

What is the Set of Images of an Object Under All Possible Lighting Conditions?

Peter N. Belhumeur*

David J. Kriegman†

Dept. of Electrical Engineering, Yale University, New Haven, CT 06520-8267

Abstract

The appearance of a particular object depends on both the viewpoint from which it is observed and the light sources by which it is illuminated. If the appearance of two objects is never identical for any pose or lighting conditions, then – in theory – the objects can always be distinguished or recognized. The question arises: What is the set of images of an object under all lighting conditions and pose? In this paper, we consider only the set of images of an object under variable illumination (including multiple, extended light sources and attached shadows). We prove that the set of n -pixel images of a convex object with a Lambertian reflectance function, illuminated by an arbitrary number of point light sources at infinity, forms a convex polyhedral cone in $\mathbb{R}^n$ and that the dimension of this illumination cone equals the number of distinct surface normals. Furthermore, we show that the cone for a particular object can be constructed from three properly chosen images. Finally, we prove that the set of n -pixel images of an object of any shape and with an arbitrary reflectance function, seen under all possible illumination conditions, still forms a convex cone in $\mathbb{R}^n$. These results immediately suggest certain approaches to object recognition. Throughout this paper, we offer results demonstrating the empirical validity of the illumination cone representation.

1 Introduction

One of the complications that has troubled computer vision recognition algorithms is the variability of an object's appearance from one image to the next. With slight changes in lighting conditions and viewpoint often comes large changes in the object's appearance. To handle this variability methods usually take one of two approaches: either measure some property in the image of the object which is, if not invariant, at least insensitive to the variability in the imaging conditions, or model the object, or part of the object, in order to predict the variability.

Nearly all approaches to object recognition have handled the variability due to illumination by using the first approach; they have concentrated on edges in the images, i.e. the discontinuities in the image intensity. Because discontinuities in the albedo on the surface of the object or discontinuities in albedo across the boundary of the object generate edges in images, these edges tend to be insensitive to a range of illumination conditions.

Yet, edges do not contain all of the information useful for recognition. Furthermore, objects which are not composed of piecewise constant albedo patterns often produce inconsistent edge maps. The reason most approaches have avoided using the rest of the intensity information is because its variability under changing illumination has been difficult to tame. Only recently have "appearance-based" approaches been developed in an effort to use intensity information to model or learn a representation that captures a large set of the possible images of an object under pose and/or illumination variation [6, 9, 12, 13].

In this paper, we consider some of the fundamental issues that must be addressed if recognition is to be performed under variation in illumination. Our goal is to develop methods that are insensitive to extreme variation in lighting conditions – variation for which edge based methods would surely fail. To this end, we answer a number of questions: If an image with n pixels is treated as a point in $\mathbb{R}^n$, what is the set in $\mathbb{R}^n$ of all images of an object under varying illumination? Is this set an incredibly complex, but low-dimensional manifold in the image space? Or does the set have a simple, predictable structure? Can a finite number of images characterize this set? If so, how many images are needed?

The image formation process for a particular object can be viewed as a function of pose and lighting. Since an object's pose can be represented by a point in $\mathbb{R}^3 \times SO(3)$ (a six dimensional manifold), the set of n -pixel images of an object under constant illumination, but over all possible poses, is at most six dimensional. Murase and Nayar take advantage of this structure when constructing appearance manifolds [6]. However, if the pose is held constant, the variability due to illumination may be much larger as the set of possible lighting conditions is infinite dimensional.

Nonetheless, we will show that the set of all images of a convex object with a Lambertian reflectance function is a convex polyhedral cone in $\mathbb{R}^n$ where n is the number of pixels in each image, and that the dimension of the cone is equal to the number of distinct surface normals, independent of the albedo pattern. Furthermore, we will show that this cone can be determined from three properly chosen images of the object. Empirical investigations support the validity of this representation, and these results suggest some specific recognition algorithms.

2 The Illumination Cone

In this section, we will develop our illumination cone representation. The development will be based

*P. N. Belhumeur was supported by ARO.

†D. J. Kriegman was supported by NSF under an NYI, IRI-9257990 and by ONR N00014-93-1-0305.

upon two assumptions: First, we assume that the surfaces of objects have Lambertian reflectance functions. Second, we assume that the shape of an object's surface is convex. While the theorems we state and prove are based upon these two assumptions, we will demonstrate the validity of the illumination cone representation by presenting results on images of objects which have neither purely Lambertian surfaces, nor convex shapes.

