# Eigenfaces and Beyond

Matthew Turk
Computer Science Department
University of California
Santa Barbara, CA 93106
mturk@cs.ucsb.edu

**Abstract**

*Automated face recognition has a long history within the field of computer vision, and there have been several different classes of approaches to the problem. It has been about fifteen years since the "Eigenfaces" method first made an impression on the computer vision research community and helped spur interest in appearance-based recognition, biometrics and vision-based human-computer interface. In this chapter I give a personal view of the original context and motivation for the work, some of the strengths and limitations of the approach, and progress in the years since. The original Eigenfaces approach was in many respects a reaction to the feature-based approaches to face recognition prevalent in the mid-1980s. Appearance-based approaches to recognition complement feature- or shape-based approaches, and a practical face recognition system should have elements of both. Eigenfaces should not be viewed as a general approach to recognition, but rather one tool out of many to be applied and evaluated in the appropriate context.*

## 1. Introduction

The subject of visual processing of human faces has received attention from philosophers and scientists for centuries. Aristotle devoted several chapters of the *Historia Animalium* to the study of facial appearance. Physiognomy, the practice or art of inferring intellectual or character qualities of a person from outward appearance[1], particularly the face and head, has had periods of fashion in various societies [65]. Darwin considered facial expression and its identification to be a significant advantage for the survival of species [66]. Developmental studies have focused on strategies of recognition or identification and the differences between infant and adult subjects. Neurological disorders of face perception have been isolated and studied, providing insight into normal as well as abnormal face processing.

The ability of a person to recognize another person (e.g., a mate, a child, or an enemy) is important for many reasons. 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 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 (and even between humans and robots [67]). 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.

[1] For example, from Aristotle's *Historia Animalium*: "Straight eyebrows are a sign of softness of disposition; such as curve in towards the nose, of harshness; such as curve out towards the temples, of humour and dissimulation; such as are drawn in towards one another, of jealousy."

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. Many people say "I'm not very good at names, but I never forget a face." Although this quote confuses recall (generally more difficult) with recognition (generally less difficult), the point is valid that our face recognition capabilities are quite good. Lest we marvel too much at human performance, though, it should also be noted that the *inability* to recognize a face is a common experience as well. Quite often we strain to see the resemblance between a picture (e.g., a driver's license photo) and the real person; 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.

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 clothing, hairstyle, gait, silhouette, skin, 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, a 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 mode of person identification is as compelling as face recognition.

Until rather recently, face recognition was often pointed to as one of those things that "computers can't do," even by such luminaries as Marvin Minsky [68] and David Hubel [69]. This was a motivating factor for many computer vision students and researchers. In addition, facial processing had already interested human and biological vision researchers for years, and there were many interesting and curious results and theories discussed in the literature.

There has been a great deal of scientific investigation into human face recognition performance, seeking to understand and characterize the representations and processes involved. However, a thorough understanding of how humans (and animals) represent, process, and recognize faces remains an elusive 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.

There are several aspects of recognizing human identity and processing facial information that make the problem of face recognition somewhat 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. There are many ways in which these "face processing" tasks may interrelate. For example, the face may be initially detected and then recognized. Alternatively, detection and recognition may be performed in tandem, so that detection is merely a successful recognition event. Or facial feature tracking may be performed and facial expression analyzed before attempting to recognize the normalized (expressionless) face. There are, of course, many additional variations possible.

For the purposes of this chapter, "face recognition" and "face identification" describe the same task.[2] 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.

Face recognition began to be a hot topic in computer vision in the late 1980s and early 1990s. In the past two decades, the field has made substantial progress: starting with a limited set of slow techniques with questionable accuracy and applicability, there are now real-time systems installed in public spaces and sold in shrink-wrapped boxes. (Whether or not their performance is good enough for the intended applications will not be a subject of discussion in this chapter, nor will we discuss the significant policy issues that these systems generate. Clearly, however, these are topics of great importance, as argued in [62, 63, 64].)

In retrospect, the "Eigenfaces" approach to face recognition popularized initially by Turk and Pentland [25, 26, 27] appears to have played a significant role in the field's emerging popularity. Maybe it was due to the part-catchy, part-awkward name given to the technique (never underestimate good PR!) or maybe it was due to the simplicity of the approach, which has been re-implemented countless times in graduate and undergraduate computer vision courses over the years. One could argue that it was the right approach at the right time – different enough from previous approaches to the problem that it caught the attention of researchers, and also caught (and perhaps even influenced to some small degree) the emerging trends in appearance-based vision and learning in vision.

Eigenfaces attracted people to the topic and pushed the state of the art at the time, and it has been a very useful pedagogical tool over the years, as well as a useful benchmark for comparison purposes. But it has been about fifteen years since the first publications of this work, and many other techniques have been introduced; in addition, there have been several modifications and improvements to the original Eigenfaces technique. In this chapter we take a brief look back at face recognition research over the years, including the motivations and goals of early work in the field and the rationale for pursuing Eigenfaces in the first place; we discuss some of the spinoffs and improvements over the original method; and we speculate on what may be in store for the future of automated face recognition, and for Eigenfaces in particular.

## 2. Original Context and Motivations of Eigenfaces

### 2.1. Human and biological vision

Face recognition can be viewed as the problem of robustly identifying an image of a human face, given some database of known faces. It has all the difficulties of other vision-based recognition problems, as the image changes drastically due to several complex and confounding factors such as illumination, camera position, camera parameters (e.g., the specific lens used, the signal gain),

[2] The distinction may be made between classifying an object into its general category ("recognition") and labeling the object as a particular member of the category ("identification"), but we will follow the common terminology and use these terms interchangeably, with the precise meaning depending on the context.

and noise. In addition, human faces change due to aging, hairstyle, facial hair, skin changes, facial expression, and other factors. These all make the face recognition problem quite difficult and unsolvable by direct image matching. Early researchers in the area took several different approaches attempting to deal with the complexity; these approaches were motivated by several different factors.

Much of the interest in face recognition in the mid-1980s was motivated by an interest in human and biological vision, and especially some exciting findings related to agnosia and object-selective visual recognition. Visual agnosia is a neurological impairment in the higher visual processes which leads to a defect in object recognition [70]. Agnosic patients can often "see" well, in that there is little apparent deficit in spatial vision or perception of form. The dysfunction is specific to some class of objects or shapes, such as perceiving letters or any object from an unusual viewpoint. Etcoff et al. [71] report a patient's description of his agnosia to be like "attempting to read illegible handwriting: you know that it is handwriting, you know where the words are and letters stop and start, but you have no clue as to what they signify."

