To appear in the Proceedings of CVPR98.

Comparing Images Under Variable Illumination

Peter Belhumeur
Yale University
New Haven, CT
belhumeur@yale.edu

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

Abstract

We consider the problem of determining whether two images come from different objects, or the same object under different illumination conditions. We show that this problem cannot be solved using hard constraints: even under a Lambertian reflectance assumption, there is always an object and a pair of lighting conditions consistent with any two images. However, there is probabilistic information available to help us solve this problem. For point sources, we show that the ratio of two images from the same object will be 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. Finally, we show how to incorporate these insights into an algorithm, presenting experiments on face recognition under variable illumination.

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 will be used to derive some definite information about the shape of an object that will accurately predict some properties of its appearance in new images; 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 extrapolating from a small set of training images. Yet, recognition systems do not always have the luxury of large training sets. A basic question arises from such approaches: 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 just due to lighting variation, but must indicate a difference in object identity? We first focus on this question for the case in which the illumination of local patches of the image can be modeled as coming from a point source at infinity. We will show that it is always possible to account for any pair of images with a single object and two different lighting conditions. While this result is demonstrated using point sources, it will of course apply to any more general lighting model that includes point sources as a subset. We conclude that 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 lighting due to two different known point sources. We show that in this case, there is always a family of objects consistent with the images, and that we can determine three components of the Hessian matrix that characterizes the surface of this object. We can extrapolate from this result to the set of objects that are consistent with the two images and unknown lighting conditions.

At this point, we will have shown that there always exist a family of objects that could produce any pair of images and derived some of the characteristics of these possible objects. We next show that in some cases, an object can account for two images only if we posit an unlikely set of coincidences. We can explicitly understand these coincidences by examining properties of the ratio of the two images (the usefulness of this representation was previously pointed out in Wolff and Angelopoulou[28] and Fan and Wolff[9]). We show that in the case where the object is relatively simple (can be approximated locally by a low order polynomial) the ratio of two images of the same object must be even simpler than either of the individual images. However, the ratio of images produced by two different, but equally simple objects will be, in general, more complex. We then use these insights to derive a very simple local filter to apply to each image, before comparing them. We will show that these methods can be used to produce greater accuracy on a real task, face recognition under variable lighting conditions, than have previously used methods.

Finally, we will briefly consider a case of more general lighting conditions, when multiple light sources are present. We will 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 [3] have shown that from as few as three images of an object, where each 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 will show that if the training images have multiple unknown light sources instead of point sources, it is not possible to exactly determine the illumination cone.

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[16], Ullman[26]). 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[5], based on earlier work in computer vision (e.g., Lowe[15]), 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 [19]). 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([6, 7, 18]).

Partly for these reasons, *appearance-based* methods of recognition have also been explored. In this approach, 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., [25, 14, 20]). This works well when the training images densely samples 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.

Alternately, some approaches analytically predict the images an object can produce from a small number of training images (e.g., [27, 24, 17, 3]). 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 a 3-D structure to predict all its possible images (e.g., [4]). 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 our own in terms of what information they hope to extract from an image. The model-based approaches hope to derive 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 very 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 simple image space. On the other hand, our approach 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, 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 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 will also show what aspects of object structure can be determined from two images with point sources. These results will suggest a direction we may take to gauge the likelihood that two images are produced by the same object.

So first, we 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, simple sources, Lambertian material) 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 will 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 will be consistent with a true surface. Let the albedo of the object be written as a function $\alpha(x,y)$ also. Then, for the surface normals of the object we have $\frac{(f_x, f_y, 1)}{\sqrt{f_x^2+f_y^2+1}}$ and we have the two equations

$$\begin{aligned} I &= \alpha \frac{-(s_x, s_y, s_z)\cdot(f_x, f_y, 1)}{\sqrt{f_x^2+f_y^2+1}} \\ J &= \alpha \frac{-(l_x, l_y, l_z)\cdot(f_x, f_y, 1)}{\sqrt{f_x^2+f_y^2+1}}. \end{aligned}$$

Our problem is to determine which functions f and α may satisfy these equations.

