Object Recognition Based on Reflectance and Geometry *

Shree K. Nayar

Department of Computer Science
Columbia University
New York, N.Y. 10027

Ruud M. Bolle

Exploratory Computer Vision Group
IBM T. J. Watson Research Center
Yorktown Heights, N.Y. 10598

ABSTRACT

In the past, recognition systems have relied solely on geometric properties of objects. This paper discuss the simultaneous use of geometric as well as reflectance properties for object recognition. Neighboring points on a smoothly curved surface have similar surface orientations and illumination conditions. Hence, 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 computationally efficient as it computes ratios for all image regions in just two raster scans. The region reflectance ratio represents a physical property of a region that is invariant to the illumination conditions.

The reflectance ratio invariant is used to recognize three-dimensional objects from a single brightness image. Object models are automatically acquired and represented using a hash table. Recognition and pose estimation algorithms are presented that use the reflectance ratios 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. The proposed approach to recognition is very effective for objects with printed characters and pictures. We conclude with experimental results on the invariance of reflectance ratios and their application to object recognition.

1 INTRODUCTION

Object recognition has been an active area of machine vision research for the past two decades [1, 2]. 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 image. Edges and vertices are examples of geometric features often used by recognition systems. During image formation, however, a substantial amount of information is lost regarding the geometry of the scene. Hence, geometric features are not always adequate for robust recognition. In the past, little attention has been given to the use of other properties of objects for recognition. In addition to its geometry, an object may be characterized by physical properties such as reflectance, roughness, and material type. Clearly, the representation of an object using all of these properties is useful only if the recognition system is able to compute them from images.

In this 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 intensity and direction of illumination. This photometric invariant, referred to as the *reflectance ratio*, provides valuable information for recognition tasks. The reflectance ratios (photometric features) of object regions and the spatial configuration (geometric features) of the regions are used to represent the object. A serious problem associated

*This research was conducted at the Center for Research in Intelligent Systems, Department of Computer Science, Columbia University. This research was supported in part by the David and Lucile Packard Fellowship and in part by ARPA Contract No. DACA 76-92-C-0007.

0-8194-1280-5/93/$6.00

with existing recognition systems is the acquisition of object models. Most often, the vision programmer is forced to select an appropriate representation, model the objects using this representation, and input this information to the recognition system. This approach is tedious and impractical when dealing with a large number of objects, or objects with complex geometric and reflectance properties. It is clear, that recognition systems of the future must be capable of acquiring object models without human assistance. The recognition method developed in this paper is capable of automatically learning an object's reflectance ratios and their spatial configuration.

The problem of computing the reflectance of regions in a scene was first addressed by Land [3]. In general, image brightness is the product of surface reflectance and illumination. Hence, 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 ingenious 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, the relative reflectance (or "lightness") of regions can be computed. 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. As a result, it is possible to filter out the effects of illumination. Several hardware implementations for the retinex theory have been proposed [4], [5], [6].

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. Each one of these factors can cause a region boundary to go undetected or the computed lightness of a region to be erroneous. Such errors can greatly affect the lightness values computed for all other regions in the image. For this reason, Land's global method is not applicable to most real images. In this paper, we develop an alternative scheme for computing the ratio of the reflectance of a region to that of its background. The image is first segmented into regions of constant (but unknown) reflectance. Next, a reflectance ratio is computed for each region and its background using only points that lie close to the region's boundary. In this 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 [4] restricted themselves to planar (two-dimensional) scenes with patches of constant reflectance. In contrast, our derivation of the reflectance ratio is based on the analysis of regions that lie on curved surfaces. 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).

Recently, Finlayson [7] proposed computing histograms using ratios in different color channels for object recognition. Histograms, however, are in general sensitive to the scale and rotation of objects in the scene and thus not effective for three-dimensional object recognition and pose estimation. Here, we use the reflectance ratio invariant to recognize 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 region are used to represent objects using a hash table. Recognition and pose estimation algorithms are presented that use the reflectance ratios of scene regions 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 reflectance ratios and locations of other regions in the scene. Recognition results are presented for realistic scenes with occlusion, shadows, and illumination variations.

The paper is organized as follows. First, we derive the reflectance ratio for two neighboring points on a curved surface and state the conditions under which it is invariant to illumination and imaging parameters. Next, we introduce the notion of a region reflectance ratio and present an algorithm for computing region ratios in digital images. The algorithm is efficient; it produces region ratios in just two raster scans of the image. We then develop object recognition algorithms that use reflectance ratios as well as geometric constraints for the recognition objects. We conclude with experimental results. The first set of experiments demonstrate the invariance of reflectance ratios to imaging and illumination parameters. Finally, experimental results on object recognition are presented. These results show the simultaneous use of geometric and photometric information to be an effective approach to recognition.