*Prosopagnosia*, from the Greek *prosopon* (face) and *agnosia* (not knowing), refers to the inability to recognize familiar faces by visual inspection [72, 73, 74]. Prosopagnosic patients, although very few in number, have proved to be a valuable resource in probing the function of face recognition. Prosopagnosics can typically identify the separate features of a face, such as the eyes or mouth, but have no idea to whom they belong. They may recognize the sex, age, pleasantness, or expression of a face, without an awareness of the identity:

> I was sitting at the table with my father, my brother and his wife. Lunch had been served. Suddenly… something funny happened: I found myself unable to recognize anyone around me. They looked unfamiliar. I was aware that they were two men and a woman; I could see the different parts of their faces but I could not associate those faces with known persons.... Faces had normal features but I could not identify them. (Agnetti et al., p. 51, quoted in [75])

There is evidence that damage to a particular area of the right hemisphere has a predominant role in producing face recognition difficulties. The question arises, is face recognition a special, localized, subsystem of vision? One way to approach this question, and additionally to learn about the neural mechanisms involved in face recognition and object recognition in general, is by recording the activity of brain cells while performing visual tasks including observing and recognizing faces. Through single cell recording, a number of physiologists found what seem to be "face" neurons in monkeys, responding selectively to the presence of a face in the visual field. Perrett et al. [76, 77, 78] found cells in area STS of the rhesus monkey which were selectively responsive to faces in the visual field. Many of these cells were insensitive to transformations such as rotation. Different cells responded to different features or subsets of features, while most responded to partially obscured faces. Some cells responded to line drawings of faces. About 10% of the cells were sensitive to identity. Other researchers (e.g. [4, 79, 80]) have found cells with similar properties in monkey inferior temporal cortex, concluding that there may be specialized mechanisms for the analysis of faces in IT cortex. Kendrick and Baldwin [81] even found face-selective cells in sheep.

Psychologists have used both normal and prosopagnosic subjects to investigate models of face processing, recognition, and identification. In addition to the theoretical and clinical pursuits of neuroscience, the validity and limitations of eye-witness testimony in criminal proceedings has spurred much face recognition research in cognitive psychology. Yin [82] presented pictures of faces in various orientations and tested subsequent recall, finding that the recall performance for inverted faces was degraded more than that of other configuration-specific stimuli such as landscapes or animals. He argued for a special face-processing mechanism to account for this effect. Others have furthered these techniques to experiment with face images which have been modified in several ways. Developmental studies (e.g. [83]) have observed the development of face recognition from infant to adult. Carey and Diamond [84] found that the effect of inversion on face recognition described by Yin increases over the first decade of life, suggesting that young children represent faces in terms of salient isolated features ("piecemeal representation"), rather than in terms of configurational properties used by older children and adults ("configurational representation"). In recent years there seems to be a growing consensus that both configurational properties and feature properties are important for face recognition [4]. Carey and Diamond [85] claimed that face recognition is not a special, unique system, and that the inversion effect may be due to a gain in the ability to exploit distinguishing "second-order relational features."

A number of experiments have explored feature saliency, attempting to discern the relative importance of different features or areas of the face. Although the early experiments generally agreed on the importance of face outline, hair, and eyes – and the relative unimportance of the nose and mouth – there is evidence that these results may be biased by the artifacts of the techniques and face presentations used [4]. Along with stored face images, a number of researchers [86, 87] have used face stimuli constructed from Identikit or Photofit2 to explore strategies of face recognition. Use of these kits may actually bias the experiments, however, since there is an underlying assumption that a face can be properly decomposed into its constituent features: eyes, ears, nose, mouth, etc. One lesson from the study of human face recognition is that approaches which treat faces as a collection of independent parts are unlikely to be relevant to the perception of real faces, where the parts themselves are difficult or impossible to delimit [4]. Consequently artists' sketches are better than face construction kits in reproducing the likeness of a target face. Faces grow and develop in a way such that features are mutually constraining. In fact these growth patterns can be expressed mathematically and used to predict the effects of aging [88]. Such techniques have already been used successfully in the location of missing children years after their disappearance [89]. Other studies have shown that expression and identity seem to be relatively independent tasks [90, 91], which is also supported by some neurological studies of prosopagnosics.

### 2.2. Compelling applications of face recognition

The research that led to Eigenfaces had several motivations, not the least of which was industry funding with the aim of developing television set-top boxes that could visually monitor viewers as an automated "people meter" to determine television ratings. This required a real-time system to locate, track, and identify possible viewers in a scene. Besides identifying typical members of the household, such a system should also detect unknown people (and perhaps ask them to enter relevant information, such as their sex and age) and distinguish between valid viewers and, for example, the family dog. It was also important to know when the television was on but no one was viewing it, and when people were in the vicinity but not actively watching. For this application, computation efficiency and ease of use (including training) were primary considerations.

General surveillance and security applications were also a motivating application, especially criminal mugshot identification. Whereas recognizing members of a family for use as an automated people meter required a small database of known identities, many security applications required a very large database, so memory and storage requirements were important; full images could not be stored and matched for each entry.

Other areas of interest that motivated work in face recognition at the time included image compression, film development, and human-computer interaction. In the areas of image compression for transmission of movies and television, and in general any "semantic understanding" of video signals, the presence of people in the scene is important. For example, in partitioning the spatial-temporal bandwidth for an advanced HDTV transmission, more bandwidth should be given to people than to cars, since the audience is much more likely to care about the image quality and detail of the human actors than of inanimate objects. The detection of faces in photograph negatives or originals could be quite useful in color film development, since the effect of many enhancement or noise reduction techniques depends on the picture content. Automated color enhancement is desirable for most parts of the scene, but may have an undesirable effect on flesh tones. (It is fine for the yellowish grass to appear greener, but not so fine for Uncle Harry to look like a Martian!) In human-computer interaction, there was a growing interest in creating machines that understand, communicate with, and react to humans in natural ways; detecting and recognizing faces is a key component of such natural interaction.

Interest in computer-based automated face recognition in the mid-1980s was motivated by several factors, including an interest in the mechanisms of biological and human vision, general object recognition, image coding, and neural networks. Today, the field is largely driven by applications in security and surveillance, perceptual interfaces [92], and content-based query of image and video. Questions of relevance to biological vision (mechanisms of human face recognition) are still of interest, but they are not a major motivation.

### 2.3. Object recognition strategies

Object recognition has long been a 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.

