

Reflectance Based Object Recognition

Shree K. Nayar

Department of Computer Science
Columbia University
New York, N.Y. 10027
nayar@cs.columbia.edu

Ruud M. Bolle

Exploratory Computer Vision Group
IBM T. J. Watson Research Center
Yorktown Heights, N.Y. 10598
bolle@watson.ibm.com

Abstract

Neighboring points on a smoothly curved surface have similar surface normals and illumination conditions. Therefore, their brightness values can be used to compute the ratio of their reflectance coefficients. Based on this observation, we develop an algorithm that estimates a reflectance ratio for each region in an image with respect to its background. The algorithm is efficient as it computes ratios for all image regions in just two raster scans. The region reflectance ratio represents a physical property that is invariant to illumination and imaging parameters. Several experiments are conducted to demonstrate the accuracy and robustness of ratio invariant.

The ratio invariant is used to recognize objects from a single brightness image of a scene. Object models are automatically acquired and represented using a hash table. Recognition and pose estimation algorithms are presented that use ratio estimates of scene regions as well as their geometric properties to index the hash table. The result is a hypothesis for the existence of an object in the image. This hypothesis is verified using the ratios and locations of other regions in the scene. This approach to recognition is effective for objects with printed characters and pictures. Recognition experiments are conducted on images with illumination variations, occlusions, and shadows. The paper is concluded with a discussion on the simultaneous use of reflectance and geometry for visual perception.

Index Terms: Object representation, physical properties, retinex theory, reflectance ratio, photometric invariant, region ratios, indexing, model acquisition, object recognition, pose estimation.

1 Introduction

Object recognition has been an active area of machine vision research for over two decades [Besl and Jain-1985, Chin and Dyer-1986]. The traditional approach has been to recover geometric features from images and then use these features to hypothesize and verify the existence of three-dimensional objects in the scene. Lines, curves, and vertices are examples of geometric features often used by recognition algorithms. During image formation, a three-dimensional scene is mapped onto a two-dimensional plane, causing a significant amount of information to be lost regarding the geometry of the scene. As a result, geometric features are not always adequate for robust recognition of objects.

Compared to the attention given to object geometry, little effort has been directed towards the explicit use of other scene properties for recognition. In addition to its shape, an object may be characterized by physical attributes such as reflectance, roughness, and material type (Figure 1). Clearly, the representation of an object based on these properties is meaningful only if the recognition system is able to compute them from images. In this

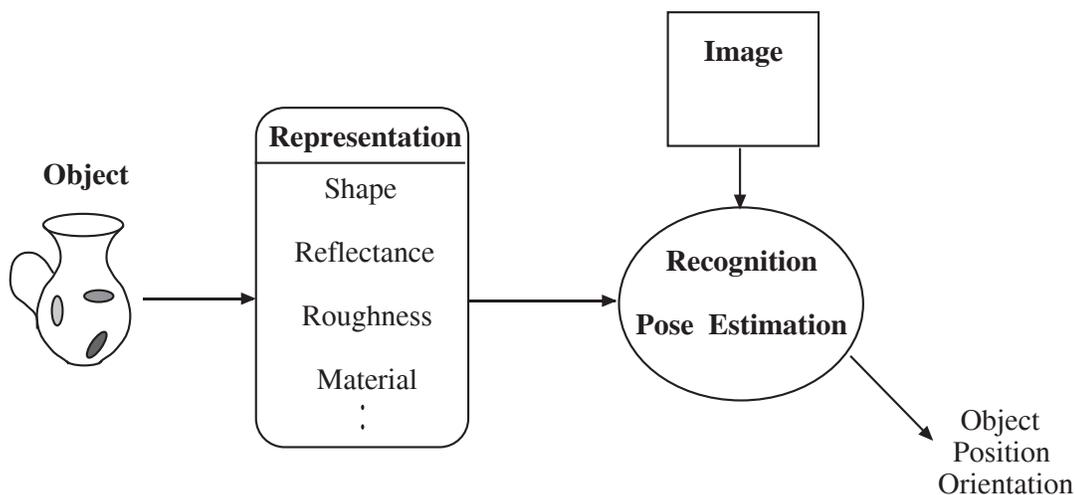


Figure 1: A physical approach to object representation and recognition.

paper, we present a method for computing the reflectance of regions in a scene, with respect to their backgrounds, from a single image. The result is a physical property of each scene region that is invariant to the shape of the object it lies on, as well as the intensity and direction of illumination. This photometric invariant, referred to as the *reflectance ratio*, provides valuable information for recognition tasks. Reflectance ratios of object regions and their spatial configurations are used to represent and recognize the object.

The problem of computing the reflectance of regions in a scene was first addressed by Land [Land-1964, Land-1983, Land-1986b]. Surface radiance, and hence image brightness, is the product of surface reflectance and illumination. Consequently, it is impossible to separate the contributions of reflectance and illumination at a single image point if the point is treated in isolation. Land constructed a set of clever experiments to show that humans are able to perceive the reflectance of scene regions even in the presence of non-uniform and unknown illumination. He developed the *retinex theory* that suggests computational steps for recovering the reflectance of scene regions. Though it is not possible to determine the absolute reflectance of regions, their relative reflectance (or “lightness”) can be estimated. The retinex theory is based on the assumption that the scene is subjected to smoothly varying illumination and consists of patches with constant reflectance. Under these assumptions, reflectance values change abruptly at region boundaries while illumination variations are small. It is therefore possible to filter out the effects of illumination and estimate reflectance. Later, Land and McCann [Land and McCann-1971] proposed a one-dimensional hardware implementation for computing lightness. Subsequently, Horn [Horn-1974] extended these ideas to two dimensions and presented several analog methods for implementing the lightness computation. Improvements to these implementations were suggested by Blake [Blake-1985].

The main idea underlying Land’s lightness computation is global consistency. The lightness value computed for any particular region must be consistent with those computed elsewhere in the image. However, realistic images include shadows, occlusions, and noise. Any one of these factors can cause a region boundary to be missed or erroneously detected. Such errors can greatly affect the lightness values computed for all regions in the image. For this reason, Land’s global method is not applicable to most real images. In [Land-1986a], Land suggests a “local designator” for computing lightness. This designator, inspired by Mach bands [Land-1986a], is effectively the brightness of an image point divided by a weighted average of surrounding brightness values within a fixed radius. Variants of this type of center-surround operator, based on the Laplacian of the Gaussian, are discussed by Horn [Horn-1986], Hurlbert [Hurlbert-1989], and Finlayson [Finlayson-1992]. Such operators treat all image points equally. This proves to be a disadvantage. Brightness values of edge pixels in discrete images are known to be sensitive to the exact location of the edge. Therefore, like most local operators, center-surround operators produce unreliable results in the vicinity of an edge. At the same time, edges are critical to the computation of reflectance estimates as they are sole representatives of abrupt reflectance changes. In addition, center-surround operators tend to have large masks that produce the undesirable effect of image blurring prior to reflectance estimation.

Here, we develop an alternative scheme for computing the ratio of the reflectance of a region to that of its background. To overcome the above problems faced by center-surround operators, the image is first segmented into regions of constant (but unknown) reflectance.

Next, a reflectance ratio is computed for each region with respect to its background using only points that lie close to the region’s boundary. Since all regions are pre-segmented, image points that lie on, or very close to, edges can be avoided. Furthermore, in this local approach, the reflectance ratio computed for any particular region is not affected by those computed for regions elsewhere in the image. In their analysis and experiments, Land and McCann [Land and McCann-1971] restricted themselves to planar (two-dimensional) scenes with patches of constant reflectance. These scenes are similar in appearance to the paintings of the Dutch artist Mondrian¹ and produce images such as the one shown in Figure 2(a)² In contrast, our derivation of the reflectance ratio is based on the analysis of regions on curved surfaces. Such regions are commonplace in realistic scenes like the one shown in Figure 2(b). In the case of curved surfaces, image brightness variations result from both illumination variations as well as surface normal changes. For curved surfaces, our reflectance ratio invariant is valid when a region and its background have the same distribution (scattering) function but different reflectance coefficients (albedo). We present an efficient algorithm that computes reflectance ratios of all scene regions in just two raster scans of an image.

The idea of using ratios of adjacent brightness values for recognition was also explored by Finlayson [Finlayson-1992]. Finlayson uses the ratios in the different channels of a color image to obtain measurements that are more robust to illumination variations. Histograms of the color ratios are computed for recognizing objects. Finlayson argues that histograms of ratios are more robust to illumination variations than the histograms of raw color measurements used by Swain [Swain and Ballard-1991]. Recently, Strickler [Strickler-1992] suggested the use of color ratios along boundaries in an image as well as the lengths of boundaries for recognition. The above approaches are effective in two-dimensional recognition problems where colorful objects are only subjected to rotations in the plane of the image. Further, only limited amounts of occlusion can be tolerated. For recognition of three-dimensional objects, color histograms are of limited use as they do not preserve detailed geometrical information; given a color histogram of an object one cannot infer the position and orientation of the object with respect to the sensor.

Our goal is to achieve not just object identification but also accurate object location and pose estimates that are necessary for an intelligent system to actively interact with its environment. In this paper, we use the reflectance ratio invariant for recognition and pose estimation of objects from a single image. This approach is very effective in the case of man-made objects that have printed characters and pictures. Each object is assumed to have a set of regions, each with constant reflectance. The reflectance ratio and center of each

¹Actually, they are closer in appearance to the paintings of van Doesburg.

²The Mondrian world has also been extensively used by researchers interested in color constancy; the problem of determining the color of objects when the color of illumination is unknown. References to work in this area can be found in [Healey *et al.*-1992].



(a)



(b)

Figure 2: Images of (a) a Mondrian scene; and (b) a realistic scene with three-dimensional objects.

region are used to represent the object. We address both two-dimensional as well as three-dimensional recognition problems. In both cases, techniques are presented for automatically acquiring object models. The recognition algorithms are based on the geometric hashing technique [Lambdan and Wolfson-1988]. The indices [Califano and Mohan-1991] of the hash table are the reflectance ratios of three regions on an object as well as their relative positions. The entries of the table include an object identifier followed by reflectance ratios of other regions on the object and their relative positions. The indices provide a hypothesis for the object, while the entries are used to verify the hypothesis and compute object pose.

Two sets of experimental results are reported. The first set demonstrates the invariance of reflectance ratios to imaging parameters (viewing direction, aperture setting, magnification, and defocus) and illumination parameters (source direction, and number of sources). The next set is on recognition and pose estimation in the presence of varying illumination, severe occlusions, and shadows. The results reported here represent an initial attempt at using physical properties (such as reflectance) for high-level perception tasks (such as recognition). We conclude with a discussion on how this paradigm can be advanced further.

