# Comparing Images Under Variable Illumination

David W. Jacobs
NEC Research Institute
Princeton, NJ 08540
dwj@research.nj.nec.com

Peter N. Belhumeur
Center for Computational Vision and Control
Yale University
New Haven, CT 06520
belhumeur@yale.edu

Ronen Basri
The Weizmann Institute
Rehovot, 76100, Israel
ronen@wisdom.weizmann.ac.il

## Abstract

*We consider the problem of determining whether two images come from different objects or the same object in the same pose, but under different illumination conditions. We show that this problem cannot be solved using hard constraints: even using a Lambertian reflectance model, there is always an object and a pair of lighting conditions consistent with any two images. Nevertheless, we show that for point sources and objects with Lambertian reflectance, the ratio of two images from the same object is simpler than the ratio of images from different objects. We also show that the ratio of the two images provides two of the three distinct values in the Hessian matrix of the object's surface. Using these observations, we develop a simple measure for matching images under variable illumination, comparing its performance to other existing methods on a database of 450 images of 10 individuals.*

## 1 Introduction

A central problem of visual object recognition is to use information about an object derived from sample images in order to recognize that object under novel viewing conditions. In model-based approaches, it is typically assumed that training images are used to derive some definite information about the shape of an object that accurately predicts some properties of its appearance in any new image; in appearance-based vision an effort is made to represent the set of all images an object can produce, either by sampling them or by generating a representation from a small set of training images. Both approaches encounter difficulties when few training images are available, or when the images are taken under uncontrolled conditions. A basic question arises: Given only a single training image of an object how can one determine whether a test image is of the same object taken under different conditions, or of a new object altogether?

This paper addresses this question for the case where variation in appearance is due to illumination. In particular, how does one look at intensity images of two completely different objects and determine that the difference between these images could not be due to lighting variation, but must indicate a difference in object identity or position? We show that given only two images, one cannot determine with certainty whether they arise from the same or different objects, for there always exists a continuous surface with Lambertian reflectance that could have produced both images. While this result is demonstrated using point sources, it of course applies to more general lighting models that include point sources as a subset. Thus when two images appear to come from different objects, this is because it is unlikely, not impossible (under a Lambertian reflectance model) that they come from the same object. Our challenge then is to justify and quantify this unlikeliness.

To do this, we must first characterize what we can determine about an object from two images. We do this for the case of an object with Lambertian reflectance and lighting due to two different known point sources. We show that from the ratio of the two images, we can determine three components of the Hessian matrix that characterizes the surface of this object. We then extrapolate from this result to the case where the object is illuminated by unknown light sources.

To develop a measure for determining whether two images arise from the same object, we examine properties of the ratio of the two images (the usefulness of this representation was previously pointed out in Wolff and Angelopoulou [30] and Fan and Wolff [10]). We show that in the case where the object is relatively simple the ratio of two images of the same object must be even simpler than either of the individual images (where, as an example, we define simplicity based on the complexity of the algebraic function needed to locally approximate the shape or image). However, the ratio of images produced by two different, but equally simple objects is more complex for generic shapes and lighting conditions. We then use these insights to derive a simple local measure of comparison. We show that these methods can be used to produce greater accuracy on a real task – face recognition under variable lighting conditions – than widely used existing methods.

Finally, we briefly consider the case of more general lighting conditions, when multiple light sources are present. We show the difficulty of extrapolating from a small number of training images to the entire set of images an object can produce. Specifically, Belhumeur and Kriegman [4] have shown that from as few as three images of an object, where each is produced by a single point light source, one can determine the *illumination cone* that describes the set of all images that object can produce with multiple light sources. We show that if the training images have multiple unknown light sources instead of point sources, it is not possible to determine the illumination cone exactly.

## 2 Background

In model-based approaches to object recognition, it has been assumed that one can construct a precise 3-D model of an object to use for recognition. This is suitable for some applications, but it has proven difficult to build accurate 3-D models using only images taken in uncontrolled circumstances. This also raises questions about the suitability of approaches based on 3-D models as explanations of human vision (e.g., Marr [18], Ullman [28]).

