COMPUTER VISION AND IMAGE UNDERSTANDING
Vol. 65, No. 2, February, pp. 269–295, 1997
ARTICLE NO. IV970573

Physics-Based Segmentation of Complex Objects Using Multiple Hypotheses of Image Formation*

BRUCE A. MAXWELL†

Department of Computer Science, University of North Dakota, Grand Forks, North Dakota 58202

AND

STEVEN A. SHAFER

Microsoft, 1 Microsoft Way, Redmond, Washington 98052

Received November 14, 1995; accepted November 20, 1996

We present a general framework for the segmentation of complex scenes using multiple physical hypotheses of image formation. These hypotheses specify broad classes for the shape, illumination, and material properties of simple image regions. Through analysis, merging, and filtering of hypotheses the framework generates a ranked list of segmentations. We have implemented an algorithm based upon this framework and show example segmentations of scenes containing multicolored piece-wise uniform dielectric objects. By using this new approach we can intelligently segment scenes with objects of greater complexity than previous physics-based algorithms. The results show that by using general physical models we can obtain segmentations that correspond more closely to objects in a scene than segmentations found using only color. © 1997 Academic Press

1. INTRODUCTION

The objective of physics-based segmentation is to divide an image of a scene into regions that are meaningful in terms of the objects constituting that scene. This means the computer must generate and reason about one or more descriptions of the scene elements that formed the image—the illumination, material optics, and geometry—in order to form an interpretation. Forming such an interpretation is a relatively simple task for humans. For example, a person can easily generate a comprehensive physical description of Fig. 1, a picture containing numerous objects with many different reflective properties.

For a computer, however, images containing multicolored objects and multiple materials such as Fig. 1 are difficult to understand and segment intelligently. Simpler scenes like Fig. 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 [2, 13, 17]. 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.

Multicolored objects, like the mug in Fig. 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 [2, 7, 13, 17].

The difficulty inherent in segmenting images with multiple materials and multicolored objects is that by expanding the space of physical models considered for the shape, illumination, and material optics, a single 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 Fig. 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

* The U.S. Government's right to retain a nonexclusive royalty-free license in and to the copyright covering this paper, for governmental purposes, is acknowledged.

† Correspondence should be sent to: Bruce A. Maxwell, Department of Computer Science, University of North Dakota, University & Tulane, Grand Forks, ND 58202. Phone (W): (701) 777-4982; phone (H): (701) 795-9286; fax: (701) 777-3330.

1077-3142/97 $25.00
Copyright © 1997 by Academic Press
All rights of reproduction in any form reserved.



FIG. 1. Complex scene containing multiple materials and multicolored objects (Color Plate 1).

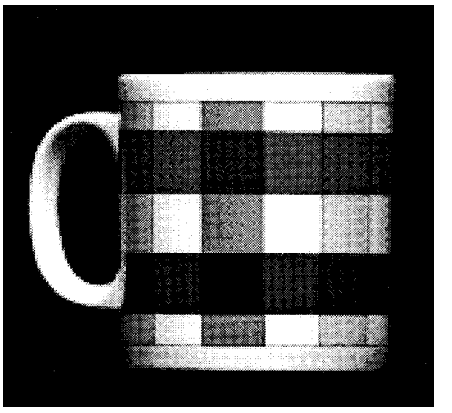

FIG. 3. Multicolored piece-wise uniform dielectric object (Color Plate 3).

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 is part of a photograph and the variation is due to changes in the material properties (color and intensity). 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 changes 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; we can never expect to get *the* single correct interpretation of Fig. 4, only a *possible* correct interpretation.

Considering multiple interpretations of an image, however, runs the risk of getting bogged down in the very large search space of possibilities. We present two major ideas with the intent of avoiding this computational quagmire. First, we present a framework within which knowledge and assumptions about the physics of image formation can be used to heavily prune the set of possible interpretations. Second, we abstract the problem to a simpler domain of broad classes and use reasoning in this domain to narrow the number of physical descriptions which must be considered. The ultimate goal is to narrow the number of physical descriptions to a few likely candidates. Determining this small number of likely physical descriptions is the key to segmenting and understanding image data.

1.1. Previous Work in Segmentation

Early work in segmentation was based upon straightforward statistical models of the image data and did not search for the underlying semantic meaning. They modeled images as regions of uniform color and intensity, and variations in these characteristics as noise [6]. Researchers knew that using information about the scene was important, but they incorporated such knowledge (such as trees are beside a road) on top of their statistical models [42].

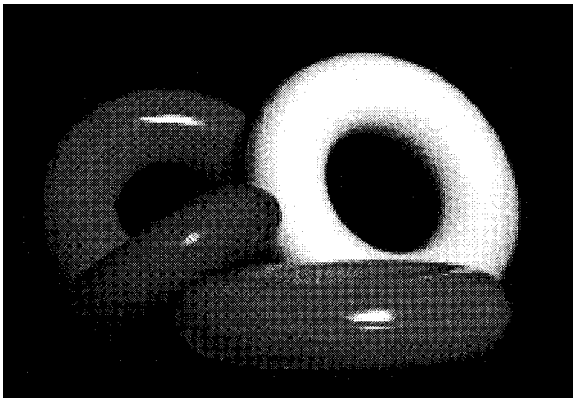

FIG. 2. Uniformly colored dielectric objects with highlights (Color Plate 2).



FIG. 4. Image of an object, a reflected image of the object, and a photograph of the object (Color Plate 4).

A statistical approach was taken partly because of the optimism of the 1970s surrounding symbolic reasoning and artificial intelligence, which relegated to low-level vision the task of dividing an image into simple regions based upon color and brightness. More extensive low-level processing was considered unnecessary because it was assuming interreflection, but both methods had limitations. To correctly model interreflection using the methods of Bajcsy *et al.* a white reference plate is necessary in order to eliminate the effect of colored illumination. Furthermore, there are still a large number of materials and lighting conditions that cannot be handled by these models and

propose that the very forms of the models are to be instantiated in order to accommodate qualitatively different shapes, materials, colors, and illumination environments. In this, we are moving the analysis from the primitive level 1 analysis of Rissanen [35]—estimating parameters of a previously established model—to a level 3 analysis—selecting the model class—with a resultant increase in perceptual power.

From the above summary of work in physics-based segmentation, it is clear that model selection has only recently been examined by Breton *et al.,* and only for illumination and shape. Multiple models are needed because of ambiguity in an image. Figure 2, for example, shows three identical image regions that have very different physical explanations. Some unifying framework is needed to bring together the myriad of physics-based vision techniques and reason about when, where, and how they should be applied. Some of the questions that must be answered include: what models do we use, what parameters do we need to consider, how do we choose an initial set of models, and how do they merge and interact?

These are the questions we deal with in this paper. In Section 2 we present a general model, showing all of the possible parameters for the space of model classes. In Section 3 we suggest a method for narrowing the number of physical interpretations of an image region and choosing a subset of the possible models with which to begin segmentation. In Section 4 we analyze the process of merging different model hypotheses to obtain global segmentations.

The second half of this paper describes an initial implementation of our framework using a limited set of hypotheses. With this limited set, we are able to generate segmentations of images containing multicolored piece-wise uniform dielectric objects that more closely correspond to objects in the scene than segmentations found using only color. In Section 5 we present the implementation details and outline our initial segmentation algorithm for finding simple image regions. In Section 6 we 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 7 we present our 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, in Section 8 we show how a multilevel region graph can be created and used to find a set of segmentations for the image. Finally, in Sections 9 and 10 we present the results of our segmentation method on two test images, discuss these results, and identify directions for future work.

2. A GENERAL MODEL OF IMAGE FORMATION

Images are formed when light strikes an object and reflects towards an imaging device such as a camera or an eye. The color and brightness of a point in an image is the result of the color and intensity of the incident light, and the shape and optical properties of the object. This section presents a formal model of these elements, how they interact, and how they are related to what we see in an image. Note that this description of image generation neglects camera effects such as those described by Wilson [40]. For now we assume these effects are small and realize that, for completeness, they should be incorporated into this framework in the future.

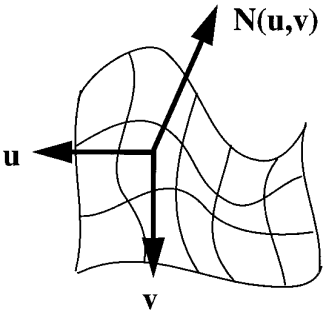

FIG. 5. Local coordinate system on a surface patch.

2.1. The Elements of a Scene

The elements constituting our model of a scene are surfaces, illumination, and the light transfer function or reflectance of a point in 3D space. 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 [37]. We begin by providing a formal notation for each of these elements.

2.1.1. *Surfaces*

We model objects in the real world using 2D 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 3D 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 3D global coordinates. This global coordinate system is also anchored to an arbitrary origin, often specified relative to an imaging device. As shown in Fig. 5, 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 3D coordinate system at each surface point with two axes on the tangent plane and one in the direction of the surface normal. Other useful properties, such as curvature, can also be defined and specified for each point using the surface embedding function.

It is important to note that we do not view the world as

consisting of surfaces to be found, but as objects to be modeled. It is commonly presumed in machine vision that "surfaces" exist in nature and that the job of the vision system is to discover them. We reject that view, believing instead that surfaces are artifacts of the interpretation process and exist only within the perceptual system that is attempting to build a model of the world. Given this view, there is no "correct" surface with which to model an object. Instead, the choice of which manifold and surface embedding function will be used to represent a given object is made by the modeler and depends largely upon the task and information at hand. Given a brick wall, for example, if the application is obstacle avoidance, a single plane could be chosen to model the entire wall. For other situations, such as segmentation, it might be necessary to model each brick as well as the troughs between them. At an even smaller scale, understanding the image texture in detail may require a model of each bump on each brick in order to interpret the wall. All are potentially useful "surfaces" to model the same wall, and all might be needed at various points in the visual process. Thus one object in the world can be modeled by many different surfaces, and the choice of model, or surface, is made by the interpreter. This view allows us to conceive of a perceptual process that incorporates numerous differing surfaces to describe an object, an important capability that other computational vision systems, which seek for a single "correct" surface, lack.