2 Reflectance Ratio: A Photometric Invariant

The reflectance of a surface depends on its roughness and material properties. In general, incident light is scattered by a surface in different directions. This distribution of reflected light can be described as a function of the angle of incidence, the angle of emittance, and the wavelength of the incident light. Consider an infinitesimal surface patch with normal \mathbf{n} , illuminated with monochromatic light of wavelength λ from the direction \mathbf{s} , and viewed from the direction \mathbf{v} . The reflectance of the surface element can be expressed as:

$$r(\mathbf{s}, \mathbf{v}, \mathbf{n}, \lambda) \quad (1)$$

Now consider an image of the surface patch. If the spectral distribution of the incident light is $e(\lambda)$ and the spectral response of the sensor is $s(\lambda)$, the image brightness value produced by the sensor is:

$$I = \int s(\lambda) e(\lambda) r(\mathbf{s}, \mathbf{v}, \mathbf{n}, \lambda) d\lambda \quad (2)$$

For the purpose of discussion, let us assume the surface patch is illuminated by “white” light and the spectral response of the sensor is constant within the visible-light spectrum, then $s(\lambda) = s$ and $e(\lambda) = e$. We have:

$$I = s e \rho R(\mathbf{s}, \mathbf{v}, \mathbf{n}) \quad (3)$$

where $\rho R(\mathbf{s}, \mathbf{v}, \mathbf{n})$ is the integral of $r(\mathbf{s}, \mathbf{v}, \mathbf{n}, \lambda)$ over the visible-light spectrum. We have decomposed the result into $R(\cdot)$ which represents the dependence of surface reflectance on the geometry of illumination and sensing, and ρ which may be interpreted as the fraction of the incident light that is reflected in all directions by the surface. Incident light that is not reflected by the surface is absorbed and/or transmitted through the surface. Two surfaces with the same distribution function $R(\cdot)$ can have different reflectance coefficients ρ .

As a result of the white-light assumption, the reflectance coefficient ρ is independent of wavelength. This enables us to represent the reflectance of the surface element with a single constant. The same can be achieved by using an alternative approach which does not require making assumptions about the spectral distribution of the incident light and the spectral response of the sensor. Consider a narrow-band filter with spectral response $f(\lambda)$, placed in front of the sensor. Image brightness is then:

$$I = \int f(\lambda) s(\lambda) e(\lambda) r(\mathbf{s}, \mathbf{v}, \mathbf{n}, \lambda) d\lambda \quad (4)$$

Since the filter is narrow-band, it essentially passes a single wavelength λ' of reflected light. Its spectral response can therefore be expressed as:

$$f(\lambda) = \delta(\lambda' - \lambda) \quad (5)$$

The image brightness measured with such a filter is:

$$I = s' e' r(\mathbf{s}, \mathbf{v}, \mathbf{n}, \lambda') \quad (6)$$

where $s' = s(\lambda')$ and $e' = e(\lambda')$. Once again, the reflectance function can be decomposed into a scattering function and a reflectance coefficient:

$$I = s' e' \rho' R'(\mathbf{s}, \mathbf{v}, \mathbf{n}) \quad (7)$$

In this case, $R'(\cdot)$ represents the distribution of reflected light for a particular wavelength of incident light. On the other hand, for white-light illumination, $R(\cdot)$ represents the distribution computed as an average over the entire visible-light spectrum. However, the individual terms in both (3) and (7) represent similar effects. Since we do not use narrow-band filters in our experiments, we will use the following expression for image brightness in our discussion:

$$I = k \rho R(\mathbf{s}, \mathbf{v}, \mathbf{n}) \quad (8)$$

The constant $k = s.e$ accounts for the brightness of the light source and the response of the sensor. The exact functional form of $R(\mathbf{s}, \mathbf{v}, \mathbf{n})$ is determined to a great extent by the microscopic structure of the surface; generally $R(\cdot)$ includes a diffuse component and a specular component [Nayar *et al.*-1991b]. Once again, the reflection coefficient ρ is the fraction of incident light that is reflected by the surface. It represents the reflective power of the surface and is sometimes referred to as surface albedo.

2.1 Reflectance Ratio of Neighboring Points

Consider two *neighboring* points on a surface (Figure 3). For a smooth continuous surface, the points may be assumed to have the same surface normal vectors ($\mathbf{n}_1 \approx \mathbf{n}_2$). Further, the points have the same source and sensor directions. Hence, the brightness values, I_1 and I_2 , of the two points may be written as:

$$I_1 = k \rho_1 R_1(\mathbf{s}, \mathbf{v}, \mathbf{n}) \quad (9)$$

$$I_2 = k \rho_2 R_2(\mathbf{s}, \mathbf{v}, \mathbf{n}) \quad (10)$$

The main assumption made in computing the reflectance ratio is that the two points have the same scattering functions ($R_1 = R_2 = R$) but their reflectance coefficient ρ_1 and ρ_2 may differ. Then, the image brightness values produced by the points are:

$$I_1 = k \rho_1 R(\mathbf{s}, \mathbf{v}, \mathbf{n}) \quad (11)$$

$$I_2 = k \rho_2 R(\mathbf{s}, \mathbf{v}, \mathbf{n})$$

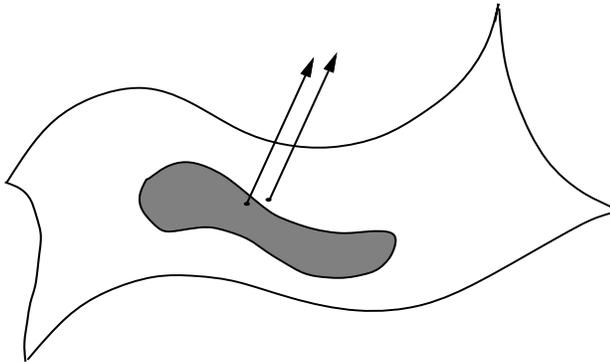


Figure 3: Neighboring points on a surface.

The ratio of the reflectance coefficients of the points is:

$$p = I_1/I_2 = \rho_1/\rho_2 \quad (12)$$

Note that p is independent of the reflectance function, illumination direction and intensity, and the surface normal of the points. It is a photometric invariant that is easy to compute and does not vary with the position and orientation of the surface with respect to the sensor and the source. Further, it represents an intrinsic surface property that can be effectively used for visual perception..

We have assumed that the scene is illuminated by a single light source. Now consider the same scene illuminated by several light sources. The brightness of any point can be written as:

$$I = \rho [k_1 R(\mathbf{s}_1, \mathbf{v}, \mathbf{n}) + k_2 R(\mathbf{s}_2, \mathbf{v}, \mathbf{n}) + \dots + k_n R(\mathbf{s}_n, \mathbf{v}, \mathbf{n})] \quad (13)$$

where $\mathbf{s}_1, \mathbf{s}_2, \dots, \mathbf{s}_n$ are the directions of the n sources that are visible to the surface point under consideration and k_1, k_2, \dots, k_n are proportional to the brightness of the n sources. Since the reflectance ratio is computed using neighboring points, it can be assumed that both points are illuminated by the same set of sources. Then, from (12) and (13) we see that the reflectance ratio p is unaffected by the presence of multiple light sources.

Note that by definition p is unbounded; if the second surface point is black, $I_2 = 0$, then $p = \infty$. From a computational perspective, this poses problems. Hence, we use a different definition for p to make it a well-behaved function of the reflectance coefficients ρ_1 and ρ_2 :

$$p = (I_1 - I_2)/(I_1 + I_2) = (\rho_1 - \rho_2)/(\rho_1 + \rho_2) \quad (14)$$

Now, we have $-1 \leq p \leq 1$. In the following section, we discuss the effects of reflectance mechanisms, colored illumination, and interreflections on the above ratio invariant.

2.2 Reflectance Mechanisms and the Ratio Invariant

We have shown that the ratio invariant is valid when the two neighboring points have the same scattering function. It is worthwhile exploring when this situation arises in practice. To this end, we briefly discuss reflectance mechanisms that are exhibited by large classes of real surfaces. Reflection of light from surfaces can be classified into two broad categories: diffuse (body) and specular (surface). The diffuse component results from light rays penetrating the surface and re-emerging at the surface after multiple reflections and refractions inside the surface medium. This component is distributed about the surface normal in a wide range of angles, giving the surface a matte appearance. The specular component, on the other hand, is a surface phenomenon. Light rays incident on the surface are reflected such that the angle of reflection equals the angle of incidence. If the specular surface is rough, the light rays are scattered in a lobe around the specular direction. Except for extreme cases of surface roughness, the specular component is concentrated in a small range of angles [Nayar *et al.*-1991b].

If the image brightness values of the neighboring pixels include both diffuse and specular components, it is improbable that their ratio will produce a meaningful invariant. Since the two components arise from different mechanisms, it is unlikely that two surface elements, with both components, would have exactly the same scattering functions and yet different reflectance coefficients [Torrance and Sparrow-1967]. This does not, however, imply that the ratio invariant is invalid if the surface has a non-zero specular coefficient. We simply assume that the image brightness values of the two neighboring points do not have a significant specular component; i.e., the points do not reflect specularly in the sensor direction and hence the scattering functions are determined solely by the diffuse components. Under this assumption, the ratio invariant will not be useful in the highlight regions of the image³. This leaves us with the diffuse component.

A large class of real surfaces have a diffuse component that is Lambertian-like; the diffuse component does not vary much with the viewing direction. The brightness of a Lambertian [Horn-1986] surface element may be written as $I = k \rho \mathbf{n} \cdot \mathbf{s}$ where ρ , \mathbf{n} , \mathbf{s} are the reflectance coefficient, normal vector, and source direction, respectively. Note that here the scattering function is $R = \mathbf{n} \cdot \mathbf{s}$ and the ratio given by (14) is an invariant. In practice, points lying on matte regions with different shades (or colors) would fall in

³Several algorithms have been proposed for separating the diffuse and specular components of images (see [Nayar *et al.*-1993] for references). In theory, one of these algorithms can be used to remove specularities from an image before reflectance ratios are computed. However, separation algorithms are still in the research stage and as yet do not guarantee accurate removal of specularities for complex scenes.

this category. In the case of surfaces with macro-scopic roughness, the diffuse component can deviate from the Lambertian model as recently shown by Oren and Nayar [Oren and Nayar-1993]. In such cases, image brightness depends not only on the source direction but also the viewing direction. The ratio invariant holds well for rough diffuse surfaces if the neighboring points lie on regions with different reflectance coefficients but the same roughness. This situation arises in the case of paints that have different colorants but similar surface structure. The above arguments in favor of the ratio invariant are valid also for multiple source illumination and extended source illumination [Nayar *et al.*-1990], since the ratio is computed using neighboring points that may be assumed to have the same illumination conditions. Also note that if a narrow-band filter is used at the sensor end, the reflectance ratio of neighboring points remains invariant even in the presence of interreflections (colored or not) [Koenderink and van Doorn-1983] [Forsyth and Zisserman-1989] [Nayar *et al.*-1991a]; we are only concerned with a single wavelength and the exact spectral distribution of incident light is not relevant.

For colored surfaces, the reflectance coefficients of the neighboring points may vary with the wavelength of incident light. Hence, for any two neighboring points, multiple reflectance ratios can be computed using a set of narrow-band filters each passing through it a different wavelength. If, for instance, we use three filters that pass wavelengths close to the ones humans perceive as red, green, and blue colors, respectively, we obtain three ratio estimates:

$$\mathbf{p} = [p^r, p^g, p^b]^T \quad (15)$$

This ratio vector is invariant to the spectral distribution of the illuminant since we have used narrow-band filters. Multiple ratios clearly provide more information regarding the relative reflectance of neighboring points, as observed by Finlayson [Finlayson-1992].

