Physics-Based Segmentation: Looking Beyond Color

Bruce A. Maxwell and Steven A. Shafer

CMU-RI-TR-95-37

Robotics Institute
Carnegie Mellon University
Pittsburgh, Pennsylvania 15213

25 October 1995

This research was partially supported by the Advanced Research Projects Agency of the Department of Defense and was monitored by the Air Force Office of Scientific Research under contract F49620-92-C-0073, ARPA Order No. 8875. The views and conclusions contained in this document are those of the authors and should not be interpreted as representing the official policies, either expressed or implied, of the U.S. government. The United States Government is authorized to reproduce and distribute reprints for government purposes notwithstanding any copyright notation hereon.

Abstract

We previously presented a framework for segmentation of complex scenes using multiple physical hypotheses for simple image regions. A consequence of that framework was a proposal for a new approach to the segmentation of complex scenes into regions corresponding to coherent surfaces rather than merely regions of similar color. Herein we present an implementation of this new approach and show example segmentations for scenes containing multi-colored piece-wise uniform objects. By using this new approach we are able to intelligently segment scenes with objects of greater complexity than previous physics-based segmentation algorithms. The results show that by using general physical models we can obtain segmentations that correspond more closely to objects in the scene than segmentations found using only color.

Figure 1 Complex scene containing multiple materials and multi-colored objects

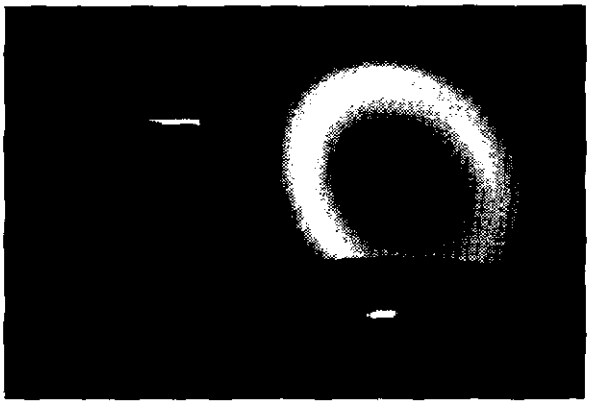


Figure 2 Uniformly colored objects.

Figure 3 Multi-colored object.

Figure 4 Image of an object, a reflected image of the object, and a photograph of the object.

Section 1. Introduction

Images containing multi-colored objects and multiple materials such as Figure 1 are difficult to understand and segment intelligently. Simpler scenes like Figure 2 with only uniformly colored objects of known material type can be segmented into regions that correspond to objects using color and one or two known physical models to account for color variations due to geometry and phenomena such as highlights [1] [7] [8]. Using these methods, a discontinuity in color between two image regions is assumed to imply discontinuities in other physical characteristics such as the shape and reflectance.

Multi-colored objects, like the mug in Figure 3, violate this assumption. The change in color between two image regions does not necessarily imply a discontinuity in shape, illumination, or other characteristics. To correctly interpret more complex scenes such as this, multiple physical characteristics must be examined to determine whether two image regions of differing color belong to the same object. The most successful physics-based segmentation methods to date do not attempt to solve this problem. Instead, they place strong restrictions on the imaging scenario they can address--especially material type and illumination--to permit the effective use of one or two easily distinguished models [1] [4] [7] [8].

The difficulty inherent in segmenting images with multiple materials and multi-colored objects is that by expanding the space of physical models considered for the shape, illumination, and material optics, a given image region can be described by a subspace of the general models; each point within this subspace is a valid explanation for the image region. In Figure 1, for example, the reflection of the bucket in the copper kettle may be part of the kettle (copper reflecting colored illumination) or it could be a separate object (painted metal reflecting white illumination). Likewise, the shadow on the large ceramic vase could be due to differing illumination or could be painted on the vase itself. Either is a valid explanation for the image region in isolation.

Figure 4 is an even more graphic example of this. The boxes show three roughly identical image regions. The region on the right that is part of a photograph and the variation is due to changes in the material properties (color and inten-

sity). The variation in the middle region is due to the geometry of the object surface and the illumination. Finally, the variation in the left-most box is due to variation in the illumination over the surface of the mirror.

Therefore, to segment an image with numerous possible materials, shapes, and types of illumination, we must select not only the model parameters, but also the models themselves. Furthermore, we have to realize that the image may be ambiguous; we cannot simply select a single hypothesis, but must entertain several possibilities. In other words, we can never expect to get *the* single correct interpretation of Figure 4, only a *possible* correct interpretation.

Model selection, or instantiation has only recently been introduced to physics-based vision. Breton *et al.* have presented a method for instantiating models for both the illumination and shape, however, they still consider only a single model for material type (Lambertian) [3]. In [11] we presented a framework for segmentation using multiple physical hypotheses for shape, illumination, and material properties. This framework was based upon the division of a model space comprised of general parameterizations of the transfer function, illumination, and shape into broad classes, or subspaces. By reasoning about these subspaces, we proposed a method for accepting or rejecting mergers between the hypotheses of adjacent regions.

This paper describes an initial implementation of that framework using a limited set of those hypotheses. With this limited set, images containing multi-colored piece-wise uniform dielectric objects can be segmented so that the final segmentation more closely corresponds to objects in the scene than segmentations found using only color.

In section 2 we summarize the fundamental hypotheses and show how the limited hypothesis set used in this implementation fits into the general framework. In section 3 we then discuss direct instantiation of the hypotheses using analysis of individual image regions. We show that this is a very hard problem given existing vision tools. In section 4 we present a partial solution to this problem by exploring physical invariants that measure the similarity of the elements of adjacent hypotheses without requiring direct instantiation. Using these tools of analysis, in section 5 we show how a multi-level region graph can be created and used to find a set of segmentations for the image. Finally, in sections 6 and 7 we discuss the results of our segmentation method on two test images, discuss these results, and present some directions for future work.

Section 2. Modeling Scenes

Our model for a scene consists of three elements: surfaces, illumination, and the light transfer function or reflectance of a point or surface in 3-D space. These elements constitute the *intrinsic characteristics* of a scene, as opposed to *image features* such as pixel values, edges, or flow fields [17]. The combination of models for these three elements is a *hypothesis* of image formation. By attaching a hypothesis to an image region we get a *hypothesis region*: a set of pixels and the physical process which gave rise to them. When an image region has multiple hypotheses, we call the combination of the image region and the set of hypotheses a *hypothesis list*.

Without prior knowledge of image content, no matter how an image is divided there are numerous possible and plausible hypotheses for each region. Variation in the color of an image region can be caused by changes in the illumination, the transfer function, or both. Likewise, variation in intensity can be caused by changes in the shape, illumination, transfer function, or any combination of the three. Many algorithms (in particular shape-from-shading) work because they assume the image variation is due to changes in only one element of the hypothesis (shape) [5].

Section 2.1. Taxonomy of the Scene Model

In [11] we proposed a general parametric representation for each element of a hypothesis based upon the known physical parameters. Because of their generality, however, the raw parametric models do not provide any guide to segmentation. Unlike the method of Breton *et. al.*, there are too many parameters in our models to undertake a brute-force discretization of the space of possible models. Instead, we divide the parameter space into a set of broad classes, or subspaces. These subspaces are broad enough to allow coverage of a large portion of the general element models, and yet they provide enough specificity to allow reasoning about the relationships of adjacent hypothesis regions. We quickly review the broad classes for each hypothesis element.

For this implementation, we limit the transfer function's parametric model to being piece-wise uniform over a surface. Our taxonomy then divides the transfer function into two classes: metals and dielectrics. Metals display only surface reflection, while dielectrics possess body reflection and possibly surface reflection as illustrated by Shafer [16].

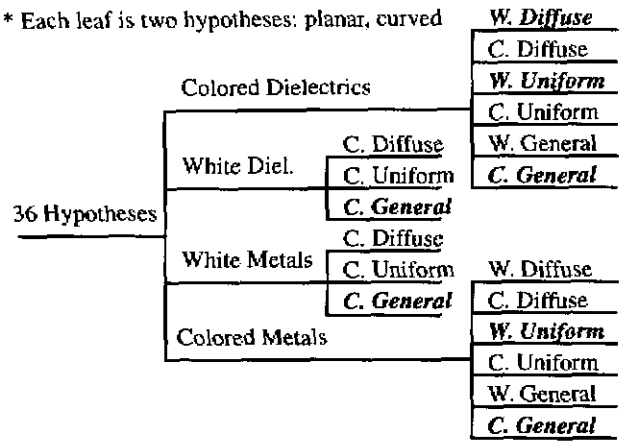


Figure 5 36 feasible combinations of the broad classes for colored regions. The 14 "common" hypotheses are bold-faced.