As Wolff and Angelopoulou[28] and Fan and Wolff[9] 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[21] use the ratio of two regions in the same image for a different 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 will 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[29]). 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 $l = (0,0,1), s = (1,0,0)$. Then we have $f_x = r$. The characteristic curves in this case are horizontal lines across the image. The value of $f(x_0, y_0)$ is given by

$$f(x_0, y_0) = f(0, y_0) + \int_0^{x_0} r(x, y_0)dx \qquad (1)$$

where we have no obvious source of knowledge available to provide the initial condition, $f(0, y_0)$. 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

$$f_x = r \quad f_y = g_y + \frac{\partial \int_0^x r dx}{\partial y}$$

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.

Similarly, we have

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

Similarly, 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 will be no cast shadows. The source s casts no shadows on the surface because the surface, f is monotonically increasing along each characteristic curve, since $f_x = r$ is always positive. Since the second light source is also the viewing direction, it cannot cast shadows on any visible object points. Note also that for any given f that satisfies equation 1, we can choose α to satisfy the equations given by the two images (such an α 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 $\frac{rl_y - s_y}{rl_x - s_x}$. These characteristics will not be straight lines, but will 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 will always be 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 [29]). Again, there will be a whole family of these solutions, one for any function that provides an initial condition. And, in general, we will be able to determine the partial derivative of the surface in the direction of the characteristic curve, but not orthogonal to it. Then, again, we will be able to determine three of the four components of the Hessian of the surface; those that involve taking a partial derivative in the direction of the characteristic.

Similar issues have been considered in work on photometric stereo. However, this work addresses settings in which the reconstruction problem is not underconstrained. For example, Coleman and Jain[8] discuss recovery of structure for textured shapes with specularities, using four images. Onn and Bruckstein[22] show how to use integrability to recover structure from two images when the scene has a uniform albedo. And Fan and Wolff[9] 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 will vary throughout the image, for lights in general position. Finally, we can see that even for unknown lighting conditions, the partial derivatives of the ratio image, r_x and r_y, provide a measure of the smoothness of the surface. They provide us with two of the three independent components of the surface's Hessian, although we do not know which components of the Hessian they provide without knowing the lighting conditions to tell us the characteristic curves. Therefore we may note that the magnitude of the gradient of the ratio image will provide us with some information about the magnitude of the curvature of the surface that produced it.

4 Determining the Simplicity of Interpretations

We cannot tell for sure that two images come from different objects. In this section, we turn instead to determining whether explaining the images with a common object would require an implausible coincidence. This approach to image interpretation has been applied to other vision problems by, for example, Rock[23], Lowe[15], and Freeman[10].

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

As previously noted, the image of an object is described by

$$I = \alpha \frac{-(s_x, s_y, s_z) \cdot (f_x, f_y, 1)}{\sqrt{f_x^2 + f_y^2 + 1}} \qquad (2)$$

and the ratio of two images from the same object is given by

$$\frac{I}{J} = r = \frac{s_x f_x + s_y f_y + s_z}{l_x f_x + l_y f_y + l_z} \qquad (3)$$

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

$$\frac{I}{J'} = r' = \frac{s_x f_x + s_y f_y + s_z}{l_x g_x + l_y g_y + l_z} \left(\frac{\alpha}{\beta} \frac{\sqrt{g_x^2 + g_y^2 + 1}}{\sqrt{f_x^2 + f_y^2 + 1}} \right) \qquad (4)$$

In many instances Equation 3 will describe a simpler ratio image than Equation 4, simpler even than the image (Equation 2), for many possible definitions of simplicity. For example, suppose that f and g are functions with smooth derivatives that do not vary rapidly. Then the ratio of two images from the same object (Equation 3) will also be smooth. However, the ratio of images from two different objects will only be equally smooth if the variability in the extra multiplicative term

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

is either small, or happens to cancel the variability in the rest of the equation.