In order to parameterize light striking and reflecting from a surface, we also need to define a parameterization of direction. In the global coordinate system we use two angles (θ_x, θ_y), where θ_x specifies the angle between the direction vector and the x-axis, and θ_y corresponds to the angle between the direction vector and the y-axis. To specify directions, or a ray, in the local coordinate systems, we will use normal spherical coordinates, as shown in Fig. 6, specified by the ordered pair (θ, φ). θ is the *polar angle,* defined as the angle between the surface normal and the ray, and φ is the *azimuth,* defined as the angle between a perpendicular projection of the ray onto the tangent plane and a reference line on the surface (usually defined to be either the u or v axis).

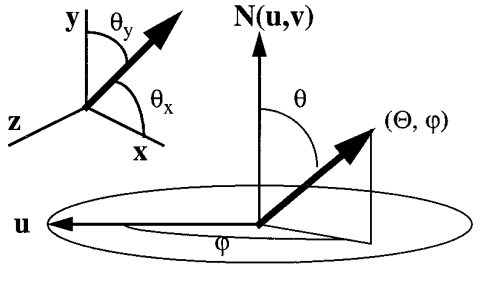

FIG. 6. Specifying direction in the global and local coordinate systems.

2.1.2. *Illumination*

Much research in machine vision assumes a single light source, often a relatively large distance away from the scene being imaged. More recently, Langer and Zucker have proposed a computational illumination model for many forms of direct illumination [19]. However, many visual phenomena arise because of reflection from nearby objects acting as additional light sources. The field of computer graphics has long incorporated this idea into systems such as ray tracing and radiosity.

To begin examining general images we cannot assume point lighting, three independent light sources, or other constructed illumination setup. While for Lambertian surfaces, we can represent arbitrary illumination in a more compact manner—see, for example, [5 or 34]—for a general framework we cannot assume an image will contain only Lambertian surfaces. A general model must allow us to specify any type of illumination, including interreflection from other objects, and still have identifiable subsets that fit with our traditional conceptions of illumination. We develop our model by first defining and specifying the parameters of a single ray of light and then extending this model to describe the light arriving at a point.

A *photon* is a quantum of light energy that moves in a single direction unless something—like matter, or a strong gravity field—affects its motion. Thanks to the sun and artificial light sources, there are many photons moving in many directions at any given time. Collections of photons moving in the same direction at the same place and time constitute *rays* of light. As photons move, they oscillate about their direction of travel at a spectrum of *wavelengths* λ which specify the distance traveled in a single oscillation. The human eye is sensitive to photons with wavelengths that fall between approximately 380 and 760 nm, and the spectral distribution of wavelengths present in a collection of photons determine what color we see. A charge-coupled device [CCD] camera responds to a slightly different range of wavelengths, and infrared color filters are normally used to approximately match the color response of the human eye. The *polarization* of a population of photons specifies their oscillation and orientation with respect to the direction of travel, and it can affect the manner of reflection and transmission when light interacts with matter. Polarization is commonly represented using a set of parameters, such as the Stokes parameters [8], which we indicate by the variable $s \in \mathbb{h}1, 2, 3, 4\mathrm{j}$ that indexes the Stokes parameters to specify the relative energy of photons oscillating at different orientations.

In a scene, light is being emitted or reflected in numerous directions, entering and leaving points throughout the area of interest. Using the parameters described above, 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, can be specified by the 8-tuple $(x, y, z, \theta_x, \theta_y, \lambda, s, t)$.

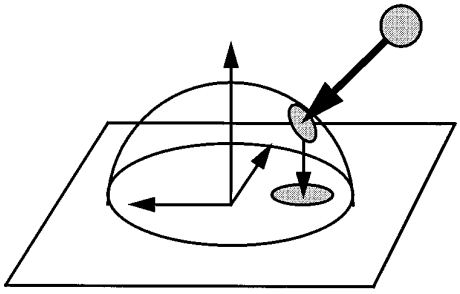

FIG. 7. Orthogonal mapping of the illumination environment onto a plane.

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 3D coordinates. We can describe the light energy arriving at a point from all directions by the *incident light energy field* function $L^+(x, y, z, \theta_x, \theta_y, \lambda, s, t)$, which specifies the radiance 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* [28]. 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. As a result, we consider only the subspace of the incident light energy field $L^+(x, y, z, \theta_x, \theta_y, \lambda, s)$.

For a point in free space, we note that rays arriving at that point can be mapped onto a sphere of unit radius [10]. In this manner, the incident light on a surface point can be visualized on the unit sphere. The brightness and color of a point (θ_x, θ_y) on the sphere indicates the brightness and color of the incident light from that direction. We define this representation of the light energy field on the unit sphere for a 3D point (x, y, z) to be the *global illumination environment* for that point. It is important to note that on opaque surfaces some of the incident light is blocked by the object matter itself, limiting the illumination environment to the hemisphere above the tangent plane. If the surface is transparent, the illumination environment will be the complete sphere, as light can be incident on the surface point from below as well as above. We can visualize the illumination environment for opaque surfaces by orthogonally projecting it onto a plane as in Fig. 7. Two example illumination environments are shown in Fig. 8. A rendering of what such illumination environments might look like is shown in the inset image beside each figure.

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* $L^+(u, v, \theta, \varphi, \lambda, s)$, which also can be visualized on a hemisphere above the tangent plane to the local surface point for opaque surfaces. This representation we call the *local illumination environment* for the surface point (u, v). The global and local illumination functions are distinguished by their parameters.

The total radiance of a patch of the illumination environment hemisphere with polarization specification s at wavelength λ, specified by the angles (θ, φ) and subtending $d\theta$ and $d\varphi$ is given by $L^+(u, v, \theta, \varphi, s, \lambda) \sin \theta d\theta d\varphi d\lambda$ [14]. The total irradiance at a point (u, v) is given by (1). The sine term is part of the solid angle specification, and the cosine term reflects the foreshortening effect as seen by the surface point:

$$E = \bigcirc\!\!\!\!\!\!\int_s \int_\lambda \int_{-\pi}^{\pi} \int_0^{\pi/2} L^+(u, v, \theta, \varphi, s, \lambda) \cos\theta \sin\theta d\theta d\varphi d\lambda. \tag{1}$$

2.1.3. *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 parameterization as the incident light energy field and describes an intensity for every direction and wavelength. As with the incident light energy field, we can define a local coordinate version of the exitant light energy field $L^-(u, v, \theta, \varphi, s, \lambda)$.

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

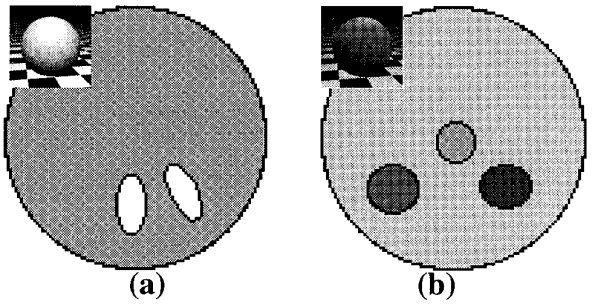

FIG. 8. Examples of (a) white uniform illumination; (b) general function illumination (see Color Plate 5).



FIG. 9. Some special cases of the light transfer function: Lambertian reflection, fluorescence, polarization, transmittance, and specular or surface reflection.

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 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* $\Re(u, v; \theta^+, \varphi^+, s^+, \lambda^+; \theta^-, \varphi^-, s^-, \lambda^-)$.

The relationship between the incident light energy, the exitant light energy, and the transfer function can be written using local coordinates as the integral in (2). This integral says that the exitant light energy field is the sum of the self-luminance of the point, L_0, and the product of the transfer function and the incident light energy field integrated over the parameters of the incident light. The cosine term is due to foreshortening, and the sine term from the solid angle specification. The result of this integral is a function of the exitant light variables.

$$L^-(u, v; \ldots - \ldots) = L_0^-(u, v; \ldots - \ldots) + \sum_{s^+} \int_{\lambda^+} \int_{-\pi}^{\pi} \int_{0}^{\pi} L^+(u, v, \ldots + \ldots)\, \Re(u, v; \ldots + \ldots; \ldots - \ldots) \cos\theta^+ \sin\theta^+ \, d\theta^+ \, d\varphi^+ \, d\lambda^+. \quad (2)$$

A structured analysis of the transfer function shows how it subsumes several common special cases, sketched in Fig. 9. We give a brief description of the parameter constraints that correspond to these special cases: fluorescence, polarization, transmittance, and surface or specular reflection. These descriptions demonstrate the framework provided by the general transfer function:

• For a nonfluorescing surface, if the incident light is of wavelength λ_0, then the exitant light energy field will also have wavelength λ_0, and no other wavelengths will be present. If, on the other hand, the same incident light strikes a fluorescent surface, there may be other wavelengths present in the exitant light energy field. In terms of the parameters of the transfer function, fluorescence implies there exists some pair of wavelengths (λ^+, λ^-), where $\lambda^+ \neq \lambda^-$ for which $\Re > 0$.

• Polarizing transfer functions modify the polarization of the incoming light. This effect can be seen in sunglasses, which often block the horizontal polarization mode. For non-polarizing surfaces, $\Re = 0$ whenever $s^+ \neq s^-$. For a polarizing transfer function, there exists some pair of stokes parameters (s^+, s^-) where $s^- \neq s^+$ for which $\Re > 0$.

• Transmitting surfaces allow some light to pass through them. Conversely, an opaque surface limits both the incident and exitant light energy fields to a hemisphere above the tangent plane for that surface. Transmittance occurs when either the exitant or incident light energy field bounds (θ^-, φ^-) and (θ^+, φ^+) are extended beyond the hemisphere above the tangent plane of the surface, implying that at least some of the exitant or incident light energy is passing through the material. In terms of the parameters, a surface is transmitting if $\Re > 0$ when $\theta^- > 90°$ or $\Re > 0$ when $\theta^+ > 90°$.

