INVITED PAPER *Special Issue on Machine Vision Applications*

A Random Walk through Eigenspace

Matthew TURK[†], ***Nonmember***

SUMMARY It has been over a decade since the "Eigenfaces" approach to automatic face recognition, and other appearance-based methods, made an impression on the computer vision research community and helped spur interest in vision systems being used to support biometrics and human-computer interface. In this paper I give a personal view of the original motivation for the work, some of the strengths and limitation of the approach, and progress in the years since. Appearance-based approaches to recognition complement feature- or shape-based approaches, and a practical face recognition system should have elements of both. Eigenfaces is not a general approach to recognition, but rather one tool out of many to be applied and evaluated in the appropriate context.

key words: *face recognition, eigenfaces, computer vision*

1. Introduction

It is often observed that the human ability to recognize faces is remarkable. Faces are complex visual stimuli, not easily described by simple shapes or patterns; yet people have the ability to recognize familiar faces at a glance after years of separation. Lest we marvel too much at human performance, it should also be noted that the inability to recognize a face is sometimes remarkable as well. Quite often we strain to see the resemblance between a picture (e.g., a driver's license photo) and the real person, and sometimes we are greeted in a friendly, familiar manner by someone we do not remember ever seeing before. Although face recognition in humans may be impressive, it is far from perfect. Yet there is something about the perception of faces that is very fundamental to the human experience. Early in life we learn to associate faces with pleasure, fulfillment, and security. As we get older, the subtleties of facial expression enhance our explicit communication in myriad ways. The face is our primary focus of attention in social intercourse; this can be observed in interaction among animals as well as between humans and animals. The face, more than any other part of the body, communicates identity, emotion, race, and age, and is also quite useful for judging gender, size, and perhaps even character.

The subject of visual processing of human faces has received attention from philosophers and scientists such as Aristotle and Darwin for centuries. The ability of a person to recognize another person (e.g., a mate, a child, or an enemy) is important for many reasons. Recognition is not only visual; it may occur through a variety of sensory modalities, including sound, touch, and even smell. For people, however, the most reliable and accessible modality for recognition is the sense of sight. Using vision, a person may be recognized by one's face, but also by one's clothing, hairstyle, gait, silhouette, hands, etc. People often distinguish animals not by their faces but by characteristic markings on their bodies. Similarly, the human face is not the only, and may not even be the primary, visual characteristic used for person identification. For example, in a home or office setting, the person's face may be used merely in verifying identity, after identity has already been established based on other factors such as clothing, hairstyle, or a distinctive moustache. Indeed, the identification of humans may be viewed as a Bayesian classification system, with prior probabilities on several relevant random variables. For example, a parent is predisposed to recognize his child if, immediately prior to contact, he sees a school bus drive by and then hears yelling and familiar light footsteps. Nevertheless, because faces are so important in human interaction, no other avenue to person identification is as compelling as face recognition.

There has been a good deal of investigation into human face recognition performance, seeking to understand and characterize the representations and processes involved. Face-specific cells (cells that appear to respond selectively to the presence of faces) have been found in monkeys and sheep [1]–[3]. Prosopagnosia, the specific inability to recognize faces, has been identified and studied in human patients. There have been many interesting studies in experimental and developmental psychology that have probed the limits of human face recognition, suggesting models and constraints on representation and processing (e.g., [4]–[6]). Nevertheless, it is still the case that a thorough understanding of how humans (and animals) represent, process, and recognize faces remains a distant goal. Although studies of face recognition in physiology, neurology, and psychology provide insight into the problem of face recognition, they have yet to provide substantial practical guidance for computer vision systems in this area.

What does it mean to recognize a face? There are several aspects of recognizing human identity and processing facial information that make the problem some-

Manuscript received June 22, 2001.

[†] The author is with the Computer Science Department, University of California, Santa Barbara, CA 93106 USA.

what ill-defined. As mentioned above, recognition of a person's identity is not necessarily (and perhaps rarely) a function of viewing the person's face in isolation. In addition, face recognition is closely related to face (and head and body) detection, face tracking, and facial expression analysis. Figure 1 shows a few typical engineering approaches to the overall problem. In the first example, a face is initially detected, then recognized. In the second example, detection and recognition are performed in tandem; detection is merely a successful recognition. In the third example, facial feature tracking is performed and expression analysis occurs before attempting to recognize the normalized (expressionless) face. There are, of course, many additional variations possible.

Just as the human task of face recognition is neither clearly defined nor clearly differentiated from related tasks, automatic face recognition by computers is not a single defined problem. Face recognition systems may be useful in several contexts, for example:

- Given a database of standard face images (e.g., criminal mug shots), determine whether or not a new mug shot is of one of the people in the database.
- In the same situation, determine possible identity when the new image originates from a completely different source (e.g., a surveillance camera at a bank), with different (and probably unknown) imaging conditions.
- Identify the new computer user as one of the registered users in order to allow login access.
- Determine that a face is present in an image, at a particular location and scale, in order to correctly color balance the image, or to compress the image properly.

