to law enforcement applications. Possible applications of FRT developed by the FBI are shown in Table 1. These applications range from static matching of controlled format photographs in mugshot matching (application 1) to real-time matching of surveillance video images (application 3). The applications have different advantages and disadvantages and present a wide range of different technical challenges.

In addition to the separation of images into static and real-time image sequences several other parameters are important in evaluating these applications. In any pattern recognition problem the accuracy of the solution will be strongly effected by the limitations placed on the problem. It is impractical to match a blurry photo against a database of photographs of the entire US population. 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. One of the objectives of this report is to investigate existing algorithms and ask if the restrictions allowed by the applications present in Table 1 make it likely that these algorithms can be used to solve practical law enforcement problems.

Three different kinds of problems are presented 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, 5, and 6 involve finding or creating a face image which is similar to the human recollection of a face. Finally, applications 7 , 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.

Over the past twenty 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 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. A brief survey of the psychophysics and neurophysiology literature is presented in Section 3.

Application	Advantages	Disadvantages
1. Mugshot Matching	Controlled image Controlled segmentation Good quality Consistent with IPS/III	No existing database Large potential database Rare search type
2. Bank/Store Security	High value Geographic search limits	Uncontrolled segmentation Low image quatity
3. Crowd Surveillance	High value Small file size	Uncontrolled segmentation Low image quality Real-time
4. Expert Identification	High value Enhancement possible	Low image quality Legal certainty required
5. Witness Face Reconstruction	Genetic optimization	Unknown similarity
6. Electronic Mugshot Book	Descriptor search limits Genetic optimization	Unknown similarity
7. Electronic Lineup	Descriptor search limits	Unknown similarity
8. Reconstruction of Face from Remains	High value	Requires physiological input
9. Computerized Aging	Missing children	Requires example input

Table 1: Law Enforcement Usage of Faces.

In Sections 4 and 5 we give a detailed account of face recognition reported in the engineering literature. Barring a few exceptions [18, 20, 107], research on machine recognition of faces has developed independently of studies in psychophysics and neurophysiology. During the early and mid-seventies, typical pattern classification techniques, which use measured attributes between features in faces or face profiles, were used. During the eighties, work on face recognition remained largely dormant. Since the early nineties, research interest in FRT has grown very significantly. One can attribute this to several reasons: a reduction in defense related research funds, with a corresponding increase in emphasis on civilian research projects; the re-emergence 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 [9, 77, 124, 99, 115], and more recently HyperBF networks [18], have been used. Barring a few exceptions [99], many of the existing approaches have been tested on relatively small datasets, typically less than 100 images.

The large amount of published work on face recognition has treated the problem as a generic image recognition problem. However, there are several papers that have specifically

addressed law enforcement applications. We briefly summarize these at the end of Section 4.

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 mugshot problem, due to the availability of side views of the face.

All of the discussion thus far has focussed on recognizing faces from still images. The still image problem has several inherent advantages and disadvantages as noted in Table 1. For applications such as mugshot 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 [124, 108] have addressed the issue of segmenting a face image from a video sequence. However, there is a significant amount of work reported in the Image Understanding (IU) literature [1, 2] on segmenting a moving object from a sequence. Also, there is a significant amount of work on the analysis of non-rigid moving objects, including faces, in the IU [1, 2] as well as the image compression literature [3]. We briefly discuss those techniques that have potential applications to recovery and reconstruction (in 3-D) of faces from a video sequence. The reconstructed image will be useful for recognition tasks when disguises and aging are present. Section 6 addresses these issues.

A last, but not least important, group of issues involves data collection, evaluation and benchmarking of existing algorithms and systems. These are addressed in Section 7.

Finally, a summary and conclusions are given in Section 8.

2 Applications

As indicated in Table 1, applications of FRT 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, 5 and 6). 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.

2.1 Static matching

Mugshot matching is the most common application in this group. Typically, in mugshot 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 mugshot images are given in Figure 1.

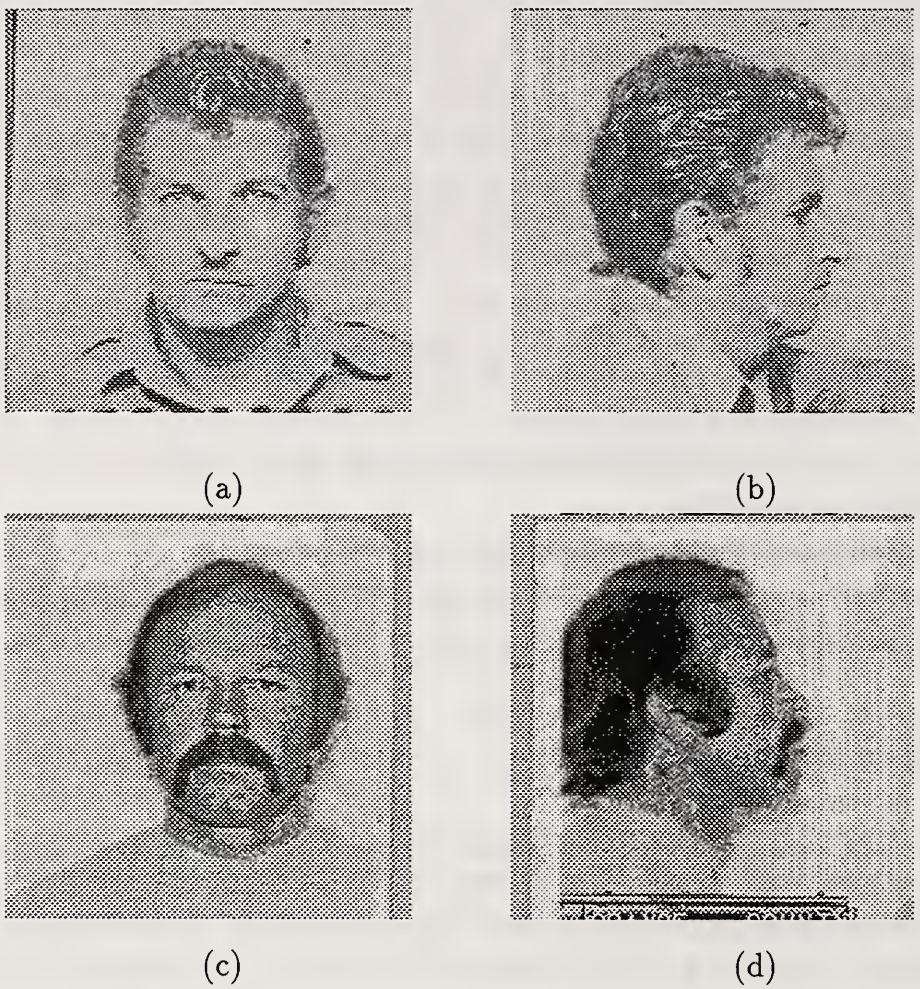


Figure 1: Frontal and profile mugshot images

Simple versions of the mugshot matching problem are recognition of faces in driver's licenses, personal ID cards, and passports. Typical examples of face images in drivers licenses or personal ID cards are shown in Figure 2. The images in these documents seem to be acquired with more control than in mugshots.

Typically, images in mugshot applications are of good quality, consistent with IPS/III standards. Given the reasonably controlled imaging conditions, segmentation/location of

(a) (b)

Figure 2: Face images in a controlled background, as in passport or identification documents

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 Figure 3. Applications 1 and 2 do not require real time recognition. 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 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 through 7 involve finding or creating a face image which is similar to the human recollection of a face. 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 mugshot book, they tend to make more recognition errors as the number of mugshot presentations increases. It is difficult to completely quantify the degradation in machine implementation of algorithms developed for applications 4 through 6. Another issue is the incorporation of mechanisms for recalling faces that humans use in the algorithms. Applications 4 through 6 need a strong interaction of algorithms and known results in psychophysics and neuroscience studies.

Applications 8 and 9 involve transformations of images from current data to what they

(a) (b)

Figure 3: An approximate illustration of an uncontrolled environment for face images corresponding to application 2

could have been (application 8) or to what they will be (applications 9). These are even more difficult than applications 4 through 6, since "smoothing" or "predictive" mechanisms need to be incorporated into the algorithms.

2.2 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 [82] 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 for image sequence based segmentation, structure estimation, non-rigid 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.

3 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 the human brain. 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 law enforcement applications the images, e.g., mugshots, are present in single or multiple views of 2-D 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 report. 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 [33, 35, 16] and survey papers [12, 37, 17, 61].

The issues that are of interest to designers are:

- **Is face recognition a dedicated process? [37]**: 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 prewired 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 mugshot matching applications, where profiles can be extracted from side views), several fiducial points ("features") are around the nose region (see Section 5). 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:** **[17]** Perkins [100] 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 [60, 43] concluded that information in low spatial frequency bands plays a dominant role in face recognition. Recent studies [110] show that, depending on the specific recognition task, the low, bandpass and high frequency components may play different roles. For example the sex judgement 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:** **[36]** 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 showed that 73% of the victims had unilateral right hemisphere lesions, 17% had bilateral lesions, and 10% had unilateral left hemisphere lesions. This survey 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 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. Language processing is usually situated in the left hemisphere and may have reduced the available space for processing of visuospatial information. 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. [25, 26]** 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, hairstyle, hats, etc. Ten year old children exhibit this behavior less frequently, while children older than twelve 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 [25].

- **Facial expression: [16]** 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 also 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 [46] 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 [46]. Using the same data collected in [46], some studies have been done to quantify the role of gender in face recognition. It has been found [45] that in a Japanese population, 65% 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. The variation of women's faces is generally slight when compared to men's faces [45].

- **Image quality:** In [116] 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 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 [116].

3.1 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. As the evidence suggests that analysis of facial expressions is parallel to face recognition, designers interested only in face recognition should evaluate the usefulness of analyzing expressions. 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 mugshot books and lineups. Interpreting face recognition using Marr's computational vision paradigm may point to new algorithms and systems; see Chapter 6 of [16]. Other issues, such as organization of face memory, are very pertinent for the design of large databases such as mugshot albums. Issues such as the role of the brain, prosopagnosia, etc. may be marginal to designers.

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 is a pioneering example of such an effort. Such efforts are not very common, as a majority of engineering researchers lack the required background in psychophysics and neurosciences. While many would argue that good understanding of human perception is a necessary condition for designing computer vision algorithms, most simply take an engineering approach and design an algorithm or system that accepts available data and produces the reasonable results.

Our general feeling is that 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 point of view.

4 Face Recognition from Still Intensity and Range Images

In this section we summarize the state of the art in face recognition in the engineering literature. In Section 4.1, existing work on segmenting a face from a still intensity image is summarized. Extraction of features such as eyes, mouth, points of high curvature and singular values are reviewed in Section 4.2. Section 4.3 is a very detailed review of over twenty years of work in face recognition in the engineering literature. Statistical, structural and neural network approaches are discussed. Finally, a summary section is included.

4.1 Segmentation

Craw et al. in [30] 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 $8 \times 8, 16 \times 16, 32 \times 32, 64 \times 64$ 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. Since connectivity is based on similar edge direction and magnitude, head edges can be easily confused with all the other edges present in the image. To compensate for this, the template is used as a guide to find the head outline. The results obtained at a lower resolution are used as guidelines for the next higher resolution, since the chances of error are at their minimum for the lowest resolution. 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 results obtained show reasonably good success in detecting the head outline, though the search for the eyes did not succeed as well. An improvement in the system could be achieved by using Canny's [24] or Burr's [21] edge finder. In later work by the author, this technique was replaced by a wire mesh method.

In [31] Craw, Tock and Bennet describe a system to recognize and measure facial features. Their work was motivated in part by automated indexing of police mugshots. They endeavor to locate 40 feature points from a grey-scale image; these feature points were chosen according

to Shepherd [111], 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. [52]. 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 [42]. After a rough idea of the location of the head is obtained, refinement is done by transforming individual vectors of the polygon; details of this are given in [31]. 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. Feature experts (templates) were used at this stage. 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. The authors mention further work related to obtaining the eye blink rate from a sequence of images.

