ARMY RESEARCH LABORATORY

An Approach to Feature-Based Face Recognition

Michael J. Vrabel

ARL-TR-2215

December 2000

Approved for public release; distribution unlimited.

20001218 104

Army Research Laboratory

Adelphi, MD 20783-1197

ARL-TR-2215 **December 2000**

An Approach to Feature-Based Face Recognition

Michael J. Vrabel

Sensors and Electron Devices Directorate

Approved for public release; distribution unlimited.

Abstract

I detail a scheme for searching an unknown scene for occurrences of an object. The approach is independent of object size, location, and orientation and is tolerant of significant changes in object shape and appearance.

Contents

1 Introduction **1**

2 Brief Critique of Face-Recognition Literature **3**

2.1 Faces in Profile 3

2.1.1 Geometric, Feature-Based Matching 3

2.1.2 Holistic Face Recognition 4

2.2 Faces in Frontal View 5

2.2.1 Face Recognition Based on Use of Local Image Primitives 5

2.2.2 Face Recognition Based on Three-Dimensional Models 7

2.2.3 Face Recognition Based on Profile Feature Extraction 9

2.2.4 Face Recognition—Geometric, Feature-Based Matching, and Template Matching 10

2.2.5 Face Recognition—Neural Networks 16

2.2.6 Face Recognition—Gabor Functions 19

2.2.7 Cortical Thought Theory 20

3 Phase One—Algorithm Development **21**

3.1 Symmetry Breaking 25

3.2 Node Averaging 25

3.3 Enforced Solution Convergence 25

3.4 Global Model Performance 26

4 Phase Two—Algorithm Development **27**

5 Phase Three—Algorithm Development **30**

6 Conclusion **34**

References **35**

Distribution **39**

Report Documentation Page 41

Figures

1 Canonical face stored in memory for demonstration of phase one algorithm 21

2 Phase one test scene 26

3 Canonical image for use with demonstration of phase two algorithm 29

4 Phase two test scene 29

5 Exemplar configuration 30

6 Phase three test configuration for use with table 4 32

Tables

1 Relationship of equation (7) in tabular form, where $D_{K,N} \equiv 1$. 23

2 Performance of phase one algorithm with randomly selected state space trajectories 26

3 Phase two results: best node-feature matches 29

4 Phase three results: best node-feature matches 33

1. Introduction

Solving a difficult automatic target recognition (ATR) problem is beyond our present technological capability. "Difficult" here excludes from consideration those classes of problems that can be considered "toy" problems—constrained and artificial constructs of limited military interest. And, while recognizing the need to deal with difficult imagery, every attempt is usually made to simplify the problem. For instance, few in the ATR community (if any) would attempt to duplicate the sophisticated hierarchical understanding of the content of a scene that is natural to the human visual system. Rather, ATR has come to mean simply detecting the presence of an object (which may be any one of a diverse class of objects) in a scene and, perhaps, identifying it and determining its orientation. Achieving error-free ATR may require a duplication of the human or comparable visual system. Implied questions like these will not be answered in the near future.

I present an approach to image recognition that was intended to be very different from that in the current literature. The only related work I am aware of is that of Yow and Cipolla (1997). In addition to being original, my work also was intended to demonstrate an approach requiring no image database and therefore no training. Like the human visual system, my approach seeks not only to identify an object but to create a hierarchical description of its attributes. While the example I use is the human face, the approach can be applied to other objects. This report was also intended to add to the growing base of image recognition techniques that will contribute to the eventual solution of this class of problems.

The first phase of this report develops an algorithm to perform a global search of a scene to find an optimum match between scene content and phase one memory. This procedure can be repeated with the same scene to generate a ranked listing of potential multiple occurrences of a desired image. This recognition phase is performed without regard to image size, location, or orientation within the scene and is flexible enough to recognize noisy, occluded, or significantly distorted images. The first phase locates image candidates within a scene. The second phase builds upon this information by performing a more detailed analysis of the global characteristics of the image candidate. At each phase, a quantum increase in the information is available about a candidate, information that can be used not only to increase confidence in identifying the object but also to extract information about its characteristics. In the third phase, the most detailed analysis of the image candidates is performed. This phase examines individual feature shapes. Using a concise mathematical model stored in memory, the phase three algorithm performs the most detailed level of feature and, hence, image analysis and identification.

As an introduction to the face-based algorithm, I provide a brief critique of the many approaches to this problem. While face recognition is a highly specialized area within the field of image recognition, no other area has seen the same breadth of theory and algorithm development. Face recognition can be considered image recognition in microcosm. Developments in the face recognition field evolved from developments in autonomous image recognition (or its military equivalent—target recognition).

2. Brief Critique of Face-Recognition Literature

Although face recognition is a specialized area, it covers a broad spectrum of overlapping approaches, thereby lending itself to many different classification schemes. The classification scheme I have chosen is intended to allow for an organized, coherent treatment of face recognition.

A critical aspect of face-recognition algorithms is recognition accuracy. Because of the great variability in the characteristics of the face data sets used to test these algorithms, a rigorous comparison of recognition performance would be virtually impossible. Nevertheless, I will present a general discussion of performance accuracy.

Face-recognition performance (experimental subject study), whether based on a profile or full-face view, does not vary greatly (Ellis, 1975). Yet different views present dramatically different problems from a theoretical viewpoint and are reflected in the published face-recognition algorithms.

Much less literature on face recognition in profile exists than for faces in frontal view. I will discuss face profiles first.

2.1 Faces in Profile

Algorithms for recognizing faces in profile have the difficulty of dealing with a potentially dominant and highly variable hairline. This is done by avoiding all interior detail and operating upon the silhouette of the face only. This approach poses potential problems, which may explain why it has received little attention.

2.1.1 Geometric, Feature-Based Matching

Geometric, feature-based matching emerged from the discovery that facial identification is possible even when facial detail is marginally resolved. It is assumed that the overall geometric configuration of the face is sufficient for recognition. Depending on how facial detail is defined (this is often as simple as eyes, nose, and mouth), a feature vector is associated with this detail. This feature vector, the components of which define a space in which all face images can be placed, uniquely defines the face. The purpose of all feature-based matching is to establish the optimum description of this feature-vector-defined space.

Harmon (1976) (see also Harmon et al (1978)) is considered the classic feature-based treatment of faces in profile. Their technique resembles approaches to be discussed in more detail later. Harmon's approach to face recognition used the distances and angles between profile fiducial points (such as tip of chin and bottom of nose) as features and, using principal-component analysis, isolated a subset of optimal features. They achieved

a recognition accuracy approaching 100 percent with a large population of **manually** segmented faces. This work proved that with well-defined features, an accurate profile face recognizer is possible.

Wu and Huang (1990) use a similar technique except that their entire process is automated and applied to significantly different profiles of Asian rather than European subjects. With a backlit image, cubic B-splines are used to locate six profile turning points. From these points, a 24-dimension feature vector is produced. The training set comprises three images of 18 individuals. From these images, a mean and standard deviation are computed for each component of the feature vector. First, the feature vector components of an unknown face are compared to the stored feature vector components of the known image. If for any known image, the unknown image falls within certain prescribed distance criteria based on standard deviation, then the distance between known and unknown is computed:

$$d = \sum_{i=1}^{24} \mid X_i - u_i \mid / \sigma_i \ , \tag{1}$$

where u_i and σ_i are the ith feature vector component mean and standard deviation and X_i is the unknown feature vector component. The smallest value of d determines the match. A 100 percent success rate was achieved for the 18 subjects.

The strengths of this approach are (1) algorithmic simplicity, (2) computational speed, and (3) apparent accuracy. The weaknesses are (1) the need for face profiles with no confusing internal detail (e.g., dark image on a uniform background), and (2) the questionable algorithm performance with changes in facial expression.

2.1.2 Holistic Face Recognition

Holistic face recognition avoids the difficulty of locating fiducial points in a face profile by appropriately processing all profile boundary points. Two distinct holistic approaches are reviewed.

Kaufman and Breeding (1976) use a set of correlation coefficients that serve as feature vectors. Their approach uses the circular autocorrelation function. Properly defined, this function can be made invariant to scaling and translation with a simple relationship for rotating the face image. The major shortcoming of this approach is that it requires a closed-face contour. This was achieved by taking a face silhouette and shifting a copy horizontally so that the face profile of the copy fell behind that of the original. The portion of the original silhouette covered by the shifted duplicate was deleted. Then, by repeating the process of shifting duplicates up and down, a set of final silhouettes emphasizing the face profile was obtained. Using a weighted K-nearest neighbor decision rule as a classifier, Kaufman and Breeding showed that a recognition accuracy of 90 percent for a 10-class problem could be achieved. They made a performance comparison using moment invariants rather than the circular autocorrelation function. The

maximum accuracy for the moment invariants approach was 70 percent, with the qualification that the results are for a limited and particular face data set. The strengths of this approach are algorithmic simplicity and computational efficiency. The weakness is as follows: Because of the requirement for a closed contour and the potentially large variability of the hairline, the result of the above procedure for closed contouring introduces unavoidable accuracy-limiting distortions.

This requirement for a closed contour was circumvented in a later paper by Aibara et al (1991). Subjects were photographed in profile against a uniform background. A vertical scan was used to locate the tip of the nose, assuming the nose to be the most forward-projecting part of the face. Then 146 pixels corresponding to the boundary of the face silhouette clustered above and below the nose were selected. This defined a simple open-curve representation of the face. The face-evaluating function was the P-type Fourier descriptor. This function is both translation- and scale-invariant and has a simple relationship between original and rotated curves. The primary advantage of this descriptor over the circular autocorrelation function is that it can operate upon open curves. Another benefit is that it is sensitive to the low frequencies characteristic of the smooth curve of the human face profile. The Fourier coefficients of the descriptor at the bottom of the frequency range were used to define the components of the feature vector. Four photographs each of 90 subjects were used to define the test database. The average Fourier coefficients of three of the photos for each of the 90 were used to define the reference image, and the fourth became the unknown input data. Under the test conditions, a recognition accuracy of about 95 percent was achieved. The strengths of this approach are (1) it is algorithmically simple, (2) it is computationally efficient, (3) the elimination of the need to find fiducial points (always difficult to perform and a source of error) can enhance overall performance, and (4) eliminating the closed contour requirement of the previous approach eliminates sources of image distortion. The weaknesses of this approach are (1) the ever-present problem of separating the face contour from any background, and (2) the effect that expression changes have on solution accuracy. The latter problem requires storing many face images to cover variations in expression.

2.2 Faces in Frontal View

There are two types of approaches to face recognition: (1) geometric, feature-based matching, and (2) template matching. The overlap in the application of these techniques is so great that no attempt will be made to distinguish between them in this review, except to list the more significant attributes of each.