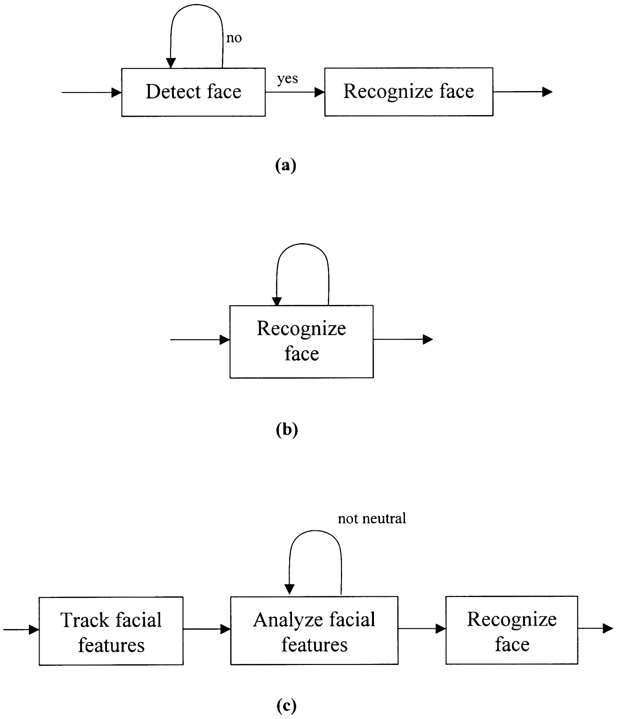

Fig. 1 Typical approaches to face recognition.

For the purposes of this paper, "face recognition" and "face identification" describe the same task[†]. That is, given an image of a human face, classify that face as one of the individuals whose identity is already known by the system, or perhaps as an unknown face. "Face detection" means detecting the presence of any face, regardless of identity. "Face location" is specifying the 2D position (and perhaps orientation) of a face in the image. "Face tracking" is updating the (2D or 3D) location of the face. "Facial feature tracking" is updating the (2D or 3D) locations, and perhaps the parameterized descriptions, of individual facial features. "Face pose estimation" is determining the position and orientation (usually 6 degrees of freedom) of a face. "Facial expression analysis" is computing parametric, and perhaps also symbolic, descriptions of facial deformations.

2. Recognition Strategies

Object recognition has long been a primary goal of computer vision, and it has turned out to be a very difficult endeavor. The primary difficulty in attempting to recognize objects from imagery comes from the immense variability of object appearance due to several factors, which are all confounded in the image data. Shape and reflectance are intrinsic properties of an object, but an image of the object is a function of several other factors, including the illumination, the viewpoint of the camera (or, equivalently, the pose of the object), and various imaging parameters such as aperture, exposure time, lens aberrations and sensor spectral response. Object recognition in computer vision has been dominated by attempts to infer from images information about objects that is relatively invariant to these sources of image variation. In the Marr paradigm [7], the prototype of this approach, the first stage of processing extracts intrinsic information from images; i.e., image features such as edges that are likely to be caused by surface reflectance changes or discontinuities in surface depth or orientation. The second stage continues to abstract away from the particular image values, inferring surface properties such as orientation and depth from the earlier stage. In the final stage, an object is represented as a three dimensional shape in its own coordinate frame, completely removed from the intensity values of the original image.

[†] It is often useful to distinguish between classifying as belonging to the general class of objects ("recognition") and labeling as a particular member of the class ("identification"), but we will follow the common terminology and use the terms interchangeably, with the precise meaning depending on the context.

This general approach to recognition can be contrasted with *appearance-based* approaches, such as correlation, which matches image data directly. These approaches tend to be much easier to implement than methods based on object shape—correlation only requires a stored image of the object, while a full 3D shape model is very difficult to compute—but they tend to be very specific to an imaging condition. If the lighting, viewpoint, or anything else of significance changes, the old image template is likely to be useless for recognition.

The idea of using pixel values, rather than features that are more invariant to changes in lighting and other variations in imaging conditions, was counter-intuitive to many. After all, the whole point of the Marr paradigm [7] of vision was to abstract away from raw pixel values to higher level, invariant representations such as 3D shape. Mumford [8] illustrated some of these objections with a toy example: recognizing a widget that comprises a one-dimensional black line with one white dot somewhere on it. He shows that for this example the eigenspace is no more efficient than the image space, and a feature-based approach (where the feature is the position of the white dot) is a much simpler solution. This example, however, misses the point of the eigenfaces approach, which can be seen in the following counter-example.

Imagine starting with images of two different faces. They would typically differ in the precise location of facial features (eyes, nostrils, mouth corners, etc.) and in grayscale values throughout. Now warp one image so that all the extractable features of that face line up with those of the first face. (Warping consists of applying a two-dimensional motion vector to every pixel in the image and interpolating properly to avoid blank areas and aliasing.) The eyes line up, the noses line up, the mouth corners line up, etc. The feature-based description is now identical for both images. Do the images now look like the same person? Not at all—in many (perhaps most) cases the warped image is perceived as only slightly different from its original. Here is a case where an appearance-based approach will surely outperform a simple feature-based approach.

Soon after this toy example was introduced, Brunelli and Poggio [9] investigated generic feature-based and template-based approaches to face recognition and concluded that the template-based approach worked better, at least for their particular database of frontal view face images. Of course, both of these examples are extreme cases. A face is nothing like a black line with a white dot. Nor is the variation in facial feature locations and feature descriptions so small as to be insignificant. Clearly, both geometric and photometric information can be useful in recognizing faces.

Both strategies—*feature-based* methods and *appearance-based* methods—have more practical versions than these simple characterizations may indicate. In the feature-based approach it is not necessary to go all the way from image features to 3D shape. Alternatively, features can be compiled from 3D models of the known objects, so that the task is reduced to matching computed features with expected features. This approach has led to several strategies for selecting good features and performing efficient feature matching (e.g., [10] and [11]). Similarly, appearance-based approaches are not constrained to matching raw image with a single raw image template. Several techniques attempt to first eliminate some of the expected variation (e.g., in scale, orientation, and overall brightness level) before matching occurs, or to use multiple image templates for a given object.

