Pattern Recognition, Vol. 25, No. 1, pp. 65–77, 1992
Printed in Great Britain

0031-3203/92 $5.00 + .00
Pergamon Press plc

AUTOMATIC RECOGNITION AND ANALYSIS OF HUMAN FACES AND FACIAL EXPRESSIONS: A SURVEY

ASHOK SAMAL and PRASANA A. IYENGAR
Department of Computer Science and Engineering, University of Nebraska–Lincoln, Lincoln, NE 68588-0115, U.S.A.

(*Received* 1 *June* 1990; *in revised form* 24 *May* 1991; *received for publication* 7 *June* 1991)

Abstract—Humans detect and identify faces in a scene with little or no effort. However, building an automated system that accomplishes this task is very difficult. There are several related subproblems: *detection* of a pattern as a face, *identification* of the face, *analysis* of facial expressions, and *classification* based on physical features of the face. A system that performs these operations will find many applications, e.g. criminal identification, authentication in secure systems, etc. Most of the work to date has been in identification. This paper surveys the past work in solving these problems. The capability of the human visual system with respect to these problems is also discussed. It is meant to serve as a guide for an automated system. Some new approaches to these problems are also briefly discussed.

Face detection   Face identification   Facial expressions   Classification   Facial features

1. INTRODUCTION

The face is a unique feature of human beings. Even the faces of "identical twins" differ in some respects. We locate, identify and distinguish between faces with very casual inspection. This is one of the reasons early postage stamps had human faces (of Queen Victoria) on them. It was assumed that it would be easier to detect forgeries. The uniqueness of faces is also the reason for their widespread use in applications where identification of people is important.

At the outset, one must distinguish between face identification and person identification. In face identification only the visual stimuli are used. In the case of person identification, however, other cues also play an important role. Some common cues are gait, posture, voice, clothing, etc. While the face and the facial features are widely used to identify persons, they are by no means the only means for identification. For example, in the last century, a system based on measurement of other body features was used in France to identify prisoners. Such features include head length, head breadth, foot length, and middle-digit lengths of the left foot and the left hand of a person. Recently, techniques have been devised to identify humans based on the structure of the genes.[1] In this paper we are only concerned with the problem of face recognition.

A system to correctly locate and identify human faces would be useful in several applications. It can be used for identification of criminals, in much the same way as it is done now. However, the process can be made considerably faster and more robust since it will involve only a minimal role for the humans. It can also be used for authentication in secure systems. Examples of such systems include computer systems and bank teller machines. Virtually all facilities that allow access based on the identification of the person fall into this category. Many systems currently use such items as cards, code numbers, and/or passwords to allow access. A system based on face recognition will be much more robust and much less inconvenient. Some other applications in which face recognition has been used include assistance in speech recognition,[2] in systems for better man–machine interaction[3] and in visual communication over telephone and other low bandwidth lines.[4]

The goal of this paper is to survey the work done in the automatic recognition and analysis of human faces. Some aspects of the problem that have not been seriously studied, e.g. analysis of facial expressions, are also discussed here. Section 2 identifies the five basic problems that must be addressed. The inherent difficulties in solving these problems are also discussed. In Section 3 the capability of the human visual system with respect to these problems is described. It, in some ways, defines the expectations for an automatic system. In Section 4 the past work in these problems is surveyed in detail and then summarized. Some possible directions of research are discussed in Section 5. Finally, the paper concludes with a summary and directions for future research in Section 6.

2. BASIC PROBLEMS IN FACE RECOGNITION

Our main goal is to explore the issues in the design and implementation of a system that can be used for human face recognition. Before a face can be identified it must first be located in a scene. Thus the first problem is to detect the presence of a face and locate it. The next problem is to devise mechanisms to identify the face on the basis of known information. A related problem is to decide how the face must be represented, e.g. a 2D intensity array or a set of feature values. There are two more interesting problems that are related to human faces and the information content of the faces: analysis of the facial expressions and classification based on physical features. These problems are elaborated next.

2.1. *Representation of faces*

Representation of faces is important in a face recognition system. All the known faces must be stored in some format. The new (detected) faces must also be represented appropriately so that the matching procedure for identification can be used efficiently. A face is difficult to describe completely and accurately in terms of features alone, e.g. long nose, dark hair and round face. This is the reason that police departments use face reconstruction artists to derive a complete face from the descriptions of eye witnesses, who describe the face as a set of features. Attempts must be made to keep the representation compact without losing any (or much) information. For example, if one decides to use an intensity image (say 512×512), 6 bits per pixel (compared to 8 bpp) may be sufficient. Similarly, images smaller than 512×512 may be sufficient both for detection and identification. The nature of the representation affects and is affected by the matching scheme used in identification.

2.2. *Detection of faces*

In some instances the condition under which the image (photograph) is obtained is controlled, e.g. the mugshots taken by the police. Hence, the location of the face in the scene can be easily determined. In most cases, however, the location of a face in the image is not known a priori. The first step, therefore, is to determine if the scene has any faces. If a face is present, its location in the image must also be determined.

Several factors make this problem more complex. The first is the presence of facial hair, make-up, etc., which obscure the facial features. Another problem is variation in scale and orientation of the face in the image. Faces occur in images at many different sizes and angles. Thus, it is difficult to search for a fixed pattern (template) in the image. Usually, the faces in the image are either frontal (camera focused on the front of the face) or in profile (camera pointing towards one side of the face). Presence of noise and occlusion also make the problem harder.

2.3. *Identification of faces*

The next step is to identify the faces detected in the image. The known faces, and possibly their features, must be stored in a database. The goal here is to match the face in the image with the known faces in the database. One can use either a wholistic approach to matching or an approach based on the features. The set of features used in a system based on face profiles is very different from one used in a system based on frontal images. However, the central problem in all cases is *matching*. Some additional complications arise due to the change in a face and its features over time.

2.4. *Analysis of facial expressions*

The last two problems are not necessarily components of a face recognition unit, but are probably very desirable. The first is the analysis of facial expressions in the face. For example, one would like to determine, based on the information in the face, if the expression is one of happiness, sorrow, or surprise. While humans do it with relatively little effort, it is a very complex problem and has so far not been seriously addressed. The problem here is to model human emotions and correlate them with the facial features and their changes.

2.5. *Classification based on physical features*

Another operation humans do rather effortlessly when they look at faces is to classify them as male or female. Other classifications are based on the age of the person, race, expected profession, etc. It would be useful for a face recognition system to be able to perform these classifications as well. This step may need a large amount of knowledge and is also very complex. The problem is to derive a model, which may be empirical, to describe the classes as functions of facial features.

3. CAPABILITIES OF A HUMAN VISUAL SYSTEM

Before an automated face recognition and analysis system is built, one must first decide on its functionalities. A good reference point is the human visual system. After all, it is the best known face recognition system. The capability of the human visual system with respect to the problems mentioned in the last section is discussed here. It may not be possible to incorporate all features of the human system into an automatic system, and some features may even be undesirable. However, it can certainly serve as the reference point.