2.2.1 Face Recognition Based on Use of Local Image Primitives

Seitz (1989) explored image primitives as a basis for object recognition. Image primitives are locally defined characteristics of small clusters of pixels. Seitz used an array of 3×3 pixels. He concluded that local orientation

represents a powerful image primitive (local orientation is derived from variations in gray scale) more suited to image recognition than are primitives, such as local curvature, corner points, line ends, and crossings, as proposed by others (Asada and Brody, 1986; Bigun and Granlund, 1987; Grimson, 1989; Heitger et al, 1989).

In a later paper, Seitz and Bichsel (1991) exploited their approach used in (Seitz, 1989) to develop a practical face-recognition concept. A face is stored in memory as a low-resolution (24 × 32 pixels) representation of its local orientation. The local orientation is an angular measure indicating the direction of greatest gray-level change. To test for a match between an unknown image and a stored face representation, a sum of squared orientation differences (pixel by pixel) is used. This requires an accurate geometric normalization of the unknown face—its location, size, and orientation must be well defined. To do this, the starting point was an array of multiresolution representations of the unknown consisting of a power-of-two pyramid in which each resolution level contains one-fourth the number of pixels of the previous level. At each resolution level, specific features can be detected. At the lowest level, only a rough outline of the head is searched for. At higher levels, more detail is sought by the use of the results from the lower levels to progressively refine the definition of the face. With proper normalization of the face, identification can proceed entry by entry as with template matching. A second identification procedure based on facial geometry was used: from the results of the previously defined normalization, a set of human face landmarks (e.g., center of pupil, left and right ear lobe) can be established. The relationships between these landmarks is used to define a 62-dimension face feature vector. A test of algorithm performance with 397 face images of 70 different subjects with template matching, feature matching, or a combination of both yielded an accuracy comparing favorably with the previously reported results of others.

The strengths of this approach are as follows: (1) because it deals with intensity gradients rather than absolute gray-scale levels (as with the more traditional template-matching techniques), it is far less sensitive to illumination variations, and (2) it is capable of extracting more information because it operates over the whole face image, unlike the less illumination-sensitive techniques that use binary or edge-face models and are therefore restricted to areas of the face that have edges. The weaknesses of this approach are as follows: (1) because it exploits only a single characteristic of an image (intensity gradient), the amount of information about that image is severely limited and affects recognition accuracy; (2) like many other face-recognition procedures, it is sensitive to variations caused by image rotation, expression change, and hair style or other physical changes; and (3) it is computationally intensive.

Spacek et al (1994) developed a distinctly different approach with low-level descriptors. They used an edge finder to convert a face image to a binary representation. For arrays of 3 × 3 pixels, they described each boundary point as belonging to 1 of 36 possible types based on local boundary shapes

(called attributes). These attributes are a measure of local boundary curvature and orientation. The population of each of the 36 attribute types was summed over the entire face image. The normalized frequency distribution of the boundary points over these 36 types formed the basis for face recognition. To identify the face, five classifiers were tested: (1) a decision tree, (2) a Bayesian classifier on the full attribute set, (3) a Bayesian classifier on a reduced (optimum) set, (4) a learning vector quantizer on the full attribute set, and (5) a learning vector quantizer on a reduced (optimum) set.

I can draw the following significant conclusions from this work:

- No clear winner emerged from among the five classifiers, although the decision tree did seem to perform a bit better statistically.
- For discrimination purposes, only those pixels where change occurred were significant. This means that descriptors comprising straight lines were relatively unimportant and both right and sharp angles were more relevant than obtuse angles. Image information appears concentrated in regions where the local curvature is greatest—an experimental demonstration of a conclusion that can be arrived at based on theoretical considerations (for instance, see Resnikoff (1989)).

The strong points of this approach are algorithmic simplicity and computational speed. The primary weak point is a seeming lower recognition accuracy, which may be due to the limited information set the algorithm is capable of extracting from the face image.

2.2.2 Face Recognition Based on Three-Dimensional Models

All algorithms for face recognition must cope with the large variability in the human face. Accounting for variations due to lighting conditions and angle, not to mention expression, is a challenging task. For methods that operate off two-dimensional imagery, the solution is to either store and test a large database of images for each individual or to develop an algorithm that allows a classifier to be trained to multiple poses, an approach that has its own special problems.

Three-dimensional face recognition involves efficiently storing a three-dimensional model of a face and then extracting from it two-dimensional models for face recognition, as opposed to using a special class of face-recognition algorithms. This creates an extremely challenging constraint: because it typically requires a range-finding laser scanner in a laboratory environment, three-dimensional face modeling demands cooperative subjects.

Using concepts in differential geometry, Gordon and Vincent (1992), explored the use of morphological operators for feature extraction. They followed two general procedures:

1. Identify connected part boundaries for convex structures such as the outline of the nose and eye sockets.

2. Identify connected ridge lines for structures like the browline and chin/jaw line.

Two features referred to as ridge and valley lines can be derived from the principal curvature of a surface. Ridge lines are local maxima in the maximum normal curvature at a point along the line of maximum curvature. Valley lines are local minima in the minimum normal curvature along the line of minimum curvature. The procedure tends to produce unconnected line segments for these features. Further processing with dilation, thinning, and skeletonizing joins the line segments. Finally, a morphological seedfill algorithm known as geodesic reconstruction is used to extract the feature. No attempt is made to demonstrate the performance of this model in a face-recognition algorithm. The appearance of the images produced with this procedure does little to increase our confidence in its performance when used with unknown two-dimensional face images. More basically, it is uncertain how to link the three-dimensional database and the unknown image. Even if this procedure were successful, it may be limited to answering classes of questions such as: Is this person who she (or he) claims to be?

A second paper by Gordon (1992) applies this technique to a population of faces. Results appear promising, but additional work is needed.

A second, more practical treatment is that of Akamatsu et al (1991). They propose to laser scan the face of a cooperative subject, store the three-dimensional representation, and, using modern computer graphics techniques, produce two-dimensional synthesized images. The advantage here is that two-dimensional face images can be generated for any lighting condition and angle. This procedure simply provides a compact way of storing the equivalent of many two-dimensional face images of a subject.

The final paper reviewed in this section is not based on a three-dimensional face model. Rather it identifies faces based on isodensity maps (Nakamura et al, 1991), which are families of isodensity lines created by joining contiguous pixels of the same gray level after image quantizing. While isodensity maps do not define an exact relationship to the underlying three-dimensional structure of the face, the relief of the face is reflected well in the consequent binary image. The structure of the face-recognition algorithm is as follows: the gray-level histogram of all the points of a face image is divided into eight regions (the division points are selected experimentally). These divisions are weighted to yield more isodensity lines about the center of the face because it was observed that this region yielded more stable matching lines. Stable here means relatively unresponsive to changing image conditions.

Nakamura et al's face identification procedure is based on applying template matching to the consequent binary image. The Sobel operator is used to extract the contour edges and the success of the method requires a continuous contour edge. Propagation and shrinking are applied to connect broken parts of the contour lines. Template matching is implemented for any particular isodensity line level by sliding the unknown isodensity line pixel by pixel across the registered (stored) image from top to bottom

and left to right. A pixel match occurs if, in a 5- $\times$ 5-pixel window centered about the candidate-matching pixel, an isodensity line pixel occurs in the registered image. For similar faces, the matches achieved were long and contiguous. For dissimilar faces, the matches tended to be associated with short, fragmented lines. By combining the results of the pixel-by-pixel match with the finding that best matches are associated with long, contiguous line segments, Nakamura et al derived a relationship defining a best match.

The strong points of this approach to face recognition are (1) it is algorithmically simple, (2) it is computationally efficient, (3) because the binary line image strongly reflects the underlying three-dimensional structure of the face, the information content of the image is potentially high, (4) because a binary line image of a face is stored in memory, the amount of required computer storage is minimal, and (5) the authors claim that this procedure has high discrimination accuracy even for a face with glasses or a thin beard (stubble). The weak points are (1) the algorithm cannot cope with anything but a uniform background around the face (this is a problem common to most face-recognition algorithms), (2) the algorithm has been tested for only minimal head tilting or panning and it appears that its performance could be quite sensitive to these effects, (3) all test images were obtained only under conditions of controlled lighting with registered pictures renewed every several months to account for changes in the physical structure of the face; again it appears that the algorithm is sensitive to these sources of image variation, and (4) the algorithm has not been fully developed, compelling the use of experimental procedures to set a number of important parameters (hence, a potential dependency on the choice of faces).

2.2.3 Face Recognition Based on Profile Feature Extraction

Profile feature extraction refers not to faces in profile but rather to a unique approach of Jia and Nixon (1992) to recognizing frontal views of faces based on an analysis of a narrow vertical band of the center of the face. This vertical pixel intensity array encompasses the center of the forehead, the center of the nose (avoiding the sides of the nose and nostrils), the central area of the mouth, and continuing below the chin. To extract this band, it is assumed that the eyes can be sufficiently resolved in the face image to accurately locate and scale the image. The intensity distribution of the image so defined can be represented by an intensity projection. The intensity projection of an image $f(x, y)$ along the direction w on the line z is defined as

$$p_w(z) = \int_z f(x, y) dw. \tag{2}$$

This projection reflects the peaks and valleys of the intensity along the length of this vertical band. Although it resembles the face in profile, it is not the same. An efficient description of the profile is required and Jia and Nixon tested seven potential feature descriptors: (1) the resampled projection, (2) the autocorrelation function, (3) the dyadic autocorrelation

function, (4) the Fourier transform, (5) the Walsh transform, (6) the Fourier power spectrum, and (7) the Walsh power spectrum.

To minimize the effect of truncation on the autocorrelation function and Fourier transform, a Hamming window is used. The profile is sampled along its length at 128 points. To measure the performance of the seven descriptors, Jia and Nixon used two differently defined relative differences:

$$d_1 = \sum_{i=1}^{128} \mid x_i - y_i \mid /(\mid x_i y_i \mid)^{0.5} \text{ and} \tag{3}$$

$$d_2 = \sum_{i=1}^{128} \mid (x_i - y_i)/(x_i + y_i) \mid, \tag{4}$$

where x_i and y_i are the feature elements of a face image and the relative match between the images is given by the inverse of the relative differences. An analysis of the match among 40 subjects demonstrated the Walsh transform to be the best of the seven descriptors. It is difficult to ascertain the performance of the algorithm from the data presented by its authors except to note their statement: "These results have shown . . . sufficient reliability to discriminate between different persons' faces and to match different pictures of the same person."