In the past decade, learning has become a very significant issue in visual recognition. Rather than constructing 3D shape models or expected features manually, it would be a beneficial for a system to learn the models automatically. And rather than enumerating all the conditions that require new templates, it would be helpful for the system to analyze the imaging conditions to decide on optimal correlation templates, or to learn from a collection of images what attributes of appearance will be most effective in recognition. It is likely that no recognition system of any reasonable complexity—that is, no system that solves a non-trivial recognition problem—will work without learning as a central component. For learning to be effective, enough data must be acquired to allow a system to account for the various components of the images, those intrinsic to the object and otherwise.

The concept of robustness, i.e., stability in the presence of various types of noise and a reasonable quantity of outliers, has also become very important in computer vision in the past decade or more. System performance (e.g., recognition rate) should decrease gracefully as the amount of noise increases. Noise can come from many sources: thermal noise in the imaging process, noise added in transmission and storage, lens distortion, unexpected markings on an object's surface, occlusions, etc. An object recognition algorithm that requires perfect images will not work in practice, and the ability to characterize a system's performance in the presence of noise is vital.

Learning and robustness must also be balanced with practical speed requirements. Computational complexity can also limit the usefulness. Whether the task is offline, real-time, or the intermediate "interactive-time" (with a human in the loop), constraints on processing time are always an issue in practice.

Face recognition is an example of object recognition. As with most recognition tasks, the source images comprise pixel values that are influenced by several factors such as shape, reflectance, pose, occlusion, and illumination. The human face is an extremely complex object, with both rigid and non-rigid components that vary over time, sometimes quite rapidly. The object is covered with skin, a non-uniformly textured material that is difficult to model either geometrically or photometrically. Skin can change color quickly when one is embarrassed or becomes warm or cold. The reflectance properties of the skin can also change rather quickly, as perspiration level changes. The face is highly deformable, and facial expressions reveal a wide variety of possible configurations. Other time-varying changes include the growth and removal of facial hair, wrinkles and sagging of the skin brought about by aging, skin blemishes, and changes in skin color and texture caused by exposure to sun. Add to that the many common artifact-related changes, such as cuts and scrapes, bandages, makeup, jewelry and piercings, and it is quite clear that the human face is much more difficult to model (and thus recognize) than most industrial parts.

Partly because of this difficulty, face recognition has been considered a challenging problem in computer vision for some time, and the amount of effort in the research community devoted to this topic has increased significantly over the years.

3. A Brief History of Automated Face Recognition, Part One

Attempts to automate human face recognition by computers began in the late 1960s and early 1970s. Bledsoe [12] developed a system to automatically classify features extracted by human operators from face images. Kelly [13] and Kanade [14] built probably the first fully automated face recognition systems, extracting feature measurements from digitized images and classifying the feature vector. Harmon et al. [15], Gordon [16] and others have investigated using facial profiles (side views). Yuille et al. [17] and others have used deformable templates, parameterized models of features and sets of features with given spatial relations. Various approaches using neural networks (e.g., [18], [19]) have attempted to move away from purely feature-based methods. Moving beyond typical intensity images, Lapresté [20], Lee and Milios [21], Gordon [22] and others used range data to build and match models of faces and face features.

By the late 1980s, there had been several feature-based approaches to face recognition. For object recognition in general, the most common approach was to extract features from objects, build some sort of model from these features, and perform recognition by matching feature sets. Features, and the geometrical relationships among them, are stable under varying illumination conditions and pose—if they can be reliably calculated. However, it is often the case that they cannot, so the problem became more and more complex. Indexing schemes and other techniques were developed to cope with the inevitable noisy, spurious, and missing features.

In late 1987, as Sandy Pentland and I began to think about face recognition, we looked at the existing feature-based approaches and wondered if they were erring by discarding most of the image data. If extracting local features was at one extreme, what might be an effective way of experimenting with the other extreme, i.e., working with a global, holistic face representation? We began to build on work by Sirovich and Kirby [23] on coding face images using Principal Components Analysis (PCA). Around the same time, Burt [24] was developing a system for face recognition using pyramids, multiresolution face representations. The era of appearance-based approaches to face recognition had begun.

The "Eigenfaces" approach, based on PCA, was never intended to be the definitive solution to face recognition. Rather, it was an attempt to re-introduce the use of information "between the features"; that is, it was an attempt to swing back the pendulum somewhat to balance the attention to isolated features.

4. Image Space

Appearance-based approaches to vision often start with the concept of *image space*. A two-dimensional image $I(x, y)$ may be viewed as a point (or vector) in a very high dimensional space, called image space, where each coordinate of the space corresponds to a sample (pixel) of the image. For example, an image with 32 rows and 32 columns describes a point in a 1024-dimensional image space. In general, an image of r rows and c columns describes a point in N-dimensional image space, where $N = rc$. This representation obfuscates the neighborhood relationship (distance in the image plane) inherent in a two-dimensional image. That is, rearranging the pixels in the image (and changing neighborhood relationships) will have no practical effect on its image space representation, as long as all other images are identically rearranged. Spatial operations such as edge detection, linear filtering, and translation are not local operations in image space. A 3×3 spatial image filter is not an efficient operation in image space; it is accomplished by multiplication with a very large, sparse $N \times N$ matrix. On the other hand, the image space representation helps to clarify the relationships among collections of images.

With this image representation, the image becomes a very high dimensional "feature," and so one can use traditional feature-based methods in recognition. So, merely by considering an image as a vector, feature-based methods can be used to accomplish appearance-based recognition; that is, operations typically performed on feature vectors, such as clustering and distance metrics, can be performed on images directly. Of course, the high dimensionality of the image space makes many feature-based operations implausible, so they cannot be applied without some thought towards efficiency. As image resolution increases, so does the dimensionality of the image space. At the limit, a continuous image maps to an infinite-dimensional image space. Fortunately, many key calculations scale with the number of sample images rather than the dimensionality of the image space, allowing for efficiency even with relatively high resolution imagery.