3.1. *Representation*

The first step in recognition of a face is likely to be the structural encoding of the visual stimulus (the face). The encoding may be in the form of a primal sketch and may be hardwired.[5] There is some evidence to suggest that the internal representation of known faces is different from unknown or unfamiliar faces. However, not much else is known in terms of the nature of internal representations, e.g. the size of the representation.

3.2. *Detection*

Humans detect a face pattern by casual inspection of the scene even if the face is unfamiliar. It indicates that either there are special purpose units to detect faces (there is currently very little evidence to support this),[5] or this task is done very efficiently, probably using massively parallel computation. Also, the face is perceived as a whole (face Gestalt) as opposed to a collection of features that constitutes a face. For example, when a face is partially occluded, we perceive a whole face, as if our perceptual system fills in the missing parts.

We also detect faces effortlessly in a wide range of conditions, e.g. in bad lighting conditions, from great distances, and when they are partially occluded. Our face detection system/mechanism is very robust in this regard. There is also a strong perceptual bias towards seeing patterns as faces. We often "see" faces in clouds and flames. Furthermore, there is a hysteresis effect, i.e. once we recognize a pattern as a face, it is difficult not to see it as one afterwards.

We detect some faces better by squinting or seeing from a distance. These processes essentially blur the image, which is equivalent to filtering out high spatial frequencies. The details of sharp features, e.g. eyeballs and eyelashes, become less prominent. The presence of features and their geometrical relationships with each other appears to be more important.

One of the issues in detection is the minimum size for a pattern to be recognized as a face. Figure 1 shows a face (of Albert Einstein) at different spatial resolutions. A face can be detected rather easily in the 32×32 image. However, in the 16×16 image there is little resemblance to a face. It is generally believed to be the lower limit for detection.[7] Others have suggested even lower limits (100–200 bits per face).[8] Note that these limits are for detecting a pattern as a face and not for identification.

A related issue is the minimum gray scale resolution necessary for face detection. Figure 2 shows the same face (of Albert Einstein) at different gray level resolutions (bits per pixel). In all cases, the presence of a face is immediately noticed. It suggests that for detection (not identification) only 1 bpp images may be sufficient if the pattern has a good spatial resolution. There is experimental evidence[9] to suggest that $32 \times 32 \times 4$ bpp is sufficient for detection.

3.3. *Identification*

After a pattern has been perceived as a face, the next step is to identify it. Identification is the process of association of a name to the face. Our primary concern is the identification based only on the face image. Correct identification means different images of the same face are identified as one. It has been estimated that one knows about 700 people personally and thousands in general.[5] We are also capable of identifying a large number of strange faces with relatively brief exposures. Typically, however, one sees hundreds of thousands of faces over one's lifetime.

The identification of known faces or faces of famous people occurs almost instantaneously. However, for unfamiliar faces and familiar faces at unusual orientations (e.g. upside-down) the time taken for identification is usually longer. The human identification system is also very robust. We identify faces under a wide range of lighting conditions, under occlusion, and across a range of transformations. It has been suggested[5] that for every known face there exists a unit that registers its familiarity under all conditions. More familiar faces need less evidence for identification, compared to the less familiar ones. Consequently, familiar faces are less likely to be falsely recognized. Bartlett[10] postulated that all new faces have the same (zero) level of familiarity and new typical faces produce a stronger impression of familiarity than unusual faces. Increase in the degree of familiarity after a single exposure to a face is higher for unusual faces than typical faces.

Despite our ability to identify faces rather effortlessly, it is difficult to describe a human face. The most common method is to enumerate different facial features, e.g. the hair is blond, the eyes are green, the nose is flat. The features play a very important role in the identification process. Some of the important features are eyes, nose, lips, ears and hair. The spatial relationships of the features are also important. However, the face is not merely a sum or a collection of the internal features. The identification of faces involves synthesis of the facial features.

It has been suggested[7] that the saliency of the facial features decreases from top to bottom. It would imply that the role of features such as hair and eyes is more important in the identification process than features like chin and lips. However, faces with unusual features, e.g. a very prominent chin, are always easily identified.

Since features play an important role in the identification process, one question that has been asked is, "How many features are necessary to discriminate faces?" It has been estimated[7] that the number of features needed grows logarithmically with the number of faces. It means that for a person who can identify 1000 faces, the number of discriminating features is ten, which is reasonably low. However,

