COLOR AND FACE RECOGNITION

by

Janet Dueck

Bachelor of Science (B.Sc.)
Computing Science
Simon Fraser University 1993

A THESIS SUBMITTED IN PARTIAL FULFILLMENT
OF THE REQUIREMENTS FOR THE DEGREE OF
MASTER OF SCIENCE
in the School
of
Computing Science

SIMON FRASER UNIVERSITY
September 1995

APPROVAL

Name: Janet Dueck

Degree: Master of Science

Title of thesis: Color and Face Recognition

Examining Committee: Dr. Tom Calvert
Professor, Computing Science, SFU
Chair

Dr. Stella Atkins, External Examiner
Associate Professor, Computing Science, SFU

Dr. Brian V. Funt, Senior Supervisor
Professor, Computing Science, SFU

Dr. Ze-nian Li, Supervisor
Associate Professor, Computing Science, SFU

Date Approved: September 28, 1995

SIMON FRASER UNIVERSITY

PARTIAL COPYRIGHT LICENSE

I hereby grant to Simon Fraser University the right to lend my thesis, project or extended essay (the title of which is shown below) to users of the Simon Fraser University Library, and to make partial or single copies only for such users or in response to a request from the library of any other university, or other educational institution, on its own behalf or for one of its users. I further agree that permission for multiple copying of this work for scholarly purposes may be granted by me or the Dean of Graduate Studies. It is understood that copying or publication of this work for financial gain shall not be allowed without my written permission.

Title of Thesis/Project/Extended Essay

Color and Face Recognition.

Author: _____

(signature)

Janet Dueck

(name)

November 16, 1995

(date)

Abstract

The machine recognition of faces is useful for many commercial and law enforcement applications. Two typical examples would be security systems and mug-shot matching. A real-time method which has been developed in the last few years is the eigenface method of recognition. The eigenface method uses the first few principal components (the eigenfaces) of the database images to characterize the known faces. Images are classified by their weights, the weights are found by projecting each image onto the eigenfaces.

The goal of this thesis was to improve the recognition of faces by using color. We started by looking at the limitations of the eigenface method as applied to grey-scale images. Next, color ratios, chromaticities and color band normalized images were used. Images were compared using both the eigenface method and doing a direct picture-to-picture comparison. Last but not least, a method was developed using color which would correct for illumination direction when there are gross differences in illumination between two images.

For similar images, ie. images in which there was little variation in head size, orientation or illumination, the eigenface method with grey-scale images performed very well with a recognition rate of 95%. Of the color representations that were tried, only color band normalized images performed as well as the eigenface method for grey-scale images. When there was a gross change in the illumination the performance of the eigenface method declined to a recognition rate of 73%. Correcting for the illumination differences through the use of color allowed reliable recognition independent of illumination.

Acknowledgments

Several people have contributed to this work, and I would like to thank them.

I would like to thank my senior supervisor, Dr. Brian Funt. His support and guidance were essential to the development of this work.

Drs. Graham Finlayson and Mark Drew, as well as the other students, Subho and Kobus, were fun and stimulating to work with. Dr. Finlayson has made valuable contributions to both the content of the work and my approach to it. Dr. Drew has encourage me in many ways, but the most important to me is his willingness to spend long hours discussing statistical analysis. I feel fortunate to have worked with these people.

I would also like to acknowledge the financial support of the Natural Sciences and Engineering Research Council of Canada and Simon Fraser University.

Last, but not least, this work would not have been possible without the support and encouragement of my husband, Scott, and the understanding of my son, Neil.

Thank you all.

Contents

List of Tables

List of Figures

Chapter 1

Introduction

People can distinguish thousands if not millions of faces. We can identify faces under a myriad of conditions—with changes in color and direction of illumination, at different distances, from different perspectives, with changing facial expressions or with cosmetic differences such as haircuts and makeup. Complete recognition involves high cognitive functions, taking into account such features as color, context, characteristic movements and voice. Ultimately, machine vision seeks to duplicate this complex process. To date, no system exists which can do all of these things. We are seeking an analytic method for face recognition.

This thesis addresses one small portion of the problem, the identification of frontal face images taken under varying illuminations, using image processing techniques. The face size and position do not vary from image to image. The goal of this thesis was to test whether the use of color could improve face recognition.

Turk and Pentland [25, 24] have developed a near real-time face recognition technique for grey-scale images which they call eigenfaces. The eigenface method has been used as the metric against which all other techniques are compared. The starting point for my work was an exploration of the limitations of this method. I found that the eigenface method works well when applied to images with a fixed head location, size and orientation, and when there is a limited variation in the illumination conditions under which the pictures were taken. It does not work as well when the changes in illumination are more pronouced, for example when one side of the face is

quite heavily shaded.

Color provides additional information that we may be able to exploit to improve recognition. Work has been done which shows that if an object is colorful enough then color alone can be used to identify it. The limited colors seen in the face make it unlikely that color can be used as the only identifying characteristic—although it could perhaps be used to segment a database of face images. A color image of a face contains both color and shape information, the shape information being implied by the changes in pixel intensity. The properties of color may be exploited to obtain quasi-invariant descriptors of the shape. Color edges, chromaticities and color band normalization emphasize different aspects of the image; all three representations will be used for recognition.

When we compare two images of a face, differences in the illumination conditions will change the information in the image. Any method which does a comparison between images will break down unless the change in illumination is taken into account. Color can be used under certain conditions to correct illumination differences, so that the comparison can be made between images with equivalent lighting conditions.

1.1 Goals

The goal of this thesis was to test whether the use of color could improve face recognition by machines. The specific goals were:

1. to determine the limitations of the eigenface method
2. to test whether color alone is a sufficient basis on which a face database could be segmented. This was done using color angle invariants.
3. to see whether color can be used to improve the eigenface method. This was done by using the eigenface method on chromaticity, color ratio and color band normalized images.
4. to use color to improve recognition rates under grossly varying illumination conditions.

1.2 Thesis Outline

1. Introduction

   In this introductory chapter, the problem of face recognition is described and a brief outline of the work contained in this thesis given. This chapter also describes the layout of the thesis itself, including a summary of its contents.

   - Goals and Approach
   - Thesis Outline

2. Literature Survey

   Chapter 2 surveys some of the previous work that uses principal component analysis for face recognition. A brief survey of color representation and color constancy is also included.

   - Using Eigenvectors for face representation and recognition
   - Eigenfaces for Recognition
   - Color

3. The Eigenface Method

   Chapter 3 describes the eigenface method developed by Turk and Pentland. Experimental work is done with Turk and Pentland's original data. The goal of this chapter is to explore the limitations of the method.

   - The Eigenface Method
   - The Data Set
   - Experimental Results

4. Color and Face Recognition

   Chapter 4 explores some of the ways in which color might be used in conjunction with the eigenface method.

- The Data Set
- Color Angle Invariants
- Color Representations
  - Color Band Normalization
  - Chromaticity
  - Color Ratios
- Experimental Results

5. Color and Changing Illumination

   Chapter 5 describes the relationship between a single-colored object and its shape. It also shows how this relationship can be used to correct for illumination differences in images of the same single-colored object.

   - Properties of color images of surfaces under multiple illuminants
   - Adjusting for changing illumination

6. Illumination Correction and Faces

   Chapter 6 describes the application of the relationship developed in chapter 5 to face images and presents results of test of the methods developed.

   - Implementation
   - Data Set
   - Comparison of Methods
   - Experimental Results

7. Discussion

   In chapter 7 the work is summarized and discussed.

Chapter 2

Literature Survey

2.1 Face Recognition

There are many problems which fall under the heading of face recognition, each with its own technical challenge. Chellappa et al. [4] divide the application of face recognition into three categories:

1. applications which require matching one face image to another face image
2. applications for finding or making a face image which is similar to one which is remembered
3. applications for generating a face image from data (for example, computerized aging for identifying missing children and reconstruction of the face for the identification of remains).

We will addressing only techniques which fall into the first category. Typical applications of the first category are mug-shot matching or security systems. Mug-shot matching has the advantage of having good image quality with images taken under controlled conditions. A security system may not present the same amount of control as the mug-shot case.

In general, two approaches are taken to mug-shot matching. The first applies a global transform to the entire image. The second bases its recognition on features

extracted from the image. The work contained in this thesis uses the global approach.

2.1.1 Using Eigenvectors

Sirovich and Kirby [20] have developed a representation for a face image in which the face is characterized by a relatively low-dimensional vector. Using the Karhunen-Loeve expansion (also known as Principal Component Analysis) they show that roughly 40 numbers are needed to characterize a single face image from a data set of 115 to within 3% error. The error ϵ is defined to be

$$\epsilon_N = \frac{\|\varphi - \varphi^N\|}{\|\varphi\|}$$

where φ is the original image and φ^N is the reconstruction of the image using N eigenvectors. They claim that, in principle, any collection of face images could be represented by a small set of numbers for each face image and a small set of eigenvectors.

Hallinan [10] has taken the use of Principal Component Analysis (PCA) one step further and developed a model for the face which explains the illumination conditions under which an image is taken. The model splits the variations in image intensity due to illumination into images which combine linearly (boundary images). PCA is applied to the boundary images and the resulting eigenvectors are used to approximate an arbitrary range of lighting conditions. Hallinan reports that the first five eigenvectors consistently represent five common lighting conditions. For example, the 0^{th} basis represents the contribution made by frontal/ambient lighting.

Cheng et al. [5] have developed a method for face recognition which uses Singular Value Decomposition (SVD) and thresholding of eigenvalues. The average face is calculated using three images of a person's face (the training set). The eigenvectors and eigenvalues are determined for this average face by SVD. Thresholding is used to discard those eigenvectors with eigenvalues close to zero. Feature vectors are calculated by projecting each of the training set images for a person onto that person's

eigenvectors. The feature vector used to identify person A is the average of all the feature vectors calculated for person A. A new image to be identified is projected onto the eigenspace spanned by the feature vectors and identification is made using the Frobenius norm. The authors report a 100% recognition rate for a database of 8 people, using eight pictures for each person. Further work would be needed to show whether this method would work well with a large face image database.

2.1.2 Eigenfaces

Turk and Pentland [24] have used the results of Sirovich and Kirby as the basis of their eigenface recognition method. The eigenface method is used both to recognize and to find a face in a scene. Turk and Pentland's starting premise was that if 40 eigenvectors are needed to characterize the facial information of 115 people, fewer should be needed to recognize them. The details of how to find the eigenvectors (which Turk and Pentland call eigenfaces) are described in chapter 3. Once the eigenfaces have been found, each image in the database is projected onto the eigenfaces, producing a vector of weights. These weights form a face-class vector and are used to identify the face. When a new image is to be identified, it is also projected onto the eigenfaces. The resulting weight vector is compared to the face-class vectors and identified as the person whose face-class vector is closest in Euclidean distance. Faces are discriminated from non-faces based on the distance between the new image and face space; if the new image is too far away it must not be a face.

Experiments were conducted on a database of 2592 images of 16 people. Each person's picture was taken with all combinations of three different head sizes, three different head orientations and three different illumination directions. The images were then put into a 6-level pyramid. Turk and Pentland have chosen to use 7 of a possible 16 eigenvectors, one for each person in their data set. These 7 eigenvectors encode approximately 80% of the available information. They report that the eigenface approach is robust under changes in illumination but that it degrades when there are changes in size or orientation.

Pentland et al. [15] have extended the capabilities of the eigenface system in several ways. The database size has increased to 7562 images of approximately 3000 people and is annotated with additional information about sex, race, age and facial expression. For several people the database contains many images with different facial expressions, hair styles, etc. Twenty eigenfaces were calculated from a randomly chosen subset of 128 people.

A recognition rate of approximately 95% was achieved when the method was tested using 200 selected images. Recognition accuracy as function of race was also tested with recognition rates of 90%, 95% and 80% for white, black and Asian males respectively. In addition to recognition, the authors also addressed interactive database search. Asked for images of certain types (ie. adult Hispanic males) the system presents images which fit this criteria in groups of 21. The user can choose any one of these images, and the system will present the 21 images which are most similar to it. The similarity search uses the eigenface descriptors.

The authors took two approaches to multiple views of the same person. The first was to pool all of the images and build a set of eigenfaces which would represent all images from all views. The second built a separate eigenspace for each characteristic view; this is known as view-based eigenfaces and seems to work better than pooling all images from all views.

The eigenface technique was also extended to describe and encode facial features. The eyes, nose and mouth were detected in the same way as faces were detected using eigenfaces. On a limited data set (45 people, 2 views per person), recognition was tested using eigenfaces only, eigenfeatures only and the two combined. With a low number of eigenvectors, eigenfeatures performed better than eigenfaces. Combining the two techniques produced only a slight improvement over eigenfeatures alone.

Tistarelli [23] has used a space-variant descriptor with the eigenface method. People foveate on those parts of an object which contain the most interesting features. Tistarelli has mimicked the foveation of the eye with a vision system based on a retina-like space-variant CCD sensor like that used by Sandini and Dario [19]. The sensor

is fixated at three distinct points on the face and the resulting images are used to compute the eigenfaces. Projecting each of the three images back onto the eigenfaces gives a vector of descriptive triplets. Each known person in the database is represented by this vector of triplets and an unknown individual is identified by finding the minimum Euclidean distance between the vectors of triplets in the database and the vector of triplets for the image we wish to identify. The author reports a 100% recognition rate for a training set comprised of seven people. Further work would be needed to assess whether space-variant eigenfaces could scale up for use with a larger data set.

2.2 Color

The human eye has the ability to see colors as relatively constant under changes of illumination. This is not true for color pictures taken with a camera; a red sweater will not appear to be the same red in two images if the images were taken under different illuminations. Having the computer "see" the color of the sweater as the same color for images taken under different illuminations is the goal of color constancy. Intuitively it seems that color should be able to help the computer recognize a face. We will need to determine whether this is true, and to do so an illumination independent description of the colors in the face image must be used.

