A Framework for Segmentation Using Physical Models of Image Formation

Bruce A. Maxwell and Steven A. Shafer

The Robotics Institute, Carnegie Mellon University
Pittsburgh, PA 15213

color slides available via mosaic: http://www.cs.cmu.edu:8001/afs/cs/user/maxwell/ftp/vision/cvpr94_slides.html

Abstract

This paper presents a new approach to segmentation using explicit hypotheses about the physics that creates images. We propose an initial segmentation that identifies image regions exhibiting constant color, but possibly varying intensity. For each region, we propose a set of hypotheses, each of which specifically models the illumination, reflectance, and shape of the 3-D patch which caused that region. Each hypothesis represents a distinct, plausible explanation for the color and intensity variation of that patch. Hypotheses for adjacent patches can be compared for similarity and merged when appropriate, resulting in more global hypotheses which group elementary regions.

1. Introduction

The goal of physics-based segmentation is to find image regions that correspond to semantic scene elements. In practical terms, this means finding one or more physical descriptions of the illumination, materials, and geometry that created the image so that subdivisions of the image correspond to objects in the scene. In this presentation, we focus upon the problem of segmenting a single color image. That a solution exists for humans is obvious: an individual can look at a picture such as Figure 1 and provide a fairly detailed physical description of the objects in the scene. We believe that postulating such a physical description is the key to understanding image data.

One of the key steps towards achieving a physical understanding of images was Shafer's dichromatic reflection model [18]. It allowed researchers to begin looking at a large class of actual materials: inhomogeneous dielectrics (paints, plastics, acrylics, ceramics, and paper). Klinker *et al*. [10] demonstrated the power of this model, and the physics-based vision approach, by using it in tandem with a model for noise and camera effects to segment real images of inhomogeneous dielectrics.

Healey [7] subsequently proposed the unichromatic reflection model for metals, and showed that it could be used with the dichromatic reflection model to segment images with both metals and inhomogeneous dielectrics under specific lighting conditions. In parallel with this work, the computer vision community has looked into determining light source color [12], and continued to work on determining shape, although mostly with range data (e.g. see [4]).

Figure 1 A complex scene of many materials

Figure 2 An object, its mirror image, and a photograph of it

Recently, Breton *et al.* [3] have combined shape, light source direction, and material consistency into a single segmentation system. They propose families of models for light source direction and shape, but they assume a single model--Lambertian--for the reflectance properties of the material.

Unfortunately, no current system can deal with a picture such as Figure 1. It contains grey and colored metals reflecting multi-colored illumination, and numerous dielectrics with differing reflectance properties. In order to

begin to understand general images such as this, we must consider families of possible models for all three elements of a scene--illumination, reflectance, and shape or geometry.

Selecting among multiple physical hypotheses has recently been examined by Breton *et al.*, but only for illumination and shape (they assume Lambertian surfaces). Model selection is necessary because of inherent ambiguity in an image. As seen in Figure 2, there can be several different explanations for identical image regions. But what are the general models we should use? What are the parameters of the model classes we need to consider, and do we need to consider them all? If not, how do we choose an initial set of models, and how do they merge and interact? These are the questions we consider in this paper.

2. The elements of a scene

The elements constituting our model of a scene are surfaces, illumination, and material optics. These elements can be thought of as the *intrinsic characteristics* of a scene, as opposed to *image* features such as edges or regions of constant color [19]. We begin by providing a formal notation for each of these elements.

2.1. Surfaces

We model objects in the real world using 2-D manifolds we call *surfaces*. On a given surface, we can define local coordinates as a two-variable parameterization (u, v) relative to an arbitrary origin. The shape of the manifold in 3-D space is specified by a *surface embedding function* $S(u, v) \rightarrow (x, y, z)$, defined over an extent $E \subseteq (u, v)$. The surface embedding function maps a point in the local coordinates of the manifold to a point in 3-D global coordinates. This global coordinate system is also anchored to an arbitrary origin, often specified relative to an imaging device. The surface embedding allows us to define a tangent plane $T(u, v)$ and surface normal $N(u, v)$ at each point on the manifold, and thereby to define a local 3-D coordinate system at each surface point with one axis the surface normal and two axes on the tangent plane.