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 patch can be expressed as:

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

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:

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

If we 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:

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

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 $R(.)$ 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 or transmitted through the surface. Two surfaces with the same distribution function $R(.)$ 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 that assumptions be made 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:

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

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

$$f(\lambda) = \delta\,(\lambda' - \lambda) \tag{5}$$

The image brightness measured with such a filter is:

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

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

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

In this case, $R'(.)$ represents the distribution of reflected light for a particular wavelength of incident light. On the other hand, for white-light illumination, $R(.)$ 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 have used the white-light illumination assumption in our experiments, we will use the following expression for image brightness in our discussion:

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

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 microscopic structure of the surface; generally $R(\cdot)$ includes a diffuse component and a specular component [10]. 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.

Consider two *neighboring* points on a surface (Figure 1). For a smooth continuous surface, the two points may be assumed to have the same surface normal vectors. Further, the two 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:

$$I_1 = k\,\rho_1 R_1(\mathbf{s}, \mathbf{v}, \mathbf{n}) \tag{9}$$
$$I_2 = k\,\rho_2 R_2(\mathbf{s}, \mathbf{v}, \mathbf{n}) \tag{10}$$

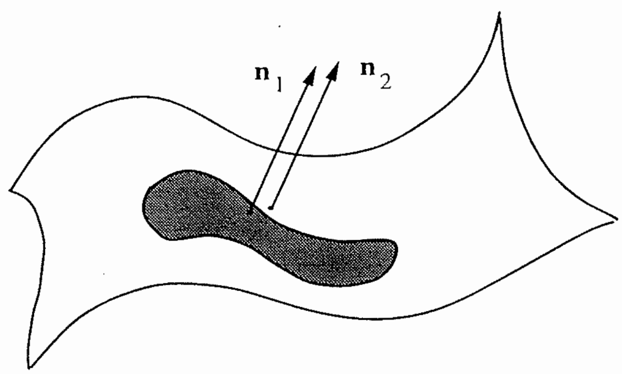


Figure 1: Neighboring points on a surface.

The main assumption made in computing the reflectance ratio is that the two points have the same reflectance functions ($R_1 = R_2 = R$) but their reflectance coefficient ρ_1 and ρ_2 may differ. An example is that of two neighboring Lambertian points [11] that have different albedo values because they lie in regions that have different shades or colors. Then, the image brightness values produced by the two points are:

$$\begin{aligned} I_1 &= k\,\rho_1 R(\mathbf{s}, \mathbf{v}, \mathbf{n}) \\ I_2 &= k\,\rho_2 R(\mathbf{s}, \mathbf{v}, \mathbf{n}) \end{aligned} \tag{11}$$

The ratio of the reflectance coefficients of the two points is:

$$p = I_1 / I_2 = \rho_1 / \rho_2 \tag{12}$$

Note that p is independent of the reflectance function, illumination direction and intensity, and the surface normals of the two 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 object recognition.

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:

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

where $\mathbf{s}_1, \mathbf{s}_2, \ldots, \mathbf{s}_n$ are the directions of the n sources that are visible to the surface point under consideration and k_1, k_2, ..., 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 implementation 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:

$$p = (I_1 - I_2)/(I_1 + I_2) = (\rho_1 - \rho_2)/(\rho_1 + \rho_2) \tag{14}$$

Now, we have $-1 \leq p \leq 1$. We will use this definition of the reflectance ratio in the following sections.

3 COMPUTING REFLECTANCE RATIOS OF REGIONS

To this point, we have focused on two neighboring points. We now consider a surface region 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 surface region S with respect to its background. The 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. However, 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. The reflectance ratio for a region can then be determined as an average of the reflectance ratios computed along the boundary of the region. The computed 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 a scene and provides important information regarding the physical properties of surface regions in the scene.

In this section, we present an algorithm that computes reflectance ratios for scene regions. The algorithm can be divided in two parts. 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 image regions. The algorithm is computationally efficient in that reflectance ratios of all scene regions are computed in just two raster scans of the image. In the following section, we use computed reflectance ratios to represent and recognize objects.

3.1 Sequential Labeling Algorithm

Sequential labeling is a well-known technique for efficient segmentation of images [11]. It has been widely used in the context of binary images [12, 13] where it is 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 [14]. Here, we use the reflectance ratio 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 local shading effects. The connectivity between two pixels is defined as:

$$c(A, B) = \begin{cases} 1 & \text{if } |p(A, B)| < T \\ 0 & \text{otherwise} \end{cases} \tag{15}$$

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 in 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. 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.