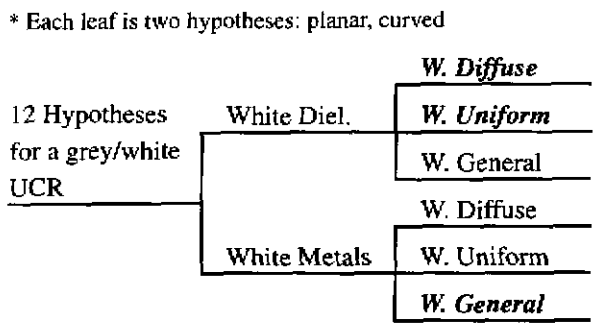


Figure 6 12 Fundamental hypotheses for a white/grey region. The 6 "common" hypotheses are bold-faced.

For illumination we identify three subspaces--in order of increasing complexity--which we term diffuse, uniform, and general illumination. The class diffuse illumination contains all illumination environments that have the same intensity and color from all directions. Diffuse illumination is a good approximation to objects in shadow or not directly lit [6]. The uniform illumination class contains all illumination environments whose representations are separable into a geometric component and a radiometric component, and whose geometric component takes on one of two values $\{1,\alpha\}$, where α could be 0. An example of uniform illumination is a point light source with ambient illumination of intensity α. Uniform illumination is a reasonable approximation of many man-made and natural light sources. All remaining possible illumination environments fall into the complex illumination class. We must include complex illumination because in some situations it is necessary to model illumination environments with both varying intensity and color (e.g., when interreflection is present).

We divide the shape into two subclasses--curved, and planar--because it separates the shape of the hypothesis into a highly constrained class (planar) and a more general class (curved). The planar class is highly constrained because it limits the number of free parameters for the surface patch to five: a unit vector and a point in 3-D space. It also strongly constrains the interaction of that hypothesis region with adjacent ones.

Finally, we divide both the illumination and transfer function classes into colored and white/grey subclasses. By then considering all possible combinations of the broad classes we get 2 x 6 x 4 = 48 possible hypotheses. Note that 12 of these hypotheses can only explain a white or grey region as they contain no colored elements. Therefore, we must consider at most 36 possible hypotheses for a colored image region.

Section 2.2. Fundamental Hypotheses

This set of 36 possible combinations of the broad classes we define to be the set of *fundamental hypotheses* for a colored region. The set of fundamental hypotheses are shown in Figure 5. For a given region, each of these hypotheses is a valid explanation for its appearance in the image.

The remaining set of 12 fundamental hypotheses, shown in Figure 6, explain white or grey regions of an image. (Note: it is possible for a white region to be the result of colored hypothesis elements if the illuminant and the transfer function have inverse spectral curves, but we assume this is rare and does not occur in our image set).

To denote a specific hypothesis we use the notation (<transfer function>, <illumination>, <shape>). The three elements of a hypotheses are defined as:
<transfer function> $\in$ {Colored dielectric, White dielectric, Col. metal, Grey metal},
<illumination> $\in$ {Col. diffuse, White diffuse, Col. uniform, White uniform, Col. complex, White complex}, and
<shape> $\in$ {Curved, Planar}.

Of the set of 36 fundamental hypotheses for a colored region, we select a smaller, but representative subset of 14

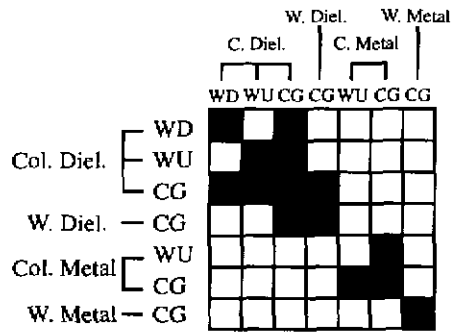


Figure 7 Table of desired hypothesis merges for colored regions

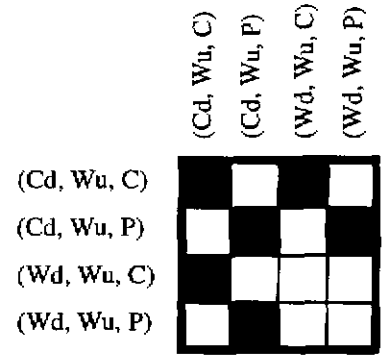


Figure 8 Desired mergers of implemented hypotheses

hypotheses, highlighted in Figure 5, to be considered as an initial set for each image region. The rules used to select these 14 hypotheses are:

1. If a subspace is both common and a good approximation of a larger encompassing space, include the subspace and exclude the larger space.
2. If a subspace is both uncommon and not a good approximation of a common larger space, exclude the subspace and include the larger space.

We can likewise select 6 of the 12 fundamental hypotheses for a white/grey region as highlighted in Figure 6.

Section 2.3. Merging the Fundamental Hypotheses

Using physical constraints and several rules, identified below, we can now create a table of all possible mergers of the subset of 14 colored hypotheses as shown in Figure 7. The key finding of this table is that it is sparse, strongly constraining which hypotheses can be merged and considered to be part of the same object.

The rules for merging are as follows.

1. For adjacent hypothesis regions to belong to the same object the discontinuity between them must be a simple one and *must involve only one of the hypothesis elements*.
2. Hypotheses of different materials should not be merged (including differently colored metals).
3. Hypotheses with incoherent shape boundaries should not be merged.
4. Hypotheses of differing color that propose the physical explanation to be colored metal under white illumination should not be merged.
5. Hypotheses proposing different color diffuse illumination should not be merged.

For more discussion on the models, taxonomy, and fundamental hypotheses, see [11].

Section 2.4. Implementation details

For our initial implementation of the segmentation method we consider the hypothesis set

H_c = {(Colored dielectric, White Uniform, Curved), (Colored dielectric, White uniform, Planar)}

for colored regions and the hypothesis set

H_w = {(White dielectric, White uniform, Curved), (White dielectric, White uniform, Planar)}

for white/grey regions. We are in the process of expanding the size of these initial hypothesis sets to include more of the fundamental hypotheses. Currently a region is labeled as white/grey if

$$(c_{nr} - 0.333)^2 + (c_{ng} - 0.333)^2 + (c_{nb} - 0x333)^2 < 0.0016,$$

where (c_{nr}, c_{ng}, c_{nb}) is the average normalized color of the region defined by equation (1).

$$(c_{nr}, c_{ng}, c_{nb}) = \left(\frac{r}{r+g+b}, \frac{g}{r+g+b}, \frac{b}{r+g+b}\right) \quad (1)$$

The threshold was set based upon the images in the test set. As the set of hypotheses considered in our current implementation all require white illumination, the exposure times for the different color bands were set so that a white board appeared white under the illumination used for the test images. This removed the need for color constancy and was found to be sufficient for white regions of the test objects to be classified as white using the above test.

Finally, for this implementation we only consider objects with piece-wise uniform transfer functions, such as the mug in Figure 3 and the objects in Figure 9 and Figure 11. Figure 8 shows all of the potential mergers of the hypotheses we implement.

Section 3. Initial segmentation

To test the segmentation method, we use simple pictures of multi-colored objects on a black background. Figure 9 and Figure 11 are two example test images. Figure 9 is a synthetic image created using Rayshade (a public domain ray tracer). Figure 11 was taken in the Calibrated Imaging Laboratory at Carnegie Mellon University. While obtaining the real image, an attempt was made to include examples of only the broad hypothesis classes used in this implementation.

The initial segmentation of images is accomplished using a simple region growing method with normalized color, defined by equation (1), as the descriptive characteristic. Because the segmentation method emphasizes discontinuities between hypothesis regions, the initial segmentation method uses local information to grow the regions and stops growing when it reaches discontinuities in the normalized color.

The algorithm traverses the image in scanline order looking for seed regions where the current pixel and all of its 8-connected neighbors have similar normalized color and none of these pixels already belong to another region or are too dark. When it finds such a seed region, it puts the current pixel on a stack and begins the region growing process. The growing algorithm is as follows.

1. Pull the top pixel off of the stack, make it the current pixel, and mark it in the region map as belonging to the current region (all pixels in the region map are initialized to the null region).
2. For each of the current pixel's 4-connected neighbors, if the neighbor's normalized color is close to the current pixel as specified by a threshold, and the neighbor is not part of another region nor is it too dark, then put it on the stack.
3. Repeat from 1 until the stack is empty.

When a region has finished growing, the search for another seed region continues until all pixels in the image have been checked. In the end, all pixels that are part of region are marked with their region id in the region map. All other pixels are either too dark, or are part of a discontinuity or rapidly changing region of the image. For now we simply ignore these pixels and concentrate on the found regions.

The dark threshold used on the test images was a pixel value of 35 (out of 255), and two pixels were found to have similar normalized colors if the Euclidean distance between the normalized colors was less than 0.3.

The overall goal of the initial segmentation algorithm is to find regions that can be considered part of the same object. By locally growing the image regions, some variation in the region color is allowed, but the regions do not grow through most discontinuities caused by variation in the transfer function or illumination. One problem with using normalized color as the growth parameter is that discontinuities in shape can be overlooked if the transfer function on both sides of an edge is the same. An example of this would be the edges of a uniformly colored cube. It is possible to compensate for this problem by using an edge detector or other filter which can identify intensity discontinuities prior to region growing. By not allowing regions to grow through intensity discontinuities, some shape discontinuities can also be identified in the initial segmentations.