To be more concrete, consider the case in which local surface patches of f and g are well approximated by second order polynomials in x and y. In this case, the image I in Equation 2 is a sixth order polynomial, with terms that are quadratic in α, I, and x and y. The ratio image, r, from Equation 3 is only second order; it is bilinear in r, x and y. The ratio of images from different objects, r', however (Equation 4) is eight order, with second order terms in r', α and β, and fourth order terms in x and y. The order of the polynomial needed to fit a function is often used as a measure of that functions simplicity. By that measure, in this case we see that the ratio of images from the same object is simpler than either image itself, and much simpler than the ratio of images from two different objects. This simplicity will generally translate into simpler properties, such as fewer local extrema, and in many cases into 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 will be constant. However, the ratio of images from different objects will only be 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 absolute 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 with 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 will experiment with only the simplest of these on the task of recognizing faces under variable illumination conditions. A very simple measure of the complexity of the ratio image is the integral of the magnitude of its gradient, squared. 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 [16]). Although simple, this measure has the advantage that it is very local. Therefore, the analysis we have done assuming point sources of light and low order polynomial surfaces must hold only very locally in the object to apply.

At the same time, since large portions of a face are roughly spherical, as we discuss in Section 6 it may be a good idea to filter each image's log with a high-pass filter. Now notice that the magnitude squared of the derivative of the ratio image in a particular direction, the x direction for example, is

$$\frac{\partial}{\partial x}\frac{I}{J} = \frac{(I_x J - J_x I)^2}{J^4}$$

If we take the sum of square differences between the magnitude of the partial derivatives of the logs of the images, we get

$$\left(\frac{\partial(logI)}{\partial x} - \frac{\partial(logJ)}{\partial x}\right)^2 = \left(\frac{I_x J - I J_x}{IJ}\right)^2$$

This is just the geometric mean of the partial derivative of $\frac{I}{J}$ and that of $\frac{J}{I}$. It is useful to take the geometric mean of these quantities to treat the two images symmetrically, and because if one image is generally dark, then it will produce large gradients when in the denominator of the ratio, and comparably small gradients when in the numerator.

This discussion suggests that we compare images by taking the sum-of-squared differences of the gradient of the log of the Gaussian of each image. This measure reflects the smoothness of the ratio of the images. At the same time, it may be computed extremely efficiently, because each image may be filtered separately with a single filter that combines the gradient and log operations with Gaussian smoothing. Then, the filtered images need only be compared for Euclidean distance.

Method	Error Rate (%)			
	15°	30°	45°	60°
Correlation	0.0	0.0	26.2	49.4
PCA (10 dimensions)	0.0	0.0	42.2	52.9
Mag. of Gradient	0.0	0.0	7.7	22.3
Mag. of Gradient of the Log	0.0	0.0	6.2	15.2

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.

We have experimented with this measure on the task of face recognition under variable illumination. We used a publicly available data base of faces constructed by Hallinan[13]. In these images, faces were lit from an angle of 15°, 30°, 45° or 60°. Figure 1 shows an example of one face lit from each of these angles. Faces lit at a 15° angle were used as a training set, and then tested using images in which the lighting had greater eccentricity. Belhumeur, Hespanha and Kriegman[2] report experiments comparing a number of different recognition methods on this data base. We have followed the same procedure that they describe, and report results for a few different methods in Table 5.

These results show that our new method works dramatically better than simple correlation, or correlation after projecting onto the ten principal components of the training images (as described in [25]). Our method also performs significantly better than simply correlating with the magnitude of the gradient, although it should be noted that that method also works much better than correlation of the intensities. Adini, Moses, and Ullman[1] have also reported experiments on face recognition under variable illumination. We have not yet been able to compare our methods to the results achieved on their data base.

In comparing our filter to others, one should note that it does not attempt to combine information from a number of training images to build up some representation of a face, as do some methods such as Eigenfaces, Fisherfaces or the linear subspace method described in [2]. 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 methods that attempt to build a more complete model of the face or its important features. Rather, our method simply compares a new image independently to each previously seen image. In this it is more comparable to simple correlation, or other methods that compare images directly after filtering. 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 fully characterize its possible appearances. In this case, we feel it is appropriate to apply methods that determine the likelihood that two images come from the same object.

Finally, we wish to stress that the filter we have applied to images before comparison is very simple. We certainly do not wish to claim that we have "invented" such a simple method. Rather, we feel that our contribution in this section is to point out that this particular simple filter has a close connection to the gradient of the ratio of the two images, and that for this reason it can serve as a measure of the simplicity of the ratio image; which we have seen reflects the likelihood that the images come from the same object under different lighting conditions.