Govindaraju et al. [51] 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. [136]. 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.

In [114] the face is segmented from a moderately cluttered background. The approach taken involves working with both the intensity image of the face as well as the edge image found using Canny's edge finder [24]. Pre-processing 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.

$$2x_i a_0 - y_i^2 a_1 + 2y_i a_2 - a_3 = x_i^2 \tag{1}$$

where

$$a_0 = x_0 \quad , \quad a_1 = \frac{a^2}{b^2}$$

$$a_2 = \frac{a^2}{b^2} y_0 \quad , \quad a_3 = x_0^2 + \frac{a^2}{b^2} - a^2$$

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.

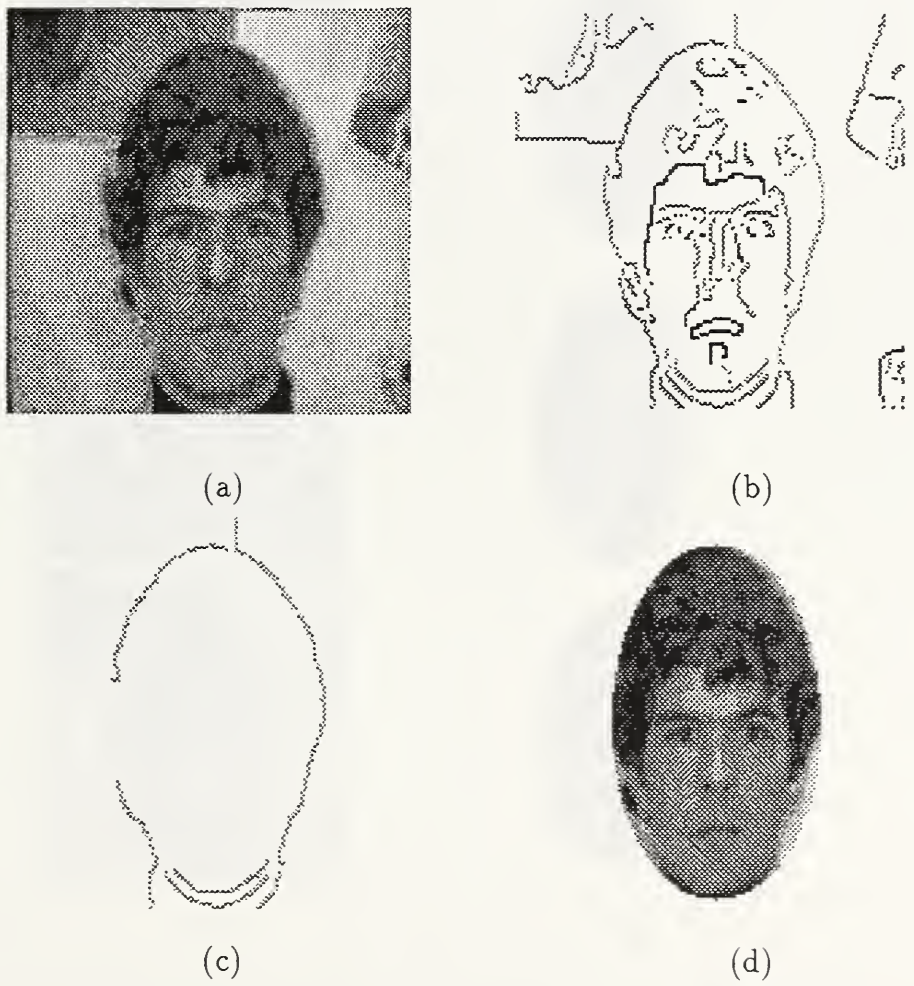
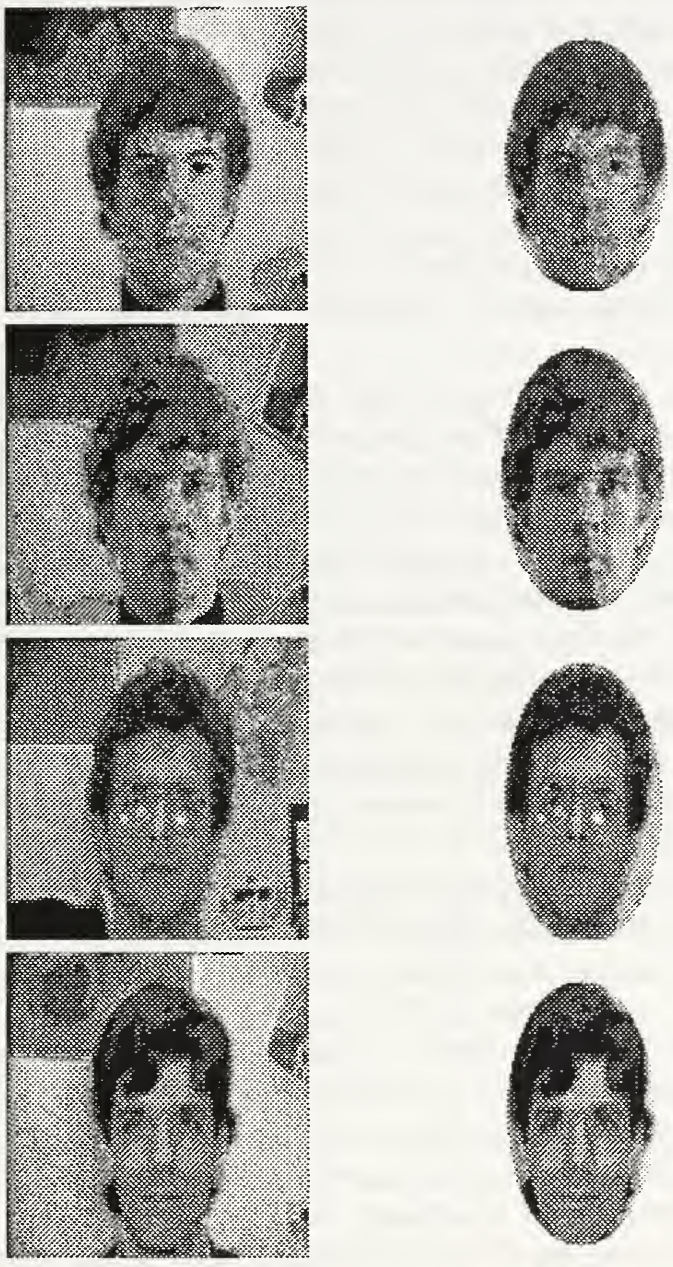


Figure 4: (a) Input image, (b) Edge image, (c) Linked segments, (d) Segmented image

Figure 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. Figure 5 shows some of the results of the segmentation algorithm. The image size was 128×128 pixels. Further work on locating the internal features is in progress; resulting improvement in the accuracy is expected.



(a) (b)

Figure 5: Good results of segmentation. (a) Input image, (b) Extracted image

4.2 Feature Extraction

Early work on face recognition was done by Sakai et al. in [105]. They used a digitized image with eight grey levels. The work was conducted on a data set consisting of frontal face images. A 3×3 mesh was used to determine pixels having the greatest gradient values so that the amount of information is reduced to the bare essentials. These pixels were then connected to neighboring pixels exhibiting similar characteristics to form line and contour segments. The recognition principle used was pattern matching using a pre-defined face template. A coarse to fine approach is used to determine individual features of the face. It should be noted that recognition does not differentiate among different faces; rather, it determines the existence of a face in the image. The authors note that the procedure they employ has a dependency on the illumination direction. Changes made to the illumination direction cause problems to their approach. The same commented has been made by various other researchers.

Reisfeld and Yeshurun in [104] 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 [104]. From their paper it is not apparent how they determine whether a feature is the eye or the nose, i.e. no qualifiers are given for this determination. They indicate their procedure's superiority over other correlation based schemes like that of Baron [12] 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.

Yuille, Cohen and Hallinen in [136] 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; it takes 5 to 10 minutes on a SUN 4 even with a good starting point selected. 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.

In [95] Nixon uses the Hough transform to achieve facial recognition. 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 almost certainly round perimeter of the iris is attractive because detection of a circular shape is an established task in image processing. The perimeter of the eye's sclera is a distinct part and may also be employed in detection. The sclera has the advantage that its shape is reflected by the region below the eyebrows. This allows for eye spacing measurement even when the eyes are closed or the iris is not round due to illness. 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 procedure for locating each eye is constrained to that half of the image. The Hough transform is applied to detect the instance of each shape in a six subject data set. 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. Detection of the position of the eyebrows provides a measurement of eye spacing which is greater than that provided by the other techniques but which can be used when the others cannot be discriminated [95].

In [65], 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. The singular values (SV) extracted by the SVD have good performance as shape descriptors. The SVs of an image are very stable and represent the algebraic attributes of the image, being intrinsic but not necessarily visible. [65] proves their stability and invariance to proportional variance of image intensity in the optimal discriminant vector space, to transposition, rotation, translation, and reflection which are important properties of the SV feature vector. Representing an image as an n-dimensional SV feature vector, the recognition problem is solved in an n-dimensional feature space. A facial photo of 32mm $\times$ 27mm typically needs a 70-dimensional SV feature vector to describe it. The original high-dimensional SV feature vector may be compressed to a lower dimensional feature space (2-D or 1-D) using various transforms. The Foley-Sammon transform is used to obtain the optimal set of discriminant vectors spanning the Sammon discriminant plane. With a small set of photos, Hong uses only two of the vectors for recognition; more of the discriminant vectors will be needed for recognition with more photos. The small set of photos consists of nine facial photos of size

50mm $\times$ 35mm. Each photo is sampled five times by varying the relative position between the photo and the TV camera used for scanning, for a total of 45 image samples with five images in each class. The SVD operation is applied to each image matrix for extracting SV features and the SV vector. 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 [65].

Classification approaches based on structure parameters are generally not robust for recognizing complex human face images. They are sensitive to changes in rotation, scaling, and facial expression. The development of shape descriptors [50] aims at describing the shape of an object independently of translation or rotation. In [28] 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 1-dimensional space. Three SV feature vectors are extracted from the training set in [28]. 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 [50]. 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% [28].

Manjunath et al. [83] 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 2-D Gabor function used and its Fourier transform are:

$$g(x, y : u_0, v_0) = \exp(-[x^2/2\sigma_x^2 + y^2/2\sigma_y^2] + 2\pi i[u_0 x + v_o y]) \quad (2)$$

$$G(u, v) = \exp(-2\pi^2(\sigma_x^2(u - u_0)^2 + \sigma_y^2(v - v_0)^2)) \quad (3)$$

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 non-orthogonal 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:

$$\Phi_\lambda(x, y, \theta) = \exp\left[(-\lambda^2(x'^2 + y'^2)) + i\pi x'\right] \quad (4)$$

$$x' = x\cos\theta + y\sin\theta \quad (5)$$

$$y' = -x\sin\theta + y\cos\theta \quad (6)$$

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

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

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 are sensitive to oriented lines and step edges of short lengths, and their response decreases if the lengths are increased.

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

where g is a sigmoid non-linearity, 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 maxima 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}:

$$I_{m,n}(x, y) = \max_{(x',y') \in N_{xy}} I_{m,n}(x', y') \quad (9)$$

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

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 [56]. Hallinan [56] 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

Figure 6: (a) Input image, (b) Feature points extracted

(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. The main point here is the estimation of the amount of degradation which determines the value of α. α is easily optimized since the percentage of trimming and the area of the trimmed template have a 1-1 correspondence.

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. Cost functionals for the template are developed which consist of a prior functional and a data functional. The data functional acts on both the intensity image and the edge potential field. The prior is developed by hand fitting the template to a training set and approximating the empirical distributions. The search for candidates uses a coarse to fine approach. The ordering of the search involves five steps: 1) dark center region, 2) edge bounding the iris, 3) orientation of the white region, 4) size of the white region, and 5) edges bounding the white region.