The strengths of Jia and Nixon's approach are (1) algorithmic simplicity, (2) computational speed, and (3) reasonable accuracy under controlled conditions with a cooperative subject. Weaknesses of the approach are (1) since the algorithm works with only a limited number of pixels in a narrow band on the face, the performance can be adversely affected by changes in the chosen band that might not otherwise affect an algorithm operating over the entire face, (2) tests indicate that an up-and-down movement (more than a side-to-side movement) can adversely affect algorithm performance, and (3) extreme changes in lighting will cause the technique to fail.

2.2.4 Face Recognition—Geometric, Feature-Based Matching, and Template Matching

This classification represents an intermixing of template-matching techniques and matching based on geometric features. The latter class is subdivided into feature-based algorithms that exploit discrete facial markers such as the nose and the eye, and those that take a more holistic approach. For recent papers on feature-based and template matching not referenced here, see Brunelli and Poggio (1992a), Robb (1989), Smith (1986), Sutherland et al (1992), and Wong and Calia (1992).

Holistic, Feature-Based Matching: Turk and Pentland (1991) take an information theory approach to face recognition that uses principal component analysis, more commonly referred to in the literature as the Karhunen-Loeve expansion. This treatment postdates an earlier set of papers by Kirby and Sirovich (1990) and Sirovich and Kirby (1987) who use a similar approach. Other work using the Karhunen-Loeve transform in face recognition can be found in (Suarez, 1991). This approach can be thought of as

decomposing face images into a set of characteristic features called eigenfaces. In mathematical terms, the eigenvectors of the covariance matrix of the set of face images are determined. These eigenvectors represent a set of features that characterize the variation between face images. Each image location contributes to some degree to each eigenvector, so that each eigenvector appears as a ghostly face. These eigenvectors are therefore called eigenfaces. The eigenfaces can be viewed as a map of the variations between faces. Each face in the training set is represented as a linear (weighted) combination of the eigenfaces. This approach leads to a concept of face recognition that is based on a set of features lacking any correspondence to an intuitive sense of facial components. The idea is to find that set of eigenfaces that most efficiently account for the distribution of training face images within a complete image space.

The algorithm training for this approach proceeds as follows:

1. Select a set of face training images. Each individual can be represented many times under various lighting conditions, head orientation, and so on.
2. From this training set, calculate the subset of eigenfaces that correspond to the highest eigenvalues; these define the optimized face space. As new faces are added to the training set, these eigenfaces can be recalculated.
3. Calculate the weight distribution for each individual for each eigenface corresponding to its distribution in the defined face space. These weights form a vector that describes the contribution of each eigenface in representing the face image.

To recognize a new face—

1. By projecting the input image onto each of the eigenfaces, calculate its weight set.
2. With any standard pattern-recognition classification algorithm, determine whether the image is a face. If the image is a face, determine whether it is known or unknown.

A test of the algorithm involved a large image database of 16 subjects. The independent variables for this data set were differences in lighting, size of the head, orientation of the head, and combinations of these three variables. The algorithm achieved 96 percent correct classification averaged over lighting variation, 85 percent correct averaged over orientation variation, and 64 percent correct averaged over size variation.

The strengths of this approach are (1) it is algorithmically simple, (2) it works very well within the limitations of the algorithm, (3) it is insensitive to small changes in face image, or at least in the ability to train to small face variations, and (4) there is some indication that the procedure can be scaled to handle a large population without an excessive number of eigenfaces. The weaknesses of this approach are (1) it is computationally expensive,

(2) the background (including hair) can significantly affect recognition performance, and (3) the algorithm is particularly sensitive to size variation (it requires a good geometrical normalization procedure).

The discrete cosine transform is another holistic, feature-based approach that has been applied to face recognition; this approach is strongly analogous to that used by Turk and Pentland (1991). A comparison of the discrete cosine transform with the discrete Fourier transform and the Karhunen-Loeve transform was made by Goble (1991). That author determined that the discrete cosine transform results were superior in all tested cases to those obtained with the discrete Fourier transform and in some cases were superior to those obtained with the Karhunen-Loeve transform. The strengths and weaknesses of this approach are similar to those of the Karhunan-Loeve transform, but the former appears to be somewhat inferior.

Discrete Feature-Based and Template Matching: The major difficulty in evaluating the performance of the various face-recognition algorithms is the differences in the training and test data sets used by each author. Even a reasonable comparison is difficult, if not impossible. A paper by Brunelli and Poggio (1993) attempts to address this difficulty, at least with regard to template matching versus geometric, feature-based matching.

The authors created a database of 188 images of 47 subjects. Photographs of each individual were taken over a period of weeks. The illumination was only partially controlled, and the scale in face size was varied by as much as 30 percent. While only frontal views were used, no effort was made to ensure perfectly frontal images.

To apply geometric, feature-based matching to this data set, one must normalize the faces properly, that is, the features to be extracted from the images must be independent of position, scale, and rotation of the face in the image plane. This is achieved by locating the eyes in each image. (See Stringa (1993) for a description of a more sophisticated eye-detection algorithm.) To do this, a set of five eyes templates were used. The five spanned a range of sizes reflecting the uncertainty in face image sizes. Brunelli and Poggio (1993) performed the template matching using a normalized cross-correlation coefficient. Having located the eyes and, thus, normalized the geometry of the face, one can approximately locate, size, and orient various facial features as the nose and mouth by using anthropometric measures. Integral projection is used to achieve the best possible definition of the various facial features. Let $I(x, y)$ be the image. The vertical integral projection is

$$V(x) = \sum_{y} I(x, y). \tag{5}$$

Similarly, the horizontal integral projection is defined as

$$H(y) = \sum_{x} I(x, y). \tag{6}$$

These equations are applied to a binary (edge) representation of the face. By using edge-projection analysis, one can create two maps for each face

image, one where the horizontal edges dominate (in the natural reference frame of the face) and the other where the vertical edges dominate. By applying the integral projections to the edge-dominance maps, one can conduct a careful analysis of the resulting profiles and with reasonable accuracy locate the facial feature points needed to generate the feature vector. A feature vector of length 35 was then created. Face recognition is then performed with a Bayesian classifier. The effectiveness of the selected features (components of the feature vector) in describing the images was investigated with the Karhunen-Loeve expansion. The results of this study suggested that performance could be improved with more accurate feature detectors, but, as Brunelli and Poggio (1993) pointed out, it is not clear how to design them.

A template-matching scheme was implemented with whole-image gray-level templates. Each subject in the database is represented by a pixel array of four masks representing eyes, nose, mouth, and face (the region from the eyebrows downward). The masks are positioned with the results described previously. The unknown (unclassified) image is compared with all the database images in turn, and a vector of matching scores computed through normalized cross correlation is returned. The unknown subject is identified as the one giving the highest cumulative score.

Correlation-based recognition is sensitive to illumination gradients. To determine whether some form of image preprocessing could minimize this problem, Brunelli and Poggio tried four schemes:

1. No preprocessing.
2. Intensity normalization with the ratio of the local value over the average brightness in a suitable neighborhood.
3. The intensity of the gradient: $\mid \partial_x I \mid + \mid \partial_y I \mid$.
4. The Laplacian of the intensity image: $\partial_{xx} I + \partial_{yy} I$.

The best results were obtained with gradient information (scheme 3).

The relationship between image resolution and recognition accuracy was investigated. The results indicate that correlation-based (template-matching) recognition is possible with window templates as small as 36×36 pixels. At least for the database examined and the matching procedure used, the feasibility of such small templates tends to negate the common objection that recognition through template matching is computationally too expensive.

How effective were the individual windowed templates? The experimental ranking in order of decreasing performance is (1) eyes, (2) nose, (3) mouth, and (4) whole face template.

How were the rankings of the individual templates combined? The scores were simply added together. The results of combining all windowed results had a beneficial effect on recognition and increased the robustness of the classification.

The conclusion drawn from the vector of geometric features versus template matching was that template matching is superior in recognition performance. This result must be qualified by adding that it is specific to the approach and the database used. As feature-detection and template-matching schemes, these approaches are fairly sophisticated.

The weaknesses of this approach are that (1) the computational complexity of the scheme is high, and (2) the way in which the eye-detection procedure (part of the normalization scheme) was implemented is relatively unsophisticated and can be improved. Overall, this paper represents a competent and thorough treatment of the approaches taken.

Brunelli and Poggio's paper (1993) used a relatively simple classification scheme. In a second paper by the same author (1992b) in which they follow a geometric, feature-based matching scheme, a more sophisticated classifier called a Hyper Basis Function network is used. The paradigm used with this network is learning from examples, which can be regarded as the reconstruction of an unknown function from sparse data whenever the input and output can be expressed as numerical vectors.

Among the components of the feature vector used are (1) pupil-to-nose vertical distance, (2) pupil-to-mouth vertical distance, (3) pupil-to-chin vertical distance, (4) nose width, (5) mouth width, (6) zygomatic breadth, (7) biogonial breadth, (8) chin radius, (9) mouth height, (10) upper lip thickness, (11) lower lip thickness, (12) pupil-to-eyebrow separation, and (13) eyebrow thickness.

For face recognition, 35 features were used. The above list was expanded by eliminating the consequences of facial bilateral symmetry and expanding some features. This list is reasonably representative of what is generally used for facial feature vectors and gives good insight into the kind of information needed and, hence, the difficulties in extracting it.

The number of Hyper BF networks necessary for identification is the same as the number of subjects to be recognized. For proper training, a large number of facial images for each subject are required. During training, each network undergoes a competitive learning stage, in which the weights of the different features and prototypes are changed to maximize the response to inputs corresponding to the subject represented. A Hyper BF network has three significant quantities (1) the unknown coefficients to be determined and associated with a scaler function to be approximated, (2) the vector defining the network centers, and (3) the weights assigned to each input coordinate, which determine the importance of each input. The Hyper BF network can be considered as a memory representation in which the distinctive (or discriminating) facial features are exaggerated, creating a caricature.

The strengths of this classifier are that it (1) is reasonably accurate (a cited recognition performance of 95 percent), and (2) provides insight into facial caricatures. The weakness is that it requires a large number of training images for each individual.

Kanade (1973) describes a rather interesting approach to face recognition based on feature matching. The approach applies a flexible analysis scheme that combines local processing with global recognition. Backup procedures are included so that if difficulties are encountered during the recognition process, previous steps can be retried. The approach is claimed to be flexible and adaptive.

The procedure starts with a binary face representation. A thresholded Laplacian operator is used with local pixel averaging to produce a smooth image. The Laplacian was superior to either the Robertz operator or a maximum-of-differences operator for this purpose. No additional operators such as thinning or elimination of isolated points are used.

Face recognition is a two-stage process. First, the face and its features are located. Integral projections of horizontal and vertical slits (or windows) are used to localize facial features, a procedure described earlier. The contours of these projections are matched to stored families of contours to identify features.