How do we describe color? In the classic "desert island" experiment, colors are described as having hue, lightness and chroma [2, pp. 18–19]. Suppose that someone who knows nothing about color is on a desert island surrounded by colored pebbles. To while away the time, the person decides to arrange the pebbles by their color in some logical way. How might they do this? First the pebbles are divided into two piles, one chromatic, the other achromatic. The achromatic pile is sorted by its lightness, ie. into all the shades from black to white. The chromatic pile is then sorted by hue, or into piles of what we commonly think of as red, yellow, blue etc. Each of these piles is then ordered by lightness. At this point the person notices that some

pebbles which the same hue and lightness are different. This difference is the third descriptor of color, the chroma. The chroma describes how much color each stone has.

The von Kries model of chromatic adaptation can be described as a diagonal linear transform between sensor responses for sensor responses under two different illuminations. This model will not work for wide-band sensors; however, if the sensor response functions are first transformed to a more narrow-band sensor basis then the von Kries model is acceptable for color constancy [7].

Finlayson et al. [6] have used this result to show that color angle invariants are sufficient to recognize objects if the objects are colorful. Color angles are invariant when a linear diagonal transformation is sufficient to describe the relationship between two images taken under different colors of illumination, which is the case when the sensors are sufficiently sharp. Under these conditions a change in the color of the illumination will induce a change in the length of the red, green and blue vectors, but the angle between the vectors remains the same.

If an object is colorful then color alone can be used to recognize it. Swain and Ballard [22] have shown that under controlled illumination, objects can be identified solely on the basis of the histogram of their colors. Under varying illuminations their method deteriorates and they recommend that correction for color constancy be applied before processing. Finlayson and Funt [8] have extended the work by Swain and Ballard by histogramming of same-band color ratios. Color ratios are shown to be invariant to changing illuminations.

Petrov [17] has suggested that all perceived surface colors can be described as a set of nine-dimensional 3x3 matrices. He defines the color C of a surface σ as linear function $C \cdot \sigma_0 = \sigma$, where C is a 3x3 matrix and σ_0 is a white surface with the same shape as σ. This equation assumes a fixed illumination.

Petrov [16] uses this definition of color to relate an image to the real world. Here the world is viewed as having three distinct characteristics: shape, described by the normal of the surface, $\vec{n}$ (as a function of x and y); color, C, as described above; and illumination. Illumination consists of two components, H_0, which is attributable to three chromatic light sources and h^d to diffuse lighting. These characteristics are related to the image P by the equation

$$C \cdot (H_0 \cdot \vec{n}(x,y) + h^d) = P(x,y). \tag{2.1}$$

Petrov uses Equation 2.1 to solve for the shape of the object.

Petrov and Kontsevich [18] use Petrov's description of color and its relationship to a surface to define the properties of color images of surfaces under multiple illuminants. In this work, the sensor response is

$$\rho = M\vec{n}. \tag{2.2}$$

M is nine-dimensional, 3x3 matrix which linearly maps the shape of the object to the receptor response. This work will be discussed in more detail in chapter 5.

Chapter 3

The Eigenface Method

The eigenface method of face recognition assumes that faces can be recognized by how different they are from an "average face". Consider a set of NxN images. Each NxN image is a point or vector in N^2 space. However, face images are very similar and should produce a small cloud of points in this very large space. The eigenface method uses the difference from an average face to emphasize differences rather than similarities. What we would like to do is find a small set of basis vectors which represent a face subspace of our image space. The application of principal component analysis (PCA) reduces the number of required basis vectors to a very small number.

The eigenface approach was motivated by work done on image compression by Sirovich and Kirby [20]. In this work they showed that, by using PCA, they could reconstruct approximately 98% of the image data found in 115 face images from only 40 eigenvectors. The amount of lost information was determined by

$$\epsilon_N = \frac{\|\varphi - \varphi^N\|}{\|\varphi\|}$$

where φ is the original image and φ^N is the reconstruction of the image using N eigenvectors. Turk and Pentland investigated the possibility that even fewer eigenvectors are sufficient to identify a face, using seven of the sixteen possible eigenvectors in their work. (Their data set was comprised of images of sixteen people.)

3.1 The Eigenface Method

Let Γ be an NxN face image. We will view Γ as a vector of length N^2. Let our training set be comprised of M images; Γ_1, Γ_2, . . . ,Γ_M. Each of the faces in the training set is assumed to be centered, to be the same size, and to have the same orientation and illumination. The average face image, Ψ, is

$$\Psi = \frac{1}{M}\sum_{i=1}^{M}\Gamma_i$$

and $\Phi_i = \Gamma_i - \Psi$ the deviation of face image i in the training set from the average face image. An N^2xN^2 matrix A can be formed with the Φ_i's as its columns.

$$A = [\Phi_1\Phi_2...\Phi_M]$$

We are looking for a few ($<< N^2$) basis vectors which preserve most of the information given by the variation in the vectors of A. Principal component analysis will give us this information [12].

What is principal component analysis? Let $\mathbf{x}$ be a vector of length N. Principal component analysis first looks for N-vector α_1 such that $|\alpha_1| = 1$, $\alpha_1^T\mathbf{x}$ has maximum variance and

$$\alpha_1^T\mathbf{x} = \alpha_{11}x_1 + \alpha_{12}x_2 + ... + \alpha_{1N}x_N.$$

It next looks for a linear function $\alpha_2^T\mathbf{x}$ which has maximum variance and is uncorrelated with $\alpha_1^T\mathbf{x}$. The third principal component α_3, is similarly found by looking for a linear function $\alpha_3^T\mathbf{x}$ which has maximum variance and is uncorrelated with $\alpha_1^T\mathbf{x}$ and $\alpha_2^T\mathbf{x}$. The k^{th} such derived vector is the k^{th} principle component. The k^{th} principal component of A is given by maximizing

$$\lambda_k = \frac{1}{M}\sum_{n=1}^{M}(\alpha_k^T\Phi_n)^2$$

and subject to

$$\alpha_k^T\alpha_l = \begin{cases} 0 & k \neq l \\ 1 & k = l \end{cases}$$

Another way of mathematically describing the same basis vectors is as the eigenvectors of the covariance matrix of A. By definition the covariance matrix C = AA^T. Consequently, C is a real symmetric matrix and can be factored into $C = Q\Lambda Q^T$, where the columns of Q are the orthonormal eigenvectors and Λ contains the eigenvalues of C [21, pp. 290–297]. Turk and Pentland call the eigenvectors eigenfaces because they resemble ghostly faces.

Once we have the eigenfaces, α_k, face classes are defined for the images in the training set. Face classes are specified by a vector Ω_i,

$$\Omega_i = [\omega_1, \omega_2, ..., \omega_{M'}],$$

where each $\omega_k = \alpha_k^T \Phi_i$ and $M' < M$. M' refers to the number of eigenfaces with which we have chosen to do our recognition. Fewer than M eigenfaces may be used because principal component analysis preserves most of the information in the first few eigenfaces. For example, choosing M' to be 7 preserves approximately 80% of our image information. ω_k is just a projection of Φ_i onto the k^{th} eigenface basis vector.

To identify a face Γ, we find the face class which minimizes the Euclidean distance

$$\varepsilon_k = ||(\Omega - \Omega_k)||^2$$

where $\Omega = [\omega_1, \omega_2, ..., \omega_{m'}]$ with $\omega_k = \alpha_k^T(\Gamma - \Psi)$ and Ω_k is the vector describing the k^{th} face class.

The computation of the eigenvectors of the covariance matrix is prohibitively expensive because the covariance matrix is N^2xN^2. A fast way to find the eigenvectors of C is to compute them from the eigenvectors of matrix $L = A^T A$ [24]. To find the eigenvectors of the covariance matrix using L, let ν = the eigenvectors of L. Then,

$$L\nu = \lambda\nu$$

$$A^T A\nu = \lambda\nu$$

$$AA^T A\nu = A\lambda\nu$$

$$CA\nu = \lambda A\nu.$$

If α are the eigenvectors of C, where $C = AA^T$, then $\alpha = A\nu$ and we can calculate the eigenfaces by first finding the eigenvectors of L, which is computationally a much simpler problem.

Summarized Procedure

- Collect a set of N face images for known individuals.
- Calculate the NxN matrix L, find its eigenvectors and eigenvalues and choose the M' eigenvectors with the highest eigenvalues.
- Calculate the eigenfaces using the eigenvectors of L and the mean adjusted face images in the training set.
- For each known individual calculate the face class vector Ω_k
- For each new face to be identified calculate its image vector Ω and find the distance between Ω and each of the face classes.

Implementation

I implemented the eigenface method of face recognition in C using the image processing package *hips*[13]. The code used by Turk and Pentland was available over the network but would not run at our installation since their code depended heavily on other in-house software which was not available over the network. I chose to rewrite the program using locally available software. *S-PLUS* [11], a math package, was used to find the eigenvectors of the correlation matrix L.

I encountered two main problems while implementing the method. The first was a problem with rounding errors. As the values of the eigenvalues of the covariance matrix decreased so did the length of the eigenvectors. After normalizing the eigenvectors the smallest of these basis vectors was no longer orthogonal. I got around this problem by checking the length of each eigenvector and discarding those which were too short. The second problem was finding a system which would run the experiments reliably. While running, each experiment required approximately 3 Meg of RAM but this memory was not constantly held—it was allocated and freed as needed. If someone else was on the system and used that memory in the short time that it was free then the experiment would abort and I would be forced to start over again.

Each experiment consisted of trying to identify all 432 face images with each of the 27 possible training sets of sixteen people. This is equivalent to trying to identify 11,664 face images. The pared down code produced an output file giving the "recognized" face for each image. On a Sun SS10-41 approximately 2 images were processed per second, while on a Sun SparcLX the rate decreased to approximately 1 image per second.

3.2 The Data Set

The data set used for these experiments consisted of 432 images, 27 images each of sixteen people. Each person's picture was taken with three head sizes (large, medium and small), three head orientations (upright, tilted to the left, and tilted to the right) and with three illuminations (light along the optical axis, at 45 degrees to the optical axis and at ninety degrees to the optical axis). The work originally done by Turk and Pentland placed the 432 images into a six-level pyramid[1] giving them more than 2500 images. The second level of the pyramid was available on the Internet and this is what I used for my experiments. Figure 3.1 (from [24]) shows all 27 images for one person in the data set.

Each training set consisted of 16 images, one for each person in the data set. All 16 images in the training set had the same head size, orientation and lighting. Figure 3.2 (from [24]) shows an example of the training set consisting of large upright heads illuminated along the optical axis. I evaluated the data in five different ways:

1. original images,
2. images scaled by a gaussian with $\sigma = 100$. Let $I(x,y)$ be a face image. Then the scaled image, $S(x,y)$,will be

$$\begin{aligned} S(x,y) &= I(x,y) * G(x,y) \\ &= I(x,y) * \frac{1}{2\pi\sigma^2} e^{\frac{-(x^2+y^2)}{2\sigma^2}} \end{aligned} \tag{3.1}$$

3. images scaled (Equation 3.1) by a gaussian with $\sigma = 50$,
4. images with the background masked out and
5. images cropped with the center of the image positioned at the tip of the nose.

The scaling, masking and cropping were all used to get rid of some of the background information. Turk and Pentland apply a gaussian to the images to mimic the effect

[1] An image pyramid consists of a set of images where the dimensions of the image at each level is half that of the level below it. At level 0 the image is NxM, at level 1 it is $N/2$x$M/2$, at level 2 it is $N/4$x$M/4$, etc.

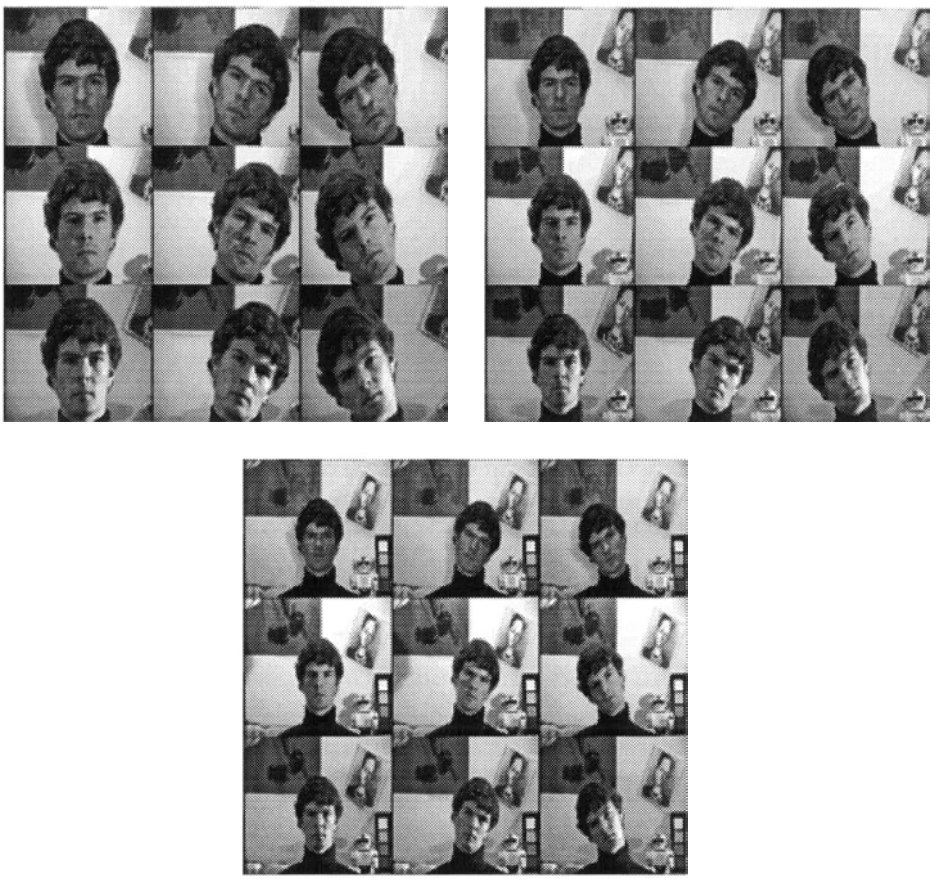

Figure 3.1: The 27 images for one person in the data set

Figure 3.2: Training Set Example

of foveation by reducing the intensity of the background pixels. Figure 3.3 shows an example of one face cropped and masked.