2.1 Developing the Illumination Cone

To begin let us assume a Lambertian model for reflectance with a single point light source at infinity. Let $\mathbf{x}$ denote an image with n pixels. Let $B \in \mathbb{R}^{n \times 3}$ be a matrix where each row of B is the product of the albedo with the inward pointing unit normal for a point on the surface projecting to a particular pixel; here we assume that the surface normals for the set of points projecting to the same image pixel are identical. Let $\mathbf{s} \in \mathbb{R}^3$ be a column vector signifying the product of the light source strength with the unit vector for the light source direction. Thus, a convex object with surface normals and albedo given by B , seen under illumination $\mathbf{s}$, produces an image $\mathbf{x}$ given by the following equation

$$\mathbf{x} = \max(B\mathbf{s}, 0), \quad (1)$$

where $\max(\cdot, 0)$ zeros all negative components of the vector $B\mathbf{s}$ [5]. Note that the negative components of $B\mathbf{s}$ correspond to the shadowed surface points and are sometimes called *attached shadows* [11]. Also, note that convexity of the object's shape is needed to guarantee that the object does not cast shadows on itself.

If the object is seen under illumination by k point light sources at infinity, the image $\mathbf{x}$ is given by the superposition of images which would have been produced by the individual light source, i.e.

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

where $\mathbf{s}_i$ is a single light source. Note that extended light sources at infinity can be handled by allowing an infinite number of point light sources.

The product of B with all possible light source directions and strengths sweeps out a 3-D subspace in the n dimensional image space [11]; we call the subspace created by B the illumination subspace $\mathcal{L}$, where

$$\mathcal{L} = \{\mathbf{x} \mid \mathbf{x} = B\mathbf{s}, \forall \mathbf{s} \in \mathbb{R}^3\}.$$

When a single light source is parallel with the camera's optical axis, all visible points on the surface are illuminated, and, consequently, all pixels in the image have non-zero values. The set of images created by scaling the light source strength and moving the light source away from the direction of the camera's optical axis such that all pixels remain illuminated can be found as the relative interior of a set $\mathcal{L}_0$ defined by the intersection of $\mathcal{L}$ with the non-negative orthant of $\mathbb{R}^n$.¹ We will show presently that $\mathcal{L}_0$ is a convex cone

¹By orthant we mean the high-dimensional analogue to quadrant, i.e. the set $\{\mathbf{x} \mid \mathbf{x} \in \mathbb{R}^n, \text{ with certain components}$

in $\mathbb{R}^n$. However, to show this, we first need the definition of convexity and the definition of a cone [2, 10].

Definition 1 $X \subset \mathbb{R}^n$ is convex iff for any two points $\mathbf{x}_1, \mathbf{x}_2 \in X$, $\lambda\mathbf{x}_1 + (1 - \lambda)\mathbf{x}_2 \in X$ for every $\lambda \in [0, 1]$.

Definition 2 $X \subset \mathbb{R}^n$ is a cone iff for any point $\mathbf{x} \in X$, $\alpha\mathbf{x} \in X$ for every $\alpha \geq 0$.

With these definitions, we state the following lemma:

Lemma 1 The set of images $\mathcal{L}_0$ is a convex cone in $\mathbb{R}^n$.

Proof. To prove that $\mathcal{L}_0$ is convex, we appeal to the definition of convexity. Now $\mathcal{L}_0 = \mathcal{L} \cap \{\mathbf{x} \mid \mathbf{x} \in \mathbb{R}^n, \text{ with all components of } \mathbf{x} \geq 0\}$. Because $\mathcal{L}$ is a linear subspace, if $\mathbf{x}_1, \mathbf{x}_2 \in \mathcal{L}$, then $\lambda\mathbf{x}_1 + (1 - \lambda)\mathbf{x}_2 \in \mathcal{L}$. And, if $\mathbf{x}_1$ and $\mathbf{x}_2$ both have all components non-negative, then $\lambda\mathbf{x}_1 + (1 - \lambda)\mathbf{x}_2$ has all components non-negative for every $\lambda \in [0, 1]$. Thus, $\lambda\mathbf{x}_1 + (1 - \lambda)\mathbf{x}_2 \in \mathcal{L}_0$.

To prove that $\mathcal{L}_0$ is a cone, we appeal to the definition of a cone. Because $\mathcal{L}$ is a linear subspace, if $\mathbf{x} \in \mathcal{L}$ then $\alpha\mathbf{x} \in \mathcal{L}$. And, if $\mathbf{x}$ has all components non-negative, then $\alpha\mathbf{x}$ has all components non-negative for every $\alpha \geq 0$. Therefore $\alpha\mathbf{x} \in \mathcal{L}_0$. So it follows that $\mathcal{L}_0$ is a cone. ■

As we move the light source direction further from the camera's optical axis, points on the object will fall into shadow. Naturally, which pixels are the image of shadowed or illuminated surface points depends on where we move the light source direction. If we move the light source all the way around to the back of the object so that the camera's optical axis and the light source are pointing in opposite directions, then all pixels are in shadow.

Let us now consider all possible light source directions, representing each direction by a point on the surface of the sphere; we call this sphere the *illumination sphere*. The set of light source directions for which a given pixel in the image is illuminated corresponds to an open hemisphere of points; the set of light source directions for which the pixel is shadowed corresponds to the other hemisphere of points. A great circle on the illumination sphere divides these sets.