Using this algorithm, the complete image can be segmented in a single raster scan. Following the raster scan, the 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.

3.2 Algorithm for Computing Reflectance Ratios

Sequential labeling provides a set of image regions that correspond to surface regions. Each region is assumed to have a constant reflectance coefficient. The brightness of the region may vary due to variations in surface normal and illumination. However, a completely connected region is obtained since local brightness variations are small; neighboring pixels have nearly equal surface normal vectors and illumination conditions. The following is an example one-dimensional image of a region and the result of sequential labeling:

(a)	35	37	39	41	64	77	85	87	89	89	91	92	94	96
(b)	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 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 [11]. 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 [15]. Thus, pixels on and around the edge end up with different labels from the region and the background.

One way of computing the reflectance ratio $P(S)$ of a region S is by using the average brightness of the region and the average brightness of the background. Since, brightness generally varies over both the region and the background, this will not yield an accurate reflectance ratio. Instead, we can obtain an accurate estimate of the reflectance ratio by using only points on the boundary of the region. Consider a pixel that lies in the region but on its boundary. An estimate of the reflectance ratio can be computed using this pixel and the background pixel that is closest to it. Note that the surface normal and illumination of the region boundary pixel and its closest background pixel can be assumed to be equal. A robust estimate for the reflectance ratio of a region can be computed as an average of the ratios computed using all its boundary pixels.

As mentioned earlier, the sequential labeling algorithm 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 value. We focus only on larger regions that are referred to as *valid* regions. The 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 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 if K lies in a valid region, i.e, we 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:

$$p_i(Label(X)) = p(X, K) = (I_X - I_K)/(I_X + I_K) \tag{16}$$

This is the ith 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 of the image, 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 the ratios in its list:

$$P(S) = 1/N \sum_{i=1}^{N} p_i(S) \tag{17}$$

where, N represented the total number of ratio estimates obtained for the region 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 $P(S)$ of a region is defined as:

$$\gamma(S) = N^2/A(S) \tag{18}$$

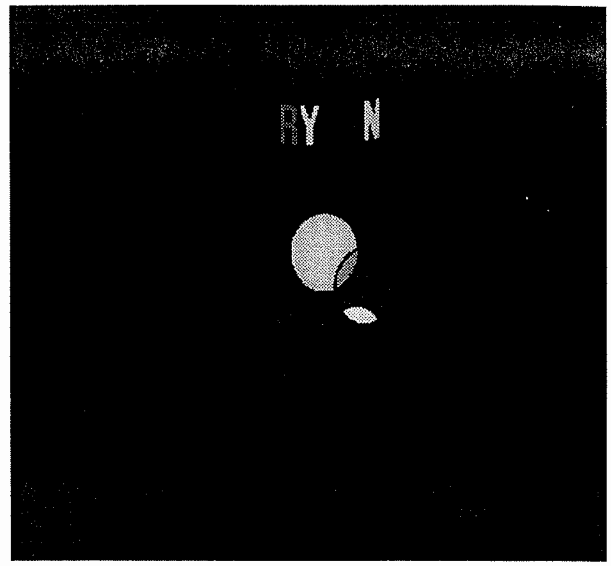
where $A(S)$ is the area of the region S. This confidence may be used as a measure of the accuracy of the reflectance ratio computed for the region. If $\gamma(S)$ is small, few ratios have contributed to the estimate and it may be unreliable for recognition purposes.

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 the viewing direction since the fraction of the region boundary shared with any particular background region can vary with viewing direction.

