A framework for segmentation using physical models of image formation

Bruce A. Maxwell and Steven A. Shafer

CMU-RI-TR-93-29

Robotics Institute
Carnegie Mellon University
Pittsburgh, Pennsylvania 15213

10 December 1993

This research was partially supported by the Avionics Laboratory, Wright Research and Development Center, Aeronautical Systems Division (AFSC), U.S. Air Force, Wright-Patterson AFB, OH 45433-6543 under contract F33615-90-C-1465, ARPA Order No. 7597. 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.

Abstract

Most approaches to computer image segmentation group sets of pixels according to visible features of an image such as edges, color, brightness, and curvature. Such approaches exploit specialized object properties to obtain satisfactory groupings, which can force those techniques to be domain specific. Furthermore, they do not provide a physical explanation for the image, nor do they group regions that have a single physical structure yet differing visible features.

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

This approach to segmentation has the potential to provide a list of possible explanations for a given image; to group together regions with coherent physical properties; and to provide a framework for applying specific operators such as shape-from-shading, color constancy, and roughness evaluation as part of the overall process of low-level vision. However, many profound unsolved problems are raised in determining the most "plausible" explanations for a given image region. In this paper, we present the approach, working through an example by hand, and discuss the implications of this approach for physics-based vision.

1. Introduction

The goal of physics-based segmentation is to find image regions that correspond to semantic scene elements. In practical terms, this means finding one or more physical descriptions of the illumination, materials, and geometry that created the image. In this presentation, we focus upon the problem of segmenting a single color image. That a solution exists for humans is obvious: an individual can look at a picture such as Figure 1 and not only comprehend what the picture is about, but provide a fairly detailed physical description of the scene. We believe that postulating such a physical description is the key to understanding image data.

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 realized that using information about the scene was important, but they tried to incorporate such knowledge (such as trees are above and beside a road) on top of their statistical models [41].

The statistical approach was taken partly because of the optimism of the 70's surrounding symbolic reasoning and artificial intelligence, which relegated to low-level vision the straightforward task of dividing an image into simple regions based upon color and brightness. More extensive low-level processing was considered unnecessary because it was assumed that programs using higher level reasoning would be able to understand, identify, and merge these simple regions as appropriate [35].

In the mid-70's, Horn proposed using physical models of image formation--the interaction of light and matter--to analyze and understand images [16]. Theoretically, using Horn's model some physical characteristics of a surface, including shape, could be estimated from a single image. Unfortunately, Horn's model was limited to perfectly diffuse, perfect reflecting surfaces (also called Lambertian surfaces) and point light sources, and assumed a single surface and light source in the scene. Furthermore, as it did not allow for noisy images or camera limitations--i.e. clipping of the color values to the camera's range--it was not easily applicable to real images.

In the mid-80's, Shafer's dichromatic reflection model [33] allowed researchers to begin looking at a large class of actual materials: inhomogeneous dielectrics. Inhomogeneous dielectrics include paints, plastics, acrylics, ceramics, and paper. Klinker *et al.* [20][21] demonstrated the power of this model, and the physics-based vision approach, by using it in tandem with a model for noise and camera effects to segment real images of inhomogeneous dielectrics.

Despite the power of this segmentation program, it was still applicable to a limited class of images. Metals or multicolored objects could not be correctly segmented. Furthermore, the assumptions of Klinker *et al.* included a single color of illumination. This resulted in incorrect segmentations in regions with colored interreflection from nearby objects.

Finding solutions for these limitations was the next step in physics-based vision. Bajcsy *et al.* [2] attempted to model interreflection and improve the parameter estimation methods of Klinker *et al.* by using hue, saturation, and intensity. Brill [7] proposed a slightly different model for inhomogeneous dielectrics and demonstrated its use in segmentation. Healey [14] proposed the unichromatic reflection model for metals, and showed that it could be used with the dichromatic reflection model to segment images with both metals and inhomogeneous dielectrics under specific lighting conditions.

As a result of these efforts, the vision community could claim it could segment images containing two materials--inhomogeneous dielectrics and metals--and images containing interreflection, but both methods had limitations. To correctly model interreflection, for example, a white reference plate was necessary in order to negate the effects of the global illumination. Furthermore, there are still a large number of materials and lighting conditions that cannot be handled by these models and their variations. More comprehensive reflection models, and models for different types of materials are being researched, but no general reflection model yet exists (e.g. see [39], [29], [31], [8], and [15]). Up to the present, physics-based segmentation routines for single color images have been based upon one, or at most two, specific models of reflection with a set number of parameters. Furthermore, the issue of differing types of illumi-