The procedure starts by finding the top of the head and then proceeding to the following facial features in the order listed (1) sides of face at cheeks, (2) vertical regions of nose, mouth, and chin, (3) chin contour, (4) nose end points and cheek areas, and (5) eye positions. If in performing the above sequence of search operations, an error is encountered (i.e., a poor match is achieved based on criteria intrinsic to the algorithm), we go back one or more steps and retry the process. A test of the localization procedure for the following four classes of faces was made (1) full face with no glasses or beard, (2) full face with glasses, (3) face with turn or tilt, and (4) face with beard. The first class, the training set for this algorithm, performed reasonably well, achieving a judged correct performance of approximately 92 percent for 670 faces. The third class also performed reasonably well at 80 percent for 79 faces. The other two classes performed poorly.

After the face is localized, a more detailed examination of the facial features is made (the second stage). The procedure used is similar to the first stage, but because facial features are now localized, it is computationally feasible to perform a much higher resolution and detailed search. The second stage outputs a set of fiducial points for the face. From these points a set of 16 feature-vector components are generated that comprise ratios of distances, areas, and angles and represent enough geometric information about a face to permit some recognition.

A simple measurement of distance between known and unknown face was used as a test of the identification algorithm. It is difficult to judge the performance of the model since so little information is given about the variability among faces. A database in excess of 600 faces was available, but the number of subjects this represents is unknown. It was noted, though, that performance improved when ineffective feature vector components were omitted, not an uncommon finding. The first stage face-localization algorithm appears to have worked better than the second stage face-identification algorithm.

The strengths of this approach are that it (1) represents a comprehensive treatment of face localization (or segmentation and geometric normalization) that relies not on single facial features like the eyes but on a distributed set of facial characteristics (potentially a more robust approach), and that it (2) is an adaptive procedure that uses feedback to improve algorithm performance, a technique not commonly used. The weaknesses of this approach are (1) algorithmic complexity, (2) computational intensity, and (3) questionable face-identification capability.

2.2.5 Face Recognition—Neural Networks

Neural networks are used in face-recognition algorithms as feature detectors (including segmentation), classifiers, or both. The first paper I review uses a three-layered feed-forward neural network as a classifier.

Lim et al (1992) describe an algorithm for feature-vector extraction and classification. The first step is to transform all images to a binary representation using the Sobel operator with an experimentally selected threshold. To geometrically normalize the images, an eye-detection scheme is used that relies on eye blinking. If a large number of frames of an image are grabbed, a big difference in gray level at the regions of the pupils is detected for those frames with closed eyes. This procedure assumes reasonable frame-to-frame face-image registration. Having detected the pupils, the algorithm can determine the approximate location of various characteristic points of the face. The authors provide no details on the refined feature-extraction method, except that a final feature vector of 17 elements resulted.

To train the classifier, Lim et al computed three feature vectors derived from three images of each individual. The image with the largest Euclidean distance from the remaining image was discarded and the mean of the remaining two used for training. The neural network consisted of 17 input units, 25 hidden layer units, and 4 output units. A back-propagation algorithm was used for training. A 100 percent recognition rate was achieved for a data set of 10 subjects.

The strength of this approach is its algorithmic simplicity, except for the uncertainty in the feature-extraction procedure. The weaknesses are that (1) the data set was too small and poorly defined to permit proper evaluation, (2) it is not clear what benefits derive from using this neural network, and (3) the ability of the neural network classifier to be scaled to handle large numbers of individuals remains uncertain.

A paper by Soulie et al (1993) describes a neural network model for both face segmentation and identification. The authors look not just at recognition performance but also at rejection performance. Rejection performance is the ability of the neural network to detect and reject unknown faces. The scenes used to test the face-segmentation model contained a reasonable amount of background clutter with multiple faces of varying size. Both the face-segmentation and face-identification modules were of a similar neural network design. Both were time-delay neural networks that use

a multilayer perceptron architecture. The networks were trained with a gradient back-propagation algorithm (a gradient-descent method). The authors used the stochastic gradient version of the Widrow-Hoff rule. The main difference in the architecture between the face-segmentation and face-identification networks was the output layer. The recognition network used as many units in the output layer as there were subjects to identify. The segmentation network had two outputs—face or no face. As is typical of neural networks, good performance requires many examples of each subject. With fewer than 150 images per person, performance was seriously degraded.

Two databases were used to test the network. The first contained face images for 20 individuals that were centered and normalized to 20×24 pixels. Of those, 14 were known faces and their image sets were divided into training and test sets. The remaining 6 were unknown faces used to test the network's ability to detect and reject unknown faces. The second database contained 250 scenes, with various groupings of individuals sitting or standing in a home-like setting. Searches of these scenes were restricted to persons looking almost directly at the camera.

A serious problem of the face segmentation module was variation in face size; this was solved with multiresolution decomposition of the image. This decomposition gives multiple scene views at different scales, thus ensuring scale-invariant detection. A postneural network-processing algorithm was used to statistically select those windows containing faces from among the many resultant segmentation windows. The system was demonstrated to be robust to partial face occlusion and proved to be effective in locating faces of varying sizes in a complex scene, although face segmentation failures occurred (or could occur) with faces that were too close to the border of the image.

Test results with the first database showed that the identification error rate increased with the number of subjects to identify. For the full 14-person set, the error rate was 1.3 percent. When tested to reject the faces of unknown subjects, the network performed reasonably well. For instance, 85 percent of the unknown faces were detected with a 5 percent rejection of known faces.

The strengths of this approach are (1) robust to variations in face rotation, expression, lighting, and noise, (2) capable of segmentation with a complex background scene, and (3) can be made insensitive to face size variations. The weaknesses are (1) requires many images of an individual for proper neural network training, (2) scales poorly in training time and related network complexity as the number of individuals to be recognized increases (as presently configured, the network is restricted to tens of individuals and cannot cope with hundreds), and (3) multiresolution decomposition is computationally very expensive.

In another neural network treatment, Runyon (1992) compares a neural network with a nonneural network classifier. Both use the same automated segmentation and preprocessing algorithms. The nonneural network version uses a Karhunen-Loeve transform feature extractor and a K-nearest

neighbor (KNN) classifier. The neural network version uses the same feature extractor but with a multilayer perceptron classifier having a back-propagation learning rule. Runyon's thesis is significant because it looks at classes of real-world face-recognition problems that have been downplayed or even ignored in the past, and it does so within the context of this classifier comparison.

The segmentation processor uses multiple images combined with a motion detector, with the assumption that the only motion in front of the camera is that of the subject. The relative motion isolates the individual from the background. A correlation is then performed between a reference image and the unknown input image, which permits centering and scaling of the unknown image. After normalization, the image is multiplied by a positioned gaussian window to emphasize the inner region of the face and deemphasize the outer area, thus reducing the problem with hair and its variability.

A K-nearest neighbor classifier uses a scoring technique that compares a feature of the unknown with a known face from memory and assigns a score of K to its closest match. A score of K-1 is assigned to the next closest match, and so on. After scoring is completed, the scores are summed and the known image with the highest value identifies the unknown face. K can be any value; at its extreme, it can be assigned a value of 1.

The first test was performed with 23 users over two days. Its purpose was to determine the classification accuracy of competing models for a large number of subjects when training, and test images were collected on different days. For the K-nearest neighbor classifier, the recognition rate was 29 percent. For the back-proprogation neural network (BPNN) classifier, the recognition rate was 34 percent. A baseline was generated with same-day training and testing. The results were KNN = 78 percent and BPNN = 76 percent. An investigation was made to determine the effect of segmentation inaccuracy on recognition accuracy. With manually segmented images, the same-day results were KNN = 90 percent and BPNN = 97 percent. In general, segmentation error contributed roughly 20 percent to the decrease in recognition accuracy. A final test was made with this data set. For training, it used images collected on both days, and for testing, it used images collected on both days. The results were KNN = 62 percent and BPNN = 74 percent.

The second test was with four subjects over seven days. Its purpose was to study the effects of time on recognition accuracy, albeit for a smaller set of subjects. Four training and three test images were collected for each subject each day for a total of seven days (a total of 28 training images and 21 test images per person). The test used an iterative procedure: first, each system was trained on each person's four training images from day one and then tested on all 21 images of each person. The system was then retrained with the images from the second day in addition to those of the first. The accuracy of this system was again tested with all 21 test images of each subject. The system was then trained with three days of training images,

tested, trained again, and so on, until the training images for all seven days had been used. For the KNN model, improvement in performance accuracy was less than monotonic with training spanning one to seven days. The initial recognition accuracy was 62 percent and the final was 90 percent. For BPNN, the performance improved monotonically starting at 82 percent and ending at 100 percent. The overall performance of the neural network classifier was superior, particularly for the seven-day test.

No particularly good rule appears to exist that would specify the structure or connectivity of a neural network. Hancock and Smith (1990) apply a genetic algorithm to specify the structure of a BPNN. The network is feed-forward and has a single hidden layer with full connectivity to the output units. When the genetic algorithm was applied to simple face models, the best score was 57 percent, compared with a score of 44 percent for the fully connected network and the best score from an initial net random population of 41 percent. The results demonstrate that a genetic algorithm can improve the internal structure of a neural network. The major drawback of this approach is that it is CPU-intensive. Runs are defined in terms of CPU days or weeks.

For additional recent neural network treatments of face recognition, see Allinson and Ellis (1992); Bouattour et al (1992); Frasconi et al (1992); Kerin and Stonham (1990); Krepp (1992); Sander (1988); Turk and Pentland (1991).

2.2.6 Face Recognition—Gabor Functions

The approach of Petkov et al (1993) to face recognition was motivated by a desire to duplicate processes of the primary visual cortex in mammals. (For a similar approach using the Gabor wavelet transformation, see Manjunath et al, 1992.) Experimental results indicate that two-dimensional Gabor functions can be made to fit the receptive fields of simple cells in the primary visual cortex of mammals. The projection (functional inner product) of a two-dimensional image on a Gabor function is performed. This projection is then integrated over all pixel locations of the input face image. Discretization is used with eight discrete angles (or orientations) and eight basic spatial frequencies. A feature vector of 64 Gabor functions is thus generated.

Tested on a set of 205 face images of 30 subjects, a recognition rate of 94 percent was achieved. Individual images of a person showed differences in facial expression and orientation, but both size and lighting variations were limited. The quoted accuracy of this approach must be qualified by the test conditions since each test image had a set of five to nine images of the same individual with which to make a match.

The strengths of this approach are (1) the preliminary information it has provided about biological processes, and (2) the relative accuracy it achieves under the constrained test conditions. The weaknesses are (1) the computational intensity, even when constrained to Gabor functions of only eight orientations and eight scales, and (2) the need for good geometric normalization of the face image.