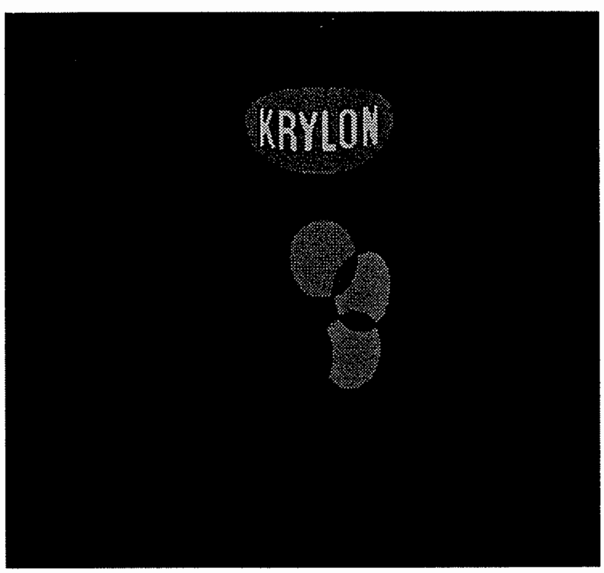
We have conducted several experiments to demonstrate the invariance of reflectance ratios. These results will be presented later. At this point, we show the result of applying reflectance ratio algorithm to the image of a single object. Figure 2 shows an object with several regions that have different reflectance coefficients from their backgrounds. The image in Figure 2a was obtained under ambient lighting conditions. Figure 2b shows several connected regions extracted using the sequential labeling algorithm. A reflectance ratio threshold of $T = 0.05$ (see (15)) 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 2c shows the reflectance ratios of the labeled regions computed using the algorithm presented in this section. 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 though they are illuminated differently. 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 2d shows the centroids of the object regions. If a region is near planar in 3-D space, its centroid in the image can be assumed to be the projection of its centroid in 3-D space. This assumption is valid for both orthographic as well as perspective image formation models.

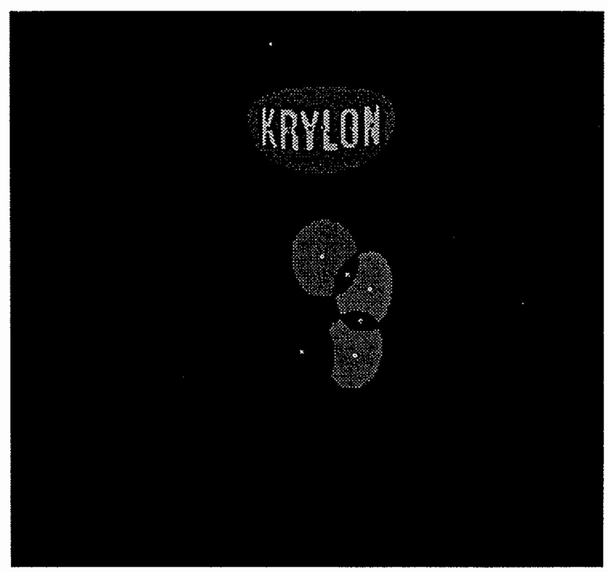
(a) Brightness image.



(b) Labeled image.



(c) Reflectance ratios of object regions.



(d) Centroids of object regions.

Figure 2: Reflectance ratios and region centroids computed for a sample object.

4 OBJECT RECOGNITION

In this section, we apply reflectance ratios to the problem of object recognition. The recognition method presented here is 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.

4.1 Learning Object Models

Since, our objective is to recover the three-dimensional pose of an object from a single brightness image, the object model must include reflectance ratios of the object as well as the three-dimensional coordinates of the centroids of each region. This is done using a range finder. We use the image sensor of the range finder to also obtain a brightness image of the object. As a result, the range and brightness images of the object are registered. The reflectance ratio algorithm is applied to the brightness image and the ratios ($\hat{P}_m$) and centroids ($\hat{\mathbf{x}}_m$) (in the image) of the object's regions are determined. Next, the range map is used to obtain the three-dimensional coordinates ($\hat{\mathbf{X}}_m$) of points of 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 and hence can be assumed to be near-planar. Under this assumption, centroids of regions in the image correspond to centroids of the regions in the 3-D scene. Using the above approach, a ratio-centroid list $L_A = ((\hat{\mathbf{X}}_1, \hat{P}_1), (\hat{\mathbf{X}}_2, \hat{P}_2), \ldots, (\hat{\mathbf{X}}_m, \hat{P}_m), \ldots)$ is obtained for each object. Here, $\hat{\mathbf{X}}_m$, $m = 1, \ldots, M$ are the 3-D centroids of the regions and $\hat{P}_m$, $m = 1, \ldots, M$ are the reflectance ratios.

Next, a hash table [16] is initialized as shown below. All object models are stored in the same hash table. The indices in the hash table are invariants that can be computed from a single image of the scene. There are no useful geometric invariants that can be computed from the spatial arrangement of the region centroids. This is because object rotation in the scene changes the relative configuration of the region centroids in the image. Hence, we rely on the photometric invariance of reflectance ratios for indexing into the hash table. We select three regions, i, j, and k on the object and use their reflectance ratios to obtain an index $< \hat{P}_i, \hat{P}_j, \hat{P}_k >$. Indices are formed using only those region triplets (i, j, k) whose centroids in 3-D space lie within the radius of coherence D_A. This ensures that the number of indices generated is $O(N)$, with N the number of visible regions on the object, and not combinatorial in N.