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 experiment were frontal face intensity images. Three sets of data were used for the experiments. 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. A tabular representation of results for three sets of values for the α's is shown in [56]. An increase in hit rate is reported when using α-trimmed distribution. The overall best hit rate reported was 80%.

[29] 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. A controller activates tasks according to the information available to it from the blackboard. It operates so as to accomplish a set of goals, with the ultimate goal being that of detecting the face. Once a task has been selected by the controller, the relevant knowledge sources are activated. A knowledge source consists of the programs that determine how the segments are to be classified. Finally, thresholding is done to determine the Face with the highest confidence level. The author reports good results in detecting faces using this method. The modularity of the system makes it possible to expand it by adding other knowledge sources such as eyebrows, ears, forehead, etc. Utilization of higher order knowledge from lower order resources forces the system to follow a particular path in an ordered manner. 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 [114].

4.3 Recognition

In [76] a basic study in classifying human faces is reported. The photographs used are front views of human faces, with mouth closed, no beards, and no eyeglasses. The individuals are looking into the camera when the photographs are taken. Euclidian distances between singular points on the face are used as parameters for characterizing a face; these parameters are resistant to changes in lighting and degree of development and to small changes of facial expression, and carry as much information as possible. Kaya et al. [76], estimate that the number of parameters should be greater than $\log_2 N$ bits where N equals the number of faces to be classified. They use the Euclidian distances between singular points on the face as the

characteristic parameters. The characteristic parameters are normalized by dividing by the nose length to account for any differences due to the size of the photograph and the distance of the subject from the camera. Photographs of 62 Japanese adults between the ages of 20 and 30 were used. The photographs were taken under the same lighting conditions. The characteristic parameters were measured by hand. The mean and standard deviation are calculated for the parameters. The correlation of the parameter indicates that the actual dimension of the parameter vector may be smaller than 9, the number of parameters. The parameter matrix is dominated by the principal eigenvector as a result of the normalization. The number of classifiable faces depends largely on the algorithm of classification. One metric for the effectiveness of the classification algorithm is the average number of parameters used to identify the face. A small number of parameters is desirable as it decreases the amount of memory needed to store the patterns. The workability of the algorithm is based on the probability of finding the correct face and the expected number of steps needed until its identity is established. For the nine geometric parameters used in [76], it was found that the amount of stored information may range from 8 to 14 bits depending on the noise estimate. The noise components are the inherent noise D_i and the noise D_m involved in extracting the parameters from the photograph. D_m consists of noise in detecting the outlines of ambiguous facial parts and quantization noise from the digitization of the photograph. D_m depends on the camera characteristics and on the algorithm used for extracting the parameters from the photograph. D_i is the variation of the parameters of a face from photograph to photograph; it is caused by variations in rotation, expression, etc. Variations in the geometry of the face over a short period of time are also considered to be D_i. The amount of information carried by the parameters is the average of the mutual information of the parameters and the estimated total noise [76].

In one of the earliest studies [74], three crucial steps were identified for face recognition. These are the selection of effective features to identify the faces, machine extraction of those features, and decision-making based on the feature measurements. 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 [74]. 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. Kanade regards the rescanning process as a type of visual accommodation in both light intensity and location. After the finer image data acquisition, more precise information is extracted from each region using thresholding, differentiation, and integral projection. The two stage process may reduce memory and time. Kanade uses a set of sixteen facial parameters which are ratios of distances, areas, and angles to compensate for the varying size of the pictures. 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 reference set. The remaining

20 pictures are used as a test set. Kanade uses a simple distance to measure the similarity between an image of the test set and the image in the reference set. Kanade's results range from 45% to 75% correct, depending on the parameters used. When several of the ineffective parameters are not used better results are obtained [74].

Recently, the use of the Karhunen-Loeve (KL) expansion for the representation [77, 115] and recognition [99, 124] of faces has generated renewed interest. The KL expansion has been studied for image compression for more than thirty years [131, 70]; its use in pattern recognition applications has also been documented for quite some time [41]. 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 [70]. In [115], Sirovich and Kirby revisit the problem of KL representation of images (cropped faces). Noting that the number of images M usually available for computing the covariance matrix of the data is much less than the row or column dimensionality of the covariance matrix, leading to the singularity of the matrix, they use a standard method from linear algebra [14] that will calculate only the M eigenvectors that do not belong to the null space of the degenerate matrix. Once the eigenvectors (referred to as eigenpictures) are obtained, any image in the ensemble can be 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 was on the representation of human faces.

In a subsequent extension of their work, Kirby and Sirovich in [77] 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 eigencomputations 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.

Turk and Pentland [124] used eigenpictures (also known as eigenfaces in [124]) 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 Euclidian distance) to the weights of the test image. By using the observation that the projection of a face image and a non-face image are quite different, a method for detecting the presence of a face in a given image is given. Turk and Pentland illustrate their method using a large database of 2500 face images of sixteen 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. Impressive recognition rates are reported. It is reported 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 eigenfaces approach 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 spatio-temporal 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.

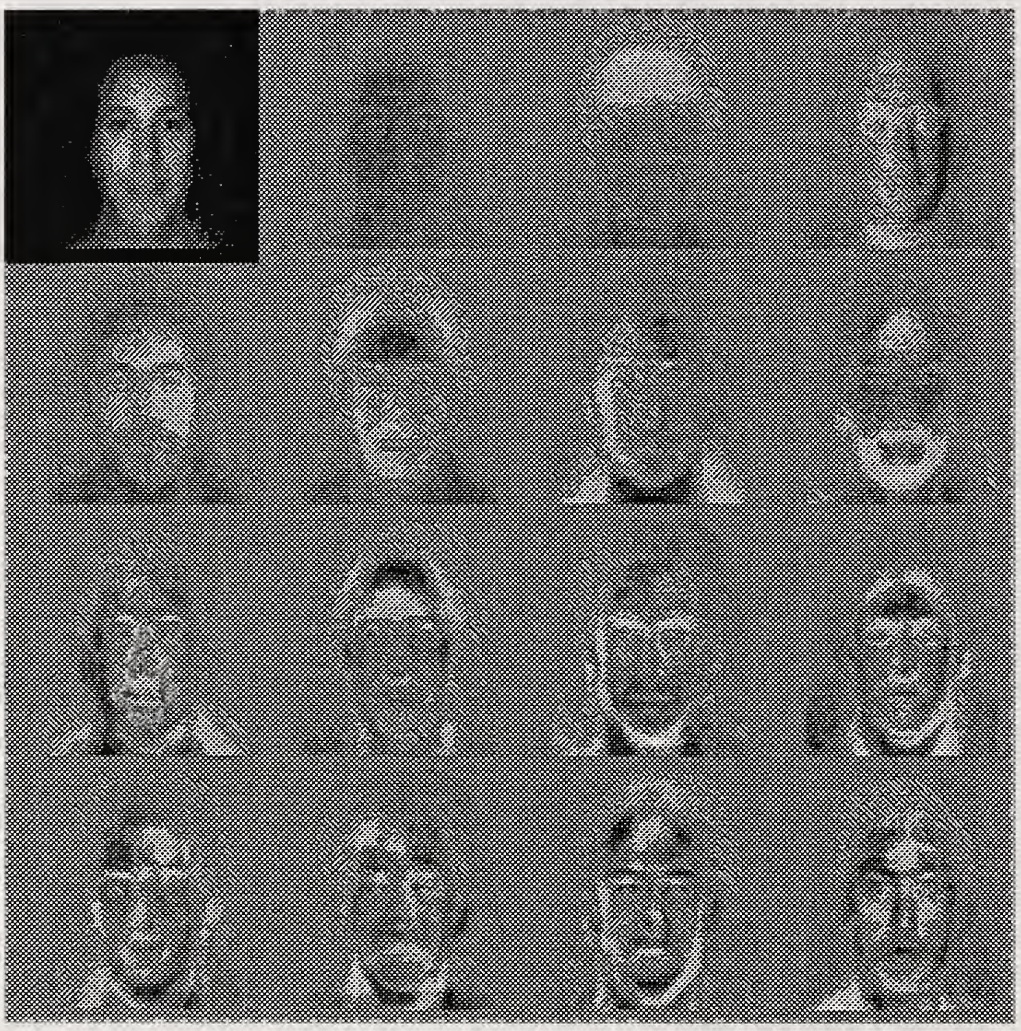


Figure 7: Eigenfaces [M. A. Turk and A. P. Pentland, "Face Recognition Using Eigenfaces," in *Proceedings, IEEE Computer Society Conference on Computer Vision and Pattern Recognition*, pp. 586–591, 1991. ©1991 IEEE]

In [9], the KL transform is used for the extraction of features from face images. 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. [9] 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 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 image (KL-IPAT). The KL-IPAT was extracted from 269 images with 100 eigenfaces. It greatly improves image recognition when compared with the eigenface approach using the KL on the raw image. The second step is to apply the Fourier Transform to the standardized image and use the resulting Fourier spectrum instead of the spatial domain of 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 image (KL-FSAT). The robustness of the KL-IPAT and KL-FSAT was checked against geometrical variations using the standard features for 269 face images. 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 missidentified 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 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 [9].

In [99], Pentland et al. extended the capabilities of their earlier system [124] 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 tested 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 mugshot 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, headwear, 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 thirty 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 mugshot applications, usually a frontal and a side view of a person are available. In some other applications, more than two views may be applicable. 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 mugshot 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 [124], 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 vs. 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 3, both wholistic 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 use of isodensity lines, i.e. curves of constant gray level, for face recognition has been investigated in [93]. Such lines, although they are not directly related to the 3-D 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 postprocessing 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. Good recognition accuracy was reported on this small data set.

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 [79]. 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 [4].

A single layer adaptive NN (one for each person in the database) for face recognition, expression analysis and face verification is reported in [119]. Named WISARD (Wilkie, Aleksander and Stonham's Recognition Device), 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 sixteen 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 [47], 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 $900 \times 40 \times 900$ 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 K-L 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.

Using a vector of sixteen numerical attributes (Figure 8) such as eyebrow thickness, widths of nose and mouth, six chin radii, etc., Brunelli and Poggio [18] also develop a NN approach for gender classification. They train two HyperBF networks [101], 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 16-dimensional 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 16-dimensional feature vector is used. The database consists of 21 males and 21 females. The leave-one-out strategy [41] 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 35-dimensional 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.

(a) (b)

Figure 8: Radii vectors and other feature points [R. Brunelli and T. Poggio, "Face Recognition: Features versus Templates," *IEEE Trans. Pattern Analysis and Machine Intelligence*, Vol. 15, pp. 1042–1052, 1993. ©1993 IEEE]

The use of HyperBF networks for face recognition is also reported in [18]. To remove variations due to changing viewpoint, the images are first transformed using 2-D 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. Reasonable results are reported. By feeding the outputs of sixteen HyperBFs to another HyperBF, significant reductions is error rates were reported.

[103] 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 backpropagation 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. In conclusion, 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 are not as important as the upper and lower portions and may be used to develop a focus of attention [103].