Masking and cropping were done to verify the recognition rates for faces only. Initially, my experiments were made with data which had been scaled by a gaussian with sigma = 50. While the overall recognition rates were similar to those reported by Turk and Pentland, the breakdown of the results made me wonder whether the results were from recognizing the face or the background. Principal component analysis is not invariant to translation, rotation or scaling and so the large variations in the placement of the face in the image were removed by centering the cropped images on the tip of the nose.

3.3 Experimental Results

In all experiments, seven of the possible 16 eigenfaces were used. Each experiment consisted of comparing all 432 faces to each of the 27 possible training sets. The experiments were run five times, once for each type of image data. The results for all five runs can be found in Tables 3.1– 3.3. These results include only the recognition

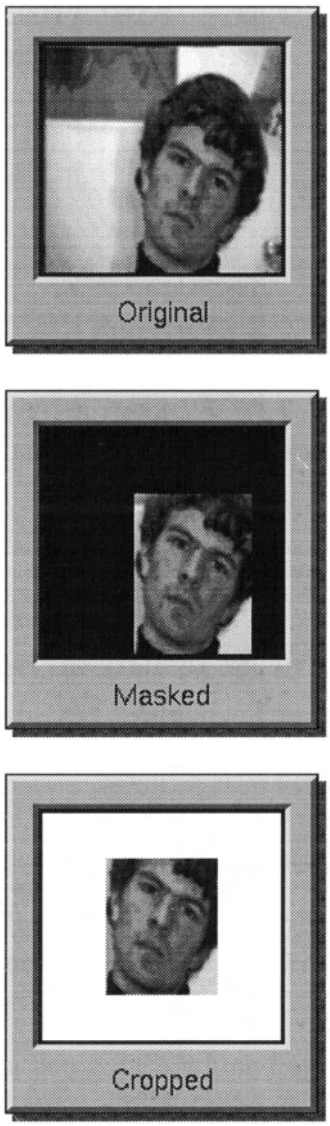

Figure 3.3: Original, Masked and Cropped Images

Table 3.1: Recognition Rates (in %) for Varying Illumination

	Overall	Large	Med	Small
Original	93.5	87.1	93.4	100
Gaussian(σ=100)	92.8	85.4	93.1	100
Gaussian(σ=50)	89.8	80.6	89.6	99.3
Masked	76.5	68.4	77.4	83.7
Cropped	78.4	66.7	78.5	89.9

Table 3.2: Recognition Rates (in %) for Varying Face Size

	Overall	Large	Med	Small
Original	34.0	28.5	48.6	25.0
Gaussian(σ=100)	32.4	27.1	44.4	25.7
Gaussian(σ=50)	28.1	23.9	36.1	24.3
Masked	12.2	12.1	11.8	12.5
Cropped	18.9	18.4	22.9	15.2

rates for faces which varied from the training set by one variable, so that the effect of varying each variable could be determined. The results reported in Tables 3.1–3.3, Figure 3.4 and Figure 3.5 do not include the recognition of the faces that were already in the training set. In all cases, the members of the training set were identified correctly and I felt that excluding these numbers would give a clearer picture of the reliability of the method. Figure 3.6 shows the same results as Figure 3.5 but includes the recognition of the training set members in the percentages as in the original results reported by Turk and Pentland.

The first experiment was run using data which had been scaled by a gaussian with $\sigma = 50$ (Equation 3.1). The results were interesting when broken down by head size (See Figure 3.4). As you can see from Figure 3.4, the rate of recognition increased as the size of the head decreased when the illumination or orientation of the face was varied. As we will see a little later, this trend of increased recognition as head size decreased was true for all experiments when the illumination or orientation varied.

Table 3.3: Recognition Rates (in %) for Varying Face Orientation

	Overall	Large	Med	Small
Original			63.2	95.1
Gaussian(σ=100)	60.9	33.3	57.3	92.0
Gaussian(σ=50)	47.6	24.7	39.6	78.5
Masked	10.7	8.7	7.9	15.6
Cropped	21.2	17.1	18.4	28.1

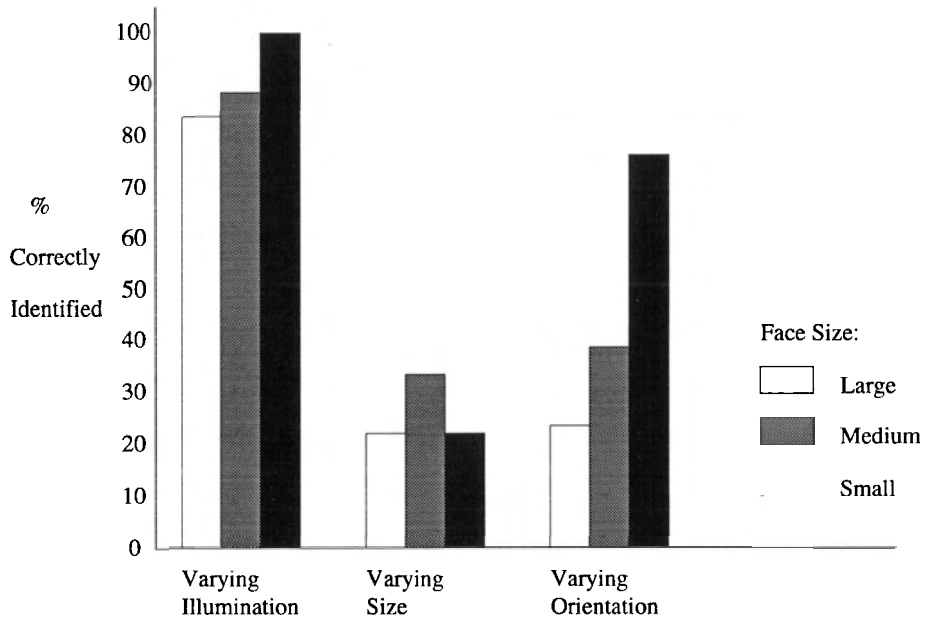

Figure 3.4: % of Images Correctly Identified (Excluding Training Set Data)

The results shown in Figure 3.4 for varied head size were as expected. It is reasonable that a training set comprised of medium faces will have a better chance of recognizing either a large or a small face than does a training set comprised of large faces of recognizing a small face or vice versa, and the results showed that this was indeed true.

I next ran the experiments on original data and with the data scaled by a gaussian with $\sigma = 100$ (Equation 3.1). Both of these experiments provided more background information than the images scaled with $\sigma = 50$. If the background was indeed being recognized as I hypothesized, then the recognition rate should improve. It did.

Next, I ran the experiment on the masked data. While I was masking the images I noticed that the faces were not uniform in their placement in the images. There were gross differences between heads of the different orientations and sizes. Faces for which only illumination varied were not uniform in placement but the differences were quite small (a reference point was within 5 or 6 pixels). Because of these two considerations cropped images were centered on the tip of the nose.

Figure 3.5 shows the overall recognition rates for all five methods. The results clearly identify that the background information is used to help identify faces for the original images and images scaled with a gaussian. This can be seen in the decreasing recognition rate as the amount of background information decreases, that is, as one moves from the original image to images scaled with a gaussian with $\sigma = 100$ to images scaled with a gaussian with $\sigma = 50$ to the masked images. The increase in recognition from the masked to the cropped images is due to the centering of each face and shows quite clearly that principal component analysis is not invariant to translation. The minimal improvement in recognition rates when going from masked to cropped images for varying illumination is due to the small variation in the placement of the face within the image when only the illumination varied. The much larger improvement in recognition rates when going from masked to cropped images for varying size or orientation is due to the large variation in the placement of the face within the image.

Table 3.1 shows both the overall recognition rates and the recognition rates broken down by the size of face for variation in illumination. For all experiments, the recognition rate increases as the size of the face moves from large to small. This is

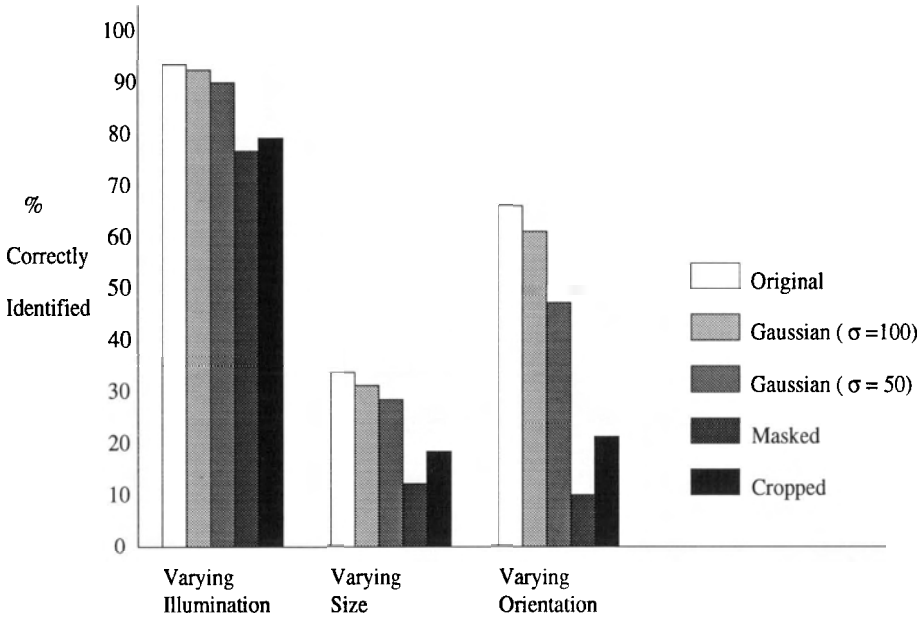

Figure 3.5: % of Images Correctly Identified (Excluding Training Set Data)

expected since masking or cropping a large face removes more of the background (ie. hair or shirt) than masking or cropping a small face.

Table 3.2 shows the overall recognition rates and the recognition rates broken down by face size for variation in size. For all 5 experiments, a medium face recognized its neighbors better than a large face recognized a small one or vice versa.

Table 3.3 shows the overall recognition rates and the recognition rates broken down by face size for variation in face orientation. Once again, the recognition rate increases as the size of the face moves from large to small.

Figure 3.6 shows the overall recognition rates for all 5 methods when the recognition of the faces in the training set is included in recognition data. The trends in Figure 3.6 are the same as those shown in Figure 3.5 although the overall recognition rates are much higher. This was included for comparison purposes only because this is the method of reporting chosen by Turk and Pentland.

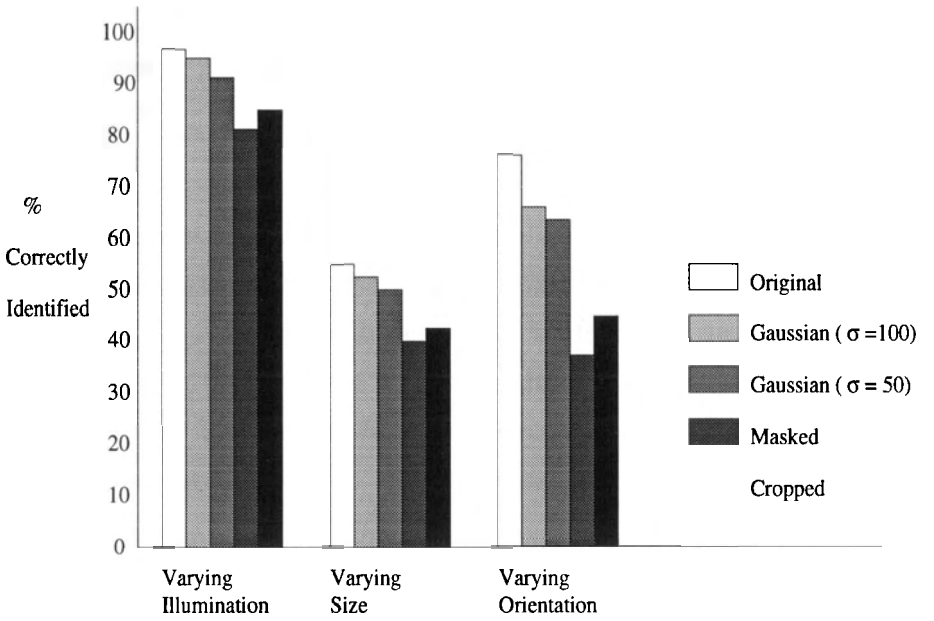

Figure 3.6: % of Images Correctly Identified (Including Training Set Data)

Discussion

This work implements work previously done by Turk and Pentland on face recognition using eigenfaces. This method is fast and my experiments gave comparable results. However, a careful examination of these results illustrated the limitations of the eigenface method of recognition when identifying a face which differed from the known face in the training set by size, illumination or orientation. This work does not show how many faces could be reliably identified or how well the method would work if these three parameters were controlled.

When the eigenface method of face recognition was used to recognize faces which varied in size or orientation the recognition rates were very poor and would not be acceptable in any practical application. If a reliable technique was found which could overcome the principal components' variance with translation of the face, it might be possible to recognize these images in log polar space. Tistarelli [23] has used an active/space-variant sensor for eigenface recognition of faces but does not deal with the issue of face translation.

The recognition rates for varying illumination were much better than those for varying size and orientation and with further work it may be possible to improve the recognition rates for varying illumination. One possible extension would be to use color.