Each of the n pixels in the image has a corresponding great circle on the illumination sphere. The collection of great circles carves up the surface of the illumination sphere into cells. See Fig 1.

The collection of light source directions contained within a single cell on the illumination sphere produces a set of images, each with the same pixels in shadow and the same pixels illuminated; we say that these images have the same "shadowing configurations." Different cells produce different shadowing configurations.

We call the cell on the illumination sphere containing the collection of light source directions which produce images with all pixels illuminated the *bright cell*.

of $\mathbf{x} \geq 0$ and the remaining components of $\mathbf{x} < 0$. By non-negative orthant we mean the set $\{\mathbf{x} \mid \mathbf{x} \in \mathbb{R}^n, \text{ with all components of } \mathbf{x} \geq 0\}$.

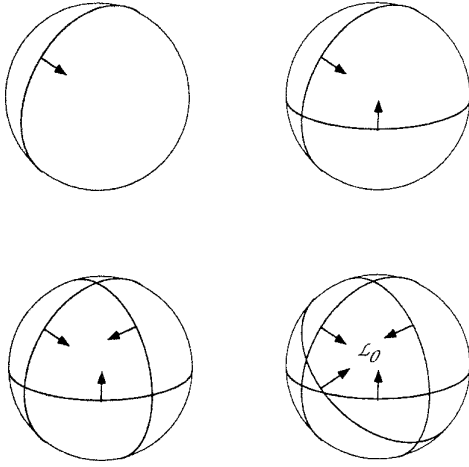


Figure 1: The Illumination Sphere: The set of all possible light source directions can be represented by points on the surface of a sphere; we call this sphere the *illumination sphere*. Great circles corresponding to individual pixels divide the illumination sphere into cells of different shadowing configurations. The arrows indicate the hemisphere of light directions for which the particular pixel is illuminated. The cell of light source directions which produce $\mathcal{L}_0$, the set of images in which all pixels are illuminated, is labeled in the figure. Each of the other cells produce the $\mathcal{L}_i$, $0 < i \leq n(n-1)+1$.

Thus, the collection of light source directions from the interior and boundary of the bright cell produces the set of images $\mathcal{L}_0$. To determine the set of images produced by another cell on the illumination sphere, we need to return to the illumination subspace $\mathcal{L}$.

The illumination subspace $\mathcal{L}$ not only slices through the non-negative orthant of $\mathbb{R}^n$, but other orthants in $\mathbb{R}^n$ as well. Let $\mathcal{L}_i$ be the intersection of the illumination subspace $\mathcal{L}$ with an orthant i in $\mathbb{R}^n$ through which $\mathcal{L}$ passes. Certain components of $\mathbf{x} \in \mathcal{L}_i$ are always negative and others always greater than or equal to zero. Not surprisingly, each $\mathcal{L}_i$ has a corresponding cell of light source directions on the illumination sphere. Note that $\mathcal{L}$ does not slice through all of the 2^n orthants in $\mathbb{R}^n$, but at most $n(n-1)+2$ orthants (see the proof of Theorem 1). Thus, there are at most $n(n-1)+2$ sets $\mathcal{L}_i$, or cells on the illumination sphere.

The set of images produced by the collection of light source directions from a cell other than the bright cell can be found as a projection P_i of all points in a particular set $\mathcal{L}_i$. The projection P_i is such that it leaves the non-negative components of $\mathbf{x} \in \mathcal{L}_i$ untouched, while the negative components of $\mathbf{x}$ become zero. We denote the projected set by $P_i(\mathcal{L}_i)$.

Lemma 2 *The set of images $P_i(\mathcal{L}_i)$ is a convex cone in $\mathbb{R}^n$.*

Proof. By the same argument used in the proof of Lemma 1, $\mathcal{L}_i$ is a convex cone. Since the linear pro-

jection of convex cone is itself a convex cone, $P_i(\mathcal{L}_i)$ is a convex cone. ■

Since $P_i(\mathcal{L}_i)$ is the projection of $\mathcal{L}_i$, it is at most three dimensional. Each $P_i(\mathcal{L}_i)$ is the set of all images such that certain pixels are illuminated, and the remaining pixels are shadowed. Let P_0 be the identity, so that $P_0(\mathcal{L}_0) = \mathcal{L}_0$ is the set of all images such that all pixels are illuminated. The number of possible *shadowing configurations* is the number of orthants in $\mathbb{R}^n$ through which the illumination subspace $\mathcal{L}$ passes, which in turn is the same as the number of sets $P_i(\mathcal{L}_i)$.