2.2.7 Cortical Thought Theory

In 1985, Routh of the Air Force Institute of Technology proposed cortical thought theory as an attempt at a unified brain theory. The theory was applied to a sequence of face-recognition systems. It has since fallen out of favor and is now considered obsolete. (See Lambert, 1987; Russel et al, 1986; Russel, 1984.)

3. Phase One—Algorithm Development

Image recognition can be characterized as an optimization procedure—a best matching between memory and the contents of a scene. The computational complexity of the approach to be taken requires that much of the redundant, low-information content data in the scene be eliminated before proceeding. This is done by processing the scene to produce a family of contours. If the original scene is a gray-scale image, this requires a gray scale to contour transformation. All contours are further processed, creating a set of contour nodes (or points). (The procedure for doing this will be the subject of another report.)

If a postprocessed scene composed of a node set can be said to subjectively represent the preprocessed scene fairly well, then it is reasonable to expect that the presence of an object can be ascertained. The example image used throughout this report is the frontal view of a face. Statements made about this image can be generalized to almost any other class of imagery. A face is recognizable as a face because it possesses certain attributes that we refer to here as features. A feature is a small attribute compared to the size of the overall image. For instance, the contour outlining the image of a face is not a feature but a small continuous segment is a feature (as, for instance, the chin). The context of the chosen features forms the basis for image recognition. The image of a canonical or standard face to be stored in memory is arrived at by averaging a large population of faces. Several standard faces can be stored by dividing the population into subgroups based on facial characteristics. For demonstration purposes, the shape of the face is chosen arbitrarily.

Figure 1 is a phase one representation of a canonical face. It is composed of six features. The choice of features and feature count is somewhat arbitrary. The center of each line segment is assumed to be the most probable location of each feature. Line segments indicate the relative orientation of features

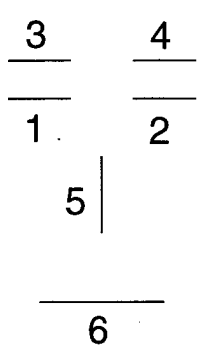


Figure 1. Canonical face stored in memory for demonstration of phase one algorithm. Numbers are feature designations (location = center of line). Orientation of lines represents relative orientation of features.

and have no other significance. Figure 1 barely resembles a face because the computational demands of phase one require that a scene be searched without *a priori* knowledge of image size, location, or orientation. This simple representation can be used as the basis for a more complex search scheme, one that efficiently scans a scene in a hierarchical fashion searching for any of a large number of types of images.

The initial stage of the scene-analysis algorithm involves a global search for an optimum match between a pattern formed from information extracted from the features of the canonical face and the unknown scene. The overall image-recognition algorithm is organized hierarchically with progressive phases that allow an increasingly focused examination of smaller regions of the scene. The initial data set will be superseded by an increasingly detailed data set as the search progresses.

Equations (7) through (10) are key to the global image search. The equation parameters are taken from memory and define the characteristics of the object being sought for this phase:

$$Q_1^{K,N,M} = D_{K,N} + D_{K,M} - D_{N,M} \ , \tag{7}$$

where

$K, N, M = 1, \ldots n,$

$K \neq N, K \neq M, N \neq M,$

K, N, and M are feature numbers,

n is the feature count, and

$D_{i,j}$ is the relative distance between features i and j with the smaller of $D_{K,N}$ or $D_{K,M}$ always normalized to 1.

Table 1 gives the characteristics of $Q_1^{K,N,M}$ as a function of both the angle formed by vectors between features K, N, and M with the vertex at K and the larger of $D_{K,N}$ and $D_{K,M}$.

$$Q_2^{K,N} = X_K X_N + Y_K Y_N \ , \tag{8}$$

where

$K, N = 1, \ldots n,$

$K \neq N$, and

X_i and Y_i are the x, y components of the unit tangent vector defining the orientation of feature i.

Table 1. Relationship of equation (7) in tabular form, where $D_{K,N} \equiv 1$.

	$D_{K,M}$				
θ^*	1	2	3	5	10
0	2.00	2.00	2.00	2.00	2.00
15	1.74	1.93	1.95	1.96	1.96
30	1.48	1.76	1.81	1.84	1.85
45	1.23	1.53	1.60	1.65	1.68
60	1.00	1.27	1.35	1.42	1.46
75	0.78	1.01	1.09	1.16	1.21
90	0.59	0.76	0.84	0.90	0.95
105	0.41	0.54	0.60	0.65	0.70
120	0.27	0.35	0.39	0.43	0.46
135	0.15	0.20	0.23	0.25	0.27
150	0.07	0.09	0.10	0.11	0.12
165	0.02	0.02	0.03	0.03	0.03
180	0.00	0.00	0.00	0.00	0.00

$^*\theta = \cos^{-1}\left([D^2_{K,N} + D^2_{K,M} - D^2_{N,M}]/2ab\right)$.

For features without a clear orientation (e.g., a circle), any orientation can be selected. This equation is nothing more than a vector dot product:

$$Q_3^{K,N} = D_{K,N} + D_{K,R} - D_{N,R}\,, \tag{9}$$

where

$D_{K,R} = D_{K,N}$, and

R is the location of the projection of the unit tangent vector at feature K closest to feature N.

Note that equation (9) has the same functional form as equation (7). With the normalization requirements on the distance terms, equation (9) can also be written as

$$Q_3^{K,N} = 2 - D_{N,R}\ . \tag{10}$$

Equation (9) is a late addition to this equation set and was included to cure defects in the performance of the phase one model. It is left to the reader to determine the class of problems this equation was designed to resolve.

$$Q_4^{K,N,M} = D_{K,N}/D_{K,M} \tag{11}$$

if

$$Q_4^{K,N,M} > 1 : Q_4^{K,N,M} = 1/Q_4^{K,N,M}\ . \tag{12}$$

Equation (11) resolves the obvious ambiguity that exists in the previous equations.

An analogous set of relationships, derived in this case from the set of nodes constituting the unknown scene, can also be defined and is designated as

$$q_1^{a,b,c},\ q_2^{a,b},\ q_3^{a,b}, \text{and } q_4^{a,b,c}\ . \tag{13}$$

There are three feature designations, K, N, and M, and three unknown scene node designations, a, b, and c. The following pairings between features and nodes are used: $a \Rightarrow K, b \Rightarrow N$, and $c \Rightarrow M$. The relationships of the previous equations are combined:

$$E_1^{a,b,c} = 1 - 0.5 \mid Q_1^{K,N,M} - q_1^{a,b,c} \mid, \tag{14}$$

$$E_2^{a,b} = 1 - \mid Q_2^{K,N} - q_2^{a,b} \mid, \tag{15}$$

$$E_3^{a,b} = 1 - \mid Q_3^{K,N} - q_3^{a,b} \mid, \text{and} \tag{16}$$

$$E_4^{a,b,c} = 1 - Q_4^{K,N,M} / q_4^{a,b,c}. \tag{17}$$

If $E_4^{a,b,c} < 0$, then

$$E_4^{a,b,c} = 1 - q_4^{a,b,c} / Q_4^{K,N,M}. \tag{18}$$

Written as above, all E terms have a range of 0 to 1 and provide a measure of the correlation between the nodes of the unknown scene and the features from memory. The optimization procedure attempts to find that node corresponding to a (only one per feature) contained within the unknown scene for which the following function is minimized:

$$\Lambda_m^a = -\xi_m^b \xi_m^c (E_1^{a,b,c} E_4^{a,b,c} + E_2^{a,b} E_2^{a,c} E_2^{b,c} + E_3^{a,b} E_3^{a,c}), \tag{19}$$

where $m = 1, \ldots M'$ with M' being the feature count and ξ to be subsequently defined. The E-term groupings of equation (19) are somewhat arbitrary. They are based on the anticipated significance of the individual E terms in image identificaion and backed by computer simulation.

I use an iterative procedure to find node a for all features. This procedure is computationally intensive. Any *a priori* knowledge of image size, orientation, or location in the unknown scene, however approximate, can greatly reduce these computational requirements. Under the assumption of no such knowledge, the procedure, while intensive, is nonetheless straightforward: nodes a, b, and c are selected at random from the unknown scene and Λ_m^a is computed. This allows the development of a performance history for each node. This history is given by

$$\xi_m^a = (A\xi_m^{a'} + \mid \Lambda_m^a \mid)/(A+1), \tag{20}$$

where

ξ_m^a is the new value,

$\xi_m^{a'}$ is the previous value, and

A is a constant.

Initially all nodes are assigned identical values for ξ_m^a. The choice of A is somewhat arbitrary. This choice affects the algorithm's "forgetfulness" and the solution convergence rate. Too small a value creates the risk of locking the iterating solution in a local minimum far removed from the optimum solution. Too large a value can adversely affect the solution convergence time. Rather than adjusting ξ_m^a after every random selection of nodes b and

c, I calculate Λ_m^a for a large population of selected nodes b and c and use the best performer out of this population as per equation (19). As a performance history for each node contained within each feature begins to evolve, this can be used to increasingly bias the initially random choice of nodes b and c toward the best performers. All such decisions involve this balancing between rate of solution convergence and the risk of becoming trapped in a less than optimum local minimum.

3.1 Symmetry Breaking

Many images, including the image in figure 1, have a high degree of bilateral symmetry. All unknown scene nodes that are strong candidates for feature 1 of the face image, for instance, are equally strong candidates for feature 2. The same can be said of feature pair 3 and 4. This situation must be rectified. The most direct procedure is to select the evolving optimum node candidate for any of the above features early in the optimization procedure, inhibit all nodes not in its vicinity, and reverse roles for the nodes of its matched feature pair.

3.2 Node Averaging

To ensure that isolated nodes that were inadvertently optimized to the relationship of equation (19) are not selected, I perform node averaging. If node a has n' nearest neighbors, then

$$\Lambda_m^a = (\Lambda_m^a + \sum_{n=1}^{n'} \Lambda_n^a)/(n' + 1). \tag{21}$$

This operation occurs periodically within the optimization procedure and ensures that any node response is representative of the nodes in its vicinity. Vicinity is defined as both physical proximity based on distance and a shared contiguous contour.

3.3 Enforced Solution Convergence

Initially, equation (19) places no restriction on the choice of nodes b and c. For any node a to represent an optimum fit to any feature, the measure of the goodness of that fit is its relationship to the other optimum feature nodes (nodes b and c). At the end of the solution procedure, nodes b and c must correspond to the optimal nodes for their corresponding features. To a limited extent, the form of equation (19) ensures this correspondence. To guarantee it, I must progressively enhance the contributions of ξ_m^b and ξ_m^c to equation (19) as the solution proceeds. Any number of approaches can be taken. An optimum schedule can significantly affect the solution convergence rate. Under any circumstance, the solution convergence rate must be slow enough to ensure that the solution does not become trapped in an unacceptable local state. No attempt has been made to find an optimum schedule. The output of the first phase global search is a single node associated with each feature from which the absolute size, location, and orientation of the image candidate can be derived.