Throughout this paper, to specify direction in the global coordinate frame we use the angle pair (θ_x,θ_y), which specifies the angle from the *x*-axis and *y*-axis respectively. In the local coordinate frame we use polar coordinates (θ,φ), which specify the angle from the surface normal (θ) and the angle from the *u*-axis to the projection of the direction vector on the tangent plane (φ).

2.2. Illumination

In a scene, light is being emitted or reflected in numerous directions, entering and leaving points throughout the area of interest. We can parameterize a single ray of light at time *t* at position (x, y, z), moving in direction (θ_x, θ_y), of frequency λ and polarization *s* (specifying the set of Stokes parameters [2]), by the 8-tuple $(x, y, z, \theta_x, \theta_y, \lambda, s, t)$.

For the purposes of image formation, we want to specify the intensity of visible light that is incident from all directions on points (x, y, z) in global 3-D coordinates. We describe this light by defining the *incident light energy field function* $L^+(x, y, z, \theta_x, \theta_y, \lambda, s, t)$, which specifies the radiant intensity, or radiance per unit solid angle, of light incoming to the point (x,y,z) from direction (θ_x, θ_y) of wavelength λ and Stokes parameter *s* at time *t*. This function is similar to the *plenoptic function* defined in [1], or the *helios function* [14]. In this paper we consider only single pictures taken at time *t*, making time a constant and allowing us to drop it from our parameterization of illumination functions and use the subspace of the incident light energy field $L^+(x, y, z, \theta_x, \theta_y, \lambda, s)$.

To obtain a local representation of illumination, if we substitute the local surface coordinates (u, v) for the global coordinates (x,y,z), and the local spherical coordinates (θ, φ) for the global axis angles, we obtain the *local incident light energy field* $L^+(u, v, \theta, \varphi, \lambda, s)$. Note that the global and local illumination functions are distinguished by their parameters.

2.3. Reflectance and the light transfer function

In order for a point on a surface to be visible to an imaging system, there must be some emission of light from that point. As with the incident light energy field, we are interested in describing the light energy that is leaving a surface point (x, y, z) in every direction (θ_x, θ_y) in polarization state *s* for every wavelength λ. The light leaving a point is specified by the *exitant light energy field* $L^-(x, y, z, \theta_x, \theta_y, s, \lambda)$. This function has the same form as the incident light energy field, describing an intensity for every direction and wavelength. The local coordinate version of the exitant light energy field is specified by $L^-(u, v, \theta, \varphi, s, \lambda)$.

The relationship between the incident and exitant light energy fields depends upon the macroscopic, microscopic, and atomic characteristics of the given point. It is the gross characteristics of this relationship that allow people to identify and describe surfaces in a scene. Intuitively, the reflectance, or transfer function specifies the exitant light energy field given a unit of incident light energy. Formally we represent the reflectance, or *global light transfer function* by $\Re(x, y, z;\theta_x^+, \theta_y^+, s^+, \lambda^+;\theta^-x, \theta_y^-, s^-, \lambda^-;t)$ which indicates the exitant light energy field $L^-(x, y, z, \theta_x, \theta_y, s, \lambda)$ produced by one unit of incident

light from direction (θ_x^+, θ_y^+), of polarization s^+, and wavelength λ^+ for a particular surface point (x, y, z) at time *t*. To allow us to drop time from the parameterization, we assume surfaces whose transfer functions do not change. An alternative form of the light transfer function can be obtained by substituting the local coordinates (u, v, θ, φ) for the global parameters $(x, y, z, \theta_x, \theta_y)$ resulting in the *local light transfer function* $\Re(u, v; \theta^+, \varphi^+, s^+, \lambda^+; \theta^-, \varphi^-, s^-, \lambda^-)$.