Theorem 1 *The number of shadowing configurations is at most $m(m-1)+2$, where $m \leq n$ is the number of distinct surface normals.*

Proof. Each of the n pixels in the image has a corresponding great circle on the illumination sphere, but only m of the great circles are distinct. The collection of m distinct great circles carves up the surface of the illumination sphere into cells. Each cell on the illumination sphere corresponds to a particular set of images $P_i(\mathcal{L}_i)$. Thus, the problem of determining the number of shadowing configurations is the same as the problem of determining the number of cells on the illumination sphere. If every vertex on the illumination sphere is formed by the intersection of only two of the m distinct great circles (i.e. if no more than two surface normals are coplanar), then it can be shown by induction that the illumination sphere is divided into $m(m-1)+2$ cells. If a vertex is formed by the intersection of three or more great circles, there are fewer cells. ■

Thus, the set $\mathcal{U}$ of images of a convex Lambertian surface created by varying the direction and strength of *single* point light source at infinity is given by the union of at most $n(n-1)+2$ convex cones, i.e.

$$\mathcal{U} = \{\mathbf{x} \mid \mathbf{x} = \max(Bs, 0), \forall s \in \mathbb{R}^3\} = \bigcup_{i=0}^{n(n-1)+1} P_i(\mathcal{L}_i).$$

From this set, we can construct the set $\mathcal{C}$ of all possible images of a convex Lambertian surface created by varying the direction and strength of an *arbitrary* number of point light sources at infinity,

$$\mathcal{C} = \{\mathbf{x} : \mathbf{x} = \sum_{i=1}^k \max(Bs_i, 0), \forall s_i \in \mathbb{R}^3, \forall k \in \mathbb{Z}^+\}$$

where $\mathbb{Z}^+$ is the set of positive integers.

Theorem 2 *The set of images $\mathcal{C}$ is a convex cone in $\mathbb{R}^n$.*

Proof. The proof that $\mathcal{C}$ is a cone follows trivially from the definition of $\mathcal{C}$. To prove that $\mathcal{C}$ is convex, we appeal to a theorem for convex cones which states that a cone $\mathcal{C}$ is convex iff $\mathbf{x}_1 + \mathbf{x}_2 \in \mathcal{C}$ for any two points $\mathbf{x}_1, \mathbf{x}_2 \in \mathcal{C}$ [2]. So the proof that $\mathcal{C}$ is convex follows trivially from the above definition of $\mathcal{C}$. ■

We call $\mathcal{C}$ the *illumination cone*, and every object has its own illumination cone. Note that each point in the cone is an image of the object under a particular lighting configuration, and the entire cone is the set of images of the object under all possible configurations of point light sources at infinity.

Theorem 3 *The illumination cone $\mathcal{C}$ can be completely determined by three properly chosen images, i.e. three linearly independent images from the set $\mathcal{L}_0$, without any knowledge of the light source directions.*

Proof. The illumination cone $\mathcal{C}$ is completely determined by the illumination subspace $\mathcal{L}$, which is determined both by B , which in turn is determined by three linearly independent images in $\mathcal{L}_0$ and by the corresponding light source directions. As has been known to the photometric stereo community for over a decade, without prior knowledge of the light sources, we cannot determine B uniquely. To see this note that for any arbitrary invertible 3×3 linear transformation $A \in GL(3)$

$$Bs = (BA)(A^{-1}s),$$

i.e. the albedo and surface normals are transformed by A , while the light source is transformed by A^{-1} . Therefore, we can only recover B up to an arbitrary invertible linear transformation $B^* = BA$. Nevertheless, even without prior knowledge of the light source directions or the object geometry, we can still use B^* to determine what the object will look like under different illumination: it is easy to show that $\mathcal{L} = \{\mathbf{x} \mid \mathbf{x} = B^*s, \forall s \in \mathbb{R}^3\} = \{\mathbf{x} \mid \mathbf{x} = Bs, \forall s \in \mathbb{R}^3\}$, see [11]. ■

Thus, for a convex object with Lambertian reflectance, we can determine its appearance under arbitrary illumination from as few as three images of the object. Knowledge of the light source strength or direction is *not* needed, see also [11]. What may not be immediately obvious is that any point within the cone $\mathcal{C}$ (including the boundary points) can be found as a convex combination of the rays (images) produced by light source directions lying at the $m(m-1)$ intersections of the great circles. Furthermore, because the cone is constructed from a finite number of extreme rays (images), the cone is polyhedral.

We should point out that nothing in the proof of Theorem 2 required assumptions about the shape of the object, the nature of the light sources, or the reflectance function for the object's surface. Thus, we can state a more general theorem about the set of images of an object under varying illumination:

Theorem 4 *The set of n -pixel images of any object, seen under all possible lighting conditions, is a convex cone in $\mathbb{R}^n$.*

2.2 Empirical Investigation of the Illumination Cone

To demonstrate the power of these concepts, we have constructed the illumination cone for two different scenes: a human face and a desktop still life. To construct the cone for the human face, we used images