INDEX	*ENTRY*
$\vdots$	$\vdots$
$< \hat{P}_i, \hat{P}_j, \hat{P}_k >$	$< \mathcal{M}_I, (\hat{\mathbf{X}}_i, \hat{\mathbf{X}}_j, \hat{\mathbf{X}}_k), \{(\hat{\mathbf{X}}_1, \hat{P}_1), \ldots (\hat{\mathbf{X}}_{\hat{M}}, \hat{P}_{\hat{M}}\}) >$
$\vdots$	$\vdots$

Associated with each index in the hash table is an entry. In the entry are stored, the object identifier $\mathcal{M}_I$, and the 3-D coordinates of the centroids $(\hat{\mathbf{X}}_i, \hat{\mathbf{X}}_j, \hat{\mathbf{X}}_k)$ of the three regions used in the index. The entry also includes the ratio-centroid pairs $(\hat{\mathbf{X}}_m, \hat{P}_m)$, of other object regions that are used for object verification and pose estimation.

The above procedure is applied to all sets of three regions in the list L_A. Each object is typically represented by several indices and entries in the hash table. This process is repeated for all objects, $\mathcal{M}_I, I = 1, \ldots, \mathcal{O}$, of interest to the recognition system. The resulting hash table represents the complete object-model database which is ready for use by the recognition system.

4.2 Recognition and Pose Estimation

Though model acquisition requires the use of both a brightness and a range image 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 $L_R = ((\mathbf{x}_1, P_1), (\mathbf{x}_2, P_2), \ldots)$. For recognition, a set of three regions is selected from the list L_R. Consider the three regions (i, j, k). This set is used only if the image centroids of the regions j and k lie within the radius of coherence D_R from the centroid of the region i. The ratios of the three regions are used to form the index $< P_i, P_j, P_k >$. If this index does not have an entry in the hash table, the next set of three regions is selected from L_R. 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 3-D centroids and image centroids using the alignment technique proposed by Huttenlocher and Ullman [17]. In general, however, there exist two solutions to the transformation [17]:

$$\mathbf{x} = \mathbf{T}_{K1}(\hat{\mathbf{X}}) \text{ and } \mathbf{x} = \mathbf{T}_{K2}(\hat{\mathbf{X}}) \tag{19}$$

Weinshall [18] has shown that instead of computing these two transformations the inverse of the Grammian of the points $\hat{\mathbf{X}}_i$, $\hat{\mathbf{X}}_j$, and $\hat{\mathbf{X}}_k$ can be used to predict the image coordinates $\hat{\mathbf{x}}_o$ of a fourth 3-D point $\hat{\mathbf{X}}_o$ in the entry. Again, two solutions to $\hat{\mathbf{x}}_o$ 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 relectance ratio $\hat{P}_o$ (in the entry) and P_o (in the list L_R) must be similar. The point $\hat{\mathbf{x}}_o$ is not gauranteed 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 case, 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 $\mathbf{T}_{K2}$.

At this point, all regions used as indices and those that are verified are removed from the list L_R. A new set of three regions is selected from the list and used to form the next index. This process is repeated until either all objects in the hash table have been recognized or all regions in the list L_R have been explained.

5 Experiments

This section presents experimental results related to the reflectance ratio as well as the application of reflectance ratios to object recognition. First, we demonstrate the invariance of reflectance ratios to the number of light sources in the scene, the direction of illumination, and the viewing direction of the sensor. These results show that the ratio invariant is very robust under the assumptions made during its derivation. Next, we present recognition results that demonstrate that the simultaneous use of photometric and geometric information is feasible and effective.

5.1 Invariance of Reflectance Ratios

In section 3, we showed results of applying the reflectance ratio algorithm to the brightness image of an object. Here, we focus on the invariance of reflectance ratios to illumination and imaging parameters. Figure 3 illustrates the experimental set-up used. Objects are illuminated using incandescent light sources and are imaged using a Nikon 50mm 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 are digitized using a Matrox frame-grabber and processed on a Sun Sparcstation 2.