• Specular reflection, described in more detail later on, occurs when the incident light is only reflected about the local surface normal in the perfect specular direction. This restriction implies that the transfer function is zero, except when $\varphi^- = \varphi^+ + \pi$ and $\theta^- = \theta^+$. It is important to note that surface reflection is relative to the local surface normal, and it is possible to have an optically rough surface where the local surface normals vary relative to the overall surface [3, 39].

• Finally, Lambertian surfaces—also called perfectly diffusing surfaces—reflect incident light equally in all directions. For a unit energy ray of light from direction (θ, φ), the exitant light energy in all directions is specified by the expression $\cos\theta$.

To illustrate a transfer function, we show a sphere with that transfer function sitting above a matte black and white checkered surface under a dark grey sky with a white point light source shining on it from above and to the right of the viewer. Because all illumination is of uniform spectrum (i.e., grey), any color in the image is due to the transfer function. The checkerboard pattern is present to highlight

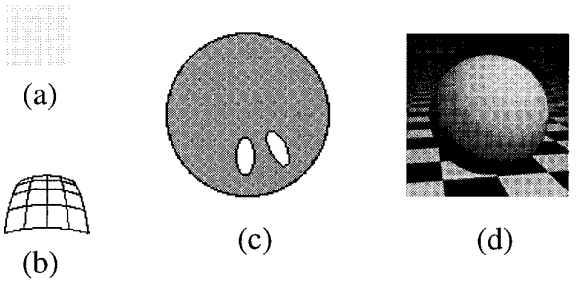

FIG. 10. Hypothesis visualization: (a) the actual region; (b) wire-diagram of the shape; (c) illumination environment; (d) transfer function (see Color Plate 6 for (a) and (d)).

the specularity of the object. Figure 10d shows a visualization of a matte plastic transfer function.

2.2. General Hypotheses of Physical Appearance

We have defined a 3D world model for individual points and their optical properties, but how does a whole surface appear in a digitized computer image? To describe a surface and its appearance, we introduce a nomenclature for the aggregation of appearance properties in the 3D world and how these aggregations map to an image.

We have defined surfaces with an extent and embedding, and we have defined a transfer function $\Re$ over a surface. The combination of a surface and a transfer function we define to be a *surface patch.* Because the transfer function can vary arbitrarily, there are no constraints on the appearance of a general surface patch in an image. Frequently, however, the transfer function at nearby points on a surface displays some type of identifiable coherence. Coherence does not imply uniformity and covers a broad scope of possible aggregations such as uniformity, repetitive patterns, or irregular textures. Some properties that commonly impart coherence include material type, color, roughness, and the index of refraction. We can model the coherence of the object's appearance with a surface patch whose transfer function is similarly coherent.

A surface patch with a coherent transfer function, however, will not always display the coherence in an image. Differing illumination over the surface patch 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, which is what is viewed by the imaging device. To achieve coherence in the exitant light energy field, we must add to the surface/transfer function pair a coherent illumination environment over the surface patch. This combination we define as an *appearance patch:* a surface patch whose points exhibit a coherent transfer function and illumination environment, whose exitant light energy field exhibits a coherence related to that of the transfer function over the entire patch, and which 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 exitant light from a surface caught by the imaging device determines the color and position of the set of pixels related to that surface. The physical explanation for a given exitant light energy field from a given surface patch we define to be a *hypothesis* $H = \mathbb{k}S, E, \Re, L^+\mathbb{l}$. 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* $HR = \mathbb{k}P, H\mathbb{l}$, 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. For simplicity, we assume the image is formed by a pinhole camera at the origin looking at the canonical view volume. To represent the fact that a single region may have more than one possible explanation, we define a *hypothesis list* $HS = \mathbb{k}P, H_1, \ldots, H_n\mathbb{l}$ 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 $\mathbb{h}HS_i\mathbb{j}$ of hypothesis lists for pixel regions P_i, we define a *segmentation* of the pixel set $P = \cup_i P_i$ 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.

To illustrate a hypothesis, we combine the representations developed previously into a 3-panel image displaying the characteristics of S, L, and $\Re$ as shown for a yellow region in Fig. 10.

To summarize, our model for a scene consists of three elements: surfaces, illumination, and the light transfer function or reflectance of a point or surface in 3D space. These elements constitute the *intrinsic characteristics* of a scene, as opposed to *image features* such as pixel values, edges, or flow fields [37]. 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.*

It is important to realize that 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) [9].

2.3. Taxonomy of the Scene Model

In an ideal world complexity, or ''weirdness,'' would be quantifiable and could be used as the basis for generating and rank-ordering the possible hypotheses for a given region. The weirdness of a hypothesis might be represented by three axes indicating the complexity of the shape, transfor a region using only a measure of complexity should be sufficient to satisfy the MDL criteria.

As there are a large number of hypotheses for any image region, how to select the initial hypothesis set for each region is a crucial decision. One important consideration of the MDL principle is that the optimal model, or model set must be among those tested for shortest length. Three possible approaches that could be taken to generate this model set are:

1. Generate a large number of possible hypotheses and test;

2. Generate incrementally according to some search criterion;

3. Generate a small, but comprehensive set, using broad classes of the hypothesis elements; expand this set incrementally if all of its constituents are ruled out as possibil-

It is important to note that the initial hypothesis list for a region can be incrementally expanded if all of its constituents are considered unlikely. In the next three subsections we derive broad classes from the general parameterized models. These classes are simple, yet comprehensive enough to cover a wide range of possible environments and objects. Furthermore, while they are abstractions of more detailed models, they contain sufficient information to allow reasoning about different physical interpretations and the relationships of these interpretations between neighboring regions.

2.3.1. *Taxonomy of Surfaces*

Surfaces have numerous levels of complexity. A cube, for example, can be modeled as a set of planar patches, a polyhedron, or a superquadric. As noted previously, when modeling objects in the real world, surfaces can take on any amount of complexity, depending upon the needs of the modeler. To reason about merging adjacent hypotheses, we need to know whether they have compatible shapes—i.e., fit together at the boundaries. When the boundaries are compatible, we should consider merging the two regions.

We initially consider only one characteristic of a surface: is it curved, or is it planar? Clearly a curved surface can be arbitrarily close to planar, so in practical application this distinction must be made using a selected threshold. The curved/planar distinction allows for straightforward reasoning at an abstract level about merging two hypothesis regions. A finer distinction requires a specific method for modeling curved surfaces. When a surface representation method is chosen, reasoning about merging two curved surfaces can be done based on that representation, e.g., matching two spheres, superquadrics, generalized cylinders, or polynomial surfaces. Note that absolute depth is not a necessary requirement for reasoning about merging. If two regions' boundaries do not match in relative shape, they should not be merged. If the regions' boundaries do match, given some optimal offset and measure of similarity, a merger is not ruled out on the basis of shape.

2.3.2. *Taxonomy of Illumination*

Several special forms of the illumination function are often used in both computer vision and computer graphics to represent light conditions in a scene. The general form of L^+, given by $L^+(x, y, z, \theta_x, \theta_y, \lambda, s, t)$, contains these special cases as subspaces of its parameters space. Figure 11 shows the relationships of the subspaces we identify for this function. Not shown in Fig. 11 is the all-encompassing set of time-varying illumination functions. We assume time-invariant illumination, making time a constant and removing it from the parameterization. This leaves us with time-invariant illumination functions, shown as the largest space in Fig. 11. Within the space of general time-invariant illumination functions is the subspace of unpolarized time-invariant illumination $L^+(x, y, z, \theta_x, \theta_y, \lambda)$. For most images of interest all of the illumination in a scene is characterizable by this function. Scenes with illumination outside this subspace are rare and would be those illuminated by a polarized light source such as a laser, or by a time-varying source (over the course of the image capture process).

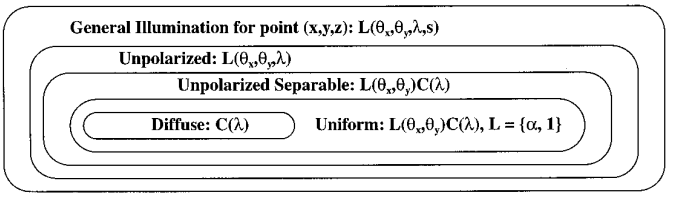

FIG. 11. Subspaces of the global incident light energy field $L(x, y, z, \theta_x, \theta_y, \lambda, s)$.

One common assumption in computer vision is that the illumination over the hemisphere is constant in its hue and saturation, but of varying brightness. Mathematically, this subspace is represented by the *separable* illumination functions. We define separable illumination functions to be those which can be expressed as $L^+(x, y, z, \theta_x, \theta_y)$ $C(x, y, z, \lambda)$, where $L^+(x, y, z, \theta_x, \theta_y)$ specifies the incoming intensity in a given direction at (x, y, z), and $C(x, y, z, \lambda)$ the color of the illumination. A more restrictive subspace of separable illumination is the *uniform* illumination subspace which we define for the point (x, y, z) to be $L^+(\theta_x, \theta_y)C(\lambda)$, where $L^+(\theta_x, \theta_y) = \mathbb{1}, \alpha\mathrm{j}$. Note that α represents the background, or ambient illumination commonly used in computer graphics. This definition states that each direction in a uniform illumination environment has the same color and one of two brightness values (light or dark). Some commonly used special cases of uniform illumination include:

- Point light source at $(\theta_{x0}, \theta_{y0})$

$$L^+(\theta_x, \theta_y) = \begin{cases} 1, & (\theta_x = \theta_{x0}) \text{ and } (\theta_y = \theta_{y0}) \\ 0, & \text{otherwise;} \end{cases}$$

- Finite disk source of apex angle α centered at $(\theta_{x0}, \theta_{y0})$

$$L^+(\theta_x, \theta_y) = \begin{cases} 1, & \text{angle between } (\theta_x, \theta_y) \text{ and } (\theta_{x0}, \theta_{y0}) < \alpha \\ 0, & \text{otherwise,} \end{cases}$$

- Perfectly diffuse "ambient" illumination $L^+(\theta_x, \theta_y) = 1$ for all θ_x and θ_y. Thus, L^+ is trivial and the illumination is fully characterized by $C(\lambda)$ at (x, y, z).