By setting constraints on the general parameters, it is possible to specify common special cases of the light transfer function. Special cases include: fluorescence, polarizing materials, transparency, surface or specular reflection, and Lambertian reflection. For a more detailed discussion, see [13].

2.4. General hypotheses

Having defined a 3-D world model for individual points and their optical properties, we introduce a nomenclature for describing the aggregation of appearance properties in the 3-D world and how these aggregations map to an image.

We begin by defining the combination of a surface and a transfer function to be a *surface patch*. We can model the coherence of an object's appearance with a surface patch whose transfer function is similarly coherent. Note that coherence does not imply uniformity, and covers a broad scope of possible aggregations including uniformity and repetitive or irregular textures.

A surface patch with a coherent transfer function, however, will not always display the coherence in an image. Uneven illumination or occluding objects can mask or modify the appearance of the patch to an imaging system. For the purposes of image analysis, we would like to specify not only coherence in the transfer function, but coherence in the exitant light energy field. To describe this, we define an *appearance patch* as a surface patch whose points possess a coherent transfer function *and* incident light energy field, and whose exitant light energy field exhibits a coherence over the entire patch and is not occluded from the imaging system.

Given an appearance patch, we can imagine that the exitant light energy field over the patch maps to a set of pixels in the image. The physical explanation for an appearance patch we define to be a *hypothesis* $H = \langle S, E, \Re, L^+ \rangle$. The four elements of a hypothesis are the surface embedding S, the surface extent E, the transfer function $\Re$, and the incident light energy field L^+. With these functions, it is possible to completely determine the exitant light energy field (assuming no self-luminance). The basic connection between a physical explanation and a group of image pixels is provided by a *hypothesis region* $HR = \langle P, H \rangle$, defined as a set of pixels P that are the image of the hypothesis H. The combination of the hypothesis elements represents an explanation for the color and brightness of every pixel in the image region. To represent the fact that a single region may have more than one possible explanation, we define a *hypothesis set* $HS = \langle P, H_1, \ldots, H_n \rangle$ to be a set of pixels *P* with an associated list of hypotheses $H_1, \ldots, H_n$, where each hypothesis H_i provides a unique explanation for all of the pixels in *P*, and only the pixels in *P*.

Finally, given a set $\{HS_i\}$ of hypothesis sets for pixel regions P_i, we define a *segmentation* of the pixel set $P = \bigcup_i P_i$ to be a set of hypotheses, containing one hypothesis from each HS_i, that explains the values of the pixels in *P*. Of course, to be physically realizable, these hypotheses must be mutually consistent. The goal of low-level vision, in terms of our vocabulary, is to produce one or more segmentations of the entire image.

3. Fundamental hypotheses

The difficulty inherent in segmentation using physical descriptions lies in determining the correct mapping between the image pixels and the scene that created them. The segmentation process involves identifying which sets of pixels correspond to which appearance patches, identifying the possible physical explanations for those patches, and then merging them with other appearance patches when their physical explanations are compatible in an identifiable fashion.

This approach to segmentation is not new--for example, Klinker *et al.* [10] and Healey [7] both identified regions of similarity of some physical properties. What is new in this presentation is the generality. These past works assumed that the scene obeyed very specific properties and looked only for a single, narrowly defined kind of coherence. In our new approach, the general illumination and transfer functions allow us to represent, reason about, and discover many different kinds of coherence in a single image. This capability is necessary for the analysis of natural or common man-made scenes such as Figure 1.

3.1. Pixel classification

The first step in segmentation is to identify pixel regions that display coherence in some feature space. In a color image, the most obvious characteristic linking together groups of pixels is their color. The simplest such groupings are aggregates of pixels with identical color. A reasonable starting assumption might be that a set of connected pixels with the same color corresponds to a single appearance patch within a scene. We believe, however, that using regions of uniform color overlooks much of the

information contained in the image.