3.4 Global Model Performance

Figure 2 is a scene to be searched for the test image. In this case the test image is that of figure 1. The phase one algorithm searches the unknown scene for the globally optimum fit. It should be recalled that the search is performed without any knowledge of the size, location, or orientation of any potential face candidate.

Table 2 demonstrates the performance of the model for a number of randomly selected state space trajectories. The table includes the effects of significant changes in the shape of the test image in figure 1. This change in shape tests the ability of the phase one algorithm to find a match when the best match between memory and the unknown scene is poor. As can be seen in table 2, the algorithm performance is reasonable.

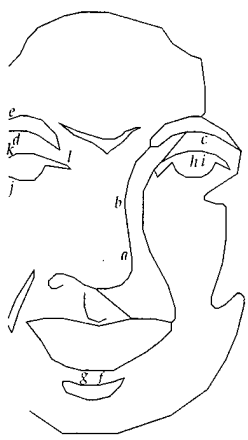


Figure 2. Phase one test scene. Letters represent image node locations for use with table 2.

Table 2. Performance of phase one algorithm with randomly selected state space trajectories. Table is an array of predicted feature nodes for figure 1 to 2 pairing. Included is model performance with a varying $\mathcal{Y}$-axis scale factor for memory image of figure 1.

	Feature designation†					
$\mathcal{Y}$*	1	2	3	4	5	6
1.00	*j*	*h*	*e*	*c*	*b*	*f*
1.00	*h*	*j*	*c*	*e*	*b*	*f*
1.00	*h*	*j*	*c*	*d*	*a*	*f*
1.00	*h*	*j*	*c*	*e*	*a*	*f*
1.00	*j*	*h*	*d*	*c*	*b*	*g*
1.00	*j*	*h*	*e*	*c*	*b*	*f*
1.00	*j*	*h*	*e*	*c*	*b*	*f*
1.00	*j*	*h*	*e*	*c*	*b*	*f*
1.00	*j*	*i*	*d*	*c*	*b*	*f*
0.50	*i*	*j*	*c*	*d*	*b*	*f*
0.75	*j*	*h*	*e*	*c*	*b*	*f*
1.25	*k*	*h*	*e*	*c*	*b*	*f*
1.50	*l*	*h*	*d*	*c*	*a*	*f*

*$\mathcal{Y}$-axis scale factor for figure 1.
†Taken from figure 1.

4. Phase Two—Algorithm Development

Because the global search algorithm uses so little information of necessity, it is unable to reliably determine if a valid image has been identified. What has been determined is the size, location, and orientation of a reputed image. What is needed now is a more detailed examination of the image candidate. A further search need not be restricted to the features used in phase one. Features can be redefined to yield more detailed morphological information. The scene search for phase two can be restricted to the region about each feature, which greatly reduces the computational requirements of the problem. The approach used for the redefined feature search parallels that used for the phase one global search. Based on the results of phase one, a shape-preserving affine transformation is performed on the memory image coordinate system. This attempts to optimize the congruence between the image characteristics in memory and the candidate image in the unknown scene. The following eight equations are substituted for equations (7) through (18):

$$F_1^{a,b,c} = s_1^{a,b} s_1^{a,c}, \tag{22}$$

where

$s_1^{a,b} = d_{K,N}/r_{a,b}$, if $s_1^{a,b} > 1 :\ s_1^{a,b} = 1/s_1^{a,b}$,

$s_1^{a,c} = d_{K,M}/r_{a,c}$, if $s_1^{a,c} > 1 :\ s_1^{a,c} = 1/s_1^{a,c}$,

K, N, M = feature designation ($K \neq N \neq M$),

a, b, c = corresponding candidate feature nodes from unknown scene,

d = (phase one) scaled memory image distance, and

r = corresponding unknown scene node-node distance.

$$F_2^{a,b,c} = E_1^{a,b,c}, \tag{23}$$
$$F_3^{a,b} = E_2^{a,b}, \tag{24}$$
$$F_4^{a,c} = E_2^{a,c}, \tag{25}$$
$$F_5^{a,b} = E_3^{a,b}, \tag{26}$$
$$F_6^{a,c} = E_3^{a,c}, \text{and} \tag{27}$$
$$F_7^{a} = \mathcal{D}_1^a/\mathcal{D}_2^a, \tag{28}$$

where

if $F_7^a > 1 :\ F_7^a = 1/F_7^a$,

$\mathcal{D}_1^a = ((X_K - x_a)^2 + (Y_K - y_a)^2)^{0.5}$,

$\mathcal{D}_2^a = (5000\mathcal{D}_2^{a'} + \mathcal{D}_1^a)/5001$,

$\mathcal{D}_2^{a'}$ = most recently computed value of $\mathcal{D}_2^a$ in the phase two iterated solution,

X_K, Y_K = internal memory coordinates of feature K after phase one scaling, translation, and rotation, and

x_a, y_a = coordinates of node a in the unknown scene coordinate system.

$$F_8^a = (X_K^t x_a^t + Y_K^t y_a^t)^2, \tag{29}$$

where

X_K^t, Y_K^t = the components of the unit tangent vector associated with each feature stored in memory after phase one rotation, and

x_a^t, y_a^t = the unit tangent vector components of the contour at node a in the unknown scene coordinate system.

These equations include concepts based on both relative and absolute relationships between unknown scene and transformed memory.

For equation (19) I substitute

$$\Lambda_m^a = -\xi_m^b \xi_m^c (F_1^{a,b,c} + F_2^{a,b,c} + F_3^{a,b} F_4^{a,c} + F_7^a + F_5^{a,b} F_6^{a,c} F_8^a). \tag{30}$$

The extended size of the phase two equation set—a direct consequence of the presence of the phase one results—allows a more sophisticated analysis of the image candidate. The goal is, as for phase one, to find a single node associated with each feature. This more detailed phase two analysis is important. In phase one, the tendency is to select features that are well distributed around the image. In phase two, of necessity features become more closely spaced, a potential source of performance degradation. The extended phase two equation set alleviates this somewhat.

The phase one and phase two algorithms are not only capable of generalizing to large variations in image shape but are also capable of a successful search even when features are missing. The algorithm is designed so that deleting features from an image (with the consequent loss of image information) will degrade algorithm performance, but this degradation should be reasonably graceful. This is an important attribute of any image-recognition algorithm.

For phase two, the canonical test image is a segment of a face—the region around the eye (fig. 3). This figure is interpreted identically to figure 1. Figure 4 is the scene to be searched for an occurrence of the figure 3 image. This image is not taken from figure 2; for demonstration purposes, something more complicated was desired. Note that the relative shape of the figure 3 image makes a poor fit to the obvious best (or proper) fit of figure 4.

Table 3 gives the output of the phase two algorithm. It lists several candidates for the best mode for each feature. These are listed in order with the best for each feature first. In addition, table 3 contains a figure of merit for the best fit node for each of the six features. This number is related to the output of equation (30). The smaller the number, the better the feature match. Note that the two nodes that make the poorest subjective match (5 and 6) also have the largest values.

Figure 3. Canonical image for use with demonstration of phase two algorithm.

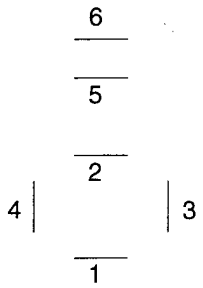


Figure 4. Phase two test scene. Letters represent image node locations for use with table 3.

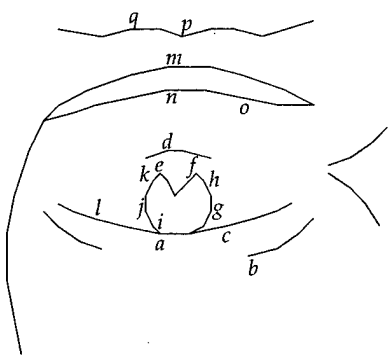


Table 3. Phase two results: best node-feature matches. Canonical shape in figure 3 and test scene in figure 4.

	Feature designation*					
	1	2	3	4	5	6
	a	d	g	j	m	p
$\mathcal{A}^\dagger$	b	e	h	k	n	q
	c	f	i	l	o	m
$\mathcal{M}^\ddagger$	0.46	0.56	0.50	0.47	0.59	0.62

*See figure 3.
†Best node fits, in order.
‡Phase two figure of merit.

5. Phase Three—Algorithm Development

Up to this point, all image identification is based on identifying a single node with a feature. In the final phase of the image-recognition algorithm, intrinsic feature shape is considered. The approach is tied to the way such information is stored in memory. This, in turn, is driven by the desire to store as little information as possible. At this point, what is stored?

- A single point identifying the location of each feature, and
- A unit tangent vector defining the orientation of the feature.

Even a feature without an overall orientation (e.g., a square) can be defined by associating the feature-identifying node with the appropriate subset of feature nodes. What is the simplest way of describing a generic shape? Remember that intrinsic feature shape is an elusive concept. Only infrequently can an image stored in memory be expected to be a good match to the content of a scene. Consider figure 5. Nodes A through F are feature-defining nodes. Now consider only node A. Store in memory the distances A to B and A to C. The tangent at A is already stored. Store in memory the distances B to D and C to E. Assume that curve segments A to D and A to E can be approximated by the relationship R^N. R is the distance from A along straight line segments A to B or A to C. Distances A to B and A to C can be conveniently normalized. Store N in memory. Curve E to A to D can be approximated from this limited information. If desired (and there are compelling reasons to do so), curve E to F to D can also be associated with node A and its shape inferred from similar considerations as given. From the limited information above, an approximation to the shape of figure 5 can be constructed.