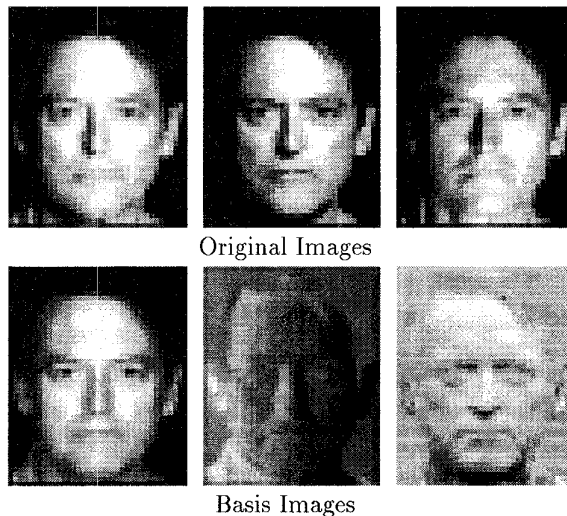


Figure 2: The top half of the figure shows three of the original images used to construct the illumination subspace $\mathcal{L}$ of the face. The bottom half of the figure shows three basis images, lying in $\mathcal{L}_0$, that span the illumination subspace $\mathcal{L}$ for the face.

from the Harvard Face Database [4], a collection of images of faces seen under a range of lighting directions. For the purpose of this demonstration, we used the images of one person, taking six images from the set $\mathcal{L}_0$ and using singular value decomposition (SVD) to construct a 3-D basis for the illumination subspace $\mathcal{L}$. Note that this 3D linear subspace differs from the affine subspace constructed using the Karhunen-Loeve transform: the mean image is not subtracted before determining the basis vectors as in the Eigenpicture methods [12, 13].

The illumination subspace was then used to construct the illumination cone $\mathcal{C}$. We generated novel images of the face as if illuminated by one, two, or three point light sources by randomly sampling the illumination cone. Rather than constructing an explicit representation of the halfspaces bounding the illumination cone, we sampled $\mathcal{L}$, determined the corresponding orthant, and appropriately projected the image onto the illumination cone. Images constructed under multiple light sources simply correspond to the superposition of the images generated by each of the light sources.

The top half of Fig. 2 shows three of the images of a person's face that were used to construct the basis of the linear subspace $\mathcal{L}$. The bottom half of Fig. 2 shows three basis images that span $\mathcal{L}$. The three columns of Fig. 3 respectively comprise sample images from the illumination cone for the face with one, two, or three light sources.

There are number of interesting points to note about this experiment. There was almost no shadowing in the training images yet there are strong attached shadows in many of the sample images. They are particularly distinct in the images generated with a single light source. Notice for example the sharp



Figure 3: **Random Samples from the Illumination Cone of a Face:** Each of the three columns respectively comprises sample images from the illumination cone with one, two and three light sources.

shadow across the ridge of the nose in column 1, row 2 or the shadowing in column 1, row 4 where the light source is coming from behind the head. Notice also the depression under the cheekbones in column 2, row 5, and the cleft in the chin revealed in column 1, row 3. For the image in column 3, row 2, two of the light sources are on opposite sides while the third one is coming from below; notice that both ears and bottom of the chin and nose are brightly illuminated while

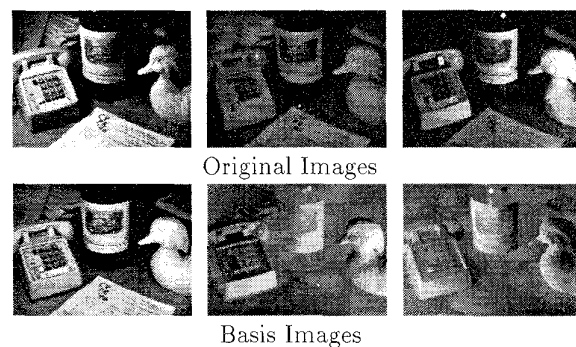


Figure 4: The top half of the figure shows three of the original images used to construct the illumination subspace $\mathcal{L}$ of the still life. The bottom half of the figure shows three basis images, lying in $\mathcal{L}_0$, that span the illumination subspace $\mathcal{L}$ for the still life.



Figure 5: **Random Samples from the Illumination Cone of a Desktop Still Life:** Each of the three columns respectively comprises sample images from the illumination cone with one, two and three light sources.

that rest of the face is darker.

To construct the cone for the desktop still life, we used our own collection of nine images from (or near) the set $\mathcal{L}_0$. The top half of Fig. 4 shows three of these images. The bottom half of Fig. 4 shows the

three basis images that span $\mathcal{L}$. The three columns of Fig. 5 respectively comprise sample images from the illumination cone for the desktop still life with one, two or three light sources.

The variability in illumination in these images is so extreme that the edge maps for these images would differ drastically. Notice in the image in column 1, row 4 that the shadow line on the bottle is distinct and that the left sides of the phone, duck, and bottle are brightly illuminated. Throughout the scene, notice that those points having comparable surface normals seem to be similarly illuminated. Furthermore, notice that all of the nearly horizontal surfaces in the bottom two images of the first column are in shadow since the light is coming from below. In the image with two light sources shown at the bottom of column 2, the sources are located on opposite sides and behind the objects. This leads to shadow line in the center of the bottle. The head of the wood duck shows a similar shadowing where the front and back of the head are illuminated, but not the side.