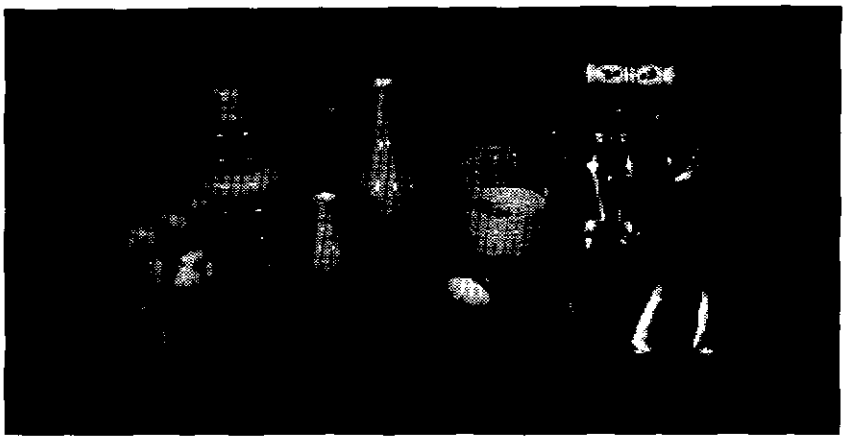


Figure 1. (Plate 1) A complex scene composed of numerous materials, textures, and shapes.

Figure 2. (Plate 2) An object, a mirror image of the object, and a picture of the object.

nation has not been examined, and all of the major work in segmentation has assumed uniformly colored objects.

In parallel with this work, the computer vision community has looked into determining light source color [24], and continued to work on determining shape, although mostly with range data (e.g. see [11] [26] [27]). Unlike the work in segmentation, which assumes all of the objects in an image conform to the same model, in the area of shape recovery model *selection* as well as parameter estimation is being used. Large families of models are initially considered for a set of data, and the best model is selected, as well as the best estimation of its parameters.

Recently, Breton *et al.* [5] have combined shape, light source direction, and material consistency into a single segmentation routine. They initially propose a family of models for light source direction and shape, but they assume a single model--Lambertian--for the reflectance properties of the material.

Unfortunately, none of these systems can deal with a pictures such as Figure 1. It contains grey and colored metals reflecting multi-colored illumination, and numerous dielectrics with differing reflectance properties. In order to begin to understand general images such as this, the next logical step is to begin looking at the families of possible models for all three elements of a scene--illumination, reflectance, and shape or geometry. The need for a general model of illumination is apparent from the metal teapot on the right side of Figure 1. Unless we can explicitly model the illumination from all directions with respect to the surface of the teapot, we cannot understand that the color variation is due to both the material type (copper), and the illumination. A comprehensive reflectance model, or at least a specification of the space of possible models, is also necessary in order to segment general images, as shown by the previous discussion.

Figure 3. (Plate 3) Image of uniformly colored inhomogeneous dielectrics.

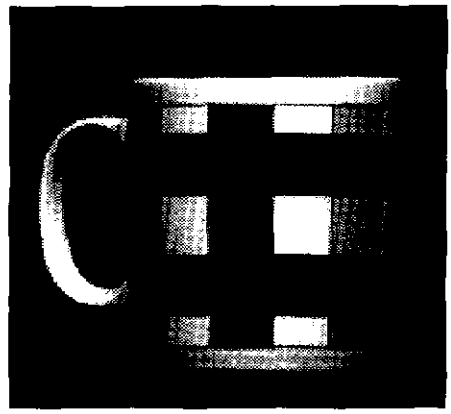


Figure 4. (Plate 4) Image of a single multi-colored object.

In the past, researchers have approached the analysis of such images by postulating particular model equations, and instantiating their parameters, with discontinuities in the parameters taken as segmentation boundaries. Instead, we 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 model of Rissanen [32]--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. Model selection is necessary because of ambiguity in an image. As can be seen in Figure 2, there can be several different physical explanations for identical image regions. But what are the general models we should use? What are the parameters of the model classes we need to consider, and do we need to consider them all? If not, how do we choose an initial set of models, and how do they merge and interact?

These are the questions we 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 choosing a subset of the possible models with which to begin segmenting an image. Finally, in section 4 we propose a method for merging and analyzing the different model hypotheses to obtain a global segmentation.

Using this method allows both multiple explanations for the same image region, and grouping together of regions that display coherence in one or more of the elements of their physical explanation. For example, physics-based vision can segment images such as Figure 3 [21]. The discontinuities in color in Figure 4, however, cause current methods to fail for this common image. Only by using more general models for segmentation can the image region corresponding to the entire cup be proposed as a single semantic entity.

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 new formal model of these elements, how they interact, and how they are related to what we see in an image.

2.1. The Elements of a Scene

The elements constituting our model of a scene are surfaces, illumination, and the light transfer function or reflec-

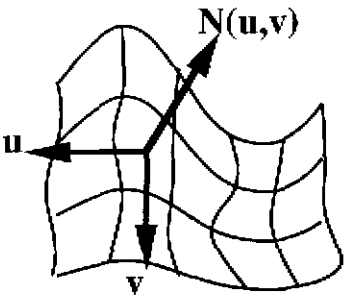


Figure 5. Local coordinate system on a surface patch.