3 Reflectance Ratio of a Region

Hitherto, we have focused on two neighboring points. We now consider a surface region S that has constant reflectance coefficient ρ_1 and is surrounded by a background region with constant reflectance coefficient ρ_2 . We are interested in computing the reflectance ratio $P(S)$ of the region with respect to its background. The image brightness of the entire region cannot be assumed constant for two reasons. First, the surface may be curved and hence the surface normal can vary substantially over the region. Second, while the illumination may be assumed to be locally constant, it may vary over the region. These factors can cause brightness variations, or shading, over the region and its background as well. Figure 4(a) shows the image of a curved region and Figure 4(b) shows image brightness values varying along the boundary (white line) of the curved region. The reflectance ratio can be accurately estimated using neighboring (or nearby) points that lie on either side of the boundary between the region and the background. Figure 4(c) shows reflectance

ratios computed along the boundary of the region. Note that while image brightness varies along the boundary, the ratio remains nearly constant. A robust estimate of the region's reflectance ratio can be obtained as an average of the ratios computed along its boundary. The region ratio is also a photometric invariant; it is independent of the shape of the surface and the illumination conditions. It is computed using a single image of the scene and provides a useful intrinsic property.

Before proceeding any further, it is worthwhile comparing our approach with Land's retinex theory [Land-1964]. Our objective is similar to that of Land's, that is, to compute the reflectance ratios of regions in a scene from variations in image brightness across region boundaries. However, Land's analysis of brightness variations does not include the problem of curved surfaces. The scene is assumed to be planar with matte patches of constant reflectance. In our development of the reflectance ratio we have studied the effects of surface orientation on image brightness. Our method for computing the reflectance ratio of a region is also different from that proposed by the Land. The retinex theory relies on global consistency. If the reflectance ratio between two regions is inaccurately estimated or if the boundary between two regions is not detected, the reflectance values computed for all other regions are inaccurate. Real images include shadows, occlusions, and noise. These factors can cause a region to be merged with another, or region boundaries to be detected where none exist. Consequently, Land's global approach is generally not applicable to real-world scenes.

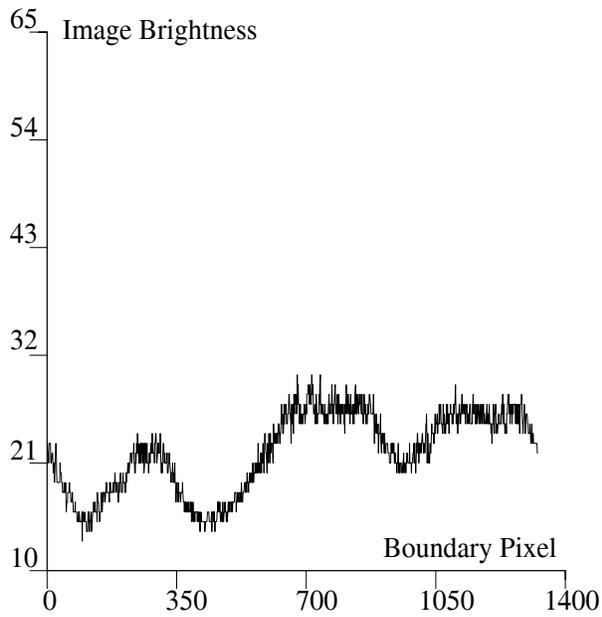
We adopt a more local approach. First, the scene is segmented into regions that are assumed to have different reflectance coefficients. Due to specularities, shadows, occlusions, and image noise, some regions may be missed and others may be erroneously created. However, these errors do not affect reflectance ratios computed for other valid regions in the scene. The extension of the reflectance ratio analysis to curved surfaces and the pre-segmented approach to the computation of region ratios are the two key differences between Land's lightness computation and our approach. We will see that these differences result in substantial improvements in the robustness of computed ratios.

4 Computing Region Ratios

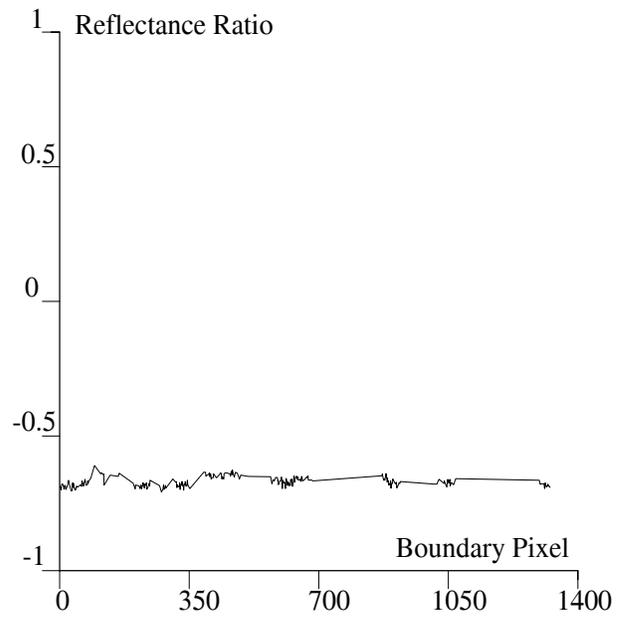
In this section, we present an algorithm that computes reflectance ratios of scene regions. Details of this algorithm can be found in [Nayar and Bolle-1993]. First, a sequential labeling algorithm is used to segment the image into connected regions. The second phase involves the computation of a reflectance ratio for each of the segmented regions. The algorithm is computationally efficient in that reflectance ratios of all scene regions are computed in just two raster scans of the image.



(a)



(b)



(c)

Figure 4: (a) Image of a curved region; (b) image brightness values along the region boundary; (c) reflectance ratios computed along the boundary. The ratios are nearly constant while the brightness values vary.

4.1 Sequential Labeling Algorithm

Sequential labeling is a well-known technique for efficient segmentation of images [Rosenfeld and Pfaltz-1966, Lumia *et al.*-1983, Horn-1986]. It has been widely used in the context of binary images [Nagy-1969, Gray-1971] where it is relatively straightforward to determine if two image pixels are “connected.” Algorithms have also been proposed that use near equal brightness values to determine the similarity between pixels in gray-scale images [Ballard and Brown-1982]. Here, we use the ratio invariant as a measure of similarity between two neighboring pixels. Let $p(A, B)$ denote the reflectance ratio $(\rho_A - \rho_B)/(\rho_A + \rho_B)$ of two neighboring pixels A and B . The pixels A and B are considered to be connected if $|p(A, B)| < T$, where T is a threshold value close to zero. A non-zero threshold is selected to account for brightness variations that result from image noise and surface shading effects. The connectivity between two pixels is defined as:

$$c(A, B) = \begin{cases} 1 & \text{if } |p(A, B)| < T \\ 0 & \text{otherwise} \end{cases} \quad (16)$$

The sequential labeling algorithm proceeds as follows. The image is examined in a raster scan fashion (left to right and top to bottom). The label of pixel A is determined by the labels of three of its neighbors; pixel B to its left, pixel C above it, and pixel D diagonal to it.

D	C
B	A

Note that for a raster scan these three neighboring pixels have already been labeled. If pixel A is connected to either B or C (not both) then it is assigned the same label as the pixel it is connected to. Else, if A is connected to both pixels B and C and pixels B and C have equal labels, then A is assigned the same label. An interesting situation arises when A is connected to both B and C and these two pixels have different labels [Horn-1986]. In this case, we assign A the label of either B or C and record the fact that the labels of B and C are “equivalent.” If none of the above cases occur and we find that A is connected to D , then we assign A the same label as D . Finally, if A is not connected to any of its neighbors, a new label is created. This labeling algorithm is based on an asymmetric 6 connectivity mask used to overcome problems associated with 4 and 8 connectivity masks (see [Horn-1986] for details).

Using this algorithm, the complete image can be segmented in a single raster scan. Following the raster scan, equivalences between different labels can be resolved such that all equivalent labels are represented by a single label. This information can either be stored as a table for future use or the image can be relabeled to account for the equivalences. A minor addition can be made to the sequential labeling algorithm so that the areas and centroids of all the labeled regions are also obtained.

4.2 Algorithm for Computing Reflectance Ratios

Sequential labeling provides a set of image regions assumed to correspond to surface regions. The following is an example one-dimensional image of a region and the result of sequential labeling:

<i>Brightness:</i>	35	37	39	41	64	77	85	87	89	89	91	92	94	96
<i>Label:</i>	1	1	1	1	2	3	4	4	4	4	4	4	4	4

Here, label 4 corresponds to a region and label 1 represents its background. Though there are smooth brightness variations within the region and the background, the labeling is robust. This is because local brightness variations within the region, or background, are small and reflectance ratios for connected pixels are close to zero.

In digital images, the edges between a region and its background are blurred for two reasons. First, the image has a finite resolution, causing the physical edge to lie within a pixel. The brightness value of the pixel therefore is a weighted average of the brightness values of the region and the background at the boundary [Horn-1986]. Second, every optical system is characterized by a blur function; due to imperfect imaging optics, every point in the scene is projected onto a small patch (not a point) on the image sensor [Born and Wolf-1965]. For these reasons, pixels on and around the edge will be assigned labels different from the region and the background.

As mentioned earlier, the sequential labeling algorithm also provides the area of each labeled region. Small regions that result from points that lie on boundaries in the scene can be ignored by using a threshold. We focus only on larger regions that are referred to as *valid* regions. Reflectance ratios for all valid regions can be computed in a single raster scan of the image. During this final raster scan, attention is given only to those pixels that lie in valid regions. If a pixel does lie in such a region, we first determine if it lies on the boundary of the region. Consider the pixel X and its four closest neighbors A , B , C , and D .

				K				
				A				
N			D	X	B			L
				C				
				M				

If X lies inside a region, it and its four neighbors have the same label. If however X lies on a region boundary, one or more of its neighbors must have different labels. Assume that the neighbor A has a different label from that of X . We examine the pixel K that lies at a distance d from X in the direction of A . We check to see if K lies in a valid region, i.e., make sure that K is not an edge pixel. If it does lie in a valid region, it is assumed to lie on the background of the region that pixel X represents. The distance d used to find the background pixel must be large enough to avoid edge pixels with unpredictable intensities and at the same time small enough to satisfy the condition that pixels X and K have near equal normals and illumination conditions. In our implementation, parameter d is selected by the user and is usually between two and five pixels in length. If the above conditions are satisfied, a reflectance ratio estimate is obtained as:

$$p_i(\text{Label}(X)) = p(X, K) = (I_X - I_K)/(I_X + I_K) \quad (17)$$

This is the i th ratio estimate computed for the region that contains X . This process is repeated for all neighbors of X whose labels differ from that of X . During the raster scan, a list of computed reflectance ratios is maintained for each valid region. After all image pixels are examined, the reflectance ratio of a region S is computed as the average of all ratios in its list:

$$P(S) = 1/N \sum_{i=1}^N p_i(S) \quad (18)$$

where N represents the total number of ratio estimates obtained for S .

Generally, N is not equal to the perimeter of the region for two reasons. First, each boundary pixel may produce more than one ratio estimate since it has four neighbors. Second, a boundary pixel may not produce any ratio estimates because it is surrounded by edge pixels that belong to invalid regions. A confidence measure for the ratio $P(S)$ is

defined as:

$$\gamma(S) = N^2/A(S) \tag{19}$$

where $A(S)$ is the area of S . This confidence may be used as a measure of the accuracy of the computed reflectance ratio. If $\gamma(S)$ is small, few ratios have contributed to the final estimate and it may be unreliable for recognition purposes. Recently, Nicol [Nicol-1993] suggested a systolic architecture that performs sequential labeling of multi-valued images in real time. This architecture may be modified to facilitate real time computation of region ratios and centroids from an image.