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 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 people's faces. They 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 Mumford's 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.

In the past decade, learning has become a very significant issue in visual recognition. Rather than laboriously 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 image models or templates, it would be helpful for the system to analyze the imaging conditions to decide on optimal representations, 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 incorporating 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 (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 or uncertainty 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.

For face recognition systems, learning and robustness must also be balanced with practical speed requirements. 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 face recognition.

As with most recognition tasks, the source images (face 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 clear that the human face is much more difficult to model (and thus recognize) than most objects.

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.

### 2.4. Face recognition through the 1980s

In general, face recognition has been viewed as both an interesting scientific direction of research (possibly helping to understand human vision) and as a useful technology to develop. The big question in the mid-1980s was how to go about solving the problem? Perhaps more importantly, how to go about defining the problem? Face recognition was viewed as a high-level visual task, while basic areas such as stereo and motion perception were still not completely understood. However, the tasks involved in face processing are reasonably constrained; some may even have a degree of “hardwiring” in biological systems. Faces present themselves quite consistently in expected positions and orientations; their configuration (the arrangement of the components) seldom changes; they are rather symmetrical. On the other hand, human face recognition and identification is very robust in the face of external changes (e.g. hair styles, tan, facial hair, eyeglasses), so a recognition scheme cannot be rigid or overly constrained.

Should one approach face recognition via the framework of the “Marr paradigm,” building a primal sketch or intrinsic images from the raw image data, then a viewer-centered 2 ½ D sketch revealing scene discontinuities, and finally a 3D object-centered representation on which to perform recognition? Or should face recognition be a very specialized task, somewhat separate from other general object recognition approaches, as perhaps suggested by some of the human vision literature? Until the late 1980s, most of the work in automated face detection and recognition had focused on detecting individual features such as the eyes, nose, mouth, and head outline, and defining a face model by the position, size, and relationships among these features, with some newer methods based on neural networks, correlation-based techniques, and shape matching from range data.

Attempts to automate human face recognition by computers began in the late 1960s and early 1970s. Bledsoe [12] was the first to report semi-automated face recognition, using a hybrid human-computer system which classified faces on the basis of fiducial marks entered on photographs by hand. Parameters for the classification were normalized distances and ratios among points such as eye corners, mouth corners, nose tip, and chin point. 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. At Bell Labs, Harmon, Goldstein and their colleagues [15, 93, 94] developed an interactive system for face recognition based on a vector of up to 21 features, which were largely subjective evaluations (e.g. shade of hair, length of ears, lip thickness) made by human subjects. The system recognized known faces from this feature vector using standard pattern classification techniques. Each of these subjective features however would be quite difficult to automate.

Sakai et al. [95] described a system which locates features in a Laplacian-filtered image by template-matching. This was used to find faces in images, but not to recognize them. A more sophisticated approach by Fischler and Elschlager [96] attempted to locate image features automatically. They described a linear embedding algorithm which used local feature template matching and a global measure to perform image matching. The technique was applied to faces, but not to recognition. The first automated system to recognize people was developed by Kelly [13]. He developed heuristic, goal-directed methods to measure distances in standardized images of the body and head, based on edge information.

Kanade's face identification system [14] was the first automated system to use a top-down control strategy directed by a generic model of expected feature characteristics of the face. His system calculated a set of facial parameters from a single face image, comprised of normalized distances, areas, and angles between fiducial points. He used a pattern classification technique to match the face to one of a known set, a purely statistical approach depending primarily on local histogram analysis and absolute gray-scale values.

In a similar spirit, Harmon et al. [15] recognized face profile silhouettes by automatically choosing fiducial points to construct a 17-dimensional feature vector for recognition. Gordon [16] also investigated face recognition using side view facial profiles. Others have also approached automated face recognition by characterizing a face by a set of geometric parameters and performing pattern recognition based on the parameters (e.g. [97, 98, 99]).

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.

A number of researchers (e.g. [100, 101]) were using faces or face features as input and training patterns to neural networks with a hidden layer, trained using backpropagation, but on small data sets. Fleming and Cottrell [19] extended these ideas using nonlinear units, training the system by back propagation. The system accurately evaluated "faceness," identity, and to a lesser degree gender, and reported a degree of robustness to partial input and brightness variations. Cottrell and Metcalfe [102] built on this work, reporting identity, gender, and facial expression evaluations by the network. The WISARD system [103] was a general-purpose binary pattern recognition device based on neural net principles. It was applied with some success to face images, recognizing both identity and expression.

Range data has the advantage of being free from many of the imaging artifacts of intensity images. Surface curvature, which is invariant with respect to viewing angle, may be quite a useful property in shape matching and object recognition. Lapresté et al. [20] presented an analysis of curvature properties of range images of faces, and propose a pattern vector comprised of distances between characteristic points. Sclaroff and Pentland [104] reported preliminary recognition results based on range data of heads.

Lee and Milios [21] explored matching range images of faces represented as extended Gaussian images. They claimed that meaningful features correspond to convex regions and are therefore easier to identify than in intensity images. Gordon [22] represented face features based on principal curvatures, calculating minimum and maximum curvature maps which are used for segmentation and feature detection. The major drawback of these approaches is the dependency on accurate, dense range data, which is currently not available using passive imaging systems, while active range systems can be very cumbersome and expensive. In addition, it is not clear that range information alone is sufficient for reliable recognition [105].

In 1988, 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.

## 3. Eigenfaces

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. Given the context of the problem at the time – the history of various approaches and the particular requirements of the motivating applications – we wanted to consider face recognition methods that would meet the basic application requirements but with a different approach than what had been pursued up to that time. Feature-based approaches seemed to solve some issues but threw away too much information. Neural network approaches at the time seemed to depend on "black box" solutions that could not be clearly analyzed. Approaches using range data were too cumbersome or expensive for our main application interests, and also did not appeal to our human vision motivations.

Much of the previous work on automated face recognition had ignored the issue of just what aspects of the face stimulus are important for identification, by either treating the face as a uniform pattern or assuming that the positions of features are an adequate representation. It is not evident, however, that such representations are sufficient to support robust face recognition. Depending too much on features, for example, causes problems when the image is degraded by noise or features are occluded (e.g., by sunglasses). We would like to somehow allow for a system to decide what is important to encode for recognition purposes, rather than specifying that initially.