As shown by the computer graphics community, these three simple cases play an important role in modeling illumination; a large number of illumination environments can be modeled using one or more point, finite disk, or ambient light sources [12]. The uniform illumination subspace also falls within the computational model of Langer and Zucker, who have shown their model to be useful for scene analysis [19]. When reasoning about hypotheses, we would like to have a small number of classes, with most of them being highly constrained. We use the three subspaces—in order of increasing complexity—diffuse, uniform, and general illumination to describe the forms of the illumination environment. Diffuse illumination is a good approximation to objects in shadow or not directly lit. Uniform illumination is an approximation of man-made and natural light sources, and we must include general illumination because in some situations it is necessary—such as the colored objects reflected by the teakettle in Fig. 1. Figure 8 illustrates both a uniform illumination environment and a general illumination environment along with their effects on white dielectric spheres.

2.3.3. *Taxonomy of the Transfer Function*

As with the illumination function, the transfer function can be subdivided into commonly occurring subspaces. These generally fall within the space of nonpolarizing, opaque, and nonfluorescing transfer functions. We assume that the transfer functions of all objects within a scene are represented within this subspace. This assumption implies three constraints:

1. the polarization parameters are separable and, as we consider only unpolarized incident light, can be removed from the parameterization;

2. λ^+ and λ^- can be combined into a single parameter λ as $\Re = 0$ whenever $\lambda^+ \neq \lambda^-$;

3. the direction of incident and exitant light is limited to a hemisphere above the tangent plane for the point (u, v).

These assumptions allow us to rewrite the transfer function as $\Re(u, v, \theta^+, \varphi^+ \theta^- \varphi^-, \lambda)$, where $0 < \theta < 90°$. They do not, however, restrict the nature of the transfer function between neighboring points. Transfer functions exhibiting coherence over the extent of (u, v) form subspaces of the more general function. Two restrictive, but common, subspaces are transfer functions exhibiting piece-wise-uniform and uniform characteristics over their extent. In the uniform surface subspace, the transfer function is constant with respect to the parameter pair (u, v) and can be rewritten as $\Re(\theta^+, \varphi^+, \theta^-, \varphi^-, \lambda)$. For this subspace the transfer function is identical to the well-known *spectral bidirectional reflectance distribution function* (spectral BRDF) for a uniform surface [31].

For the purpose of this analysis, we will concentrate on the spectral BRDF, which contains two important and overlapping subsets: surface reflection and body reflection. Their relationship within the BRDF and the interaction of the union of these subspaces is shown in Fig. 12.

Surface reflection, as noted previously, takes place at the interface between an object and its surroundings. The direction of the exitant light energy is governed by the surface normal at the point of reflection; it is reflected through the local surface normal in the ''perfect specular direction.'' The amount of light reflected is determined by Fresnel's laws, whose parameters include the angles of incidence and emittance, the index of refraction of the material, and the polarization of the incoming light. For white metals and most man-made dielectrics the surface reflection can be considered constant over the visible spectrum [15, 16]. Materials whose surface reflection fits this assumption form a useful subset, shown in Fig. 12, and are said to have *neutral interface reflection* (NIR) [24]. The surface reflection from an NIR material is approximately the same color as the illumination. Common materials for which the surface reflection is more dependent upon wavelength include ''red metals'' such as gold, copper, and bronze, all of which modify the color of the reflected surface illumination [13].

Many materials displaying surface reflection are optically ''rough.'' They possess microscopic surfaces with local surface normals that differ from the macroscopic shape. A subset of these rough surfaces are those with roughness characteristics—such as microscopic slopes or heights—that have a Gaussian distribution. Several reflection models, such as Torrance–Sparrow and Beckmann–Spizzochino, have been developed for rough surfaces using a Gaussian distribution assumption for some surface characteristics [3, 29, 39]. These models fit into our taxonomy of transfer functions as shown in Fig. 12.

Metals are an example of a material that displays only surface reflection. Because of the nature of the metal atoms, virtually no light penetrates beyond the surface of the material. Metals have been modeled by the *unichromatic reflection model* [13], and most models for rough specular surfaces apply directly to metals [3, 39].

A more complex form of reflection, body reflection, occurs when light enters a surface and strikes colorant particles. The colorant particles absorb some of the wavelengths and re-emit others, coloring the reflection. The photons that are re-emitted go in random directions, striking other colorant particles, and ultimately exiting the surface as body reflection. Surfaces whose colorant particles re-emit equally all wavelengths of visible light form the ''white'' subset of transfer functions with body reflection.

Because of the stochastic nature of this reflection, a common assumption is that the body reflection is independent of viewing direction. Surfaces whose transfer func-

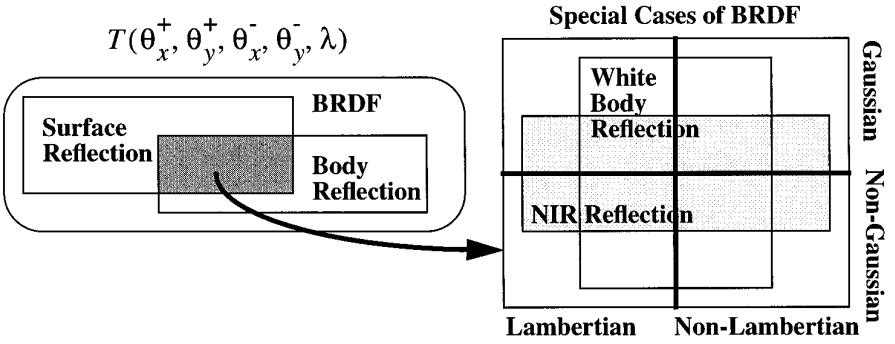

FIG. 12. Taxonomy of the bidirectional reflectance distribution function.

tions display this independence are called Lambertian because they obey Lambert's Law, which states that the reflection is dependent upon the incoming light's intensity and cosine of the angle of incidence [14]. Other models of body reflection that are dependent upon viewing direction are being researched [24, 41]. The white subset and Lambertian subset relationships are shown in Fig. 12.

Many interesting and useful transfer functions exhibit both body and surface reflection. Common materials simultaneously displaying these types of reflection include plastic, paint, glass, ink, paper, cloth, and ceramic, most of which can be modeled with the NIR assumption. Transfer functions within this overlapping region have been approximated by the *dichromatic reflection model* [38, 36].

For the purposes of our proposed segmentation method, we 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.* Metals, as noted previously display only surface reflection; dielectrics always have some body reflection, and often display surface reflection as well, although not as strongly as metals.

3. FUNDAMENTAL HYPOTHESES

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 = ℏplanar, curvedj
- Illumination environment = ℏdiffuse, uniform, generalj
- Transfer function = ℏmetal, dielectricj.

There are 12 possible combinations of these broad classes, subdividing the space of hypotheses for an image region into 12 subspaces. Each of these subspaces is parameterized by the color values (wavelength spectrum) of the illumination and the transfer function.

3.1. Generating the Fundamental Hypotheses

Because of the large number of possible color distributions, for the purpose of reasoning about hypotheses we further subdivide L^+ and $\Re$ into two classes: uniform spectrum (white or grey), and nonuniform spectrum (colored). This divides L^+ into six forms of illumination, and $\Re$ into four forms of the transfer function. We define the possible combinations of surface, illumination, and transfer function to be the set of *fundamental hypotheses* for an image region.

To denote a specific fundamental hypothesis we use the notation (<**transfer function**>, <**illumination**>, <**shape**>). The three elements of a hypotheses are defined as follows:

<**transfer function**> := Colored dielectric ℑ White dielectric ℑ Col. metal ℑ Grey metal

<**illumination**> := Col. diffuse ℑ White diffuse ℑ Col. uniform ℑ White uniform ℑ Col. complex ℑ White complex

<**shape**> := Curved ℑ Planar.

Simple combination of the classes of the hypothesis elements ($2 \times 6 \times 4$) indicates there are 48 possible hypotheses. However, not all 48 are applicable to every region. Consider first a colored region. To possess color, either L^+ or $\Re$ must have a nonuniform spectrum. If we remove from consideration the 12 uniform illumination/uniform transfer function hypotheses, 36 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 only 12 fundamental hypotheses for a uniform spectrum region. Therefore, for a given image region we have to consider at most 36 fundamental hypotheses.

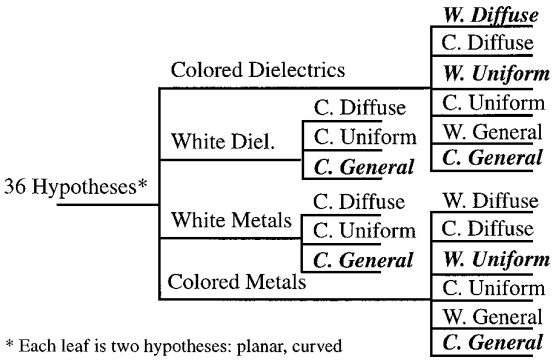

FIG. 13. Taxonomy of fundamental hypotheses. Primary (tier 1) hypotheses are emphasized.

To more explicitly show the structure of the fundamental hypotheses we arrange them as shown in Fig. 13 and Fig. 14. The trees represent taxonomies of the fundamental, or simplest hypotheses and classify the different physical explanations for gray and colored image regions. The leaves of these trees are a finite set of simple, comprehensive explanations for the color and brightness of every pixel within an image region. Using the set of fundamental hypotheses as the initial hypothesis list for each region, we can begin to reason about and merge hypothesis regions into more sensible global hypotheses that correspond more closely with what we consider to be objects in the scene that created the image.

3.2. Analyzing the Fundamental Hypotheses

The taxonomy of Fig. 13 implies that all of the fundamental hypotheses possess equivalent value for describing regions of an everyday scene. We believe this is not the case for most images. To concentrate our efforts on the more common hypotheses, we subdivide the 36 hypotheses into two groups, or tiers, reflecting how common or rare a hypothesis seems to be. Common hypotheses we place in tier one and less common hypotheses in tier two. For the purpose of brevity, we concentrate on the hypotheses for a colored image region. Note that a similar analysis applies to white and grey regions, which can also be divided into common and rare categories.