4.3 Multiple Backgrounds and Polyhedra

In the above discussion, we started by selecting a region and its background. Note that a reflectance ratio may be computed for the background as well, in which case, the region and the background reverse roles; the region is the background and vice versa. We also assumed that a region and its background have constant reflectance coefficients. In practice this assumption can be relaxed; a region of constant reflectance may be surrounded by several regions with different reflectance coefficients. The reflectance ratio computed for the region is again an average of the ratios computed along its entire boundary. In this case, however, the ratio can vary with viewing direction since the fraction of the region boundary shared with any particular background region depends on the viewing direction.

An interesting problem arises when the scene includes polyhedral objects. An insightful discussion on the interpretation of polyhedral scenes is presented by Sinha and Adelson [Sinha and Adelson-1993]. Here we are interested in polyhedral faces from the perspective of ratio computation. Each planar face on a polyhedron (Figure 5(a)) is expected to have near constant brightness, causing the sequential labeling algorithm to label it as a region. Since we do not have prior knowledge of the scene’s geometry, each face is treated as a region with constant reflectance and a ratio estimate is computed for it. Each polyhedral face is surrounded by faces with different surface normals. It is unlikely that all neighboring faces of any given face will produce the same brightness value, leading the ratio algorithm to interpret the face as a surface region with a single background. For the purpose of discussion, assume that the polyhedron shown in Figure 5(a) is Lambertian in reflectance with uniform reflectance coefficient ρ . Further, assume that the object is illuminated by a single light source in the direction \mathbf{s} . Then, the three neighboring faces of face A produce brightness values:

$$I_1 = k \rho \mathbf{n}_1 \cdot \mathbf{s}, \quad I_2 = k \rho \mathbf{n}_2 \cdot \mathbf{s}, \quad I_3 = k \rho \mathbf{n}_3 \cdot \mathbf{s} \tag{20}$$

The three faces produce the same image brightness when $\mathbf{n}_1 \cdot \mathbf{s} = \mathbf{n}_2 \cdot \mathbf{s} = \mathbf{n}_3 \cdot \mathbf{s} = I_1/k\rho$. This yields a constraint on the three normals that can be illustrated on the unit sphere (Figure 5(b)). For the three neighboring faces to produce the same brightness their

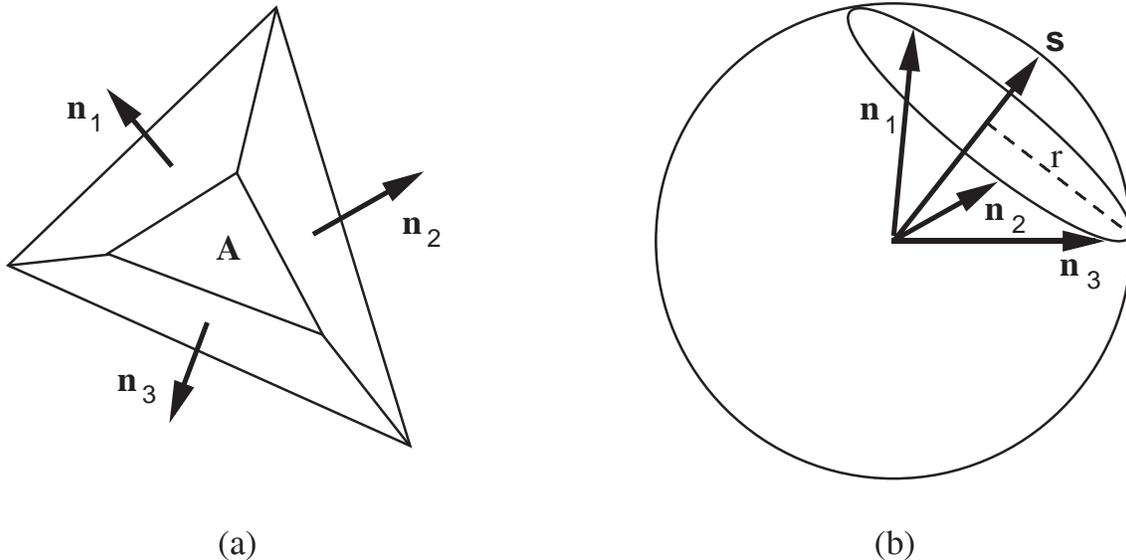


Figure 5: (a) Faces on a polyhedral object. (b) Surface normals of neighboring faces must lie on the circle C for the faces to have equal brightness.

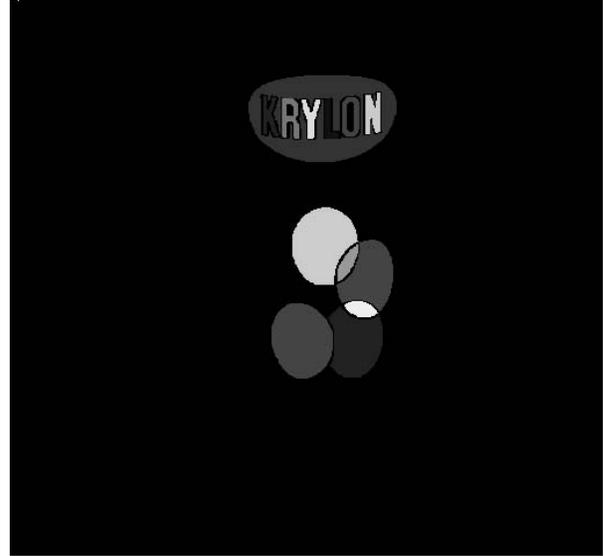
normals must lie on circle C with radius $r = \sqrt{1 - (I_1/k\rho)^2}$ as shown in the figure. In any scene, it is unlikely that this constraint is exactly satisfied. It is conceivable of course that each of the neighboring faces have different coefficients ρ or are illuminated by multiple sources from different directions. These conditions result in another set of constraints on the surface normals. Though a scenario can be contrived that results in all neighboring faces producing equal brightness, such a situation is viewed as degenerate. In our example, if any one of the normals lie outside C , the face A will be treated by the ratio algorithm as a region with multiple backgrounds. A multiple background region is easily detected by computing the variance of the ratios along the region boundary. If it is known a-priori that most regions on the objects of interest have single backgrounds, the ones with multiple backgrounds can be simply discarded to reduce the likelihood of a polyhedral face being interpreted as a constant reflectance patch.

4.4 An Example

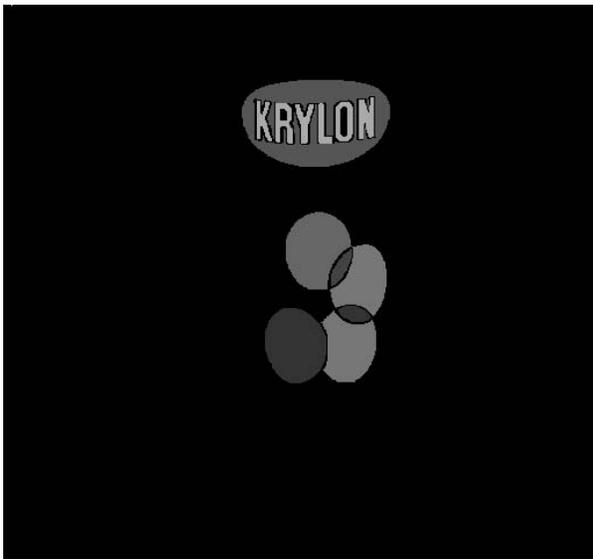
We have conducted several experiments to demonstrate the invariance of reflectance ratios. These results will be presented later. At this point, we show the result of applying the ratio algorithm to the image of a single object. Figure 6 shows an object with several regions that have different reflectance coefficients from their backgrounds. The image in Figure 6(a) was obtained under ambient lighting conditions. Figure 6(b) shows several connected



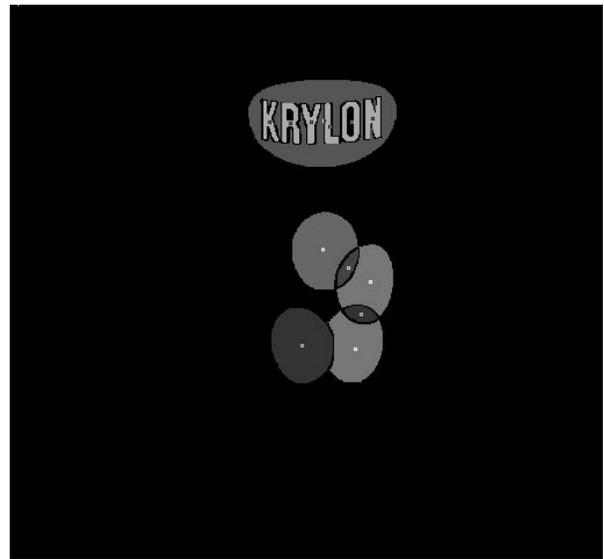
(a) Brightness image.



(b) Labeled image.



(c) Reflectance ratios of object regions.



(d) Centroids of object regions.

Figure 6: Reflectance ratios and region centroids computed for a sample object.

regions extracted using the sequential labeling algorithm. A reflectance ratio threshold of $T = 0.05$ (see (16)) was used to determine connectivity between neighboring pixels. The connected regions are displayed using different gray levels. This image shows only valid object regions, i.e., regions with areas that are neither too small nor too large.

Figure 6(c) shows reflectance ratios of the labeled regions computed using the algorithm described above. Ratio values between -1.0 and 1.0 are offset and scaled to lie between 0 and 255 image brightness levels. Note that all letters in the word “KRYLON” have similar ratio values. In the case of the circular regions, each region is surrounded by more than one background region. Hence, the ratio of each circular region is computed as a weighted average of the ratios with respect to all its background regions. Figure 6(d) shows the centroids of the object regions.

5 Recognition Using Reflectance Ratios

The recognition methods presented here are effective for objects that have markings with different reflectance coefficients. Man-made objects with pictures and text printed on them are good examples of such objects. We consider two recognition scenarios that differ in the assumptions made with respect to the constraints on the objects in the scene.

- **Two-dimensional:** Here, the object could be 2-D or 3-D in geometry and lies in a stable configuration on a plane that is parallel to the image plane of the sensor. Object appearance can therefore vary due to translations and rotations in the plane or due to scaling caused by variations in the magnification of the imaging system. The image formation model is assumed to be perspective. Hence, object features in the image are only be subjected to similarity transformations [Califano and Mohan-1991].
- **Three-dimensional:** In this case, a 3-D object can be in any arbitrary orientation and position in 3-D space. Here, the image formation model is assumed to be weak-perspective; orthographic projection followed by scaling [Huttenlocher and Ullman-1990]. Object features in the image can vary in configuration due to 3-D rotations of the object in the scene.

5.1 Two-Dimensional Object Recognition

5.1.1 Acquiring Object Models

During model acquisition, an object is placed in one of its stable configurations on a plane parallel to the image plane of the sensor. The reflectance ratio algorithm is used to compute the centroids and ratios of regions on the object. This results in a list $L_A = ((\hat{\mathbf{x}}_1, \hat{P}_1),$

$(\hat{\mathbf{x}}_2, \hat{P}_2), \dots, (\hat{\mathbf{x}}_m, \hat{P}_m), \dots$) where $\hat{\mathbf{x}}_m, m = 1, \dots, M$ are the centroids, and $\hat{P}_m, m = 1, \dots, M$ the ratios, of the regions.

From list L_A , indices are formed by taking combinations of three regions, i, j , and k , as shown in Figure 7. The three regions are selected in a counter-clockwise fashion. As photometric invariants we use the reflectance ratios of the three regions. As geometric invariants we use two angles, $\hat{\alpha}$ and $\hat{\beta}$, of the triangle formed by the three regions. The invariant index for the set of three regions is $\langle \hat{P}_i, \hat{P}_j, \hat{P}_k, \hat{\alpha}, \hat{\beta} \rangle$, which is stored in a hash table [Aho *et al.*-1974, Lambdan and Wolfson-1988, Califano and Mohan-1991] as shown below.