Another approach has been to describe 3-D objects in terms of their invariant, or quasi-invariant properties. Such descriptions of 3-D objects capture that portion of their structure that is apparent in all, or almost all images of the objects. For example, Biederman [6], based on earlier work in computer vision (e.g., Lowe [17]), proposed that the human visual system describes and retrieves 3-D objects based on non-accidental properties that can be detected in images, regardless of viewpoint. Others in computer vision have developed approaches to recognition based on invariants (for an overview, see [21]). Models of objects that are based on invariants can, by definition, be constructed from a single image of an object, although difficulties have also arisen in applying such approaches to general classes of objects ([7, 8, 20]).

Perhaps most related to the topics in this paper is the work on image "lightness" (e.g., Horn [15]). Here the image of an object is filtered in an attempt to remove or suppress the lighting effects in order to recover only the object's surface reflectance. In [15] this is done in a three step process of differentiation, thresholding, and integration. Yet this method is at best only locally reliable, as image noise and errors due to shadows are compounded in the integration process.

Partly for these reasons, *appearance-based* methods of recognition have also been explored. In these methods, an object is not described in terms of its 3-D properties, but rather in terms of the 2-D images that it produces. One approach to appearance-based recognition is to sample an object's possible images, and then to compare, in a lower dimensional image subspace, a novel image to the set of sampled images, using pattern recognition techniques such as nearest neighbors (e.g., [27, 16, 22]). This works well when the training images densely sample the space of images one hopes to recognize. In general, though, an object can produce so many different images that it is not clear how to sample them all.

Alternatively, some approaches attempt to predict the images an object can produce from a small number of training images (e.g., [29, 26, 19, 4]). This overcomes the difficulty of having to sample a great number of an object's images. However, while this may not require full recovery of an object's 3-D structure, it clearly requires a lot of knowledge of 3-D structure to predict all its possible images (e.g., [5]). Such information may not be available when an object is viewed in only a few prior images under uncontrolled conditions.

We can contrast these approaches with the one suggested in this paper in terms of the information they hope to extract from an image. The model-based and invariant-based approaches hope to derive intrinsic properties of the 3-D object structure. The second set of appearance-based approaches hope to extract a characterization of the set of images an object can produce, which may be almost as ambitious. The first appearance-based approaches we discussed, in contrast, extract relatively little information from each image. They essentially treat each image as an isolated point of information. New images are compared to these using a measure such as Euclidean distance in a dimensionally reduced image space. On the other hand, the approach proposed here seeks to make comparisons between images that are derived from the nature of the imaging variability. We think of an image as providing considerable information about what other images an object can produce, without necessarily providing any definite information about its 3-D structure.

## 3 Two Images Are Always Compatible

We are interested in comparing two images to determine whether they come from the same, unknown object, with the same pose, but under different illumination conditions. So we ask first: Is it ever the case that two images cannot come from the same object? We show that the answer to this question is no. In fact, even if we assume that the lighting in each scene is constrained to be a known point source at infinity, we can always construct an object in a fixed pose that is consistent with both images. We should point out that while our analysis in this section correctly handles shadowing, it does not account for interreflections.

We also show what aspects of a Lambertian object's structure can be determined from two images with point sources. These results suggest a direction we may take to gauge the likelihood that two images are produced by the same object.

We first assume that two images, $I$ and $J$, come from a Lambertian object lit by two known point sources at infinity, $s = (s^x, s^y, s^z)$ and $l = (l^x, l^y, l^z)$ respectively. If given these limitations (known sources and Lambertian reflectance), we can construct an object that is consistent with both images, we have poor prospects of ever telling with certainty that two images, no matter how different they may appear, could not come from the same object.

We assume that the object is viewed from the direction $(0, 0, -1)$, and therefore, that the depth of the surface can be written as $z = f(x, y)$. By writing $f$ in this form we are ensuring that $f$ describes an integrable surface, i.e., that the surface normals of $f$ are consistent with a true surface. Let the albedo of the object be written as a function $\alpha(x, y)$ also. Then, the surface normals of the object are $(f_x, f_y, 1)/\sqrt{f_x^2 + f_y^2 + 1}$, and we have the two equations

$$
\begin{aligned}
I(x, y) &= \alpha(x, y) \frac{-(s^x, s^y, s^z) \cdot (f_x, f_y, 1)}{\sqrt{f_x^2 + f_y^2 + 1}} \\
J(x, y) &= \alpha(x, y) \frac{-(l^x, l^y, l^z) \cdot (f_x, f_y, 1)}{\sqrt{f_x^2 + f_y^2 + 1}}.
\end{aligned}
$$

Our problem is to determine which functions $f$ and $\alpha$ may satisfy these equations.