3 Dimension of the Illumination Cone

In this section, we investigate the dimension of the illumination cone, and show that it is equal to the number of distinct surface normals. We also present experimental results validating this claim.

3.1 Determining the Dimension

Given that the set of images of an object under variation in illumination is a convex cone, it is natural to ask what is the dimension of the cone in $\mathbb{R}^n$? By this we mean, what is the span of the vectors in the illumination cone $\mathcal{C}$?

Why do we want to know the answer to this question? Because the complexity of the cone, will almost certainly dictate the nature of the recognition algorithm. For example, if the illumination cones are 1-D, i.e. rays in the positive orthant of $\mathbb{R}^n$, then a recognition scheme based on normalized correlation could handle all of the variation due to illumination. However, in general the cones are not one dimensional. To this end, we offer the following theorem.

Theorem 5 *The dimension of the illumination cone $\mathcal{C}$ is equal to the number of distinct surface normals.*

Proof. As for the proof of Theorem 1, we again represent each light source direction by a point on the surface of the illumination sphere. Each cell on the illumination sphere corresponds to the light source directions which produce a particular set of images $P_i(\mathcal{L}_i)$. For every image in a set $P_i(\mathcal{L}_i)$ certain pixels always equal zero, i.e. always in shadow. There exists a cell on the illumination sphere corresponding to the light source directions which produce $\mathcal{L}_0$, the set of images in which all pixels are always illuminated – we call this the bright cell. There exists a cell corresponding to the light source directions which produce a set of images in which all pixels are always in shadow – we call this the dark cell. Choose any point $\mathbf{s}_b$ within the bright cell. The point $\mathbf{s}_d = -\mathbf{s}_b$ is antipodal to $\mathbf{s}_b$ and lies within the dark cell. Draw any half-meridian connecting $\mathbf{s}_b$ and $\mathbf{s}_d$. Starting at $\mathbf{s}_b$, follow the path of

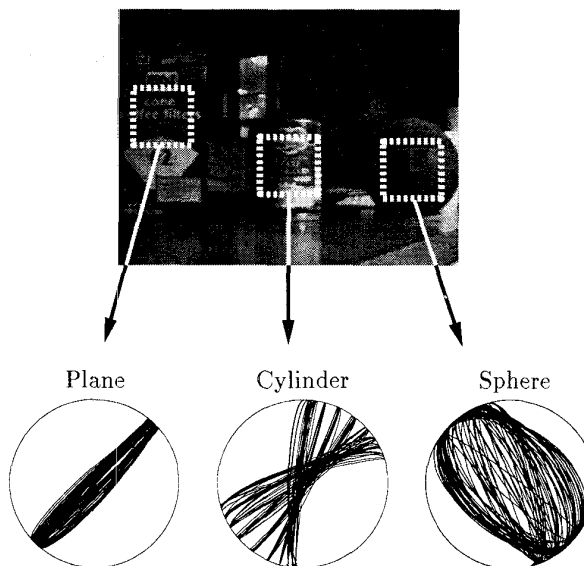


Figure 6: **Examples of Illumination Spheres:** The figure shows a view of the illumination sphere for a planar, cylindrical, and spherical object.

the half-meridian; it crosses m distinct great circles, and passes through m different cells before entering the dark cell. Note the path of the half-meridian corresponds to a particular path of light source directions, starting from a light source direction producing an image in which all pixels are illuminated and ending at a light source direction producing an image in which all pixels are in shadow. Each time the half-meridian crosses a great circle, the pixel corresponding to the great circle becomes shadowed.

Take an image produced from any light source direction within the interior of each cell through which the meridian passes, including the bright cell, but excluding the dark cell. Arrange each of these m images as column vectors in an $n \times m$ matrix M . By elementary row operations, the matrix M can be converted to its echelon form M^* , and it is trivial to show that M^* has exactly m non-zero rows. Thus, the rank of M is m , and the dimension of $\mathcal{C}$ is at least m . Since there are only m distinct surface normals, the dimension of $\mathcal{C}$ cannot exceed m . Thus, the dimension of $\mathcal{C}$ equals m . ■

Note that for images with n pixels, this theorem indicates that the dimension of the illumination cone is one for a planar object, is roughly $\sqrt{n}$ for a cylindrical object, and is n for a spherical object. But if the cone spans $\mathbb{R}^n$, what fraction of the positive orthant does it occupy? And if this fraction can be estimated, can we make claims on the upper bounds for recognition rates of databases with multiple objects?

3.2 Empirical Investigation

To empirically demonstrate the validity of Theorem 1, we have constructed the linear subspaces for three objects using an image of a box of coffee filters, a can of corn, and a bocce ball. For each object, we computed

the great circles associated with each pixel and display one hundred of them in Fig. 6. As mentioned above, we would expect the illumination cone produced by one face of the box of coffee filters to be one dimensional since all of the surface normals are identical. The illumination sphere should be partitioned into two regions by a single great circle. This is nearly seen in the figure. Due to both image noise and the fact that the surface is not truly Lambertian, there is some small deviation of the surface normals. For a cylinder, the surface normals lie on a great circle of the Gauss sphere, and so we would expect the illumination sphere to be partitioned by lines of longitude intersecting at a common pole. This structure is also seen in the figure. Finally, the visible surface normals of a sphere should cover half of the Gauss sphere and so the great circles should cover the illumination sphere.