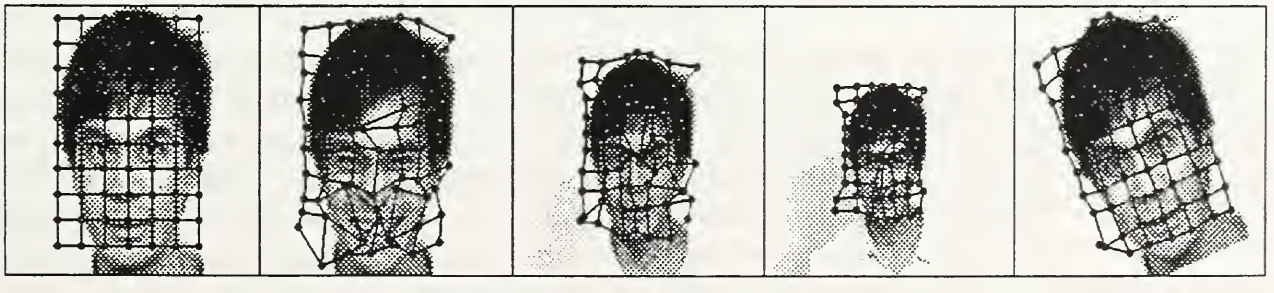


Figure 9: System for DLA

The systems presented in [20] and [80] are based on the Dynamic Link Architecture (DLA). DLAs 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 networks. DLAs 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 [20] and [80] 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 2-D array of nodes $A_x^I = (x, \alpha)$, where $\alpha = 1,...,F$. 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 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. The grid is appropriately positioned over the image to be stored and is stored with each grid point's locally determined jet and serves as the pattern classes. 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 [20] 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 [80] has a larger gallery of faces and recognizes them under different types of distortion and rotation in depth, achieving less then 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 [80, 20].

Manjunath et al. [83] 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

$$\mathbf{q_i} = [Q_i(x, y, \theta_1), \ldots, Q_i(x, y, \theta_N)] \tag{10}$$

corresponding to the i^{th} feature point. 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 Euclidian distance d_{ij} between two points V_i and V_j are said to be of consequence for the i^{th} 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 i^{th} feature point $\{V_i\}$ of $\mathcal{I}$. Search for the best feature point $\{V_{i'}\}$ in $\mathcal{O}$ using the criterion
$$S_{ii'} = 1 - \frac{\mathbf{q}_i \cdot \mathbf{q}_{i'}}{||\mathbf{q}_i||\,||\mathbf{q}_{i'}||} = \min_{m' \in N_i} S_{im'} \tag{11}$$

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
$$T_{ii'jj'} = 1 - \rho_{ii'jj'} \tag{12}$$

4. The total cost is computed as
$$C_1(\mathcal{I}, \mathcal{O}) = \sum_i S_{ii'} + \lambda_t \sum_i \sum_{j \in N_i} T_{ii'jj'} \tag{13}$$
where λ_t is a scaling parameter assigning relative importances 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
$$C(\mathcal{I}, \mathcal{O}^*) = \min_{\mathcal{O}'} C(\mathcal{I}, \mathcal{O}') \tag{14}$$

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. Figure 10 shows a set of input and identified images for this method.

Cheng et al. [27] 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 j^{th} face image of person i, then the average image for person i is given by $\frac{1}{N}\sum_{j=1}^{N} A_j^{(i)}$. 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.

Seibert and Waxman [107] 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 3-D objects [108] by combining 2-D views from different vantage points; in the case of faces, arrangement of features such as eyes and nose play the role of the 2-D 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 [109] exploit the role of caricatures and distinctiveness (summarized in Section 3 of the report) in human face recognition to enhance the capabilities of their previously reported face recognition system. Both of their papers are preliminary reports and lack experimentatal 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 [135]. Then each face is recognized by a neural network like structure called Cresceptron [130]. 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 extremely well.

4.4 Range Images

The discussion so far has revolved around face recognition methods and systems which use data obtained from a 2-D 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 involving mugshots as far as FBI datasets are concerned, 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. It is primarily for this reason that the work on this type of recognition has been included in this report. For further study, the reader is encouraged to read some of the selected papers presented in the bibliography at the end of this report.

Gordon in [48] 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 0.4mm. 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 Figure 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 [49], the process of finding the features is formalized for recognition purposes.

4.5 Forensic Applications

Starkey [118] 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 96 100% 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 Euclidian metric. The search area is 10×10. 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. Starkey concludes that neural networks will provide a viable solution to identifying criminals photographed at the crime scene, but this will not constitute an irrefutable evidence [118].

The initial forensic evidence is often a witness' recall of the culprit's appearance. Verbal descriptions of people's faces very often lack detail. Often, efforts to assist witnesses' recall

using composite techniques are unsuccessful. Face recognition, by comparison, is very good, and witnesses are usually called upon to examine mug-shot albums of known criminals. A study was designed to examine the efficiency of an experimental computerized mug-shot retrieval system and album search. Two factors which might be expected to affect retrieval were varied independently – distinctiveness of target face and position in the album. Distinctive faces are easier to remember than non-distinctive faces. Target faces occurring later in the album are believed to be more difficult to detect than those encountered earlier in the search. The FRAME prototype system with a data base of 1000 faces was used. 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. 38 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 3/4 view. Four non-distinctive and four distinctive faces were chosen from the set. Eight paid subjects were shown one of the 3/4 view test photographs for 10 sec., 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 non-distinctive faces 69%. The rate for the album searches was 78% for the distinctive faces and 44% for the non-distinctive faces [111].

In [81] Laughery et al. deal with the use of a computer in 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-file. A review of several systems developed to cope with what type of input to use in recognizing and retrieving a face from the gallery is presented first. The three prototype systems that are reviewed are characterized by varying methods of interaction and input: The "Whatsisface", "Sketch", and "Computerized Montage" systems. Laughery et al. also review the different interactive and automatic measurement algorithms for locating and measuring facial features. With the interactive system, facial images can be measured accurately with a precision which exceeds manual methods. Automatic measurement systems are not yet completely reliable. The basic approach to the mug-file 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 the case of line drawings, the quality of the image depends on the artist's ability and talent. The features selected for use in a pattern recognition system need to be discriminatory, easily obtained, stable over time, efficient and economic during data collection, storage and application. The precision and accuracy of measuring and encoding

a feature are important in terms of the useful information provided and the cost of using the feature. Precise and accurate measurement of a feature with less information content may be more cost-effective than imprecise measurement of a feature with more information content. Once a set of features has been measured for a large set of faces, statistical analyses can be performed to identify a subset of features which tend to be independent and summarize the information in the original set. The number of features is an important parameter in a pattern recognition system. Features may be described syntactically or geometrically. A geometric system of coding usually combines basic measurements into features summarizing the image information. In [76] ten feature distances are measured to code nine features with each feature being a distance divided by the nose length. Indications were found that 92% of the variation in the normalized data could be explained by five components. The components are considered high-level features when they summarize the information from the basic features. 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-file 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. The matching algorithm can use either syntactic data or geometric data. The computer is programmed to select matching features in a specified order. Algorithm design is concerned with the sequence in which selections are made and how to minimize errors in the face description. The matching algorithm [76] 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., 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. It is felt that either the matching or sequencing algorithms may be used to recognize and retrieve facial images [81].

In [55] Haig uses an image-processing computer to study the process of face recognition. The system advantages include the abilities to insert new targets into digital storage, to control and change the intensities both locally and globally, to move the targets around, to change their size and orientation, to present them for a wide range of fixed time intervals, to run experiments automatically, to collect the data, and to analyze the results. Haig's database consists of 100 target faces, taken under reasonably standardized conditions from the direct frontal aspect. The pictures are stored in 128×128 pixels, and registered such that the inter-pupilary distance is 30 pixels. The goal of Haig's 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. The observers scored 28.7% correct for the head, 24.3% for the eyes and mouth, and 22.7% for the nose. A changed head outline, while maintaining the inner features, very strongly influences the observer's responses. Fixing the head position, Haig tested the inner features. The results showed that the observers favor the eyes with 58.7%, mouths with 24.0% and noses with 17.3%. The response pattern demonstrates gratifying

consistency: Response to changed eyes is 80.3%, to changed mouths 13.8% and changed noses 1.3%. 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. Haig, in looking at the overall results, noticed 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. In the higher resolution distributed apertures experiment, it is determined that the resolution could be doubled in each orthogonal direction. This allows for a total of 162 usable apertures on each target. The median face area required for recognition in the lower aperture experiment is 5.9% and for the higher aperture experiment 4.0%. The reason for the lower threshold for the experiment is the greater variety of aperture positions and combinations [55].

4.6 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 law enforcement applications, the face recognition problem, by and large, should be viewed as a 2-D image matching and recognition problem with provisions for two or three views of a person's face. Given that in mugshots, 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 4.1. We are somewhat surprised that more attention has not been paid to the location/segmentation of a face in a given image. Segmentation becomes more difficult when beards or baldness are present; also if a face has to be segmented in a cluttered scene where other objects are present, the techniques presented in [51] and [114] 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 3, 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 eige-

nanalysis. From an aesthetic point of view, this is not satisfactory 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 wholistic 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. Studies alluded to in Section 3 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 arena the eigenfaces and eigenfeatures approach of Pentland and his students seems to be the most tested system, using several thousands of images. Although pattern recognition techniques based on eigenexpansion have been around for nearly thirty years, credit should be given to Pentland and his co-workers for pushing the technology towards face recognition applications. 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 mugshot 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 thousands of examples.

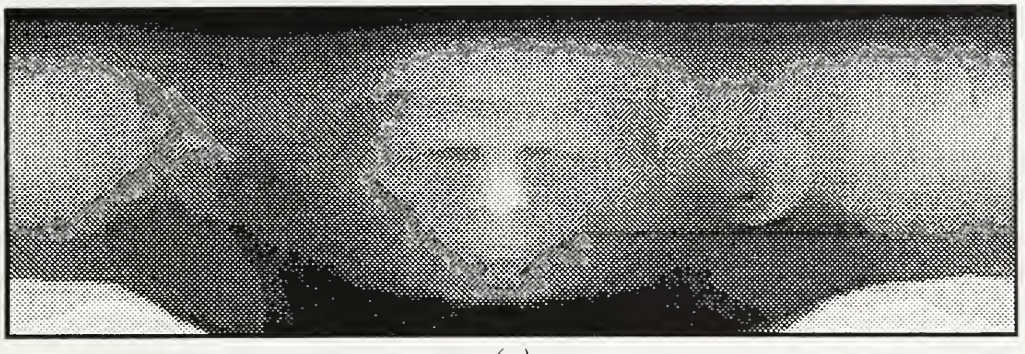
Techniques based on NN also deserve careful study. It is our view that NN-based methods can potentially incorporate both numerical and structural information relevant to face recognition; all of the existing work on NN approaches to FRT have demonstrated this on limited sets of images. In addition, the 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. Anyway, the usefulness of NN classifiers needs to be evaluated on significantly larger datasets than reported in the literature.

In the final stages of the mugshot matching problem, finer attributes of facial features are usually matched for identification purposes. The point features used in [83] 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.

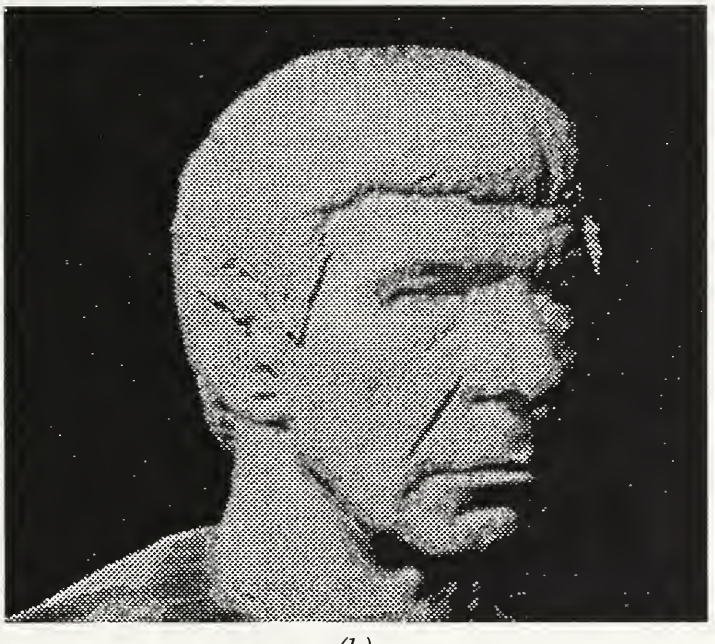
Owing to the cost of the equipment and the need for easy maintenance, intensity based systems are preferred in law enforcement applications. Although range information is richer than the 2-D intensity array, we feel that cost considerations will make range image based

techniques less attractive for field use.