Given the existence of more complex physics-based segmentation methods, a valid question is why not use a segmentation algorithm such as Healey's normalized color method [7], Klinker's linear and planar cluster algorithm [8], or Bajcsy et. al.'s normalized color method [1]? There are legitimate problems with using any of these methods. Hea-

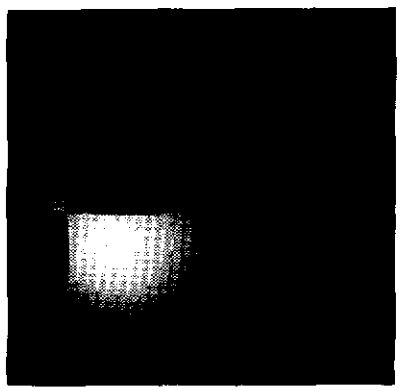


Figure 9 Synthetic test image of two spheres

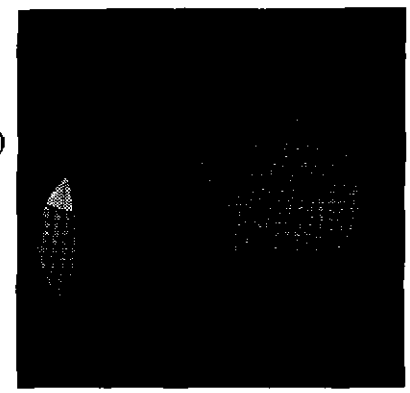


Figure 10 Initial segmentation of test image A

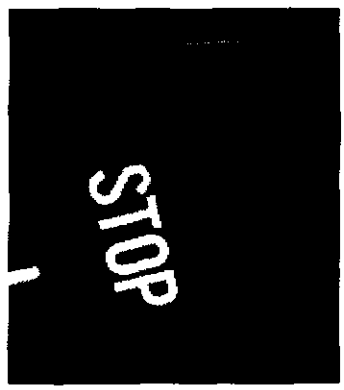


Figure 11 Real image of stop-sign and cup

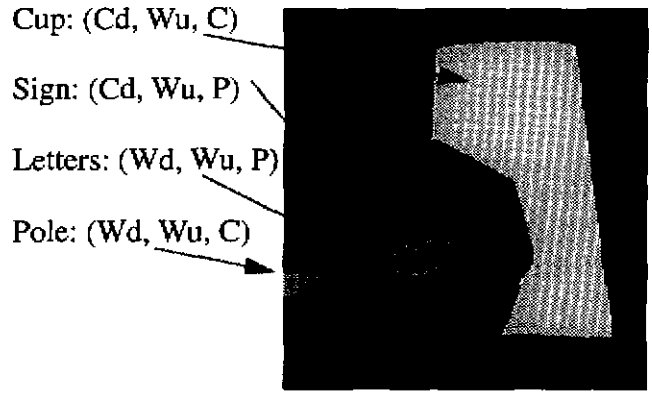


Figure 12 Initial segmentation of test image B

ley's normalized color method, while it does attempt to identify metals in an image, has two conflicts with our overall framework. First, it requires the entire scene to be illuminated by a single spectral power distribution. Interreflection, especially with respect to metals, confuses the algorithm. Second, white or grey dielectric objects can be confused for metal objects or highlights, again causing problems. We actually implemented Klinker's linear cluster algorithm and ran it on numerous test images. Two problems were found. First, without implementing all of Klinker's algorithm--which requires the assumption that all objects in a scene are dielectrics--variations in the normalized color due to highlights or noise are not well captured. Second, because of the need to find linear clusters, Klinker's algorithm breaks down on planar surfaces or regions of almost uniform color. Finally, although Bajcsy *et. al.*'s algorithm does allow identification of interreflections and shadows, it requires a white reference in the image with which to obtain the color of the illumination. We want to be able to segment images without the white reference patch or a white object.

Finally, we found that for this implementation and this set of test images the local normalized color segmentation alone was found to be fast and adequate. Figure 10 and Figure 12 show examples of the initial segmentations and are hand-labeled with the actual physical explanations.

Once the initial segmentation is completed, the four initial hypotheses are assigned to each region and the hypothesis merger process begins.

Section 4. Hypothesis Analysis

Overall, our segmentation algorithm proceeds as follows. First, we segment the image using the local normalized color algorithm described above. Then the set of initial (uninstantiated) hypotheses are assigned to each region. The next step analyzes all possible pairs of adjacent hypotheses to test if they are compatible. Finally, using the results of

this step we create a region graph with which we obtain the most likely final segmentations of the image.

Herein we identify two methods for proceeding with the analysis portion of the algorithm. The more obvious and direct method we call *direct instantiation*. This involves finding estimates of and representations for the specific shape, illumination environment, and transfer function for each region. By directly comparing the representations for two adjacent hypotheses, we obtain an estimate of how similar they are. An alternative method of analysis, *implicit instantiation*, does not attempt to directly model the hypotheses elements. Instead, as explained in section 4.2, we examine certain physical characteristics of adjacent regions that indirectly reflect the similarity of the hypothesis elements. We explore both of these alternatives and show that implicit instantiation, while less theoretically satisfying, is the more practical alternative.

Section 4.1. Direct Instantiation

If we can estimate and represent each hypothesis element, merging adjacent regions involves looking at the table in Figure 8 to find the possible mergers and then directly comparing the values of each hypothesis element. If the elements for two adjacent hypotheses h_1 and h_2 match according to a specified criteria, then the regions corresponding to these hypotheses should be considered part of the same object in any segmentation using h_1 and h_2. It is important to realize that other hypothesis pairs for the same two regions may not match.

While this approach is theoretically attractive, direct instantiation of hypotheses is difficult. We attempted to implement the direct instantiation approach for the hypotheses (Colored plastic, White Uniform illumination, Curved) and (White plastic, White Uniform illumination, Curved) for which some tools of analysis do exist for finding both the shape and illumination of a scene.

To directly instantiate the shape and illumination of the hypotheses, we implemented Bischel & Pentland's shape-from-shading [SFS] algorithm and Zheng and Chellappa's illuminant and albedo estimation algorithm [2] [19]. Bischel & Pentland's SFS algorithm was chosen because it is a local method, and, according to the survey by Zhang *et. al.*, it is the best local method when the illumination comes from the side [18]. A local SFS method is useful when analyzing small regions of an image because they need only the information contained in a small neighborhood around a given pixel to calculate depth. Zhang & Chellappa's illuminant estimator was selected because it is also a locally calculated method, and they showed their method produced better results than Pentland's or Lee & Rosenfeld's methods [14] [5] [19].

For this test, we represent the shape as a depth map, the illuminant as two angles (tilt and slant), and the transfer function as a normalized color vector. The tilt is defined as the angle the illuminant direction $\vec{L}$ makes with the x-z plane, and the slant is the angle between $\vec{L}$ and the z-axis.

The first step after the initial segmentation is to analyze each region independently. Figure 13 shows the results of SFS for the regions in the synthetic test image. For this image the illuminant and viewing directions are the same. The illuminant direction estimator was able to find the actual direction of the illumination independently for each region.

The second step is to compare the hypothesis elements of adjacent pairs. To compare the hypothesis shape of the regions, a two-step algorithm is employed. First, the optimal offset, in a least-squares sense, of the two regions is found by comparing the depth values of the two regions along the border and minimizing the square of the error between them. Second, using the optimal offset we find the sum-squared error of neighboring pixels along the border and use it to obtain the sample variance of neighboring pixels along the border.

To quantify the variance in the border pixels for a given region pair we first select a threshold variance for the surface depths by estimating the noise in the image. We then compare the variance due to noise with the sample variance using a chi-square test [9]. The chi-square test returns a probability that the error is due to noise in the depth map. This probability is an estimate of how well the region borders match. For example, if there is a 99% probability that the error is due to noise, then there is only a 1% probability that the error is due to a discontinuity in the shape of the regions. Figure 14 shows the sum squared error for each region pair in the synthetic test image. (We felt the sum-squared error was more informative in this case because the results of the chi-square test were probabilities of 1 for the small errors and 0 for the large errors for a wide range of standard variances.) For this image direct instantiation gives a clear indication of which regions' shapes match.

Comparing the illumination and transfer functions for this test case is trivial. The transfer functions are necessarily discontinuous at the borders because of the hypotheses being considered and the initial segmentation method. To

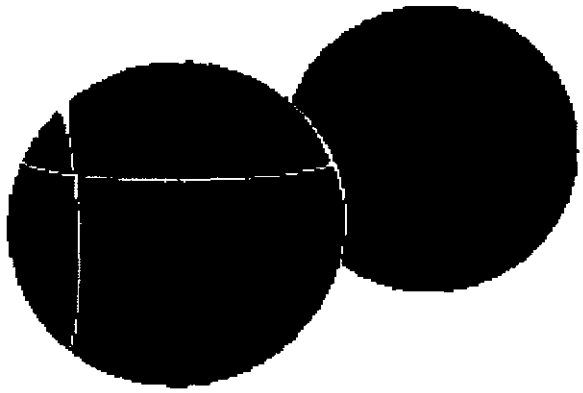


Figure 13 Shape from shading result. Displayed intensity decreases with depth.