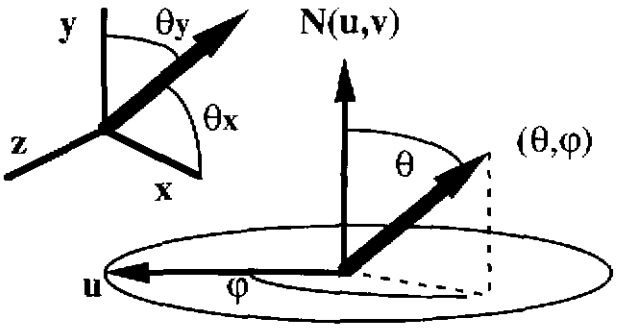


Figure 6. Specifying direction in the global and local coordinate systems.

tance of a point in 3-D 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 [36]. We begin by providing a formal notation for each of these elements.

2.1.1. Surfaces

We model objects in the real world using 2-D manifolds we call *surfaces*. On a given surface, we can define local coordinates as a two-variable parameterization (u, v) relative to an arbitrary origin. The shape of the manifold in 3-D space is specified by a *surface embedding* function $S(u, v) \rightarrow (x, y, z)$, defined over an extent $E \subseteq (u, v)$. The surface embedding function maps a point in the local coordinates of the manifold to a point in 3-D global coordinates. This global coordinate system is also anchored to an arbitrary origin, often specified relative to an imaging device. As shown in Figure 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 3-D 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. Throughout this presentation we use wire-frame diagrams, such as Figure 5 to show the shape of a surface patch.

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 in the local coordinate systems, we will use normal spherical coordinates, as shown in Figure 6, specified by the ordered pair (θ, φ). θ is the *polar angle*, defined as the angle between the surface normal and the direction, 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).

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. 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. In the field of machine vision, interreflection has been studied between two objects, but still no general model exists for specifying the totality of illumination on a surface point.

To begin examining general images we can't assume point lighting, three independent light sources, or other constructed illumination setup. 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, then extending this model to the 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 760nm, 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 [4], which we indicate by the variable $s \in \{1, 2, 3, 4\}$ 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 $(x, y, z, \theta_x, \theta_y, \lambda, s, t)$.

For the purposes of image formation, we want to specify the intensity of visible light that is incident from all directions on points (x, y, z) in global 3-D coordinates. We 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 radiant intensity, or radiance per unit solid angle, of light incoming to the point (x,y,z) from direction (θ_x, θ_y) of wavelength λ and Stokes parameter s at time t. This function is similar to the *plenoptic function* defined in [1], or the *helios function* [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 [9]. 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 3-D point (x, y, z) to be the *global illumination environment* [GIE] 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 Figure 7. To give some simple examples, several common illumination environments can be visualized as in Figure 8, Figure 9, and Figure 10. A simple example 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 coordi-

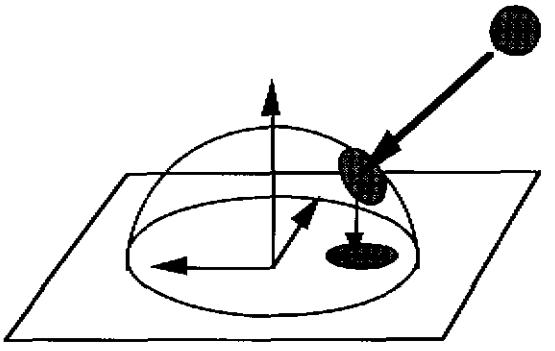


Figure 7. Orthogonal mapping of the illumination environment onto a plane.

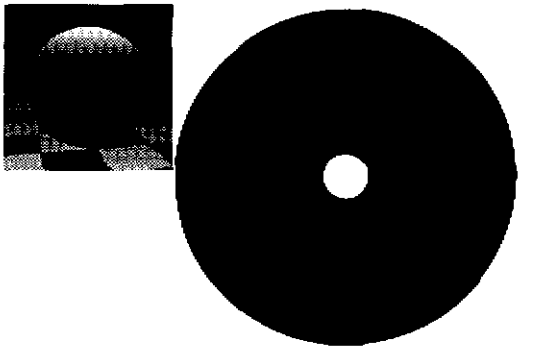


Figure 8. (Plate 5) Illumination environment for inset image: orthogonal mapping of a white light source directly overhead.

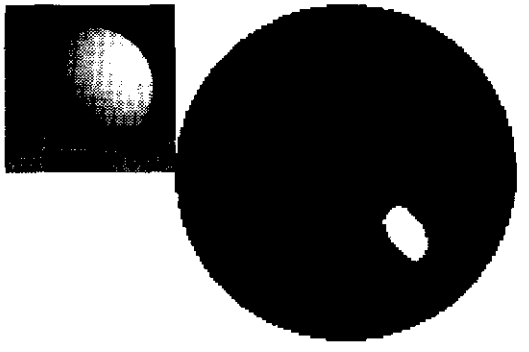


Figure 9. (Plate 6) Blue ambient light with a white circular source to the right and behind.