Figure 10: Some results of the recognition system in Manjunath et al.



(a)



(b)

Figure 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. [G. Gordon, "Face Recognition Based on Depth Maps and Surface Curvature," in *SPIE Proceedings, Vol. 1570: Geometric Methods in Computer Vision*, pp. 234–247, 1991.]

5 Face Recognition from Profiles

Kaufman and Breeding [75] 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 eleminate 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. A procedure for creating the set of stored images is also described. Experiments were performed on a total of 120 profiles of ten persons half of which were used for training. A set of twelve 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 training feature vector was four. Comparisons with features derived from moment invariants [34] show that the circular autocorrelations performed better.

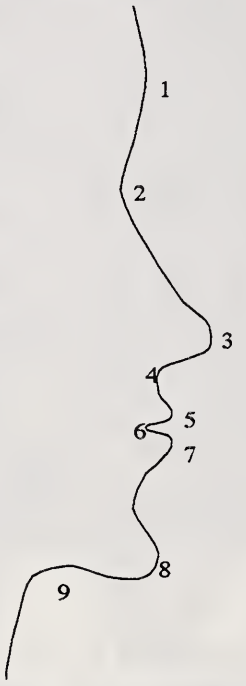


Figure 12: The nine fiducial points of interest for face recognition using profile images [similar to figure in L. Harmon and W. Hunt, "Automatic Recognition of Human Face Profiles," *Computer Graphic and Image Processing*, Vol. 6, pp. 135–156, 1977.]

Harmon and Hunt [57] 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 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 [57]. 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, a 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. [59] 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.

[59] used the feature extraction methods developed in [57] to create 11 feature vector components. The 11 features were reduced to 10, because nose protrusion is highly correlated with two other features. The 10-dimensional 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. The combination appeared to provide a robust system for identifying an unknown with the appropriate file face [59].

[58] is a continuation of the research done in [57] and [59]. 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 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 [58], 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 [58].

Wu and Huang [133] also report a profile-based recognition system using an approach similar to that of Harmon and his group [57], 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 seventeen of the eighteen 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 for recognizing closed contours. Using p-type FD's that can describe open as well as closed curves, Aibara et al. [7] 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.

5.1 Summary

Profile based recognition has not been pursed with as much vigor as face recognition. Given that mugshots have at least one side view, one could pursue a combination of the eigenapproach for sideviews of the face, as done in Pentland et al. [99] and a similar approach for the profiles. The p-type FD's, used in [7] 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 datasets such as mugshots.

6 Face Recognition from an Image Sequence

In surveillance applications, face recognition and identification from a video sequence is an important problem. Although over twenty years of active research on image sequence analysis is available [1, 2, 3, 125, 68], 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. 3-D models for faces
4. Non-rigid motion analysis

Current attempts [108, 124] 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. Also, 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 3-D 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, techniques such as motion stereo [87] 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 3-D 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 twenty years, hundreds of papers dealing with structure from motion have appeared in the literature. It is beyond the scope of this report to even include a brief summary of major techniques. We simply list books [125, 68, 64, 86, 129, 137] and review papers [90, 85, 91, 92, 6] here; papers that describe major approaches are listed in the additional bibliography.

3-D models of faces have been employed [82, 19, 8] in the model based compression literature by several research groups. Such models are very useful for applications such as witness face recontruction, reconstruction of faces from partial information and computerized aging. 3-D 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 non-rigid objects [121, 67, 98, 88]. There is emerging work in image sequence analysis dealing with non-rigid objects with emphasis on medical applications. We feel that some of the ongoing work on non rigid motion analysis will be useful in face recognition. An application of non-rigid motion to face recognition is reported in Yacoob and Davis's [134] 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 twenty five subjects displaying over 60 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, 3-D modeling of faces and non-rigid motion analysis potentially offer new solutions to various aspects of the face recognition and reconstruction problem. In the next subsection, we will briefly summerize the relevant literature on the problem of segmenting moving objects from an image sequence.

6.1 Segmentation

As noted earlier, thresholding of frame differences is one of the simplest methods for detecting moving objects. Several of the earlier papers [72, 73, 71, 91] 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 Figure 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 some form of optical flow field. Optical flow is the distribution of apparent velocities of brightness patterns in an image [66] and may arise from relative motion of objects and the viewer. Analysis of optical flow field is useful for estimating the motion 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 twenty five years, it has received significant attention in the computer vision literature only over the last fifteen 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 [123, 102]. Horn and Schunck [66] developed an iterative method for computing optical flow field using the regularization approach. Subsequently Anandan [11] 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 [38], Glazer [44], Heeger [62], Hildreth [64] and Fleet and Jepson [40]. Accurate computation of optical flow is still an unsolved problem leading researchers into computation of other flow fields such as image flow [127, 120, 113] and normal flow [10]. 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 [13].

Given that an accurate flow field is available, several techniques are available for detecting motion boundaries or clustering of flow fields. Adiv [5] 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

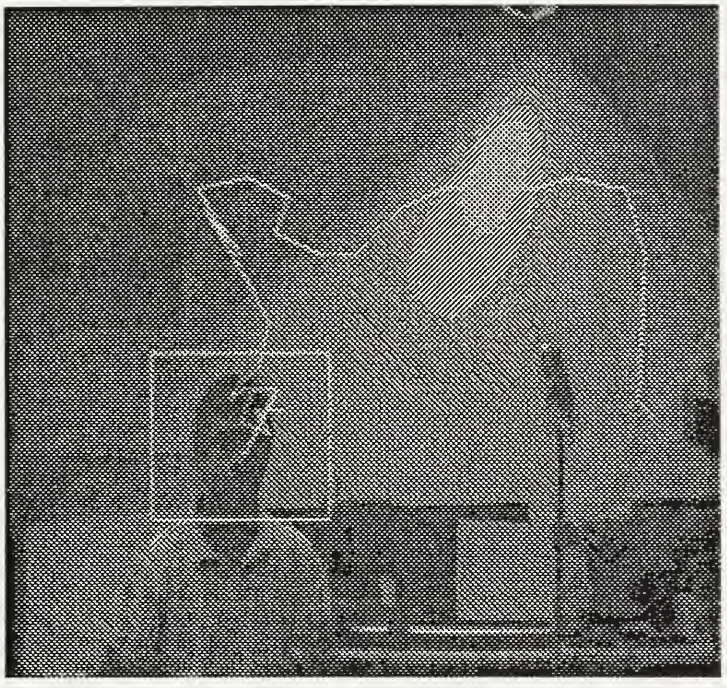


Figure 13: Locating the head from a video sequence applying the method of Pentland et al. [M. A. Turk and A. P. Pentland, "Face Recognition Using Eigenfaces," in *Proceedings, IEEE Computer Society Conference on Computer Vision and Pattern Recognition*, pp. 586–591, 1991. ©1991 IEEE]

is done by using the motion coherency of the planar surfaces. From the grouped 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 [122] 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 [89, 69, 63]. The central theme of this approach is to model the discontinuities using the "line processes" of Geman and Geman [42], 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 [42], maximum posterior marginal [84], and iterated conditional mode [15]. Implementation of these algorithms using analog VLSI hardware is addressed in [69, 78]. A recent paper [78] 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 [117, 22, 97, 32, 126, 138, 106, 132].

Two algorithms that use models of human motion for the purpose of segmentation are described in [96, 112]. An algorithm for the detection of moving persons from normal flows is also described in [94].

7 Evaluation of a Face Recognition System

In addition to the material contained in this report, previous work on the evaluation of OCR and fingerprint classification systems [54, 23] provides insights into how the evaluation of 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 law enforcement application.

7.1 Evaluation Requirements

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 trade-off 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 of existence of a matching face 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.

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 applications.

Based on this experience, we recommend that initial studies be carried out using realistic images from some specific law enforcement application. 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 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 FBI applications.

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 of the type which may be of interest to the FBI speed will be an important factor in evaluating applicability. NIST has a thousand-processor SIMD computer which has shown some capability in fingerprint classification and OCR applications. In the second Census OCR Systems Conference the time per form for the NIST system was 1 second/form where typical work-station times were 60 seconds/form. In addition to the SIMD computer, a 8-processor MIMD computer with 300 Mflop processors is being acquired to test other classes of parallel algorithms for speed and portability. In many potential law enforcement applications, parallel computers of this capacity 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.

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 effect this interface are dealt with in summary form in Section 2. 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. Until the speed and selectivity of possible machine-based recognition algorithms is known, the design and evaluation of human interfaces should be postponed.

7.2 Evaluation Methods

Database size and uniformity

As an initial evaluation sample we recommend collection of a minimum of 5000 and a maximum of 50,000 mug-shot images. 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. The minimum sample size for the test sets is based on the need to obtain accurate matching and false matching statistics. The 10 to 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.

Sample size issues - feasibility of resampling

We recommend that images be collected at relatively high resolution, 512 by 512, and using 8 bits 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.

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 report 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.

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

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.

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?

8 Summary and Conclusions

In this report, we have presented an extensive survey of face recognition technology as applicable to law enforcement applications. 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 report.

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

- Face recognition technology is an essential tool for law enforcement agencies' efforts to 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. Face recognition, in addition to fingerprint recognition, will remain a critical high-technology strategic research area with significant potential impact on reducing crime.

- Over thirty 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 mugshot books, and lineups. Even for applications 1-3, what is known about human recognition of faces could potentially 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 mugshots, drivers' licenses, personal ID's, and passport pictures, as in applications 2 and 3, segmentation in general is a non-trivial task. More effort needs to be directed to addressing the segmentation problem.

- Both global and local feature descriptions are useful. The most significant 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, 128, 53, 39] 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. We feel that 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 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 should

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 2-D 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 3-D structure from 2-D 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 mugshot 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.

In conclusion, face recognition technology is a vibrant topic, with significant potential applications for law enforcement agencies. Despite its ubiquitous usefulness, currently there is only one major research program (ARPA, Dr. Paul White). The results of this program, or a set of algorithms selected after further evaluation from those presented here, may be adequate for most law enforcement applications. Until further evaluations for specific applications are completed the most promising direction for future efforts will remain uncertain.

References

[1] "Proceedings of the DARPA Image Understanding Workshop, 1984–."

[2] "Proceedings of the IEEE Workshops on Motion, 1986, 1989, 1991."

[3] "IEEE Transactions on Pattern Analysis and Machine Intelligence," 1979–.

[4] Y. S. Abu-Mostafa and D. Psaltis, "Optical Neural Computers," *Scientific American*, Vol. 256, pp. 88–95, 1987.

[5] G. Adiv, "Determining Three-Dimansional Motion and Structure from Optical Flow Generated by Several Moving Objects," *IEEE Trans. Pattern Analysis and Machine Intelligence*, Vol. 7, pp. 525–542, 1985.

[6] J. K. Aggarwal and K. Nandhakumar, "On the Computation of Motion from Sequences of Images," in *Proc. IEEE,* Vol. 76, pp. 917–935, 1988.

[7] T. Aibara, K. Ohue, and Y. Matsuoka, "Human Face Recognition by P-type Fourier Descriptors," in *SPIE Proceedings, Vol. 1606: Visual Communications and Image Processing*, pp. 198–203, 1991.

[8] K. Aizawa and 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.), pp. 317–348, Boston, MA: Kluwer Academic Publishers, 1993.

[9] S. Akamatsu, T. Sasaki, H. Fukamachi, and Y. Suenaga, "A Robust Face Identification Scheme – KL Expansion of an Invariant Feature Space," in *SPIE Proceedings Vol. 1607: Intelligent Robots and Computer Vision X: Algorithms and Techniques*, pp. 71–84, 1991.

[10] Y. Aloimonos and et al., "Behavioral Visual Motion Analysis," in *Proceedings, Image Understanding Workshop*, pp. 521–541, 1992.

[11] 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.