A set of mathematical standards needs to be developed, as was done in phases one and two, to examine the intrinsic shape of all features. With the phase one and two efforts, a "best fit" single node is associated with each image feature. For the phase three effort, an association will be made between these feature nodes and the underlying feature shapes. An objective decision can then be made about the presence or absence of each feature. This is not intended to be template matching. I must assume that the fit between memory and feature candidate is poor, but subjectively acceptable. I must also assume that, in the presence of extraneous contours, a subjective best match can be found. The following equations, similar to what was

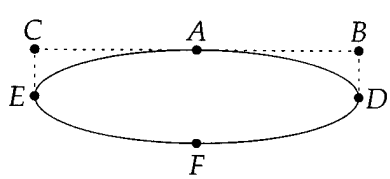


Figure 5. Exemplar configuration.

presented earlier, were found to be adequate:

$$G_1^{b,c} = 0.5 \mid Q_1^{1,N,M} - q_1^{1,b,c} \mid, \text{and} \tag{31}$$

$$G_2^{b,c} = (D_{1,N} d_{1,c})/(D_{1,M} d_{1,b}), \tag{32}$$

if $G_2^{b,c} > 1 : \; G_2^{b,c} = 1/G_2^{b,c}$, and

$$G_3^{b,c} = 1 - \mathcal{Z}_1 \mathcal{Z}_b \mathcal{Z}_c, \tag{33}$$

where

$\mathcal{Z}_1 = \mid X_1^t x_1^t + Y_1^t y_1^t \mid,$

$\mathcal{Z}_b = \mid X_N^t x_b^t + Y_N^t y_b^t \mid$, and

$\mathcal{Z}_c = \mid X_M^t x_c^t + Y_M^t y_c^t \mid.$

$$G_4^b = 0.5 \mid Q_3^{1,N} - q_3^{1,b} \mid, \text{and} \tag{34}$$

$$G_5^b = (\mid (x_1 - x_b)(X_1 - X_N) \mid) + (\mid (y_1 - y_b)(Y_1 - Y_N) \mid)/r_{1,b} d_{1,N}. \tag{35}$$

See equations (22) and (28) for the definition of terms. The next standard G_6 is different. If node b shares a common contour with node a, then

$$G_6^b = 0 \; ; \quad \text{else} \quad G_6^b = 1. \tag{36}$$

$$G_7^b = d_{1,N}/r_{1,b} \tag{37}$$

or the inverse, whichever is largest.

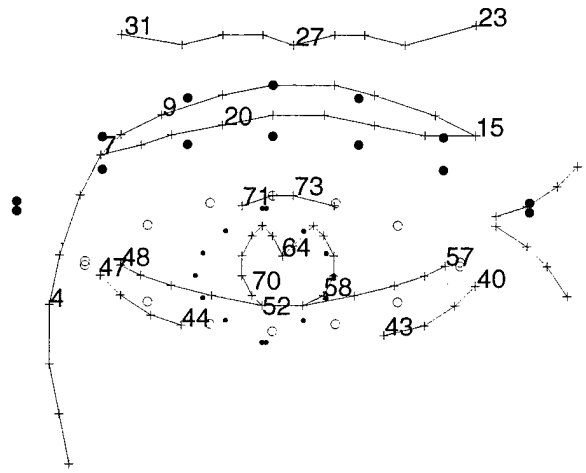
One can see that G_1 through G_7 are essentially either the terms or variations on the terms presented for phases one and two. The optimization procedure finds that set of 14 nodes for which the following relationship is minimized:

$$\Lambda_m^1 = G_1^{b,c} + G_2^{b,c} + G_3^{b,c} + G_4^b + G_5^b + G_6^b + G_7^b, \tag{38}$$

where $m = 1, \ldots 14$.

The performance of equation (38), applied to figure 4 and the results in table 3, is demonstrated with the help of figure 6. A feature extracted from memory is represented by 14 nodes (an arbitrary choice) distributed along its contour (or contours). Based on the results of phases one and two, a scale factor and an orientation for all features are established. From table 3, a scene-node candidate for node 1 of the memory-generated feature is selected. This memory-generated feature is translated until its node 1 is congruent with the selected node from table 3. In figure 6, this is done for three features. The solid circles define feature 6 (see fig. 3), and node 1 of this feature is placed congruent with node 11 of the eye-eyebrow scene (node 11 is identical to node m of fig. 4 and table 3). Similarly, the open circles define feature 2 and are placed congruent with scene node 72 (node d of fig. 4 and table 3). The dots are feature 3 and are placed congruent with scene node 59 (node g of fig. 4 and table 3). All remaining 13 feature nodes are numbered

Figure 6. Phase three test configuration for use with table 4.



sequentially counterclockwise from 1. Equation (38) is applied with a reasonably localized search around the 14 memory-derived feature nodes and finds its corresponding optimum match from the test scene. A numerical rating proportional to the result of equation (38) is generated and stored and provides an absolute scale basis for establishing the confidence in the node match.

Table 4 presents phase three results for those nodes of table 3 that best satisfy the requirements of equation (38). The phase three algorithm assigns a best candidate from the scene for all 14 nodes, even when all or part of the scene feature is missing. Examples of such poor "hits" can be readily discerned in table 4. While a simple thresholding of the data would eliminate the poor nodes, this is not the best solution to this problem. As can be seen from the phase three equation set, the poor data points will degrade the overall performance of the model. To eliminate poor data requires an iterative approach with these points progressively suppressed from the equation set.

Nevertheless, the performance of the simple phase three model implementation is impressive and demonstrates the adequacy of the approach.

Table 4. Phase three results: best node-feature matches.

	Node† / performance‡													
$\mathcal{B}^*$	1	2	3	4	5	6	7	8	9	10	11	12	13	14
1-a	52	58	55	57	57	61	62	66	21	7	48	48	49	51
	0.66	0.86	0.66	0.49	1.14	1.30	1.18	1.31	1.01	1.04	1.46	0.75	0.68	0.52
2-f	62	67	50	49	49	51	52	53	54	55	57	57	57	57
	1.01	1.02	1.58	1.08	0.61	0.78	0.73	0.70	0.75	0.87	0.28	0.91	1.37	1.79
3-g	59	60	62	66	66	67	68	69	70	52	52	53	53	58
	0.58	0.66	0.95	1.19	0.84	0.71	0.67	0.67	0.73	0.54	0.91	0.90	0.59	0.68
4-j	69	70	52	53	53	58	59	59	60	61	62	62	66	68
	0.54	0.62	0.58	1.01	0.64	0.63	0.64	0.66	0.63	0.66	1.13	0.96	0.64	0.63
5-n	19	17	15	15	15	14	12	11	10	8	7	7	7	21
	0.39	0.58	0.66	0.54	0.59	0.68	0.62	0.48	0.64	0.61	0.52	0.47	0.50	0.66
6-m	11	10	8	7	7	7	21	19	17	16	15	15	16	13
	0.46	0.61	0.59	0.52	0.46	0.50	0.67	0.44	0.49	0.51	0.64	0.66	0.81	0.49

*Feature node (see figs. 3 and 4 for definitions).
†See figure 6 for scene node-designation scheme.
‡Phase three figure of merit (proportional to eq (32)).

6. Conclusion

This report demonstrates the feasibility of performing a sophisticated scene search for a complex object. The search is intended to both recognize the occurrence of an object and to create a labeled subset of the search-scene nodes in preparing to answer the question: What does the object look like? This report develops the theory for the object search and presents the results of a series of computer simulations.

References

Aibara, T., K. Ohue, and Y. Matsuoka (1991). "Human face recognition by P-type Fourier descriptor," *Visual Communications and Image Processing*, Proc. SPIE **1606**, pp 198–203.

Akamatsu, S., T. Sasaki, N. Masui, H. Fukamachi, and Y. Suenaga, "A new method for designing face image classifiers using 2D CG model," *Visual Communications and Image Processing*, Proc. SPIE **1606** (1991), pp 204–216.

Allinson, N. M., and A. W. Ellis, "Face recognition: Combining cognitive psychology and image engineering," *Electron. Commun. Eng. J.* **4**, No. 5 (1992), pp 291–300.

Asada, H., and M. Brody, "The curvature primal sketch," *IEEE Trans. Pattern Analysis and Machine Intelligence* **PAMI-8**, No. 1 (1986), pp 2–14.

Bigun, J., and G. H. Granlund, "Optimal orientation detection of linear symmetry," *Proc. 1st International Conference Computer Vision* **1987** (1987), pp 433–438.

Bouattour, H., F. F. Soulie, and E. Viennet, "Neural nets for human face recognition," *Proc. International Joint Conference on Neural Networks* **3** (1992), pp 700–704.

Brunelli, R., and T. Poggio, "Face recognition: Features versus templates," *IEEE Trans. Pattern Analysis and Machine Intelligence* **15**, No. 10 (1993).

Brunelli, R., and T. Poggio, "Face recognition through geometrical features," *Proc. Second European Conference on Computer Vision* (1992a), pp 792–800.

Brunelli, R., and T. Poggio, "Caricatural effects in automated face perception," preprint (1992b).

Ellis, H. D., "Recognizing faces," *Br. J. Psychol* **66**, No. 4 (1975), pp 409–426.

Frasconi, P., M. Gori, G. Soda, A. De Curtis, and F. Innocenti, "Face recognition using multi-layered perceptrons which also reject never seen faces," *Proc. Fifth Italian Workshop–Neural Nets* (1992), pp 286–292.

Goble, J. R., "Face recognition using the discrete cosine transform," thesis, Air Force Institute of Technology, Wright Patterson AFB (1991).

Gordon, G. G., "Face recognition based on depth and curvature features," *Proc. IEEE Computer Society Conference on Computer Vision and Pattern Recognition* (1992), pp 808–810.

Gordon, G. G., and L. Vincent, "Application of morphology to feature extraction for face recognition," *Nonlinear Image Processing III*, Proc. SPIE **1658** (1992), pp 151–163.

Grimson, W.E.L., "On the recognition of curved objects," *IEEE Trans. Pattern Analysis and Machine Intelligence* **II**, No. 6 (1989), pp 632–643.

Hancock, P.J.B., and L. S. Smith, "GANNET: Genetic design of a neural net for face recognition," *Proc. Parallel Problem Solving from Nature*, 1st Workshop (1990), pp 292–296.

Harmon, L. D., "Automatic recognition of human face profiles," *Proc. 3rd International Joint Conference–Pattern Recognition* (1976), pp 183–188.

Harmon, L. D., S. C. Kuo, P. F. Ramig, and U. Raudkivi, "Identification of human face profile by computer," *Pattern Recognition* **10** (1978), pp 301–312.

Heitger, F., G. Gerig, L. Rosenthalter, and O. Kubler, "Extraction of boundary key features and spontaneous figure completion," *Proc. 6th Scandinavian Conference on Image Analysis* (1989), pp 1090–1097.

Jia, X., and M. S. Nixon, "Profile feature extraction via the Walsh transform for face recognition," *Intelligent Robots and Computer Vision XI*, Proc. SPIE **1824** (1992), pp 46–52.

Kanade, T., "Computer recognition of human faces," Ph.D. dissertation, Kyoto University, Kyoto, Japan (1973).

Kaufman, G. J., and K. J. Breeding, "The automatic recognition of human faces from profile silhouettes," *IEEE Trans. Syst., Man, Cybern.* **SMC-6**, No. 2 (1976).