If an image of an object is a point in image space, a collection of M images gives rise to M points in image space; these may be considered as samples of a probability distribution. One can imagine that all possible images of the object (under all lighting conditions, scales, etc.) define a manifold within the image space. How large is image space, and how large might a manifold be for a given object? To get an intuitive estimate of the vastness of image space, consider a tiny 8×8 binary (one bit) image. The number of image points (the number of distinct images) in this image space is 2^{64}. If a very fast computer could evaluate one billion images per second, it would take almost $600 years$ to exhaustively evaluate the space. For grayscale and color images of reasonable sizes, the corresponding numbers are unfathomably large. It is clear that recognition by exhaustively enumerating or searching image space is impossible.

This representation brings up a number of questions relevant to appearance-based object recognition. What is the relationship between points in image space that correspond to all images of a particular object, such as a human face? Is it possible to efficiently characterize this subset of all possible images? Can this subset be learned from a set of sample training images? What is the "shape" of this subset of image space?

Consider an image of an object to be recognized. This image $I(r, c)$ is a point x in image space, or, equivalently, a feature in a high-dimensional feature space. The image pixel $I(r, c)$ can be mapped to the ith component of the image point (x_i) by $i = r \cdot width + c$. A straightforward pattern classification approach to recognition involves determining the minimal distance between a new face image x and pre-existing face classes $\tilde{x}$. That is, given k prototype images of known objects, find the prototype $\tilde{x}$ that satisfies

$$\min_i d(x, \tilde{x}_i), \quad i = 1, \ldots, k$$

A common distance metric is merely the Euclidian distance in the feature space,

$$d(x_1, x_2) = \|x_1 - x_2\|$$
$$= \sqrt{(x_1 - x_2)^T (x_1 - x_2)} = \sqrt{\sum_{i=1}^{rc} (x_{1i} - x_{2i})^2}$$

This is the L2 norm, the mean squared difference between the images. Other metrics, such as the L1 norm, or other versions of the Minkowski metric, may also be used to define distance. However, these are relatively expensive to compute. Correlation is a more efficient operator, and under certain conditions maximizing correlation is equivalent to minimizing the Euclidian distance, so it is often used as an approximate similarity metric.

If all images of an object clustered around a point (or a small number of points) in image space, and if this cluster were well separated from other object clusters, object recognition—face recognition, in this case—would be relatively straightforward. In this case, a simple metric such as Euclidian distance or correlation would work just fine. Still, it would not be terribly efficient, especially with large images and many objects (known faces). The "Eigenfaces" approach was developed in an attempt to improve on both performance and efficiency.

5. PCA and Eigenfaces

Considering the vastness of image space, it seems reasonable to begin with the following presuppositions:

(1) Images of a particular object (such as an individual's face), under various transformations, occupy a relatively small but distinct region of the image space.
(2) Different objects (different faces) occupy different regions of image space.
(3) Whole classes of objects (all faces under various transformations) occupy a still relatively small but distinct region of the image space.

These lead to the following questions about face images:

(1) What is the shape and dimensionality of an individual's "face space," and how can it be succinctly modeled and used in recognition?
(2) What is the shape and dimensionality of the complete face space, and how can it be succinctly modeled and used in recognition?
(3) Within the larger space, are the individual spaces separated enough to allow for reliable classification among individuals?
(4) Is the complete face space distinct enough to allow for reliable face/non-face classification?

The Eigenfaces framework [25]–[27] provided a convenient start to investigating these and related issues. Let us review the basic steps in an eigenfaces-based recognition scheme. Principle Component Analysis (PCA) [28] provides a method to efficiently represent a collection of sample points, reducing the dimensionality of the description by projecting the points onto the principal axes, an orthonormal set of axes pointing in the directions of maximum covariance in the data. PCA minimizes the mean squared projection error for a given number of dimensions (axes), and provides a measure of importance (in terms of total projection error) for each axis. Transforming a point to the new space is a linear transformation.

Let a set of face images $\{x_i\}$ of several people be represented as a matrix X, where

$$X = [x_1\, x_2\, x_3\, \cdots\, x_M]$$

and X is of dimension $N \times M$, where N is the number of pixels in an image, the dimension of the image space which contains $\{x_i\}$. The difference from the average face image (the sample mean) $\overline{x}$ is the matrix X',

$$\begin{aligned} X' &= [(x_1 - \overline{x})(x_2 - \overline{x})(x_3 - \overline{x}) \cdots (x_M - \overline{x})] \\ &= [{x_1}'\, {x_2}'\, {x_3}'\, \cdots\, {x_M}'] \end{aligned}$$