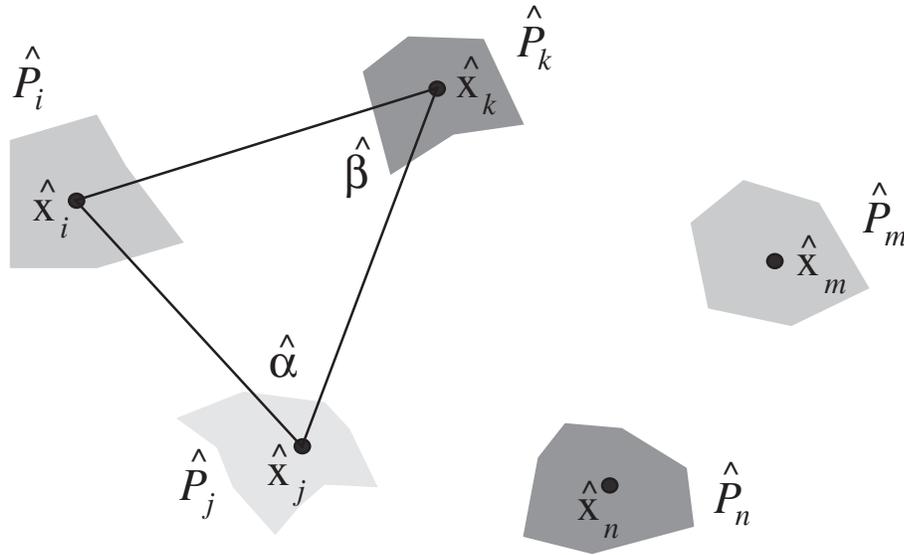


Figure 7: Three regions form an index for the two-dimensional recognition problem. Other regions are used for verification.

<i>INDEX</i>	<i>ENTRY</i>
\vdots	\vdots
$\langle \hat{P}_i, \hat{P}_j, \hat{P}_k, \hat{\alpha}, \hat{\beta} \rangle$	$\langle \mathcal{M}_I, (\hat{\mathbf{x}}_i, \hat{\mathbf{x}}_j, \hat{\mathbf{x}}_k), \{(\hat{\mathbf{x}}_1, \hat{P}_1), \dots, (\hat{\mathbf{x}}_M, \hat{P}_M)\} \rangle$
\vdots	\vdots

Not all combinations of regions are used as indices. We only use those triplets that lie within a small distance relative to the size of the image. This ensures that the number of indices generated is $O(N)$ and not combinatorial in N , where N is the number of visible object regions. Associated with each index in the hash table, is an entry (see table). In the entry, \mathcal{M}_I is the object label and $(\hat{\mathbf{x}}_i, \hat{\mathbf{x}}_j, \hat{\mathbf{x}}_k)$ are the image coordinates of the centroids of the three regions used in the index. The pairs $(\hat{\mathbf{x}}_m, \hat{P}_m), m = 1, \dots, \hat{M}$, are the centroids and reflectance ratios of other regions to be used for object verification. All centroids are expressed in the reference frame of the model image of the object \mathcal{M}_I .

This procedure is applied to all sets of three regions on the object. Hence, object \mathcal{M}_I may be represented by several index-entry pairs in the hash table. The above process is repeated for each object of interest to the recognition system.

5.1.2 Recognition and Pose Estimation

During recognition, the scene may include several objects and the objects may occlude one another. For an object to be recognized at least four of its regions must be completely visible to the sensor, three regions for object hypothesis and at least one for verification.

The recognition process is very similar to model acquisition. The reflectance ratio algorithm is applied to the scene image. The result is a list of region centroids and ratios: $L_R = ((\mathbf{x}_1, P_1), (\mathbf{x}_2, P_2), \dots)$. Each region in the list is a valid region, i.e., it is neither too small (possible edge region) or too large (scene background). Recognition proceeds by selecting three of these regions, say (\mathbf{x}_i, P_i) , (\mathbf{x}_j, P_j) , and (\mathbf{x}_k, P_k) . These regions are selected such that they lie close to one another, again for computational efficiency. From the three selected regions, the index $\langle P_i, P_j, P_k, \alpha, \beta \rangle$ is formed and used to find an entry in the hash table (model database). Here the angles α and β are at the vertices \mathbf{x}_j and \mathbf{x}_k , respectively, of the triangle formed by the three region centers. If this set of ratios and angles $\langle P_i, P_j, P_k, \alpha, \beta \rangle$ represents an index in the hash table, we check to see if the corresponding entry is empty. If so, a new set of three regions are selected from the list L_R .

If the index has a non-empty entry, we have a hypothesis for the object (say \mathcal{M}_K) that the three selected regions lie on. We also have a correspondence between the centroids $(\hat{\mathbf{x}}_i, \hat{\mathbf{x}}_j, \hat{\mathbf{x}}_k)$ in the entry and the centroids $(\mathbf{x}_i, \mathbf{x}_j, \mathbf{x}_k)$ in the scene image. To verify the object using other region ratios and centroids, we need to compute the similarity transformation between the corresponding triplets. The similarity transformation includes a scaling factor s , rotation matrix R , and a translation vector \mathbf{t} :

$$\mathbf{x}' = \mathbf{T}_K(\hat{\mathbf{x}}) = s \mathbf{R} \hat{\mathbf{x}} + \mathbf{t} \quad (21)$$

Using the three corresponding centroids in the image and the table entry, the scaling, rotation and translation of the hypothesized object are determined. This gives a hypothesis

pose of the object in the image. To verify this hypothesis, the entries $(\hat{\mathbf{x}}_m, P_m)$, $m = 1, \dots, M$, are used. Each centroid $\hat{\mathbf{x}}_m$ in the entry is transformed as:

$$\mathbf{x}'_m = \mathbf{T}_K(\hat{\mathbf{x}}_m) \quad (22)$$

If the hypothesis is correct, the scene image is expected to include a region with ratio \hat{P}_m at the image coordinates \mathbf{x}'_m . Hence, the centroid-ratio pair $(\mathbf{x}'_m, \hat{P}_m)$ is compared (using small tolerances) with each of the pairs in the list L_R . If a match is found, we have a region in the image that verifies the object and its pose. Note that this verification process is very reliable since it uses both photometric as well as geometric constraints; we are looking for a region at a particular location in the image with a particular reflectance ratio. Due to occlusions in the scene, all centroid-ratio pairs in the entry of the hash table may not be verified. However, if at least one pair is verified, the object has been recognized and its pose is given by \mathbf{T}_K .

5.2 Three-Dimensional Object Recognition

5.2.1 Acquiring Object Models

The 3-D scenario is more general than the 2-D one as it allows for arbitrary rotations and translations of objects in the scene. Since, our objective is to recover the 3-D pose of an object from a single brightness image, the object model must include reflectance ratios of objects as well as the 3-D coordinates of the centroids of each region. This is done using a range finder which also produces a registered brightness image. The reflectance ratio algorithm is applied to the brightness image and the ratios (\hat{P}_m) and centroids $(\hat{\mathbf{x}}_m)$ (in the model image) of the object's regions are determined.

Next, the range map is used to obtain the 3-D coordinates $(\hat{\mathbf{X}}_m)$ of points on the object surface that correspond to the region centroids in the image. We assume that though the object surface may be curved, each constant reflectance region is small compared to the size of the object. Under this assumption, centroids of regions in the image correspond to centroids of regions in the 3-D scene. Using the above approach, a ratio-centroid list $L_A = ((\hat{\mathbf{X}}_1, \hat{P}_1), (\hat{\mathbf{X}}_2, \hat{P}_2), \dots, (\hat{\mathbf{X}}_m, \hat{P}_m), \dots)$ is obtained for each object. Here, $\hat{\mathbf{X}}_m$, $m = 1, \dots, M$ are the 3-D centroids of the regions and \hat{P}_m , $m = 1, \dots, M$ are the reflectance ratios.

As in the 2-D case, a hash table is initialized. The indices in the hash table must be invariants that can be computed from a single image of the scene. In the 3-D case, there are no useful geometric invariants, such as the angles $\hat{\alpha}$ and $\hat{\beta}$ in the 2-D case, that can be computed from the spatial arrangement of region centroids [Burns *et al.*-1990]. This is because object rotation in the scene changes the relative configuration of region centroids in the image. Thus, we rely on the photometric invariance of reflectance ratios for indexing into the hash table. Three regions, i , j , and k on the object are selected and their reflectance

ratios used to construct an index $\langle \hat{P}_i, \hat{P}_j, \hat{P}_k \rangle$ (see table). Indices are formed using only those region triplets (i, j, k) whose centroids in 3-D space lie close to one another.

<i>INDEX</i>	<i>ENTRY</i>
\vdots	\vdots
$\langle \hat{P}_i, \hat{P}_j, \hat{P}_k \rangle$	$\langle \mathcal{M}_I, (\hat{\mathbf{X}}_i, \hat{\mathbf{X}}_j, \hat{\mathbf{X}}_k), \{(\hat{\mathbf{X}}_1, \hat{P}_1), \dots, (\hat{\mathbf{X}}_M, \hat{P}_M)\} \rangle$
\vdots	\vdots

In the entry, \mathcal{M}_I is the object identifier and $(\hat{\mathbf{X}}_i, \hat{\mathbf{X}}_j, \hat{\mathbf{X}}_k)$ are the 3-D centroids of the three regions used in the index. The entry also includes the centroid-ratio pairs $(\hat{\mathbf{X}}_m, \hat{P}_m)$, $m = 1, \dots, M$ to be used for object verification. The above procedure is applied to all sets of three regions in the list L_A and for all objects. The resulting hash table represents the complete object-model database.

5.2.2 Recognition and Pose Estimation

Though model acquisition requires the use of both brightness and range images of each object, recognition and pose estimation is accomplished using a *single* brightness image. The reflectance ratio algorithm is applied to the scene image to obtain the list $L_R = ((\mathbf{x}_1, P_1), (\mathbf{x}_2, P_2), \dots)$. A set of three regions (i, j, k) is selected from this list. The ratios of the three regions are used to form the index $\langle P_i, P_j, P_k \rangle$. If this index does not have an entry in the hash table, the next set of three regions is selected. If an entry does exist, we have a hypothesis for the object (say \mathcal{M}_K). The entry includes the 3-D centroids of the regions (i, j, k) and a set of centroid-ratio pairs for other regions on the object \mathcal{M}_K . Assuming the object hypothesis is correct, we have a correspondence between the image centroids $(\mathbf{x}_i, \mathbf{x}_j, \mathbf{x}_k)$ and the 3-D centroids $(\hat{\mathbf{X}}_i, \hat{\mathbf{X}}_j, \hat{\mathbf{X}}_k)$ in the entry. Under the weak-perspective assumption, the transformation \mathbf{T} from the 3-D scene points to 2-D image points can be computed from the three corresponding pairs using the alignment technique proposed by Huttenlocher and Ullman [Huttenlocher and Ullman-1990]. In general, there exist two solutions to the transformation [Huttenlocher and Ullman-1990]:

$$\mathbf{x}' = \mathbf{T}_{K1}(\hat{\mathbf{X}}) \quad \text{and} \quad \mathbf{x}' = \mathbf{T}_{K2}(\hat{\mathbf{X}}) \quad (23)$$

Weinshall [Weinshall-1991] has shown that instead of computing these two transformations the inverse of the Gramian of the points $\hat{\mathbf{X}}_i$, $\hat{\mathbf{X}}_j$, and $\hat{\mathbf{X}}_k$ can be used to predict the image

coordinates \mathbf{x}'_m of a fourth 3-D point $\hat{\mathbf{X}}_m$ in the entry. Again, two solutions to \mathbf{x}'_m exist but if the initial object hypothesis is correct, one of the two solutions is likely to be close to one of the centroids in the list L_R . Further, the reflectance ratio \hat{P}_m (in the entry) and P_m (in the list L_R) must be similar. The point \mathbf{x}'_m is not guaranteed to be in the list L_R since it may not be visible to the sensor or it may be occluded by other objects in the scene. In any event, for the object to be verified, one or more projections of the 3-D regions in the entry must match in location and ratio with regions in the list L_R . If so, the object \mathcal{M}_K has been recognized and its pose is given by either \mathbf{T}_{K1} or \mathbf{T}_{K2} .