Figure 4(a) shows the object discussed earlier in the paper. Figure 4(b) shows the invariance of computed ratios to object illumination using multiple light sources. The reflectance ratios for the region "K" on the object 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. The sensitivity of computed ratios to source direction is illustrated in Figure 4 (c). The direction of a single light source is varied from $\theta_i =$ -70 degrees to $\theta_i =$

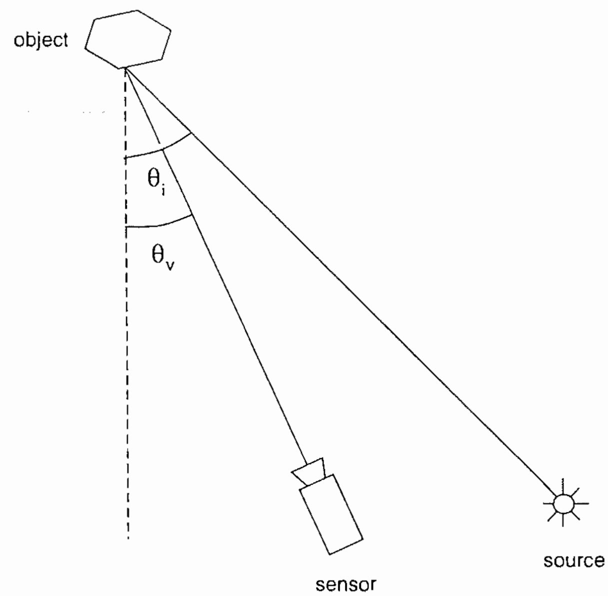


Figure 3: Experimental set-up used to test the invariance of reflectance ratios.

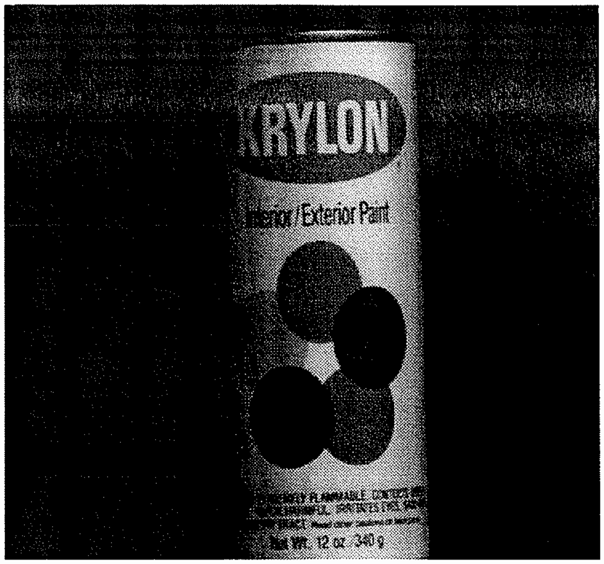
20 degrees in increments of 10 degrees. As seen from the figure, the reflectance ratio for region "K" demonstrates remarkable invariance to illumination direction.

The effects of varying the sensor direction are shown in Figure 4(d). As the viewing direction is varied, the projected area and shape of an object region changes. As a result, the boundary of the region also varies. The reflectance ratio of region "K" is computed for different sensor directions starting from θ_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 directions. Hence, for regions with more than one background region, the computed ratio is expected to vary with viewing direction.

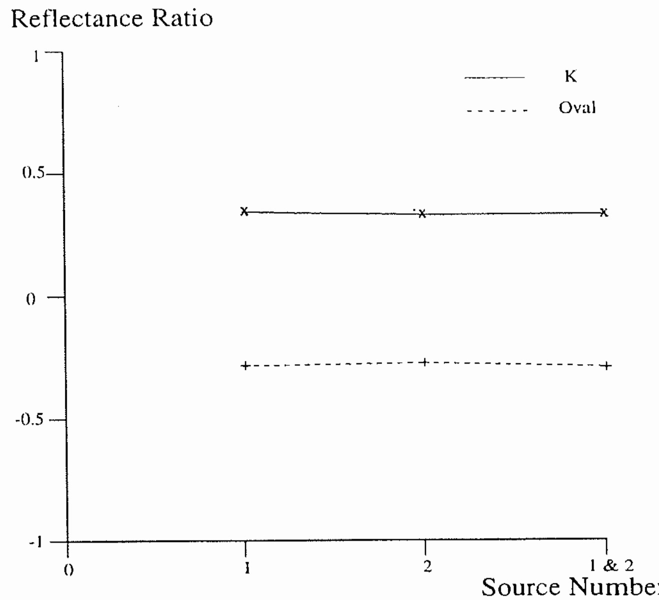
5.2 Object Recognition Using Reflectance Ratios

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. Figure 5(a) shows the model acquisition results 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 that are included in the hash table entry for object verification and pose estimation are indicated by their centroids (black boxes). All regions used for model acquisition and recognition are assumed to be near-planar though the objects they lie on may be curved. Under this assumption, the centroid of an image region corresponds to the projection of the centroid of the region in 3-D space. The planarity assumption is generally found to be reasonable for regions that are small compared to the size of the object.