Principal Components Analysis seeks a set of $M - 1$ orthogonal vectors, e_i, which best describes the distribution of the input data in a least-squares sense, i.e., the Euclidian projection error is minimized. The typical method of computing the principal components is to find the eigenvectors of the covariance matrix C, where

$$C = \sum_{i=1}^{M} x_i' x_i'^T = X' X'^T$$

is $N \times N$. This will normally be a huge matrix, and a full eigenvector calculation is impractical. Fortunately, there are only $M - 1$ non-zero eigenvalues, and they can be computed more efficiently with an $M \times M$ eigenvector calculation. It is easy to show the relationship between the two. The eigenvectors e_i and eigenvalues λ_i of C are such that

$$C e_i = \lambda_i e_i$$

These are related to the eigenvectors $\hat{e}_i$ and eigenvectors μ_i of the matrix $D = X'^T X'$ in the following way:

$$\begin{aligned} &D \hat{e}_i = \mu_i \hat{e}_i \\ &X'^T X' \hat{e}_i = \mu_i \hat{e}_i \\ &X' X'^T X' \hat{e}_i = \mu_i X' \hat{e}_i \\ &C X' \hat{e}_i = \mu_i X' \hat{e}_i \\ &C(X' \hat{e}_i) = \mu_i (X' \hat{e}_i) \\ &C e_i = \lambda_i e_i \end{aligned}$$

showing that the eigenvectors and eigenvalues of C can be computed as

$$\begin{aligned} &e_i = (X' \hat{e}_i) \\ &\lambda_i = \mu_i \end{aligned}$$

In other words, the eigenvectors of the (large) matrix C are equal to the eigenvectors of the much smaller matrix D, premultiplied by the matrix X'. The nonzero eigenvalues of C are equal to the eigenvalues of D.

Once the eigenvectors of C are found, they are sorted according to their corresponding eigenvalues; a larger eigenvalue mean that more of the variance in the data is captured by the eigenvector. Part of the efficiency of the Eigenfaces approach comes from the next step, which is to eliminate all but the "best" k eigenvectors (those with the highest k eigenvalues). From there on, the "face space," spanned by the top k eigenvectors, is the feature space for recognition. The eigenvectors are merely linear combinations of the images from the original data set. Because they appear as somewhat ghostly faces, as shown in Fig. 2, they are called Eigenfaces.

PCA has been used in pattern recognition and classification systems for decades. Sirovich and Kirby [23], [29] used PCA to form *eigenpictures* to compress face images, a task for which low mean-squared error reproduction is important. Turk and Pentland [25] used PCA for representing, detecting, and recognizing faces. Nayar and Murase [30] used a similar eigenspace in a parametric representation that encoded pose and illumination variation, as well as identity. Finlayson, et al. [31] extended grayscale eigenfaces to color images. Craw, et al. [32], Moghaddam [33], Lanitis et al. [34] and others have subsequently used eigenfaces as one component of a larger system for recognizing faces.

The original eigenface recognition scheme involves two main parts, creating the eigenspace and recognition using eigenfaces. The first part (described above) is an off-line initialization procedure; that is, it is performed initially and only needs to be recomputed if the training set changes. The eigenfaces are constructed from an initial set of face images (the training set) by applying PCA to the image ensemble, after first subtracting the mean image. The output is a set of eigenfaces and their corresponding eigenvalues. Only the eigenfaces corresponding to the top M eigenvalues are kept—these define the *face space*. For each individual in the training set, the average face image is calculated (if there is more than one instance of that individual), and this image is projected into the face space as the individual's class prototype.

The second part comprises the ongoing recognition procedure. When a new image is input to the system, the mean image is subtracted and the result is projected into the face space. This produces a value for each eigenface; together, the values comprise the image's eigenface descriptors. The Euclidian distance between the new image and its projection into face space is called the "distance from face space" (DFFS), the reconstruction error. If the DFFS is above a given threshold, the image is rejected as not a face—in other words, it is not well enough represented by the eigenfaces to be deemed a possible face of interest.

If the DFFS is sufficiently small, then the image is classified as a face. If the projection into face space is if sufficiently close to one of the known face classes (by some metric such as Euclidian distance) then it is recognized as the corresponding individual. Otherwise, it is considered as an unknown face (and possibly added