6 Experiments

In this section, we present two sets of experimental results. First, we demonstrate the invariance of reflectance ratios to several illumination and imaging parameters. These include source direction, number of sources, viewing direction, and optical parameters such as focus, zoom, and aperture. These results show the ratio invariant to be very robust under the assumptions made during its derivation. Next, we present two-dimensional and three-dimensional recognition results that demonstrate that the simultaneous use of photometric and geometric invariants is a powerful approach to recognition.

6.1 Invariance of Reflectance Ratios

Figure 8 illustrates the experimental set-up used. Objects are illuminated using incandescent light sources and are imaged using a Nikon zoom lens and a CCD camera. The illumination and viewing directions are varied by moving the light source and the sensor in a plane that passes through the object. The source direction is represented by the angle θ_i and the viewing direction of the sensor by θ_v . Images (512 x 480 pixels) are digitized using a Matrox frame-grabber and processed on a Sun Sparcstation 2. A simple calibration procedure was used to ensure that the imaging device is linear in its response to scene radiance. In the experiments reported in this section, the object was positioned so as to avoid specularities since the ratio invariant was shown not to hold true in the presence of specularities. In the recognition experiments, however, no such precaution was taken.

Figure 9(a) shows the object discussed earlier in the paper. Figure 9(b) shows the invariance of computed ratios to object illumination using multiple light sources. In this and all following experiments the ratio algorithm was applied using the connectivity threshold $T = 0.05$ and pixel distance $d = 4$. The reflectance ratios for the letter “K” (in the word “KRYLON”) and its oval-shaped background region are computed for source 1 in the direction $\theta_i = 40$ degrees, source 2 in the direction $\theta_i = 70$ degrees, and simultaneous illumination by both sources. Note that the ratio estimates for both regions are unaffected by the use of multiple sources. This is because neighboring points in the scene are

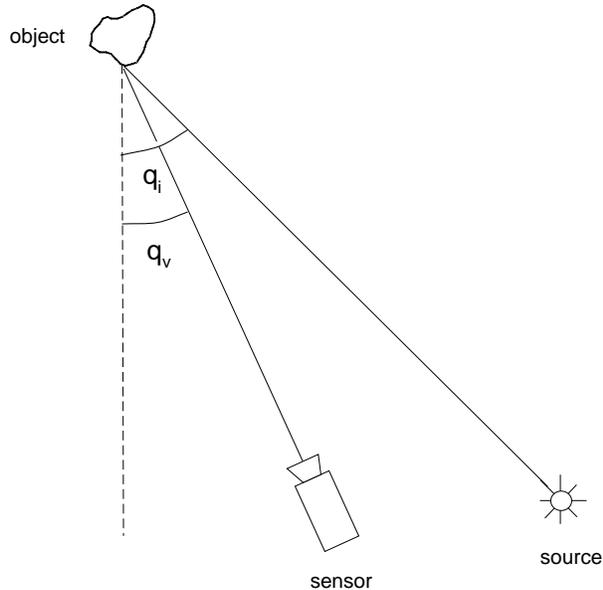


Figure 8: Experimental set-up used to test the invariance of reflectance ratios.

subjected to similar illumination conditions. The algorithm would therefore give accurate region ratios even in the presence of shadows as long as the shadows do not produce sharp brightness discontinuities, i.e., they have a visible penumbra that causes shadow edges to be blurred. The sensitivity of computed ratios to source direction is illustrated in Figure 9(c). The direction of a single light source is varied from $\theta_i = -70$ degrees to $\theta_i = 20$ degrees in increments of 10 degrees. As seen from the figure, the reflectance ratio for region “K” demonstrates remarkable invariance to illumination direction despite the fact that the average brightness as well as the shading within the region varies dramatically with source direction.

The effects of varying the sensor direction are shown in Figure 9(d). As the viewing direction is varied, the projected area and shape of an object region change. As a result, the image boundary of the region also varies. The reflectance ratio of region “K” is computed for different sensor directions starting from $\theta_v = -70$ degrees to $\theta_v = 20$ degrees. In this case, the region is surrounded by a background region with constant reflectance. If on the other hand, a region is surrounded by several regions with different reflectance coefficients, the boundary between the region and any one of the background regions will vary with viewing direction. As described in Section 4.3, for regions with more than one background region, the computed ratio is expected to vary with viewing direction.

We now present results on the sensitivity of computed ratios to the optical parameters of the imaging system. In Figure 10, the aperture of the camera lens is varied and ratio estimates for the region in the image computed. Since varying the aperture has the effect

of varying the sensor gain (see Eq. (9)), the ratio estimates remain more or less unaffected. In Figure 11, the magnification of the imaging system is varied using the zoom setting of the lens. Again, the region ratio is seen to remain near constant as the region is magnified. Finally, in Figure 12 the effect of image defocus is studied. Defocusing has the effect of low-pass filtering the image, resulting in the blurring of region boundaries. In extreme cases of defocus, the sequential labeling algorithm will not be successful in segmenting the region. Furthermore, even if the region is detected, the region ratio estimates tend to be inaccurate since the ratio algorithm uses close by points on either side of the region boundary. As seen in Figure 12, the region ratio varies with image defocus. Interestingly, this phenomenon is also observed in the case of biological vision systems. Psychophysical experiments [Gilchrist-1979] show that the ability of humans to estimate the relative reflectance of a region deteriorates as the edge of the region is blurred.

6.2 Object Recognition

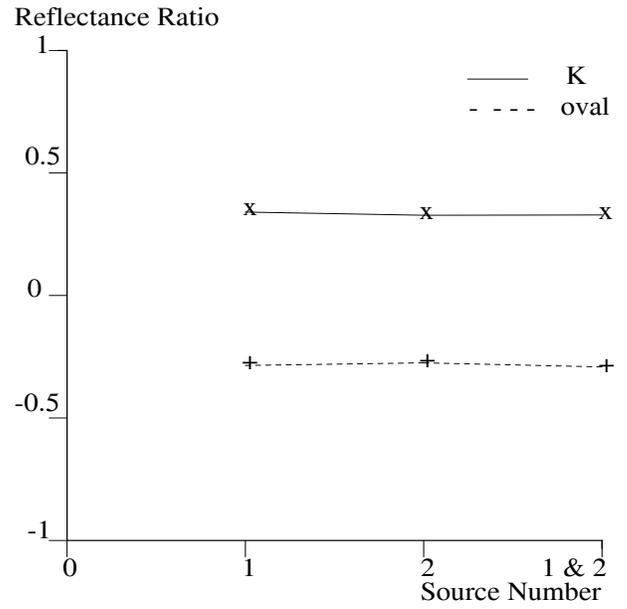
The recognition experiments were conducted on man-made objects with letters and pictures printed on them. The printed regions have reflectance coefficients that depend on the shade or color of the paint used to print them. The approach proposed here is particularly effective for such objects. The reflectance ratio algorithm produces a set of detected regions, each region represented by its centroid in the image and its reflectance ratio. This compact representation of objects is used to automatically acquire object models as well as recognize them in unknown images. The recognition and pose estimation stages are efficient as they are based on the indexing scheme described in Section 5. All images used for model acquisition and object recognition were obtained under ambient lighting conditions.

Figure 13 shows results for a 2-D recognition problem. The image shown in Figure 13(a) is used for model acquisition. One of the sets of three regions used to form an index is shown by the triangle. Other regions on the object are included in the entry of the hash table to be used for object verification and pose estimation. The centroids of these verification regions are indicated by black squares in Figure 13(a).

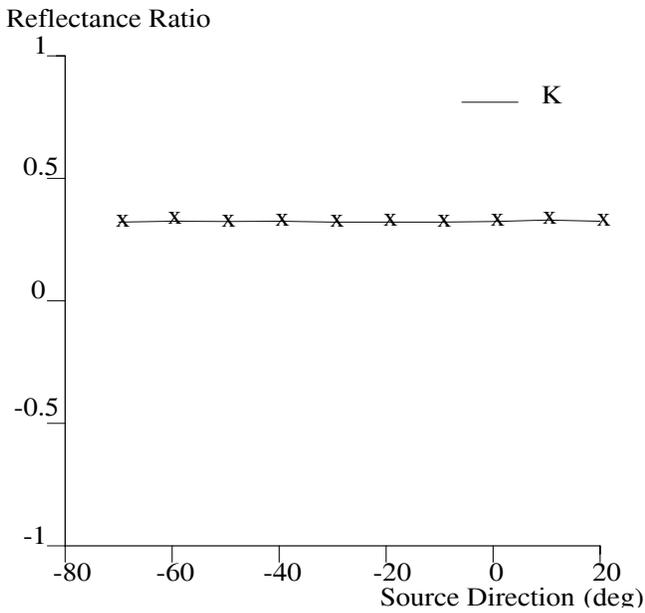
Figure 13(b) shows a scene with several objects. The ratio algorithm produces a total of 58 valid regions with constant reflectance. By valid region we mean those that are neither too small (edge regions) nor too large (object background regions). The index triangle shown in Figure 13(a) is detected in the scene image and is shown as a triangle. This provides a hypothesis for the object and its pose in the scene image. This hypothesis is verified by projecting the verification regions in the hash table to the scene image. The projected regions are shown as white squares in the scene image. The actual centroids of the verification regions in the scene image are shown as black squares. The reflectance ratios and centroids of the actual and projected regions are very close to one another, thus providing several positive verifications for the object in the scene image. Not all of the verification regions shown in the model image are found in the scene image since the



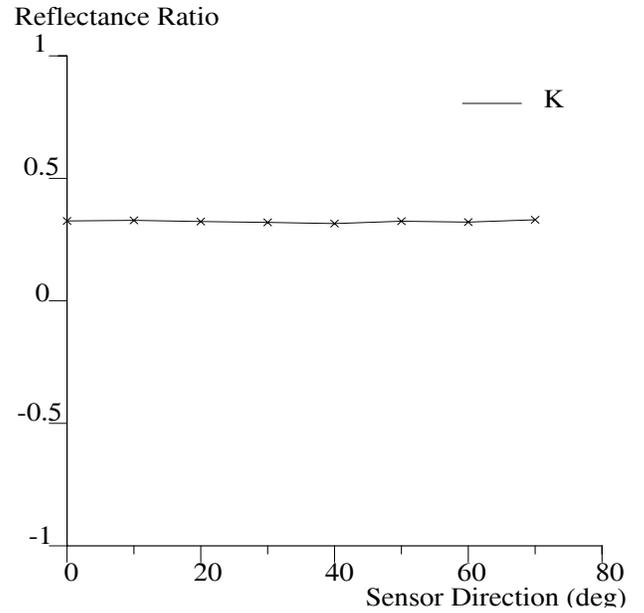
(a)



(b)



(c)



(d)

Figure 9: Invariance of reflectance ratios to: (b) multiple source illumination; (c) direction of illumination; and (d) viewing direction.