Overall, the eigenface method of face recognition suffers from the same constraints from which any image-to-image method of recognition suffers. To get reliable recognition the two images of the same object must be from the same camera angle, have the same illumination, be the same size in the image, etc. When one thinks about it this is not surprising. Let A and B be two images of the same object with B part of the training set and let $A \simeq B$. Subtracting another image such as the average face, Ψ, from both A and B and projecting both $A - \Psi$ and $B - \Psi$ must maintain this quasi-equality in order for Euclidean distance to be used. Murase and Hayar [14] show that for two images used to compute the eigenspace, the closer the projections are in the eigenspace, the higher the correlation between the two images.

Chapter 4

Color and Face Recognition

In chapter 3 we looked at the eigenface method in some detail. The work presented in this chapter is an exploration of some of the ways in which color might be used in conjunction with the eigenface method. The hope is that the addition of color will improve the recognition rate.

There seems to be some evidence that color is useful for object recognition. Wurm et al. [27] have shown that color improves object recognition by measuring the recognition response rate. Wurm et al. were not able to show that "color is a distinctive feature that can be traded off with shape or texture features". They concluded that color and shape act additively, not interactively.

4.1 Data Set

The data set used for the experiments presented in this chapter consisted of 135 images, nine images each of fifteen people. All images were taken with a Sony 3-CCD DXC-930 color camera, attached to a Parallax 24-bit frame grabber card installed on a SUN SparcLX workstation. Figure 4.1 shows the fifteen people used to build the data set. For each person, pictures were taken using all combinations of three illumination directions (from the left, from the right and from the left and the right) and three colors of illumination (white, blue and orange). The light sources for all images were halogen lamps. The blue and orange illuminations were generated by

Figure 4.1: People used to build Data Set

placing a colored filter in front of the camera lens. The spectral characteristics of the colored filters can found in Barnard's thesis [1]. Figure 4.2 shows all nine images for one person in the data set. The directional changes in the illumination are not readily apparent in the images because the lamps were shone from beside the person onto the opposing wall. The light reflected from the wall effectively produced a diffuse light source, limiting the amount of shading shown on each face. Shining the lamps directly at a person's face blinded them, producing many interesting facial contortions but few reproducible pictures.

When the eigenface method was used for recognition, each training set consisted of fifteen images — one from each person in the data set. All fifteen images in the training set were taken under the same lighting conditions. Figure 4.1 shows the training set used to test the recognition rates for grey-scale images.

Figure 4.2: Illuminations used for testing

4.2 Color Angle Invariants

Recent work done by Finlayson et al. [6] has shown that color angle invariants are sufficient to recognize classes of objects for which there are distinct color characteristics.

Consider I_1, a 3xN color image. If the response of our sensors is sufficiently sharp then

$$I_1 \approx DI_2 = \begin{bmatrix} d_1 & 0 & 0 \\ 0 & d_2 & 0 \\ 0 & 0 & d_3 \end{bmatrix} \begin{bmatrix} r_1 & r_2 & \cdots & r_n \\ g_1 & g_2 & \cdots & g_n \\ b_1 & b_2 & \cdots & b_n \end{bmatrix} \tag{4.1}$$

where I_2 is the set of sensor responses under a different illumination than I_1 and D is a linear diagonal transform [3]. We can think of the change in illumination as a change in the length of the red, green and blue vectors. If this is so, then the angles between the red, green and blue vectors must be invariant to changes in illumination. Normalizing each of these vectors reduces D to the identity matrix.

The range of color seen in faces is limited, skin tones range from white to brown. Eye color, while more distinctive, comprises a very small part of the face image. Consequently color angle invariants would probably not be sufficient for identification but could allow us to segment our face database into smaller subclasses. Such segmentation could limit the search and reduce the time needed for identification.

Experimental Results

To test whether color angle invariants could be used to segment a face database, I tried to recognize two people, one with light pink skin and blue eyes, the other with brown eyes and skin. An area around the right eye was segmented out from the database images, the hope being that eye color would provide additional color cues. Color angle invariants were able to distinguish between these two people 75% of the time for the images shown in Figure 4.3. I was hoping for better results than this since the two people whose images were chosen have such a large difference in coloring. Based on this results, it appears unlikely that color is a sufficient basis on which to segment a database of face images.

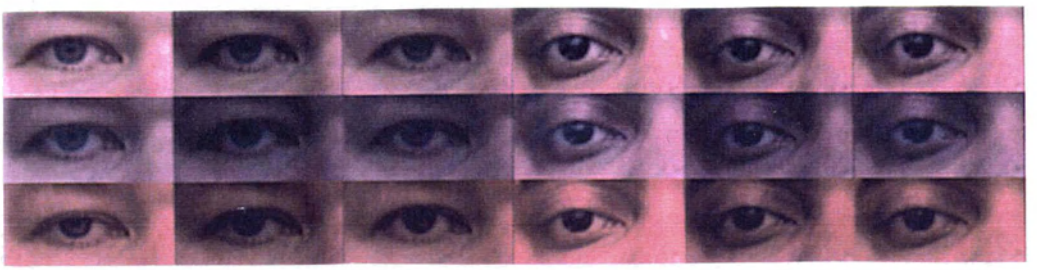

Figure 4.3: Eyes used to test Color Angle Invariance

4.3 Color Representations

Whenever images are compared, either directly or by some other methodology such as eigenfaces, it is important that all images appear to taken under the same illumination. Sirovich and Kirby [20] account for illumination intensity differences in grey-scale images by adjusting the pixel values relative to a set patch from the cheek. Normalizing a grey-scale image to length one will have the same effect. It is important to note that this normalization accounts for only a global intensity variation, it does not deal with changes in illumination color or direction. With color images both the illumination color and intensity must be accounted for. Three color respresentations that might account for illumination color changes are presented in this section.

4.3.1 Color Band Normalization

Normalizing two color images of the same object taken under different colors of illumination will not produce images which are approximately the same, the differences in the color bands will be preserved. As outlined in section 4.2.1, when the camera sensors are sufficiently sharp, a change in illumination color can be thought of as a change in the length of the red, green and blue vectors [3]. Therefore, normalizing each band should give us images for which the differences in color and intensity are

corrected. Changes in illumination direction will not be corrected for by this method and will be dealt with in chapters 5 and 6.

Using band normalized images for eigenface recognition should produce results at least as good as the results achieved with grey-scale images—better if color helps the recognition process.

4.3.2 Chromaticity

An image is formed by the interaction between the sensor, incident light and reflectance of the surface. If the surface is lambertian then this relationship for the i^{th} sensor is described by

$$\rho_i = \int_0^\infty S(\lambda)R(\lambda)\nu_i(\lambda)d\lambda,$$

where S is the spectral power density function, R the spectral reflectance function and ν the response function of the sensor. S, R and ν are functions of the wavelength λ. An eye or camera has three sensors, usually described as red, green and blue.

The International Commission on Illumination (Commission International de l'Éclairage, or CIE) describes color using chromaticity coordinates x, y and z [2], where x, y and z are defined as

$$\begin{aligned} x &= X/(X+Y+Z) \\ y &= Y/(X+Y+Z) \\ z &= Z/(X+Y+Z). \end{aligned}$$

Here

$$\begin{aligned} X &= k\int S(\lambda)R(\lambda)\bar{x}(\lambda)d\lambda \\ Y &= k\int S(\lambda)R(\lambda)\bar{y}(\lambda)d\lambda \\ Z &= k\int S(\lambda)R(\lambda)\bar{z}(\lambda)d\lambda \end{aligned}$$

with

$$k = \frac{100}{\int S(\lambda)\bar{y}(\lambda)d\lambda}.$$

Figure 4.4: Face as seen in chromaticity space

$\bar{x}$, $\bar{y}$ and $\bar{z}$ are the sensor response functions for the 2^o 1931 CIE standard observer.

Using chromaticity should give us a "flat" color description of the face. As can be seen in Figure 4.4 a few values describe much of the facial area. The original image used to build Figure 4.4 can be found in Figure 4.2. Chromaticity has gained us a more stable color description at the expense of the shading (shape) information.

For the actual testing a pseudo-chromaticity space was used with $x \approx r/(r+g+b)$ and $y \approx g/(r+g+b)$. All r, g and b values have been corrected for camera offset. In an ideal world, an image taken with the lens cap on will consist solely of zero value pixels. In reality, the pixel values range from zero to approximately twenty-two, with an average value of about thirteen. This noise is commonly referred to as the camera offset and has been corrected for by subtracting the average in each band from the color image and then setting all negative-valued pixels to zero.

4.3.3 Color Ratios

Finlayson and Funt [8] have shown that the histogramming of color ratios works well for identifying objects under uncontrolled illumination. Color ratios are the ratios of neighboring same band pixels. The ratio of two values x and y, can be found by

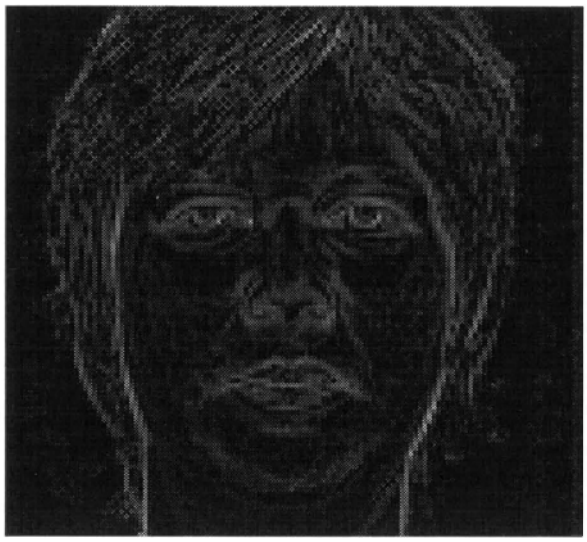

Figure 4.5: Face as seen in ratio space

solving either x/y, or $exp(log(x) - log(y))$. I convolved the log of the color image with a gradient mask to build the ratio images. As can be seen in Figure 4.5 the ratios are the edges of color change. The original image used to build Figure 4.5 can be found in Figure 4.2.

4.4 Experimental Results for Grey-scale, Chromaticity, Ratio and Band Normalized Images

Image recognition experiments were run for grey-scale, chromaticity, ratio and band normalized versions of the data set using the eigenface method and a picture-to-picture comparison. A picture-to-picture comparison identifies a face Γ by finding the image of the known person (from the database) which minimizes the Euclidean distance between it and Γ. The eigenface method also "recognizes" by looking for the minimum Euclidean distance. In all experiments, the training set or database of known people was formed by taking one image of each person from the data set described earlier. All images in the training set/database were taken under the same illumination. Table 4.1 gives the recognition rates for each of the methods. When

Table 4.1: Recognition Rates (in %)

	Picture to Picture	Eigenface
Grey-scale	95	96
Normalized *rgb*	95	95
Ratios	43	43
Chromaticity	68	53

using the eigenface method the number of eigenvectors used was adjusted so that approximately 80% of the available information was encoded in each case. For the grey-scale images seven eigenvectors were used; for chromaticities, nine; for color ratios, eleven; and for color band normalization, eight.

The color ratio images gave the worst results; people were recognized only 43% of the time. This is not all that surprising since the ratio images are essentially line images and even a small variation in head position could produce a mismatch of lines between two images of the same person. Chromaticities did a little better with a recognition rate of 53% using eigenfaces and 68% when compared picture-to-picture. These recognition rates for chromaticities are still very poor when compared to those obtained using grey-scale or band normalized color images.

Both the grey-scale and band normalized color images performed very well for both the picture-to-picture comparison and eigenfaces. Looking only at recognition rates does not give us a complete sense as to how well the grey-scale or band normalized color would scale up for use on a larger database. These numbers in themselves mean little, what is important is the relative distance between an image we wish to identify and the correct database image versus the distance between the image to be identified and all of the wrong database images. One means of representing this information is as ratios.

An example of how ratios are found is shown in Figure 4.6. It shows the relative placement of an image that we wish to identify, Γ, and the database images (for a small database of size four). Γ is recognized as TS_1 because it is closest to Γ, unfortunately this is wrong. TS_2 is the correct image. For image Γ the worst case

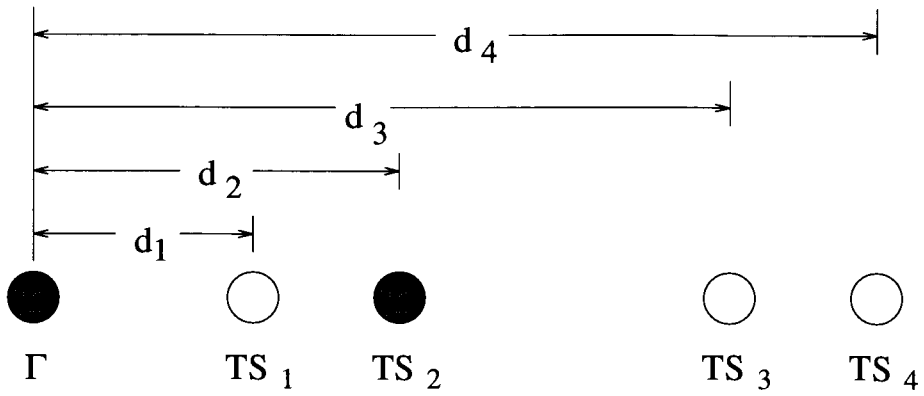

Figure 4.6: Relative position of image Γ to a database of known images; worst case ratio = d_1/d_2

ratio is d_1/d_2. The median of the wrong answer ratios is d_3/d_2. In order to assess whether one method was superior to another, the mean and standard deviation of the worst case ratios (closest wrong answer/correct answer) was calculated. The mean and standard deviation of the median of all wrong answer ratios was also calculated.