From these three examples we make the following observations. The partitioning of the illumination sphere has the expected qualitative structure. However, due to a combination of image noise, use of extended light sources near the object, and surfaces that are not truly Lambertian, none of the surface normals are truly coincident. Therefore, n distinct surface normals are computed, and so there actually are $n(n-1)-2$ regions for all three objects. Thus, the corresponding cones will span n dimensions. However, it is expected that the illumination cone for the coffee filter box would be small, that the cone for the can would be flat except in $\sqrt{n}$ directions, and that the cone for the sphere would have the largest volume.

4 Discussion

In this paper, we have shown that the set of images of a convex object with a Lambertian reflectance function, under all possible lighting conditions, is a convex, polyhedral cone. Furthermore, we have shown that for convex object with Lambertian reflectance functions the cone be learned from three properly chosen images and that the dimension of the cone equals the number of distinct surface normals. These are fundamental results which any recognition system that operates under variable illumination should consider.

Nevertheless, there remains a number of extensions and open issues which we discuss below. While we have focused this paper solely on *characterizing* the set of images under varying illumination, we believe the ideas presented within have natural applications to recognition algorithms.

4.1 Shape of Illumination Cone

While we have shown that the illumination cone is a convex, polyhedral cone that can span n dimensions if there are n distinct surface normals, we have not said how big it is in practice. The illumination cone for a cylinder would ideally span $\sqrt{n}$ dimensions. When it is computed from noisy data, it is likely to actually span n dimensions. However, it is expected to be rather "flat" in most dimensions and only have significant extent in $\sqrt{n}$ directions. Clearly if the illumination cones for objects are small and well separated, then recognition should be easy – in theory – even under extreme lighting conditions.

On the other hand, is it conceivable that the illumination cone could completely cover the positive orthant of $\mathbb{R}^n$. Note that having the cone span n dimensions does not mean that it covers $\mathbb{R}^n$, since the cone is defined only by positive combinations of the basis vectors. The existence of an object geometry whose images cover the positive orthant seems unlikely. For such an object, it must be possible to choose n light source directions; each light source must illuminate a single pixel while all others are in shadow, and each light source must illuminate a different pixel. If this were possible, the n images produced by n such light source directions would form basis vectors aligned with a coordinate axes of $\mathbb{R}^n$.

4.2 Relaxing Assumptions

Two important assumptions have been made along the way: First, we assumed that the object's shape is convex; second, we assumed the object has a Lambertian reflectance function. The first assumption was needed so that we could ignore the effects of cast shadows. However, if cast shadows do not dominate, we hope that the illumination cone is a fairly accurate approximation of the set of images of the object under all possible lighting conditions. If the object is highly non-convex and cast shadows do dominate, we should again point out that the set of images of the object under varying illumination remains a convex cone! The question remains: Can the cone be learned from as few as three images?

The second assumption was needed so that we could ignore the effects of specularities. Yet, as the recent work of Yuille and Epstein has shown [3], even for surfaces which appear to be highly specular, the Lambertian component of the reflectance function still dominates. So again, we believe that for most objects the illumination cone is a fairly accurate approximation of the set of all images. In the experiments shown in Section 2.2, the objects were neither purely Lambertian, nor completely convex. Yet, the illumination cone appears to give an accurate representation of the set of images under varying illumination, even though the training images had little variation in illumination.

4.3 Color

The images produced by a color camera with narrow-band spectral filters for each of color channel can be considered as a point in $\mathbb{R}^{3n}$. It is straight forward to show that the set of images of a convex, colored Lambertian surface produced by a single light source at infinity with arbitrary spectral distribution (color) and without shadowing is a five dimension set in $\mathbb{R}^{3n}$; this set is not a linear subspace. However, through superposition, the set of images produced by two light sources without shadowing lies in a 9-D linear subspace of $\mathbb{R}^{3n}$. This 9-D linear subspace can be considered the Cartesian product of three 3-D linear subspaces, one for each color channel. When attached shadows are considered, the set of possible images is given by the Cartesian product of the illumination cones for each color channel. Again, three color images without shadowing are sufficient for constructing this color illumination cone.