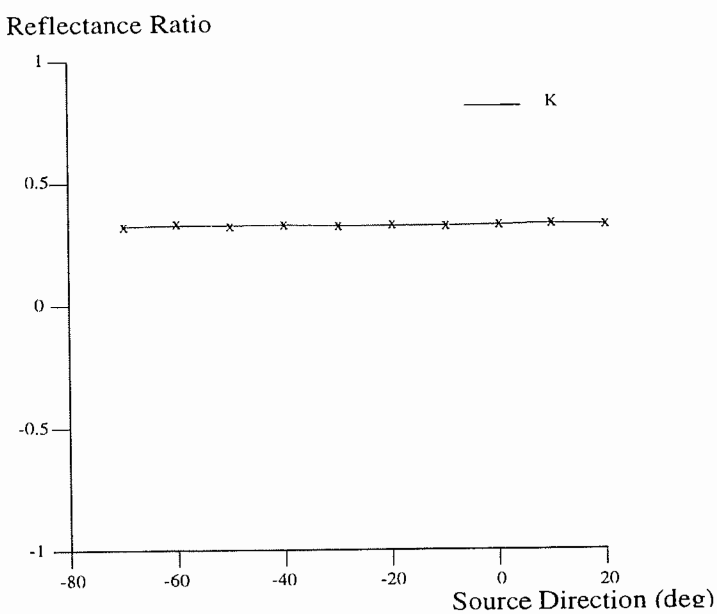
Recognition and pose estimation is done using a single brightness image of the scene. The scene shown in Figure 5(b) consists of several 3-D objects in different orientations and positions. It includes occlusions, shadows, and non-uniform illumination. The reflectance ratio algorithm was applied to the scene image 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 set of three regions in the scene image produce a hypothesis for the object. Other regions in the object model are used to verify this hypothesis using the alignment technique (19). The actual and projected centroids of the verification regions are indicated by black and white boxes, respectively. Some of the verification regions are not found in the scene image since they are occluded by other objects.



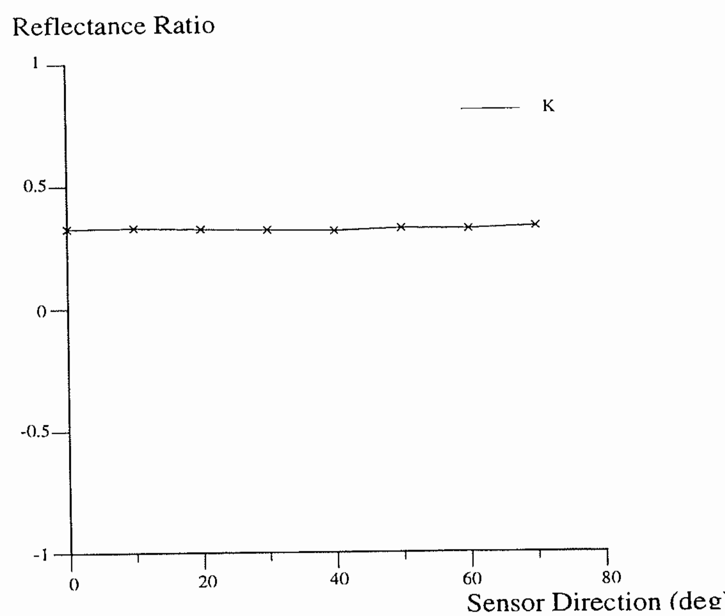
(a) Object.



(b) Reflectance Ratio vs. source number.



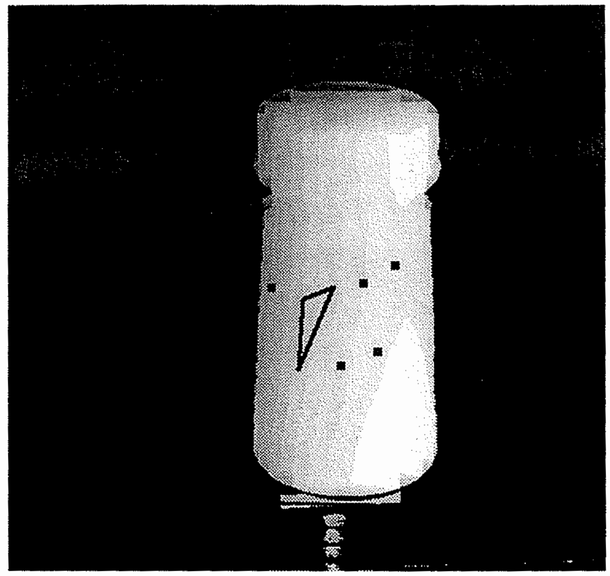
(c) Reflectance ratio vs. source direction.