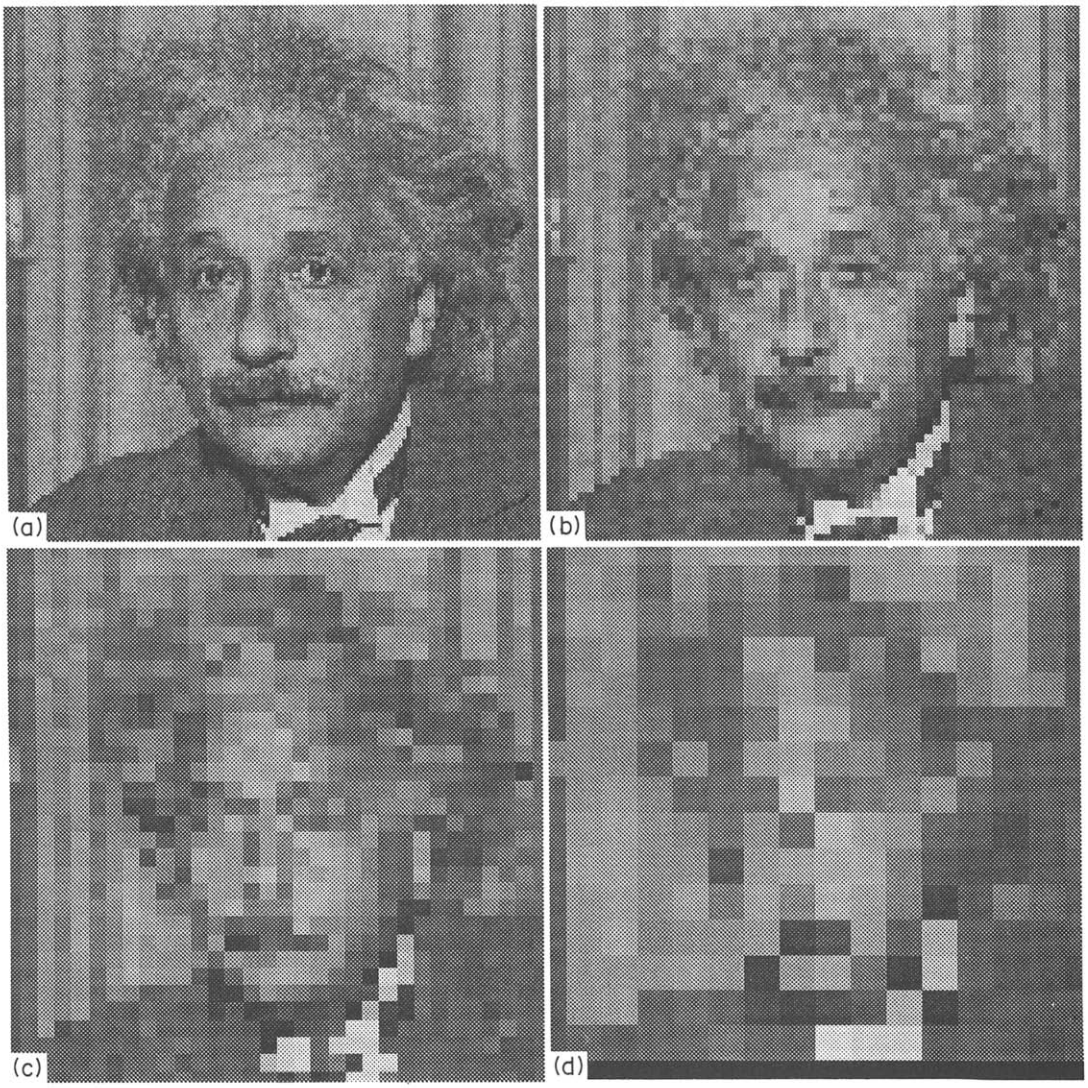

Fig. 1. A human face at different spatial resolutions: (a) 128 × 128; (b) 64 × 64; (c) 32 × 32; (d) 16 × 16. All images are enlarged to the same size.

for automatic systems which may deal with much larger sets of faces, it may not be as efficient.

Another aspect that has not been explored in great detail is the minimum spatial and gray level resolutions necessary to identify known faces. Experiments[9] show that $32 \times 32 \times 4$ bpp may be sufficient for identification.

3.4. *Analysis of facial expressions*

While our mechanisms for detecting and identifying faces are very robust, the same is not the case for analyzing facial expressions. It is often very difficult to determine the exact nature of the expression on a person's face. The recognition depends on several factors, e.g. familiarity with the face or person, general experience with different types of expressions, attention given to the face, and non-visual cues. Not a great deal is known, for certain, about how we accomplish this difficult task. It is generally believed to be an early perceptual process and is separate from those involved in feature detection.[5]

3.5. *Classification based on physical features*

Usually, when we see faces we associate some labels to them, e.g. young, male or infant. Classification can be based on many aspects. We are mainly concerned with classification based on sex: male or female; age: infant, young, middle-aged or old; race: caucasian, negro or oriental. We perform some classification with apparent ease and without many errors, e.g. the male/female classification. However, other classifications are more error prone. In this

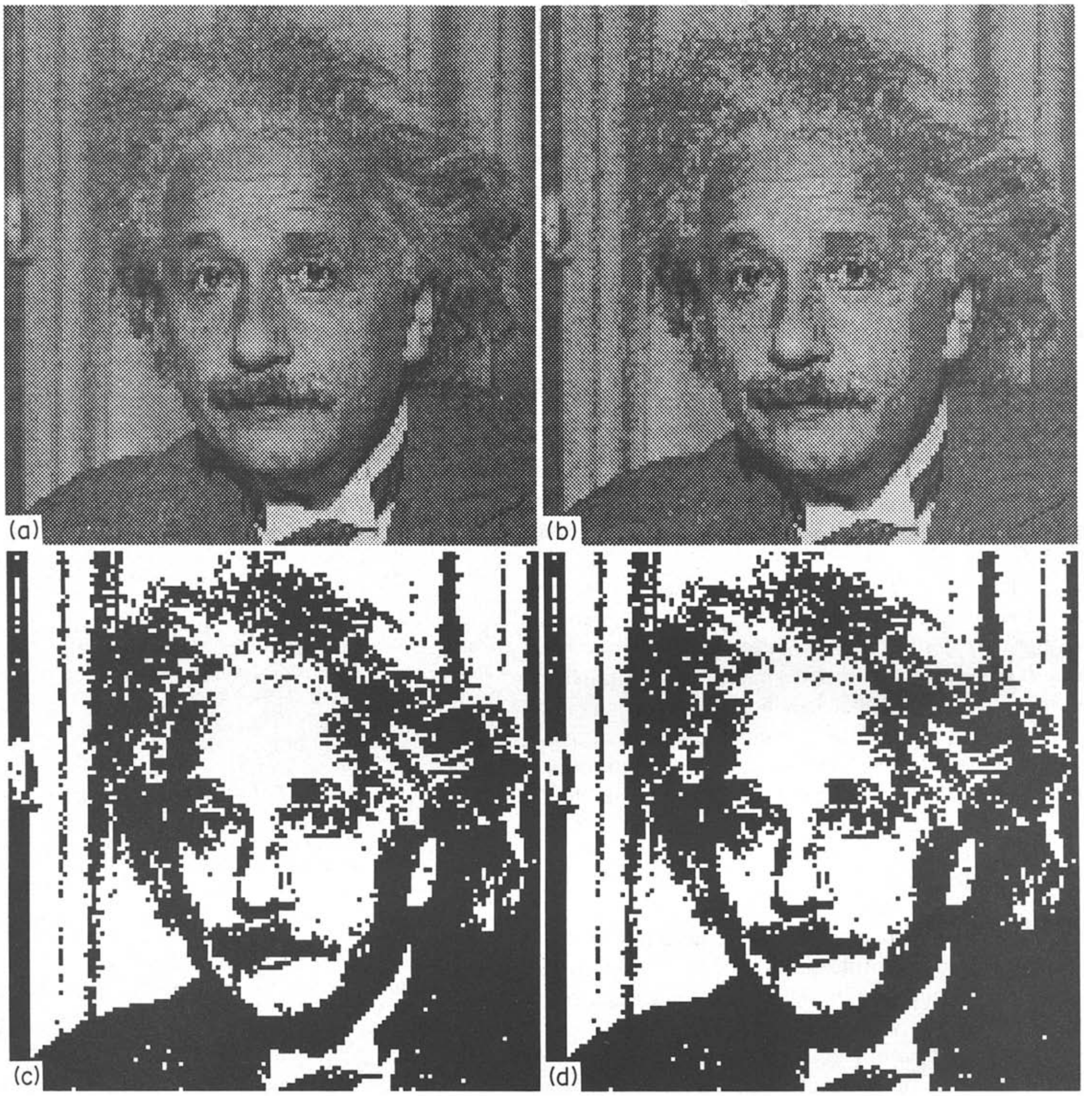

Fig. 2. A human face at different gray level resolutions: (a) 8 bpp; (b) 4 bpp; (c) 2 bpp; (d) 1 bpp.

case one may need a lot of experience or knowledge to get a reasonably correct classification. The process of classification based on physical features is believed to be a higher level (cognitive) process performed after the face has been identified. The nature of these processes, however, is generally not well understood.