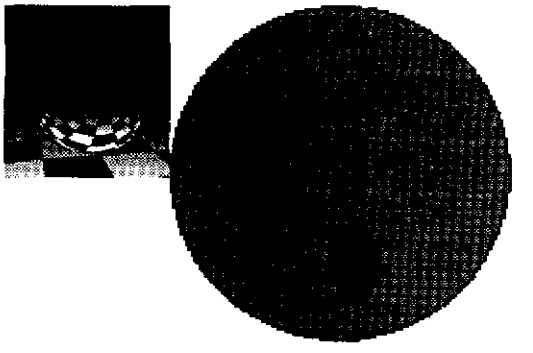


Figure 10. (Plate 7) Grey ambient light with red light reflected off another object.

nates (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* [LIE] for the surface point (u, v). Note that 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 $L^+(u, v, \theta, \varphi, s, \lambda)\sin\theta d\theta d\varphi d\lambda$ [16]. 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.

$$E = \sum_{s} \int_{\lambda} \int_{-\pi}^{\pi} \int_{0}^{\frac{\pi}{2}} L^+(u, v, \theta, \varphi, s, \lambda)\cos\theta\sin\theta d\theta d\varphi d\lambda \qquad (1)$$

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 $L^-(u, v, \theta, \varphi, s, \lambda)$.

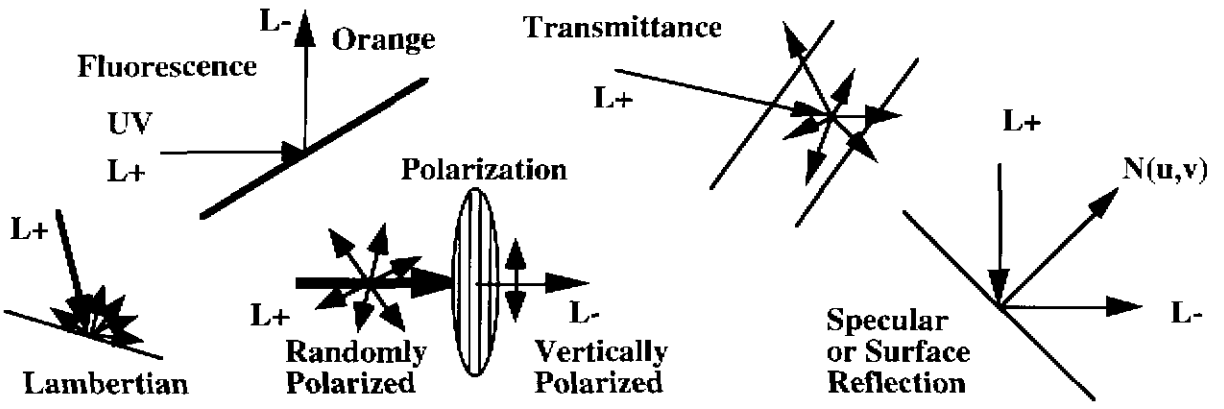


Figure 11. Some Special Cases of the Light Transfer Function: Fluorescence, Polarization, Transmittance, and Specular or Surface Reflection

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

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.

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

A structured analysis of the transfer function shows how it subsumes several common special cases, sketched in Figure 11. 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 non-fluorescing 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

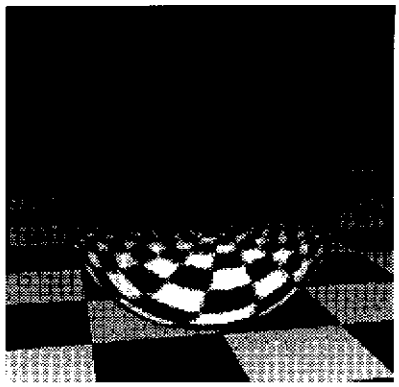


Figure 12. (Plate 8) Illustration of the transfer function for a slightly rough metal object.

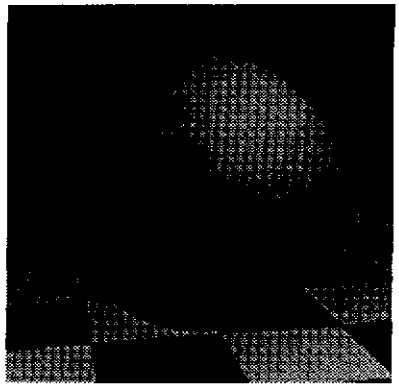


Figure 13. (Plate 9) Illustration of the transfer function for a slightly rough plastic object.

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 in Section 3.2.4., 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][38].

- Finally, Lambertian surfaces--also called perfectly diffusing perfect reflectors--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 in the environment shown in Figure 12 and Figure 13. The sphere sits 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 the specularity of the object. Figure 12 is an illustration of a highly specular material with no body reflection, and Figure 13 shows a matte colored material with a small amount of surface reflection.

2.2. General Hypotheses of Physical Appearance

We have defined a 3-D 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 3-D world and how these aggregations map to an image.

We have defined surfaces with an extent an 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.