This suggested that an information theory approach of coding and decoding face images may give insight into the information content of face images, emphasizing the significant local and global features, which may or may not be directly related to our intuitive notion of face features such as the eyes, nose, lips, and ears. This may even have important implications for the use of construction tools such as Identikit and Photofit [4], which treat faces as "jigsaws" of independent parts. Such a system motivated by information theory would seek to extract the relevant information in a face image, encode it as efficiently as possible, and compare one face encoding with a database of models encoded similarly. One approach to extracting the information contained in an image of a face is to somehow capture the variation in a collection of face images, independent of any judgment of features, and use this information to encode and compare individual face images.

### 3.1. Image Space

Appearance-based approaches to vision begin 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 3x3 spatial image filter is not an efficient operation in image space; it is accomplished by multiplication with a very large, sparse NxN 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 8x8 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 $\boldsymbol{x}$ in image space, or, equivalently, a feature in a high-dimensional feature space. The image pixel $I(r, c)$ can be mapped to the $i^{th}$ 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 $\boldsymbol{x}$ and pre-existing face classes $\tilde{x}$ . That is, given $k$ prototype images of known objects, find the prototype $\tilde{x}_i$ 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,

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

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.

### 3.2. PCA

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

- 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.
- Different objects (different faces) occupy different regions of image space.
- 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:

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

The Eigenfaces framework [25, 26, 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. A simple example of PCA is shown in Figure 1. Projecting face and non-face images into face space is shown in Figure 2.

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

$$\boldsymbol{X} = \left[\boldsymbol{x_1}\ \boldsymbol{x_2}\ \boldsymbol{x_3} \cdots \boldsymbol{x_M}\right]$$

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

$$\boldsymbol{X'} = \left[(\boldsymbol{x_1} - \bar{\boldsymbol{x}})(\boldsymbol{x_2} - \bar{\boldsymbol{x}})(\boldsymbol{x_3} - \bar{\boldsymbol{x}}) \cdots (\boldsymbol{x_M} - \bar{\boldsymbol{x}})\right] = \left[\boldsymbol{x_1}'\ \boldsymbol{x_2}'\ \boldsymbol{x_3}' \cdots \boldsymbol{x_M}'\right]$$

Principal Components Analysis seeks a set of $M$-1 orthogonal vectors, $\boldsymbol{e_i}$, 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 $\boldsymbol{C}$, where

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

is $N$ x $N$. 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 $M$ x $M$ eigenvector calculation. It is easy to show the relationship between the two. The eigenvectors $\boldsymbol{e_i}$ and eigenvalues $\lambda_i$ of $\boldsymbol{C}$ are such that

$$\boldsymbol{C e_i} = \lambda_i \boldsymbol{e_i}$$

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

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

showing that the eigenvectors and eigenvalues of $\boldsymbol{C}$ 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 non-zero 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 means 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 Figure 3, 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. Murase and Nayar [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 [32], Moghadam [33], Lanitis et al. [34] and others have subsequently used Eigenfaces as one component of a larger system for recognizing faces.

The original Eigenfaces 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 to the training set). Figure 4 shows a DFFS map corresponding to an input scene image; the face is located at the point where the DFFS map is a minimum.

Eigenfaces was originally used both for detection (via DFFS) and identification of faces. Figure 5 shows the overall system that was first constructed to use simple motion processing to indicate the likely area for the head; then within a smaller image segment, DFFS was used to determine the most likely face position, where face recognition was attempted. The complete tracking, detection, and recognition ran at a few frames per second on a 1990 workstation.

Although the early Eigenfaces papers did not clearly articulate how the background was handled, Figure 6 shows the two standard mechanisms used to eliminate or minimize background effects. Given a particular expected scale (size of face), masking out the background around the face (possibly including most of the hair and ears) adequately removes the background from consideration.

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 introduces 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.

### 3.3. Eigenobjects

An obvious question in response to approaching face recognition via Eigenfaces is what about recognizing other objects and entities using a similar approach? Is this method particular to faces, or to some particular class of objects, or is it a general recognition method? What recognition problems are best tackled from a view-based perspective in general?

Since the initial Eigenface work, several researchers have introduced PCA-based approaches to recognition problems. These include "eigeneyes," "eigennoses," and "eigenmouths" for the representation, detection, and recognition of facial features [27, 57, 56, 55]; "eigenears" as a biometric modality [58]; "eigenexpression" for facial expression analysis [27, 53]; "eigenfeatures" for image registration and retrieval [51, 50]; "eigentracking" as an approach to view-based tracking, and "eigentexture" as a 3D texture representation [59]. In addition to these vision techniques, there have been similar approaches, some inspired by the Eigenfaces work, in other areas: for example, "eigenvoices" for speech recognition and adaptation [48, 49] and "eigengenes" and "eigenexpression" in genetics.

Most notably, the Nayar and Murase work [30], mentioned above, utilized an eigenspace approach to represent and recognize general 3D objects at various poses, formulating object and pose recognition as parameterized appearance matching. They recognized and determined the pose of 100 objects in real-time [60] by creating appearance manifolds based on the learned eigenspace.

## 4. Improvements to and extensions of Eigenfaces

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 [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 (e.g., Zhao et al. [106]). 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 Eigenfaces 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 Eigenfaces 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 made 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 Eigenfaces 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.

## 5. Summary

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 bring this 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. Reports of dismal performance of commercial face recognition systems in real-world scenarios [61] seem to confirm this.

Eigenface (and other appearance-based) approaches must be coupled with feature- or shape-based approaches to recognition, possibly including 3D data and models, 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.

**Acknowledgements**
The writing of this chapter was supported in part by NSF grant #0205740. I would like to acknowledge the vital contributions of Sandy Pentland in the original Eigenfaces work as well as the support of Tony Gochal and the Arbitron Company, and to salute everyone who has re-implemented and improved the method over the years. Portions of this chapter were expanded with permission from M. Turk, "A Random Walk through Eigenspace," IEICE Trans. Inf. & Syst., Vol. E84-D, No. 12, December 2001.