Figure 4.7 shows the mean distance ratios for the worst case and median case results. As you can see, the worst case distance ratios are similar for the picture-to-picture comparison of both the grey-scale and color band normalized images. The eigenface grey-scale median distance ratio is better than that for the grey-scale picture-to-picture comparison or either of the normalized color band methods. The worst case distribution for grey-scale images using the eigenface method has the best result, having both the largest mean and highest value at one standard deviation below the mean. The large standard deviation for the grey-scale eigenface ratios is caused by skew. This can be seen by comparing the mean (10.14) and the median (8.66) of the median sample. The mean and standard deviation for the distance ratios of the wrong answers for all four methods can be found in Table 4.2.

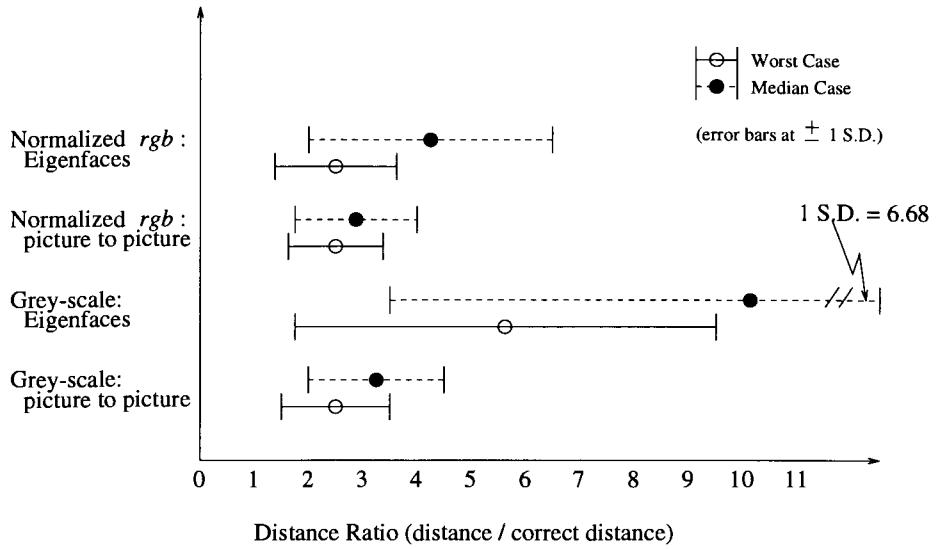

Figure 4.7: Mean Distance Ratios for Worst and Median Case Results

Table 4.2: Mean and Standard Deviation for Distance Ratios

	Method	Worst Case		Median	
		Mean	SD	Mean	SD
Grey-scale	Picture to Picture	2.55	1.03	3.29	1.35
	Eigenfaces	5.63	3.89	10.14	6.68
Normalized *rgb*	Picture to Picture	2.53	0.91	2.90	1.22
	Eigenfaces	2.50	1.10	4.28	2.20
Ratios	Picture to Picture	1.03	0.25	1.12	0.33
	Eigenfaces	0.96	0.53	1.44	1.95
Chromaticity	Picture to Picture	1.16	0.26	1.38	0.29
	Eigenfaces	1.37	0.93	2.05	1.45

4.5 Discussion

The hope was that the addition of color to the eigenface method would clearly improve recognition for images taken under similar directions of illumination. Unfortunately, this was not so. Grey-scale and color band normalized images gave comparable recognition rates but the distance ratio distribution for the grey-scale eigenface method was the best, it had the largest mean for both the worst case and median of the wrong answers. Under controlled conditions Pentland, Moghaddam and Starner [15] have achieved a recognition acuracy of 95% for a database of approximately 3,000 people. (Their images were taken with controlled illumination, scale and position—the eyes were accurately aligned in the picture taking process.) From this we know that the grey-scale eigenface method will scale up for use with a large database. Color band normalized recognition should be useable with a larger database but there is no evidence which suggests that it would do a better job than that already being done with grey-scale images.

Chapter 5

Colour and Changing Illumination

The color representations in chapter 4 were applied to images in which the direction of illumination was approximately constant (the changes in direction were small and barely discernable to the eye). Under these illumination conditions the use of color did not improve face recognition rates. In this chapter we investigate the use of color for images of faces illuminated from different directions.

The eigenface method essentially compares two images—the distance between two images will be reflected in the distance between their projections onto face space (as outlined in chapter 3). Therefore, when there is a gross change in the lighting conditions from one image to another, the distance between the two images and between their projections onto face space should increase. If this change is large enough it seems reasonable that, without correction for the differences in illumination direction, recognition rates would fall.

Under the appropriate conditions, color information may be used to correct for differences in illumination. Assume for a moment that we have two images of the same single color matte object, I_1 and I_2, where the only difference between the two images is that they were taken under different illuminations, with no image differences due to changes in scale, translation or rotation. Then, according to Petrov [18], if I_1 is an image of sufficient rank (where the image color rank is the rank of the matrix M shown in Equation 2.2) it can be transformed by a 3x3 matrix to look as if it were

taken under the same illumination as I_2.[1] Petrovs' work will be explained in the next section.

5.1 Properties of color images of surfaces under multiple illuminants

The following description of the relationship between sensor responses and surface normals is a paraphrasing of the basic definitions and relations used by Petrov and Kontsevich [18]. Let us first look at a simplified case. Suppose that we have a small lambertian patch illuminated by a single light source such as the one shown in Figure 4.1. The orientation of the surface can be described by its unit normal, $\vec{n}$, and the reflectance by the spectral reflectance function, $R(\lambda)$. The incident light is described by the unit direction vector, $\vec{q}$, and the spectral power density function, $S(\lambda)$. Now suppose that this patch is "seen" by a trichromatic sensor such as an eye or a colour camera, with response sensitivities $\nu_i(\lambda)$, where $i = 1, 2, 3$. The response of the i^{th} sensor to the view of the patch is given by

$$\rho_i = \vec{n}^T \vec{q} \int_0^\infty S(\lambda) R(\lambda) \nu_i(\lambda) d\lambda.$$

This equation can be rewritten as

$$\vec{\rho} = [\rho_1 \rho_2 \rho_3]^T = M \vec{n},$$

where each element of M, m_{ij}, is

$$m_{ij} = \int_0^\infty S(\lambda) R(\lambda) \nu_i(\lambda) q_j d\lambda.$$

These equations are valid for all surface orientations as long as the single color matte surface is externally lit. The need for external lighting can be explicitly expressed [26] as

$$\vec{n}^T \vec{q} \geq 0.$$

[1] This work was drawn to the author's attention by Graham Finlayson.

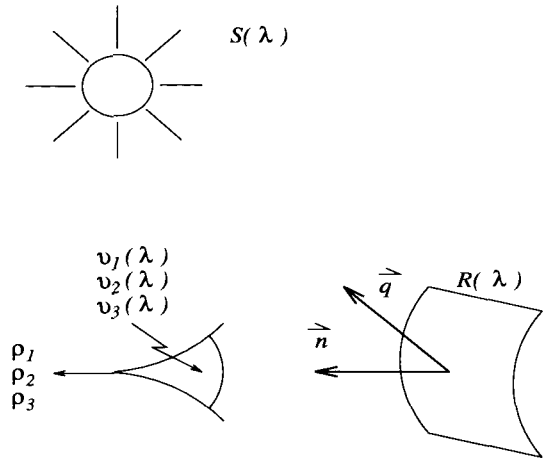

Figure 5.1: Illumination

Another way of saying this is that the surface normals and the sensor responses are linked by a linear transformation.

What if we have more than one lamp illuminating the scene? Then, assuming that the sensor responses are linear and not saturated, the response to the more complex illumination is equal to the sum of the responses to each of the individual light sources. In this case, M will be equal to the sum of the M's for each of the light sources.

There is one more fact to keep in mind; the distribution of light corresponding to any set of light sources can be emulated by three chromatic light sources [18]. Let $L_i, i = 1, 2, 3$, be such a set of chromatic light sources. Then any other light source is just a linear transform of L.

So far we know that, under certain conditions,the surface normals and the sensor responses are linked by a linear mapping, and that this mapping is dependent on the illumination. It seems reasonable that if a linear mapping exists between two illuminations, then there also exists a linear mapping between two images of the same unicolour object, assuming that the only thing to change is the illumination.

Is there any work which supports this conjecture? Petrov [16] uses a image taken under trichromatic lights to derive the shape of the object from the colour picture. The basis of this work is the equation

$$C(H_0\vec{n}(x,y) + h^d) = P(x,y).$$

Here x and y are the coordinates in the image P; $\vec{n}$ is the normal of the surface as a function of x and y; H_0 and h^d are the characteristics of the illumination with H_0 attributable to the three chromatic light sources and h^d to any diffuse lighting; C is the colour matrix which Petrov defines as the mapping which occurs when a white surface is substituted in the image for the corresponding coloured one. He uses this equation as the basis of the work in which he solves for $\vec{n}$.

Assuming we have two images of the same single colour object taken under different nondiffuse illuminations then $\vec{n}$, C and h^d (if it exists) are all constants. As long as one of the two images was taken under three chromatic light sources, say H_0, then we know that there must exist a linear mapping from H_0 to H_1, the other illumination. Therefore, there should be linear mapping from $P_0(x, y)$ (taken under H_0) and $P_1(x, y)$ (taken under H_1). In the work which follows this linear mapping will be referred to as M.

5.2 Preliminary results when adjusting for changing illumination

If the conjecture outlined in the previous section holds, we should be able to transform an image (S) of a single-color object taken under one illumination to its image taken under another illumination (D), by multiplying S by the 3x3 matrix M. This assumes that S was taken under three chromatic light sources. (An image which is taken under three chromatic light sources is defined as having a colour rank of three [18]. In such an image the response in each colour band is independent of the others.)

5.2.1 Pictures of an Egg

To test the method on a single-color object two pictures, both rank-3, were taken of an egg (Canada Grade A white). Figure 4.2 shows egg1, the egg illuminated by a faint red light, a faint white light and daylight and egg2, the egg illuminated by red light, yellow light and daylight. M was estimated in two ways.

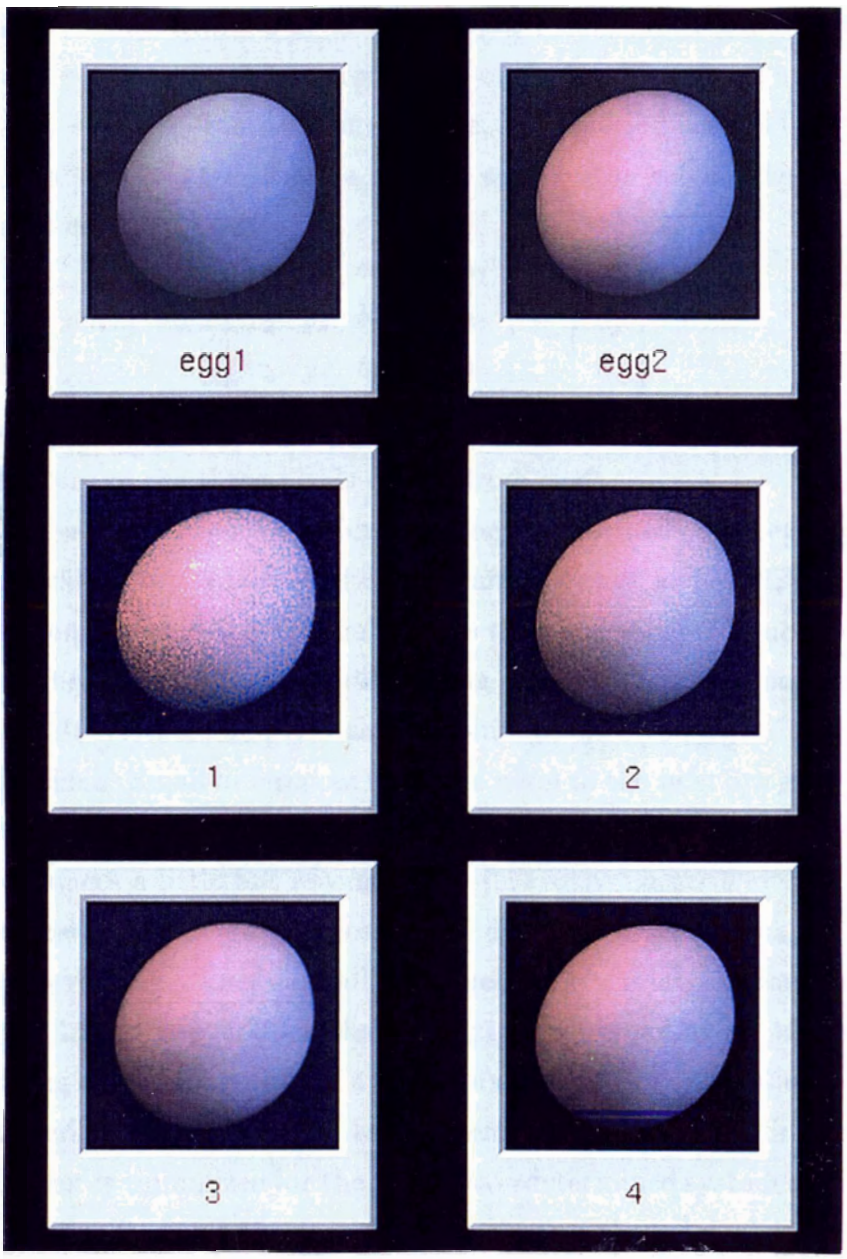

Figure 5.2: Egg; egg1 and egg2 are images of the same egg taken under different illuminations; 1 is the transformed egg1 with M calculated using 3 points; 2 has M calculated using a least squares fit; in 3 and 4 egg1 was smoothed before transformation

Initially M was determined by taking the pixel values from three arbitrary points well separated on the egg surface from both of the egg images. Three linear vector equations were solved to find M. For example, the equation needed to find the vector $[m_1, m_2, m_3]^T$ which will transform a pixel in egg1 to the value of the red element of a pixel in egg2 is

$$\begin{bmatrix} r_1 & g_1 & b_1 \\ r_2 & g_2 & b_2 \\ r_3 & g_3 & b_3 \end{bmatrix} \begin{bmatrix} m_1 \\ m_2 \\ m_3 \end{bmatrix} = \begin{bmatrix} r'_1 \\ r'_2 \\ r'_3 \end{bmatrix}$$

where $[r_i, g_i, b_i]$, $i = 1, 2, 3$ is the value of the i^{th} pixel chosen from egg1 and $[r'_1, r'_2, r'_3]^T$ are the red values of the three pixels chosen from egg2.