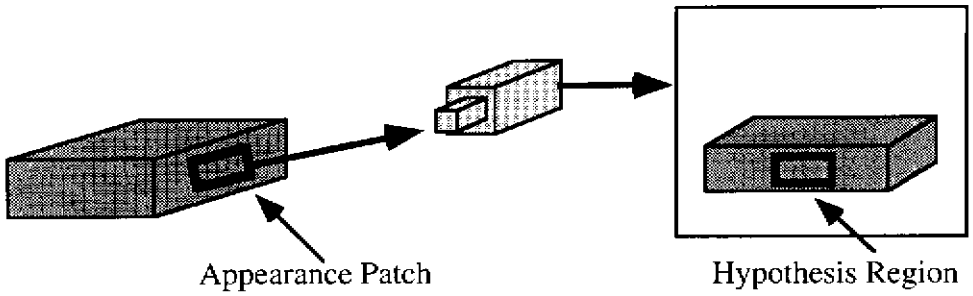


Figure 14. Mapping from an *appearance patch* to a *hypothesis region.*

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, and 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. As sketched in Figure 14, 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 patch we define to be a *hypothesis* $H = \langle S, E, \Re, L^+ \rangle$. The four elements of a hypothesis are the surface embedding S, the surface extent E, the transfer function $\Re$, and the incident light energy field L^+. With these functions, it is possible to completely determine the exitant light energy field (assuming no self-luminance). The basic connection between a physical explanation and a group of image pixels is provided by a *hypothesis region* $HR = \langle P, H \rangle$, defined as a set of pixels P that are the image of the hypothesis H. The combination of the hypothesis elements represents an explanation for the color and brightness of every pixel in the image patch. 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 set* $HS = \langle P, H_1, \ldots, H_n \rangle$ to be a set of pixels P with an associated list of hypotheses $H_1, \ldots, H_n$, where each hypothesis H_i provides a unique explanation for all of the pixels in P, and only the pixels in P.

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

To illustrate a hypothesis, we combine the representations developed previously into a 3-panel image displaying the characteristics of S, L, and $\Re$. Returning to the image of the cup in Figure 4, we can examine a single uniformly colored region (shown at the far left of Figure 15) and visualize two hypotheses for it: a mirror reflecting some illumination environment, or a plastic object under white illumination. We can illustrate the metal hypothesis in Figure 15, and the plastic hypothesis in Figure 16. Both hypotheses describe the same image region, and the combination of the two forms a hypothesis set.

3. Fundamental Hypothesis Regions

The difficulty inherent in segmentation using physical descriptions lies in determining the correct mapping between

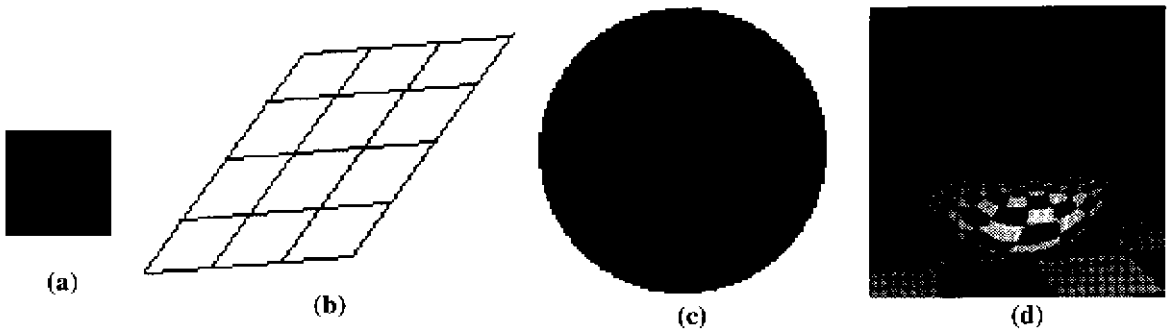


Figure 15. (Plate 10) Illustration of a metal hypothesis: (a) actual region (from Figure 4), (b) wire frame surface representation (planar), (c) illumination environment (diffuse), (d) transfer function (metal).

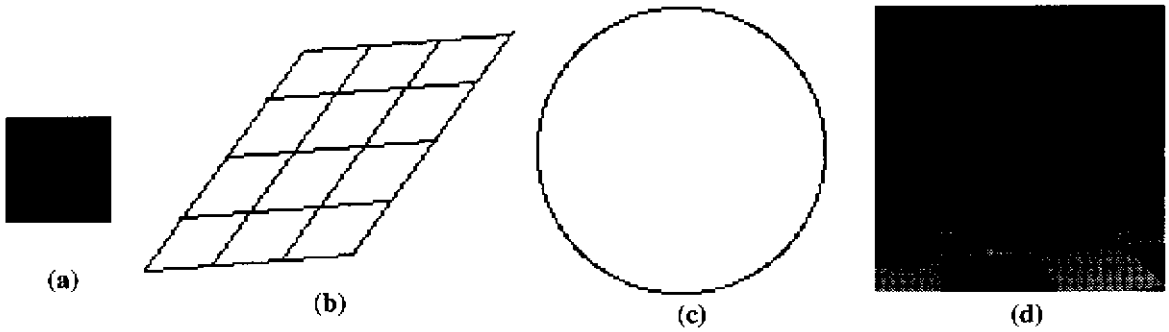


Figure 16. (Plate 11) Illustration of a dielectric hypothesis: (a) actual region (from Figure 4), (b) wire frame surface representation (planar), (c) illumination environment (diffuse), (d) transfer function (dielectric).