As Wolff and Angelopoulou [30] and Fan and Wolff [10] have pointed out, we can deal with image pairs more simply by taking their ratio, since this causes the effect of albedo to cancel. However, they go on to use the ratio image for stereo matching and for reconstructing a surface from three

images, quite different purposes than ours. Nayar and Bolle [23] use the ratio of two regions in the same image for yet another purpose, to cancel the effects of lighting, under the assumption that the regions come from coplanar portions of the object.

Taking these ratios, and defining $r = I/J$, we have

$$\frac{I}{J} = r = \frac{s^x f_x + s^y f_y + s^z}{l^x f_x + l^y f_y + l^z}$$

which implies

$$(rl^x - s^x)f_x + (rl^y - s^y)f_y + (rl^z - s^z) = 0.$$

(Throughout this section, for simplicity we assume that neither image is zero at any point, so that the ratio is well-defined). Since $f$ is our only unknown, this is a first order, partial differential equation with variable coefficients. We can solve it using well-known methods (see, for example, Zauderer [31]). In brief, we may divide the image into characteristic curves. Along each characteristic curve, we change variables so that $f$ is a function of a single variable. Then we may find the value of $f$, along a characteristic curve, up to an unknown initial condition, by integrating along this curve.

As a simple example of this method, consider the case of $l = (0, 0, 1)$ and $s = (1, 0, 0)$. Then we have $f_x = r$. The characteristic curves in this case are horizontal lines across the image. The value of $f(x, y)$ is given by

$$f(x, y) = f(0, y) + \int_0^x r(w, y)dw \qquad (1)$$

where we have no obvious source of knowledge available to provide the initial condition $f(0, y)$. We denote this initial condition $f(0, y)$ as $g(y)$. This shows that we can recover the value of $f$ up to an unknown initial condition given by $g$. Note that we have

$$f_x = r \qquad f_y = g_y + \frac{\partial \int_0^x r\,dw}{\partial y}.$$

Thus, we can recover $f_x$ directly from the ratio image. We cannot recover $f_y$, however, since $g_y$ is unknown. Moreover, even if we did know $g$, any straight-forward recovery of $f_y$ from a real image would be extremely unstable, since we would have to numerically integrate $r$ along adjacent characteristic curves, and then take its derivative.

Taking further partial derivatives, we have

$$f_{xx} = r_x, \quad f_{xy} = f_{yx} = r_y, \quad f_{yy} = g_{yy} + \frac{\partial^2 \int_0^x r\,dw}{\partial y^2}.$$

Thus, we can recover $f_{xx}, f_{xy}, f_{yx}$ by taking derivatives of the ratio image, but again we cannot recover $f_{yy}$.

For this example, then, we see that we can use two images with known light sources to recover three components of the Hessian matrix of the surface of the object. Moreover, these equations always have a solution for $f$, which is given explicitly as Equation 1. We can also note that for these light sources, there are no shadows. The source $s$

casts no shadows on the surface because $f$ is monotonically increasing along each characteristic curve, since $I$, $J$ and therefore $f_x = r$ are, by nature, non-negative. Since the second light source is also the viewing direction, it cannot cast shadows on any visible object point. Note also that for any given $f$ that satisfies Equation 1, we can choose $\alpha$ to satisfy the equations given by the two images (such an $\alpha$ may have values greater than one. To avoid this, we must scale the intensity of each light source by an appropriate constant).

For other lighting conditions, we get similar results. In general, the slope of the characteristic curve is $(rl^y - s^y)/(rl^x - s^x)$. These characteristics are not straight lines, but vary their direction as a function of $r$. Assuming general lighting, so that there is no value of $r$ that satisfies both the equations $rl^y - s^y = rl^x - s^x = 0$, this direction is always unique and well-defined. In this case, the characteristic curves can never intersect, and so there always exists a surface that satisfies the ratio image's PDE (see [31]). Again, there is a whole family of these solutions, one for any function that provides an initial condition.