As you can see from Figure 4.2 picture 1, the transformed (SM) egg image appears to be very speckled. If you look at the pixel values of egg1 and egg2, you will see that the inherent bumpiness of the egg shell causes the pixel values to fluctuate slightly — the image of the surface is not smooth. Using three arbitrary points on this bumpy surface makes M unstable and it is this instability which causes the transformed egg to appear speckled. Small differences from one pixel to the next are greatly magnified when M is applied. Smoothing the original images improved the appearance of the transformed images a little but not much.

To overcome the problem of errors in M due to the noisy data, a least squares method was developed which uses all the pixels over a relatively large region of the image. Figure 4.2 pictures 2, 3 and 4 show egg1 transformed by an M which has been calculated using a least squares fit of a 65x65 pixel patch of the egg shell from egg1 and egg2. So instead of a direct solution to a system of three equations in three variables, the squared error is minimized for the highly overdetermined system of 65^2 equations. Figure 4.2 picture 2 shows the transformation using the original images, picture 3 shows the transformation after the original images have been smoothed using a 2x2 averaging mask and picture 4 shows the transformation after the original images have been smoothed using a median filter.

While looking at the images gives an intuitive sense for how well the transformation works, a more objective measure is also needed. The sum of squares of the difference between a pair of images is used as the distance measure. Table 4.1 gives the results

Table 5.1: Distance between images when M is calculated from 3 arbitrary points

	unsmoothed	smoothed with average	smoothed with median
$(egg1 - egg2)^2$	0.078	0.075	0.101
$(egg1 - egg1^T)^2$	0.176	0.168	0.104
$(egg2 - egg1^T)^2$	0.096	0.083	0.026

Table 5.2: Distance between images when M is calculated using a least squares fit

	unsmoothed	smoothed with average	smoothed with median
$(egg1 - egg2)^2$	0.078	0.075	0.101
$(egg1 - egg1^T)^2$	0.111	0.108	0.080
$(egg2 - egg1^T)^2$	0.053	0.046	0.027

when M is calculated using three arbitrary points and Table 4.2 the results when M is calculated using a least squares fit. The transformed image of egg1 will be shown as $egg1^T$ in the tables.

When using three arbitrary points, only the images smoothed by a median filter produced an M which transformed egg1 to a small distance away from egg2. This transformed image is still too speckled to give a visually pleasing image to the eye. When the least squares fit was used to calculate M the transformed egg1 was much closer to egg2 than to the original image of egg1.

5.2.2 Pictures of an Ellipsoid

Speculation that the roughness of the egg shell was the cause of the instability in M led to taking the picture of a white painted wood ellipsoid. Figure 4.3 shows from left to right; S(ource), the ellipsoid illuminated by yellow, green and daylight; D(estination), the ellipsoid illuminated by red, green and daylight; and S^T(ource transformed), S after the transform matrix M has been applied.

M was found using a least squares fit between the pixel values of a 89x159 rectangle clipped from the images of S and D. The results of transforming S can be seen on

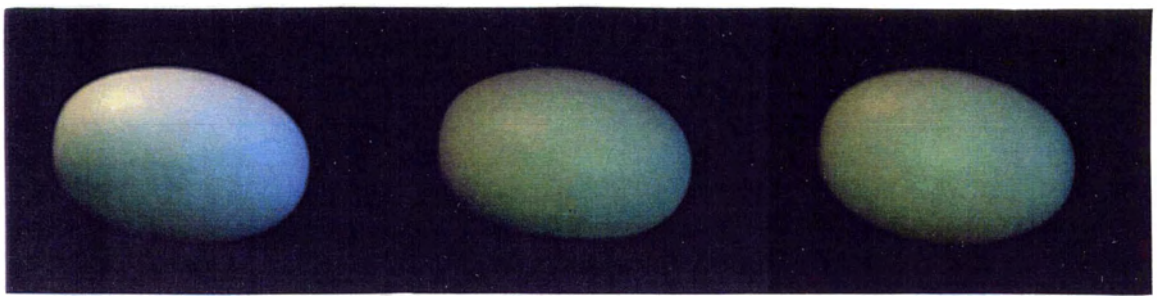

Figure 5.3: Ellipsoid: S(ource), D(Destination), S^T(ourse transformed)

Table 5.3: Distance between ellipsoid images

	offset intact	offset removed
$(S-D)^2$	0.131	0.183
$(S-S^T)^2$	0.132	0.167
$(D-S^T)^2$	0.023	0.020

the righthand side of Figure 4.3. Visually, the match between D and S^T is very good. Removing the image offset from S and D before calculating M did not appreciably change S^T to the eye.

Once again the more objective measure of the match was found by taking the sum of squares of the difference between a pair of normalized images. The distance results between the pairs of images is listed in Table 4.3 and D and S^T are clearly much closer to each other than S and D or S and S^T. This distance measure clearly shows that the matrix M does transform the image of the ellipsoid seen under one illumination (S) to the image that is seen under another illumination (D).

Chapter 6

Illumination correction and faces

A technique to correct for illumination differences between two images of the same single-color object was presented in chapter 5. For most people, the majority of the face image is a picture of essentially a single-color object—the skin. Hopefully, we can use the illumination correction technique to correct the database images so that we are doing a comparison of images that appear to be taken under approximately the same illumination.

6.1 Implementation

6.1.1 Procedure

- Collect a set of N images for known individuals (the database)
- For each new face to be identified
  - Calculate M_i, for $i = 1...N$, the transform matrix for the ith known individual
  - Find the distance between the new face and each of the transformed known individuals

6.1.2 Finding M

Doing a least squares fit between images of arbitrary people can be difficult. Ideally, we want the fit to be done between a patch of a single-color object illuminated by all three chromatic lights in one image and another equal-sized patch taken under a different illumination in the other. For most people, the skin comprises the greatest area of a single color. Even finding an equal sized patch of skin for an arbitrary person can be problematic—how do you guarantee that you have skin? For face images I hand-segmented out equal sized patches of skin. In general, I found that the forehead gave the best patch for men while the patch needed to come from the chin or cheek for women.

Is there a way to calculate a usable M using the entire image? Doing so would avoid any segmentation problems. Let S be a rank-3 image of an object and D be another image of the same object taken under different illumination conditions than S. The only difference in images S and D is the change induced by the change in the illumination. We want to calculate M such that the Euclidean distance between SM, the transformed S image, and D is minimized. Using the minimization of the squared error, $|SM - D|^2$,

$$M = Q^{-1}P,$$

where $Q = S^T S$ and $P = S^T D$. The derivation of this formula is given in Appendix A.

Both of the above approaches were used to find M in the work that follows.

6.1.3 Using Eigenfaces to Reduce Computational Costs

The Eigenface Method presented in chapter 3 can be used to reduce the computational cost of calculating SM. It is based on the assumption that the approximation

$$I \approx \sum_{i=1}^{n} c_i E_i, \tag{6.1}$$

where I is an image of a face, E_i the eigenfaces, n the number of eigenfaces and c_i the weight calculated by projecting I onto E_i, is sufficient for recognition.

Finding the distance between transformed images means solving $|SM - D|^2$, where S is an image from the database, M is the transform matrix and D is the new face which we wish to identify. If we use the eigenface approximation as given in equation 6.1 for S

$$\begin{aligned} SM &\approx \sum_{i=1}^{n} c_i E_i M \\ &= \sum_{i=1}^{n} c_i E_i Q^{-1} P \\ &= \sum_{i=1}^{n} c_i [(E_i Q^{-1}) \sum_{i=1}^{n} c_i (E_i D)] \end{aligned}$$

where $(E_i Q^{-1})$ can be precalculated directly from the eigenfaces and $(E_i D)$ needs to calculated only once for each new face. If the approximation in Equation 6.1 is sufficient to replace the database images, we can significantly reduce the computational costs and time involved in finding the distance between an incoming image and each of the known faces which make up our database.

6.1.4 Checking for translation problems

Anytime you do a picture-to-picture comparison, translation of the object from one image of the object to another will cause problems. Principal component analysis is also not invariant to translation. To avoid problems due to object translation between images, the correlation between the images of the same subject was maximized. The correlation between two images, $f(x, y)$ and $g(x, y)$ is defined [9] to be (discrete case)

$$f(x, y) \circ g(x, y) = \frac{1}{MN} \sum_{m=0}^{M-1} \sum_{n=0}^{N-1} f^*(m, n) g(x + m, y + n) \tag{6.2}$$

for $x = 0, 1, 2, ..., M - 1$ and $y = 0, 1, 2, ..., N - 1$. The correlation can be found in Fourier space by

$$f(x, y) \circ g(x, y) \Longleftrightarrow F^*(u, v) G(u, v). \tag{6.3}$$

where $*$ denotes the complex conjugate for both equations 6.2 and 6.3. The correlation check was implemented using the Fourier routines in *hips*[13] and applied prior to recognition.

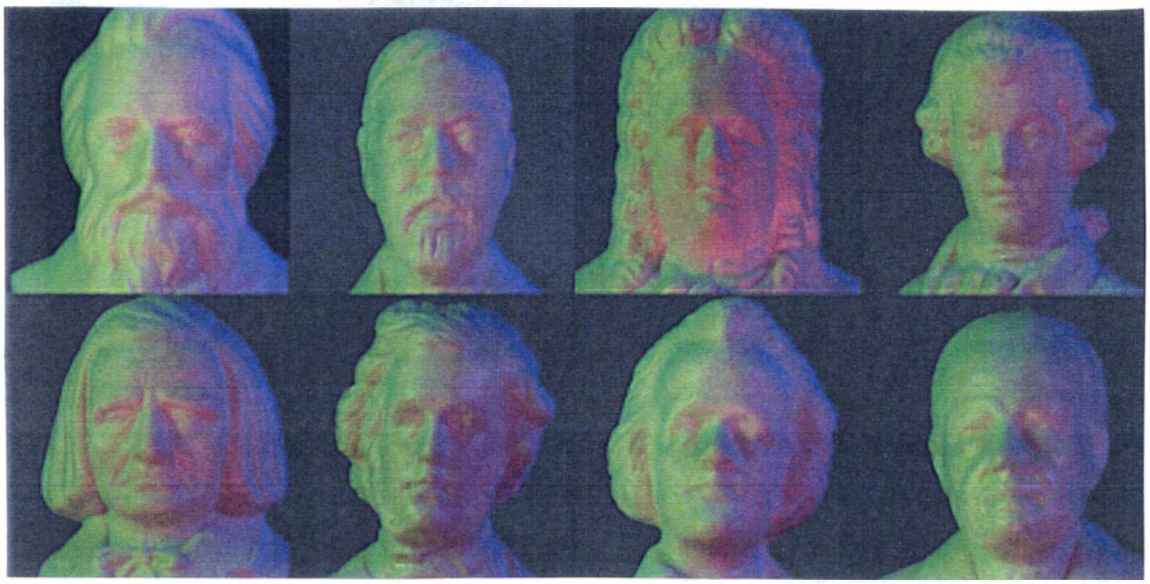

Figure 6.1: Rank 3 Images of Composer Busts

6.2 Data Set

Two sets of data were used to test the conjecture that an image of an object taken under three chromatic light sources can be linearly transformed to an image of the object under some other illumination. The first data set was a set of eight busts of famous composers, which provided true single colored objects, the second a set of fifteen students and staff in the department. 100 watt incandescent bulbs were used to illuminate the subjects. All pictures were taken with a Sony 3-CCD DXC-930 color camera and a Parallax 24-bit frame grabber card attached to a Sun Sparc LX.

The database for the busts consisted of eight rank-3 images of composers. The data set which was used to test for recognition consisted of twenty-four pictures, three of each composer. Figure 6.1 shows the rank-3 images of the busts of the composers. These rank-3 images were used as the database of composers, that is, the images against which all comparisons were made. Each of the busts in the database was illuminated by three 100 watt bulbs; green from the left, blue from the right and red from below and in front.

Figure 6.2: Busts: additional illumination

Three additional pictures were also taken of each bust; illuminated from the left with white light, illuminated from the right with white light, and illuminated by the fluorescent room light. The room lighting has a decidedly reddish tinge. An example of these three illuminations can be seen in Figure 6.2.

The database of faces consisted of fifteen rank-3 images. The data set used to test recognition consisted of forty-five pictures, three of each person. Figure 6.3 shows the rank-3 images of the faces which were used as the database or training set of faces. Once again, each of the faces appears to be illuminated by red, green and blue light. I was not able to take rank-3 pictures of everyone using the three colored lights. Illuminating a pale face with the three colors gave good results. People with a darker skin color did not produce a rank-3 image when illuminated by the bulbs so, instead I took three images under white lights (from the three different directions) and combined the red bands from each of these three images to produce the rank-3 training set image ($Image_{rank3} = [R_{image1} R_{image2} R_{image3}]$). Figure 6.4 shows an example of the three images which were combined to produce one rank-3 image. These three images were also used as a data set to test for recognition.

I found that it was very difficult to build a good database of face images of real people. People move constantly. It was the rare soul who could sit still for more than a couple of minutes. One of the reasons that taking each picture was so slow was that in order to save each picture the image had to be sent over the network. Taking pictures on video would have greatly simplified this task; however, the necessary tools were not available.

