

Department of Electrical
and
Computer Systems Engineering

Technical Report
MECSE-7-2004

A Study of The Illumination Cones Method for Face
Recognition Under Variable Illumination

T.J. Chin and D. Suter

MONASH
UNIVERSITY

A Study of The Illumination Cones Method for Face Recognition Under Variable Illumination

Tat-Jun Chin*and David Suter

25th August 2004

Abstract

Earlier face recognition algorithms have performed sufficiently well under tight environment constraints. However, simplistic and crude algorithms such as the geometric feature-based matching method and the template matching method fail catastrophically under slight environment changes, such as illumination variations. More sophisticated approaches like the *Eigenface* method or other *Appearance-Based Methods*, which had worked extremely well for face images with considerable changes in facial expression and occlusions, failed to tackle the issue of lighting changes. More recently, developments of methods that involve creating generative models of face images and extracting three-dimensional shape of faces have shown promising results. This report shall attempt to provide a detailed account of the *Illumination Cones* method that has shown a vast accuracy improvement compared to earlier methods in terms of robustness against illumination variation.

Keywords: Face recognition, lighting variation, *Illumination Cones* method.

1 Introduction

A well known face and often quoted statement in face recognition literature is 'the variation between the images of the same face due to illumination and viewing direction are almost always larger than images variations due to change in face identity [1]'. This condition severely compounds the difficulty of creating automatic face recognition algorithms that are robust across images corrupted by arbitrary lighting changes. As an illustration, Figure 1 shows a face being subjected to an increasing degree of illumination variations. The human visual system is still capable of recognizing the same face from all images, but most face recognition algorithms would already be impractical for usage in the second image.

*Tat-Jun Chin was a holder of the Australia-Asia Awards conferred by the Department of Education, Science and Training (DEST) of the Government of Australia since Feb 2004.



Figure 1: Illumination changes across the same face.

Contemporary research into face recognition algorithms that are robust against illumination variations usually adopted one of the following methods: either extract and measure some property in the face images which are invariant or at least insensitive to illumination changes; or model the three-dimensional face object in order to predict the variability of the images under changing illumination conditions [9].

Early face recognition algorithms, such as geometric feature-based matching methods [13, 20, 21] and template matching methods [13, 25] were just concerned with producing workable face recognition systems, and efforts were not concentrated towards incorporating robustness against illumination variations. Moreover, accuracy of such methods hinges greatly on the effectiveness of the feature extraction and measurement stage, and existing pre-processing techniques are still not reliable enough [14]. An example of such methods is the usage of edge maps as representations of face images. By looking at Figure 2, which is a series of images of the same face illumination under almost similar lighting configurations and their corresponding edge maps, we can deduce that a crude face recognition algorithm can be constructed by using the edge maps as templates for identity matching.

However, if the same face is subjected to a larger degree of illumination variations, illustrated by Figure 3, we can see that a straightforward template matching of the edge maps is not even remotely possible to be used as a face recognition algorithm. This suggests that simple image cues such as edges do not contain sufficient information [8], and are too susceptible to illumination variability, to be used to perform recognition under practical situations.

The *Eigenface* approach [24], being one of the first genuinely workable face recognition algorithm, was capable of accounting for facial expression variations and additions of minor occlusions, but fell short of being unaffected by lighting changes. The cornerstone of the *Eigenface* method was the utilization of the *Principal Component Analysis (PCA)* [22] to perform dimensionality reduction of face images from their huge image space to a much smaller subspace that maximizes the variations among face images of different identities. Later algorithms were also developed in similar spirit to the *Eigenface* approach in that they consider face images holistically and dimensionality reduction schemes were employed to obtain low-dimensional representations of face images. These are categorized as *Appearance-Based Methods (ABMs)*.

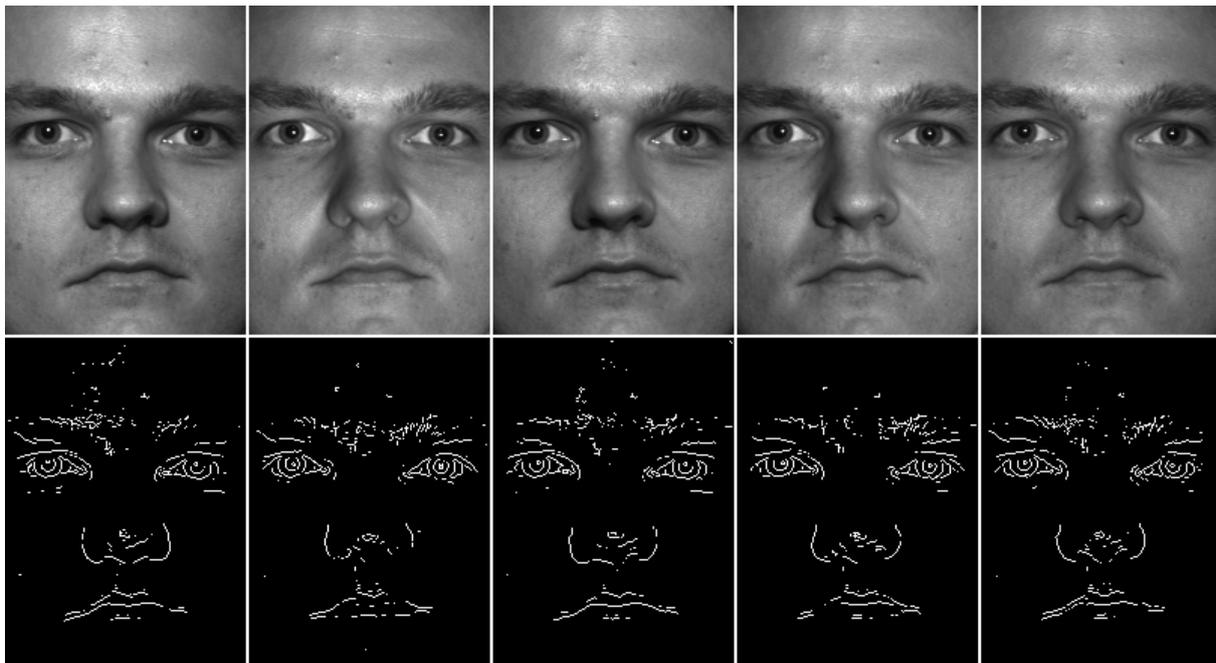


Figure 2: Edge Maps: Minor Illumination Changes.