Similar issues have been considered in work on photometric stereo. However, the photometric stereo work addresses settings in which the reconstruction problem is not underconstrained. For example, Coleman and Jain [9] discuss recovery of structure for textured shapes with specularities, using four images. Onn and Bruckstein [24] show how to use integrability to recover structure from two images when the scene has a uniform albedo. And Fan and Wolff [10] consider recovery of structure and albedo from three images.

In contrast, we have shown that given two images of an object with unknown structure and albedo, there is always a large family of solutions. In fact, for any pair of point light sources there is a family of possible solutions. We have shown that given known light sources, we can determine two independent components of the Hessian of the surface at any position, but not the third. The direction in which we can determine these components may vary throughout the image, depending both on the light sources and the ratio image. Finally, even for unknown lighting conditions, the ratio image $r$ still provides information – albeit imprecise information – about the local nature of the object's surface.

## 4 Determining the Simplicity of Interpretations

We cannot tell that two images *must* come from different objects. We now turn instead to determining whether in some cases, explaining the images with a common object would require an unlikely coincidence. This approach to image interpretation has been applied to other vision problems by, for example, Rock [25], Lowe [17], and Freeman [11].

Specifically, we begin by showing that the ratio of two images from the same object is generally simpler than either of the individual images, while the ratio of images from different objects is generically more complex than either image.

As previously noted, the image of an object is

$$I = \alpha \frac{-(s^x, s^y, s^z) \cdot (f_x, f_y, 1)}{\sqrt{f_x^2 + f_y^2 + 1}} \qquad (2)$$

and the ratio of two images from the same object is

$$\frac{I}{J} = r = \frac{s^x f_x + s^y f_y + s^z}{l^x f_x + l^y f_y + l^z}. \tag{3}$$

However, suppose our second image $J'$ is of a different object, whose surface is described by the function $z = g(x, y)$, and whose albedo is described by the function $\beta(x, y)$. Then we have as the ratio image

$$\frac{I}{J'} = r' = \frac{s^x f_x + s^y f_y + s^z}{l^x g_x + l^y g_y + l^z} \left( \frac{\alpha \sqrt{g_x^2 + g_y^2 + 1}}{\beta \sqrt{f_x^2 + f_y^2 + 1}} \right). \tag{4}$$

In many instances Equation 3 describes a simpler ratio image than Equation 4, unless the extra multiplicative term

$$\left( \frac{\alpha \sqrt{f_x^2 + f_y^2 + 1}}{\beta \sqrt{g_x^2 + g_y^2 + 1}} \right)$$

happens to cancel the rest of the equation.

Consider the case in which local surface patches of $f$ and $g$, and their respective albedo functions $\alpha$ and $\beta$ are well approximated by second degree polynomials in $x$ and $y$. In this case, the ratio image $r$ from Equation 3 is a rational function with both a linear numerator and denominator. However, the ratio of images from different objects $r'$ (Equation 4) is the product of an algebraic function and a rational function with third degree polynomials in the denominator and numerator. The degree of a polynomial (or parameters in an algebraic expression) needed to approximate a function is often used as a measure of that function's simplicity. By that measure, in this case we see that the ratio of images from the same object is far simpler than the ratio of images from two different objects. This simplicity generally translates into simpler properties, such as fewer extrema and, in many cases, less overall variability in the ratio image.

Similar reasoning holds in other cases as well. Of special interest is the case in which the surfaces $f$ and $g$ are locally planar. In this case, the ratio of two images from the same object is constant. However, the ratio of images from different objects is only constant if their albedos are identical up to a scale factor, in which case the differences in the albedo patterns cannot in principal be discerned unless the magnitude of the lighting is known.

We can now relate these results to those in the previous section in a brief, intuitive form. We showed that the shape of an object lit by point sources could be derived by integrating the ratio image along its characteristic curves. These results suggest that we can attempt to measure the likelihood that two images come from the same object by measuring the simplicity of the ratio image.