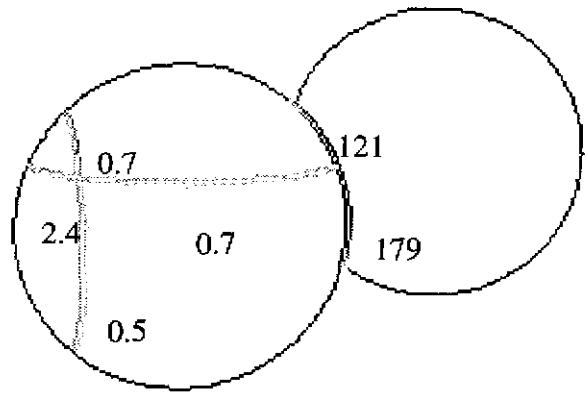


Figure 14 Border shape comparison. Darker borders indicate larger errors. Average sum-squared error per pixel shown for each region pair.

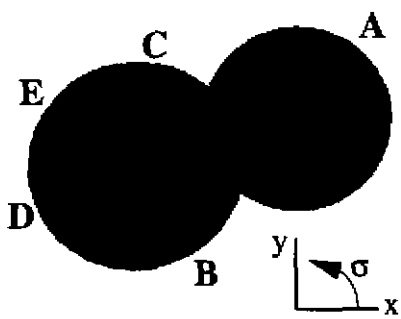


Figure 15 Synthetic test image with illumination from (tilt, slant) = (45°,27°)

Table 1 Illuminant direction estimation results

Region	Estimated Tilt	Estimated Slant	Tilt error	Slant error
A	44°	34°	-1°	7°
B	80°	27°	35°	0°
C	-57°	0°	-102°	-27°
D	20°	49°	-25°	22°
E	-20°	16°	-65°	-11°
All	46.6°	24.5°	1.6°	2.5°

compare the illuminant direction estimates of adjacent regions we convert the tilt and slant angles for each region to a 3-D vector and find the angle between the two vectors. For the synthetic test image the illuminant direction was correctly estimated for each region and the illumination was found to be the same for all region pairs. Thus, the results shown in Figure 14 are unchanged when the transfer function and illumination are considered.

As nicely as the direct instantiation method worked on the synthetic test image, the analysis tools were found to have serious problems with slightly more complicated images. First, Bischel & Pentland's SFS algorithm requires an accurate indication of the illuminant direction and albedo and also requires good initial point selection [18]. We found that small regions of an image (especially those corresponding to parts of an object) do not necessarily have good initial points, and depth maps generated for them do not correspond well with the actual shape except under certain conditions, namely, that the illuminant direction is such that there are maxima, or points close to a maxima, within the regions. Thus, despite Zhang *et. al.*'s claim as to the ability of Bischel & Pentland's SFS algorithm to handle illumination from the side, because of the maxima point problem the SFS algorithm was not able to deal with illumination that was not close (within 10°) to the viewing direction. For more general images, or real images such as the test image of the cup and stop-sign, the SFS algorithm breaks down because of the single point light source assumption and sensitivity to noise (a limitation also mentioned in [18]).

The second serious problem is with the illuminant direction estimator. Besides the assumption that the illumination is

a point source, Zhang & Chellappa's algorithm requires a good distribution of surface normals to correctly estimate the tilt and slant [19]. While this is a reasonable assumption for an entire image, it is not a valid assumption when analyzing small image regions, some of which are only part of a single object. What we found is that when the illumination is very close to the viewing direction, the illuminant estimator is better able to divine the correct direction because Zhang & Chellappa's slant estimator is dependent upon intensity variation rather than the distribution of gradients. However, for the test image in Figure 15 showing the two spheres illuminated from above and to the right, the illuminant estimator does not work as well.

Our conclusion from these experiments is that the basic problem with the direct instantiation method is that it requires region-based analysis. Existing tools for analyzing the intrinsic characteristics of a scene cannot, in general, be used on small regions of an image because it violates basic assumptions necessary for the tools to function properly. Furthermore, if we attempt to generalize direct instantiation to other hypotheses, we are currently limited by the lack of image analysis tools. While approaches to SFS like that of Breton *et. al.* [3], may overcome some of these difficulties in the future, for now we take a different approach.

Section 4.2. Implicit Instantiation

An alternative to direct instantiation of hypotheses is to use the knowledge constraints provided by the hypotheses to find physical characteristics that can differentiate between pairs of regions that are part of the same object and pairs of regions that are not. As these physical characteristics are generally local, they are more appropriate for region-based analysis than the previously mentioned direct-instantiation techniques. We call this method *implicit instantiation*.

Section 4.2.1. Reflectance Ratio

One physical characteristic we use is the reflectance ratio for nearby pixels as defined by Nayar and Bolle [12].

Consider two adjacent hypotheses h_1 and h_2 that both specify (Colored dielectric, White uniform, Curved). If h_1 and h_2 are part of the same piece-wise uniform object and have a different color, then the discontinuity at the border must be due to a change in the transfer function, and this change must be constant along the border between the two regions. Furthermore, along the border the two regions must share similar shape and illumination. If h_1 and h_2 belong to different objects, than the shape and illumination do not have to be the same.

The reflectance ratio is a measure of the difference in transfer function between two pixels that is invariant to illumination and shape so long as the latter two elements are similar. If the shape and illumination of two pixels p_1 and p_2 are similar, then the reflectance ratio, defined in equation (2), where I_1 and I_2 are the intensity values of pixels p_1 and p_2, reflects the change in albedo between the two pixels [12].

$$r = \left(\frac{I_1 - I_2}{I_1 + I_2} \right) \tag{2}$$

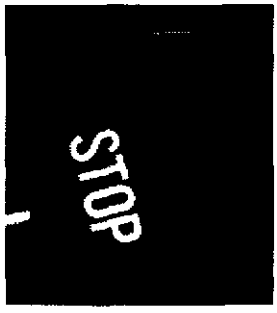
For each border pixel p_{1i} in h_1 that borders on h_2 we find the nearest pixel p_{2i} in h_2. If the regions belong to the same object, the reflectance ratio should be the same for all pixel pairs (p_{1i},p_{2i}) along the h_1,h_2 border. A simple measure of constancy is the variance of the reflectance ratio defined by

$$Var = \sum_{i=1}^{N} \frac{(r_i - r_{avg})}{N - 1} \tag{3}$$

where r_{avg} is the average reflectance ratio along the border and N is the number of border pixels. If h_1 and h_2 are part of the same object, this variance should be small, due mostly to the quantization of pixels and noise in the image and scene.

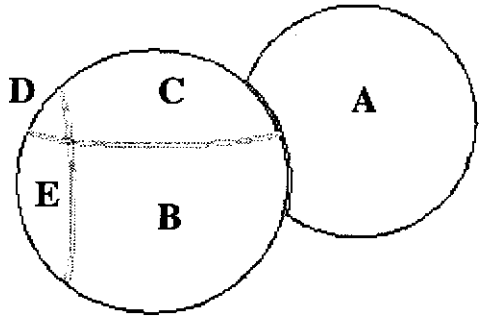
If, however, h_1 and h_2 are not part of the same object, then the illumination and shape are not guaranteed to be similar for each pixel pair, violating the specified conditions for the characteristic. This should result in a larger variance in the reflectance ratio. Ideally, we should be able to find a standard variance based upon the noise and quantization effects and use this standard variance to differentiate between these two cases. Table 2 shows the variances in the bor-

Table 2 Reflectance Ratio Results for Var_N = 0.004. The last column shows the probability that the variance is ≤ the variance due to noise.



Region A	Region B	Reflectance Ratio	Refl. Ratio Variance	$P(Var_R<Var_N)$
Red region	S region	.4463	.0004	1.0
Red region	T region	.4449	.0005	1.0
Red region	O region	.4503	.0004	1.0
Red region	P region	.4541	.0006	1.0
Red region	Cup region	.2107	.0125	0.0
O hole	O region	-.4358	.0008	1.0
P hole	P region	-.4562	.0004	1.0
While pole	Red region	.1709	.0710	0.0

Table 3 Results of Gradient Direction Comparison



Region 1	Region 2	Variance	$P(V_n > v)$
A	B	7.51	0.0
A	C	7.23	0.0
B	C	0.0812	1.0
B	E	0.0191	1.0
C	D	0.0326	0.998
D	E	0.0397	0.989

Figure 16 Result of gradient direction analysis. Darker borders indicate greater error. Last column of table shows the result of a chi-square test with Var_N =.2 radians.

der reflectance ratios of the region pairs for the test image of the stop-sign and cup. This example shows an order of magnitude difference in the reflectance ratio variances for region pairs that belong to the same object versus region pairs that do not.