**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, 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, J. Grainger and A.M. Jacobs, eds., pp. 75-100. London: Lawrence Erlbaum Associates, 1998.
[6] A. M. Burton, V. Bruce, and P. J. B. Hancock, “From pixels to people: a model of familiar face recognition,” Cognitive Science, 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, 3(1), pp. 87-8, 1991.
[9] R. Brunelli and T. Poggio. "Face recognition: Features versus templates," IEEE Transactions on Pattern Analysis and Machine Intelligence, 15, 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, December 1988.
[12] W.W. Bledsoe. *Man-machine facial recognition*. Technical Report PRI 22, Panoramic Research Inc., Palo Alto, CA, August 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. of 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, Vo. 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,” International Journal of Computer Vision 8 (2), pp. 99-111, 1992.
[18] D. Valentin, H. Abdi, A. J. O'Toole, G. W. Cottrell, “Connectionist models of face processing: A survey,” Pattern Recognition, 27, pp. 1209-1230, 1994.
[19] M. Flemming and G. Cottrell, “Face recognition using unsupervised feature extraction,” Proc International Neural Network Conference, Paris, 1990.
[20] J.T. Laprest, J.Y. Cartoux, M. Richetin, “Face recognition from range data by structural analysis,” in Syntactic and Structural Pattern Recognition, G. Ferrat et al., eds., 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. on Computer Vision, pp. 722-726, Osaka, Japan, December 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,” Journal of Optical Society of America, Vol. 4, No. 3, pp. 519-524, 1987.
[24] P. Burt, “Smart sensing within a pyramid vision machine,” Proc. of the IEEE, 76(8), pp. 1006-1015, 1988.
[25] M. Turk and A. Pentland, “Face recognition without features,” Proc. IAPR Workshop on Machine Vision Applications, Tokyo, pp. 267-270, November 1990.
[26] M. Turk and A.P. Pentland, “Eigenfaces for recognition,” J. Cognitive Neuroscience, 3(1):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, September 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 Analysis and Machine Intelligence, 12(1), pp. 103-108, 1990.

[30] H. Murase and S. Nayar, “Visual learning and recognition of 3D objects from appearance,” Intl. J. of Computer Vision, 14:5-24, 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, November 1996, Manchester, UK.

[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, Zurich, 1995, pp. 53-58.

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

[34] A. Lanitis, C. J.Taylor, and T. F.Cootes, “A unified approach to coding and interpreting face images,” Proc. Fifth Intl. Conf. on Computer Vision, pages 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 Analysis and Machine Intelligence, 19(7):743--756, 1997.

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

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

[38] M. Lades, et al., “Distortion invariant object recognition in the dynamic link architecture,” IEEE Trans. Computers, 42(3):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, H. G. Schuster, ed., 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. of 5th Intl. Conf. on 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. On Automatic Face and Gesture Recognition, Grenoble, France, pp. 300-305, March 2000.

[42] T. Ezzat and T. Poggio, “Facial analysis and synthesis using image-based models,” Proc. Second Intl. Conf. on Automatic Face and Gesture Recognition, Killington, VT, pp. 116-121, 1996.

[43] M. Bichsel, “Automatic interpolation and recognition of face images by morphing,” Proc. Second Intl. Conf. on Automatic Face and Gesture Recognition, Killington, VT, pp. 128-135, 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 on Pattern Analysis and Machine Intelligence, 19 (7):711-720, July 1997.

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

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

[48] R. Kuhn, J.-C. Junqua, P. Nguyen, and N. Niedzielski, “Rapid Speaker Adaptation in Eigenvoice Space,” IEEE Transactions on Speech and Audio Processing, 8(4):695–707, Nov 2000.
[49] J. Kwok, B. Mak, S. Ho, “Eigenvoice speaker adaptation via composite kernel principal Component Analysis,” Proc. NIPS-2003, Vancouver, B. C., Dec. 9-11, 2003.
[50] D. L. Swets and J. Weng, “Using Discriminant Eigenfeatures for Image Retrieval,” IEEE Transactions on Pattern Analysis and Machine Intelligence, Vol. 18, No. 8, pp. 831-836, 1996.
[51] H. Schweitzer. Optimal eigenfeature selection by optimal image registration. In Conference on Computer Vision and Pattern Recognition, pages 219--224, 1999.
[52] O. Alter, P. Brown, and D. Botstein, “Singular value decomposition for genome-wide expression data processing and modeling,” Proc. Natl Acad Sci 97 (18), pp. 10101–10106, Aug. 2000.
[53] Cao, X. and Guo, B. “Real-time tracking and imitation of facial expression,” Proc. SPIE Intl. Conf. on Imaging and Graphics, Hefei, PRC, Aug., 2002.
[54] M. Black and A. Jepson, “Eigentracking: robust matching and tracking of articulated objects using a view-based representation,” International Journal of Computer Vision 26(1), pp. 63-84, 1998.
[55] T. de Campos, R. Feris, and R. Cesar, “Eigenfaces versus Eigeneyes: First Steps Toward Performance Assessment of Representations for Face Recognition,” Lecture Notes in Artificial Intelligence, vol. 1793, pp. 197-206, April 2000.
[56] E. Hjelmås and J. Wroldsen, “Recognizing Faces from the Eyes Only,” Proceedings of the 11th Scandinavian Conference on Image Analysis, 1999.
[57] A. Pentland, B. Moghaddam, and T. Starner, “View-based and modular eigenspaces for face recognition,” Proc. Computer Vision and Pattern Recognition, pp. 84-91, June 1994.
[58] K. Chang, K. Bowyer, S. Sarkar, and B. Victor, **“**Comparison and Combination of Ear and Face Images in Appearance-Based Biometrics,” IEEE Trans. PAMI, Vol. 25, No. 9, pp. 1160-1165, Sept. 2003.
[59] K. Nishino, Y. Sato and K. Ikeuchi, “Eigen-texture method: appearance compression and synthesis based on a 3D model,” IEEE Transactions on Pattern Analysis and Machine Intelligence, Vol.23, No.11, pp. 1257-1265, Nov 2001.
[60] S. Nayar, S. Nene, and H. Murase, “Real-Time 100 Object Recognition System,” Proceedings of ARPA Image Understanding Workshop, San Francisco, Feb. 1996.
[61] S. Murphy and H. Bray, “Face recognition fails at Logan,” Boston Globe, September 3, 2003.
[62] P. Agre, “Your face is not a bar code: arguments against automatic face recognition in public places,” Whole Earth 106, pp. 74-77, 2001.
[63] “Face Recognition,” Electronic Privacy Information Center, http://www.epic.org/privacy/facerecognition/.
[64] “Privacy,” American Civil Liberties Union, http://www.aclu.org/Privacy/.
[65] D. Balme (ed.), *Aristotle: Historia Animalium*, Volume 1, Books I-X, Cambridge University Press, 2002.
[66] C. Darwin, *The Expression of the Emotions in Man and Animals*, Oxford University Press, Third edition, 2002.
[67] C. Breazeal, *Designing Sociable Robots*, MIT Press, 2002.
[68] M. Minsky, *The Society of Mind*, Simon and Schuster, New York, 1986.
[69] D. Hubel, *Eye, Brain, and Vision* (Scientific American Library, No 22), W. H. Freeman and Company, 1989.
[70] M. J. Farah, *Visual Agnosia*, MIT Press, Cambridge, MA, 1990.