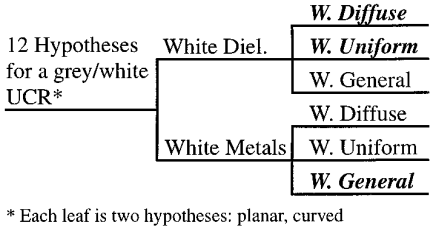

FIG. 14. Twelve fundamental hypotheses for a white/grey region. The six tier-one hypotheses are emphasized.

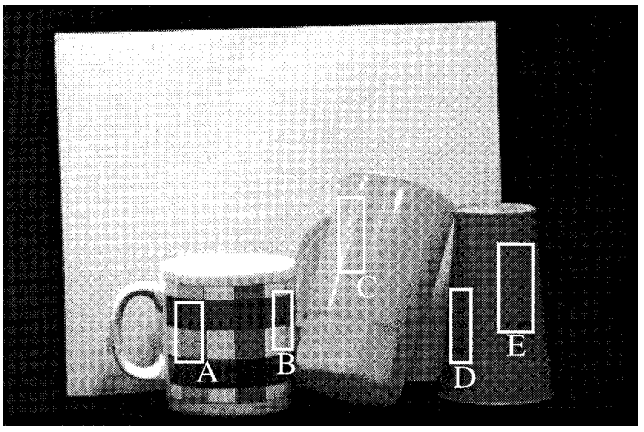

FIG. 15. A typical picture of single and multicolored dielectric objects with various illumination environments (Color Plate 7).

We begin with a structured analysis of each subtree of the taxonomy for the hypotheses of a colored image region, considering in turn each of the four classes of material. We are guided in our analysis by two general rules which take into consideration the estimated size relationships of subspaces of the taxonomies:

1. If a subspace is both common and a good approximation of a larger encompassing space, place the subspace in tier one, and the larger space in tier two.

2. If a subspace is both uncommon and not a good approximation of a common larger space, place the subspace in tier two and the larger space in tier one.

We begin by looking at the hypotheses concerned with colored dielectrics. These 12 hypotheses are grouped into six pairs according to the illumination environment. The first two, curved and planar dielectrics under diffuse white lighting are often used as a model for surfaces in shadow where no light source is directly incident [12]. An example of this case appears within box D of Fig. 15. Such situations are common in everyday pictures, compared with colored diffuse illumination such as might exist in a darkroom. Therefore, we place curved and planar colored dielectrics under diffuse white illumination in tier one, and colored dielectrics under colored diffuse illumination in tier two. Tier one hypotheses are highlighted in both Fig. 13 and Fig. 14.

The next two hypotheses, curved and planar colored dielectrics under uniform white illumination, represent a significant subset of surfaces in a typical scene such as Fig. 15. Boxes A and E are two examples. Sunlight can also be approximated by a uniform source when considering dielectrics because its effect on dielectric surfaces usually overwhelms any other illumination. Conversely, we argue that colored dielectrics under colored uniform illumination are rare and belong to tier two. Again, a darkroom would

be an example where there would be a colored light source and all diffuse illumination would have the same chromaticity.

Curved and planar dielectrics under general function white illumination are an interesting pair of hypotheses. In the real world, they are probably the most common hypotheses, as uniform and diffuse lighting are only approximations. In the case of dielectrics, however, uniform and diffuse lighting models are probably sufficient for most situations. The reason is that dielectrics, unlike metals, have a strong body reflection component; they reflect some of the light from each incident direction in each exitant direction. In the extreme case, a perfectly Lambertian surface reflects the incident light from a single direction equally in all directions. The exitant light energy field caused by a single strong incident light source can overshadow any additional exitant light energy due to illumination from other directions. Therefore, in scenes where there are one or more white light sources of possibly varying intensity, we propose that the illumination can be adequately modeled as a set of uniform brightness white sources. This obviates most of the need for white general illumination, allowing us to place it in tier two.

Curved and planar colored dielectrics under general colored illumination, however, are not well modeled by any other hypotheses in tier one. In everyday scenes these hypotheses are needed to model interreflection such as occurs in boxes B and C in Fig. 15. Because of this, we must place them in tier one.

The next major branch corresponds to the six hypotheses for white dielectrics under colored illumination. In common scenes we suggest that situations corresponding to these hypotheses are rare (e.g., darkroom). The most common occurrence of these is probably interreflection between a colored object and a white dielectric object such as a white wall. In these cases, the white object is lit by both a direct light source and some type of colored reflection from a nearby object. The illumination environment corresponding to this case can only be represented by a general function illumination environment as both the direct illumination and the interreflection are significant. The hypotheses corresponding to colored diffuse reflection are less common, generally occurring when the white object is in shadow from direct sources but still experiences reflection from a nearby colored object. Colored uniform sources—blue light bulbs, for example—are not common in human environments. Given this analysis, we propose that curved and planar dielectrics under general function colored illumination be placed in tier one, and the other four hypotheses in tier two.

White metals under colored illumination form the next major branch of the taxonomy. Unlike dielectrics, incident light from almost all directions is significant to the appearance of a metal surface patch. This can be seen in box G of Fig. 16, where interreflected light that is dim relative to the global light source still has a significant effect on the appearance of the metal object. For this reason, the hypotheses with general function colored illumination are the most common. It is rare for a metal surface to be lit only by colored uniform illumination, or to have the same color and intensity light incident from all directions as under diffuse illumination. Furthermore, unlike dielectrics, diffuse illumination environments are not good approximations because the exitant light energy field in a given direction is dependent on only one direction of the incident light energy field. Therefore, the two hypotheses with colored general function illumination belong to the first tier, and the other four hypotheses—colored diffuse and uniform illumination—belong to the second tier.

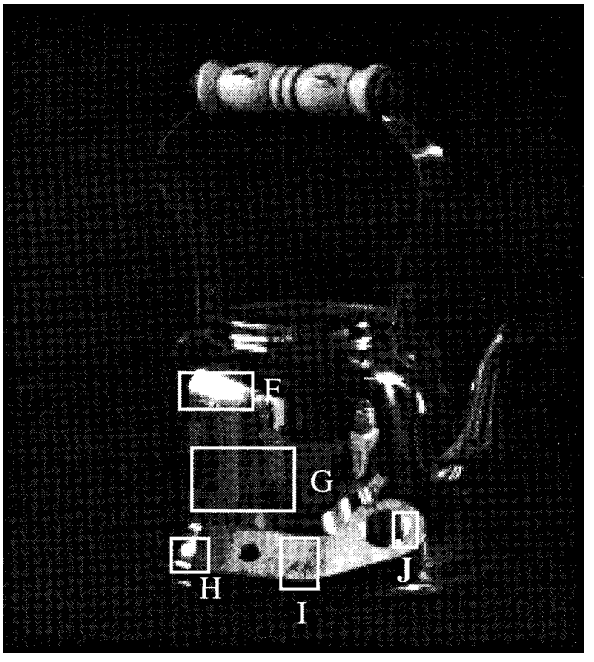

FIG. 16. A picture of a copper teakettle and a machined piece of steel (Color Plate 8).

The final branch of hypotheses contains the colored metals under white and colored illumination. Consider first the six hypotheses of colored metal under colored illumination. As with grey metals, hypotheses with colored general function illumination such as box G are the most common situations for colored metal objects. Colored uniform and diffuse illumination are not good approximations. This places the colored general illumination hypotheses in tier one, and the other four in tier two. With regard to the six white illumination hypotheses, we propose that uniform illumination is sufficient for modeling colored metal under white illumination such as box F of Fig. 16. True diffuse illumination is rare—the metal object will at least be reflecting the camera! We realize that the approximation of

general white illumination by white uniform may not be valid for all cases, but it is sufficient for our current discussion. From this analysis, the two hypotheses with uniform illumination belong in tier one; the other four belong in tier two.

The overall result of this analysis is that there are 14 common fundamental hypotheses in tier one, and 22 less common or rare fundamental hypotheses in tier two. Note that all seven illumination/transfer function combinations are present in either Fig. 15 or Fig. 16; all of these fundamental hypotheses can exist in deceptively simple images.

4. MERGING THE FUNDAMENTAL HYPOTHESES

Having developed a small set of physical hypotheses for describing a given image region, we attach this hypothesis list to each of the simple regions initially found in an image. In general, we define a segmentation of the image to be a set of hypotheses, one from each initial region that covers the image. To obtain a good segmentation, we need to minimize the number of hypotheses in the segmentation by combining hypotheses that are compatible (i.e., that appear to belong to the same object by some criterion). The combination of compatible hypotheses is the key to obtaining an intelligent segmentation.

A brute force approach would look at all combinations of the fundamental hypotheses for each pair of adjacent regions. Unfortunately, a brute force method is, not only unreasonable, but also too computationally expensive for even simple images because of the exponential explosion of the number of hypotheses. For this segmentation method to be tractable, the interaction between hypothesis regions and the nature of the physical explanations must provide constraints.

For a merger between regions to be desirable, there must be some coherence between the hypothesized physical explanations. This coherence manifests itself in the three general variables: shape, illumination, and transfer function. If two neighboring hypotheses are sufficiently similar, it may be a desirable merger. By definition there must be a discontinuity between neighboring regions. The particular form of this discontinuity is dependent upon the initial segmentation method. This implies a discontinuity in at least one of the hypothesis elements. Because of the general viewpoint principle—things do not line up for almost all viewpoints [37]—having two simultaneous discontinuities along the border of adjacent hypothesis regions is an unlikely occurrence if the regions belong to the same object. Therefore, we propose that 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.*

In addition to this general postulate, we propose four other rules:

1. hypothesis regions of differing materials should not be merged (this includes differently colored metals such as Box I in Fig. 16),

2. hypothesis regions with incoherent shape boundaries should not be merged,

3. hypothesis regions of differing color that propose the physical explanation to be colored metal under white illumination should not be merged, and

4. hypothesis regions proposing different color diffuse illumination should not be merged.