As described previously, we can use a chi-squared test to compare the variance for a particular region pair to a standard variance based upon the noise and quantization error. The result of the chi-squared test is a probability that the variance in the reflectance ratio along the border is caused by noise and not by a change in the illumination or shape. While this test does not directly compare the shape and illumination of the two regions, the variance of the reflectance ratio along the border does implicitly measure their similarity.

The reflectance ratio can be used to compare several different hypothesis pairs as shown in Table 5.

Section 4.2.2. Gradient Direction

The direction of the gradient of image intensity can also be used in a similar manner to the reflectance ratio. The direction of the gradient is invariant to the transfer function for piece-wise uniform dielectric objects (except due to

border effects at region boundaries). Therefore, by comparing the gradient direction of border pixel pairs for two adjacent regions we obtain an estimate of the similarity of the shape and illumination.

To try and reduce noise in the gradient direction estimate caused by the discontinuity in the transfer function, the gradient direction for all pixels in the region except the border pixels is first calculated. We then grow the region by assigning to each border pixel the average gradient direction of its previously calculated neighbors.

As with the reflectance ratio, we sum the squared difference in the gradient directions of adjacent border pixels from two hypotheses to find the sample variance for each hypothesis pair and then use the chi-squared test to compare the sample variance to a threshold variance. Because of the conditions required for the gradient directions of adjacent borders to be similar, we interpret the result as a probability that the illumination and shape are similar along the border of the two regions.

Not surprisingly, the effectiveness of this characteristic is limited to regions with well-defined gradient directions. For planar or almost uniform surfaces with small gradients the angle of the gradient is very sensitive to noise and quantization errors.

An advantage the gradient direction has over the reflectance ratio is that it is not particularly sensitive to absolute magnitude. So long as the gradient is not small and the gradient direction can be accurately estimated, the absolute magnitude of a given pixel is irrelevant.

Figure 16 shows the results of applying the gradient direction characteristic to the synthetic test image.

Section 4.2.3. Intensity Profile Analysis

So far, we have examined only examined calculated characteristics of the image, not the actual image intensities. The intensity profiles contain a significant amount of information, however, which we attempt to exploit with the following assertion: if two hypotheses are part of the same object and the illumination and shape match at the boundary of the hypotheses, then, if the scale change due to the albedo difference is taken into account, the intensity profile along a scanline crossing both hypotheses should be continuous. Furthermore, we should be able to effectively represent the intensity profile across both regions with a single model. If two hypotheses are not part of the same object, however, then the intensity profile along a scanline containing both hypotheses should be discontinuous and two models should be necessary to effectively represent it.

To demonstrate this property, consider Figure 17, which shows the intensity profile for the scanline from A to A'. We can calculate the average reflectance ratio along the border to obtain the change in albedo between the two image regions. By multiplying the intensities from A" to A' by the average reflectance ratio we adjust for the difference in albedo. As a result, for this particular case the intensity profile becomes C^1 continuous. On the other hand, for the scanline B to B', the curves are not C^1 continuous even when the reflectance ratio is used to adjust the intensities.

Rather than use the first or second derivatives of the image intensities to find discontinuities in the intensity profiles, we take a more general approach which maximizes the amount of information used and is not as sensitive to noise in the image. Our method is based upon the following idea: if two hypotheses are part of the same object then it should require less information to describe the intensity profile for both regions with a single model than to describe the regions individually using two. We use the Minimum Description Length [MDL], as defined by Rissanen [15], to measure complexity, and we use polynomials of up to order 5 to approximate the intensity profiles. The formula we use to calculate the description length of a polynomial model is given in equation (4), where x^n is the data, θ is the set of model parameters, k is the number of model parameters, and n is the number of data points [15].

$$DL = -\log P(x^n|\theta) + \frac{k}{2}\log n \tag{4}$$

Our method is as follows.

1. Model the intensity profile on scanline s_0 for hypothesis h_1 as a polynomial. Use the MDL principle to find the best order polynomial (we stop looking after order 5). Assign to M_a the minimum description length for of the best polynomial found for h_1.
2. Model the intensity profile on scanline s_0 for hypothesis h_2 as a polynomial. Again, use the MDL principle to find

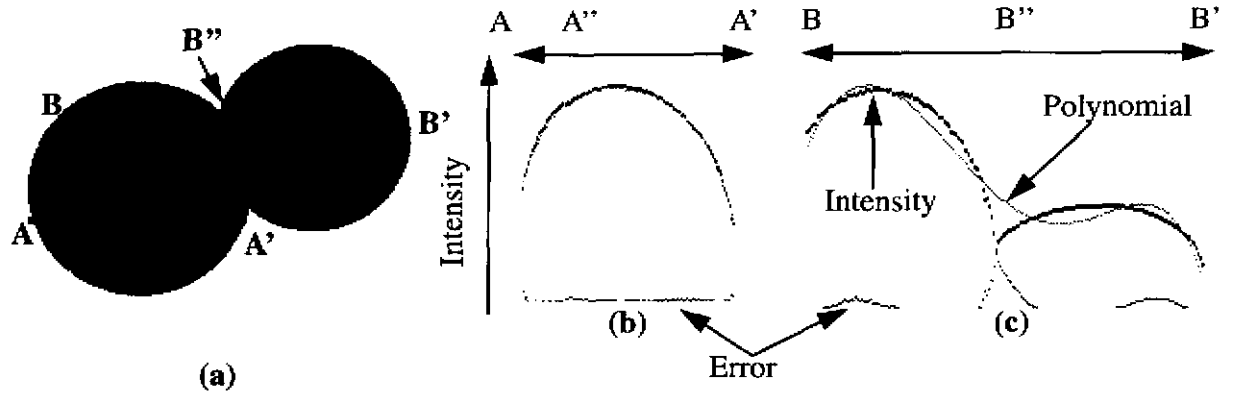


Figure 17 Test image shown in (a). Graphs (b) and (c) are the intensity profiles and least-squares polynomial for the image segments A-A' and B-B', respectively.

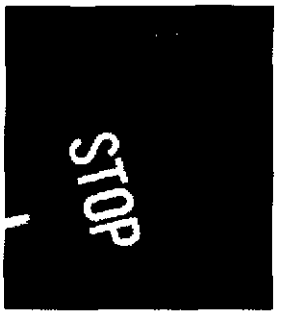


Table 4 Results of intensity profile analysis for stop-sign & cup image. If the far right column is close to or greater than 0, then the regions are better modeled by a single polynomial.

Region A	Region B	MDL A	MDL B	MDL C	A+B-C
Red region	S region	6.8	12.3	35.2	-16.1
Red region	T region	6.1	10.2	23.8	-7.8
Red region	O region	6.9	18.6	31.8	-6.2
Red region	P region	8.7	94.5	82.2	21.06
Red region	Cup region	9.7	6.8	56.9	-40.4
O hole	O region	5.2	6.4	9.1	2.4
P hole	P region	3.0	7.0	5.6	4.3
While pole	Red region	10.4	32.7	409.3	-366.2

the best order, and assign M_b be the minimum description length.

3. Model the scaled intensity profile of scanline s_0 for both h1 and h2 as a polynomial, and find the best order using MDL. Assign the smallest description length to M_c.

4. If $M_a + M_b \geq M_c$, according to an "equality" threshold ΔM, then we consider the two hypotheses to be part of the same object.

The result of this test is a merge/don't merge finding. For the purpose of integrating this result with the rest of the tests--each of which return a probability based upon a chi-square test--we represent a no-merge finding as a 5% probability, and a merge finding as a 95% probability that the two hypotheses are part of the same object.

Table 4 shows the results of this analysis applied to the stop-sign and cup test image. Note that a ΔM of 8 would represent an adequate threshold for correctly merging all but one region pair. For the synthetic image, a ΔM of 1.0 is sufficient for all region pairs. By using a more robust method for estimating the polynomials (such as least-median of squares), we believe a smaller ΔM could be used for all region pairs.

Table 5 Hypothesis Pairs and Their Tools of Analysis

Hypothesis 1	Hypothesis 2	Tools of Analysis
(C. dielectric, W. Uniform, Curved)	(C. dielectric, W. Uniform, Curved)	Reflectance Ratio, Gradient Direction, intensity analysis
(C. dielectric, W. Uniform, Curved)	(W. dielectric, W. Uniform, Curved)	Reflectance Ratio, Gradient Direction, intensity analysis
(C. dielectric, W. Uniform, Planar)	(C. dielectric, W. Uniform, Planar)	Reflectance Ratio, intensity analysis, border shape
(C. dielectric, W. Uniform, Planar)	(W. dielectric, W. Uniform, Planar)	Reflectance Ratio, intensity analysis, border shape

Section 5. Creating the Hypothesis Graph

We have seen that for the hypotheses used in our initial implementation we can use one or more tests to obtain an estimate of whether region pairs are part of the same object. Table 5 shows which tests can be used for which hypothesis pairs. Note, some of these tests (in particular, border shape) have not yet been implemented and are part of ongoing research.