the image pixels and the scene that created them. The problem is that for a single pixel in an image, there are an infinite number of physical explanations for its color, and, in isolation, it is not possible to distinguish between those explanations. A single red pixel can be a red object under white light, a white object under red light, a mirror reflecting red light, a mirror reflecting a red object, or numerous other possibilities, and it is impossible to discriminate between them given only the one pixel value. Fortunately, we are not analyzing pixels in isolation, but images, which represent collections of appearance patches from the real world. These appearance patches possess coherence in their transfer function and their illumination environment. The segmentation process is thus the act of identifying which sets of pixels correspond to which appearance patches, identifying the possible physical explanations for those patches, and then merging them with other appearance patches when their possible physical explanations are compatible in some identifiable fashion.

Such a concept for segmentation is not new--for example, Klinker *et al.* [20] and Healey [14] both identified regions of similarity of some physical properties. What is new in this presentation is the generality. These past works assumed that the scene obeyed certain properties and looked only for a single, narrowly defined kind of coherence. In our new approach, the general illumination and transfer functions allow us to represent, reason about, and discover many different kinds of coherence in a single image. This capability is necessary for the analysis of natural or common man-made scenes such as Figure 1.

3.1. Pixel Classification

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

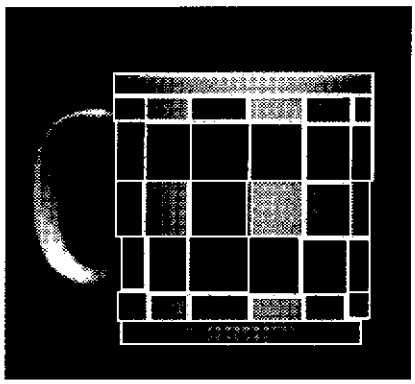


Figure 17. (Plate 12) Mug divided into idealized uniform chromaticity regions.

3.1.1. Uniform Chromaticity Regions

An approach of slightly greater complexity is to group together pixels displaying the same basic color ratios, or chromaticity, but with varying brightness. Mathematically, *chromaticity* is defined by "normalized color" coordinates, as defined in (3) [19]. Chromaticity can also be thought of as the hue and saturation of a color without the intensity information.

$$(r, g, b) = \left(\frac{r}{r+g+b}, \frac{g}{r+g+b}, \frac{b}{r+g+b}\right) \tag{3}$$

We define a *uniform chromaticity region* [UCR] to be a connected set of pixels that possess uniform chromaticity and possibly varying brightness. A UCR corresponds to a linear cluster, as defined by Klinker *et al.*[21]. As such, a more general definition of a UCR is a connected set of pixels whose covariance matrix in color space has a single non-zero eigenvalue, whose eigenvector is related to the chromaticity of the region. Because it allows for varying brightness within a region, a UCR is able to capture more of the relevant coherence between neighboring pixels than simple uniform regions.

Klinker *et al.* [20] note that a UCR, or linear cluster, can represent two distinct objects if both are dark or poorly illuminated. In this segmentation method, however, we initially assume that a UCR represents a single surface patch under a single illumination environment. This requires a form of coherence from the physical elements generating the UCR. Clearly, it is possible to construct an image with UCRs that do not have such coherence in the physical world, and we realize that our current approach will not correctly handle such situations.

The benefit derived by using UCRs is that they are groupings of pixels that we can reasonably assume to correspond to a single appearance patch in the physical world, setting constraints on the associated hypotheses. These constraints are that over the patch the transfer functions are coherent and the illumination environments are similar. Because it is a single appearance patch, it is, by definition, a single surface. Figure 17 shows an idealization of the cup image divided into UCRs.

By identifying UCRs in the image, we have taken the first step in the segmentation process by linking pixels with appearance patches in the scene. The next step is to begin to identify the relevant physical explanations, or hypotheses, for the appearance patches corresponding to the identified UCRs.

3.2. Generating Hypotheses

If a UCR does correspond to a single appearance patch in the scene, what are the possible hypotheses for that appearance patch given the constraints identified previously? Clearly, the relationship between appearance patches and hypotheses is not one-to-one. As demonstrated by Figure 2, it is possible for identical image regions to have differing

hypotheses specifying their physical description. Therefore, given a UCR and its related appearance patch, we must consider multiple physical descriptions.

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

The possible sources of color for an appearance patch are the illumination, the transfer function, or both. Intuitively, the simplest hypotheses attribute the color to a single element of the hypothesis. As an example, consider a UCR of uniform pixel values. A simple hypothesis is one that specifies the surface as red plastic under diffuse white illumination. Such a hypothesis is intuitively plausible, and simple to express. A more complex hypothesis, attributing the color to two elements, is one where both the illumination and the color of the object vary over the surface, but in such a way that their combination produces the same color and intensity at each point. This hypothesis is much more difficult to express, and is not automatically accepted by our intuition as a plausible explanation.