An approach of slightly greater complexity is to group together pixels into *uniform chromaticity region*s [UCRs] which we define to be a connected set of pixels that possess a linear relationship in color space. This approach was also used by Klinker *et al*., who termed such regions *linear clusters*. Note that under common white illumination this approach groups pixels with the same chromaticity, or normalized color coordinates as defined in [9].

Klinker *et al.* observe that a UCR can represent two distinct objects if both are dark or poorly illuminated. In this segmentation method, however, we initially assume that a UCR represents an appearance patch under a single incident light energy field. This requires a form of coherence from the physical elements generating the UCR. Clearly, it is possible to construct an image with UCRs that do not have such coherence in the physical world, and we realize that our current approach will not correctly handle such situations.

The benefit derived by using UCRs is that they are groupings of pixels that we can reasonably assume to correspond to a single appearance patch in the physical world. The next step is to identify the relevant physical explanations, or hypotheses, for an appearance patch corresponding to a UCR.

3.2. Generating Hypotheses

The first question we examine is: how many physical descriptions must be considered? We begin to answer this question by noting that a UCR has two characteristics that make it interesting: it is not necessarily white, and it is not necessarily uniform in intensity. Any hypothesis that explains a UCR has to explain what element or elements are causing the color and the brightness variation.

The possible sources of color for an appearance patch are the illumination, the transfer function, or both. Intuitively, the simplest hypotheses attribute the color to a single element of the hypothesis, but o.13 0 Td Td (o6.96 0 Tdh)Tj 4.9

objects. Furthermore, the broad classes contain sufficient information to engage in reasoning about and merging of hypotheses regions.

3.3. Taxonomy of surfaces

Surfaces can be described at many levels of complexity. When reasoning about hypotheses, what we are most interested in is how the surfaces of adjacent hypothesis regions are related. When they show similar qualities, it is reasonable to consider merging the two regions.

To simplify this reasoning process, we initially consider only two classes of surfaces: curved and planar. These two classes provide a simple distinction that can be used to reason about merging hypotheses. A finer distinction we leave for future exploration. When a surface representation method is developed, reasoning about merging two curved surfaces will be done based on that representation.

3.4. Taxonomy of illumination

There are several simplified special forms of the incident light energy field function that represent useful models of illumination. Recall that the general form of the global incident light function is given by $L^+(x, y, z, \theta_x, \theta_y, \lambda, s, t)$. The largest subspace we consider is that of time-invariant illumination, where we consider time to be a constant and drop it from our parameterization. Within that space, a subspace we highlight is unpolarized time-invariant illumination $L^+(x, y, z, \theta_x, \theta_y, \lambda)$. For now, we assume all illumination in a scene will fall into this category.

Within the unpolarized, time-invariant subspace are those illumination functions in which the color of the light is independent of the direction of incidence. The hue and saturation of such illumination functions are the same in all directions and only the brightness varies over the illumination hemisphere. These illumination functions are separable into the form $L^+(x, y, z, \theta_x, \theta_y)C(x, y, z, \lambda)$, where $L^+(x, y, z, \theta_x, \theta_y)$ denotes the incoming intensity in a given direction at (*x,y,z*), and $C(x, y, z, \lambda)$ the color of the illumination. Within the subspace of separable functions is the *uniform* illumination subspace which can be written for the point (x,y,z) as $L^+(\theta_x, \theta_y)C(\lambda)$, where for all θ_x and θ_y, $L^+(\theta_x, \theta_y) = \{1, \alpha\}$, and α represents the background diffuse illumination. Uniform illumination thus implies that all illumination in the environment has the same color and one of two brightness values. Some special cases of uniform lighting include: point sources, finite disk sources, and perfectly diffuse ambient illumination.

These three simple cases play an important role in modeling illumination [5]. For the purpose of reasoning about hypotheses, we use the subspaces--in order of increasing complexity--diffuse, uniform, and general illumination to describe possible forms of incident light energy.