Figure 6.3: Rank 3 Images of Faces

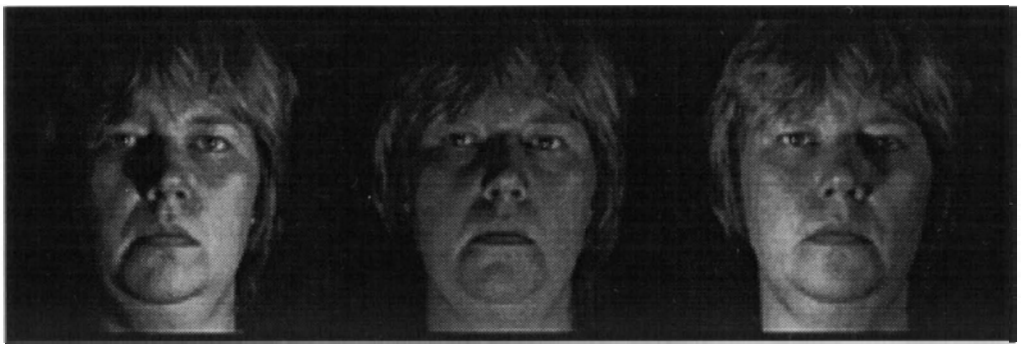

Figure 6.4: Faces: additional illuminations

6.3 Comparison of Methods

Each experiment consisted of comparing the data images to the database images. Comparisons were made for both grey-scale and color images. Grey-scale images were compared picture-to-picture and using the eigenface method of recognition. Color images were compared picture-to-picture between the transformed training set image and the data image and picture-to-picture between the transformed training set image approximated using eigenfaces and the data image. M was calculated for both the busts and the faces as described in Appendix A. For the face database, M was also calculated with a least squares fit between patches of skin.

While the recognition rates give us some idea of how well the technique is working, the most interesting thing is not the recognition rates but is rather the clustering characteristics of the distance between the image we wish to identify and the images in the database. Looking only at recognition rates for such a small sample size tells us very little about how the method would scale up for a larger database or data set. So, what characteristics do we want our comparison to have?

Ideally, we would like the image we want to identify to be much closer to the correct matching image than to any of the other images in the database. The distance values do not in themselves mean much. We are really interested in the ratio of the distance between the database images and the image to be identified, and the distance between the correct database image and the image to be identified. The correct database image will always have a ratio of 1. Figure 6.5 shows a typical

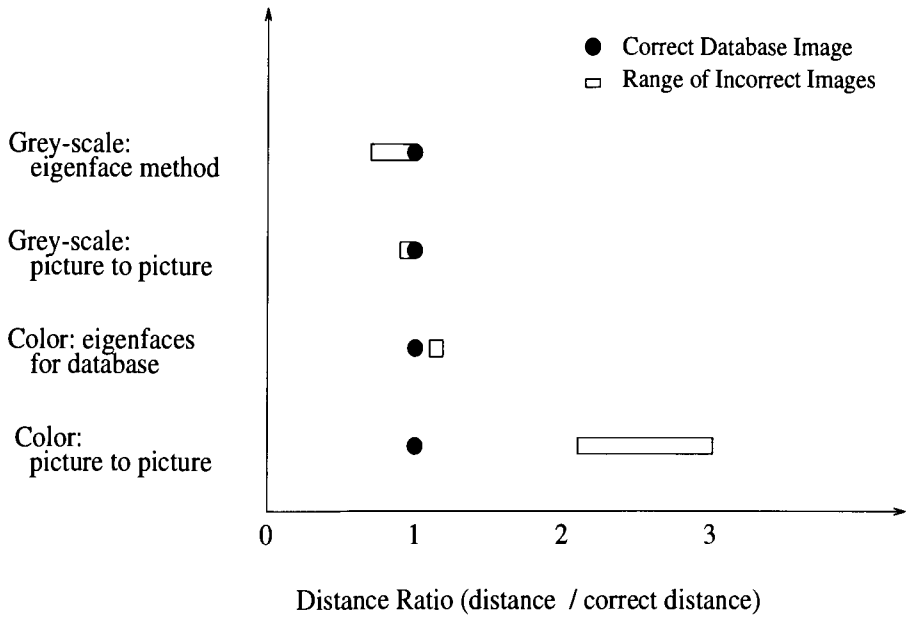

Figure 6.5: Distance Ratios for Brahms

case, the distance ratios between the image of Brahms illuminated from the left and the database images. As you can clearly see the best performance in this case was achieved when the transformed color database image was compared directly to the incoming image of Brahms. All incorrect images from the database were far away from the correct image. Using eigenfaces to approximate the transformed color image SM also gave a correct result, however, the resulting cluster of ratios are quite close to the correct image ratio of one. The grey-scale image of Brahms illuminated from the left was not recognized correctly by either the picture-to-picture comparison or the eigenface method.

While this example does not show the distribution of values for all of the images we used to test for recognition, it does give an idea of the way distance values clustered in the different experiments. To get an understanding of the expected range of incorrect distance ratios I looked at two things, the distribution of the worst case wrong values (ie. the wrong answer with the lowest ratio) and the distribution of the median case of the wrong answers. These distributions are characterized by their mean and standard deviation.

6.4 Experimental Results

Busts

Recognition rates for the composer busts are given in Table 6.1. The transformed

Table 6.1: Recognition Rates (in %) for Busts

	Method	Recognition Rate
Grey-scale	Picture to Picture	79
	Eigenfaces	67
Transformed Color	Picture to Picture	100
	Eigenfaces for SM	79

color images gave the best results—all twenty-four data images were recognized. Using eigenfaces to approximate the transformed color image gave as good a recognition rate (79%) as the grey-scale picture-to-picture comparison. Grey-scale eigenfaces did not do as well as any of the other methods (67%). When using eigenfaces to recognize either grey-scale or color images, the recognition rates increased as the number of eigenfaces used increased. In Table 6.1 the reported recognitions are for five of the possible eight eigenfaces. These five eigenfaces encoded approximately 84% of the available information. I chose to use five because the eigenface method as reported in chapter 3 used seven of a possible fifteen, and these seven also encoded just over 80% of the available information.

Figure 6.6 shows the mean and standard deviation for the worst case ratios and median case ratios from each experiment. The means and standard deviations for these distance ratios can be found in Table 6.2.

For the worst case, the picture-to-picture comparison using transformed color images was the only method which had a clear separation between the correct and incorrect images from the database. Using eigenfaces to approximate the transformed color images or doing a grey-scale picture-to-picture comparison produced values less than one that were within one standard deviation of the mean. For grey-scale images using the eigenface method, the worst case standard deviation predicts that images

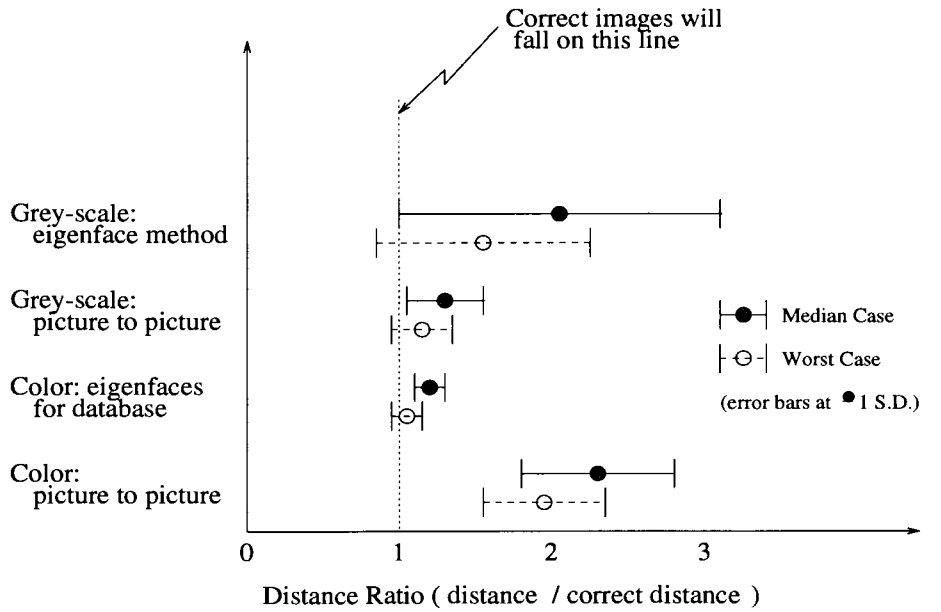

Figure 6.6: Mean Distance Ratios of Busts for Worst and Median Results

may be recognized incorrectly.

When we look at the median of the incorrect answers we see that in all cases there was an improvement over the worse case. Only eigenfaces for grey-scale images had values which fell to one within one standard deviation of the mean. Using the eigenface method to build the color database images or doing a picture-to-picture comparison with grey-scale images produced only a small gap between the correct answer and one standard deviation from the mean. Once again the color picture-to-picture comparison produced the best results, with a large gap between the correct

Table 6.2: Mean and Standard Deviation for Distance Ratios of Busts

	Method	Worst Case		Median	
		Mean	SD	Mean	SD
Grey-scale	Picture to Picture	1.16	0.19	1.28	0.25
	Eigenfaces	1.55	0.72	2.05	1.05
Transformed Color	Picture to Picture	1.95	0.39	2.29	0.49
	Eigenfaces for SM	1.06	0.09	1.18	0.10

answer at one and the values within one standard deviation of the mean.

Faces

The first question which needs to be addressed for faces is whether the illumination correction technique presented in chapter 5 will suffice for faces, because faces are not a single color. By visual assessment of the images shown in Figure 6.7, the transformation appears to work quite well when applied to faces. Regions where the color differs from the predominant color, such as the lip, eye and shirt colors are incorrect but overall the colors are quite close.

Figure 6.7: DataBase, Data and Transformed Images

For the picture-to-picture comparison of color images, M was calculated either from a patch of skin or from the entire image, as described in 6.1.2. Results will be presented for both of these calculations. Table 6.3 gives the recognition rates for each of the experimental methods and, once again, the picture-to-picture comparison of the transformed color images gave a 100% recognition rate. The grey-scale picture-to-picture comparison also gave very good results. Use of eigenfaces as a recognition method with grey-scale images or to build the transformed database color image degraded the rate of recognition to only 73%. In Table 6.3 seven of the possible fifteen eigenfaces were used to encode approximately 80% of the available information.

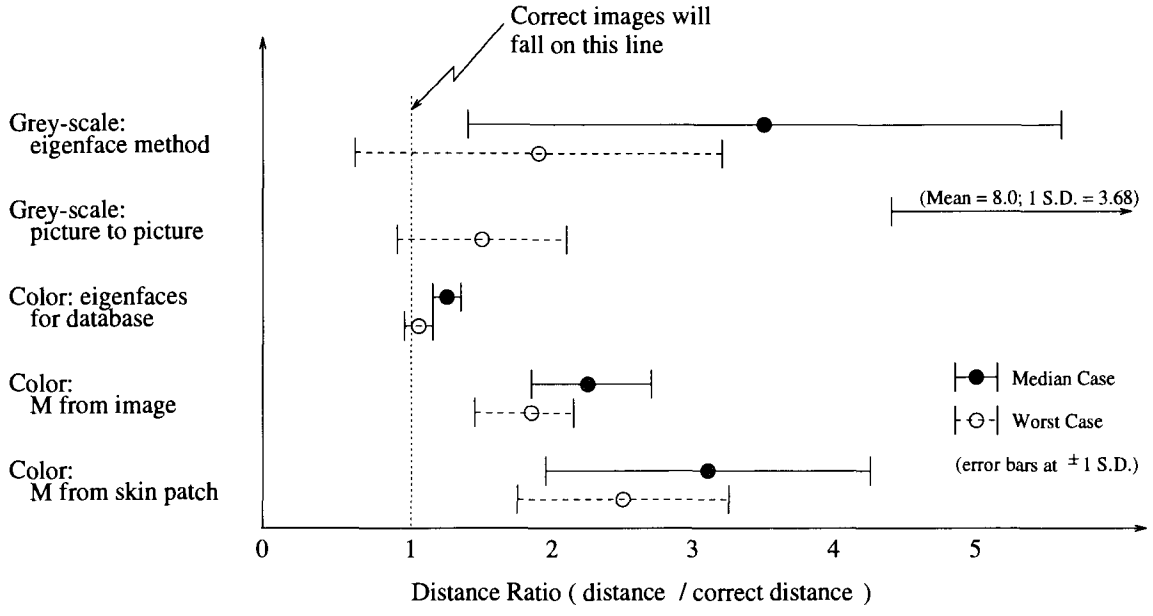

Figure 6.8: Mean Distance Ratios of Faces for Worst and Median Results

The distance ratios for faces were much more spread out than for the busts. Figure 6.8 shows the mean and standard deviation of the distance ratios for the worst case wrong values and the median case wrong values. The mean and standard deviation are also shown in Table 6.4.

For the worst case, the picture-to-picture comparison of the transformed color images gave a clear separation between the correct and incorrect images from the database. This was true for both methods of calculating M. Using eigenfaces to approximate the transformed color image produced a tight cluster of worst case distance

Table 6.3: Recognition Rates (in %) for Faces

	Method	Recognition Rates
Grey-scale	Picture to Picture	93
	Eigenfaces	73
Transformed Color Picture to Picture	M from skin patch	100
	M from image	100
	Eigenfaces for SM	73

ratios which overlapped the the critical ratio of one. This indicates that in general there was very little difference in the distance ratios for the correct database image and the incorrect database images. Grey-scale images recognized by the eigenface method and by a picture-to-picture comparison both had values within one standard deviation of the mean which were less than one.

The median of the incorrect answers for all of the color and grey-scale methods produced the expected ratio distributions — results greater than those of the worst case. The distributions for the grey-scale picture-to-picture comparisons are badly skewed. For the worst case the skew is caused by values much lower than the mean and for the median case the skew is caused by values greater than the mean. In general, all large standard deviations shown in Figure 6.8 and Figure 6.6 occur because of skew.