[12] R. Baron, "Mechanisms of Human Facial Recognition," *International Journal of Man-Machine Studies*, Vol. 15, pp. 137–178, 1981.

[13] J. L. Barron, D. J. Fleet, and S. S. Beauchermin, "Performance of Optical Flow Techniques," *International Journal on Computer Vision*, Vol. 12, pp. 43–77, 1994.

[14] R. Bellman, *Introduction to Matrix Analysis*, New York: McGraw-Hill, 1960.

[15] J. Besag, "On the Statistical Analysis of Dirty Pictures," *Journal of Royal Statistical Society*, Vol. 48, pp. 259–302, 1986.

[16] V. Bruce, *Recognizing Faces*, London: Lawrence Erlbaum Associates, 1988.

[17] V. Bruce, "Perceiving and Recognizing Faces," *Mind and Language*, pp. 342–364, 1990.

[18] R. Brunelli and T. Poggio, "HyperBF Networks for Gender Classification," in *Proceedings, DARPA Image Understanding Workshop*, pp. 311–314, 1992.

[19] 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.), pp. 285–315, Boston, MA: Kluwer Academic Publishers, 1993.

[20] J. Buhmann, M. Lades, and C. v. d. Malsburg, "Size and Distortioin Invariant Object Recognition by Hierarchiacal Graph Matching," in *Proceedings, International Joint Conference on Neural Networks*, pp. 411–416, 1990.

[21] D. Burr, "A Dynamic Model for Image Registration," *Computer Graphics and Image Processing*, Vol. 15, pp. 102–112, 1981.

[22] 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.

[23] G. T. Candela and R. Chellappa, "Comparative Performance of Classification Methods for Fingerprints," tech. rep., National Institute of Standards and Technology, Gaithersburg, MD, 1993.

[24] J. Canny, "A Computational Approach to Edge Detection," *IEEE Trans. Pattern Analysis and Machine Intelligence*, Vol. 8, pp. 679–689, 1986.

[25] S. Carey, "A Case Study: Face Recognition," in *Explorations in the Biological Language* (E. Walker, ed.), pp. 175–201, New York: Bradford Books, 1987.

[26] S. Carey, R. Diamond, and B. Woods, "The Development of Face Recognition - A Maturational Component?," *Developmental Psychology*, Vol. 16, pp. 257–269, 1980.

[27] Y. Cheng, K. Liu, J. Yang, and H. Wang, "A Robust Algebraic Method for Human Face Recognition," in *Proceedings, 11th International Conference on Pattern Recognition*, pp. 221–224, 1992.

[28] 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 Proceedings, Vol. 1607: Intelligent Robots and Computer Vision X: Algorithms and Techniques*, pp. 85–95, 1991.

[29] M. J. Conlin, "A Ruled Based High Level Vision System," in *SPIE Proceedings, Vol. 726: Intelligent Robots and Computer Vision*, pp. 314–320, 1986.

[30] I. Craw, H. Ellis, and J. Lishman, "Automatic Extractrion of Face Features," *Pattern Recognition Letters*, Vol. 5, pp. 183–187, 1987.

[31] I. Craw, D. Tock, and A. Bennett, "Finding Face Features," *in Proceedings, Second European Conf. on Computer Vision*, pp. 92–96, 1992.

[32] T. Darrell and A. Pentland, "Robust Estimation of a Multi-Layered Motion Representation," in *Proceedings, IEEE Conference on Visual Motion*, pp. 173–178, 1991.

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

[34] S. Dudani, K. Breeding, and R. McGhee, "Aircraft Identification by Moment Invariants," *IEEE Trans. Computers*, Vol. 26, pp. 39–45, 1977.

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

[36] H. D. Ellis, "The Role of the Right Hemisphere in Face Perception," in *Function of the Right Cerebral Hemisphere* (A. W. Young, ed.), pp. 33–64, London: Academic Press, 1983.

[37] H. D. Ellis, "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.), pp. 3–13, Dordrecht: Nijhoff, 1986.

[38] W. Enklemann, "Investigations of Multigrid Algorithms for the Estimation of Optical Flow Fields in Image Sequences," *Computer Vision, Graphics and Image Processing*, Vol. 43, pp. 150–177, 1988.

[39] O. Faugeras, *Three-dimensional Computer Vision, A Geometric Viewpoint*, Boston: MIT Press, 1990.

[40] D. J. Fleet and A. D. Jepson, "Stability of Phase Information," *IEEE Trans. Pattern Analysis and Machine Intelligence*, Vol. 15, pp. 1253–1269, 1993.

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

[42] S. Geman and D. Geman, "Stochastic Relaxation, Gibbs Distribution and Bayesian Restoration of Images," *IEEE Trans. Pattern Analysis and Machine Intelligence*, Vol. 6, pp. 721–741, 1984.

[43] A. P. Ginsburg, "Visual Information Processing Based on Spatial Filters Constrained by Biological Data," *AMRL Technical Report*, pp. 78–129, 1978.

[44] F. Glazer, *Hierarchical Motion Detection*, Ph.D. dissertation, University of Massachusetts, Amherst, MA, 1987.

[45] A. Goldstein, "Facial Feature Variation: Anthropometric Data II," *Bulletin of the Psychonomic Society*, Vol. 13, No. 3, pp. 191–193, 1979.

[46] A. G. Goldstein, "Race-Related Variation of Facial Features: Anthropometric Data I," *Bulletin of the Psychonomic Society*, Vol. 13, pp. 187–190, 1979.

[47] B. A. Golomb and T. J. Sejnowski, "SEXNET: A Neural Network Identifies Sex From Human Faces," in *Adavances in Neural Information Processing Systems 3* (D. S. Touretzky and R. Lipmann, eds.), pp. 572–577, San Mateo, CA: Morgan Kaufmann, 1991.

[48] G. Gordon, "Face Recognition Based on Depth Maps and Surface Curvature," in *SPIE Proceedings, Vol. 1570: Geometric Methods in Computer Vision*, pp. 234–247, 1991.

[49] G. G. Gordon and L. Vincent, "Application of Morphology to Feature Extraction for Face Recognition," in *SPIE Proceedings, Vol. 1658: Nonlinear Image Processing*, 1992.

[50] A. Goshtasby, "Description and Discrimination of Planar Shapes Using Shape Matrices," *IEEE Trans. Patterns Analysis and Machine Intelligence*, Vol. 7, pp. 738–743, 1985.

[51] V. Govindaraju, S. N. Srihari, and D. B. Sher, "A Computational Model for Face Location," in *Proceedings, Third International Conference on Computer Vision*, pp. 718–721, 1990.

[52] U. Grenander, Y. Chow, and D. Keenan, "Hands: A Pattern Theoretic Study of Biological Shapes," in *Research Notes in Neural Computing*, New York: Springer-Verlag, 1991.

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

[54] P. J. Grother and G. T. Candela, "Comparison of Handprinted Digit Classifiers," tech. rep., National Institute of Standards and Technology, Gaithersburg, MD, 1993.

[55] 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.), pp. 410–425, Dordrecht: Nijhoff, 1986.

[56] P. W. Hallinan, "Recognizing Humam Eyes," in *SPIE Proceedings, Vol. 1570: Geometric Methods In Computer Vision*, pp. 214–226, 1991.

[57] L. Harmon and W. Hunt, "Automatic Recognition of Human Face Profiles," *Computer Graphic and Image Processing*, Vol. 6, pp. 135–156, 1977.

[58] L. Harmon, M. Khan, R. Lasch, and P. Raming, "Machine Identification of Human Faces," *Pattern Recognition*, Vol. 13, pp. 97–110, 1981.

[59] L. Harmon, S. Kuo, P. Ramig, and U. Raudkivi, "Identification of Human Face Profiles by Computer," *Pattern Recognition*, Vol. 10, pp. 301–312, 1978.

[60] L. D. Harmon, "The Recognition of Faces," *Scientific American*, Vol. 229, No. 5, pp. 71–82, 1973.

[61] D. C. Hay and A. W. Young, "The Human Face," in *Normality and Pathology in Cognitive Function* (A. W. Ellis, ed.), pp. 173–202, London: Academic Press, 1982.

[62] D. Heeger, "Optical Flow from Spatio-temporal Filters," *International Journal on Computer Vision*, Vol. 1, pp. 279–302, 1988.

[63] F. Heitz and P. Bouthemy, "Multimodal Motion Estimation and Segmentation Using Markov Random Fields," in *Proceedings, International Conference on Pattern Recognition*, pp. 378–383, 1990.

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

[65] Z. Hong, "Algebraic Feature Extraction of Image for Recognition," *Pattern Recognition*, Vol. 24, pp. 211–219, 1991.

[66] B. K. P. Horn and B. G. Schunck, "Determining Optical Flow," *Artificial Intelligence*, Vol. 17, pp. 185–203, 1981.

[67] T. S. Huang, "Modeling, Analysis and Visualization of Non-rigid Object Motion," in *Proceedings, International Conference on Pattern Recognition*, pp. 361–364, 1990.

[68] T. S. Huang and (ed), "Image Sequence Analysis Vol. 5." Springer Series in Information Sciences, New York:Springer-Verlag, 1981.

[69] J. Hutchinson, C. Koch, J. Luo, and C. Mead, "Computing Motion Using Analog and Binary Resistive Networks," *Computer*, Vol. 21, pp. 52–63, 1988.

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

[71] R. Jain, "Extraction of Motion Information from Peripheral Processes," *IEEE Trans. Pattern Analysis and Machine Intelligence*, Vol. 3, pp. 489–503, 1981.

[72] R. Jain, W. N. Martin, and J. K. Aggarwal, "Segmentation Through the Detection of Changes due to Motion," *Computer Graphics and Image Processing*, Vol. 11, pp. 13–34, 1979.

[73] R. Jain and H. H. Nagel, "On the Analysis of Accumulative Difference Pictures from Image Sequence of Real World scenes," *IEEE Trans. Pattern Analysis and Machine Intelligence*, Vol. 1, pp. 206–214, 1979.

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

[75] G. J. Kaufman, Jr., and K. J. Breeding, "The Automatic Recognition of Human Faces from Profile Silhouettes," *IEEE Trans. Systems, Man, Cybernetics*, Vol. 6, pp. 113–121, 1976.

[76] Y. Kaya and K. Kobayashi, "A Basic Study on Human Face Recognition," in *Frontiers of Pattern Recognition* (S. Watanabe, ed.), pp. 265–289, New York: Academic Press, 1972.

[77] M. Kirby and L. Sirovich, "Application of the Karhunen-Loeve Procedure for the Characterization of Human Faces," *IEEE Trans. Pattern Analysis and Machine Intelligence*, Vol. 12, 1990.

[78] 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 *Proceedings, IEEE Workshop on Visual Motion*, pp. 111–122, 1991.

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

[80] 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.

[81] 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.), pp. 250–269, New York: Academic Press, 1981.

[82] H. Li, P. Roivainen, and R. Forchheimer, "3-D Motion Estimation in Model-Based Facial Image Coding," *IEEE Trans. Pattern Analysis and Machine Intelligence*, Vol. 15, pp. 545–555, 1993.

[83] B. S. Manjunath, R. Chellappa, and C. v. d. Malsburg, "A Feature Based Approach to Face Recognition," in *Proceedings, IEEE Computer Society Conference on Computer Vision and Pattern Recognition*, pp. 373–378, 1992.

[84] J. Marroquin, S. Mitter, and T. Poggio, "Probabalistic Solutions for Ill-posed Problems in Computational Vision," *Journal of American Statistical Association*, Vol. 82, pp. 76–89, 1987.

[85] W. N. Martin and J. K. Aggarwal, "Dynamic Scene Analysis: a Survey," *Computer Vision, Graphics and Image Processing*, Vol. 7, pp. 356–374, 1978.

[86] W. N. Martin and J. K. Aggarwal, *Motion Understanding, Robot and Human Vision*, Boston, MA: Kluwer Academic Publishers, 1988.