## 5  Experiments Using a Simple Comparison Method

Our results suggest a number of ways of attempting to measure whether the difference between two images is due to a difference in lighting or in object structure. In this section,

we experiment with only the simplest of these on the task of recognizing faces under variable illumination conditions. A simple measure of the complexity of the ratio image is the integral of the magnitude of its squared gradient. This measures the smoothness of the ratio image. Such measures have often been used in vision, for example in interpolating surfaces by minimizing the curvature of the interpolation (some early methods are reviewed in [18]). This measure has the advantage of being local, and therefore, the analysis we have done assuming point light sources and low degree polynomial surfaces must hold only locally to apply.

Now notice that the squared magnitude of the gradient of the ratio image

$$\left\| \nabla \left( \frac{I}{J} \right) \right\|^2$$

has two significant disadvantages: first, it is asymmetric in $I$ and $J$, and second, it behaves poorly for regions of image $J$ which are in shadow. To correct for the asymmetry, we instead use the geometric mean

$$\left\| \nabla \left( \frac{I}{J} \right) \right\| \left\| \nabla \left( \frac{J}{I} \right) \right\|.$$

To correct for the behavior in the shadowed regions, we weight the measure by the $\min(I, J)$. Thus to compare images $I$ and $J$, we simply integrate this quantity over the image region to get

$$\int \int \min(I, J) \left\| \nabla \left( \frac{I}{J} \right) \right\| \left\| \nabla \left( \frac{J}{I} \right) \right\| dx dy.$$

With straight-forward algebraic manipulation, one can show the surprising similarity of this measure to a measure simply comparing image edges. In some sense, we have come full circle, using a Lambertian model for image formation to justify an edge-based measure of comparison. Yet, this measure has two important differences from simple edge-based matching: first, this measure does not make any hard decisions about the presence or absence of an edge, and second, it normalizes the response by the local image intensity. The measure proposed here is probably most closely related to the log filter described in Ballard and Brown [2] followed by a gradient highpass filter.

We experimented with this measure on the task of face recognition under variable illumination. We used a publicly available database of faces constructed by Hallinan [14]. From this database we used 450 images of 10 individuals. The images were divided into four subsets in which the lighting directions within the subsets were 15° (Subset 1), 30° (Subset 2), 45° (Subset 3), and 60° (Subset 4) from the camera's optical axis. Figure 1 shows images of one face from each of these subsets. Faces lit at a 15° angle (Subset 1) were used as a training set, and then tested using images in which the lighting had greater eccentricity. To discount the effects of improper alignment, each image was aligned manually and cropped as shown in Figure 1. Figure 2 shows the comparative performance of the method proposed here and three competing methods.

These results show that the squared magnitude of the gradient of the ratio image works dramatically better than simple correlation, or correlation after projecting onto the
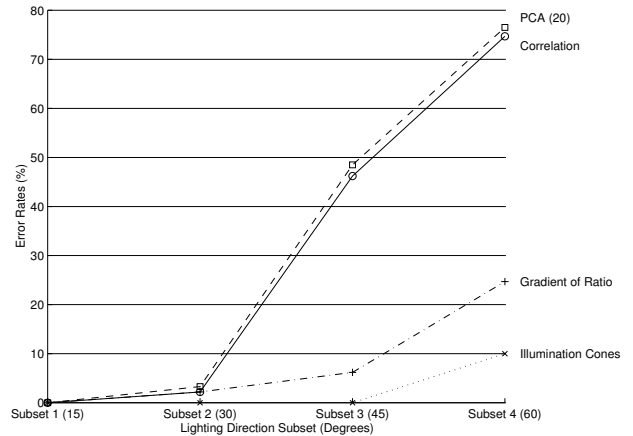
| 15° | 30° | 45° | 60° |

Figure 1: Pictures from the Harvard Face Database. The pictures are of the same individual lit by varying the direction of a point light source. The angle of the light source with the optical axis (15, 30, 45, and 60 degrees) is the same in each column.



| Method | Error Rate (%) | | | |
|---|---|---|---|---|
| | 15° Subset 1 | 30° Subset 2 | 45° Subset 3 | 60° Subset 4 |
| Correlation | 0.0 | 2.2 | 46.2 | 74.7 |
| PCA (20) | 0.0 | 3.3 | 48.5 | 76.5 |
| Gradient of Ratio | 0.0 | 2.2 | 6.2 | 24.7 |
| Illumination Cones | 0.0 | 0.0 | 0.0 | 10.0 |

Figure 2: A measure of comparison based on the magnitude of the gradient of the ratio image significantly outperforms both correlation and principal component analysis (PCA) on database of 450 faces of 10 individuals taken under extreme variation in lighting conditions.

twenty principal components of the training images (as described in [27]). However, it does not perform as well as the illumination cones method which builds a representation for the set of possible images from a small set of training images [4, 13]. (Adini, Moses, and Ullman [1] have also reported experiments on face recognition under variable illumination. We have not yet been able to compare these methods on their database.)