The *ABMs* can be extended to take into account face images under illumination variations, provided that such images have been recorded under similar illumination conditions [11]. This prerequisite is the major drawback of the *ABMs* [7], as the task of creating a training set that incorporates face images seen under every possible illumination configuration is hardly achievable.

Yet, the modest success of some of the *ABMs* triggered a plethora of research work that attempt to characterize the subspace that encompasses images of a face subjected under all possible illumination configurations. Research results such as those published in [3, 4, 8, 15, 16] concluded empirically and theoretically that the images concentrate in a low-dimensional subspace, provided certain assumptions, such as face surfaces being purely *Lambertian* and face shapes being convex, are enforced. These developments paved the way for research into the underlying generative structure of a face image that can be extrapolated to create novel images under every possible illumination condition, without the face object having being viewed under such conditions.

The *Illumination Cones* method, first surfaced in [9], is strictly an appearance-based method for recognizing faces under extreme variability in illumination. It differs substantially from other *ABMs* in that a small number of images of each face under slight illumination changes are used to generate a representation, termed the *Illumination Cone*, of all images of the face under every possible illumination configuration. Results using *Illumination Cones* for face recognition across considerable illumination variation published in [6, 7] have shown promising results.

This technical report is organized as follows: Section 2 will describe in detail the *Illumination Cones* method, including the basics from the viewpoint of physics-based vision, construction of the cones representation and other issues pertaining to the construction of

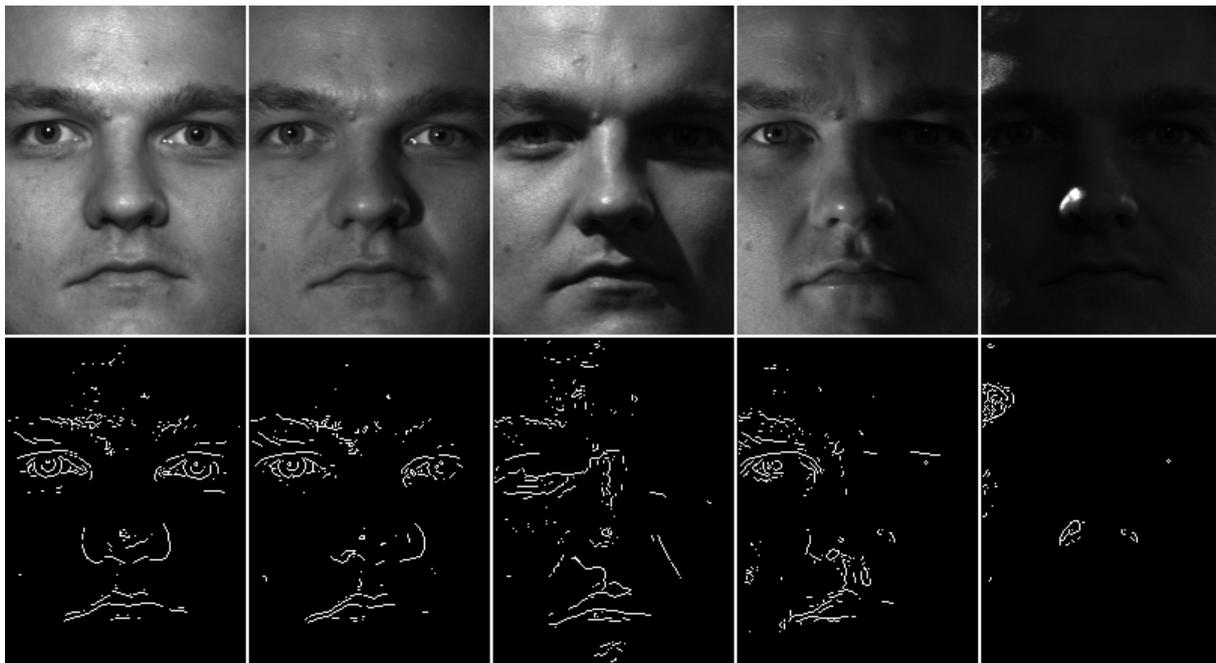


Figure 3: Edge Maps: Major Illumination Changes.

them. Section 3 will explain the setup of our experiment used to gauging the performance of this method and the results obtained. Finally, the conclusion is presented in Section 4.

2 The Illumination Cones Method

The *Illumination Cones* method is based on the theory of *Photometric Stereo*. This theory attempts to recover depth information by examining shading of two or more images of an object subjected to different illumination conditions. There are two main simplifying assumptions in order for *Photometric Stereo* to be valid for our applications: the surface of the face has *Lambertian*(diffuse) reflectance properties; and the shape of the face is convex.

A surface that has pure *Lambertian* reflectance properties appears equally bright (having the same shading pattern) from all viewing directions for a fixed lighting configuration. This behaviour is dictated by Lambert's cosine law, which says that the perceived brightness of a diffuse surface patch illuminated by a point light source varies with the incident angle (direction of light source) relative to the surface normal of the patch. It is summarized by the following equation:

$$L(\theta_e, \phi_e) = \frac{I}{\pi} \cos(\theta_s), \quad (1)$$

where $L(\theta_e, \phi_e)$ is the perceived brightness from the viewing angle determined by the polar coordinate values (θ_e, ϕ_e) , I is the intensity of the incident light source, and θ_s is the angular difference between the vector representing incident radiation and the normal of the surface patch. More detailed discussion of the surface reflectance properties can be found in [19].