How best to combine the results of different tests is still an open question. As shown previously, by estimating the population variances for the different analysis tests we obtain likelihoods that hypotheses should be merged. For our current implementation, if two or more tests are used to compare a hypothesis pair we use the average of the likelihoods of the results. How best to combine test results is still an issue of active research.

Once all possible hypothesis pairs are analyzed we generate a hypothesis graph in which each node is a hypothesis and edges connect all hypotheses that are adjacent in the image. We then assign to each edge the likelihood that the two hypotheses it connects are part of the same object. We use the results of the analysis tests to assign weights to edges that represent compatible hypotheses as specified by Figure 8. All other edges have a weight of 0.0, indicating that they should not be merged in any segmentation.

Note, however, that each edge actually has two weights associated with it. The weight assigned to the edge is a likelihood that the two hypotheses are part of the same object and should be merged in a segmentation. However, there always exists the alternative that the two hypotheses are not part of the same object and should not be merged in a segmentation. In order to find "good" segmentations, we must somehow assign a weight to the not-merge alternative.

We could define the likelihood that two connected hypotheses should not be merged as one minus the likelihood of a merger. This would present a quandary, however, as then the most likely segmentation of the image would be to select incompatible hypotheses for each region, resulting in a global likelihood of 1 (remember, incompatible hypotheses have a merge likelihood of 0). Therefore, that definition of the likelihood of not merging needs to be altered to allow merging at all!

For this implementation we turn once again to the principle of Minimum Description Length for guidance. Incompatible hypothesis pairs are different in at least two of the three elements, whereas compatible pairs differ by at most one element. When we merge two compatible hypotheses, we are in essence saying that we could represent the each of the two unchanging elements as a single model for both hypotheses. This is not unlike the intensity analysis described previously. Therefore, the cost of representing a segmentation where incompatible hypotheses are selected is greater than the cost of representing a segmentation where compatible hypotheses are used (so long as the tools of analysis return high likelihoods of a merger for the compatible hypotheses).

Because we use the indirect instantiation method, however, we do not have an accurate estimate of the representation costs or description length of any models we might use to represent the hypothesis elements. Instead, we select a value of 0.5 as the cost of not merging two hypotheses.

This value is selected for the following reason. Consider the situation shown in Figure 20. Hypothesis A for region 1

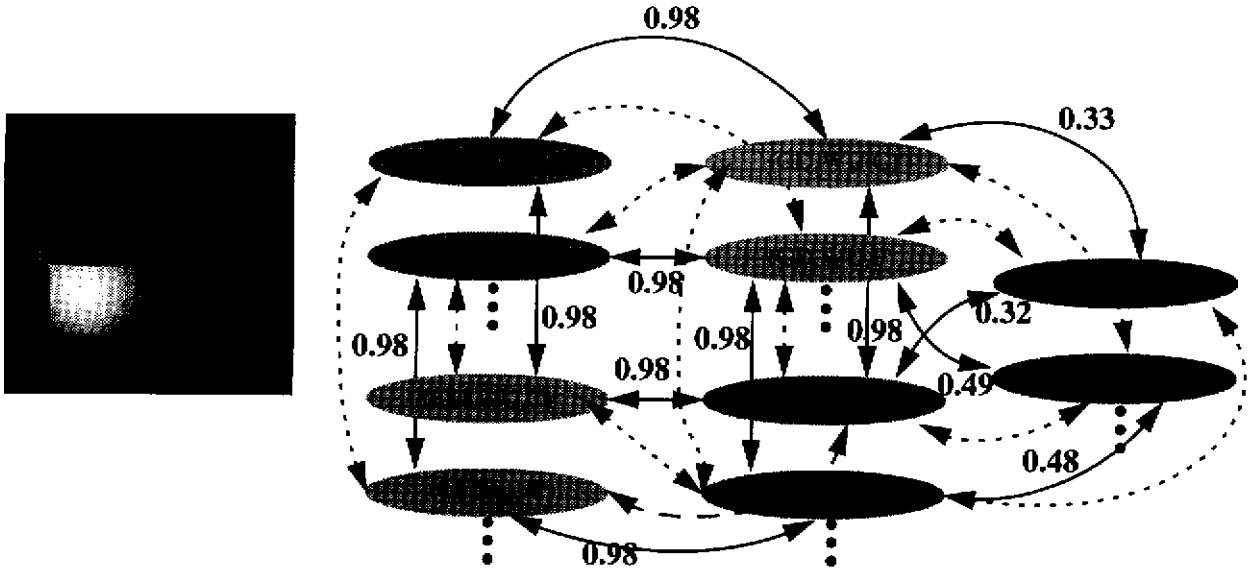


Figure 18 Two layer hypothesis graph for the synthetic test image. Dashed edges indicate incompatible hypotheses with a merge likelihood of 0, and a not-merge likelihood of 0.5. Note, as more hypotheses are included, the region graph simply gets more levels.

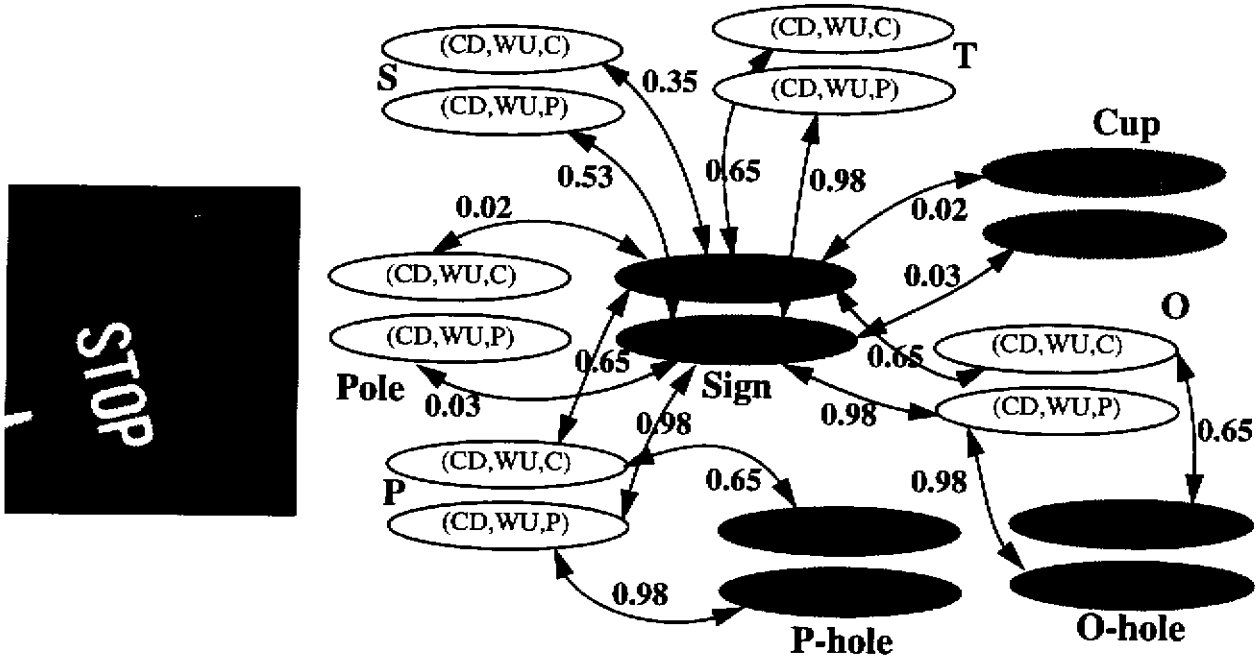


Figure 19 Two layer hypothesis graph for the stop-sign and cup image. Zero edges not shown. No edges exist between hypotheses for the same region.

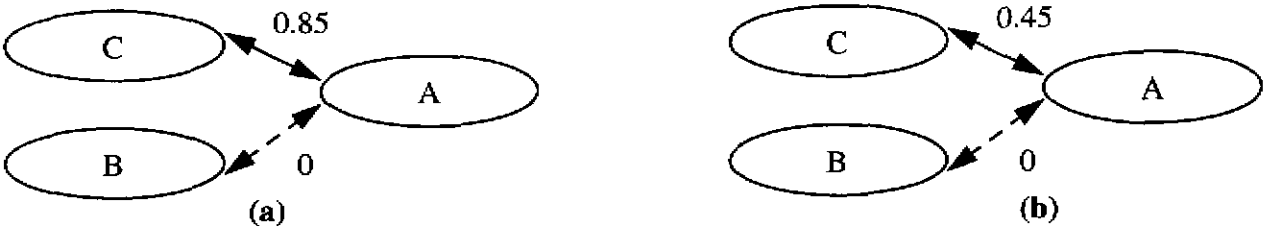


Figure 20 Potential hypothesis graphs. In (a) the best choice is to merge A and C. In (b) the best choice is to select incompatible hypotheses.

has to select the best hypothesis for region 2 with which to form a "best" segmentation of the image. Hypotheses A and C are compatible and have an edge weight of 0.85. This means it is better for hypotheses A and C to merge than not. Hypotheses A and B are incompatible. If the not merge probability is 0.5, then in Figure 20 (a) the segmentation A-C is the best. In the case shown in Figure 20 (b), because the merge likelihood of A and C is only.45, then hypotheses A and C are more likely to correspond to separate objects in the scene. This means that the segmentations A-B and A-C where neither pair are merged are better than the segmentation A-C where A and C are merged, and they have equal likelihoods of being true.