In comparing this measure to others, one should note that it does not attempt to combine information from a number of training images to build up as representation of a face, as do methods such as Fisherfaces [3], the linear subspace method [3], or the illumination cones method [13]. When enough training images are available to well characterize the entire set of images that a face can produce, we expect that one should achieve better performance with these methods, and our experiments seem to support this. However, our method simply compares a new image independently to each previously seen image. We feel that this type of approach, and indeed the results of our paper in general, are most suited to the situation in which one does not have enough prior information about an object to attempt to accurately characterize its possible appearances.

## 6 Difficulties with Representation

We have considered an approach to recognition in which two images are compared to judge whether they appear to come from the same object. An alternative approach is to use a number of training images to build a representation of the set of all images that an object can produce. This has been done effectively to account for viewpoint variation (e.g., [29]), and to account for lighting variation when the training images are each lit with a single point source ([19, 26, 4]). The illumination cones method [13] tested in the previous section is one such method. We now consider the difficulties in generating a representation of an object's possible images when the training images are taken under uncontrolled conditions, containing multiple light sources.

Belhumeur and Kriegman show how to build an exact representation of the images that a polyhedral object can produce when lit with multiple sources, which they call the *illumination cone*. Their method requires at least three images of the object each illuminated by a single light source. It is also evident from their results that given many images of the object, each of which contains multiple light sources, taking the convex combination of these images produces a subset of the illumination cone which can provide a good approximation to it. It is not clear, however, whether it is possible to build the illumination cone exactly using training images that contain multiple light sources. In this section we show that this is not possible. We take this as one indication of the potential difficulty of building a complete representation of an object's possible images using a small number of training images taken under uncontrolled viewing conditions. For such situations, it remains valuable to develop methods for directly comparing images.

We show that the illumination cone cannot be constructed from a set of images that contain multiple, unknown lights, by showing that a much simpler problem is not solvable. We show that even if one knows the 3-D structure of a convex-shaped Lambertian object exactly, one cannot in general determine the albedo of the object points exactly, even from a large set of images.

Let $p_j$ be a facet on a polyhedral object, let $J_j$ be the corresponding intensity produced in the image, $J$. Denote $p_j$'s surface normal pointing inward towards the object as $n_j$. Let the albedo at this object facet be $\alpha_j$. Assume that there are $m$ light sources, denoted $s_1, ..., s_m$. Then, with a Lambertian surface, we have

$$J_j = \sum_{k=1}^{m} max(0, \alpha_j n_j \cdot s_k).$$

Belhumeur and Kriegman [4] have shown that with this lighting model, the set of all images that a Lambertian object can produce forms a convex cone in the space, $\mathbb{R}^n$, of all images, where each coordinate of the space is the intensity value of a different pixel in the image. Let $\mathcal{C}$ denote the convex cone of images that could be produced by this object if the albedo of all its points were equal and set to unity; we refer to this as the constant albedo object. We can then describe the set of images the actual object can produce as follows. Let $A$ be a diagonal matrix, with diagonal entries $\alpha_1, ...\alpha_n$, denoting the albedo. Then the actual object can produce all images of the form $Ac$ such that $c \in \mathcal{C}$ is a column vector describing one of the images that the constant albedo object could produce. This tells us that if $J$ is a column vector whose entries are the pixel intensities of an image of the actual object, we must have

$$J = Ac, c \in \mathcal{C} \Rightarrow A^{-1}J \in \mathcal{C}.$$

$\mathcal{C}$ is a convex polytope that is defined by a set of $N$ bounding half-planes, which all pass through the origin. We can define each half-plane by a normal vector $H_i$, so that a point, $p$, is inside the half-plane when $p \cdot H_i \geq 0$. So

$$c \in \mathcal{C} \equiv c \cdot H_i \geq 0, i = 1, ..., N.$$

This tells us that

$$(A^{-1}J) \cdot H_i \geq 0, i = 1, ..., N.$$

This is a series of inequalities that are linear in the inverse of the object albedos, since the image $J$ and the illumination cone (i.e., the $H_i$) of the constant albedo object are both known. *This tells us that a single image of an object that has known surface normals constrains the albedos of the object to lie inside a convex polytope in the space of all possible inverse albedos.*