[71] N. L. Etcoff, R. Freeman, and K. R. Cave, “Can we lose memories of Faces? Content specificity and awareness in a prosopagnosic,” Journal of Cognitive Neuroscience, Vol. 3, No. 1, pp. 25-41, 1991.
[72] H. D. Ellis and M. Florence, “Bodamer’s (1947) paper on prosopagnosia,” Cognitive Neuropsychology, 7 (2), pp. 81-105, 1990.
[73] O. Sacks, The Man Who Mistook His Wife For a Hat, Harper & Row, New York, 1987.
[74] A. R. Damasio, H. Damasio, and G. W. Van Hoesen, “Prosopagnosia: anatomic basis and behavioral mechanisms,” Neurology Vol. 32, pp. 331-41, April 1982.
[75] G. M. Davies, H. D. Ellis, and J. W. Shepherd (eds.), Perceiving and Remembering Faces, Academic Press, London, 1981.
[76] D. I. Perrett, E. T. Rolls, and W. Caan, “Visual neurones responsive to faces in the monkey temporal cortex,” Exp Brain Res, Vol. 47, pp. 329-342, 1982.
[77] D. I. Perrett, A. J. Mistlin, and A. J. Chitty, “Visual neurones responsive to faces,” TINS, Vol. 10, No. 9, pp. 358-364, 1987.
[78] D. I. Perrett, A. J. Mistlin, A. J. Chitty, P. A. J. Smith, D. D. Potter, R. Broennimann, and M. Harries, “Specialized face processing and hemispheric asymmetry in man and monkey: evidence from single unit and reaction time studies,” Behavioural Brain Research, 29, pp. 245-258, 1988.
[79] R. Desimone, T. D. Albright, C. G. Gross, and C. J. Bruce, “Stimulus-selective properties of inferior temporal neurons in the macaque,” Neuroscience 4, pp. 2051-2068, 1984.
[80] E. T. Rolls, G. C. Baylis, M. E. Hasselmo, and V. Nalwa, “The effect of learning on the face selective responses of neurons in the cortex in the superior temporal sulcus of the monkey,” Exp Brain Res, Vol. 76, pp. 153-164, 1989.
[81] K. M. Kendrick and B. A. Baldwin, Science 236, pp. 448-450, 1987.
[82] R. K. Yin, “Looking at upside-down faces,” J. Exp. Psychol., 81, pp. 141-145, 1969.
[83] S. Leehey, “Face recognition in children: evidence for the development of right hemisphere specialization,” Ph.D. Thesis, Dept. of Psychology, Massachusetts Institute of Technology, May 1976.
[84] S. Carey and R. Diamond, “From piecemeal to configurational representation of faces,” Science, Vol. 195, pp. 312-313, Jan. 21, 1977.
[85] R. Diamond and S. Carey, “Why faces are and are not special: an effect of expertise,” J. Exp. Psych: G, Vol. 115, No. 2, pp. 107-117, 1986.
[86] J. L. Bradshaw and G. Wallace, “Models for the processing and identification of faces,” Perception and Psychophysics, Vol. 9 (5), pp. 443-448, 1971.
[87] K. R. Laughery and M. S. Wogalter, “Forensic applications of facial memory research,” in A. W. Young and H. D. Ellis (eds.), Handbook of Research on Face Processing, Elsevier Science Publishers B. V. (North-Holland), 1989.
[88] J. B. Pittenger and R. E. Shaw, “Ageing faces as viscal-elastic events: Implications for a theory of nonrigid shape perception,” Journal of Experimental Psychology: Human Perception and Performance, 1, pp. 374-382, 1975.
[89] “Faces From the Future,” Newsweek, Feb. 13, 1989, p. 62.
[90] D. C. Hay and A. W. Young, “The human face,” in A. W. Ellis (ed.), Normality and Pathology in Cognitive Functions, Academic Press, London, 1982.
[91] A. W. Young, K. H. McWeeny, D. C. Hay, and A. W. Ellis, “Matching familiar and unfamiliar faces on identity and expression,” Psychol Res, Vol. 48, pp. 63-68, 1986.
[92] M. Turk and M. Kölsch, “Perceptual Interfaces,” G. Medioni and S.B. Kang (eds.), Emerging Topics in Computer Vision, Prentice Hall, 2004.
[93] A. J. Goldstein, L. D. Harmon, and A. B. Lesk, “Identification of human faces,” Proc. IEEE, Vol. 59, pp. 748-760, 1971.

[94] A. J. Goldstein, L. D. Harmon, and A. B. Lesk, “Man-machine interaction in human-face identification,” The Bell System Technical Journal, Vol. 41, No. 2, pp. 399-427, Feb. 1972.
[95] T. Sakai, M. Nagao, and M. Kidode, “Processing of multilevel pictures by computer – the case of photographs of human faces,” Systems, Computers, Controls, Vol. 2, No. 3, pp. 47-53, 1971.
[96] M. A. Fischler and R. A. Elschlager, “The representation and matching of pictorial structures,” IEEE Trans. Computers, Vol. c-22, No. 1, pp. 67-92, January 1973.
[97] S. R. Cannon, G. W. Jones, R. Campbell, and N.W. Morgan, “A computer vision system for identification of individuals,” Proc. IECON, Vol. 1, pp. 347-351, 1986.
[98] I. Craw, H. Ellis, and J. R. Lishman, “Automatic extraction of face features,” Pattern Recognition Letters 5, pp. 183-187, 1987.
[99] K. Wong, H. Law, and P. Tsang, “A system for recognising human faces,” Proc. ICASSP, pp. 1638-1642, May 1989.
[100] H. Midorikawa, “The face pattern identification by back-propagation learning procedure,” Abstracts of the First Annual INNS Meeting, Boston, p. 515, 1988.
[101] E. I. Hines and R. A. Hutchinson, “Application of multi-layer perceptrons to facial feature location,” IEE Third International Conference on Image Processing and Its Applications, pp. 39-43, July 1989.
[102] G. W. Cottrell and J.Metcalfe, “EMPATH: Face, gender and emotion recognition using holons,” In R.P. Lippman, J. Moody, and D.S. Touretzky (Eds.), Advances in neural information processing systems 3, San Mateo, CA: Morgan Kaufmann.
[103] T. J. Stonham, “Practical face recognition and verification withWISARD,” in H. Ellis, M. Jeeves, F. Newcombe, and A. Young (eds.), Aspects of Face Processing, Martinus Nijhoff Publishers, Dordrecht, 1986.
[104] S. Sclaroff and A. Pentland, “Closed-form solutions for physically-based shape modeling and recognition,” Proc. CVPR, Maui, Hawaii, pp. 238-243, June 1991.
[105] V. Bruce and A. Young, “Understanding face recognition,” British Journal of Psychology, 77, pp. 305-327, 1986.
[106] W. Zhao, R. Chellappa, P. J. Phillips, and A. Rosenfeld, “Face recognition: a literature survey,” ACM Computing Surveys, Vol. 35, Issue 4, pp. 399-458, Dec. 2003.

