Image segmentation through reasoning about optics

Bruce A. Maxwell
University of North Dakota
Grand Forks, ND

Steven A. Shafer
Microsoft
Redmond, WA

Abstract

We have developed a general segmentation framework and algorithm that divides images of multi-colored piece-wise uniform dielectric objects into regions corresponding to coherent surfaces by reasoning about the underlying physics rather than color features or statistics. We present our algorithm and demonstrate its relevance to model acquisition and scene analysis.

Introduction

Solving the General Vision Problem is the Holy Grail of vision research. Broadly stated, the General Vision Problem is understanding a color image: the problem of identifying coherent surfaces or objects from a single image and explaining why they appear the way they do. A solution to the General Vision Problem would provide a description of objects and their incident illumination in an image that could contain multiple items of differing materials and shape, many displaying interreflection between themselves and their neighbors. Ultimately, a solution requires understanding the physical phenomena that create a given image.

In the past decade much of the vision community has moved away from the General Vision Problem and focused on more constrained images or image sets such as those used for stereo, motion, active vision, and photometric stereo. For tasks like robot navigation, obstacle avoidance, and object tracking analyzing single images is not only slower, but also less accurate than using multiple camera systems and multiple image algorithms such as stereo.

Understanding single images, however, is still essential for any task where an active agent or video imagery is not available. Searching image data-bases by content is one example of an important task whose input consists of only single images. Furthermore, for unlabeled images there is no information about the materials, illumination, or shape of the objects in these images. We are presented with only the image data. Therefore, we do not have the luxury of using multiple image algorithms, active vision techniques, or tightly controlled image environments. Instead, we must return to our study of the General Vision Problem and understanding single images.

One of the first, and most important tasks in single image analysis is segmentation: grouping pixels that appear to "belong" together. A segmentation provides regions of an image that can be reasoned about and analyzed as a whole. The prior segmentation of an image is a prerequisite for analysis methods such as shape-from-shading to work. This paper describes a framework and system for segmenting a single color image by reasoning about the physics of optics and image formation. Previous work in physics-based segmentation includes [1], [5], and [6].

Framework for Segmentation

Our model for a scene consists of three elements: surfaces S, illumination L^+, and the light transfer function $\Re$ 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 [14]. 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 simple image region. Variation in the color of an image region can be caused by changes in

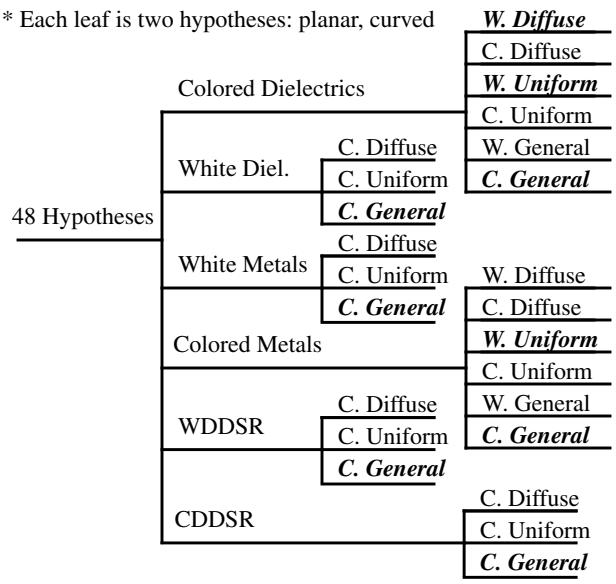

Figure 1. Taxonomy of fundamental hypotheses. Emphasized hypotheses are the most important.

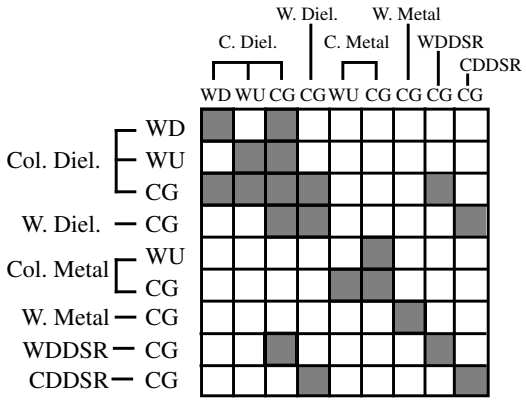

Figure 2. Table showing possible merges between hypotheses (shaded boxes). Other hypothesis merges are unlikely.

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 that extract information from single images--e.g., shape-from-shading and illuminant direction estimation--work because they assume the image variation is due to changes in only one element of the hypothesis (e.g. shape) [4].

Fundamental Hypotheses

In [10] 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 [2]. Instead, we divide the parameter space for each element 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.