Suppose we have many images of the same object available. The true albedos of the object lie inside the intersection of a set of convex polytopes in albedo space. The intersection of these polytopes gets smaller and smaller as we have more images available, constraining the possible object albedos. However, the true albedos do not lie on the boundary of any of these convex polytopes unless a point

in the object has a light source lying in its tangent plane[1]. Hence, the intersection of these convex polytopes is still an open set in albedo space, and the albedos of the object are not uniquely determined. *On the other hand, given images from objects with the same structure but different albedos, these convex cones in inverse albedo space can be non-intersecting, revealing that the objects are different.*

These results illustrate the following point: it may be impossible to determine a complete representation of the images that an object can produce, using multiple images taken under uncontrolled conditions. However, we still may be able to tell whether a new image is consistent with one or more previous images we have seen. We have shown this to be true for the simple case of a convex object with known structure but unknown albedo.

The results in this section are closely related to Forsyth's [12] color constancy algorithm. That work dealt with a very different problem, that of determining the color of patches of a planar scene from a single image in which the spectrum of the illumination is unknown. However, our derivation is similar. In Forsyth's case the appearance of all possible color chips under a known light source plays the role that is played by the illumination cone of a known model in our case. For Forsyth, the appearance of each patch of uniform color in the scene plays the role that each image plays in our derivation. Forsyth uses these to constrain the unknown illuminant function in a scene in much the same way that we constrain the unknown albedo. The key difference is that for color constancy one derives a convex constraint from every different color in the scene, where in our case there is a comparable constraint produced by every image, so that many images may be required to narrow down the solution.

## 7  Conclusions

Model-based recognition methods have achieved considerable success when they have adequate prior knowledge to build a precise object model that captures its 3-D structure. However, in some applications, this prior knowledge is not available. It is also an open question whether human object recognition routinely functions with sufficient prior knowledge of objects to construct and use precise 3-D models.

For this reason, we argue that the information that an image provides about an object may be best thought of as information about what other images are likely, or unlikely to come from the same object. We have made this concrete for the case of illumination variation. We have shown that it can be difficult to exactly recover the properties of an object, such as its albedo, from images in which the lighting conditions are unknown. At the same time, we have shown that there may be considerable information about whether two images either do or do not come from the same object. Therefore we may be able to use previously seen images of an object to recognize it in new images under variable lighting conditions, even without using these prior images to perform any sort of explicit or implicit reconstruction of any object properties.

---

[1] Strictly speaking, this is true only when $\mathcal{C}$ has volume in $\mathbb{R}^n$. If it doesn't we must restate our argument to focus on only the subset of object points that have distinct surface normals. Our basic argument still holds, however.

We have also used these insights in a recognition system. We have shown that the simplicity of the ratio of two images provides a good indication of whether they come from the same object; we measure this simplicity by looking at the gradient of the ratio image. This approach is simple and local, but provides good results.

# References

[1] Y. Adini, Y. Moses, S. and Ullman, 1997. "Face Recognition: The Problem of Compensating for Changes in Illumination Direction," IEEE Trans. PAMI **19**(7):721–732.

[2] D. Ballard and C. Brown, 1982, *Computer Vision*, Prentice Hall, Englewood Cliffs, New Jersey.

[3] P. Belhumeur, J. Hespanha, and D. Kriegman, 1997. "Eigenfaces vs. Fisherfaces: Recognition Using Class Specific Linear Projection," IEEE Trans. PAMI **19**(7):711–720.

[4] P. Belhumeur and D. Kriegman, 1996. "What is the Set of Images of an Object Under All Possible Lighting Conditions?", *IEEE Conf. on Comp. Vis. and Pat. Rec.*:270–277.