4.4 Interreflection

A surface is not just illuminated by the light sources but also through interreflections from points on the surface itself. For a Lambertian surface, the image with interreflection $\mathbf{x}'$ is related to the image that would be formed without interreflection $\mathbf{x}$ by

$$\mathbf{x}' = (I - PK)^{-1}\mathbf{x}$$

where I is the identity matrix, P is a diagonal matrix with $P_{i,i}$ equal to the albedo of pixel i , and K is known as the interreflection kernel [7]. When there is no shadowing, all images lie in a 3-D linear space that would be generated from (1) by a pseudo-surface whose normals and albedo B' are given by $B' = (I - PK)^{-1}B$ [7, 8]. From Theorem 4, the set of all possible images is still a cone. While B' can be learned from only three images, the set of shadowing configurations and the partitioning of the illumination sphere is generated from B , not B' . So, it remains an open question how the cone can be constructed from only three images.

4.5 Effects of Change in Pose

All of the previous analysis in the paper has dealt solely with variation in illumination. Yet, a change in the object's pose creates a change in the perceived image. If an object undergoes a rotation or translation, how does the illumination cone deform? The illumination cone of the object in the new pose is also convex, but almost certainly different from the illumination cone of the object in the old pose. Which raises the question: Is there a simple transformation, obtainable from a small number of images of the object seen from different views, which when applied to the illumination cone characterizes these changes? Alternatively, is it practical to simply sample the pose space constructing an illumination cone for each pose? Currently, Nayar and Murase are extending their appearance manifold representation by modeling illumination variation for each pose as a 3-D linear subspace [8]. However, this representation does not account for shadowing.

4.6 Object Recognition

Ultimately, we intend to apply the illumination cone concept to recognition. In earlier face recognition work, we implemented a classification method based on the minimum distance of an image to a three dimensional linear subspace (i.e. $\mathcal{L}$) that was constructed for each person. Experimentally, recognition rates were perfect when there was only moderate variation in illumination with minor shadowing [1]. While the distance to illumination subspace algorithm performed impressively, the experiments also indicated that its performance deteriorates as the test images move further and further away from the illumination subspace $\mathcal{L}$. We believe that nearly perfect recognition rates could be achieved under extreme variation in illumination, by measuring distance to the illumination cone, rather than distance to the illumination subspace. Measuring the squared distance to a convex cone can be posed as a non-negative least-squares problem which is easily solved by convex programming since both the objective function and cone are convex.

It is important to stress that the illumination cones are convex. If they are non-intersecting, then the cones are linearly separable. That is, they can be separated by a $n - 1$ dimensional hyperplane in $\mathbb{R}^n$ passing through the origin. Furthermore since convex sets remain convex under linear projection, then for any projection direction lying in the separating hyperplane, the projected convex sets will also be linearly separable. For d different objects represented by d linearly separable convex cones, there always exists a linear projection of the image space to an $d - 1$ dimensional space such that all of the projected sets are again linearly separable. So an alternative to classification based on measuring distance to the cones in $\mathbb{R}^n$ is to find a much lower dimensional space in which to do classification. In our Fisherface method for recognizing faces under variable illumination and facial expression, projection directions were chosen to maximize separability of the object classes [1]; a similar approach can be taken here.

References

- [1] P. N. Belhumeur, J. P. Hespanha, and D. J. Kriegman. Eigenfaces vs. Fisherfaces: Recognition using class specific linear projection. In *Proc. European Conf. on Computer Vision*, 1996.
- [2] M. Canon, C. Cullum Jr., and E. Polak. *Theory of Optimal Control and Mathematical Programming*. McGraw-Hill, New York, 1970.
- [3] R. Epstein, P. Hallinan, and A. Yuille. 5 ± 2 eigenimages suffice: An empirical investigation of low-dimensional lighting models. Technical Report 94-11, Harvard University, 1994.
- [4] P. Hallinan. A low-dimensional representation of human faces for arbitrary lighting conditions. In *Proc. IEEE Conf. on Comp. Vision and Patt. Recog.*, pages 995-999, 1994.
- [5] B. Horn. *Computer Vision*. MIT Press, Cambridge, Mass., 1986.
- [6] H. Murase and S. Nayar. Visual learning and recognition of 3-D objects from appearance. *Int. J. Computer Vision*, 14(5-24), 1995.
- [7] S. Nayar, K. Ikeuchi, and T. Kanade. Shape from interreflections. *IJCV*, 6(3):173-195, August 1991.
- [8] S. Nayar and H. Murase. Dimensionality of illumination in appearance matching. *IEEE Conf. on Robotics and Automation*, 1996.
- [9] A. Pentland, B. Moghaddam, and Starner. View-based and modular eigenspaces for face recognition. In *Proc. Conf. Computer Vision and Pattern Recognition*, pages 84-91, 1994.
- [10] R. Rockafellar. *Convex Analysis*. Princeton University Press, Princeton, 1970.
- [11] A. Shashua. *Geometry and Photometry in 3D Visual Recognition*. PhD thesis, MIT, 1992.
- [12] Sirovitch, L. and Kirby, M. Low-dimensional procedure for the characterization of human faces. *J. Optical Soc. of America A*, 2:519-524, 1987.
- [13] M. Turk and A. Pentland. Eigenfaces for recognition. *J. of Cognitive Neuroscience*, 3(1), 1991.