Another factor that affects the observed brightness of a surface patch is the radiometric property of the *albedo* (whiteness) factor. This effect causes only a fraction of incident light to be radiated from a surface patch. Thus, the perceived brightness of a surface patch can be more precisely stated as

$$L(\theta_e, \phi_e) = \rho \frac{I}{\pi} \cos(\theta_s), \quad (2)$$

where ρ is the *albedo* factor and $0 \leq \rho \leq 1$.

By relating that the cosine of the angular difference between two vectors is equals to the inner product of the unit vector of the two vectors, the normalized intensity of a pixel in an image coordinate (x, y) corresponding to it's own actual surface patch can be written as

$$I(x, y) = a(x, y) \mathbf{n}(x, y) \cdot \mathbf{s} \equiv \mathbf{b}(x, y) \cdot \mathbf{s}, \quad (3)$$

where $a(x, y)$ is the albedo, $\mathbf{n}(x, y)$ is the surface normal, $\mathbf{b}(x, y) \equiv a(x, y) \mathbf{n}(x, y)$ and \mathbf{s} is the light source direction (the light is assumed to be at infinity).

2.1 Fundamentals of the Illumination Cone

With the above treatment, let \mathbf{x} denote an image with n pixels. Let \mathbf{B} denote a matrix of dimension $n \times 3$ (i.e. $\mathbf{B} \in \mathcal{R}^{n \times 3}$) where each row of \mathbf{B} is the product of the albedo with the unit normal at a particular pixel, and $\mathbf{s} \in \mathcal{R}^3$ be a column vector representing the product of the light source strength with the unit vector of the light source direction. Thus, we can write

$$\mathbf{x} = \mathbf{B}\mathbf{s}. \quad (4)$$

Since it is not possible for images to register negative values, we have to zero all the negative components of the answer. Hence,

$$\mathbf{x} = \max(\mathbf{B}\mathbf{s}, \mathbf{0}). \quad (5)$$

Take note that the negative components correspond to the surface points under shadows and are called *attached shadows*. Furthermore, since we assumed that the face shape is convex, we have avoided taking into account *cast shadows*, i.e. shadows that the object casts on itself.

If the face is illuminated by more than one light source, the resulting image \mathbf{x} would be the superposition of the images illuminated separately by the individual light sources, i.e.

$$\mathbf{x} = \sum_{i=1}^k \max(\mathbf{B}\mathbf{s}_i, \mathbf{0}), \quad (6)$$

where k is the total number of the individual point light sources.

If the matrix $\mathbf{B} \in \mathcal{R}^{n \times 3}$ is fixed and the light source matrix $\mathbf{s} \in \mathcal{R}^3$ is allowed to permute, the set $\mathbf{B}\mathbf{s}$ is the range of the matrix \mathbf{B} . Furthermore, this set is a subspace of the n -dimensional image space, with the notation

$$\mathcal{L} = \{ \mathbf{x} \mid \mathbf{x} = \mathbf{B}\mathbf{s}, \forall \mathbf{s} \in \mathcal{R}^3 \}. \quad (7)$$

\mathcal{L} is termed the *illumination subspace* by [9]. From linear algebra, we can see that the dimension of \mathcal{L} equals the rank of \mathbf{B} . If \mathbf{B} has full rank, \mathcal{L} is 3-dimensional, and it is generally assumed to be so because a face will surely have more than 2 linearly-independent surface normals.

It was proposed in [9] that the total number of unique images obtained by subjecting a fixed-posed object to every possible illumination configuration equals to $m(m-1)+2$, where m is the number of distinct surface normals of the object. Therefore, the set \mathcal{L} would contain *at most* $n(n-1)+2$ images, where n is the number of pixels in the face image. We can construct the set \mathcal{L} as a union of *extreme rays*, with the definition of an *extreme ray* as

$$\mathbf{x}_{ij} = \max(\mathbf{B}\mathbf{s}_{ij}, \mathbf{0}), \quad (8)$$

where

$$\mathbf{s}_{ij} = \mathbf{b}_i \times \mathbf{b}_j \quad (9)$$

and both \mathbf{b}_i and \mathbf{b}_j corresponding to rows of \mathbf{B} .

Consequently, we can construct the set \mathcal{C} , which comprises all images of a hypothetically *convex* and *Lambertian* face under all possible illumination configuration, by varying the direction and strength of an *arbitrary number* of point light sources at infinity, i.e.

$$\mathcal{C} = \{ \mathbf{x} \mid \mathbf{x} = \sum_{i=1}^k \max(\mathbf{B}\mathbf{s}_i, \mathbf{0}), \forall \mathbf{s}_i \in \mathcal{R}^3, \forall k \in \mathcal{Z}^+ \}. \quad (10)$$

\mathcal{C} is termed the *illumination cone* by [9], and was proven to be a *convex cone* in \mathcal{R}^n .

2.2 Constructing the Illumination Cone

The *illumination cone* \mathcal{C} is completely determined by the *illumination subspace* \mathcal{L} . If both matrices \mathbf{B} and \mathbf{s} are known, \mathcal{L} can be determined uniquely. If only \mathbf{s} is known, the problem is similar to classical photometric stereo, and least squares approximation can be used to obtain \mathbf{B} with the light source information and face images as inputs [5]. However, for the task of building representations of faces from training images, we will almost always not have *a priori* knowledge of the light-source directions or intensities. Therefore, we need a means of estimating both \mathbf{B} and \mathbf{s} from our training images.