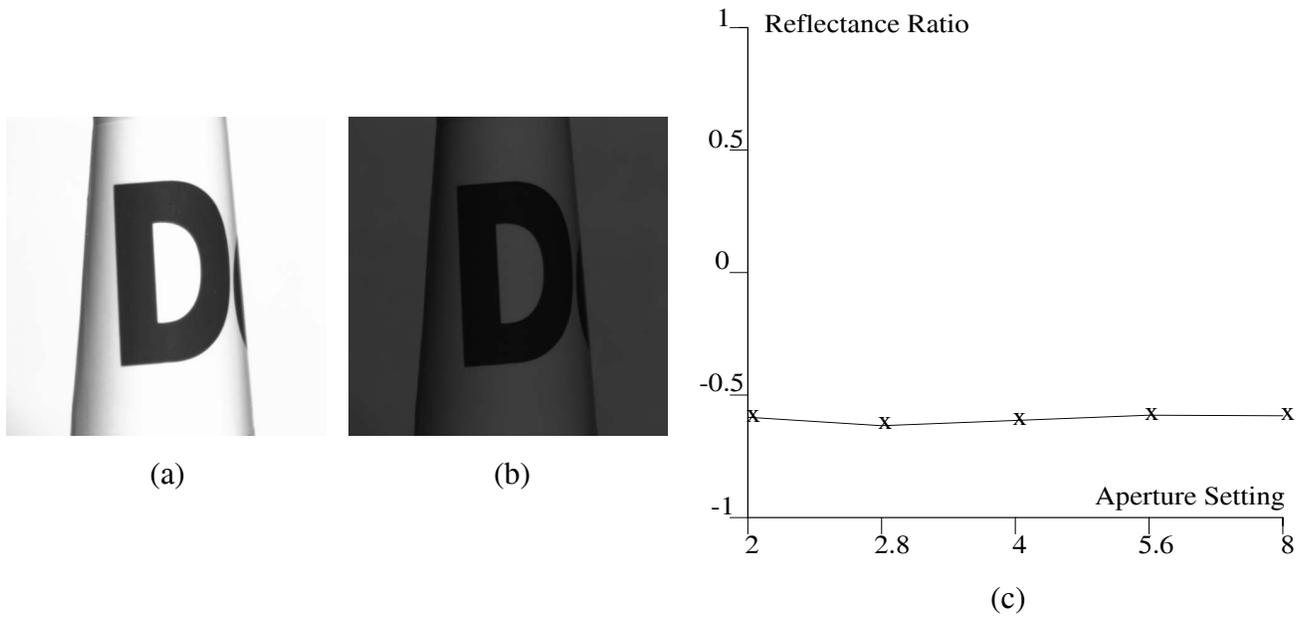


Figure 10: (a) Aperture setting = 2.0. (b) Aperture setting = 5.6. (c) Invariance of reflectance ratios to the aperture of the lens.

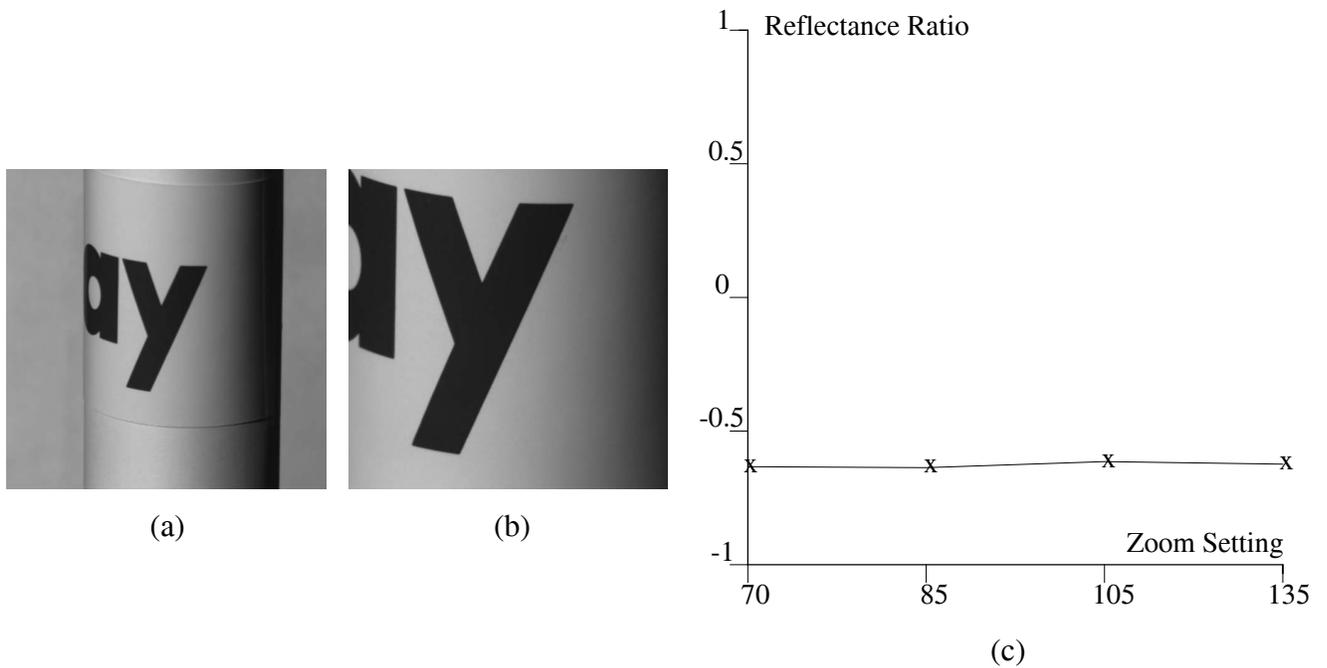


Figure 11: (a) Zoom setting = 70. (b) Zoom Setting = 135. (c) Invariance of reflectance ratios to zoom.

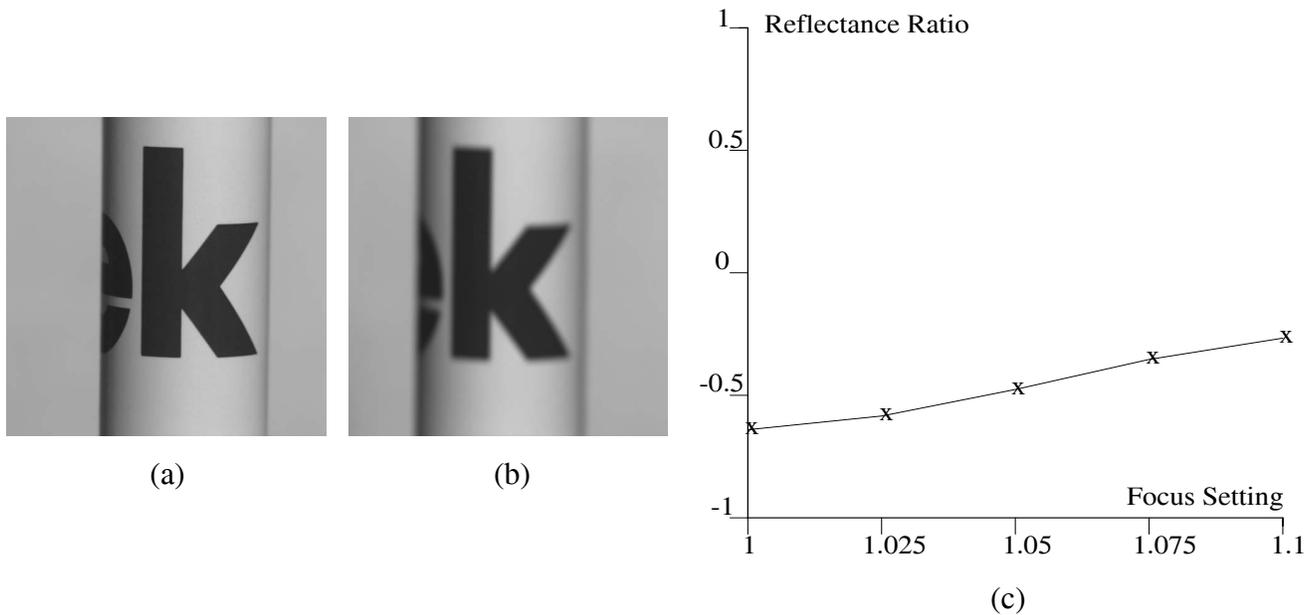


Figure 12: (a) Focus setting = 1. (b) Focus setting = 1.1. (c) Effect of image defocus on computed ratios.

object is partially occluded by another object. It is worth pointing out that the recognition algorithm can also recognize multiple objects and/or multiple instances of the same object in a scene.

Figure 14(a) shows model acquisition for a 3-D object. The range image was obtained using a light stripe range finder. The vertices of the triangle displayed are the centroids of three regions whose reflectance ratios were used as indices in the hash table. Other nearby regions used for verification and pose estimation are indicated by their centroids (black squares). Recognition and pose estimation is done using a single brightness image of the scene. The scene shown in Figure 14(b) consists of several 3-D objects in different orientations and positions. It includes occlusions, shadows, and non-uniform illumination. The reflectance ratio algorithm was applied and a total of 18 constant reflectance regions were detected. The index triangle shown in the model image is found and verified in the scene image. The three index regions produce a hypothesis for the object and its pose. Other regions in the object model are used to verify this hypothesis using the alignment technique (Eq. (23)). Again, some of the verification regions are not found in the scene image since they are occluded by other objects. Further, the actual and projected centroids do not overlap exactly since the assumption that the regions are small compared to the size of the object is never exactly satisfied in practice.



(a) Model acquisition.

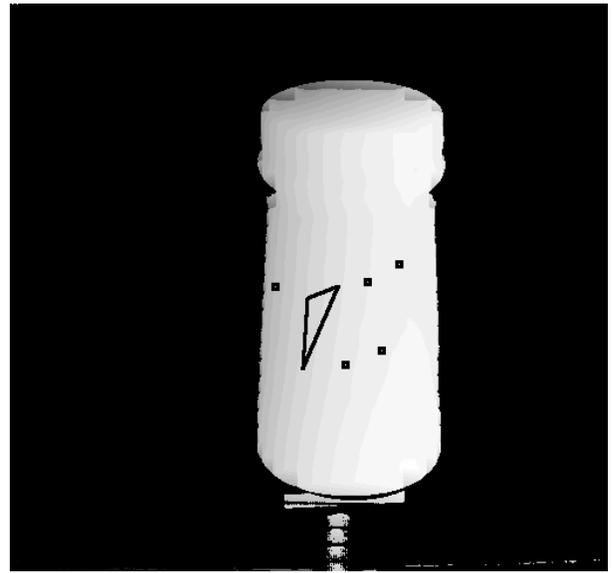


(b) Recognition and pose estimation.

Figure 13: Model acquisition and object recognition results obtained for a two-dimensional recognition problem.



Brightness image.



Range image.

(a) Model acquisition.



(b) Recognition and pose estimation.

Figure 14: Model acquisition and object recognition results obtained for a three-dimensional recognition problem.

7 Discussion

We conclude with a brief discussion on the ideas and results presented in this paper. Some directions for future work are also mentioned.