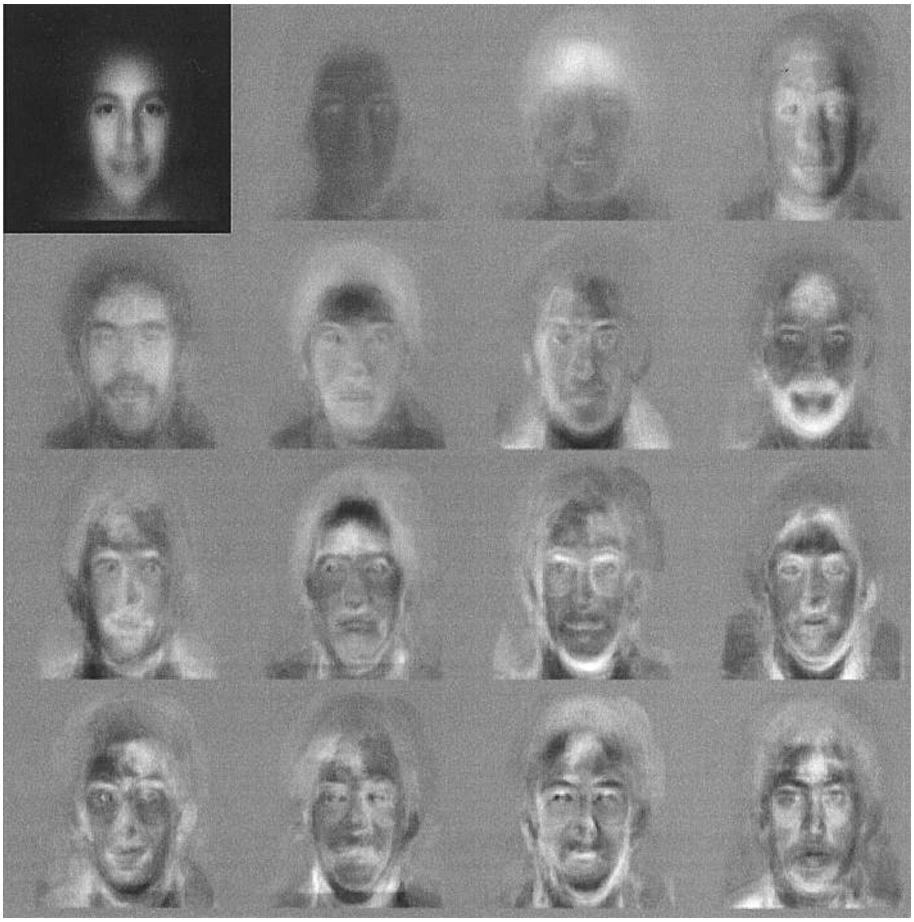

Fig. 2 The average face image $\overline{x}$ and a set of eigenface images. The eigenfaces are real-valued images scaled so that a value of zero displays as a medium gray, negative values are dark, and positive values are bright.

to the training set).

The basic Eigenfaces technique raises a number of issues, such as:

- How to select k, the number of Eigenfaces to keep
- How to efficiently update the face space when new images are added to the data set
- How best to represent classes and perform classification within the face space
- How to separate intraclass and interclass variations in the initial calculation of face space
- How to generalize from a limited set of face images and imaging conditions

There are obvious shortcomings of the basic Eigenfaces technique. For example, significant variation in scale, orientation, translation, and lighting will cause it to fail. Several appearance-based recognition methods first scale the input image to match the scale of the object in a prototype template image. While this is usually an effective approximation, one must consider that scaling an image is equivalent to changing a camera's focal length, or performing an optical zoom, but it is not equivalent to moving a camera closer to the object. A translated image has introduce occlusion, while a zoomed image does not. In addition, the reflectance is different for a translated image because of a slightly different angle of incidence. For an object with significant depth and nearby light sources, approximating translation with an image zoom may not work well. In other words, an image from the database of a face taken from one meter away will not perfectly match another image of the same face taken five meters away and zoomed in an appropriate amount.

6. A Brief History of Automated Face Recognition, Part Two

Despite its shortcomings, there are a number of attractive aspects to Eigenface methods, especially including the progress of the past decade. Since Burt [24], Turk and Pentland [26], Craw, et al. [32] and others began to use appearance-based methods in detecting and recognizing faces, there has been a voluminous amount of work on the topic, motivated by several factors. Applications of computer vision in human-computer interaction (HCI), biometrics, and image and video database systems have spurred interest in face recognition (as well as human gesture recognition and activity analysis). There are currently several companies that market face recognition systems for a variety of biometric applications, such as user authentication for ATM machines, door access to secure areas, and computer login, as well as a variety of HCI/entertainment applications, such as computer games, videoconferencing with computer-generated avatars, and direct control of animated characters (digital puppeteering). Conferences now exist, which are well attended, devoted to face recognition and related topics, and several good survey papers are available that track the various noteworthy results. The state of the art in face recognition is exemplified both by the commercial systems, on which much effort is spent to make them work in realistic imaging situations, and by various research groups exploring new techniques and better approaches to old techniques.

The Eigenface approach, as originally articulated, intentionally threw away all feature-based information in order to explore the boundaries of an appearance-based approach to recognition. Subsequent work by Moghaddam [33], Lanitis et al. [35] and others have moved toward merging the two approaches, with predictably better results than either approach alone. The original Eigenface framework did not explicitly account for variations in lighting, scale, viewing angle, facial expressions, or any of the other many ways facial images of an individual may change. The expectation was that the training set would contain enough variation so that it would be modeled in the Eigenfaces. Subsequent work has make progress in characterizing and accounting for these variations (e.g., [36] and [37]) while merging the best aspects of both feature-based and appearance-based approaches.

A few approaches in particular are significant in terms of their timing and impact. Craw et al. [32] were among the first to combine processing face shape (two dimensional shape, as defined by feature locations) with eigenface-based recognition. They normalized the face images geometrically based on 34 face landmarks in an attempt to isolate the photometric (intensity) processing from geometric factors. Von der Malsburg and his colleagues [38], [39] introduced several systems based on elastic graph matching, which utilizes a hybrid approach where local grayscale information is combined with global feature structure. Cootes and Taylor and colleagues [40] presented a unified approach to combining local and global information, using flexible shape models to explicitly model both shape and intensity.

Recent results in appearance-based recognition applied to face recognition and other tasks include more sophisticated learning methods (e.g., [41]), warping and morphing face images [42], [43] to accommodate a wider range of face poses, including previously unseen poses, explicitly dealing with issues of robustness [44], and better methods of modeling interclass and intraclass variations and performing classification [45]. Independent Component Analysis (ICA), for example, is a generalization of PCA that separates the high-order dependencies in the input, in addition to the second-order dependencies that PCA encodes [46]. The original eigenface method used a single representation and transformation for all face images, whether they originated from one individual or many; it also used the simplest techniques possible, nearest-neighbor Euclidian distance, for classification in the face space. Subsequent work has improved significantly on these first steps. Moghaddam et al. [33] developed a probabilistic matching algorithm that uses a Bayesian approach to separately model both interclass and intraclass distributions. This improves on the implicit assumption that the images of all individuals have a similar distribution. Penev and Sirovich [47] investigated the dimensionality of face space, concluding that for very large databases, at least 200 eigenfaces are needed to sufficiently capture global variations such as lighting, small scale and pose variations, race, and sex. In addition, at least twice that many are necessary for minor, identity-distinguishing details such as exact eyebrow, nose, or eye shape.