Kerin, M. A., and T. J. Stonham, "Face recognition using a digital neural network with self-organizing capabilities," *Proc. 10th International Conference on Pattern Recognition* **1** (1990), pp 738–741.

Kirby, M., and L. Sirovich, "Application of the Karhunen-Loeve procedure for the characterization of human faces," *IEEE Trans. Pattern Analysis and Machine Intelligence* **12**, No. 1 (1990), pp 103–108.

Krepp, D. L., "Face recognition with neural networks," thesis, Air Force Institute of Technology, Wright Patterson AFB (1992).

Lambert, L. C., "Evaluation and enhancement of the AFIT autonomous face recognition machine," thesis, Air Force Institute of Technology, Wright Patterson AFB (1987).

Lim, K., Y. Sim, and K. Oh, "A face recognition system using fuzzy logic and artificial neural network," *Proc. IEEE International Conference on Fuzzy Systems* (1992), pp 1063–1069.

Manjunath, B. S., R. Chellappa, and C. von der Maldburg, "A feature based approach to face recognition," *Proc. IEEE Computer Society Conference on Computer Vision and Pattern Recognition* (1992), pp 373–378.

Nakamura, O., S. Mathur, and T. Minami, "Identification of human faces based on isodensity maps," *Pattern Recognition* **24**, No. 3 (1991), pp 263–272.

Petkov, N., P. Kruizinga, and T. Lourens, "Biologically motivated approach to face recognition," *Proc. Workshop on Artificial Neural Networks* (1993), pp 68–77.

Resnikoff, H. L., *The Illusion of Reality*, New York, Springer-Verlag (1989).

Robb, B. C., "Autonomous face recognition machine using a Fourier feature set," thesis, Air Force Institute of Technology, Wright Patterson AFB (1989).

Runyon, K. R., "Automated face recognition system," thesis, Air Force Institute of Technology, Wright Patterson AFB (1992).

Russel, R. L., "Performance of a working face recognition machine using cortical thought theory," thesis, Air Force Institute of Technology, Wright Patterson AFB (1984).

Russel, R. L., R. L. Routh, J. R. Holton, and M. Kabriski, "A face recognition system based on cortical thought theory," *Proc. IEEE National Aerospace and Electronics Conference* **4** (1986), pp 1377–1385.

Sander, D. D., "Enhanced autonomous face recognition machine," thesis, Air Force Institute of Technology, Wright Patterson AFB (1988).

Seitz, P., "Using local orientational information as image primitive for robust object recognition," *Visual Communications and Image Processing IV*, Proc. SPIE **1199** (1989), pp 1630–1639.

Seitz, P., and M. Bichsel, "The digital doorkeeper–automatic face recognition with the computer," *Proc. IEEE International Carnahan Conference on Security Technology* (1991), pp 77–83.

Sirovich, L., and M. Kirby, "Low-dimensional procedure for the characterization of human faces," *J. Opt. Soc. Am. A* **4**, No. 3 (1987), pp 519–524.

Smith, E. J., "Development of an autonomous face recognition machine," thesis, Air Force Institute of Technology, Wright Patterson AFB (1986).

Soulie, F. F., E. Viennet, and B. Lamy, "Multi-modular neural network architectures: Applications in optical character and human face recognition," *International Journal of Pattern Recognition and Artificial Intelligence* **7**, No. 4 (1993), pp 721–755.

Spacek, L., M. Kubat, and D. Flotzinger, "Face recognition through learned boundary characteristics," *Appl. Artificial Intelligence* **8** (1994), pp 149–164.

Stringa, L., "Eyes detection for face recognition," *Appl. Artificial Intelligence* **7** (1993), pp 365–382.

Suarez, P. F., "Face recognition with the Karhunen-Loeve transform," thesis, Air Force Institute of Technology, Wright Patterson AFB (1991).

Sutherland, K., D. Renshaw, and P. B. Denyer, "Automatic face recognition," *Proc. First International Conference on Intelligent Systems Engineering* (1992), pp 29–34.

Turk, M., and A. Pentland, "Eigenfaces for recognition," *Journal of Cognitive Neurosciences* **3**, No. 1 (1991), pp 71–88.

Wong, R. Y., and J. Calia, "PC-based human face recognition system," *Proc. 34th Midwest Symposium on Circuits and Systems* **2** (1992), pp 641–644.

Wu, C. J., and J. S. Huang, "Human face profile recognition by computer," *Pattern Recognition* **23**, No. 3/4 (1990), pp 255–259.

Yow, K. C., and R. Cipolla, "Feature-based human face detection," *Image and Vision Computing* **15** (1997), pp 713–735.

Distribution

Admnstr
Defns Techl Info Ctr
ATTN DTIC-OCP
8725 John J Kingman Rd Ste 0944
FT Belvoir VA 22060-6218

DARPA
ATTN S Welby
3701 N Fairfax Dr
Arlington VA 22203-1714

Ofc of the Secy of Defns
ATTN ODDRE (R&AT)
The Pentagon
Washington DC 20301-3080

Ofc of the Secy of Defns
ATTN OUSD(A&T)/ODDR&E(R) R J Trew
3080 Defense Pentagon
Washington DC 20301-7100

AMCOM MRDEC
ATTN AMSMI-RD W C McCorkle
Redstone Arsenal AL 35898-5240

US Army TRADOC
Battle Lab Integration & Techl Dirctrt
ATTN ATCD-B
ATTN ATCD-B J A Klevecz
FT Monroe VA 23651-5850

NVESD
ATTN AMSEL-RD-NVOD L Garn
10221 Burbeck Rd Ste 430
FT Belvoir VA 22060-5806

Dir for MANPRINT
Ofc of the Deputy Chief of Staff for Prsnnl
ATTN J Hiller
The Pentagon Rm 2C733
Washington DC 20301-0300

SMC/CZA
2435 Vela Way Ste 1613
El Segundo CA 90245-5500

TECOM
ATTN AMSTE-CL
Aberdeen Proving Ground MD 21005-5057

US Army ARDEC
ATTN AMSTA-AR-TD
Picatinny Arsenal NJ 07806-5000

US Army Info Sys Engrg Cmnd
ATTN AMSEL-IE-TD F Jenia
FT Huachuca AZ 85613-5300

US Army Natick RDEC Acting Techl Dir
ATTN SBCN-T P Brandler
Natick MA 01760-5002

US Army Simulation Train & Instrmntn Cmnd
ATTN AMSTI-CG M Macedonia
ATTN J Stahl
12350 Research Parkway
Orlando FL 32826-3726

US Army Soldier & Biol Chem Cmnd
Dir of Rsrch & Techlgy Dirctrt
ATTN SMCCR-RS I G Resnick
Aberdeen Proving Ground MD 21010-5423

US Army Tank-Automtv Cmnd RDEC
ATTN AMSTA-TR J Chapin
Warren MI 48397-5000

Nav Surfc Warfare Ctr
ATTN Code B07 J Pennella
17320 Dahlgren Rd Bldg 1470 Rm 1101
Dahlgren VA 22448-5100

Hicks & Assoc Inc
ATTN G Singley III
1710 Goodrich Dr Ste 1300
McLean VA 22102

Palisades Inst for Rsrch Svc Inc
ATTN E Carr
1745 Jefferson Davis Hwy Ste 500
Arlington VA 22202-3402

Director
US Army Rsrch Ofc
ATTN AMSRL-RO-D JCI Chang
ATTN AMSRL-RO-EN W D Bach
PO Box 12211
Research Triangle Park NC 27709

Distribution (cont'd)

US Army Rsrch Lab
ATTN AMSRL-D D R Smith
ATTN AMSRL-DD J M Miller
ATTN AMSRL-CI-AI-R Mail & Records Mgmt
ATTN AMSRL-CI-AP Techl Pub (3 copies)
ATTN AMSRL-CI-LL Techl Lib (3 copies)

US Army Rsrch Lab (cont'd)
ATTN AMSRL-IS-EP P Gillespie
ATTN AMSRL-SE J Pellegrino
ATTN AMSRL-SE-SA J Eicke
ATTN AMSRL-SE-SE M Vrabel (10 copies)
Adelphi MD 20783-1197

REPORT DOCUMENTATION PAGE

Form Approved
OMB No. 0704-0188

Public reporting burden for this collection of information is estimated to average 1 hour per response, including the time for reviewing instructions, searching existing data sources, gathering and maintaining the data needed, and completing and reviewing the collection of information. Send comments regarding this burden estimate or any other aspect of this collection of information, including suggestions for reducing this burden, to Washington Headquarters Services, Directorate for Information Operations and Reports, 1215 Jefferson Davis Highway, Suite 1204, Arlington, VA 22202-4302, and to the Office of Management and Budget, Paperwork Reduction Project (0704-0188), Washington, DC 20503.

1. AGENCY USE ONLY *(Leave blank)*

2. REPORT DATE
December 2000

3. REPORT TYPE AND DATES COVERED
Final, 1998–1999

4. TITLE AND SUBTITLE An Approach to Feature-Based Face Recognition

5. FUNDING NUMBERS
DA PR: A305
PE: 61102A

6. AUTHOR(S) Michael J. Vrabel

7. PERFORMING ORGANIZATION NAME(S) AND ADDRESS(ES)
U.S. Army Research Laboratory
Attn: AMSRL-SE-SE email: vrabel@arl.army.mil
2800 Powder Mill Road
Adelphi, MD 20783-1197

8. PERFORMING ORGANIZATION REPORT NUMBER
ARL-TR-2215

9. SPONSORING/MONITORING AGENCY NAME(S) AND ADDRESS(ES)
U.S. Army Research Laboratory
2800 Powder Mill Road
Adelphi, MD 20783-1197

10. SPONSORING/MONITORING AGENCY REPORT NUMBER

11. SUPPLEMENTARY NOTES
ARL PR: 9NEOMM
AMS code: 611102.305

12a. DISTRIBUTION/AVAILABILITY STATEMENT Approved for public release; distribution unlimited.

12b. DISTRIBUTION CODE

13. ABSTRACT *(Maximum 200 words)*
I detail a scheme for searching an unknown scene for occurrences of an object. The approach is independent of object size, location, and orientation and is tolerant of significant changes in object shape and appearance.

14. SUBJECT TERMS Image recognition, scene analysis, automatic target recognition (ATR)

15. NUMBER OF PAGES
47

16. PRICE CODE

17. SECURITY CLASSIFICATION OF REPORT
Unclassified

18. SECURITY CLASSIFICATION OF THIS PAGE
Unclassified

19. SECURITY CLASSIFICATION OF ABSTRACT
Unclassified

20. LIMITATION OF ABSTRACT
UL

NSN 7540-01-280-5500

Standard Form 298 (Rev. 2-89)
Prescribed by ANSI Std. Z39-18
298-102