(d) Reflectance ratio vs. sensor direction.

Figure 4: Invariance of reflectance ratios to: (b) multiple source illumination; (c) direction of illumination; (d) viewing direction.

Brightness image.



Range image.

(a) Object model acquisition.

(b) Object recognition and pose estimation.

Figure 5: Model acquisition and object recognition results obtained for a three-dimensional recognition problem.

6 CONCLUSION

We conclude with the main points of this paper.

- We have presented a photometric invariant, called reflectance ratio, that is computed from a single brightness image. An algorithm is developed that computes reflectance ratios of regions in the scene in just two raster scans of the image. Experimental results demonstrate the invariance of reflectance ratios to illumination and imaging parameters.

- Reflectance ratios have been used to recognize objects from a single brightness image. The recognition techniques are based on the indexing scheme. Both photometric and geometric constraints are used to hypothesize and verify objects and their poses in the image. This approach is reliable and efficient for objects that have surface patches with different reflectance coefficients.

- The recognition system described here is capable of automatically learning models of the objects of interest. Each object is represented by a set of regions, each region described by its reflectance ratio and the location of its centroid. Experimental results on model acquisition and object recognition are presented.

Through this work we have taken initial steps towards using photometric information, in addition to geometric properties, for recognition. These ideas can be extended in several ways. The present method requires the segmentation of the scene into regions. Further, recognition is possible only if an object has more than four visible regions; three for hypothesis and at least one for verification. Hence, the method is applicable only to objects with several reflectance regions (characters or pictures). We are currently exploring ways of using the reflectance ratio invariant without having to compute regions.

Acknowledgement

The authors would like to thank Ushir Shah of Columbia University for his assistance in implementing the model acquisition and recognition algorithms.

References

[1] P. J. Besl and R. C. Jain. Three-dimensional object recognition. *ACM Computing Surveys*, 17(1):75–145, 1985.

[2] R. T. Chin and C. R. Dyer. Model-based recognition in robot vision. *ACM Computing Surveys*, 18(1), March 1986.

[3] E. H. Land. The retinex. *American Scientist*, 52(2):247–264, June 1964.

[4] E. H. Land and J. J. McCann. Lightness and retinex theory. *Journal of Optical Society of America*, 61(1):1–11, January 1971.

[5] B. K. P. Horn. Determining lightness from an image. *Computer Graphics and Image Processing*, 3(1):277–299, December 1974.

[6] A. Blake. Boundary conditions for lightness computation in mondriaan world. In *Central and Peripheral Mechanisms of Color Vision*. Macmillan: New York, 1985.

[7] G.D. Finlayson. *Colour Object Recognition*. PhD thesis, Simon Fraser University, 1992.

[8] Y. Lambdan and H.J. Wolfson. Geometric hashing: A general and efficient model-based recognition scheme. In *Proc. 2nd Int. Conf. on Computer Vision*, pages 238–249, December 1988.

[9] A. Califano and R. Mohan. Multidimensional indexing for recognizing visual shapes. In *Proc. IEEE 1991 Conf. Computer Vision and Pattern Recognition*, pages 28–34, June 1991.

[10] S. K. Nayar, K. Ikeuchi, and T. Kanade. Surface reflection: Physical and geometrical perspectives. *IEEE Trans. on Pattern Analysis and Machine Intell.*, 13(7):611–634, July 1991.

[11] B. K. P. Horn. *Robot Vision.* MIT Press, Cambridge, MA, 1986.

[12] G. Nagy. Feature extraction on binary patterns. *IEEE Trans. on Systems, Science, and Cybernetics*, 5(4):273–278, October 1969.

[13] S. B. Gray. Local properties of binary images in two dimensions. *IEEE Trans. on Computers*, 20(5):551–561, May 1971.

[14] D. H. Ballard and C. M. Brown. *Computer Vision.* Prentice Hall, 1982.

[15] A. Born and E. Wolf. *Principles of Optics.* Pergamon, London, 1965.

[16] A.V. Aho, J.E. Hopcroft, and J.D. Ullman. *The Design and Analysis of Computer Algorithms.* Addison-Wesley Publishing Company, Reading, MA, 1974.

[17] D.P. Huttenlocher and S. Ullman. Recognizing solid objects by alignment with an image. *Int. Journal of Computer Vision*, 5(2):195–212, November 1990.

[18] D. Weinshall. Model-based invariants for 3d vision. Technical Report RC 17705, IBM Thomas J. Watson Research Center, December 1991.