The varying intensity of a UCR could be due to uneven illumination, uneven coloring, or curvature of the surface. Any or all of these possibilities could occur on a single patch. Again, intuitively some of these explanations are simpler than others. Attributing all brightness variation to the shape, for example, is the underlying assumption for many shape from shading algorithms.

From these observations, we can begin to answer the question "how many hypotheses must we consider" by looking at the simplest ones first. Using simplicity, or plausibility to select between alternative hypotheses has been suggested by Tympanum *et al.*, and has been used as the basis for several vision systems [23][11][22][27][26]. This requires us to distinguish between simple and complex hypotheses. Furthermore, we need to look for simplicity not only within the hypotheses, but in the representations of the elements themselves. But what constitute simple forms of the hypothesis elements, and must every possibility by entertained? Furthermore, does simplicity always imply a hypothesis is more likely, or more *plausible*? To answer these questions, we must delve into the meaning of what constitute classes of the hypothesis elements S, L, and $\Re$, and what we mean by the terms "plausibility," "complexity," or "weirdness" with respect to a hypothesis an its elements.

3.2.1. Plausibility

In an ideal world, we would be able to quantify complexity, or "weirdness" and use it 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, transfer function, and illumination environment. Less weird explanations would be those closer to the origin of the three axes. The further from the origin, the weirder the hypothesis elements would become. By generating hypotheses close to the origin, or with only one weird element, we could begin with a small set of simple hypotheses and generate weirder ones only if necessary. Weirdness is a difficult concept to measure, however, and the axes of our weirdness scale are almost certainly non-linear and not independent.

The minimum description length [MDL] principle, however, is a mathematical formalism for "weirdness." The MDL principle says that, given a parameterization for describing a model, the best model that describes a set of data is the one that can be encoded in fewest number of binary digits, or shortest length. In computer vision, the MDL principle has been used successfully by Leclerc [23], Darrell *et al.* [11], Krumm [22] and Leonardis [26]. If we postulate a language for S, E, L, and $\Re$, then a hypothesis region and its fields and subfields are a model described in that language. Our task in segmentation is to find a set of such models that describe an entire image, or data set. Based upon the MDL principle, we propose that the most desirable sets of hypotheses that describe a particular scene are the least complex ones, or the ones that can be described most succinctly.

It is important to note that the description length has two components: the complexity of the description, and how well that description fits the data. The combination of the two components is used to select the best model. When we are dealing with a set of hypotheses for an image region, they ought to fit the data about equally well, so that term of

the description length should be approximately constant. Therefore, rank-ordering the hypotheses for a region using some measure of complexity, should be sufficient to satisfy the MDL criteria.

Because there are an infinite number of hypotheses for any UCR, care must be taken in the initial selection of the hypothesis set for each UCR. One important consideration of the MDL principle is that the optimal model, or model set must be among those tested for shortest length. Following our methodology, we want to identify subspaces of our general parameterization which will are both simple and likely to occur in general images. There are at least three approaches that could be taken to generate this model set:

- Generate a large number of possible hypotheses and test
- Generate incrementally according to some search criterion
- 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 possibilities

As indicated by previous discussion, the first approach seems pointless and intractable. Breton *et al.* were able to use the approach by creating a discrete mesh of possible light source directions for a "virtual" point source, but since our model has many more parameters in both the illumination environment and the transfer function, such coverage by a discrete mesh is intractable. The second approach has merit, but it is unclear what type of search criterion is needed for this task. Instead, we propose the use of broad classes to initially assign hypotheses to a UCR, with the understanding that the particular details of a hypothesis--i.e., the actual shape, the specific colors, surface roughness, and other characteristics--will be determined at a later point in the segmentation process. It is also important to note that this set can be incrementally expanded if all of its initial constituents are considered unlikely. The broad classes, which we derive from the general model for scene description, are simple, yet comprehensive enough to cover a wide range of possible environments and objects.

3.2.2. Taxonomy of Surfaces

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

To simplify this reasoning process, we initially consider only two classes of surfaces: curved and planar. These two classes provide a simple distinction that can be used to reason about merging hypotheses. A finer distinction would require a specific method for modeling curved surfaces, which we leave for future exploration. When a surface representation method is determined for the actual segmentation system, reasoning about merging two curved surfaces could be done based on that representation--e.g. matching two spheres, superquadrics, generalized cylinders, or polynomial surfaces.