To this end, Hayakawa proposed the use of the *Singular-Value Decomposition (SVD)* to estimate the \mathbf{B} and \mathbf{s} matrices [17]. An ensemble that contains at least three images of a fixed-posed face subjected to different but unknown point light sources is first obtained. Ideally, each point on the images must be illuminated in at least three of the images in the ensemble. These images are then converted into their vector representations, normalized and concatenated to form the *image data matrix*,

$$\mathbf{I} = [\mathbf{x}_1 \ \mathbf{x}_2 \ \dots \ \mathbf{x}_m], \quad (11)$$

where $\mathbf{x}_i \in \mathcal{R}^n$ represents normalized image i and m is the number of images in the ensemble. By invoking the SVD, the *image data matrix* can be decomposed into

$$\mathbf{I} = U\Sigma V^T, \quad (12)$$

where $U \in \mathcal{R}^{n \times n}$, $\Sigma \in \mathcal{R}^{n \times m}$ and $V \in \mathcal{R}^{m \times m}$ contain the *left singular vectors*, *singular values* and *right singular vectors*, respectively. The matrix Σ is a diagonal matrix with the non-zero values as the singular values ordered decreasingly. Theoretically, only the first three singular values are significant, therefore a minimum of only three images are needed in this method.

By taking into account only the first three singular values, we are actually finding the best rank-3 approximation of the *image data matrix*. The matrices \mathbf{B} and \mathbf{s} can be approximated as

$$\mathbf{B} = U(1:n, 1:3) \cdot [\Sigma(1:3, 1:3)]^{1/2}, \quad (13)$$

$$\mathbf{s} = [\Sigma(1:3, 1:3)]^{1/2} \cdot [V(1:m, 1:3)]^T. \quad (14)$$

However, ambiguities exist in the SVD results. With a decomposition such as $\mathbf{I} = \mathbf{B}\mathbf{s}$, for any arbitrary invertible 3×3 linear transformation $A \in GL(3)$, every

$$\mathbf{I} = \mathbf{B}\mathbf{s} = (\mathbf{B}\mathbf{A})(\mathbf{A}^{-1}\mathbf{s}) \quad (15)$$

is a valid decomposition of the *image data matrix* as well. Therefore, without the light source information, we can only recover \mathbf{B} up a 3×3 linear transformation $A \in GL(3)$ [17, 5]. The next section shall attempt to explain and deal with this ambiguity.

2.3 The Bas-Relief Ambiguity

A generalized *bas-relief* (*GBR*) transformation changes both the surface shape and albedo pattern. Let (x, y) denote a coordinate point in an image plane, $z = f(x, y)$ denote the distance from an object's surface to the image plane, and $a(x, y)$ denote the albedo pattern, GBR transformation of the object's surface and albedo pattern are defined respectively as

$$\bar{f}(x, y) = \lambda f(x, y) + \mu x + \nu y, \quad (16)$$

$$\bar{a} = \frac{a}{\lambda} \left(\frac{(\lambda f_x + \mu)^2 + (\lambda f_y + \nu)^2 + 1}{f_x^2 + f_y^2 + 1} \right)^{1/2}, \quad (17)$$

with $\lambda > 0$, and f_x and f_y denote partial differentiation of $f(x, y)$ against x and y respectively. When $0 < \lambda < 1$ and $\mu = \nu = 0$, the transformed surface is equivalent to relief sculptures or so-called "flattened forms" created by artisans since antiquity. These relief sculptures are indistinguishable from their original untransformed versions when viewed from a particular vantage point. Figure 4 provides an example.

In order to view the bas-relief transformation as a linear operator, let $\mathbf{p} = (x, y, f(x, y))$ and $\bar{\mathbf{p}} = (x, y, \bar{f}(x, y))$ denote the coordinate point (according to the image frame) of the surface of the original and transformed object respectively. Therefore,

$$\bar{\mathbf{p}}^T = \mathbf{G}\mathbf{p}^T, \quad (18)$$

$$\text{with } \mathbf{G} = \begin{bmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ \mu & \nu & \lambda \end{bmatrix}. \quad (19)$$

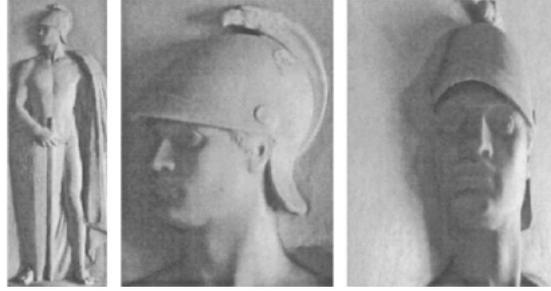


Figure 4: Frontal and side view of a relief sculpture. Notice that when viewed frontally, it appears that the sculpture has full 3-D depth. Image taken from [10].

The inward pointing surface normal for a surface point $\mathbf{p} = (x, y, f(x, y))$ can be written as

$$\mathbf{n}(x, y) = \begin{bmatrix} -f_x \\ -f_y \\ 1 \end{bmatrix}. \quad (20)$$

The following relation exists as well for surface normals of an object and the surface normals for it's GBR-transformed version:

$$\bar{\mathbf{n}} = \mathbf{G}^{-T} \mathbf{n}, \quad (21)$$

$$\text{with } \mathbf{G}^{-1} = \frac{1}{\lambda} \begin{bmatrix} \lambda & 0 & 0 \\ 0 & \lambda & 0 \\ -\mu & -\nu & 1 \end{bmatrix}, \quad (22)$$

$$\text{and } \mathbf{G}^{-T} = (\mathbf{G}^{-1})^T. \quad (23)$$

Several significant observations of this subject matter and their implications towards computer vision, especially in the field of 3-d reconstruction through photometric stereo, were outlined in [10]. Assuming that both an object and it's GBR transformation are viewed frontally, it was observed that

1. The set of cast and attached shadows produced by a surface and it's GBR transformation are identical, irrespective of the surface reflectance properties.
2. If the material of the object can be modeled as having *Lambertian* reflectance properties (one of the main assumptions of the *Illumination Cones method*), then the set of possible images including shadowing under any lighting condition for an object and a GBR-transformed version of the object are identical i.e. they have the *same illumination cones* (it should be noted that a GBR transformation alters *both* surface geometry and albedo). Hence, these objects cannot be discriminated by any recognition algorithm that uses images as inputs.
3. The generalized bas-relief transformation is the *only* transformation which has these first two properties.
4. For photometric stereo where the light source directions are unknown, the structure can only be determined up to a GBR transformation, and shadows do not provide further information.