3.5. Taxonomy of the transfer function

Numerous common cases of the transfer function arise when we consider the subset of non-polarizing, opaque, and non-fluorescing surfaces. At present, we consider only surfaces that fall into this subset. For non-polarizing materials, the polarization parameters are separable and, since we are only considering unpolarized incident light, can be removed from the overall function. For non-fluorescent materials, $\Re = 0$ whenever $\lambda^+ \neq \lambda^-$, allowing the wavelength parameters to be combined into a single parameter λ. For opaque materials, the directions of incident and exitant light energy are limited to the hemisphere above the tangent plane for the surface point (u, v) . With these restrictions, the transfer function becomes $\Re(u, v, \theta^+, \varphi^+, \theta^-, \varphi^-, \lambda)$, where $0 < \theta < 90°$.

This reduced transfer function still includes surfaces with arbitrary changes in the transfer function over (u, v) . Such surface patches can have differing color and texture within their extent. Therefore, we further identify two nested subsets: transfer functions that are piecewise-uniform, and those that are completely uniform over the extent of the (u, v) parameters. The subset of uniform transfer functions can be specified by the reduced function $\Re(\theta^+, \varphi^+, \theta^-, \varphi^-, \lambda)$, as it is constant over the relevant values of u and v. This form of the transfer function is recognizable as the well-known *spectral bi-directional reflectance distribution function* [spectral BRDF] for a uniform surface [16].

Transfer functions with surface reflection or body reflection are two important overlapping subspaces of the spectral BRDF. Surface reflection takes place at the interface between an object and the surrounding air. Incident light is reflected through the local surface normal in the "perfect specular direction. Many materials displaying surface reflection are optically "rough." They possess microscopic local surface normals that differ from the macroscopic shape. Objects displaying only surface reflection--e.g. metals--have been analyzed using the *unichromatic reflection model* [7]. For a more thorough discussion of surface reflection see, for example, [6] and [15].

Body reflection takes place when light penetrates a surface and interacts with colorant particles. During this interaction, some of the wavelengths may be absorbed, coloring the reflection. The remaining wavelengths are re-emitted in random directions with some striking other colorant particles and others exiting the surface as body reflection. Many objects are assumed to obey Lambert's Law, which states that the reflection is dependent only

upon the incoming light's intensity and cosine of the angle of incidence [8]. Other models of body reflection are being researched (e.g., see [20]). Objects displaying both body and surface reflection have been analyzed with the dichromatic reflection model.

For the purposes of our proposed segmentation method, we initially consider objects whose transfer functions fall within the union of body reflection and surface reflection. Objects with these properties naturally divide into two categories: *metals* and *dielectrics* (e.g. plastic, paint, and ceramic). Metals display only surface reflection; dielectrics always have some body reflection, and often display surface reflection as well, although not as strongly as metals. For this discussion we assume neutral interface reflection for dielectrics--i.e. that the surface reflection uniformly reflects all wavelengths.

3.6. Fundamental hypotheses

Based on the above taxonomies of S, L^+, and $\Re$, we now identify a simple, yet comprehensive set of hypotheses for explaining the color and brightness variation of a UCR. To accomplish this task, we first form a set of hypothesis classes based upon the taxonomies previously developed for the individual hypothesis elements. The broad classes for each element are:

- Surfaces = planar, curved
- Illumination = diffuse, uniform, general function
- Transfer Function = metal, dielectric

The possible combinations of these broad classes create a set of twelve hypotheses for an appearance patch corresponding to a UCR.

To account for the distribution of color between L^+ and $\Re$ we further subdivide them into two classes: uniform (white or grey) and non-uniform (colored) spectrum. This creates six illumination and four transfer function classes. The possible combinations of these classes we call the set of *fundamental hypotheses*. The fundamental hypotheses constitute a comprehensive, yet finite set of explanations for each UCR.