4. AUTOMATIC FACE RECOGNITION TECHNIQUES

It is interesting to note that work in face recognition dates back to the last century when Sir Francis Galton worked on this and some related problems. Some other early works (late 1960s) include references (11 and 12). In this section, the work done in computer recognition of human faces is surveyed in detail. The survey is divided into five parts, based on the problems discussed in Section 2: representation, recognition, identification, expression analysis and physical analysis.

It must be mentioned at the outset that the problems have not been solved in the general case. Only the problem of identification (and to some extent detection) has received serious attention. In identification research, the location of the face is usually known a priori or is derived by using simple heuristics. In addition, in many instances strong assumptions are made to make the problem more tractable. Some common assumptions are listed below:

(1) the image has the face either in the front view or in profile. So, one is certain to find all the common features in the face;
(2) the face is upright with no or very little tilt;
(3) there is no occlusion;
(4) the face has no facial hair, glasses, scars, etc.;
(5) most of the test cases are white males;
(6) the number of test cases is relatively small;

Table 1. Input data specifications for different approaches

Reference	Number	Size	Bpp	Comments
Baron[13]	150	512 × 480	6	Reduced to 128 × 120
Buhr[14]	100	512 × 512	8	—
Campbell[8]	50	256 × 256	6	—
Craw *et al.*[15]	1000	128 × 128	—	Tested on only 20 faces
Goldstein *et al.*[16]	255	—	—	Profile drawn by artist
Harmon *et al.*[17]	124	—	—	(Artist); 97 males + 27 female
Harmon *et al.*[18]	130	—	—	81 males + 49 females
Kaufman and Breeding[19]	120	240 × 356	1	—
Nixon[20]	6	128 × 128	8	—
Petajan[2]	—	244 × 248	—	Used for locating lips only
Riccia and Iserles[21]	506	—	—	Data input by user
Wong *et al.*[22]	6	256 × 256	8	—
Wu and Huang[23]	18	—	—	—

hundreds at the most, but fewer in general (see Table 1).

Table 1 summarizes the input data format used by different researchers. It shows the spatial ("size") and gray scale ("bpp") resolutions and the number of test faces used ("number") in the work. A missing entry either means that it is not reported or is not applicable. For example, sometimes the features data were derived by hand and then hand fed to the computer (see the "comments" entry in the table).

4.1. *Representation*

Two types of representations are used in the face recognition and identification research: 2D intensity image and feature vector. They are briefly described in this section.

The simplest approach to representing a face is to use a 2D array of intensity values. It is clearly not a very compact representation, but is useful when robustness in recognition is a primary goal. Systems that use other representations often keep arrays of intensity values, although they may not be explicitly used. In large systems, with perhaps thousands of faces, such a representation is not suitable. However, with appropriate compaction, e.g. reducing spatial and gray scale resolutions, using various coding schemes, this approach can still be practical. There is evidence[9] that 32 × 32 images with 4 bpp is sufficient both for detection and identification. So, a single face can be represented with only 512 bytes. Storage of thousands of face images would then require only a few megabytes of space.

The most widely used representation for faces is the feature vector. Two types of features are commonly used: (a) features derived from intensity images, and (b) features derived from face profiles. Examples of features in the first category are hair intensity, size of eyes, distance between the eyes, distance between eyes and lips, etc. In profile images, the feature values are obtained using a set of characteristic points on the profile, e.g. the notch between the brow and nose, the tip of nose, the notch between the nose and upper lip, and the tip of chin. The features are usually distances and angles between these characteristic points. See Tables 2 and 3 for a complete list of the features used in both approaches.

Table 2. Features used for identifying faces (using profile data)

Category	Features
Nose	Nose protrusion (base line to tip of nose)
Chin	Chin protrusion (base line to tip of chin)
Angles	1–2–3, 1–3–4, 4–1–5
Area	Right of base line 2–3 Right of line 6–10
Ratios	$D(1,4)/D(1,2)$
Distances	$D(1,2)$, $D(1,4)$, $D(1,5)$, $D(1,8)$ $D(3,4)$, $D(4,2)$, $D(5,10)$, $D(7,9)$
Others	Wiggliness, correlation coefficients, moment invariants

Notes: 1. $D(x,y)$ denotes the distance between features x and y.
2. For the characteristic points see Fig. 4(a).
3. Base line is the line drawn from chin (2) to forehead (3).
4. Profile triangle is the triangle formed by nose (1), chin (2) and forehead (3).
5. Wiggliness is the average of the inverse of radius of curvature of the points on the profile.

Sometimes a combination of features and intensity data is used to represent faces. Campbell[24] used several features, e.g. hair intensity and cheek intensity. In addition, a 32 × 32 pixel matix is used to represent the eyes. Another representation is proposed in reference (4) for visual communications at low data rates. A special operator, called the valley operator, is used to derive a facial image consisting of edges, which is then used for transmission.

Yet another approach is to represent a face by a single number[25] or a set of numbers.[26,27] In the late nineteenth century Galton devised a system in which portraits are described by numerical formulae. Each formula consists of four groups of figures, with five figures in each group. The shapes of the profile

Table 3. Features used for identifying faces (using intensity data)

Category	Features
Overall	height, outline, template, area, centroid
Eyes	template, shape, color/shade, area, opening, center
	intensity around pupil, D (upper eyelid, eyelid-fold)
	difference of ordinates of inner and outer corners
	inter-eye distance (center), inter-eye distance (outer corner)
Eyelids	area, length of the bounding curves
Eyebrows	thickness, separation (D (eye, brow))
Lips	thickness, shape, template, width
Mouth	template, width, height, area, area of opening
Nose	width, length, template, shape, area of nostrils
Ears	length, shape, area
Hair	intensity/shade/color, amount/coverage, length, texture
Cheeks	intensity
Distances	width of head on the eye-line, D (chin, eye-line), D (center of lips, chin)
	D (eye, hairline), D (eye, center of nose), D (tip of chin, center of face)
	D (left edge, center of nose), D (right edge, center of nose)
	D (inner corner of eye, center of face), D (center of eye, center of brow)
	D (center of face, center of brow), D (inner corner of eye, corner of mouth)
	D (center of mouth, center of eye), D (center of mouth, end of nose)
Areas ($\triangle$s)	inner corners of eyes and center of mouth
	center of eyelid fold, center of eye and center of face
	center of mouth and centers of brows
Ratios	D (center of face, opening of mouth)/D (chin, opening of mouth)
	D (chin, opening of mouth)/D (end of nose, opening of mouth)
Others	chin outline, width and height of philtrum
	forehead size, forehead outline

Notes: 1. $D(x,y)$ denotes the distance between features x and y.
2. There are two distances and two triangles involving eyes and eyebrows: one for left and one for right. Similar comments also apply for lips (upper and lower).