See [10] for proofs of these statements.

2.4 The Integrability Constraint

Reiterating the result in Equation 15, the surface normal matrix recovered from the SVD technique differs from the actual surface normal matrix by a 3×3 linear transformation. Let \mathbf{B}^* be the actual surface normal matrix, \mathbf{B} be our estimation, and $\mathbf{P} \in GL(3)$ be the linear transformation, we can write

$$\mathbf{B}^* = \mathbf{B}\mathbf{P}. \quad (24)$$

For any function $z = f(x, y)$ to correspond to a consistent surface, it must satisfy an *integrability* condition denoted by the following equation [12]:

$$f_{xy} = f_{yx}, \quad (25)$$

where the subscripts denote partial derivatives. This in turn suggests that for every row of \mathbf{B}^* which is denoted by $\mathbf{b}^*(x, y) = [b_1^* \ b_2^* \ b_3^*]$, the following condition holds:

$$\left(\frac{b_1^*}{b_3^*} \right)_y = \left(\frac{b_2^*}{b_3^*} \right)_x. \quad (26)$$

It was claimed in [10] that the GBR-transform is the *only* linear transformation that preserves the integrability in a set of transformed surface normals that was found to have satisfied the integrability condition pre-transformation. That is to say that in an operation such as

$$\bar{\mathbf{b}}(x, y)^T = \mathbf{M}\mathbf{b}(x, y)^T, \quad (27)$$

where both $\mathbf{b}(x, y)$ and $\bar{\mathbf{b}}(x, y)$ correspond to surface normals, if both $\mathbf{b}(x, y)$ and $\bar{\mathbf{b}}(x, y)$ satisfy the integrability condition as in Equation 26, the *only* valid linear transformation denoted by \mathbf{M} must be in the form specified by Equations 22 and 23. See [10] for proof of this proposition.

As a corollary, we can see that the result $\bar{\mathbf{b}}$ from the GBR-transformation of the vector \mathbf{b} would be

$$\bar{\mathbf{b}}^T = \begin{bmatrix} b_1 - b_3 \frac{\mu}{\lambda} \\ b_2 - b_3 \frac{\nu}{\lambda} \\ b_3 \frac{1}{\lambda} \end{bmatrix}. \quad (28)$$

If Equation 26 holds for \mathbf{b} , then the following equation

$$\left(\frac{b_1 - b_3 \frac{\mu}{\lambda}}{b_3 \frac{1}{\lambda}} \right)_y = \left(\frac{b_2 - b_3 \frac{\nu}{\lambda}}{b_3 \frac{1}{\lambda}} \right)_x \quad (29)$$

holds for $\bar{\mathbf{b}}$ as well.

To present the linear transformation in the form of Equation 31, Equation 30 can be written as

$$\bar{\mathbf{b}}(x, y) = \mathbf{b}(x, y) \mathbf{M}^T. \quad (30)$$

To satisfy the integrability condition for both sides of the equation, \mathbf{M} was constrained to be in the form of \mathbf{G}^{-T} . Therefore, in order to preserve integrability, we have to impose the following constraint:

$$\mathbf{B}^* = \mathbf{B}\mathbf{P}, \mathbf{P} \in (\mathbf{GBR})^{-1}. \quad (31)$$

Take note that since \mathbf{B} and $(\mathbf{B}\mathbf{A})$ in Equation 15 correspond to face surfaces, we have imposed a constraint on the possible form of \mathbf{A} , though the ambiguity has not been resolved completely. However, as far as producing images under the *Lambertian* assumption is concerned, since \mathbf{B} and $(\mathbf{B}\mathbf{A})$ differ from a GBR-transform only as a consequence of the constraint, they have identical *Illumination Cones*.

2.5 Enforcing Integrability

Using SVD to decompose the image data matrix \mathbf{I} guarantees that we have the best 3-dimensional basis based on the ranking of the singular values. However, it is not guaranteed that the estimated surface normal matrix \mathbf{B} can be used to form a consistent surface. In order to apply the integrability constraint developed in the last section on the ambiguous 3×3 linear transformation of the SVD results, we have to ensure that every row of matrix \mathbf{B} satisfies the integrability constraint.

Since there were no means of enforcing integrability during the estimation of \mathbf{B} , it has to be enforced *afterwards*. A method similar in concept to the ones published in [6, 5, 23] was developed to achieve this. Let \mathbf{T} denote the following transformation:

$$\mathbf{B} \xrightarrow{\mathbf{T}} \mathbf{B}^* \quad (32)$$

where \mathbf{B} is our estimation and \mathbf{B}^* is the surface normal matrix that satisfies the integrability constraint. The operator \mathbf{T} is obviously a 3×3 linear transformation, hence

$$\mathbf{B}^* = \mathbf{B}\mathbf{T}. \quad (33)$$

Let $\mathbf{b} = [b_1 \ b_2 \ b_3]$ be one row of \mathbf{B} and \mathbf{T}_1 , \mathbf{T}_2 and \mathbf{T}_3 be the 3 columns of \mathbf{T} , then the corresponding row of \mathbf{B}^* is

$$\mathbf{b}^* = [b_1^* \ b_2^* \ b_3^*] = [\mathbf{b}\mathbf{T}_1 \ \mathbf{b}\mathbf{T}_2 \ \mathbf{b}\mathbf{T}_3]. \quad (34)$$

Since \mathbf{B}^* , as a surface normal matrix, satisfies the integrability condition, the following condition must hold true:

$$\left(\frac{\mathbf{b}\mathbf{T}_1}{\mathbf{b}\mathbf{T}_3} \right)_y = \left(\frac{\mathbf{b}\mathbf{T}_2}{\mathbf{b}\mathbf{T}_3} \right)_x. \quad (35)$$