Table 6.4: Mean and Standard Deviation for Distance Ratios of Faces

Method		Worst Case		Median	
		Mean	SD	Mean	SD
Grey-scale	Picture to Picture	1.50	0.60	8.0	3.68
	Eigenfaces	1.90	1.30	3.52	2.14
Transformed Color Picture to Picture	M from skin	2.52	0.75	3.10	1.17
	M from image	1.85	0.32	2.26	0.43
	Eigenfaces for SM	1.06	0.08	1.22	0.10

Chapter 7

Discussion

This thesis examines one aspect of the problems encountered when machines attempt to recognize faces. Two general approaches are used when recognizing faces, using the entire image or extracting features from the image. This work has been limited to comparisons which are based on the entire image. The starting point was an examination of the eigenface method developed by Turk and Pentland [24]. Next, a number of methods were used to see if taking color information into account improves recognition. Finally, a method was developed which uses color to correct for varying directional illumination so that faces under such illuminations can be identified.

Picture-to-picture comparison and eigenface methods rely on the inherent similarity which exists between two images of the same person. This, of course, assumes that the pictures are taken such that the face size, orientation and camera angle are the same. Relying on the similarities between two pictures has its drawbacks. A different hair style, a beard versus a moustache, in fact any large change from one image to another can cause methods which rely on whole image comparisons to fail.

The eigenface method worked very well when applied to grey-scale images with a constrained head size and orientation. Similar results were achieved to those reported by Turk and Pentland [24] using their original data. Changes in head size or head orientation caused the method to fail. Changes in illumination did not degrade the results. However, the changes in illumination were small - there was no heavy shadowing of the face. So, the eigenface method worked well with constrained head size,

orientation and with limited illumination changes. One of the most interesting results is that recognition rates increase as the image head size decreases. As head size decreases small differences between face images become less apparent. Changes in facial expression have less of an impact on the similarity measure. Background information, if it remains constant, also aids recognition. It appears that the cluttered background of the original data contributed to the exceptional recognition rates reported by Turk and Pentland since masking or cropping their images reduced the recognition rates. In the real world you cannot depend on background information staying the same, even if the background is constrained to one color (ie. black or white) the clothes one wears change from day to day.

The use of color as an additional characteristic did not improve recognition rates over those obtained with the eigenface method for grey-scale images. It was also shown that faces are not colorful enough to be identified using color or even to allow the segmentation of a face database by color. None of the three color representations presented improved recognitions rates over grey-scale images and only color band normalization performed as well as grey-scale images. Using color ratios and chromaticities to represent the color images produced very poor recognition rates because of the lost shape information. The very good recognition rates obtained for both grey-scale and color band normalized images are due to the careful way in which the pictures were taken (to avoid changes in scale, rotation and translation).

The changes in illumination in images from Turk and Pentland's data set were small enough that the entire face could be seen. A new data set was created to address illuminations for which part of the face is heavily shaded. The information is still there, the pixel values are just so low that the face shape is not visible to the eye. Applying a 3x3 transform matrix to a rank-3 image of a single-color object allows us to make a new image which mimics the effect that different illumination colors or direction would have produced. If our database of faces is composed of rank-3 images, we can transform our database images so that they appear to have been taken under the same illumination as the image that we wish to identify. Illumination-corrected database images, when compared picture-to-picture to an incoming image, always recognized the person correctly. This result in itself does not mean much because the

size of the database was quite small. More importantly, when the ratio of the distance to the wrong person to the distance to the right person is evaluated for images taken under a strongly directional illumination, only the illumination corrected database images give a clear separation between the right and wrong answers. Without illumination correction grey-scale images compared picture-to-picture or via the eigenface method produce ratios whose range extends well into the region in which mistaken identities occur. These results indicate that color can be used to correct for severe illumination differences between two images of the same person.

Recognition using whole images only works when all images of the same person are taken under the same conditions. Changing any one of the conditions will cause the method to fail. For example, two of the data sets which are used in this work are comprised of images of the same people taken approximately a month apart. While both sets of pictures were taken carefully, so that the size and placement of the head in the images would be the same, some differences still exist. In particular, both a haircut and forward rotation of the face caused recognition to fail when comparisons were made between the two data sets. This failure leads me to believe that practical application of any whole-image method would be limited. Segmentation could separate the hair from the face but the forward rotation of the face can not be dealt with easily. Either a 3-D description of the face or a set of features, perhaps eigenfeatures [15], which are invariant to the forward rotation must be extracted from the image.

Actively moving a foveating camera so that images are only taken at points of interest may hold the key to reliable machine recognition. The human eye moves from point to point when recognizing an object. The amount of information received at each point is limited by the physical distribution of the receptors within the eye — at the centre the cones are packed very tightly giving a very high resolution while at the periphery the small number of cones provides a much coarser resolution. This act of foveation has been recreated in a camera by Sandini and Dario [19] and can be duplicated by converting a normal image to log-polar space. Representing images in log-polar space has the added advantage that changes in scale and rotation are seen as a translation.

Fully duplicating the human recognition system is a very complex and difficult

problem. Solving domain specific applications such as mug-shot matching can be a first step toward a more general solution. An attractive feature of image processing methods is that they are fast and relatively well understood—they allow an analytic approach. Extending the solution to include other parameters such as changing illumination which may not occur in the limited domain of mug-shots, helps to extend this work. I have shown how color information can be exploited to overcome some of the problems inherent in recognizing a face seen under varying illumination colors and direction.

Appendix A

Deriving M

Let S and D be color images of the same unicolor object for which all differences are due to a change in illumination. Furthermore, let S be a rank three color image. We want to find a transform matrix M, such that the euclidean distance between S and D,

$$F(M) = |SM - D|^2$$

is minimized. The distance is obviously minimized if both pictures appear to taken under the same illumination. For color images

$$|SM - D|^2 = trace(M^T S^T SM - 2M^T S^T D + D^T D).$$

If $Q = S^T S$ and $P = S^T D$, then

$$|SM - D|^2 = trace(M^T QM - 2M^T P + D^T D)$$

and

$$|SM - D|^2 = \sum_{i=1}^{3} M_i^T QM_i - 2M_i^T P_i + (D^T D)_i,$$

where i is the ith column vector (the comparison is only being done between bands of the same color).

To find the M which minimizes $|SM - D|^2$, we take the derivative with respect to M. Taking the derivative with respect to M means that we take the derivative with respect to each component of M and set it equal to 0.

Let

$$M = \begin{bmatrix} m_1 & m_4 & m_7 \\ m_2 & m_5 & m_8 \\ m_3 & m_6 & m_9 \end{bmatrix} = [M_1 M_2 M_3],$$

$$P = \begin{bmatrix} p_1 & p_4 & p_7 \\ p_2 & p_5 & p_8 \\ p_3 & p_6 & p_9 \end{bmatrix} = [P_1 P_2 P_3],$$

and

$$Q = \begin{bmatrix} q_1 & q_4 & q_7 \\ q_2 & q_5 & q_8 \\ q_3 & q_6 & q_9 \end{bmatrix} = [Q_1 Q_2 Q_3] = \begin{bmatrix} Q_1' \\ Q_2' \\ Q_3' \end{bmatrix}.$$

Then, for $i = 1$,

$$\begin{aligned} F(M_1) = M_1^T Q M_1 - 2M_1^T P_1 + (D^T D)_1 = \; & m_1 * (m_1 q_1 + m_2 q_2 + m_3 q_3) + \\ & m_2 * (m_1 q_4 + m_2 q_5 + m_3 q_6) + \\ & m_3 * (m_1 q_7 + m_2 q_8 + m_3 q_9) - \\ & 2 * (m_1 p_1 + m_2 p_2 + m_3 p_3) + \\ & (D^T D)_1 \end{aligned}$$

and

$$\begin{aligned} \frac{\delta}{\delta m_1} F(M_1) &= 2m_1 q_1 + m_2 q_2 + m_3 q_3 + m_2 q_4 + m_3 q_7 - 2p_1 \\ &= M_1^T \cdot Q_1 + Q_1' \cdot M - 2p_1 = 0 \end{aligned}$$

$$\begin{aligned} \frac{\delta}{\delta m_2} F(M_1) &= m_1 q_2 + m_1 q_4 + 2m_2 q_5 + m_3 q_6 + m_3 q_8 - 2p_2 \\ &= M_1^T \cdot Q_2 + Q_2' \cdot M - 2p_2 = 0 \end{aligned}$$

$$\begin{aligned} \frac{\delta}{\delta m_3} F(M_1) &= m_1 q_3 + m_2 q_6 + m_1 q_7 + m_2 q_8 + 2m_3 q_9 - 2p_3 \\ &= M_1^T \cdot Q_3 + Q_3' \cdot M - 2p_3 = 0. \end{aligned}$$

This can be rewritten as

$$\frac{\delta}{\delta M_1} F(M_1) = M_1^T Q + Q M_1 - 2 * P_1 = \begin{bmatrix} 0 \\ 0 \\ 0 \end{bmatrix}.$$

For $M = [M_1 M_2 M_3]$,

$$\begin{aligned} \frac{\delta}{\delta M} F(M) &= M^T Q + Q M - 2 * P \\ &= Q^T M + Q M - 2 * P \\ &= (Q^T + Q) M - 2 * P \\ &= 2 * Q M - 2 * P \\ &= O \end{aligned}$$

Therefore,

$$\begin{aligned} 2 * Q M &= 2 * P \\ M &= Q^{-1} P \end{aligned}$$

Bibliography

[1] Jacobus Barnard. Computational color constancy: Taking theory into practise. Master's thesis, Simon Fraser University, Burnaby, August 1995.

[2] Fred W. Billmeyer. *Principles of Color Technology.* New York: Wiley, 1981.

[3] Subho S. Chatterjee. Color invariant object and texture recognition. Master's thesis, Simon Fraser University, Burnaby, August 1995.

[4] R. Chellappa, C.L. Wilson, S. Sorohey, and C.S. Barnes. Human and machine recognition of faces: A survey. Technical Report CS-TR-3339, Computer Vision Laboratory, Center for Automation Research, University of Maryland, August 1994.

[5] Yong-Qing Cheng, Ke Liu, Jing-Yu Yang, and Hua-Feng Wang. A robust algebraic method for human face recognition. In *11th International Conference on Pattern Recognition*, pages 221–24, 1992.

[6] G.D. Finlayson, S.S. Chatterjee, and B.V. Funt. Color angle invariants for object recognition. In *The 3rd IS&T and SID Color Imaging Conference*, Scottsdale, Arizona, Nov 1995. (to appear).

[7] G.D. Finlayson, M.S. Drew, and B.V. Funt. Color constancy: Generalized diagonal transforms suffice. *J. Opt. Soc. Am. A*, 11(11):3011–19, 1994.

[8] G.D. Finlayson and B.V. Funt. Color constant color indexing. *IEEE Transactions on Pattern Analysis and Macine Intelligence*, 17(5):522–28, May 1995.

[9] Rafael C. Gonzalez and Richard C. Woods. *Digital Image Processing.* Addison-Wesley, third edition, 1992.

[10] Peter W. Hallinan. A low-dimensional representation of human faces for arbitrary lightin conditions. In *Proceedings of the IEEE Computer Society Conference on Computer Vision and Pattern Recognition*, pages 995–99, 1994.

[11] Statistical Sciences Inc. *S-PLUS, A Graphical Data Analysis System and Object-Oriented Programming Language.*

[12] I. T. Jolliffe. *Principal Component Analysis.* Springer series in statistics. Springer-Verlag New York Inc., 1986.

[13] Michael S. Landy and SharpImage Software. *HIPS Image Processing Software.*

[14] Hiroshi Murase and Shree K. Nayar. Visual learning and recognition of 3-d objects from appearnace. *International Journal of Computer Vision*, 14:5–24, 1995.

[15] Alex Pentland, Baback Moghaddam, and Thad Starner. View-based and modular eigenspaces for face recognition. In *Proceedings of the IEEE Computer Society Conference on Computer Vision and Pattern Recognition*, pages 84–91, 1994.

[16] A.P. Petrov. On obtaining shape from color shading. *COLOR research and application*, 18(6):375–79, December 1993.

[17] A.P. Petrov. Surface color and color constancy. *COLOR research and application*, 18(4):236–40, August 1993.

[18] A.P. Petrov and L.L. Kontsevich. Properties of color images of surfaces under multiple illuminants. *J. opt. soc. Am. A*, 11(10):2745–49, October 1994.

[19] G. Sandini and P. Dario. Active vision based on space-variant sensing. In *Proc. 5th Int. Symp. on Robotics Research*, pages 75–83, Toyko, 1990.

[20] L. Sirovich and M. Kirby. Low-dimensional procedure for the characterization of human faces. *J. opt. soc. Am. A*, 4(3):519–24, March 1987.

[21] Gilbert Strang. *Linear Algebra and Its Applications.* Harcourt Brace Jovanovich, third edition, 1988.

[22] Michael J. Swain and Dana H. Ballard. Color indexing. *International Journal of Computer Vision*, 7(1):11–32, 1991.

[23] M. Tistarelli. Recognition by using an active/space-variant sensor. In *Proceedings of the IEEE Computer Society Conference on Computer Vision and Pattern Recognition*, pages 833–37, 1994.

[24] M. Turk and A. Pentland. Eigenfaces for recognition. *Journal of Cognitive Neuroscience*, March 1991.

[25] M. Turk and A. Pentland. Face recognition using eigenfaces. In *Proceedings of the IEEE Computer Society Conference on Computer Vision and Pattern Recognition*, pages 586–91, 1991.

[26] R. J. Woodham. Gradient and curvature from the photometric stereo method, including local confidence estimation. *JOSA-A*, 11:3050–68, 1994.

[27] Lee H. Wurm, Gordon E. Legge, Lisa M. Isenberg, and Andrew Luebker. Color improves object recognition in normal and low vision. *Journal of Experimental Psychology*, 19(4):899–911, 1993.