**Figure Captions**

Figure 1. A simple example of principal component analysis. (a) Images with three pixels are described as points in three-space. (b) The subspace defined by a planar collection of these images is spanned by two vectors. One choice for this pair of vectors is the eigenvectors of the covariance matrix of the ensemble, u1 and u2. (c) Two coordinates are now sufficient to describe the points, or images: their projections onto the eigenvectors, ($\omega$1, $\omega$2).

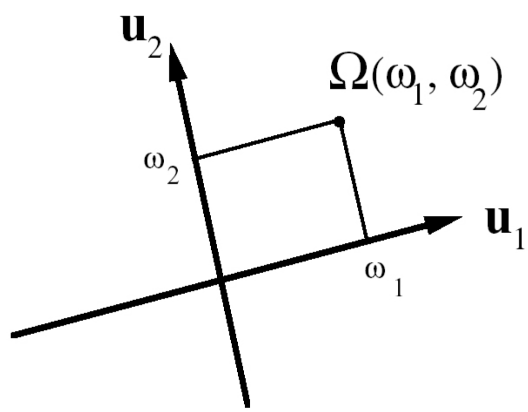
Figure 2. (a) Partially occluded face image from the test set and (b) its projection onto face space. The occluded information is encoded in the Eigenfaces. (c) Noisy face image and (d) its face space projection.

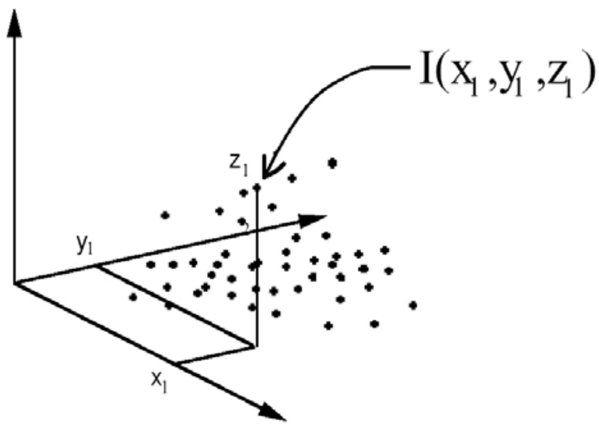
Figure 3. The average face image $\bar{\boldsymbol{x}}$ and a set of Eigenface images. The Eigenfaces are real-valued images scaled so that a value of zero displays as a medium gray, negative values are dark, and positive values are bright.

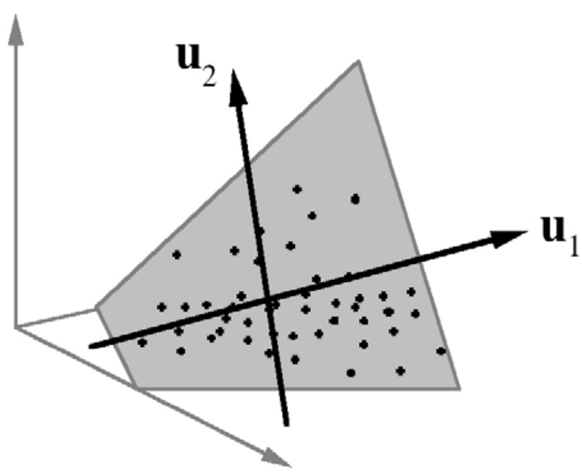
Figure 4. (a) Original image. (b) Corresponding "distance from face space" (DFFS) map, where low values (dark areas) indicate the likely presence of a face.

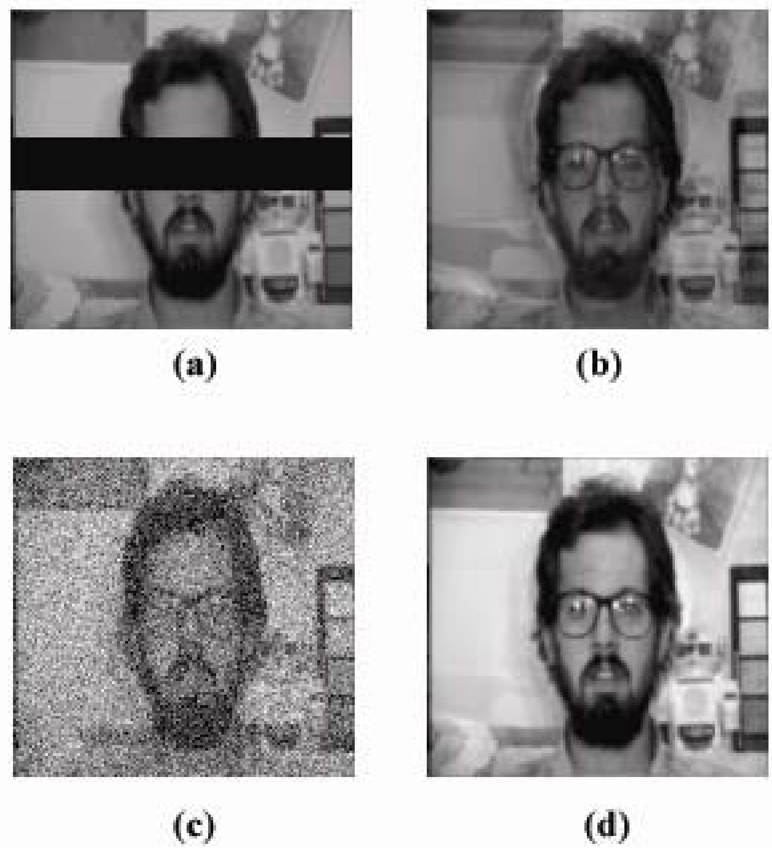
Figure 5. The initial implementation of the full almost-real-time Eigenfaces system, using simple motion analysis to restrict the search area (for the DFFS calculation).