Straightforward combination of the classes (2 x 6 x 4) gives 48 possible hypotheses. It can be shown, however, that we never have to consider more than 36 for a particular UCR. This can be seen by examining the possible explanations for a colored UCR and a grey UCR individually. For a hypothesis to postulate color, either L^+ or $\Re$ must be non-uniform. This eliminates the 12 uniform illumination/uniform transfer function combinations, resulting in 36 fundamental hypotheses for a colored UCR.

For a hypothesis to postulate no color, both L^+ and $\Re$ must be uniform spectrum. (We realize it would be possible for two non-uniform spectra to combine to give a uniform spectrum result but assume it to be sufficiently rare as to be inconsequential.) Eliminating the 36 uniform/non-uniform and non-uniform/non-uniform combinations results in 12 fundamental hypotheses for a grey UCR.

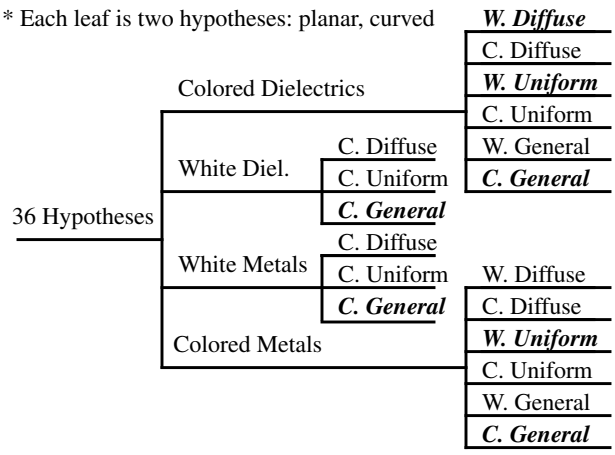

Figure 3 Taxonomy of fundamental hypotheses for a colored UCR

We can arrange the fundamental hypotheses for a UCR in a tree structure according to material type, illumination, and shape as shown in Figure 3. The resulting tree for a colored UCR with its 36 leaves represents a taxonomy of fundamental, or simple hypotheses which classify the different physical explanations for an image region. The true importance of this taxonomy is that it is a finite set of simple, yet comprehensive hypotheses for describing an appearance patch corresponding to a UCR. Therefore, we can postulate a hypothesis set with a reasonably small number of hypotheses for each UCR we identify in an image. This provides an initial segmentation and sets the stage for us to begin reasoning about and merging hypothesis regions.

4. Analysis and merging

The taxonomy shown in Figure 3 might be taken to suggest that all of the fundamental hypotheses are of equal value in explaining a scene. We do not believe this to be the case for most images. We propose dividing the 36 hypotheses into two groups, or tiers, corresponding to common and rare physical situations. Common hypotheses we specify as belonging to tier one, and rare hypotheses we place in tier two.

To accomplish this division we use two general rules. First, if a smaller subset is both common and a good approximation of a larger set, we place the smaller subset in tier one, and the larger in tier two. Second, if a smaller subset is both uncommon and not a good approximation of

a common larger subset, we place the smaller subset in tier two and the larger in tier one. Applying these guidelines, we place 14 of the fundamental hypotheses in tier one, and 22 in tier two. The tier one hypotheses are highlighted in Figure 3. For a more detailed discussion, see [13].

In our example segmentation, we consider only the fundamental hypotheses in tier one. As shown in Figure 5, the two-color Lambertian sphere divides into three UCRs: top, middle, and bottom. To each region we attach the list of fourteen hypotheses from tier one, forming three hypothesis sets of fourteen each.

The next step in the segmentation process merges small regions into big ones in order to search for coherence between regions. Our basic method is to take two adjacent hypothesis sets $HS_1 = \langle P_1, H_{11}, H_{12}, \ldots \rangle$, and $HS_2 = \langle P_2, H_{21}, H_{22}, \ldots \rangle$, and form a new hypothesis set $HS_3 = \langle P_1 \cup P_2, H_{31}, \ldots \rangle$, in which the hypotheses H_{3i} are created by merging compatible hypotheses H_{1j} and H_{2k}.