6 The Effect of Multiple Light Sources

Up until now, we have assumed that lighting, at least locally, can be modeled as coming from a point source. We now briefly examine some of the effects of multiple light sources on a single object. We do this for the special case of an object that is roughly spherical and convex. This will be of interest for particular problems, such as face recognition. We still assume that the light sources are at infinity. This means that we can describe an scene's lighting by a function on the sphere (which we'll call the illumination sphere in this context), indicating the radiance of the lighting in any direction. Let h denote this function.

In this lighting model, all object surfaces with the same surface normal have the same intensity, scaled by their albedo. So we now consider how this lighting function determines the appearance of a sphere with variable albedo. This tells us how all possible surface normals will appear; we must then consider the actual pattern of these surface normals in the object separately.

The intensity of any point on the sphere is given by integrating the effects of all lights that are visible to that point. The effect of the light coming from each point on the illumination sphere is modulated by the cosine of its angle to that surface point. Consequently, the intensity of any point on a sphere will be

$$I(n) = \int_{l \in S : l \cdot n > 0} \alpha(n) h(l) (l \cdot n) dl$$

where n is the normal of a point on the sphere, $\alpha(n)$ is the albedo at that point, l is a unit vector indicating a direction of illumination (S is the illumination sphere), and $h(l)$ is a positive function indicating the intensity of the light source in the direction l. $l \cdot n$ is the cosine of the angle between the object's surface normal and the lighting direction.

Note that if we ignore albedo, this amounts to convolving the function h with the positive part of a 2-D cosine function. The complete image is the product of albedo and the result of this convolution. If we think of the albedo of the sphere as the signal, and non-uniform illumination patterns as noise, we can see that this "noise" is multiplicative, and of low frequency. This is closely related to an observation made by Haddon and Forsyth[12]. This suggests that to attenuate the effects of illumination variation we can apply a filter that is designed to remove such noise. This can be done, for example, by convolving the log of each image with a high pass filter. If the noise is multiplicative in the image, it is additive in the log of the image, and still of low frequency, so that a high-pass filter can remove it.

7 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 alternate 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 very effectively to account for viewpoint variation (e.g., [27]), and to account for lighting variation when the training images are each lit with a single point source ([17, 24, 3]). We now consider the question of how difficult it is to build 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 that each contain a single 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 will produce 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 will be valuable to develop methods of 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 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 α_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[3] 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, $\mathcal{R}^n$, 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 set to 1. 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 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 albedo-free object could produce. This tells us that if J is a column vector whose entries are the pixel intensities of an image of the 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 of the albedo-free 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 will lie inside the intersection of a set of convex polytopes in albedo space. The intersection of these polytopes will get smaller and smaller as we have more images available, constraining the possible object albedos. However, the true albedos will 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 will still be an open set in albedo space, and the albedos of the object will not be 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 unconstrained conditions. But 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.

It is important to note that results in this section are closely related to Forsyth's[11] 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

[1] Strictly speaking, this is true only when $\mathcal{C}$ has volume in 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.

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.

8 Conclusions

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 unconstrained images. At the same time, we have shown that there may be strong, though probabilistic, evidence that 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.

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 very simple and local, but provides good results.

References

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

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

[3] Belhumeur, P. and Kriegman, D., 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.

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

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

[6] Burns, J., 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.

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

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

[9] Fan, and Wolff, L., 1997, "Surface Curvature and Shape Reconstruction from Unknown Multiple Illumination and Integrability," *Computer Vision and Image Understanding*.

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

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

[12] Haddon, J. and Forsyth, D., 1998, "Shading Primitives: Finding Folds and Shallow Grooves," *IEEE ICCV* (forthcoming).

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

[14] Kirby, M. and Sirovich, L., 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.

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

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

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

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

[19] *Geometric Invariance in Computer Vision*, edited by J. Mundy, and A. Zisserman, MIT Press, Cambridge.

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

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

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

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

[24] Shashua, A., 1992. *Geometry and Photometry in 3D Visual Recognition*, MIT TR–1401.

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

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

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

[28] Wolff, L., and Angelopoulou, E., 1994, *Eur. Conf. on Comp. Vis.*:??.

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