7. Conclusions

Appearance-based approaches to recognition have made a comeback from the early days of computer vision research, and the Eigenfaces approach to face recognition may have helped this come about. Clearly, though, face recognition is far from being a solved problem, whether by Eigenfaces or any other technique. The progress during the past decade on face recognition has been encouraging, although one must still refrain from assuming that the excellent recognition rates from any given experiment can be repeated in different circumstances. They usually cannot.

Eigenface (and other appearance-based) approaches must be coupled with feature- or shape-based approaches to recognition in order to build systems that will be robust and will scale to real-world environments. Because many imaging variations (lighting, scale, orientation, etc.) have an approximately linear effect when they are small, linear methods can work, but in very limited domains. Eigenfaces are not a general approach to recognition, but one tool out of many to be applied and evaluated in context. The ongoing challenge is to find the right set of tools to be applied at the appropriate times.

In addition to face recognition, significant progress is being made in related areas such as face detection, face tracking, face pose estimation, facial expression analysis, and facial animation. The "holy grail" of face processing is a system that can detect, track, model, recognize, analyze, and animate faces. Although we are not there yet, current progress gives us much reason to be optimistic. The future of face processing looks promising.

References

[1] D.I. Perrett, E.T. Rolls, and W. Caan, "Visual neurons responsive to faces in the monkey temporal cortex," Exp. Brain Res., vol.47, pp.329–342, 1982.

[2] K.M. Kendrick and B.A. Baldwin, "Cells in temporal cortex of sheep can respond preferentially to the sight of faces," Science, vol.236, pp.448–450, 1987.

[3] R. Desimone, "Face-selective cells in the temporal cortex of monkeys," J. Cognitive Neuroscience, vol.3, no.1, pp.1–8, 1991.

[4] V. Bruce, Recognizing Faces, Lawrence Erlbaum Associates, London, 1988.

[5] A.M. Burton, "A model of human face recognition," in Localist Connectionist Approaches to Human Cognition, ed. J. Grainger and A.M. Jacobs, pp.75–100, Lawrence Erlbaum Associates, London, 1998.

[6] A.M. Burton, V. Bruce, and P.J.B. Hancock, "From pixels to people: A model of familiar face recognition," Cognitive Science, vol.23, pp.1–31, 1999.

[7] D. Marr, Vision, W. H. Freeman, San Francisco, 1982.

[8] D. Mumford, "Parameterizing exemplars of categories," J. Cognitive Neuroscience, vol.3, no.1, pp.87–88, 1991.

[9] R. Brunelli and T. Poggio, "Face recognition: Features versus templates," IEEE Trans. Pattern Anal. & Mach. Intell., vol.15, no.10, pp.1042–1052, 1993.

[10] W.E.L. Grimson, Object Recognition by Computer: The role of Geometric Constraints, The MIT Press, Cambridge, 1990.

[11] Y. Lamdan and H.J. Wolfson, "Geometric hashing: A general and efficient model-based recognition scheme," Proc. International Conf. Computer Vision, pp.238–249, Tampa, FL, Dec. 1988.

[12] W.W. Bledsoe, "Man-machine facial recognition," Technical Report PRI 22, Panoramic Research Inc., Palo Alto, CA, Aug. 1966.

[13] M.D. Kelly, "Visual identification of people by computer," Stanford Artificial Intelligence Project Memo AI-130, July 1970.

[14] T. Kanade, "Picture processing system by computer complex and recognition of human faces," Dept. Information Science, Kyoto University, Nov. 1973.

[15] L.D. Harmon, M.K. Khan, R. Lasch, and P.F. Ramig, "Machine identification of human faces," Pattern Recognition, vol.13, no.2, pp.97–110, 1981.

[16] G.G. Gordon, "Face recognition from frontal and profile views," Proc. Intl. Workshop on Automatic Face- and Gesture-Recognition, pp.47–52, Zurich, 1995.

[17] A.L. Yuille, P.W. Hallinan, and D.S. Cohen, "Feature extraction from faces using deformable templates," Intl. J. Computer Vision, vol.8, no.2, pp.99–111, 1992.

[18] D. Valentin, H. Abdi, A.J. O'Toole, and G.W. Cottrell, "Connectionist models of face processing: A survey," Pattern Recognition, vol.27, pp.1209–1230, 1994.

[19] M. Flemming and G. Cottrell, "Face recognition using unsupervised feature extraction," Proc. Intl. Neural Network Conf., Paris, 1990.

[20] J.T. Lapresté, J.Y. Cartoux, and M. Richetin, "Face recognition from range data by structural analysis," in Syntactic and Structural Pattern Recognition, ed. G. Ferrat, et al., NATO ASI series, vol.F45, Springer-Verlag, Berlin, Heidelberg, 1988.

[21] J.C. Lee and E. Milios, "Matching range images of human faces," Proc. IEEE Third Intl. Conf. Computer Vision, pp.722–726, Osaka, Japan, Dec. 1990.

[22] G.G. Gordon, "Face recognition from depth and curvature," Ph.D. Thesis, Harvard University, 1991.

[23] L. Sirovich and M. Kirby, "Low dimensional procedure for the characterization of human faces," J. Optical Society of America, vol.4, no.3, pp.519–524, 1987.

[24] P. Burt, "Smart sensing within a pyramid vision machine," Proc. IEEE, vol.76, no.8, pp.1006–1015, 1988.

[25] M. Turk and A. Pentland, "Face recognition without features," Proc. IAPR Workshop on Machine Vision Applications, pp.267–270, Tokyo, Nov. 1990.

[26] M. Turk and A.P. Pentland, "Eigenfaces for recognition," J. Cognitive Neuroscience, vol.3, no.1, pp.71–96, 1991.