Fig. 3. Some shapes for the ridge of the nose (used by Galton).

at and immediately adjacent to some characteristic points, and those of the intermediate links, are expressed as single numerals. They are determined from a table in which different shapes are represented with an index. No more than 9–10 varieties are given for each feature. Figure 3 shows some possible shapes for the ridge of the nose. Each of them is assigned a particular number (0–10). The formula thus derived can also be used to reconstruct the profile. Another approach given by Galton[28] can be used to represent any irregular curve. Since it is difficult to describe an irregular curve, it is better described by the two bounding limits. Thus, any curve can be represented using the two (possibly regular) bounding curves.

4.2. *Detection of faces*

The first step in the identification of a face in the image is to locate it. However, for most works in identification, the location of a face in the image is either known a priori or can easily be derived based on the constraints in the imaging process, e.g. mugshots.

The face detection problem has been approached in two ways. In the first approach, the face is determined as a whole unit, usually using model-based vision techniques. Govindaraju *et al.*[29,30] have given a scheme to locate human faces in newspaper photographs. The face is represented by a very simple model comprising two straight lines (for the two sides of the face) and two arcs (for the head and chin). The size and location of the face is not known, but is determined from the caption based on heuristics from photojournalism. For example, if the caption says that there are three faces, the maximum size of each face is constrained. The knowledge in the caption is represented by a semantic network and is used in the reasoning process. A modified Hough transform is used to detect lines and arcs. To determine whether a face is present, matching is done using the structural model of the face and the lines and arcs in the image. The faces must be upright, unoccluded, and provide a front view for the system to work correctly. Craw *et al.*[15] also use the whole face in the matching procedure, but they use a template matching scheme to detect the outline of the face. The template is used at different resolutions to handle scaling. Neural networks can also be used for identifying faces.[31,32]

In the second approach, a face is located by first locating some important features. Once the features are identified, the overall location of the face is determined using the geometric information, i.e. the relative location of the features. While any feature in the face can be located first, the eyes are most commonly used. Yulie *et al.*[33] have given an approach to detect eyes in images using deformable templates. An energy function is defined with edges, peaks and valleys in the image intensity and the template as the parameters. The template interacts with the image by modifying the parameter values to minimize the energy function (the same as getting the best fit). The template is flexible enough to handle changes in the size and orientation. So, it can locate eyes despite variations in scale, tilt and lighting conditions. This approach was used to detect both eyes and lips. The eye template consists of a circle (pupil), two parabolic sections (the bounding contour) and two points (centers of whites of the eyes). This scheme is very flexible, but there are two main problems. The coefficients of the terms in the energy function are determined by experimentation and are difficult to generalize. Secondly, the time taken to get a match is likely to be high (the authors report times of several minutes).

Nixon[20] described a method to locate eyes using the Hough transform. The iris is modeled using a circle and a tailored exponential function is used to model the boundary. In reference (15) the eyes are located by looking for a pair of minima (centers of the eyes) in intensity below the eyebrows, which are located first. Another approach for locating eyes is given in reference (14). The system uses only pictures with front faces with no glasses or beard, with homogeneous illumination of the face and a homogeneous background. After a model-based segmentation is performed, the system localizes the eyes. The basis of the scheme is that eyes have a finer texture compared to their surroundings. Consequently, the edge image has many more edges around the eyes. In particular, there are two distributions: one corresponding to the eyebrows and a bigger one corresponding to the eyes themselves. Thus a projection of the edge image along the Y-axis will have two peaks corresponding to the two distributions. From this information, the eye locations are determined using symmetry and moments of the segmented regions.

The lips are also commonly used in feature based schemes. A method to detect lips is given in reference (15). After the outline of a face is determined, a vertical line is traced in the lower one-fourth of the face. Line finders are released at each pixel on this line to look for horizontal and vertical lines. If they are found and can be fitted into a thin box a lip is identified. Eyebrows are located in the same way.

4.3. *Identification of faces*

Identification is the second and the most important operation in a face recognition system. There are two aspects that are related to identification. Firstly, a name has to be associated with the face. Secondly, multiple instances, possibly with different sizes and orientations, of the same face must be identified as one. Since it can be directly used in many applications, this problem has received the most attention. In fact, it was first studied by Galton in the last century.[27,28,34] Based on the input data, the approaches can be classified into two major categories: identification using profiles of the faces and identification using intensity images.

Face profiles are more widely used than intensity images. Although the starting points for the two approaches differ, the underlying principles are similar. A typical sequence of steps in the identification procedure is as follows:

(1) determine a set of independent features to represent a face;

(2) represent the known faces in terms of their features in a database;

(3) determine the feature values of the new (unidentified) face;

(4) use a matching scheme to obtain a "best" fit with the known faces.

The approaches differ in the choice of features and the matching procedure. In most cases, the number of known faces in the database is less than a few hundred (see Table 1). The number of features varies from 10 to 40.

The earliest work related to identification was done by Galton[27,28] and is based on face profiles. A set of five cardinal points (see Fig. 4) was derived from the face profile and features derived from it were used to compare faces. The cardinal points chosen are: the notch between the brow and the nose, the tip of the nose, the notch between the nose and the upper lip, parting of the lips and the tip of the chin. Galton also invented a mechanical device to perform the comparison. Given a collection of sets of features (X) and a new set of features (Y), the goal was to determine the set $Z \in X$ that resembles Y within a certain specified degree of likeness. No two corresponding features in the matching set Z and the test set Y should differ by more than a certain specified tolerance.

Harmon *et al.*[17,18,35] also used a similar approach to identify faces. A set of fiducial marks was used to represent the profiles (see Fig. 4). The fiducial points correspond to the cardinal points in Galton's work. Nine fiducial marks are identified: chin tip, nose, forehead, bridge, nose bottom, throat, upper lip, mouth, and lower lip. From the fiducial marks a set of features was derived. The distances between the fiducials, angles between them, and areas of some triangles formed by the fiducials are used as the features. Degree of similarity is measured by a normalized Euclidean distance between the two sets of features, X and Y:

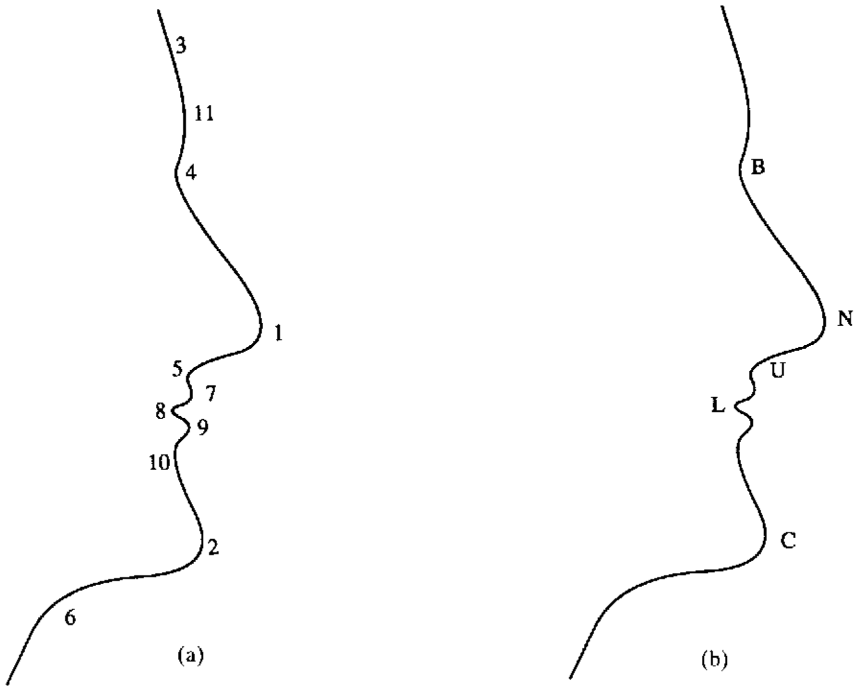

Fig. 4. Characteristic points used by: (a) Harmon; (b) Galton.

$$d^2(X, Y) = \sum_{i=1}^{n} \frac{X_i^2 - Y_i^2}{\sigma_i^2},$$

where n is the number of features and σ^2 is the variance. Profiles (of 256 faces) were drawn by artists but the fiducials were automatically derived. For example, to derive the lips and mouth fiducials, the following algorithm was used:

(1) start with the nose bottom fiducial and trace downwards;

(2) get the radius of curvature at each point on the profile curve;

(3) get a pattern of three minima, with the curvatures pointing to inside, outside and inside. These correspond to the upper lip, mouth and the lower lip, respectively.

In their later work[17] they added more test faces including some faces of females.[3] The same nine fiducial marks were used from which an 11D feature vector was derived. The features consisted of nose protrusion, area of the profile, wiggliness of the profile, and various distances: $D_{1,2}$, $D_{1,4}$, $D_{1,5}$, $D_{1,6}$, $D_{1,8}$ and $D_{7,9}$ where $D_{i,j}$ is the distance between the ith and jth fiducials. To perform matching, a set partitioning technique was also experimented with. The scheme is based on a very simple observation. The values of features of a face do not change very much in different images. In addition, if a feature is very different in two faces, the corresponding instances in images will also be very different. Therefore, if the difference between any two corresponding components of the feature vectors of the unknown face and a known face is large, the two faces must be different. These two facts guide the set partitioning approach to identification. The authors concluded that a 10D feature vector was sufficient for a robust system. A combination of set partitioning and Euclidean distance was found to be the most effective scheme. In reference (18) two more fiducial marks and some new features were used. The faces were scanned directly, using a camera and a very well contrasted background. The profiles were then automatically obtained.

Kaufman and Breeding[19] designed and implemented a system to identify faces from profiles. They used a feature based approach, where the features were coefficients computed from the polar form of the autocorrelation function. They also experimented with moment invariants as features. Baylou *et al.*[36] also used face profiles to identify faces. The number of characteristic points in their scheme was ten. Lapreste *et al.*[37] used range data to obtain face profiles. The images are acquired using a range finder. The characteristic points are then determined in the profile. Finally, the Euclidean distance is used to find a match. Wu and Huang[23] use a B-spline to extract some (six) interesting points from the face profile. A set of 24 features is derived from them which are then used for matching. Lee and Milios[38] use range images to determine correspondence between similar features of two faces.

Several attempts have been made to identify faces using features derived from intensity images. Goldstein *et al.*[16] used an approach in which the faces were obtained by an artist and the values of the features were determined by a set of jurors. A set of 34 features was initially used. It was later reduced to 22 since not all features were mutually independent, e.g. upper and lower lip thickness. Each feature was represented in a scale of 1–5 (low–medium–high). The training set consisted of 255 faces. The features measured different aspects of hair (e.g. shade, texture), eyebrows, eyes, nose, mouth, ears, cheeks and forehead. After the data were input, the system computed the (Euclidean) distance between the unidentified face and each face in the database. The face with the smallest distance was considered to be the match. The results showed that only 6–7 features were usually sufficient to identify most faces. Kaya and Kobayashi[39] also used a similar system, but used distances between the features of the face as the features.

Buhr[14] also used features derived from intensity images to identify faces. Thirty-three primary and 12 secondary features were used. Among the primary features are 21 distances (e.g. distance between center of eyes and center of mouth), 4 differences of ordinates, 4 areas of triangles, 2 ratios of distances and 2 special features (areas of the two eyes). A linear decision tree is used to determine the best match for the unknown face. Wong *et al.*[22] also used a similar approach. In addition to the Euclidean distance, they used the difference of the sum of the parameters in the matching process. Campbell *et al.*[24] used a least square fit to get the best match. Riccia[21] used clustering techniques for the same purpose.

Baron[13] used standardized images for identification. Firstly, the eyes are located by correlating with 20×23 eye templates. The images are normalized by keeping the inter-eye distance constant. The image is then reduced to 15×16 and is correlated with every face in the database. Only the faces with high correlation values are used for verification. A set of templates is kept for each feature in the face and used for this purpose. The templates are matched with the corresponding parts of the face image. A match is found if three-quarters of the features correlate with a high degree of similarity.

Another interesting approach to recognize faces was given by Preston.[40] He used an optical computer for this purpose. Photographs of six kings were used to get a matched filter, which acts as the memory for the optical computer. When one of the six pictures is presented as the input, the output produces a bright spot and the position of the spot indicates which king was used as the input. The basic mechanism used here is correlation. Taylor[41] used a network of photomultipliers to discriminate between ten photographs following 250 presentations for each photograph.

4.3.1. *Choice of features.* A system based on features must select an effective set of features for identification. A wide range of features have been used for face recognition. The features used to identify profiles are different from those used for intensity images. Tables 2 and 3 list the features used in systems using intensity data and profile data, respectively.

4.3.2. *Matching schemes.* Several different approaches are used for the final matching step. These are mostly traditional approaches and hence are not described in detail here. However, they are briefly summarized below.

(1) *Euclidean distance*: this is the most commonly used measure. The goal is to compute the distance between the feature vector of the test face and that of each face in the database. The face with the smallest distance from the unknown face is the match.

(2) *Clustering*: the clustering techniques may be used to find the degree of similarity between the unknown face and each known face.

(3) *Set partitioning*: the feature vectors for different faces usually differ significantly in at least one feature. The known faces can be reduced into a very small subset by rejecting all faces that differ significantly in one feature value. It can be repeated along several features to get a single match.

(4) *Correlation*: face templates are used to compute the similarity between the unknown face and a known face.

(5) *Others*: sometimes a combination of these methods are used. Another measure used to find the similarity between two faces is the difference between the sums of the features. Linear decision trees and least square fit are also sometimes used.