This is actually an interesting result because it reflects the actual situation. If we have a choice of two or more hypotheses for a single region in isolation, then, as discussed in the introduction, we cannot pick one hypotheses over another except by intuition and reasoning about the likelihood of certain conditions in the real world. However, when we can use the information contained in two hypotheses, as in the situation shown in Figure 20 (a) we can preferentially pick a segmentation because we are reducing the complexity of the scene. This is a powerful statement and is the essence of our approach to segmentation

The hypothesis graphs for Figure 9 and Figure 11 are shown in Figure 18 and Figure 19, respectively. The creation of hypothesis graphs is currently the extent of our implementation. The set of possible segmentations of the image given the complete hypothesis graph is the set of subgraphs such that each subgraph includes exactly one hypothesis from each region. We are currently researching methods for automatically obtaining a rank-ordered list of segmentations. As an example, we could use the hypotheses of a given region as the seed hypotheses for different segmentations of the image. We are guaranteed to get different segmentations because hypotheses for the same region cannot be part of the same segmentation.

Note that algorithms do exist for finding step-wise optimal segmentations of images given likelihoods that regions should be merged. LeValle and Hutchinson, and Panjwani and Healey have both used this algorithm to segment textured scenes [10] [13]. These algorithms would work unmodified on a single slice of a hypothesis graph (i.e. one hypothesis per region). A modification of this algorithm may be applicable to the hypothesis graphs we generate. The difference with previous applications is that the hypothesis graph created by our segmentation algorithm includes multiple hypotheses per region.

Section 6. Discussion

We conclude this paper with a brief discussion of the hypothesis graphs for our example images. For the synthetic image the compatible hypotheses for the four regions on the left sphere all have very high merge values. Conversely, the hypotheses for the right sphere have low merge values with those of the two adjacent regions of the left sphere. Therefore, the best segmentations will not merge the right sphere with the left sphere, but will merge the four regions of the left sphere. Because the values found for the planar-planar and curved-curved merges are very similar, there are four approximately equally likely segmentations for the image. The left sphere can be seen as a disk or a sphere, and the right sphere can be seen as a disk or a sphere, and the two possibilities combine with equal probability. Segmentations that divide the left sphere into planar and curved hypotheses are less likely than segmentations that do not divide it.

The hypothesis graph for the real image, however, gives a slightly more complex result. Because the gradient direction test is included in the tools for curved regions and not for planar regions, and this image includes planar regions, we get different results for the curved-curved and planar-planar hypothesis pairs for each pair of regions. The weights for the hypotheses show that the planar hypotheses for the stop-sign and letter regions are all more likely to be

merged than not. The weights also show that the cup and stop-sign regions, and the pole and the stop-sign regions are not likely to be merged for any hypothesis pairs. The interesting feature of this graph is that the weights for the curved-curved hypothesis pairs for the stop-sign and letter regions are lower than the planar-planar pairs for the same regions. Therefore, the best segmentations merge all of the stop-sign and letter planar hypotheses, and then select either planar or curved hypotheses for the cup and pole. This results in four equally likely "best" segmentations that *all* have the stop-sign as a single planar object.

Section 7. Conclusions and Future Work

Clearly, this is work in progress. However, even with only two hypotheses implemented we are able to segment images containing more complex objects than previous physics-based algorithms. Furthermore, the segmentation we generate more closely corresponds to the objects in the scene, something no other physics-based segmentation algorithm has attempted to date. Finally, the framework and algorithm are easily expandable and allow for greater complexity in images through the use of more hypotheses per region.

In order to expand the number of hypotheses per region, we are currently focusing on developing more tools for the analysis of hypothesis pairs. We are also working on automatic methods for obtaining segmentations from the hypothesis graph. As noted previously, the major challenge is dealing with multiple hypotheses per region. The other challenge is to find the n-best segmentations, not just the best. While "eyeballing" works for simple scenes and limited numbers of hypotheses, in the future, with more hypotheses per region and more complex images, having an automatic segmentation extractor will be critical.

References

[1] R. Bajcsy, S. W. Lee, and A. Leonardis, "Color image segmentation with detection of highlights and local illumination induced by inter-reflection," in *Proc. International Conference on Pattern Recognition*, Atlantic City, NJ, pp.785-790, 1990.

[2] M. Bichsel and A. P. Pentland, "A Simple Algorithm for Shape from Shading," in *Proceedings of IEEE Conference on Computer Vision and Pattern Recognition*, 1992, pp459-465.

[3] P. Breton, L. A. Iverson, M. S. Langer, S. W. Zucker, "Shading flows and scenel bundles: A new approach to shape from shading," in Computer Vision - European Conference on Computer Vision, May 1992, pp.135-150.

[4] M. H. Brill, "Image Segmentation by Object Color: A Unifying Framework and Connection to Color Constancy," *Journal of the Optical society of America* A 7(10), pp.2041-2047, 1990.

[5] M. J. Brooks and B. K. P. Horn, "Shape and Source from Shading," IJCAI, pp932-936, August 1985.

[6] J. D. Foley, A. van Dam, S. K. Feiner, J. F. Hughes, *Computer Graphics: Principles and Practice, 2nd edition*, Addison Wesley, Reading, MA, 1990.

[7] G. Healey, "Using color for geometry-insensitive segmentation," *Journal of the Optical Society of America* A 6(6), pp.920-937, June 1989.

[8] G. J. Klinker, S. A. Shafer and T. Kanade, "A Physical approach to color image understanding," *International Journal of Computer Vision*, 4(1), pp.7-38, 1990.

[9] L. Lapin, *Probability and Statistics for Modern Engineering*, Boston, PWS Engineering, 1983.

[10] S. M. LaValle, S. A. Hutchinson, "A Bayesian Segmentation Methodology for Parametric Image Models," Technical Report UIUC-BI-AI-RCV-93-06, University of Illinois at Urbana-Champaign Robotics/Computer Vision Series.

[11] B. A. Maxwell and S. A. Shafer, "A Framework for Segmentation Using Physical Models of Image Formation," in *Proceedings of Conference on Computer Vision and Pattern Recognition*, IEEE, pp361-368, 1994.

[12] S. K. Nayar and R. M. Bolle, "Reflectance Based Object Recognition," to appear in the *International Journal of Computer Vision*, 1995.

[13] D. Panjwani and G. Healey, "Results Using Random Field Models for the Segmentation of Color Images of Natural Scenes," in Proceedings of International Conference on Computer Vision, June 1995, pp.714-719.

[14] A. P. Pentland, "Finding the Illuminant Direction," *Journal of the Optical Society of America*, Vol. 72, No. 4, pp448-455, April 1982.

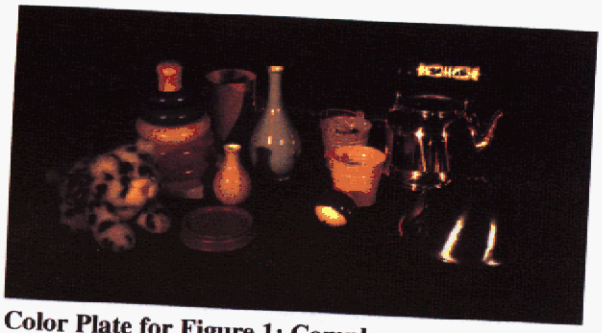
[15] J. Rissanen, *Stochastic Complexity in Statistical Inquiry*, Singapore, World Scientific Publishing Co. Pte. Ltd., 1989.

[16] S. A. Shafer, "Using Color to Separate Reflection Components," COLOR research and application, 10, pp.210-218, 1985.

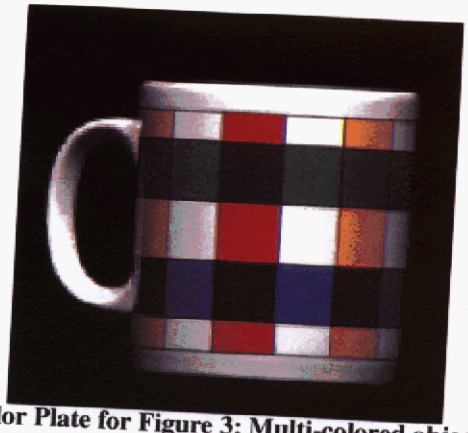
[17] J. M. Tenenbaum, M. A. Fischler, and H. G. Barrow, "Scene Modeling: A Structural Basis for Image Description," in *Image Modeling*, ed. Azriel Rosenfeld, New York, Academic Press, 1981.