While the first rule may be restrictive at a more abstract level—e.g., object recognition—it is necessary to make the problem tractable. It can also be argued that combining different material types (e.g., metals and dielectrics) is not appropriate for a low-level segmentation algorithm. The second rule is necessary so that overlapping objects with similar characteristics are not merged. The third rule results from the fact that the surface reflection, or material properties of the surface, determine the color of hypotheses proposing colored metal under white illumination. Therefore, if two of these hypothesis regions differ in color but have the same illumination environment, they must be different materials and should not be merged.

The last rule is due to the physics of illumination. Diffuse illumination specifies that the color and intensity of the illumination is constant over the illumination hemisphere. Now consider two adjacent surface patches under differently colored diffuse illumination. If the adjacent patches are at less than a 180° angle, there will be overlap between the illumination environments. This situation is impossible unless the illumination is such that each point on the illumination hemisphere appears one color from one appearance patch and a different color from the adjacent appearance patch. Such an illumination environment is unlikely at best and is reasonably discarded.

The result of applying these rules to the merger of two adjacent image regions is shown in Fig. 17. Instead of having to consider 196 combinations, we only need to look at 28. The importance of this result is that we *do not increase* the number of hypotheses being considered for the entire scene. Instead of having 14 hypotheses each for two regions we now have 28 hypotheses for the composite region. Of course, if you want to keep around the old regions as well this doubles the amount of resources you need. However, the rules reduce the number of mergers that need to be considered by a factor of 7.

4.1. Merger Analysis

As shown in Fig. 17, there are 28 potential mergers that must be considered for each pair of adjacent hypothesis regions; a merger is desirable to make if it can be ascertained that only a single discontinuity exists between the two regions.

The single discontinuity requirement imples that the shape of the two regions must be coherent in some well-defined sense. As the defining characteristic of the initial regions is coherence in color space, shape cannot be the cause of a boundary forming between two regions of an image if they are part of the same object; there must be a discontinuity in either the transfer function or the illumination. Therefore, if the borders between two regions are not continuous (e.g., at least C^1) or coherent in some manner (e.g., the edges of a cube), then no merger should be considered.

This implies that shape plays a major role in blocking or allowing mergers between regions for all 28 possible cases. Knowing that shape is a factor for all mergers, we need only analyze in detail the 10 illumination/transfer function combinations. In the interest of brevity, we preform a detailed analysis of the merge requirements for only one case: row 2, column 2 of Fig. 17. This box represents a merger between two colored dielectrics under white uniform illumination. A clear example of this case is shown in box A of Fig. 15. The reasoning process used to analyze this case can be extrapolated to the other 9 illumination/transfer function combinations.

The first task in the reasoning process is to determine which element of the hypothesis will have a discontinuity if the two regions are merged. This is equivalent to asking which element of the hypotheses causes the color change between the regions. For this case, the cause of the color change must be the transfer function as the illumination for both regions is white. Therefore, neither the illumination nor the shape of the two regions can be discontinuous for a merger to be possible.

The next task is to determine the nature of the discontinuity. In this case the discontinuity is a change of color in the transfer function. The question is whether the other characteristics changed as well. If the two regions belong to the same object in a scene, it is reasonable to assume that the surface patches have similar properties (e.g., degree of specularity, roughness). A discontinuity in these color independent properties can be used as evidence to block a merger. Likewise, strong similarity encourages a merger of the two regions.

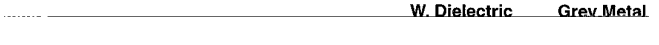

FIG. 17. Shaded boxes indicate potential merges of the 14 tier-one hypotheses. Merges are desirable if the shapes match.

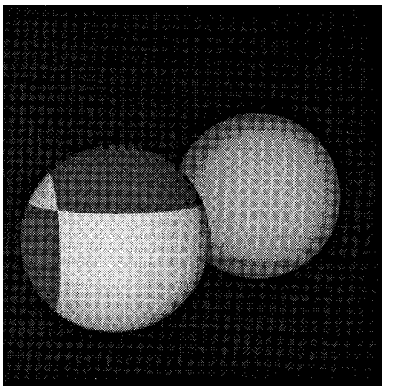

FIG. 18. Synthetic test image of two spheres (Color Plate 9).

Knowing where discontinuities are expected and where they are not is the key to applying vision operators to the two regions. We want to obtain measures of similarity which will allow or block a merger of the two regions. Clearly, there are some cases where the lack of tools may make analysis difficult. In particular, dielectrics under general illumination and metals under any illumination present a significant challenge to shape analysis. There are also cases where there are sufficient vision tools to perform the necessary calculations. In the next few sections we discuss several tools of analysis and how we use them in the merging process.

5. INITIAL SEGMENTATION

To test the segmentation method, we use simple pictures of piece-wise uniform multicolored objects on a black background. Figure 18 and Figure 19 are two example test images. Figure 18 is a synthetic image created using Rayshade (a public domain ray tracer). Figure 19 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 (3),

$$(c_{nr}, c_{ng}, c_{nb}) = \left(\frac{r}{r+g+b}, \frac{g}{r+g+b}, \frac{b}{r+g+b}\right), \quad (3)$$

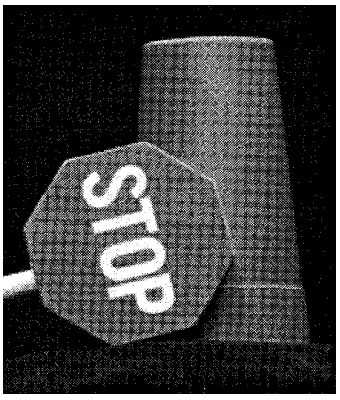

FIG. 19. Real test image of a stop-sign and a cup (Color Plate 10).

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 the 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 a 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 [13], Klinker's linear and planar cluster algorithm [17], or Bajcsy *et al.*'s normalized color method [2]? There are legitimate problems with using any of these methods. Healey'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 fast and adequate. Figure 20 and Figure 21 show examples of the initial segmentations and are handlabeled 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. For our initial implementation of the segmentation method we consider the hypothesis set $H_c = \hbar$(Colored dielectric, White uniform, Curved), (Colored dielectric, White uniform, Planar)j for colored regions and the hypothesis set $H_w = \hbar$(White dielectric, White uniform, Curved), (White dielectric, White uniform,

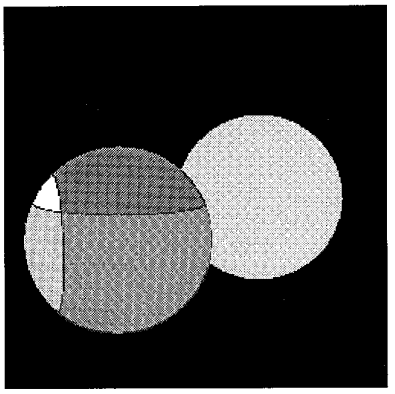

FIG. 20. Initial segmentation of test image 1.

Planar)j 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

$$(c_{nr} - 0.333)^2 + (c_{ng} - 0.333)^2 + (c_{nb} - 0x333)^2 < 0.0016, \quad (4)$$

where (c_{nr}, c_{ng}, c_{nb}) is the average normalized color of the region defined by Eq. (3). The threshold was set based upon the images in the test set.

6. 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 hypothesis graph with which we obtain the most likely final segmentations of the image.

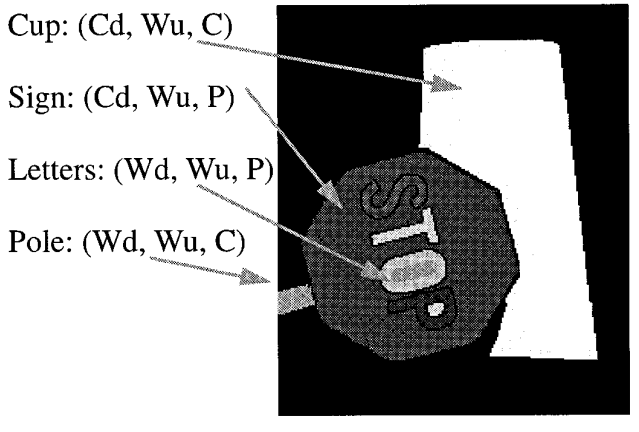

FIG. 21. Initial segmentation of test image 2.

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 hypothesis 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.

6.1. Direct Instantiation

If we can estimate and represent each hypothesis element, merging adjacent regions involves looking at the table in Fig. 17 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 Bichsel and Pentland's shape-from-shading (SFS) algorithm and Zheng and Chellappa's illuminant and albedo estimation algorithm [4, 44]. Bichsel and Pentland's SFS algorithm was chosen because, according to the survey by Zhang *et al.,* it is one of the best methods when the illumination comes from the side [43]. Zhang and Chellappa's illuminant estimate was selected because it is a locally calculated method, and they showed their method produced better results than Pentland's or Lee and Rosenfeld's methods [33, 9, 44].

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 **L** makes with the x-z plane, and the slant is the angle between **L** and the z-axis.

The first step after the initial segmentation is to analyze

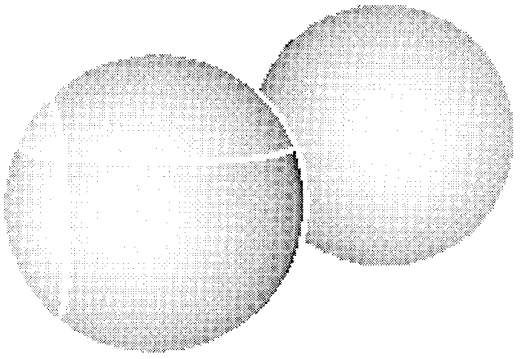

FIG. 22. Shape from shading result. Displayed intensity decreases with depth.

each region independently. Figure 22 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, we employ a two-step algorithm. First, we find the optimal offset, in a least-squares sense, of the two regions 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 [20]. The resulting 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 23 shows the sum-squared error per pixel for each region pair in the synthetic test image. We show the sum-squared error per pixel because the chi-square test results 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 compare the illuminant direction estimates of adjacent regions we convert the tilt and slant angles for each region to a 3D 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 Fig. 23 are unchanged when we consider the transfer function and illumination.