Expanding this out, we get

$$(\mathbf{b}\mathbf{T}_3)(\mathbf{b}_y\mathbf{T}_1) - (\mathbf{b}\mathbf{T}_1)(\mathbf{b}_y\mathbf{T}_3) = (\mathbf{b}\mathbf{T}_3)(\mathbf{b}_x\mathbf{T}_2) - (\mathbf{b}\mathbf{T}_2)(\mathbf{b}_x\mathbf{T}_3), \quad (36)$$

which can be expressed concisely as

$$\mathbf{b}\mathbf{S}_1\mathbf{b}_y^T - \mathbf{b}\mathbf{S}_2\mathbf{b}_x^T = 0, \quad (37)$$

where $\mathbf{S}_1 = \mathbf{T}_3\mathbf{T}_1^T - \mathbf{T}_1\mathbf{T}_3^T$ and $\mathbf{S}_2 = \mathbf{T}_3\mathbf{T}_2^T - \mathbf{T}_2\mathbf{T}_3^T$. We can see that both \mathbf{S}_1 and \mathbf{S}_2 are skew-symmetric matrices of the form

$$\mathbf{S}_1 = \begin{bmatrix} 0 & s_{11} & s_{12} \\ -s_{11} & 0 & s_{13} \\ -s_{12} & -s_{13} & 0 \end{bmatrix} \text{ and } \mathbf{S}_2 = \begin{bmatrix} 0 & s_{21} & s_{22} \\ -s_{21} & 0 & s_{23} \\ -s_{22} & -s_{23} & 0 \end{bmatrix}, \quad (38)$$

with three degrees of freedom each. Equation 38 can be expanded for every row of \mathbf{B} , and with discrete approximations of the partial derivatives, we will arrive at an over-constrained homogeneous system of linear equations

$$\mathbf{A}\mathbf{s} = \mathbf{0}, \quad (39)$$

where \mathbf{s} contains all the elements of \mathbf{S}_1 and \mathbf{S}_2 . This can be easily solved using SVD.

Furthermore, the elements of \mathbf{S}_1 and \mathbf{S}_2 are cofactors of the matrix \mathbf{T} . Therefore, we can solve for matrix \mathbf{T} by virtue of the following relation:

$$k \begin{bmatrix} T_{11} & T_{12} & T_{13} \\ T_{21} & T_{22} & T_{23} \\ T_{31} & T_{32} & T_{33} \end{bmatrix} = \begin{bmatrix} \Delta_{11} & \Delta_{21} & \Delta_{31} \\ \Delta_{12} & \Delta_{22} & \Delta_{32} \\ \Delta_{13} & \Delta_{23} & \Delta_{33} \end{bmatrix}^{-1}, \quad (40)$$

where $\Delta_{11} = T_{22}T_{33} - T_{23}T_{32}$, etc. and k is a normalization constant. Notice that since \mathbf{S}_1 and \mathbf{S}_2 contain only six of the total number of cofactors of the matrix \mathbf{T} , i.e. Δ_{11} , Δ_{12} , Δ_{21} , Δ_{22} , Δ_{31} and Δ_{32} , the remaining three cofactors, i.e. Δ_{13} , Δ_{23} and Δ_{33} are still unknown. However, according to [23], the unknown cofactors correspond to the parameters λ , α and β of the generalized bas-relief transformation, and they can be selected arbitrarily in order to solve for \mathbf{T} .

2.6 Accounting for Shadows, Specularities and Saturations

From section 2.2, it was proposed that the image data matrix \mathbf{I} be constructed by the concatenation of at least three face images that were created by subjecting the face to different but unknown point light sources. Furthermore, each point in the images used must be illuminated in at least three of the images in the ensemble. The images used must contain as little shadowing as possible.

However, it is clear that this might not be achievable. Even if an object has surface normals covering the Gauss sphere (a convex object), there is only one light source direction- the viewing direction- such as the entire visible surface is illuminated [9]. Inevitably, shadows are always present in the face images, and they will cause biased estimation of the surface normal matrix.

Saturations in the images, although having different causes than shadows, have the same effect of biasing the estimation of the surface normals. Both shadows (0 greylevel value) and saturations (maximum greylevel value) provide lower and upper limits of image intensities. These imaging shortcomings provide a constraint on the range of the surface normal matrix, and should be accounted for in the process of estimating surface normals.

A face surface is not perfectly *Lambertian*, therefore specular points will always be present in the face images. Sources of specularities might include the eyeballs, eyebrows and oily skin surfaces. These specular points can be considered as outliers in the *Lambertian* model assumed for the face surfaces, and they will cause biased estimation of the surface normals as well.

Instead of invoking a straightforward SVD to arrive at an estimation of the surface normals, an iterative method similar to the one which appeared in [6] is implemented. Their method was influenced by [18]. First of all, pixels or elements in the image data matrix \mathbf{I} that correspond to shadows and saturations are first detected by thresholding. The rows of \mathbf{I} that contain at least one of these bad pixels are sifted out, forming a sub-matrix $\tilde{\mathbf{I}}$.

The SVD is performed on $\tilde{\mathbf{I}}$ and an initial estimate of the light source matrix $\tilde{\mathbf{s}}$ is obtained. Take note that although $\tilde{\mathbf{I}} \in \mathcal{R}^{\tilde{n} \times m}$, where $\tilde{n} < n$, estimates for $\tilde{\mathbf{s}}$ would still be in $\mathcal{R}^{3 \times m}$. Each row of the matrix $\tilde{\mathbf{B}}$ is then estimated using least squares with the original image data matrix \mathbf{I} as the response variables and the initial $\tilde{\mathbf{s}}$ as the explanatory variables. By using the new estimate of $\tilde{\mathbf{B}}$, a new $\tilde{\mathbf{s}}$ is estimated, again, using least squares. The previous two steps are iterated until the estimates converge.