The taxonomies developed for S, L^+, and $\Re$ allow us to identify sets of broad classes based upon partitions of the parameter space. In summary, the broad classes for each hypothesis element are:

- Surfaces = {planar, curved}
- Illumination Environment = {diffuse, uniform, general}
- Transfer Function = {metal, dielectric, dielectric displaying surface reflection}

There are eighteen possible combinations of these broad classes, subdividing the space of hypotheses for an image region into eighteen subspaces. Each of these subspaces is parameterized by the color values (wavelength spectrum) of the illumination and the transfer function.

Because of the large number of possible color distributions, to reason about hypotheses we subdivide L^+ and $\Re$ into two classes: uniform spectrum (white or grey), and non-uniform spectrum (colored). This divides L^+ into six forms of illumination, and $\Re$ into six forms of the transfer function. The possible combinations of surface, illumination, and transfer function are defined as the set of *fundamental hypotheses* for an image region.

We denote a specific fundamental hypothesis using the 3-tuple (<**transfer function**>, <**illumination**>, <**shape**>). The three elements of a hypotheses are defined as follows.

- **<transfer function>** := Colored dielectric | White dielectric | Colored dielectric displaying surface reflection | White dielectric displaying surface reflection | Col. metal | Grey metal
- **<illumination>** := Col. diffuse | White diffuse | Col. uniform | White uniform | Col. complex | White complex
- **<shape>** := Curved | Planar

Simple combination of the classes of the hypothesis elements (2 x 6 x 6) indicates there are 72 possible hypotheses. However, not all 72 are applicable to every region. Consider first a colored region. To possess color, either L^+ or $\Re$ must have a non-uniform spectrum. If we remove from consideration the 24 uniform illumination/uniform transfer function hypotheses, 48 fundamental hypotheses remain for a colored image region.

Conversely, the elements of the hypotheses for a grey or white image region must postulate no color. A situation where both the illumination and the transfer function are colored and yet their combination is grey is possible, but we assume this situation to be rare enough to neglect it for most images. This implies there are 24 fundamental hypotheses for a uniform spectrum region. Therefore, for a given image region we have to consider at most either 48 or 24 fundamental hypotheses, if we consider all of the possible explanations. The possible combinations of the broad classes we identified are shown in Figure 1 for colored regions. A similar chart exists for white/grey regions.

Of the set of 48 fundamental hypotheses for a colored region, we can select a smaller, but representative subset of 18 hypotheses, highlighted in Figure 1, 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 10 of the 24 fundamental hypotheses for a white/grey region, also highlighted in Figure 1. For a more extensive discussion of the generation and selection of hypotheses, see [10] or [9].

Merging the Fundamental Hypotheses

Using physical constraints and several rules, identified below, we create a table of all possible mergers of the subset of 14 colored hypotheses as shown in Figure 2. One 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.

Another result of our framework is that it allows us to reason about representations of objects without completely instantiating these representations. To obtain Figure 2, for example, we did not need to specify a representation of shape or illumination, nor did we have to specify exact values for a given representation. The broad classes alone allow sufficient reasoning about the hypothesis elements to obtain the table of potential merges.

Segmentation Algorithm

The segmentation algorithm divides into four parts, as shown in Figure 3. The first task is to find simple regions in the image. The objective of this step is to find regions that can reasonably be assumed to belong to a single object. The initial segmentation of images is accomplished using a simple region growing method with normalized color as the

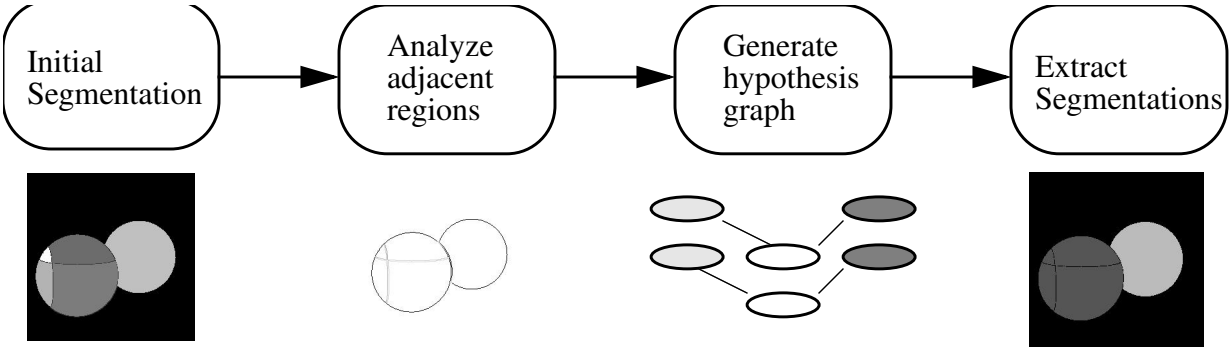

Figure 3. Algorithm outline: initial segmentation, analysis of adjacent region compatibility, hypothesis graph generation, segmentation extraction.