[87] L. Matthies and T. Kanade, "Kalman Filter-based Algorithms for Estimating Depth from Image Sequences," *International Journal of Computer Vision*, Vol. 3, pp. 209–236, 1989.

[88] D. Metaxes and D. Terzopoulus, "Recursive Estimation of Shape and Nonrigid Motion," in *Proceedings, IEEE Workshop on Visual Motion*, pp. 396–311, 1991.

[89] D. W. Murray and B. F. Buxton, "Scene Segmentation from Visual Motion Using Global Optimization," *IEEE Trans. Pattern Analysis and Machine Intelligence*, Vol. 9, pp. 220–228, 1987.

[90] H. H. Nagel, "Analysis Techniques for Image Sequences," in *Proceedings, International Conference on Pattern Recognition*, pp. 186–211, 1978.

[91] H. H. Nagel, "Overview on Image Sequence Analysis," in *Vol. 5, Springer Series in Information Sciences* (T. S. Huang, ed.), pp. 19–228, New York: Springer-Verlag, 1981.

[92] H. H. Nagel, "Image Sequences-Ten (Octal) Years - From Phenomenology Towards a Theoretical Foundation," in *Proceedings, International Conference on Pattern Recognition*, pp. 1174–1185, 1986.

[93] O. Nakamura, S. Mathur, and T. Minami, "Identification of Human Faces Based on Isodensity Maps," *Pattern Recognition*, Vol. 24, pp. 263–272, 1991.

[94] R. Nelson, "Qualitative Detection of Motion by a Moving Observer," in *Proceedings, DARPA Image Understanding Workshop*, pp. 329–338, 1990.

[95] M. Nixon, "Eye Spacing Measurement for Facial Recognition," in *SPIE Proceedings Vol. 575*, pp. 279–285, 1985.

[96] O'Rourke and N. L. Badler, "Model-Based Image Analysis of Human Motion Using Constraint Propagation," *IEEE Trans. Pattern Analysis and Machine Intelligence*, Vol. 2, pp. 522–536, 1980.

[97] S. Peleg and H. Rom, "Motion Based Segmentation," in *Proceedings, International Conference on Pattern Recognition*, pp. 109–113, 1990.

[98] A. Pentland and B. Horowitz, "Recovery of Non-rigid Motion and Structure," *IEEE Trans. on Pattern Analysis and Machine Intelligence*, Vol. 13, pp. 730–742, 1991.

[99] A. Pentland, B. Moghaddam, T. Starner, and M. Turk, "View-based and Modular Eigenspaces for Face Recognition," in *Proceedings, IEEE Computer Society Conference on Computer Vision and Pattern Recognition*, 1994.

[100] D. Perkins, "A Definition of Caricature and Recognition," *Studies in the Anthropology of Visual Communication*, Vol. 2, pp. 1–24, 1975.

[101] T. Poggio and F. Girosi, "Networks for Approximation and Learning," *Proc. IEEE*, Vol. 78, pp. 1481–1497, 1990.

[102] T. Poggio and V. Torre, "Ill-posed Problems and Regularization Analysis in Early Vision," ai memo 773, M.I.T AI Lab, 1984.

[103] A. Rahardja, A. Sowmya, and W. Wilson, "A Neural Network Approach to Component Versus Holistic Recognition of Facial Expressions in Images," in *SPIE Prcoeedings Vol. 1607: Intelligent Robots and Computer Vision X: Algorithms and Techniques*, pp. 62–70, 1991.

[104] D. Reisfeld and Y. Yeshurun, "Robust Detection of Facial Features by Generalized Symmetry," in *Proceedings, 11th International Conference on Pattern Recognition*, pp. 117–120, 1992.

[105] T. Sakai, M. Nagao, and S. Fujibayashi, "Line Extraction and Pattern Recognition in a Photograph," *Pattern Recognition*, Vol. 1, pp. 233–248, 1969.

[106] B. G. Schunck, "Image Flow Segmentation and Estimation by Constraint Line Clustering," *IEEE Trans. Pattern Analysis and Machine Intelligence*, Vol. 11, pp. 1010–1027, 1989.

[107] M. Seibert and A. Waxman, "Recognizing Faces from their Parts," in *SPIE Proceedings, Vol. 1611: Sensor Fusion IV*, pp. 129–140, 1991.

[108] M. Seibert and A. M. Waxman, "Combining Evidence from Multiple Views of 3-D Objects," in *SPIE Proceedings, Vol 1611: Sensor Fusion IV: Control Paradigms and Data Structures*, 1991.

[109] M. Seibert and A. M. Waxman, "An Approach to Face Recognition Using Saliency Maps and Caricatures," in *Proceedings, World Congress on Neural Networks*, pp. 661–664, 1993.

[110] 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.

[111] 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.), pp. 398–409, Dordrecht: Nijhoff, 1986.

[112] A. Shio and J. Sklansky, "Segmentation of People in Motion," in *Proceedings, IEEE Workshop on Visual Motion*, pp. 325–332, 1991.

[113] A. Singh, "An Estimation-Theoretic Framework for Image Flow Analysis," in *Proceedings, International Conference on Computer Vision*, pp. 167–177, 1990.

[114] S. A. Sirohey, "Human Face Segmentation and Identification," tech. rep., Center for Automation Research, University of Maryland, College Park, MD, 1993.

[115] L. Sirovich and M. Kirby, "Low-dimensional Procedure for the Characterization of Human Face," *Journal of the Optical Society of America*, Vol. 4, pp. 519–524, 1987.

[116] H. L. Snyder, "Image Quality and Face Recognition on a Television Display," *Human Factors*, Vol. 16, pp. 300–307, 1974.

[117] A. Spoerri and S. Ullman, "The Early Detection of Motion Boundaries," in *Proceedings, First International Conference on Computer Vision*, pp. 209–218, 1987.

[118] R. B. Starkey and I. Aleksander, "Facial Recognition for Police Purpose Using Computer Graphics and Neural Networks," in *Proceedings, Colloquium on Electronic Images and Image Processing in Security and Forensic Science (Digest No. 087)*, pp. 2/1–2, 1990.

[119] 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.), pp. 426–441, Dordrecht: Nijhoff, 1984.

[120] M. Subbarao, "Interpretation of Image Flow: A Spatio-Temporal Approach," *IEEE Trans. Pattern Analysis and Machine Intelligence*, Vol. 11, pp. 266–278, 1989.

[121] D. Terzopoulus, A. Watkin, and M. Kass, "Constraints on Deformable Models: 3-D Shape on Nonrigid Motion," *Artificial Intelligence*, Vol. 36, pp. 91–123, 1988.

[122] W. B. Thompson and T. C. Pong, "Detecting Moving Objects," in *Proceedings, First International Conference on Computer Vision*, pp. 201–208, 1987.

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

[124] M. A. Turk and A. P. Pentland, "Face Recognition Using Eigenfaces," in *Proceedings, IEEE Computer Society Conference on Computer Vision and Pattern Recognition*, pp. 586–591, 1991.

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

[126] J. Y. A. Wang and E. H. Adelson, "Layered Representation for Motion Analysis," in *Proceedings, IEEE Computer Society Conference on Computer Vision and Pattern Recognition*, pp. 361–366, 1993.

[127] A. M. Waxman, B. Kamgar-Parsi, and M. Subbarao, "Closed-form Solutions to Image Flow Equations for 3-D Structure and Motion," *International Journal of Computer Vision*, Vol. 1, pp. 239–258, 1987.

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

[129] J. Weng, T. S. Huang, and N. Ahuja, "Motion and Structure from Image Sequences," in *Vol. 29, Springer Series in Information Sciences*, New York: Springer-Verlag, 1993.

[130] J. J. Weng, N. Ahuja, and T. S. Huang, "Learning Recognition and Segmentation of 3D Objects from 2D Images," in *Proc. IEEE Int. Conf. on Computer Vision*, pp. 121–128, 1993.

[131] P. A. Wintz, "Transform Picture Coding," *Proc. IEEE*, Vol. 60, pp. 809–820, 1972.

[132] K. Wohn and A. M. Waxman, "The Analytic Structure of Image Flows: Deformation and Segmentation," *Computer Vision Graphics and Image Processing*, Vol. 49, pp. 127–151, 1990.

[133] C. Wu and J. Huang, "Human Face Profile Recognition by Computer," *Pattern Recognition*, Vol. 23, pp. 255–259, 1990.

[134] Y. Yacoob and L. S. Davis, "Computing Spatio-Temporal Representations of Human Faces," in *Proceedings, IEEE Computer Society Conference on Computer Vision and Pattern Recognition*, 1994.

[135] G. Yang and T. S. Huang, "Human Face Detection in a Scene," in *Proceedings, IEEE Conf. on Computer Vision and Pattern Recognition*, pp. 453–458, 1993.

[136] A. Yuille, D. Cohen, and P. Hallinan, "Feature Extraction From Faces Using Deformable Templates," in *Proceedings, IEEE Computer Society Conference on Computer Vision and Pattern Recognition*, pp. 104–109, 1989.

[137] Z. Zhang and O. Faugeras, "3D Dynamic Scene Analysis," in *Vol. 27, Springer Series in Information Sciences* (T. S. Huang, ed.), New York: Springer-Verlag, 1992.

[138] Y. T. Zhou and R. Chellappa, "A Network for Motion Perception," in *Proceedings, International Conference on Neural Networks*, pp. 875–884, 1990.

Additional Bibliography

1 Feature Extraction

[1] M.K. Fleming and G.W. Cottrell. Categorization of faces using unsupervised feature extraction. In *IJCNN: International Joint Conference on Neural Networks*, pages 65–70, 1990.

[2] N.M. Marinovic and G. Eichmann. Feature extraction and pattern classification in space - spatial frequency domain. In *SPIE Vol. 579: Intelligent Robots and Computer Vision*, pages 19–26, 1985.

[3] L. de Floriani. Feature extraction from boundary models of three-dimensional objects. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 11(8):785–798, 1989.

2 Recognition

[4] V.S. Fayn. Identification of images. In *DDC*, 1986.

[5] 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.

[6] E.R. Brocklehurst. Computer methods of signature verification. *IEE Collquium Digest (80): Colloquium on MMI in Computer Security*, pages 3/1–5, 1986.

[7] L.D. Harmon and K.C. Knowlton. Picture processing by computer. *Science*, 164(3875):19–29, April 4 1969.

[8] L.D. Harmon. The recognition of faces. *Scientific American*, 229(5):71–82, 1973.

[9] A.J. Bilson. Image size and resolution in face recognition, University of Washington, 1987.

[10] P.L. Hawkes. The commercialization of biometric techniques for automatic personal identification. *IEE Collquium Digest (80): Colloquium on MMI in Computer Security*, pages 1/1–2, 1986.

[11] H. Mannaert and A. Oosterlinck. Self-organizing system for analysis and identification of human face. In *SPIE Vol. 1349 Applications of Digital Image Processing XIII*, pages 227–232, 1990.

[12] K. Flaton. 2d object recognition by adaptive feature extraction and dynamical link graph matching. In *Faculty of the Graduate School University of Southern California*, May 1992.

[13] Midorikawa. Face pattern identification by back propagation learning procedure. *Neural Networks*, 1(Suppl 1):515, 1988.

[14] M. Nixon. Automated facial recognition and its potential for security. In *IEE Collquium Digest (80): Colloquium on MMI in Computer Security*, pages 5/1–4, 1986.

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

[16] R. Brunelli and T. Poggio. Face recognition through geometrical features. In *ECCV*, 1992.

[17] D. Forsyth and A. Zisserman. Invariant descriptors for 3-d object recognition and pose. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 13(10):971–991, October 1991.

[18] T. Sakaguchi, O. Nakamura, and T. Minami. Personal identification through facial images using isodensity lines. In *SPIE Vol 1199: Visual Communication and Image Processing IV*, pages 643–654, 1989.

[19] I. Aleksander, W. Thomas, and P. Bowden. A step forward in image processing. *Sensor Review*, pages 120–124, July 1984.