As nicely as the direct instantiation method worked on the synthetic test image, the analysis tools have serious problems with slightly more complicated images. First, Bichsel and Pentland's SFS algorithm requires an accurate indication of the illuminant direction and albedo and also it requires good initial point selection [43]. 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 Bichsel and 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 [43]).

The second serious problem is with the illuminant direction estimator. Besides the assumption that the illumination is a point source, Zhang and Chellappa's algorithm requires a good distribution of surface normals to correctly estimate the tilt and slant [44]. 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

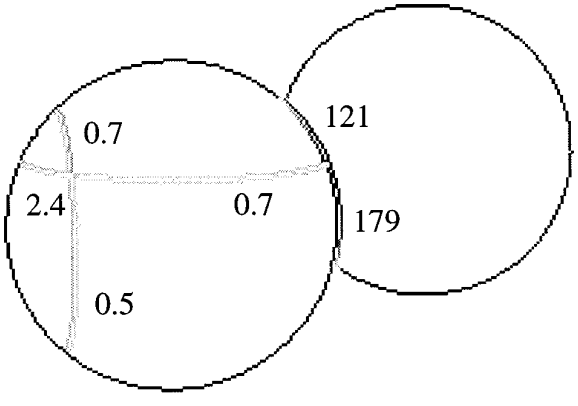

FIG. 23. Border shape comparison. Darker borders indicate larger errors. Average sum squared error per pixel shown for each region pair.

illuminant estimator is better able to divine the correct direction because Zhang and Chellappa's slant estimator is dependent upon intensity variation rather than the distribution of gradients. However, for a test image where the two spheres are illuminated from above and to the right at a 27° slant angle, the illuminant estimator does not work as well. We found tilt errors of up to 102°, and slant errors of up to 27° for the synthetic test image.

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.* [5] may overcome some of these difficulties in the future, for now we take a different approach.

6.2. 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.*

6.2.1. *Reflectance Ratio*

One physical characteristic we use is the reflectance ratio for nearby pixels as defined by Nayar and Bolle [30]. 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 Eq. (5), 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 [30]:

$$r = \left(\frac{I_1 - I_2}{I_1 + I_2}\right). \tag{5}$$

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, then the shape and illumination do not have to be the same.

TABLE 1
Reflectance Ratio Results for $\mathrm{Var}_N = 0.004$

Region A	Region B	Reflectance ratio	Refl. ratio variance	$P(\mathrm{Var}_R < \mathrm{Var}_N)$
Red region	S region	0.4463	0.0004	1.0
Red region	T region	0.4449	0.0005	1.0
Red region	O region	0.4503	0.0004	1.0
Red region	P region	0.4541	0.0006	1.0
Red region	Cup region	0.2107	0.0125	0.0
O hole	O region	−0.4358	0.0008	1.0
P hole	P region	−0.4562	0.0004	1.0
While pole	Red region	0.1709	0.0710	0.0

Note. The last column shows the probability that the variance is the variance due to noise.

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

$$\mathrm{Var} = \sum_{i=1}^{N} \frac{(r_i - r_{\mathrm{avg}})}{N-1}, \tag{6}$$

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 quantization 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. This should result in a larger variance in the reflectance ratio. We should be able to find a standard variance based upon the noise and quantization effects and use it to differentiate between these two cases. Table 1 shows the variances in the border 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.

6.2.2. *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 to reduce noise in the gradient direction estimate caused by the discontinuity in the transfer function, we first calculate the gradient direction for all pixels in the region except the border pixels. 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. 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. Table 2 shows the results of applying the gradient direction characteristic to the synthetic test image.

6.2.3. *Intensity Profile Analysis*

So far, we have examined only 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 Fig. 24, 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 continuous. On the other hand, for the scanline B to B', the curves are not 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 [35], 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 Eq. (7), 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 [35]:

$$DL = -\log P(x^n \| \theta) + \frac{k}{2} \log n. \tag{7}$$

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

TABLE 2
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

Note. The last column shows the result of a chi-square test with $\mathrm{Var}_N = 0.2$ radians.

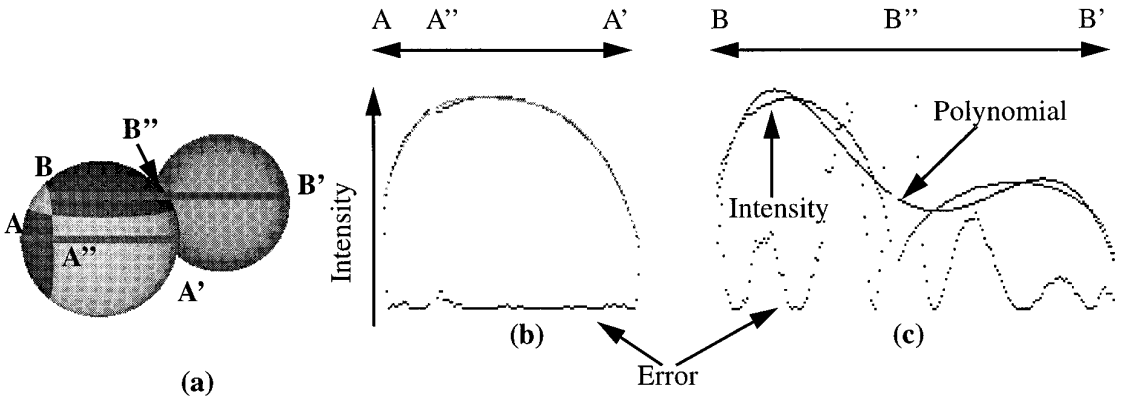

FIG. 24. Test image shown in (a). Graphs (b) and (c) are the intensity profiles, least-squares polynomial, and squared error for the image segments $A - A'$ and $B - B'$, respectively.

3. Model the scaled intensity profile of scanline s_0 for both h_1 and h_2 as a polynomial, and find the best order using MDL. Assign the smallest description length to M_c.

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

The result of this test is a merge/do not 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 3 shows the results of this analysis applied to the stop-sign and cup test image. Note that a M of 8 would represent an adequate threshold for correctly merging all but one region pair. For the synthetic image, a M 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 M could be used for all region pairs.

TABLE 3
Results of Intensity Profile Analysis for Stop-Sign and Cup Image

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

Note. If the far right column is close to or greater than 0, then the regions are better modeled by a single polynomial.

7. 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 4 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 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 Fig. 17. All other edges have a weight of 0.0, indicating that they should not be merged in any segmentation.

Note that each edge is actually two edges: a merge edge, and a not-merge edge. The weight assigned to the merge 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 likeli-

TABLE 4
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

hood 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 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 Fig. 25. Hypothesis A for region 1 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 Fig. 25a the segmentation A–C is the best. In the case shown in Fig. 25b, because the merge likelihood of A and C is only 0.45, 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 hypothesis 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 Fig. 25a 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 Fig. 18 and Fig. 19 are shown in Fig. 26 and Fig. 27, 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.

Algorithms do exist for finding step-wise optimal segmentations of images given likelihoods that regions should be merged. LeValle and Hutchinson [21] and Panjwani and Healey [32] have both used this algorithm to segment textured scenes. These algorithms would work unmodified on a single slice of a hypothesis graph (i.e., one hypothesis per region). A modified version of this algorithm may

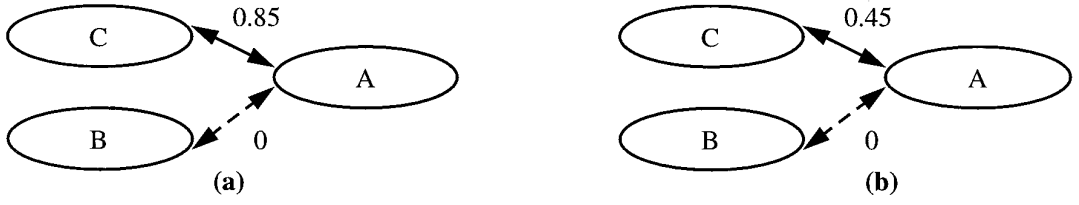

FIG. 25. Potential hypothesis graphs. In (a) the best choice is to merge *A* and *C*. In (b) the best choice is to select incompatible hypotheses.

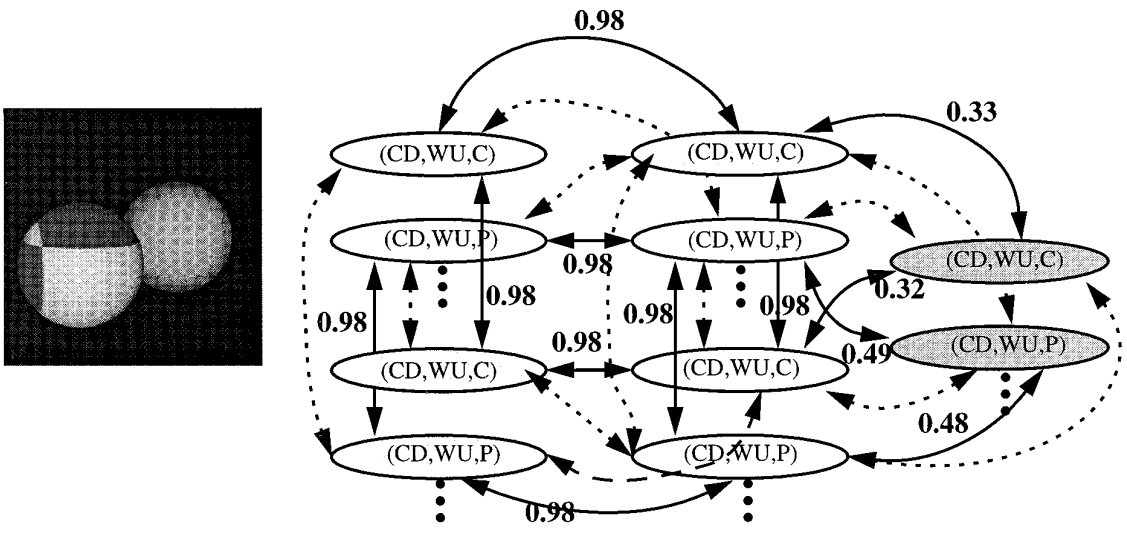

FIG. 26. Two layer hypothesis graph for the synthetic test image. Dashed edges indicate incompatible hypotheses with a merge likelihood of 0, and a not-merge likelihood of 0.5. Note, as more hypotheses are included, the region graph simply gets more levels.

be applicable to the hypothesis graphs we generate. The difference with previous applications is that our algorithm uses multiple hypotheses per region.