- **Physical Approach to Visual Perception:** Most will agree that biological vision systems use not only geometric features but also physical attributes such as reflectance for perception. We are fairly adept at distinguishing a smooth surface from a rough one, plastic from metal, cotton from silk, or bronze from copper. Some can even tell artificial wood or metal from the real thing. Machine vision systems have relied primarily on geometry for high-level tasks such as recognition and navigation. In fact, in the past, models of reflectance have been used mainly for recovering scene geometry (shape from shading, for example). Perceptual algorithms too can benefit from the explicit use of non-geometric physical attributes. Reflectance, material, and roughness are examples of such attributes.
- **Representing Physical Attributes:** While arguing in favor of physical attributes, we are faced with several significant problems. One entails the representation of an object's physical attributes. In this paper, we have used a rather simplistic representation; region ratios and centers. To accommodate a larger class of objects, richer descriptions must be explored. It is imperative that the representation be able to handle multiple properties (e.g., shape and reflectance) simultaneously, and yet be compact enough to be called a representation. The shape variations of an object may not be, in any way, correlated with its reflectance variations. For instance, a simple geometry such as a sphere may be highly textured. We may therefore need to represent geometry and reflectance at different resolutions. Further, all difficulties posed by single-attribute representations are also inherited by multi-attribute representations. For instance, one must decide a-priori the level (or scale) at which shape and reflectance variations need to be described.
- **Computing Reflectance from Images:** Representation of physical properties is meaningful only if these properties can be computed from images. We presented an algorithm that computes the relative reflectance of scene regions from a single image. The algorithm may be viewed as an extension of Land's retinex theory to three-dimensional scenes. By using segmentation first, our algorithm overcomes several problems inherent to Land's global method. Fairly straightforward hardware implementations can be envisioned to obtain real-time reflectance estimates. Ideally, we would like to compute the absolute reflectance of a region. Using a single image however only relative reflectance estimates can be obtained. This reflectance ratio was shown to be invariant to a variety of illumination and imaging parameters. The use of a single image also precludes us from being able to handle specular reflections,

faces on a polyhedron, or regions and backgrounds that have different scattering functions. An interesting extension would involve the use of multiple images of a scene obtained from different vantage points.

- **Photometric Invariants:** We demonstrated that reflectance ratios of regions can be used for robust recognition and pose estimation. This approach is of course effective only if the objects of interest have 4 or more (3 for hypothesis and at least 1 for verification) visible regions. Instead of using region ratios, it may be possible to use ratios of just neighboring points and their locations in the image. This would also avoid the need for scene segmentation prior to computing region ratios. The unresolved problem here is the selection of points both from the scene image as well as from the object model for matching and pose estimation. In the past, several other photometric invariants have been proposed for visual perception (see [Koenderink and van Doorn-1980], for examples). These invariants do not directly represent physical properties such as reflectance but rather are functions derived from image brightness that are invariant to pose and illumination for a given shape and reflectance. They are clearly useful for recognition tasks. Some of the proposed invariants are based on high-order spatial derivatives of image brightness and hence suffer from noise sensitivity. However, improvements in imaging technology are being continually made and this problem is expected to fade with time.
- **Integrating Recognition Techniques:** Several recognition techniques have been proposed in the past, each developed with a particular class of objects in mind. In the case of polyhedra, geometric features such as lines and corners provide powerful constraints and invariants. For a smoothly curved object with uniform reflectance, the occluding boundaries and the shading within provide strong cues. As shown here, for objects with constant reflectance patches (surface markings), reflectance ratios and their geometrical arrangement can be used. It is evident that a truly versatile recognition system cannot rely solely on any one of the above techniques. This is a natural consequence of the variety of objects that such a system would have to deal with. The challenge seems to lie in the integration of several recognition strategies into a single system. The broader objective of this paper has been to show that such an integrated system must also rely on physical properties in addition to geometry.

Acknowledgement

The authors would like to thank Ushir Shah for his assistance in implementing the model acquisition and recognition algorithms, and Daphna Weinshall for the alignment code. This research was conducted at the Center for Research in Intelligent Systems, Columbia University. The research was supported in part by ARPA Contract No. DACA 76-92-C-0007 and in part by the David and Lucile Packard Fellowship.

References

- [Aho *et al.*, 1974] A.V. Aho, J.E. Hopcroft and J.D. Ullman. *The Design and Analysis of Computer Algorithms*. Addison-Wesley Publishing Company, Reading, MA, 1974.
- [Ballard and Brown, 1982] D. H. Ballard and C. M. Brown. *Computer Vision*. Prentice Hall, 1982.
- [Besl and Jain, 1985] P. J. Besl and R. C. Jain. Three-dimensional object recognition. *ACM Computing Surveys*, 17(1):75–145, 1985.
- [Blake, 1985] A. Blake. Boundary conditions for lightness computation in Mondrian world. *Computer Vision, Graphics, and Image Processing*, 32:314–327, 1985.
- [Born and Wolf, 1965] A. Born and E. Wolf. *Principles of Optics*. Pergamon, London, 1965.
- [Burns *et al.*, 1990] J.B. Burns, R. Weiss and E. Riseman. View variation of point-set and line segment features. In *Proc. Image Understanding Workshop*, pages 650–659, April 1990.
- [Califano and Mohan, 1991] A. Califano and R. Mohan. Multidimensional indexing for recognizing visual shapes. In *Proc. IEEE Conference on Computer Vision and Pattern Recognition*, pages 28–34, June 1991.
- [Chin and Dyer, 1986] R. T. Chin and C. R. Dyer. Model-based recognition in robot vision. *ACM Computing Surveys*, 18(1), March 1986.
- [Finlayson, 1992] G.D. Finlayson. Colour object recognition. Technical Report M.S. Thesis, Simon Fraser University, 1992.
- [Forsyth and Zisserman, 1989] D. Forsyth and A. Zisserman. Mutual illumination. In *Proc. IEEE Conference on Computer Vision and Pattern Recognition*, pages 466–473, June 1989.
- [Gilchrist, 1979] A. L. Gilchrist. The perception of surfaces blacks and whites. *Scientific American*, pages 112–124, 1979.
- [Gray, 1971] S. B. Gray. Local properties of binary images in two dimensions. *IEEE Trans. on Computers*, 20(5):551–561, May 1971.
- [Healey *et al.*, 1992] G. E. Healey, S. A. Shafer and L. B. Wolff, editors. *Physics Based Vision: Color*. Jones and Bartlett, 1992.

- [Horn, 1974] B. K. P. Horn. Determining lightness from an image. *Computer Graphics and Image Processing*, 3(1):277–299, December 1974.
- [Horn, 1986] B. K. P. Horn. *Robot Vision*. MIT Press, Cambridge, MA, 1986.
- [Hurlbert, 1989] A.C. Hurlbert. *The computation of color*. PhD thesis, Massachusetts Institute of Technology, 1989.
- [Huttenlocher and Ullman, 1990] D.P. Huttenlocher and S. Ullman. Recognizing solid objects by alignment with an image. *International Journal of Computer Vision*, 5(2):195–212, November 1990.
- [Koenderink and van Doorn, 1980] J. J. Koenderink and A. J. van Doorn. Photometric invariants related to solid shape. *Optica Acta*, 27(7):981–996, 1980.
- [Koenderink and van Doorn, 1983] J. J. Koenderink and A. J. van Doorn. Geometrical modes as a general method to treat diffuse interreflections in radiometry. *Journal of Optical Society of America*, 73(6):843–850, 1983.
- [Lambdan and Wolfson, 1988] Y. Lambdan and H. J. Wolfson. Geometric hashing: A general and efficient model-based recognition scheme. In *Proc. International Conference on Computer Vision*, pages 238–249, December 1988.
- [Land and McCann, 1971] E. H. Land and J. J. McCann. Lightness and retinex theory. *Journal of Optical Society of America*, 61(1):1–11, January 1971.
- [Land, 1964] E. H. Land. The retinex. *American Scientist*, 52(2):247–264, June 1964.
- [Land, 1983] E. H. Land. Recent advances in retinex theory and some implications for cortical computations: Color vision and the natural image. *Proc. National Academy of Science*, 80:5163–5169, August 1983.
- [Land, 1986a] E. H. Land. An alternative technique for the computation of the designator in the retinex theory of color vision. *Proc. National Academy of Science*, 83:3078–3080, May 1986.
- [Land, 1986b] E. H. Land. Recent advances in retinex theory. *Vision Research*, 26:7–21, 1986.
- [Lumia *et al.*, 1983] R. Lumia, L. Shapiro and O. Zuniga. A new connected components algorithm for virtual memory computers. *Computer Vision, Graphics, and Image Processing*, 22:287–300, May 1983.
- [Nagy, 1969] G. Nagy. Feature extraction on binary patterns. *IEEE Trans. on Systems, Science, and Cybernetics*, 5(4):273–278, October 1969.

- [Nayar and Bolle, 1993] S. K. Nayar and R. M. Bolle. Computing reflectance ratios from an image. *Pattern Recognition*, 26(10):1529–1542, October 1993.
- [Nayar *et al.*, 1990] S. K. Nayar, K. Ikeuchi and T. Kanade. Determining shape and reflectance of hybrid surfaces by photometric sampling. *IEEE Transactions on Robotics and Automation*, 6:4:418–431, aug 1990.
- [Nayar *et al.*, 1991a] S. K. Nayar, K. Ikeuchi and T. Kanade. Shape from interreflections. *International Journal of Computer Vision*, 6:3:173–195, 1991.
- [Nayar *et al.*, 1991b] S. K. Nayar, K. Ikeuchi and T. Kanade. Surface reflection: Physical and geometrical perspective. *IEEE Trans. on Pattern Analysis and Machine Intell.*, 13(7):611–634, July 1991.
- [Nayar *et al.*, 1993] S. K. Nayar, X. Fang and T. E. Boult. Removal of specularities using color and polarization. In *Proc. IEEE Conference on Computer Vision and Pattern Recognition*, pages 583–590, June 1993.
- [Nicol, 1993] C. J. Nicol. A systolic architecture for labeling the connected components of multi-valued images in real time. In *Proc. IEEE Conference on Computer Vision and Pattern Recognition*, pages 136–141, June 1993.
- [Oren and Nayar, 1993] M. Oren and S. K. Nayar. Generalization of the Lambertian model. In *Proc. Image Understanding Workshop*, pages 1037–1048, January 1993.
- [Rosenfeld and Pfaltz, 1966] A. Rosenfeld and J. L. Pfaltz. Sequential operations in digital picture processing. *ACM*, 13:471–494, October 1966.
- [Sinha and Adelson, 1993] P. Sinha and E. Adelson. Recovering reflectance and illumination in a world of painted polyhedra. In *Proc. International Conference on Computer Vision*, May 1993.
- [Strickler, 1992] M. A. Strickler. Color and geometry as cues for indexing. Technical Report CS 92-22, University of Chicago, Technical Report, November 1992.
- [Swain and Ballard, 1991] M. J. Swain and D. H. Ballard. Color indexing. *International Journal of Computer Vision*, pages 11–32, November 1991.
- [Torrance and Sparrow, 1967] K. Torrance and E. Sparrow. Theory for off-specular reflection from rough surfaces. *Journal of Optical Society of America*, 57:1057–1114, September 1967.
- [Weinshall, 1991] D. Weinshall. Model-based invariants for 3d vision. Technical Report RC 17705, IBM Thomas J. Watson Research Center, December 1991.