[18] R. Zhang, P. S. Tsai, J. E. Cryer, M. Shah, "Analysis of Shape from Shading Techniques," in *Proceedings of IEEE Conference on Computer Vision and Pattern Recognition*, June 1994, pp.377-384.

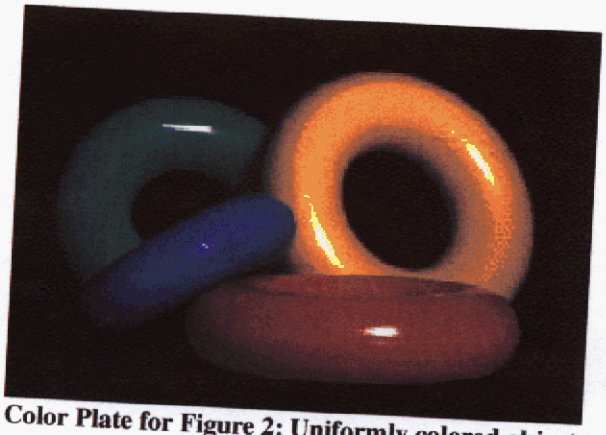
[19] Q. Zheng and R. Chellappa, "Estimation of Illuminant Direction, Albedo, and Shape form Shading," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 13, no. 7, July 1991, pp.680-702.



Color Plate for Figure 1: Complex scene containing multiple materials and multi-colored objects



Color Plate for Figure 3: Multi-colored object.



Color Plate for Figure 2: Uniformly colored objects.

Color Plate for Figure 4: Image of an object, a reflected image of the object, and a photograph of the object.

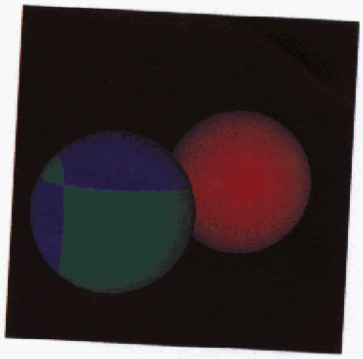
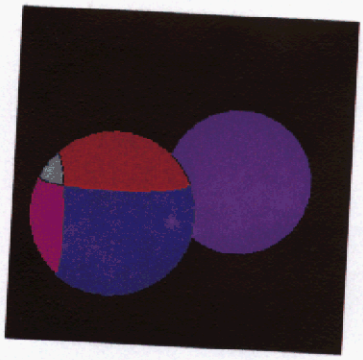
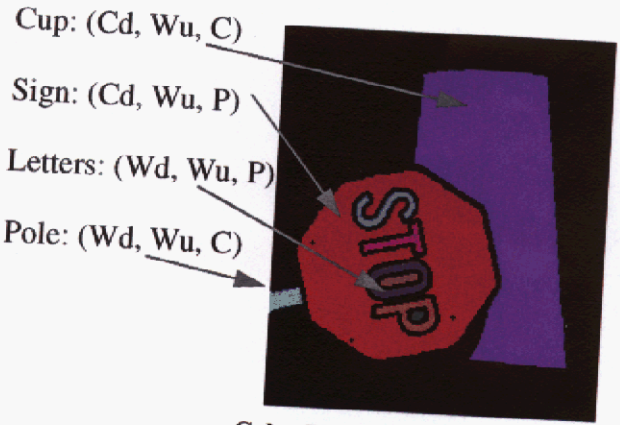


Figure 9 Synthetic test image of two spheres

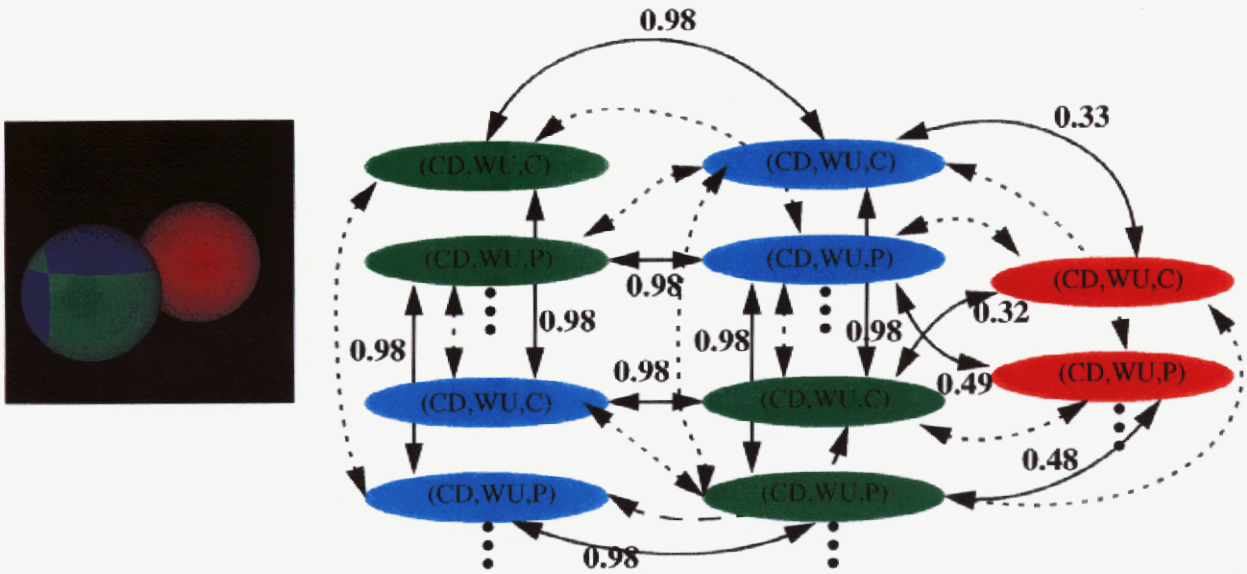


Color Plate for Figure 10: initial segmentation of test image A

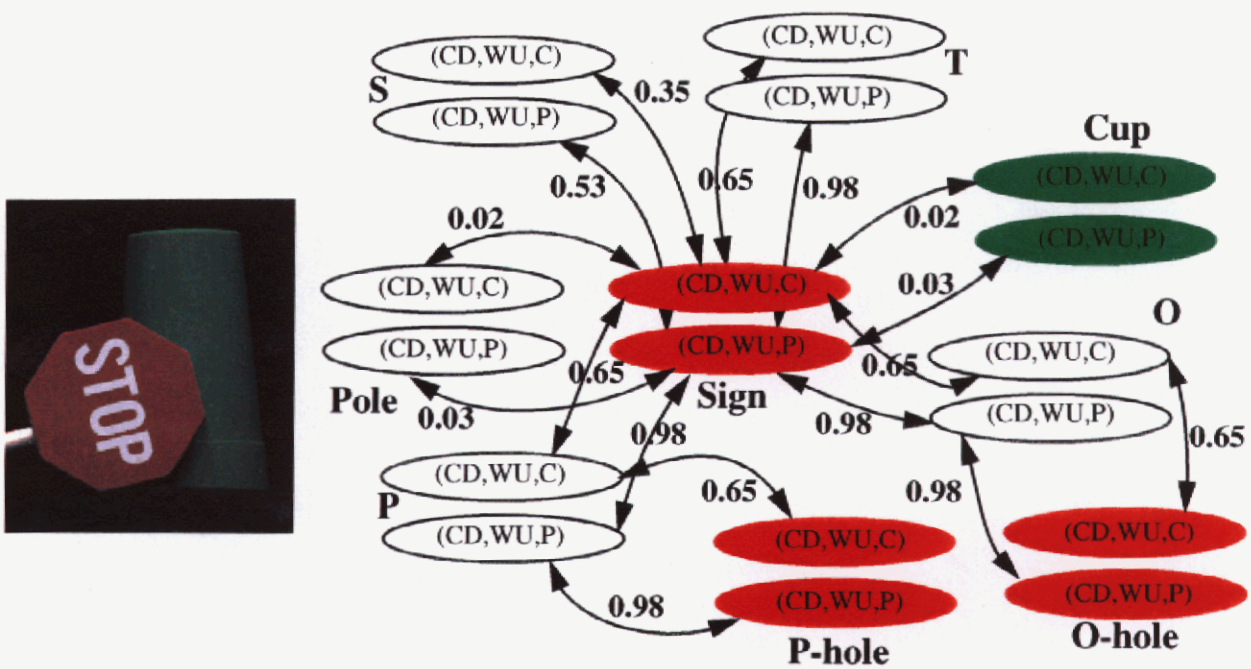
Color Plate for Figure 11: Real image of stop-sign and cup



Color Plate for Figure 12: Initial segmentation of test image B



Color Plate for Figure 18: Two layer hypothesis graph for the synthetic test image. Dashed edges indicate incompatible hypotheses with a merge likelihood of 0, and a not-merge likelihood of 0.5. Note, as more hypotheses are included, the region graph simply gets more levels.



Color Plate for Figure 19: Two layer hypothesis graph for the stop-sign and cup image. Zero edges not shown. No edges exist between hypotheses for the same region.