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.

Once these simple image regions are found, the algorithm attaches the initial list of uninstantiated hypotheses to each one. For the system described herein we use 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. These sets of hypotheses allow the system to handle multi-colored piece-wise uniform dielectrics with no specularities. For a discussion of including hypotheses proposing specular reflection, see [9].

The second task of the algorithm is to compare potentially compatible adjacent hypotheses according to the table in Figure 2. This step requires analysis of the hypothesis elements, necessitating some form of instantiation of the hypothesis elements. To make this determination, we turn once again to a physical analysis of the problem.

Our motivation for the techniques described below is the idea that it is easier to tell if regions *should not* be connected then if they *are* connected. To accomplish this we search for physical attributes of an object's appearance that have predictable relationships between different regions of the same object. By looking for these predictable relationships we can differentiate between regions that *may be* part of the same object and those that are not. We select physical characteristics that are local in nature, which are more appropriate for region-based analysis than global scene analysis methods such as shape-from-shading [10]. The three characteristics used to determine incompatibility are: reflectance ratio continuity, gradient direction continuity, and intensity continuity.

Reflectance Ratio

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. Nayar and Bolle developed the concept and have shown it to be effective for both segmentation and object recognition [11]. The reflectance ratio was originally defined for intensity images and measures the ratio in albedo between two points. The albedo of a point is the percentage of light reflected by the surface.

The principle underlying the reflectance ratio is that two nearby points in an image are likely to be nearby points in the scene. Therefore, they most likely possess similar illumination environments and geometric characteristics as shown in Figure 4(a). Nayar and Bolle show that multiple light sources do not affect the reflectance ratio so long as the assumption of similar geometries and illumination environments holds [11]. A well-behaved version of the reflectance ratio, proposed by Nayar and Bolle, is the difference in the intensities divided by their sum, as in (1) [11].

$$r = \left(\frac{I_1 - I_2}{I_1 + I_2}\right) \tag{1}$$

Given this measure of the difference between the transfer functions of adjacent pixels, how do we use it to measure hypothesis compatibility? The basic idea is to look for constant reflectance ratios along region boundaries. If the

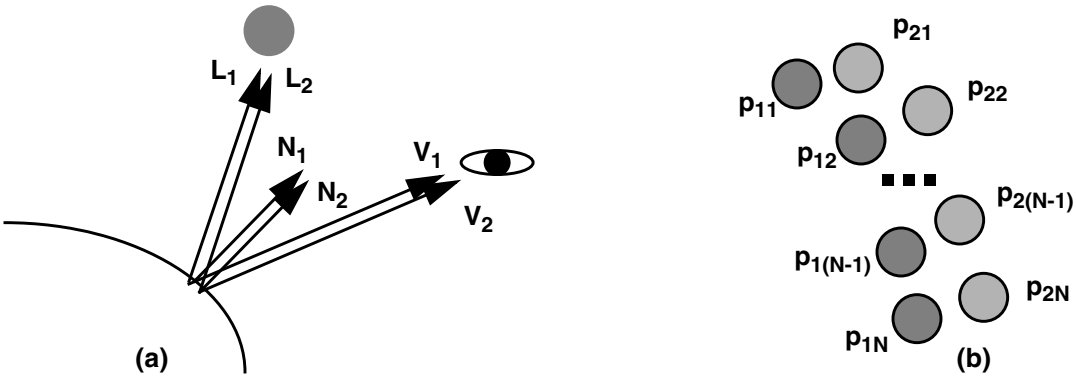

Figure 4: (a) Two nearby points. Note that the geometry is approximately the same for both points. The ratio of the intensities is, therefore, a function of the albedos of the two points.(b) Pixel pairs along the border of two regions. Two regions that are part of the same object should have a constant reflectance ratio along their boundary.

reflectance ratio along the boundary connecting two regions of different intensity is not constant, then either the shape or illumination are incompatible in addition to the transfer function. For the details of the algorithm, see [9].

Gradient Direction

The direction of the gradient of image intensity is another characteristic that reflects the geometry and illumination of a scene. For piece-wise uniform objects, the image gradient direction is invariant to the transfer function except at region boundaries. Therefore, the similarity in the gradient direction along the borders of two adjacent regions gives us a measure of the similarity of the shape and illumination of the corresponding surface patches in the scene. This property of the gradient direction, which is perpendicular to the flow-field as defined by Breton & Zucker, was recently shown by them to be a good indicator of surface continuity when used in conjunction with an albedo estimation process such as the reflectance ratio characteristic previously described [3].