4.4. *Facial features and expression analysis*

Research in the analysis of facial expressions has not been actively pursued. Schlosberg[42] gave a basis for describing emotions using three dimensions: A–R (attention–rejection), P–U (pleasantness–unpleasantness) and level of activation. For example, emotions such as contempt and disgust have a high value of rejection while anger has a high degree of unpleasantness. Ekman and Friesen[43] have identified six primary expressions: happiness, sadness, fear, disgust, surprise and anger. They have also systematically catalogued thousands of different expressions. Attempts have also been made in computer graphics (animation) to model the human facial expressions from a muscular point of view.[44, 45]

Suwa *et al.*[46] did a preliminary investigation on automatically analyzing the expressions from a series of movie frames. They approached the problem from a muscular point of view. Twenty spots in the face are identified and their movements are tracked in the sequence of video frames. They are then compared with the movements from the prototype patterns for different expressions to determine the expression on the face. Terzopoulos and Waters[47] have used a simple form of the models given by Ekman and Friesen to derive a 3D computer model to display human facial expressions and also to analyze the expressions from a video sequence. A model (physically-based synthetic tissue model) motivated by tissue biomechanics is used to model the face.

5. DISCUSSION AND SOME NEW APPROACHES

In this section, some new approaches are proposed for these problems. We are currently working on some of them and do not have any concrete results at this point. So, the details of these approaches are not presented here.

5.1. *Detection*

Perhaps the simplest way to detect a face in an image is to use a scheme based on correlation with a face template. It may, at first glance, sound computationally very unattractive. There are two mitigating factors. Firstly, this operation is massively parallel and hence can be performed rather quickly (in real time) using a multicomputer, e.g. the Connection Machine. Secondly, it is not necessary to use a large template. As indicated in Section 3 there is enough information in smaller images (32×32) to detect the presence of a face.

Another approach to detect faces is to extend the scheme given in reference (33). The extension is to model the whole face directly in 3D (Yulie *et al.* modeled only some features). This approach is computationally expensive, but with multiprocessing it may be tractable. Yet another approach is to use neural nets to detect presence of faces in an image.

The previous approaches are wholistic approaches, i.e. they determine the location of a face directly. One can also use a feature based approach. However, any such scheme must put proper emphasis on the nature and the robustness of detecting the features. Accordingly, the eyes should play a more prominent role than the ears, since they are less likely to be occluded. The importance of the features in human recognition apparently decreases from top to bottom. A system centered around detecting features in the upper half of the face is likely to be more robust. Eyes play a very important role and should perhaps be central to any detection system. Other features may then be used as secondary features.

5.2. *Identification*

Most of the work in identification has been done using face profiles. While it is easier to use profiles, it is likely that a system based on intensity images will be more robust and convenient. The features used in the two approaches are different, but the basic problems are the same. Many of the features used with profiles are somewhat artificial and are probably not used in a human visual system, e.g. area of the triangle formed by the nose tip, the forehead and the chin. Profile data are usually obtained under very controlled conditions. Using intensity images is very general and provides greater flexibility in the matching process.

One of the simpler ways to design a robust identification system is to use some form of template matching. Again, it may have been impractical in the past, but with massive parallelism it may not be such an expensive solution. Features may also be used in the identification process. One could also use the template matching procedure in this case with the features (not the whole face) stored in a database. Features with strong perceptual basis may be proved to be more useful for automatic identification.

5.3. *Facial features and expression analysis*

These two analyses are the most difficult among the set of problems discussed in this paper. The fundamental problem is to model the different aspects of the problems. For example, one must devise a model to represent human emotion. What is it in a face that makes the face look angry? An initial study may involve a large amount of empirical data to derive the appropriate parameters (e.g. reference (43)). Any system for expression analysis must also incorporate some form of learning.

The problem of classification based on physical features may not be very difficult along some dimensions. For example, classifying faces according to race may be possible using fairly simple features, e.g. average intensity, overall boundary shape, shape of eyes, nose and lips. However, it is not clear how to achieve some other classifications, e.g. male/female. Again, a system which is capable of learning is essential.

6. SUMMARY

The recognition of faces and analysis of expressions are intriguing problems which humans solve with apparent ease. We have identified the five different but related aspects of the problem: representation, detection, identification, expression analysis and classification based on physical features. Capability of a human visual system in solving these problems is also discussed. It should only serve as a reference for any computer vision system attempting to achieve the same functionality.

The past work done in these problems is surveyed in detail and then summarized. Some of the work dates back to the late nineteenth century. Among the problems, identification has been the most widely studied, due to its direct applicability in some practical problems. Most of the work, however, is based on face profiles. The last two problems have not been seriously attempted in computer vision research, due to their complexity. Some new ideas to solve these problems are also briefly discussed. One issue not discussed here, but important in many applications, is the speed of recognition. Many applications, e.g. bank teller machines, require close to real-time performance. Hence, the speed of the system must also be considered. All the problems are very interesting and none has been solved in the general case. Therefore, they should be of interest to researchers for some time. We are also currently working on all aspects of the problem and have some interesting preliminary results.