During the iterative procedures of the method, the least squares fit is used to estimate the linear model

$$\begin{bmatrix} i_1 & i_2 & \dots & i_m \end{bmatrix} = \begin{bmatrix} b_1 & b_2 & b_3 \end{bmatrix} \begin{bmatrix} s_{11} & s_{21} & \dots & s_{m1} \\ s_{12} & s_{22} & \dots & s_{m2} \\ s_{13} & s_{23} & \dots & s_{m3} \end{bmatrix}, \quad (41)$$

with elements of b and s exchanging roles as the explanatory variables. Though pixels that do not conform to the *Lambertian* linear model were excluded during the initial estimation of $\tilde{\mathbf{s}}$, elements in the image data matrix that correspond to these pixels can still influence the results during the iterative stages. An overwhelming presence of outliers can cause invalid results especially when a linear fitting algorithm with low breakdown percentage, such as least squares, is used.

Outlier influences can be mitigated by using robust fitting algorithms with high breakdown points such as the least median squares, but it is still crucial to select good training images. However, acquiring an ensemble of face images with all pixels illuminated in at least three of the images is no easy feat, especially when considering parts like the nostrils and nose edges, and furthermore, areas like eyeballs and eyebrows are simply non-*Lambertian*. At first glance, it might be wiser to use as many training images as possible to construct the image data matrix so that all pixels can be illuminated at least three times, but this will only risk corrupting the data with a high percentage of outliers. Nonetheless, results obtained by implementing the above estimation method are still adequately accurate for face recognition.

2.7 Recognition

With the estimate of the facial surface matrix \mathbf{B} at hand, given a test image \mathbf{x} , recognition using the *Illumination Cones* is performed by computing the distance of \mathbf{x} to the cones of each face image in database. Obviously \mathbf{x} is matched to the face which has the shortest cone distance. However, computation speed of the distance measure might be prohibitively slow especially when the correspond to huge subspaces.

To speed up the recognition process, dimensionality reduction is performed on the cones. In [6], it was suggested that the collection of all images in the cones are projected down to a 100-dimensional feature space using *Principal Components Analysis*. A test image \mathbf{x} is recognized by first projecting it down to the 100-dimensional feature space and then nearest neighbour classification is performed.

An alternative approach is to use the *Illumination Subspace*, \mathcal{L} , directly. To perform recognition, the distance of a test image to \mathcal{L} is computed by finding the distance of \mathbf{x} to the linear subspace spanned by the three-dimensional basis defined by the estimated \mathbf{B} matrix. This can be computed by the following equation [2]:

$$dist_{\mathcal{L}}\mathbf{x} = \|\mathbf{x} - proj_{\mathcal{L}}\mathbf{x}\|, \quad (42)$$

$$where\ proj_{\mathcal{L}}\mathbf{x} = \mathbf{B}(\mathbf{B}^T\mathbf{B})^{-1}\mathbf{B}^T\mathbf{x}. \quad (43)$$

Due to its simplicity and decent robustness against illumination variation, this approach was used in the recognition stage of our experiment.

3 Experimental Results

Face images from the Yale Face Database B [7] have been used to test the methods described previously. Subjects in the database have been exposed to systematically varied light sources while maintaining a fixed posed relative to the camera. Specifically for our experiment, face images of 14 subjects (Subject 3 till Subject 16) were used. The different lighting configurations were sampled in 15° increments in spherical angles. Five subsets were grouped to quantify the effect of illumination variation. Subset 1 (and respectively Subsets 2, 3, 4 and 5) contains images for which both the longitudinal and latitudinal angles of light source direction are within 15° (and respectively 30° , 45° , 60° , and 75°) of the camera axis.

The training ensemble consists of 7 images of Subject 3 in Subset 1. As described earlier, they are face images obtained from different illumination conditions but contain the least amount of lighting variations among them, refer Figure 5. These images were fed to the system and a surface normal matrix, \mathbf{B} , was obtained as a result of the training procedure. After this stage, novel images of the training face image under unseen before lighting configurations can be created. By using randomly generated light source directions (one at a time), the resulting novel images correspond to the elements in the *Illumination Subspace* of the face. See Figure 6 for examples.



Figure 5: The training set: Subject 3 of Yale Face Database B.



Figure 6: Novel images under unseen before illumination conditions generated from the surface normal matrix.

As a comparison of the robustness of the *Illumination Cones* against lighting variations, the face image database were trained using the *Eigenface* approach as well. The training set consisted of one face image of each subject with all of them taken from Subset 1. Ten most significant *Eigenfaces* were saved as the result of this training procedure.

Training results from both methods were run through all face images available in all the subsets. Appropriate distance thresholds were used to determine whether a test image was close enough to the training set to be considered a positive match. The results in terms of error rates of both methods are summarized in Table 1 and in Figure 7.

Table 1: Error rates of both face recognition methods.

Method	Error Rate (%)				
	Subset 1	Subset 2	Subset 3	Subset 4	Subset 5
Eigenface	8.16	50.6	86.23	91	98.49
Illumination Cones/Linear Subspace	1	2.38	7.25	10.2	22.66

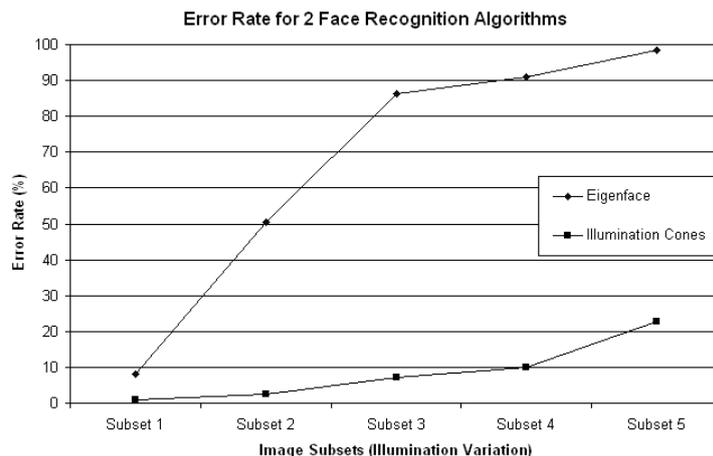


Figure 7: Error rates of both face recognition methods.