A bulldozer approach would consider all possible combinations of the fundamental hypotheses, resulting in 14^2 aggregate hypotheses. But are there really 196 plausible explanations for the combination of just two regions? Such a merging method is not only unreasonable, but also too expensive to use even on simple images because of the exponential explosion of the number of hypotheses.

Fortunately, the goal of the segmentation process provides a partial solution. The mergers we want to make during segmentation involve coherence in the general variables: material type, shape, and illumination. In other words, we want to merge hypothesis regions that are similar. Adjacent hypothesis regions, however, have differing exitant light energy fields (specifically, a different chromaticity). This implies there is a discontinuity between the two regions in at least one of the hypothesis elements. We propose that if adjacent hypothesis regions belong to the same object, this discontinuity is a simple one and *involves only one of the hypothesis elements*.

In addition to this general postulate, we propose that hypothesis regions of differing materials should not be merged. While such a merger might make sense at a more abstract level (i.e., object recognition), such a step is not appropriate for low-level segmentation.

The result of these postulates is that the combination of two hypothesis sets, each containing 14 hypotheses, results in only 28 hypotheses for the combined region. The desired mergers between hypothesis sets are indicated in Figure 4 by the shaded regions.

Returning to our example, merging the top and middle regions, and the middle and bottom regions, we obtain 28 possible hypotheses for the merger of each pair. As the hypotheses for the middle region can be matched, there are, in fact, 68 resulting hypotheses for the entire sphere.

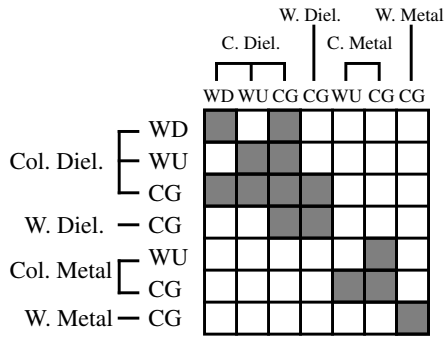

Figure 4 Desired mergers (shaded squares) of hypotheses of two colored UCRs.

Only twelve of these hypotheses, however, postulate that the material type and illumination are of the same type over all three regions.

At this point in the segmentation, we use our postulate that the simplest explanation is the best explanation for a hypothesis set. It is important to realize that this stage of the segmentation process is dependent upon the specific region being described. For example, a region of uniform pixel values can easily be described by a region of homogeneous color under diffuse illumination. Regions such as those in our example, however, require a surface of non-homogeneous color if the illumination is diffuse. By reasoning about possible realizations given specific regions and hypotheses, we believe it is possible to rank-order the resulting merged hypotheses.

We finish our example by rank-ordering the final twenty hypotheses for the example image. We realize that we are using some human reasoning in this process, but it is the first step towards developing a more rigorous, computable process. The 8 simplest explanations are shown in Figure 6. The hypotheses shown are those postulating the three regions to be a colored mirror, to be colored plastic under colored illumination, or to be inhomogeneous over the three regions, making them more complex than the homogeneous hypotheses.

The simplest explanation for this scene is hypothesis 1, proposing that each region is a colored dielectric under uniform illumination. We propose this hypothesis as the simplest explanation because its homogeneity and simplicity, allow us to specify the scene with a small number of parameters and recreate it exactly. For no other hypothesis in Table 1 is there as compact a realization.

What this analysis provides for our example image is a set of suggested segmentations. Furthermore, these segmentations are rank-ordered, giving a higher level program a sense of which are the 'best' segmentations of the image. While the criteria and reasoning used to rank-order the segmentations are not rigorous enough in this formulation to allow a computer to simulate these results, we

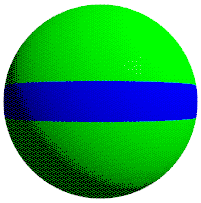

Figure 5 Example: two- color lambertian sphere.