One drawback to using the gradient direction is its sensitivity to noise and the effects of the region boundaries. Furthermore, pixels with small gradient intensities do not have reliable gradient directions as small changes in the intensity can cause large changes in the gradient direction. These problems make it a less robust measure overall than the reflectance ratio. However, with appropriate precautions it makes an effective compatibility test. For details, see [9].

Intensity Continuity

Both the reflectance ratio and the gradient direction compatibility tests share one drawback: they look only at border pixels to compare two regions. The bodies of the two regions contain a lot of information about their relative geometry and illumination characteristics. Intuitively, when we look at a multi-colored object we do more than look at points near the borders; we also look at the overall smoothness, or continuity of the shading patterns, automatically normalizing for changes in the illumination and transfer function. This intuition motivates the profile analysis compatibility test.

The test works on the following assertion: if two adjacent hypothesis regions are part of the same surface, then, if we take into account the brightness scale change due to the change in transfer function, the intensity profiles--for example, along a single scanline--of the two regions should be smooth and continuous according to some criteria. As with the other two tests, we cannot claim that hypothesis region pairs that are not part of the same surface will not be smooth and continuous.

Rather than look for discontinuities or breaks in the intensity profiles, we take a more general approach that maximizes the amount of information we use. We can think about the previous assertion from the standpoint of fitting a smooth model to the intensities of each hypothesis region. If they are part of the same surface, then a single profile model should adequately fit the intensities across both hypothesis regions. On the other hand, if their combined profile contains discontinuities or they have differently shaped profiles, then a single model will not adequately fit the data. In this case it would be better to use two models. We can use an information criterion such as the Minimum Description Length (see [13]), to test when it is more costly to use one model to fit the data, rather than two. For

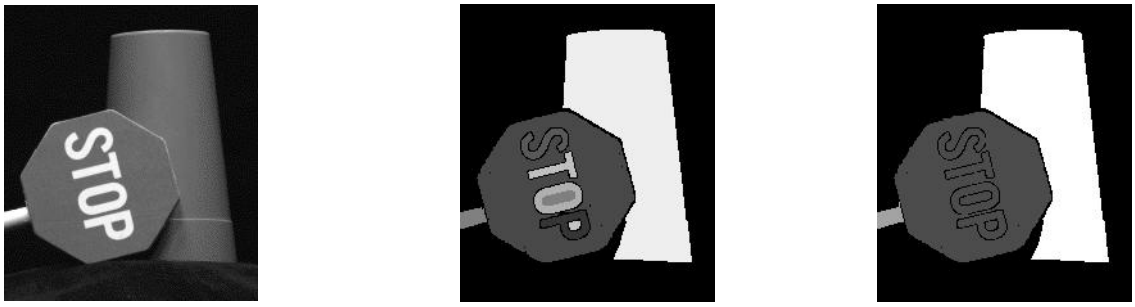

Figure 5. Left image: Original image taken in Calibrated Imaging Laboratory. Center image: initial segmentation into simple image regions according to color. Right image: final segmentation where the algorithm has merged the compatible regions.

details, see [9].

The Hypothesis Graph and Segmentation Extraction

The third task, generation of the hypothesis graph, uses the results of the analysis to generate a multi-layered graph. Each hypothesis is a node, and each edge is assigned a value based upon the results of the previous stage. For this implementation, we use a weighted average of the results of the three tests previously described.

The final task is to extract the best segmentations from the hypothesis graph. It is likely that there will be more than one "best" segmentation of the image, which reflects the naturally existing ambiguity in single image understanding. To extract globally likely segmentations from the hypothesis graph we use a modified version of the step-wise optimal algorithm used in [12] and [8] for texture segmentation. The algorithm is modified to work with multi-layer graphs (see [9] for details). Figure 5 shows a test image, the initial segmentation into simple regions, and the result of the algorithm on a test image of a wooden stop-sign and plastic cup. Note that the stop-sign and its letters form one region, while the cup and pole form separate regions. This segmentation conforms more closely to the objects in the scene, and we obtain this segmentation by examining the physics of adjacent regions and using physical relationships as the basis for our comparisons of compatibility.

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] 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.

[3] P. Breton and S. W. Zucker, "Shadows and Shading Flow Fields," in *Proceedings of IEEE Conference on Computer Vision and Pattern Recognition*, June, 1996, pp. 782-789.

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

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

[6] 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.

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

[8] 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.

[9] B. A. Maxwell, *Segmentation and Interpretation Using Multiple Physical Hypotheses of Image Formation*, Ph.D. thesis, Carnegie Mellon University, CMU-RI-TR-96-28.

[10] 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.

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

[12] 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.

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

[14] 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.