[27] M. Turk, "Interactive-time vision: Face recognition as a visual behavior," Ph.D. Thesis, The Media Laboratory, Massachusetts Institute of Technology, Sept. 1991.

[28] I.T. Jolliffe, Principal Component Analysis, Springer-Verlag, New York, 1986.

[29] M. Kirby and L. Sirovich, "Appliction of the Karhumen-Loeve procedure for the characterization of human faces," IEEE Trans. Pattern Anal. & Mach. Intell., vol.12, no.1, pp.103–108, 1990.

[30] H. Murase and S. Nayar, "Visual learning and recognition of 3D objects from appearance," Intl. J. Computer Vision, vol.14, pp.5–24, 1995.

[31] G.D. Finlayson, J. Dueck, B.V. Funt, and M.S. Drew, "Colour eigenfaces," Proc. Third Intl. Workshop on Image and Signal Processing Advances in Computational Intelligence, Manchester, UK, Nov. 1996.

[32] I. Craw, N. Costen, T. Kato, G. Robertson, and S. Akamatsu, "Automatic face recognition: Combining configuration and texture," Proc. Intl. Workshop on Automatic Face- and Gesture-Recognition, pp.53–58, Zurich, 1995.

[33] B. Moghaddam, W. Wahid, and A. Pentland, "Beyond eigenfaces: Probabilistic matching for face recognition," Proc. Third Intl. Conf. Automatic Face- and Gesture-Recognition, pp.30–35, Nara, Japan, 1998.

[34] A. Lanitis, C.J. Taylor, and T.F. Cootes, "A unified approach to coding and interpreting face images," Proc. Fifth Intl. Conf. Computer Vision, pp.368–373, 1995.

[35] A. Lanitis, C.J. Taylor, and T.F. Cootes, "Automatic interpretation and coding of face images using flexible models," IEEE Trans. Pattern Anal. & Mach. Intell., vol.19, no.7, pp.743–756, 1997.

[36] A.S. Georghiades, D.J. Kriegman, and P.N. Belhumeur, "Illumination cones for recognition under variable lighting: Faces," IEEE Conf. Computer Vision and Pattern Recognition, 1998.

[37] L. Zhao and Y.H. Yang, "Theoretical analysis of illumination in PCA-based vision systems," Pattern Recognition, vol.32, pp.547–564, 1999.

[38] M. Lades, J.C. Vorbruggen, J. Buhmann, J. Lange, C. Von der Malsburg, R.P. Wurtz, and W. Konen, "Distortion invariant object recognition in the dynamic link architecture," IEEE Trans. Comput., vol.42, no.3, pp.300–311, 1993.

[39] R.P. Würtz, J.C. Vorbrüggen, C. von der Malsburg, and J. Lange, "Recognition of human faces by a neuronal graph matching process," in Applications of Neural Networks, ed. H.G. Schuster, pp.181–200, VCH, Weinheim, 1992.

[40] A. Lanitis, C.J. Taylor, and T.F. Cootes, "A unified approach to coding and interpretting faces," Proc. 5th Intl. Conf. Computer Vision, pp.368–373, 1995.

[41] Y. Li, S. Gong, and H. Liddell, "Support vector regression and classification based multi-view face detection and recognition," Proc. Conf. Automatic Face and Gesture Recognition, pp.300–305, Grenoble, France, March 2000.

[42] T. Ezzat and T. Poggio, "Facial analysis and synthesis using

image-based models," Proc. Second Intl. Conf. Automatic Face and Gesture Recognition, pp.116–121, Killington, VT, 1996.

[43] M. Bichsel, "Automatic interpolation and recognition of face images by morphing," Proc. Second Intl. Conf. Automatic Face and Gesture Recognition, pp.128–135, Killington, VT, 1996.

[44] A. Leonardis and H. Bischof, "Robust recognition using eigenfaces," Computer Vision and Image Understanding 78, pp.99–118, 2000.

[45] P. Belhumeur, J. Hespanha, and D. Kriegman, "Eigenfaces vs. fisherfaces: Recognition using class specific linear projection," IEEE Trans. Pattern Anal. & Mach. Intell., vol.19, no.7, pp.711–720, July 1997.

[46] T.W. Lee, Independent Component Analysis: Theory and Applications, Kluwer Academic Publishers, Dordrecht, 1998.

[47] P.S. Penev and L. Sirovich, "The global dimensionality of face space," Proc. 4th Int'l. Conf. Automatic Face and Gesture Recognition, pp.264–270, Grenoble, France, 2000.

Matthew Turk received the B.S. degree in electrical engineering from Virginia Tech (VPI&SU) in 1982 and the M.S. degree in electrical and computer engineering from Carnegie Mellon University in 1984. From 1984 to 1987 he worked at Martin Marietta Aerospace on mobile robot vision for the DARPA Autonomous Land Vehicle project. In 1987 he went to the Massachusetts Institute of Technology, where he received the Ph.D. degree in Media Arts and Sciences from the MIT Media Laboratory in 1991. He was a visiting researcher in Grenoble, France in 1992, and then joined Teleos Research (Palo Alto, CA) in 1993. In 1994 he joined Microsoft Research where he co-founded the Vision Technology Group. In the fall of 2000, he moved to the University of California, Santa Barbara (UCSB) where he is an associate professor of Computer Science and Media Arts and Technology. His primary research interests are in the intersection of computer vision and human-computer interaction. At the 2000 IAPR Workshop on Machine Vision Applications (MVA2000), he received a "Most Influential Paper of the Decade" award for a paper he co-authored (with Alex Pentland) and presented at MVA1990 on using eigenfaces for representing and recognizing faces.