	Material	Illumination	Surface
1.	colored dielectric	uniform white ill.	curved
2.	colored dielectric	uniform white ill.	planar
3.	colored dielectric	diffuse white ill.	planar
4.	white dielectric	colored general ill	curved
5.	white dielectric	colored general ill	planar
6.	colored dielectric	diffuse white ill	curved
7.	grey metal	colored general ill	planar
8.	grey metal	colored general ill	curved

Figure 6 8 simplest homogeneous hypotheses

believe this method is asking the right questions and laying the foundation for a rigorous segmentation algorithm.

5. Conclusions

What we have presented herein is an abstract analysis of the problems and methods involved in segmentation of general color images. To support this analysis, we presented a general model and nomenclature describing the physics of image formation. We have also provided a rough example of our segmentation framework, demonstrating the major themes and ideas.

The value of our analysis is that for the first time we are examining where in the general segmentation process it is appropriate to apply specific physics-based vision techniques and how to integrate them into a whole. We have also explicitly identified some of the difficulties inherent in integrating and reasoning about the physics of image formation. Our analysis is both a basis and set of guidelines for future research towards the development of an integrated segmentation system.

6. Acknowledgments

This research was 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 No. F49620-92-C-0073. The United States Government is authorized to reproduce and distribute reprints for governmental purposes notwithstanding any copyright notation hereon.

7. References

[1] E. Adelson and J. Bergen, "The Plenoptic Function and the Elements of Early Vision," in *Computational Models of Visual Processing*, ed. M. S. Landy, and J. A. Movshon, Cambridge, MIT Press, 1991.

[2] M. Born and E. Wolf, *Principles of Optics*, Pergamon Press, London, 1965.

[3] P. Breton, L. A. Iverson, M. S. Langer, S. W. Zucker, *Shading Flows and Scenel Bundles: A New Approach to Shape from Shading*, TR-CIM-91-9, McGill University, Nov. 1991.

[4] T. Darrell, S. Sclaroff, and A. Pentland, "Segmentation by Minimal Description," in *Proceedings of International Conference on Computer Vision*, IEEE, pp.112-116, 1990.

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

[6] X. D. He, K. E. Torrance, F. X. Sillion, and D. P. Greenberg, "A Comprehensive Physical Model for Light Reflection," *Computer Graphics* 25(4), pp175-186, 1991.

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

[9] J. Kaufman and J. Christensen, *IES Lighting Ready Reference*, Illuminating Engineering Society of North America, 1985.

[10] 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.

[11] J. Krumm and S. A. Shafer, "Segmenting Textured 3D surfaces Using the Space/Frequency Representation," to be printed in *Spatial Vision*, 1994.

[12] H.-C. Lee, "Method for Computing the Scene-Illuminant Chromaticity from Specular Highlights," *Journal of the Optical Society of America* A 3(10), pp.1694-1699, 1986.

[13] B.A. Maxwell and S.A. Shafer, "A framework for segmentation using physical models of image formation," CMU-RI-TR-93-29, Robotics Institute, Carnegie Mellon University, Dec. 1993.

[14] P. H. Moon and D. E. Spencer, *The Photic Field*, Cambridge, MIT Press, 1981.

[15] S. Nayar, K. Ikeuchi, and T. Kanade, *Surface Reflection: Physical and Geometrical Perspectives*, CMU-RI-TR-89-7, Robotics Institute, Carnegie Mellon University, 1989.

[16] F.E. Nicodemus, J. C. Richmond, J. J. Hsia, I. W. Ginsberg, and T. Limperis, *Geometrical Considerations and Nomenclature for Reflectance*, National Bureau of Standards NBS Monograph 160, Oct. 1977.

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

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

[19] 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.

[20] L. B. Wolff, *A Diffuse Reflectance Model for Dielectric Surfaces*, The Johns Hopkins University, Computer Science TR 92-04, April 1992. (Submitted to JOSA)