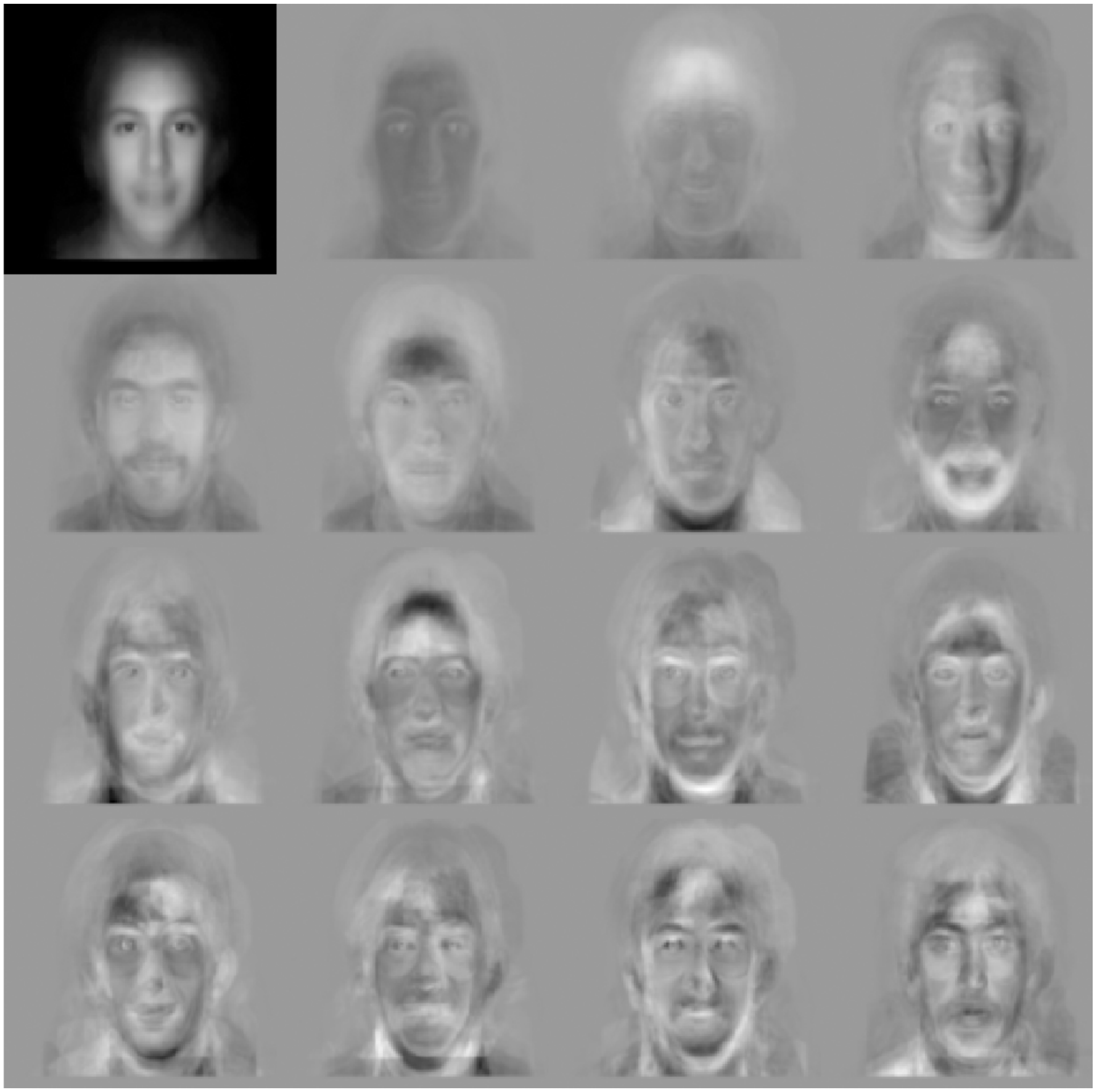
Figure 6. Two methods to reduce or eliminate the effect of background. (a) An original face image. (b) Multiplied by a Gaussian window, emphasizing the center of the face. (c) Multiplied by a binary face mask outlined by the operator (while gathering the training set).

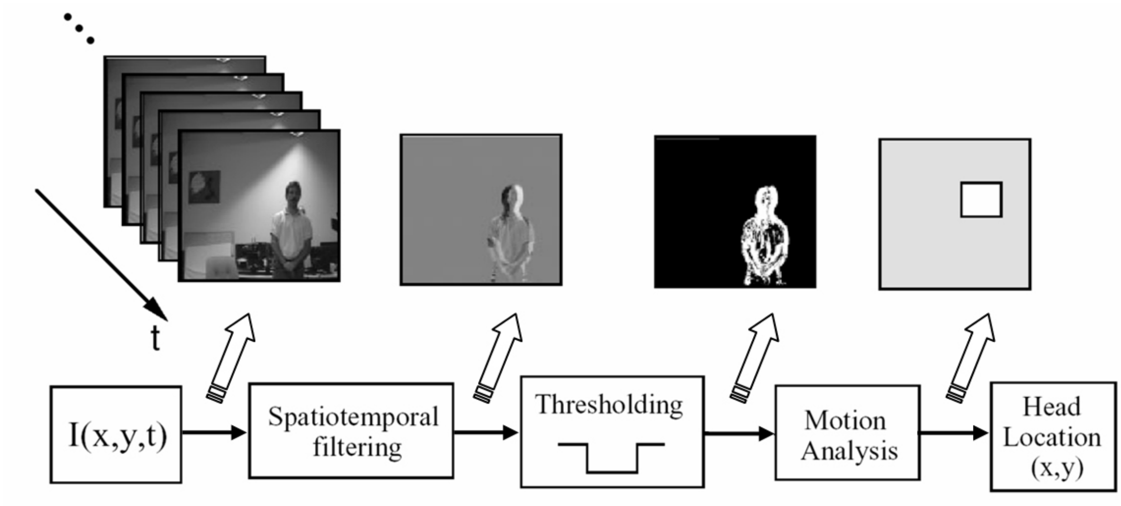

u2
Ω(ω1, ω2)
ω2
u1
ω1


I(x1 ,y1 ,z1 )
z1
y1
x1


u2
u1


(a)
(b)
(c)
(d)




t
I(x,y,t)
Spatiotemporal filtering
Thresholding
Motion Analysis
Head Location (x,y)