[5] P. Belhumeur, D. Kriegman, and A. Yuille, 1997. "The Bas-Relief Ambiguity" *CVPR*:1060–1066.

[6] I. Biederman, 1985, "Human Image Understanding: Recent Research and a Theory," *Computer Graphics, Vision, and Image Processing*, (32):29-73.

[7] J. Burns, R. Weiss, and E. Riseman, 1992, "The Non-Existence of General-Case View-Invariants," *Geometric Invariance in Computer Vision*, edited by J. Mundy, and A. Zisserman, MIT Press, Cambridge.

[8] D. Clemens and D. Jacobs, 1991, "Space and Time Bounds on Model Indexing," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, **13**(10):1007-1018.

[9] E. Coleman and R. Jain, 1982, "Obtaining 3-Dimensional Shape of Textured and Specular Surfaces Using Four-Source Photometry," *CGIP* **18**(4):309–328.

[10] J. Fan and L. Wolff, 1997, "Surface Curvature and Shape Reconstruction from Unknown Multiple Illumination and Integrability," *Computer Vision and Image Understanding* **65**(2):347–359.

[11] W. Freeman, 1994, "The Generic Viewpoint Assumption in a Framework for Visual Perception," *Nature* **368**:542–545.

[12] D. Forsyth, 1990. "A Novel Algorithm for Color Constancy," *Int. J. of Comp. Vis.* **5**(1):5–36.

[13] A. Georghiades, D. Kriegman, and P. Belhumeur, 1998. "Illumination Cones for Recognition Under Variable Lighting: Faces", *IEEE Conf. CVPR*.

[14] P. Hallinan, 1994. "A Low-Dimensional Representation of Human Faces for Arbitrary Lighting Conditions," *IEEE Conf. CVPR*:995–999.

[15] B. Horn, 1974, "Determining Lightness from an Image", *CGIP*, **3**(4):277–299.

[16] M. Kirby, and L. Sirovich, 1990, "The application of the Karhunen-Loeve procedure for the characterization of human faces", *IEEE transactions on Pattern Analysis and Machine Intelligence*, **12**(1):103-108.

[17] D. Lowe, 1985, *Perceptual Organization and Visual Recognition*, Kluwer Academic Publishers, The Netherlands.

[18] D. Marr, 1982, *Vision*, W.H. Freeman and Company, San Francisco.

[19] Y. Moses, 1993. *Face recognition: generalization to novel images*, Ph.D. Thesis, Weizmann Institute of Science.

[20] Y. Moses and S. Ullman, 1992, "Limitations of Non Model-Based Recognition Schemes," *Second European Conference on Computer Vision*:820-828.

[21] J. Mundy and A. Zisserman (eds.), 1992, *Geometric Invariance in Computer Vision*, MIT Press, Cambridge.

[22] H. Murase and S. Nayar, 1995. Visual learning and recognition of 3D objects from appearance. *Int. Journal of Computer Vision*, **14**(1):5–25.

[23] S. Nayar and R. Bolle, forthcoming, "Reflectance Based Object Recognition," *Int. J. of Comp. Vis.*

[24] R. Onn, and F. Bruckstein, 1990, "Integrability Disambiguates Surface Recovery in Two-Image Photometric Stereo," *Int. J. of Comp. Vis.* **5**(1):105–113.

[25] I. Rock, 1983. *The Logic of Perception*, MIT Press, Cambridge, MA.

[26] A. Shashua, 1997. On Photometric Issues in 3D Visual Recognition from a Single 2D Image. *IJCV*, **21**(1/2):99–122.

[27] M. Turk, and A. Pentland, 1991. "Eigenfaces for Recognition", *Journal of Cognitive Neuroscience*, 3, 71-86.

[28] S. Ullman, 1989, "Aligning Pictorial Descriptions: An Approach to Object Recognition," *Cognition* **32**(3):193-254.

[29] S. Ullman and R. Basri, 1991, "Recognition by Linear Combinations of Models," *IEEE Trans. PAMI*, **13**(10):992-1007.

[30] L. Wolff and E. Angelopoulou, 1994, *Eur. Conf. on Comp. Vis.*:247–258.

[31] E. Zauderer, 1983. *Partial Differential Equations of Applied Mathematics*, John Wiley and Sons.