[20] S. Sclaroff and A. Pentland. Closed-form solutions for physically-based shape modeling and recognition. In *Proceedings IEEE: Computer Society Conference on Computer Vision and Pattern Recognition*, pages 238–243, 1991.

[21] A.G. Goldstein and et.al. Recognition of human faces from isolated facial features: A developmental study. *Psychonomic Science*, 6(4):149–50, 1966.

[22] T.J. Stonham. *Practical Face Recognition and Verification with WISARD*, chapter In: Aspects of Face Processing; H.D. Ellis and M.A. Jeeves and F. Newcombe and A. Young, EDS, pages 426–441. Martinus Nijhoff Publishers, Dordrecht, 1986.

[23] K. Takahashi, T. Sakaguchi, T. Minami, and O. Nakamura. Description and matching of density variation for personal identification through facial images. In *SPIE Vol 1360: Visiual Communications and Image Processing*, pages 1694–1704, 1990.

[24] K.H. Wong, Hudson H.M. Law, and P.W.M. Tsang. A system for recognising human faces. In *IEEE Proceedings: International Conference on Acoustics, Speech, and Signal Processing*, pages 1638–1642, 1989.

[25] 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 Proceeding 1606: Visual Communications and Image Processing*, pages 204–216, 1991.

[26] S. Lele and J.T. Richtsmeier. Euclidean distance matrix analysis: A coordinate-free approach for comparing biological shapes using landmark data. *American Journal of Physical Anthropology*, 86:415–427, 1991.

3 Neurosciences

[27] S. Carey. *A Case Study: Face Recognition*, chapter In: Explorations in the Biology of Language, E. Walker Ed., Montgomery, Vermont, pages 175–201. Bradford Books, 1987.

[28] V. Bruce. *Recognising Faces.* Lawrence Erlbaum Associates, 1988.

[29] G. Rhodes. Lateralized process in face recognition. *British Journal of Psychology*, 76:249–271, 1985.

[30] H.D. Ellis. *Introduction: Processes Underlying Face Recognition*, chapter In: R. Bruyer, ED., The Neurophysiology of Face Perception and Facial Expression, pages 1–27. Lawrence Erlbaum, Hillsdale, NJ, 1986.

[31] J. Hochberg and R.E. Galper. Recognition of faces: I. an exploratory study. *Psychonomic Science*, 9(12):619–620, 1967.

[32] D.C. Hay and A.W. Young. *The Human Face*, chapter In: A.W. Ellis, ED., Normality and Pathology in Cognitive Function, pages 173–202. Academic Press, London, 1982.

[33] G. Davies, H. Ellis, and J. Shepherd. *Perceiving and Remembering Faces.* Perceptual and Moter Skills, 1965.

[34] S. Deutsch. Conjectures on mammalian neuron networks for visual pattern recognition. *IEEE Transactions on System Science and Cybernetics*, ssc-2(2):81–85, December 1966.

[35] P.D. McCormack and S.P. Colletta. Recognition memory for items from unilingual and bilingual lists. *Bulletin of the Psychonomic Society*, 6(2):149–151, 1975.

[36] M.D. McNeese. The boundaries of hemisperic processing in visual pattern recognition, U.S. Government Printing Office, 1987.

[37] H.D. Ellis. *The Role of the Right Hemisphere in Face Perception*, chapter In: A.W. Young, ED., Function of the Right Cerebral Hemisphere, pages 33–64. Academic Press, London, 1983.

[38] H. Ellis, J. Sheppard, and G. Davies. An investigation into the use of the photofit technique for recalling faces. *British Journal of Psychology*, 66:29–37, 1975.

[39] H. Ellis. Recognizing faces. *British Journal of Psychology*, 66:409–426, 1975.

[40] H.D. Ellis, M.A. Jeeves, F. Newcombe, H.D., and A. Young. *Aspects of Face Processing.* Mattinus Nijhoff Publishers, Dordrecht, 1986.

[41] H.D. Ellis. *Introduction to Aspects of Face Processing: Ten Questions in Need of Answers*, chapter Aspects of Face Processing; H.D. Ellis and M.A. Jeeves and F. Newcombe and H.D. and A. Young, EDS., pages 3–13. Mattinus Nijhoff Publishers, Dordrecht, 1986.

[42] S.L. Witryol, W.A. Kaess, and J.J. Danick. The kw memory for names and faces text, Department of Psychology University of Conneticut, September, 30 1957.

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

[44] J.F. Fagan III. Infants' recognition of invariant features of faces. *Child Development*, 47:627–638, 1976.

[45] M.P. Young and S. Yamane. Sparse population coding of faces in the inferotemporal cortex. *Science*, 256:1327–1331, May 29 1992.

[46] R.K. Yin. Looking at upside-down faces. *Journal Exp. Psychol.*, 81(1):141–145, 1969.

4 Profile Images

[47] G.F. Kaufman Jr. and K.J. Breeding. The automatic recognition of human faces from profile silhouettes. *IEEE Transactions on Systems, Man and Cybernetics*, smc-6(2), February 1976.

5 Range Data

[48] J.Y. Cartoux, J.T. Lapreste, and M. Richetin. Face authentification or recognitioin by profile extraction from range images. In *Proceedings-Workshop on Interpretation of 3D Scenes*, pages 194–199, 1990.

[49] P. Maragos. Tutorial on advances in morphological image processing and analysis. *Optical Engineering*, 26(7):623–632, 1987.

6 Related Issues

[50] A. Samal and P.A. Lyengar. Automatic recognition and analysis of human faces and facial expressions: A survey. *Pattern Recognition Society*, 25(1):65–77, 1992.

[51] T. Minami. Present state of the art on indentification of human face. *Journal Soc. Instrum. and Control Eng. (Japan)*, 25(8):707–713, 1986.

[52] M.A. Fischler and R.A.Elschlager. The representation and matching of pictorial structures. *IEEE Transaction of Computers*, c-22(1):67–92, January 1973.

[53] S. Carey and R. Diamond. From piecemeal to configurational representation of faces. *Science*, 195:312–314, January 1977.

[54] D.J. Burr. A dynamic model for image registration. *Computer Graphics and Image Processing*, 15:102–12, 1981.

[55] S. S. Culbert. *Object Recognition as a Function of Number of Different Views During Training.* Perceptual and Moter Skills, 1965.

[56] M.A. Jeeves. *Plenary Session. An Overview. Complementary Approaches to Common Problems in Face Recognition*, chapter In: Aspects of Face Processing; H.D. Ellis and M.A. Jeeves and F. Newcombe and A. Young; EDS., pages 445–452. Martinus Nijhoff Publishers, Dordrecht, 1986.

[57] J. Carter and M. Nixon. An integrated biometric database. In *Colloquim on Electronic Images and Image Processing in Security and Forsenic Science, Digest No. 87*, pages 8/1–6, 1986.

[58] T.Y. Tu, C. Ma, and Y.L. Ma. Image recognition by the kolmogorov complexity program. In *SPIE Vol 768: International Symposium on Pattern Recognition and Acoustical Imaging*, pages 343–350, 1987.

[59] P. O'Higgins and N.W. Williams. An investigation into the use of fourier coefficients in characterizing cranial shapes in primates. *Journal of Zoology, London*, 211:409–430, 1987.

[60] Z. De-Chen and Shen Xuan-Jing. Fast texture image segmentation. In *SPIE Proceedings 1349: Applications of Digital Image Processing XIII*, pages 277–282, 1990.

[61] C.A. Rothwell, A. Zisserman, C.I. Marinos, D.A. Forsyth, and J.L. Mundy. Relative motion and pose from arbitray plane curves. *Image and Vision Computing*, 10(4):250–262, May 1992.

[62] R.M. Haralick. A facet model for image data. *Computer Graphics and Image Processing*, 15:113–129, 1981.

[63] A. Pentland and B. Horowitz. Recovery of nonrigid motion and structure. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 13(7):730–742, 1991.

[64] M. Musen and J. Van Der Lei. *Of Brittleness and Bottelnecks: Challenges in the Creation of Pattern Recognition and Expert System Models*, chapter In: Pattern Recognition and Artificial Intelligence, E.S. Celsema and L.N. Kanal, EDS. North Holland, 1988.

[65] R.M. Haralick. Ridges and valleys on digital images. *Computer Graphics and Image Processing*, 22:28–38, 1983.

[66] Y.Q. Cheng, Y.M. Zhuang, and J.Y. Yang. Optimal fisher discriminant analysis using the rank decompostion. *Pattern Recognition*, 25(1):101–111, 1992.

[67] B.P. Yuchas, Jr. M.H. Goldstein, T.J. Sejnowski, and R.E. Jenkins. Neural network models of sensory intergration for improved vowel recognition. In *IEEE Proceedings*, pages 1658–1668, 1990.

7 Structure from Motion

[68] 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. Pattern Anal. Machine Intell.*, 6:547–554, 1984.

[69] J. W. Roach and J. K. Aggarwal, Determining the movementa of objects from a sequence of images, *IEEE Trans. Pattern Anal. Machine Intell.*, 2:554–562, 1980.

[70] R. Y. Tsai and T. S. Huang, Uniqueness and estimation of 3-D motion parameters of rigid bodies with curved surfaces, *IEEE Trans. Pattern Anal. Machine Intell.*, 6:13–27, 1984.

[71] J. Weng, T. S. Huang and N. Ahuja, Motion and structure from two prospective views: algorithms, error analysis and error estimation, *IEEE Trans. Pattern Anal. Machine Intell.*, 11:451–476, 1989.

[72] T. J. Broida and R. Chellappa, Estimation of object motion parameters from noisy images, *IEEE Trans. Pattern Anal. Machine Intell.*, 8:90–99, 1986.

[73] J. Aisbett, An iterated estimation of the motion parameters of a rigid body from noisy displacement vectors, *IEEE Trans. Pattern Anal. Machine Intell.*, 12:1092–1098, 1990.

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

[75] Y. Yasumoto and G. Medioni, Robust estimation of three-dimensional motion parameters from sequence of image frames using regularization, *IEEE Trans. Pattern Anal. Machine Intell.*, 8:464–471, 1986.

[76] C. L. Fennema and W. R. Thompson, Velocity determination in scenes containing several moving objects, *Computer Graphics and Image Processing*, 9:301–315, 1979.

[77] X. Zhuang, T. S. Huang and R. M. Haralick, Two-view motion analysis: a unified algorithm, *Journal of the Optical Society of America*, 3:1492–1500, 1986.

[78] G. S. Young and R. Chellappa, 3-D motion estimation using a sequence of noisy stereo images: models, estimation, and uniqueness results, *IEEE Trans. Pattern Anal. Machine Intell.*, 12:735–759, 1990.

[79] J. A. Webb and J. K. Aggarwal, Structure from motion of rigid and jointed objects, *Artificial Intelligence*, 19:107–130, 1982.

[80] M. Spetasakis and Y. Aloimonus, A multi-frame approach to visual motion perception, *International Journal of Computer Vision*, 6:245–255, 1991.

[81] Y. Kim and J. Aggarwal, Determining object motion in a sequence of stereo images, *IEEE J. RA*, 3:599–614, 1987.

[82] S. Liou and R. Jain, Motion detection in spatio-temporal space, *Computer Vision, Graphics and Image Processing*, 45:227–250, 1989.

[83] Z. Zhang, O. Faugeras and N. Ayache, Analysis of a sequence of stereo scenes containing multiple moving objects using rigidity constraints, in *Proceedings, Second International conference on Computer Vision*, pages 177–186, 1988.

[84] H. Chen and T. Huang, Matching 3-D line segments with applications to mutiple-object motion estimation, *IEEE Trans. Pattern Anal. Machine Intell.*, 12:1002–1008, 1990.

[85] S. Sethi and R. Jain, Finding trajectories of feature points in a monocular image sequence, *IEEE Trans. Pattern Anal. Machine Intell.*, 9:56–73, 1987.