8. 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

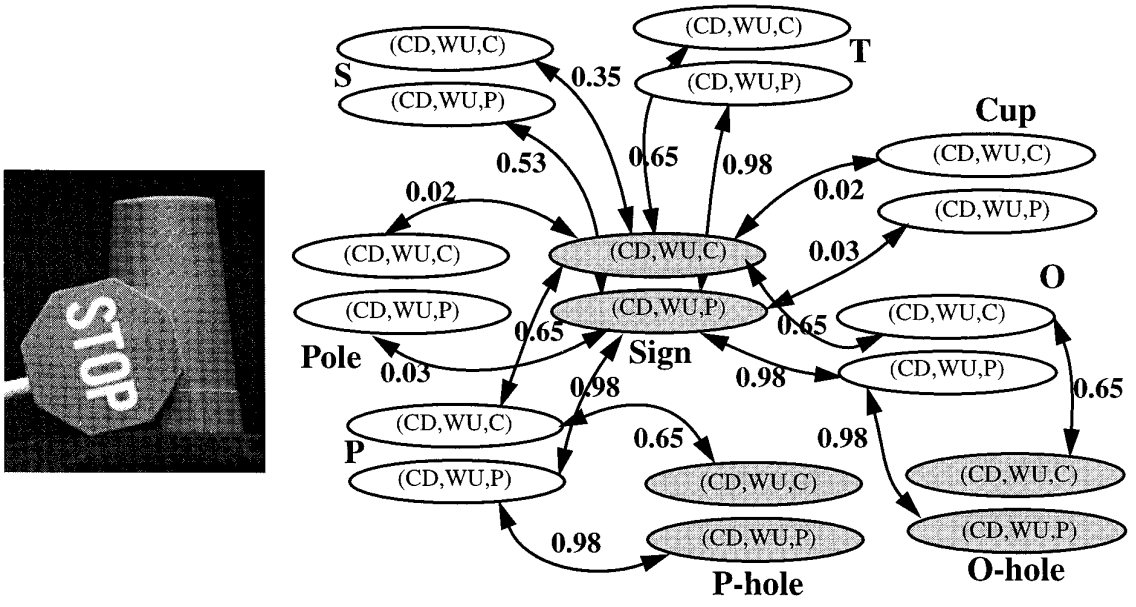

FIG. 27. Two layer hypothesis graph for the stop-sign and cup image. Zero edges not shown. No edges exist between hypotheses for the same region.

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.

9. 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 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.

ACKNOWLEDGMENTS

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.

REFERENCES

1. E. Addison and J. Bergen, The plenoptic function and the elements of early vision, in *Computational Models of Visual Processing* (M. S. Landy and J. A. Movshon, Eds.), MIT Press, Cambridge, MA, 1991.
2. R. Bajcsy, S. W. Lee, and A. Leonardis, Color image segmentation with detection of highlights and local illumination induced by interreflection, in *Proc. International Conference on Pattern Recognition,* Atlantic City, NJ, 1990, pp. 785–790.
3. P. Beckmann and A. Spizzochino, *The Scattering of Electromagnetic Waves from Rough Surfaces,* Artech House, Norwood, 1987.
4. M. Bichsel and A. P. Pentland, A simple algorithm for shape from shading, in *Proceedings, IEEE Conference on Computer Vision and Pattern Recognition, 1992,* pp. 459–465.
5. P. Breton, L. A. Iverson, M. S. Langer, and 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.
6. C. R. Brice and C. L. Fenema, Scene analysis using regions, *Artif. Intell.* 1, 1970, 205–226.
7. M. H. Brill, Image segmentation by object color: A unifying framework and connection to color constancy, *J. Opt. Soc. Am. A* **7**(10), 1990, 2041–2047.
8. M. Born and E. Wolf, *Principles of Optics,* Pergamon, London, 1965.
9. M. J. Brooks and B. K. P. Horn, Shape and source from shading, in *Proceedings, Int'l Joint Conf. on Artificial Intelligence, August 1985,* pp. 932–936.
10. M. Cohen and D. Greenberg, The hemi-cube: A radiosity solution for complex environments, *Computer Graphics Proc. of SIGGRAPH-85, 1985,* pp. 31–40.
11. T. Darrell, S. Sclaroff, and A. Pentland, Segmentation by minimal description, in *Proceedings, International Conference on Computer Vision, IEEE, 1990,* pp. 112–116.
12. J. D. Foley, A. van Dam, S. K. Feiner, and J. F. Hughes, *Computer Graphics: Principles and Practice,* 2nd ed., Addison Wesley, Reading, MA, 1990.
13. G. Healey, Using color for geometry-insensitive segmentation, *J. Opt. Soc. Am. A* **6**(6), 1989, 920–937.
14. B. K. P. Horn, *Robot Vision,* MIT Press, Cambridge, MA, 1986.
15. R. S. Hunter, *The Measurement of Appearance,* Wiley, New York, 1975.
16. D. B. Judd and G. Wyszecki, *Color in Business, Science, and Industry,* 3rd ed., Wiley, New York, 1975.
17. G. J. Klinker, S. A. Shafer, and T. Kanade, A Physical approach to color image understanding, *Int. J. Comput. Vision* **4**(1), 1990, 7–38.
18. J. Krumm, *Space Frequency Shape Inference and Segmentation of 3D Surfaces,* Ph.D. thesis, CMU-RI-TR-93-32, Carnegie Mellon University, December 1993.
19. M. S. Langer and S. W. Zucker, A ray-based computational model of light sources and illumination, in *IEEE Workshop on Physics-Based Modelling in Computer Vision, Cambridge, MA, June 1995,* pp. 93–99.
20. L. Lapin, *Probability and Statistics for Modern Engineering,* PWS Engineering, Boston, 1983.
21. S. M. LaValle and 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.
22. Y. G. Leclerc, Constructing simple stable descriptions for image partitioning, *Int. J. Comput. Vision* **3,** 1989, 73–102.

23. H.-C. Lee, Method for computing the scene-illuminant chromaticity from specular highlights, *J. Opt. Soc. Am. A* **3**(10), 1986, 1694–1699.

24. H.-C. Lee, E. J. Breneman, and C. P. Schulte, Modeling light reflection for color computer vision, *IEEE Trans. Pattern Anal. Mach. Intell.* **PAMI-12**(4), 1990, 402–409.

25. A. Leonardis, *Image Analysis Using Parametric Models: Model-Recovery and Model-Selection Paradigm,* Ph.D. thesis, LRV-93-3, University of Ljubljana, March 1993.

26. A. Leonardis, A. Gupta, and R. Bajcsy, Segmentation as the search for the best description of the image in terms of primitives, in *Proceedings, International Conference on Computer Vision, IEEE, 1990,* pp. 121–125.

27. 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, 1994,* pp. 361–368.

28. P. H. Moon and D. E. Spencer, *The Photic Field,* MIT Press, Cambridge, MA, 1981.

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

30. S. K. Nayar and R. M. Bolle, ''Reflectance Based Object Recognition,'' *Int. J. Comput. Vision,* **17,** No. 3, 1996, pp. 219–40.

31. 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.

32. D. Panjwani and G. Healey, Results using random field models for the segmentation of color images of natural scenes, in *Proceedings, International Conference on Computer Vision, June 1995,* pp. 714–719.

33. A. P. Pentland, Finding the illuminant direction, *J. Opt. Soc. Am.* **72**(4), 1982, 448–455.

34. A. P. Petrov and L. L. Kontsevich, Properties of color images of surfaces under multiple illuminants, *J. Opt. Soc. Am. A, Opt. Image Sci. Vis. (USA)* **11**(10), 1994, 2745–2749.

35. J. Rissanen, *Stochastic Complexity in Statistical Inquiry,* World Scientific, Singapore, 1989.

36. S. A. Shafer, Using color to separate reflection components, *COLOR Res. Appl.* **10,** 1985, 210–218.

37. J. M. Tenenbaum, M. A. Fischler, and H. G. Barrow, Scene modeling: A structural basis for image description, in *Image Modeling* (A. Rosenfeld, Ed.), Academic Press, New York, 1981.

38. S. Tominaga and B. A. Wandell, Standard surface-reflectance model and illuminant estimation, *J. Opt. Soc. Am. A* **6**(4), 1989, 576–584.

39. K. Torrance and E. Sparrow, Theory for off-specular reflection from roughened surfaces, *J. Opt. Soc. Am.* **57,** 1967, 1105–1114.

40. R. Willson, *Modeling and Calibration of Automated Zoom Lenses,* Ph.D. thesis, Carnegie Mellon University, CMU-RI-TR-94-03, January 1994.

41. L. B. Wolff, *A Diffuse Reflectance Model for Dielectric Surfaces,* The Johns Hopkins University, Computer Science TR 92-04, April 1992.

42. Y. Yakimovsky and J. Feldman, A semantics-based decision theory region analyzer, in *Proceedings, 3rd International Joint Conference on Artificial Intelligence, 1973,* pp. 580–588.

43. R. Zhang, P. S. Tsai, J. E. Cryer, and M. Shah, Analysis of shape from shading techniques, in *Proceedings, IEEE Conference on Computer Vision and Pattern Recognition, June 1994,* pp. 377–384.

44. Q. Zheng and R. Chellappa, Estimation of illuminant direction, albedo, and shape form shading, *IEEE Trans. Pattern Anal. Mach. Intell.* **13**(7), 1991, 680–702.