4 Conclusion

A summary of the procedures used above is presented here:

1. Construct the image data matrix \mathbf{I} using face images of fixed posed but different illumination condition.
2. Estimate an initial facial surface normal matrix \mathbf{B} using SVD coupled with the iterative procedure described in Section 2.6 to take non-*Lambertian* pixels into account.
3. Enforce integrability of the \mathbf{B} matrix using the method described in Section 2.5.
4. Perform recognition using the Linear Subspace method described in Section 2.7.

The results show that the *Illumination Cones* method outperforms the *Eigenface* method in all subsets of illumination changes. While the *Eigenface* method breaks down even with only minimal variations in Subset 2, the *Illumination Cones* method provides decent results from Subset 2 till Subset 4, before breaking down eventually in Subset 5. Though this is a big leap in improvement, the method still falls far short of matching the capability of the human visual system in face recognition. This has presented challenges and future work in the this active research area.

References

- [1] Y. Adini, Y. Moses, and S. Ullman. Face recognition: The problem of compensating for changes in illumination direction. *IEEE Transactions of Pattern Analysis and Machine Intelligence*, 19(7):721–732, 1997.
- [2] H. Anton and R.C. Busby. *Contemporary Linear Algebra*, chapter 7. Anton Textbooks. John Wiley and Sons.
- [3] R. Basri and D.W. Jacobs. Lambertian reflectance and linear subspaces. In *International Conference on Machine Vision*, 2001.
- [4] R. Basri and D.W. Jacobs. Lambertian reflectance and linear subspaces. *IEEE Transactions of Pattern Analysis and Machine Intelligence*, 25(2):218–223, 2003.
- [5] P.N. Belhumeur, R. Epstein, and A.L. Yuille. Learning object representation from lighting variations. In *Object Recognition Workshop: European Conference on Computer Vision*, 1996.
- [6] P.N. Belhumeur, A.S. Georghiades, and D.J. Kriegman. Illumination cones for recognition under variable lighting: Faces. In *IEEE Conference in Computer Vision and Pattern Recognition*, 1998.
- [7] P.N. Belhumeur, A.S. Georghiades, and D.J. Kriegman. From few to many: Illumination cones models for face recognition under variable lighting and pose. *IEEE Transactions of Pattern Analysis and Machine Intelligence*, 23(6):643–660, 2001.
- [8] P.N. Belhumeur, J.P. Hespanha, and D.J. Kriegman. Eigenfaces vs fisherfaces: Recognition using class specific linear projection. *IEEE Transactions of Pattern Analysis and Machine Intelligence*, 19(7):711–720, 1997.
- [9] P.N. Belhumeur and D.J. Kriegman. What is the set of images of an object under all possible illumination conditions? *International Journal of Computer Vision*, 28(3):245–260, 1998.
- [10] P.N. Belhumeur, D.J. Kriegman, and A.L. Yuille. The bas-relief ambiguity. *International Journal of Computer Vision*, 35(1):33–44, 1999.
- [11] P.N. Belhumeur, A.L. Yuille, and R. Epstein. Determining generative models of objects under varying illumination: Shape and albedo from multiple images using svd and integrability. *International Journal of Computer Vision*, 35(3):203–222, 2002.
- [12] M.J. Brooks and B.K.P. Horn. The variational approach to shape from shading. *Computer Vision, Graphics and Image Processing*, 33:174–208, 1986.
- [13] R. Brunelli and T. Poggio. Face recognition: Features versus templates. *IEEE Transactions of Pattern Analysis and Machine Intelligence*, 15(10):1042–1052, 1993.
- [14] I. Cox, J. Ghosn, and P. Yianilos. Feature-based face recognition using mixture distance. In *IEEE Conference of Computer Vision and Pattern Recognition*, pages 209–216, 1996.

- [15] R. Epstein, P.W. Hallinan, and A.L. Yuille. 52 eigenimages suffice: An empirical investigation of low-dimensional lighting models. In *IEEE Workshop on Physics-Based Vision*, 1995.
- [16] P.W. Hallinan. A low-dimensional representation of human faces for arbitrary lighting conditions. In *IEEE Conference on Computer Vision and Pattern Recognition*, 1994.
- [17] H. Hayakawa. Photometric stereo under a light source with arbitrary motion. *Journal of Optical Society of America*, 11(11):3079–3089, 1994.
- [18] K. Ikeuchi, R. Reddy, and H. Shum. Principal component analysis with missing data and its application to polyhedral object modeling. *IEEE Transactions of Pattern Analysis and Machine Intelligence*, 17(9):854–867, 1995.
- [19] R. Jain, R Kasturi, and B.G. Schunck. *Machine Vision*, chapter 9. Computer Science Series. McGraw-Hill International Editions, 1995.
- [20] T. Kanade. Picture processing by computer complex and recognition of human faces. Technical report, Dept. Inform. Sci., Kyoto University, 1973.
- [21] Y. Kaya and K. Kobayashi. A basic study on human face recognition. In *Frontiers of Pattern Recognition*, page 265, 1972.
- [22] M. Kirby and L. Sirovich. Application of the karhunen-loeve procedure for the characterization of human faces. *IEEE Transactions of Pattern Analysis and Machine Intelligence*, 12(1), 1990.
- [23] D. Snow and A. Yuille. Shape and albedo from multiple images using integrability. In *IEEE Conference on Computer Vision and Pattern Recognition*, pages 158–164, 1997.
- [24] M. Turk and A. Pentland. Eigenfaces for recognition. *Journal of Cognitive Neuroscience*, 3(1):71–86, 1991.
- [25] A.L. Yuille. Deformable templates for face recognition. *Journal of Cognitive Neuroscience*, 3(1):59–70, 1991.