REFERENCES

1. P. J. Neufeld and N. Colman, When science takes the witness stand, *Scient. Am.* **262**, 46–53 (May 1990).
2. E. D. Petajan, Automatic lipreading to enhance speech recognition, *Proc. CVPR*, pp. 40–47 (1985).
3. K. Mase, Y. Suenaga and T. Akimoto, Head reader—a head motion understanding system for better man-machine interaction, *Proc.* 1987 *IEEE Int. Conf. Syst. Man Cybern.*, pp. 970–974 (1987).
4. D. E. Pearson and J. A. Robinson, Visual communication at very low data rates, *Proc. IEEE* **73**, pp. 795–812 (April 1985).
5. H. D. Ellis, Processes underlying face recognition, *The Neuropsychology of Face Perception and Facial Expression*, R. Bruyer, ed., Chapter 1, pp. 1–27. Lawrence Erlbaum Associates, New Jersey (1986).
6. D. C. Hay and A. W. Young, The human face, *Normality and Pathology in Cognitive Functions*, A. W. Ellis, ed., Chapter 6, pp. 173–203. Academic Press, New York (1982).
7. L. D. Harmon, The recognition of faces, *Scient. Am.* **229**, 71–82 (October 1973).
8. F. W. Campbell, How much of the information falling on the retina reaches the visual cortex and how much is stored in the visual memory? Seminar at the Pontificae Academiae Scientiarium Scripta Varia (1983).
9. A. Samal, Minimum resolution for human face detection and identification, *Proc. SPIE/SPSE Symp. Electronic Imaging* (January 1991).
10. J. C. Bartlett, S. Hurry and W. Thorley, Typicality and familiarity of faces, *Memory Cognition* **12**, 219–228 (1984).
11. W. W. Bledsoe, Man machine facial recognition, Technical Report PRI 22, Panoramic Research Inc. (August 1966).
12. M. D. Kelley, Visual identification of people by computer, Ph.D. thesis, Department of Computer Science, Stanford University (1970).
13. R. J. Baron, Mechanisms of human facial recognition, *Int. J. Man–Mach. Stud.* **15**, 137–178 (1981).
14. R. Buhr, Analyse und klassifikation von gesichtsbildern, *ntzArchiv* **8**, 245–256 (1986).
15. I. Craw, H. Ellis and J. R. Lishman, Automatic extraction of face-features. *Pattern Recognition Lett.* **5**, 183–187 (1987).
16. A. J. Goldstein, L. D. Harmon and A. B. Lesk, Identification of human faces, *Proc. IEEE* **59**, pp. 748–760 (May 1971).
17. L. D. Harmon, S. C. Kuo, P. F. Ramig and U. Raudkivi, Identification of human face profiles by computer, *Pattern Recognition* **10**, 301–312 (1978).
18. L. D. Harmon, M. K. Khan, R. Lasch and P. F. Ramig, Machine identification of human faces, *Pattern Recognition* **13**, 97–110 (1981).
19. G. J. Kaufman Jr and K. J. Breeding, The automatic recognition of human faces from profile silhouettes, *IEEE Trans. Syst. Man Cybern.* **6**, 113–121 (1976).
20. Nixon, Eye spacing measurements for facial recognition, *SPIE Proc.* **575**, *Applications of Digital Image Processing VIII*, pp. 279–285 (1985).
21. G. Della Riccia and A. Iserles, Automatic identification of pictures of human faces, *Proc.* 1977 *Carnahan Conf. Crime Countermeasures*, pp. 145–148 (1977).
22. K. H. Wong, H. H. M. Law and P. W. M. Tsang, A system for recognizing human faces, *Proc. ICASSP* 89, pp. 1638–1642 (May 1989).
23. C. J. Wu and J. S. Huang, Human face profile recognition by computer, *Pattern Recognition* **23**, 255–260 (1990).
24. R. A. Campbell, S. Cannon, G. Jones and N. Morgan, Individual face classification by computer vision, *Proc. Conf. Modeling Simulation Microcomput.*, pp. 62–63 (1987).
25. T. Sakai, M. Nagao and T. Kanade, Computer analysis and classification of photographs of human faces, *Proc. 1st USA–Japan Comput. Conf.*, pp. 55–62 (1972).
26. Sir F. Galton, Numeralized profiles for classification and recognition, *Nature* **83**, 127–130 (31 March 1910).
27. K. Pearson, *Photographic Researches and Portrature*, Volume II, Chapter XII, pp. 283–333. Cambridge University Press, Cambridge (1924).
28. Sir Francis Galton, Personal identification and description—I, *Nature* 173–177 (21 June 1888).
29. V. Govindaraju, S. N. Srihari and D. B. Sher, A computational model for face location, *Proc. 3rd Int. Conf. Comput. Vision*, pp. 718–721 (1990).
30. V. Govindaraju, D. B. Sher, R. K. Srihari and S. N. Srihari, Locating human faces in newspaper photographs, *Proc. CVPR*, pp. 549–554 (1989).
31. J. L. Perry and J. M. Carney, Human face recognition using a multilayer perceptron, *Proc. Int. Conf. Neural Networks* **II**, pp. 413 (January 1990).
32. M. Propp and A. Samal, Human face recognition using neural networks. (In preparation.)
33. A. L. Yulie, D. S. Cohen and P. W. Hallinan, Feature extraction from faces using deformable templates, *Proc. CVPR*, pp. 104–109 (1989).
34. Sir Francis Galton, Personal identification and description—II, *Nature* 201–203 (28 June 1888).
35. L. D. Harmon, Automatic recognition of human face profiles, *Proc. 3rd Int. Joint Conf. Pattern Recognition*, pp. 183–188 (1976).
36. P. Baylou, E. H. Bouyakhf, G. Bousseau and A. Mora, Identification d'individus par analyse automatique du profil du visage, *Proc. 3rd Int. Conf.: Security Through Sci. Engng*, pp. 145–149 (September 1980).
37. J. T. Lapreste, J. Y. Cartoux and M. Richetin, Face recognition from range data by structural analysis, *Syntactic and Structural Pattern Recognition*, G. Ferrate, T. Pavlidis, A. Sanfeliu and H. Bunke, eds, pp. 303–314, NATO ASI Series (1988).
38. J. C. Lee and E. Milios, Matching range images to human faces, *Proc. 3rd Int. Conf. Comput. Vision*, pp. 722–726 (1990).
39. Y. Kaya and K. Kobayashi, A basic study on human face recognition, *Frontiers of Pattern Recognition*, pp. 265–289. Academic Press, New York (1971).
40. K. Preston Jr, Computing at the speed of light, *Electronics* **38**, 72–83 (1965).
41. W. K. Taylor, Machine learning and recognition of faces, *Electronic Lett.* **3**, 436–437 (1967).
42. H. Schlosberg, Three dimensions of emotion, *Psychol. Rev.* **61**, 81–88 (1954).
43. P. Ekman and W. V. Friesen, *Manual for the Facial Action Coding System*. Consulting Psychologists Press, Palo Alto (1977).
44. S. M. Platt and N. I. Badler, Animating facial expressions, *Comput. Graphics* **15**, 245–252 (August 1981).
45. K. Waters, A muscle model for animating three-dimen-

sional facial expression, *Comput. Graphics* **22**, 17–24 (1987).
46. M. Suwa, N. Sugie and K. Fujimora, A preliminary note on pattern recognition of human emotional expression, *Proc. 4th Int. Joint Conf. Pattern Recognition*, pp. 408–410 (1978).
47. D. Terzopoulos and K. Waters, Analysis of facial images using physical and anatomical models, *Proc. 3rd Int. Conf. Comput. Vision*, pp. 727–732 (1990).

About the Author—ASHOK SAMAL received his B.Tech. degree in Computer Science from Indian Institute of Technology at Kanpur, India, in 1983. He received his Ph.D. in Computer Science from University of Utah in 1988. Since Fall 1988 he has been with the Department of Computer Science and Engineering at the University of Nebraska–Lincoln. His primary research interests are computer vision, image processing, parallel algorithms, applications of parallel processing to computer vision, discrete relaxation and artificial intelligence. He has published articles in these areas in journals and conference proceedings. He is also interested in computer graphics, operating systems and functional languages.

About the Author—PRASANA A. IYENGAR received his B.E. in Computer Engineering from Bombay University, India, in July 1989. He was a Trainee Software Engineer for a year at Godrej & Boyce Mfg. Co. Ltd, Bombay, India. Since August 1990 he has been a graduate student at the Computer Science and Engineering Department, University of Nebraska–Lincoln. His research interests include computer vision and parallel processing.