3.2.3. Taxonomy of Illumination

There are several simplified special forms of the incident light energy field function that represent useful models of illumination. Recall that the general form of the global incident light function is given by $L^+(x, y, z, \theta_x, \theta_y, \lambda, s, t)$. Figure 18 shows the relationships of the subspaces we identify for this function. The largest subspace we consider is that of time-invariant illumination, where we consider time to be a constant and drop it from our parameterization. The second subspace we highlight is 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 will fall into this category. 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). Within the unpolarized, time-invariant subspace are those illumination functions in which the color of the light is independent of the direction of incidence. The hue and saturation of such illu-

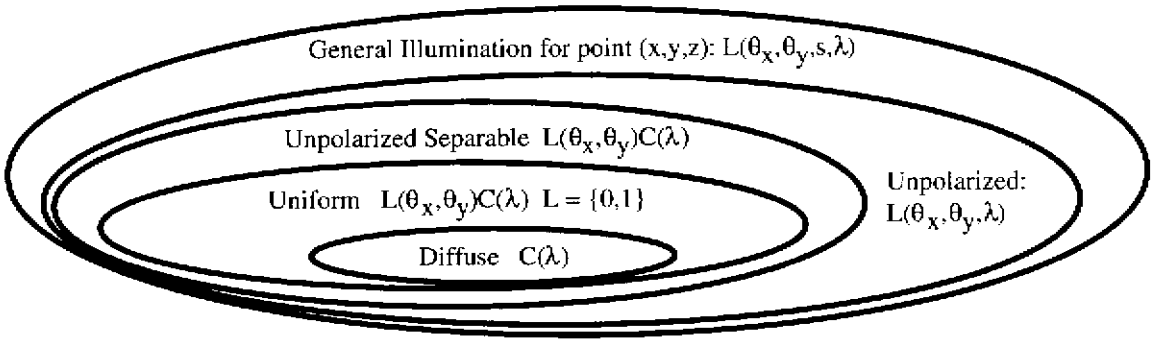


Figure 18. Subspaces of the global incident light energy field $L(x,y,z,\theta_x,\theta_y,\lambda,s)$.

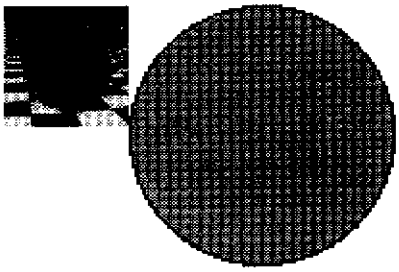


Figure 19. (Plate 13) Diffuse illumination environment.

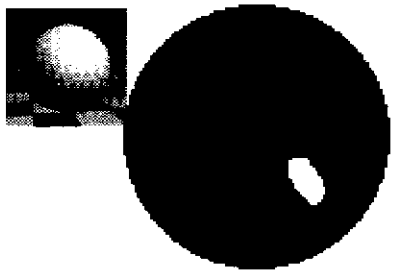


Figure 20. (Plate 14) Uniform illumination environment.

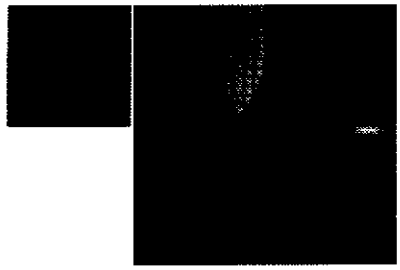


Figure 21. (Plate 15) General illumination environment.

mination functions are the same in all directions and only the brightness varies over the illumination hemisphere. These illumination functions are separable into the form $L^+(x, y, z, \theta_x, \theta_y)C(x, y, z, \lambda)$, where $L^+(x, y, z, \theta_x, \theta_y)$ denotes the incoming intensity in a given direction at (x,y,z), and $C(x, y, z, \lambda)$ the color of the illumination. Within the subspace of separable functions is the *uniform* illumination subspace which can be written for the point (x,y,z) as $L^+(\theta_x, \theta_y)C(\lambda)$, where $L^+(\theta_x, \theta_y) = \{1, 0\}$. Uniform illumination thus implies that all illumination in the environment has the same color. Some important special cases of uniform lighting include:

- Point light source at $(\theta_{x0}, \theta_{y0})$ $\quad L^+(\theta_x, \theta_y) = \begin{cases} 1 & (\theta_x = \theta_{x0}) \text{ and } (\theta_y = \theta_{y0}) \\ 0 & \text{otherwise} \end{cases}$

- Finite disk source of apex angle α centered at $(\theta_{x0}, \theta_{y0})$ $\quad L^+(\theta_x, \theta_y) = \begin{cases} 1 & \text{angle between } (\theta_x, \theta_y) \text{ and } (\theta_{x0}, \theta_{y0}) < \alpha \\ 0 & \text{otherwise} \end{cases}$

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

These three simple cases play an important role in modeling illumination. Indeed, as shown by the computer graphics community, a large number of illumination environments can be modeled using one or more point, finite disk, or ambient light sources [12]. For the purpose of reasoning about hypotheses, we use three subspaces--in order of increasing complexity--diffuse, uniform, and general illumination to describe the forms of the illumination environment. A diffuse illumination environment, uniform color and brightness over the hemisphere, is shown in Figure 19, along with its effect on a white sphere. Figure 20 illustrates a uniform illumination environment, as specified in Figure 18, and its effect also on a white sphere. Finally, a general illumination environment is illustrated in Figure 21, along with its effect on a metal sphere.

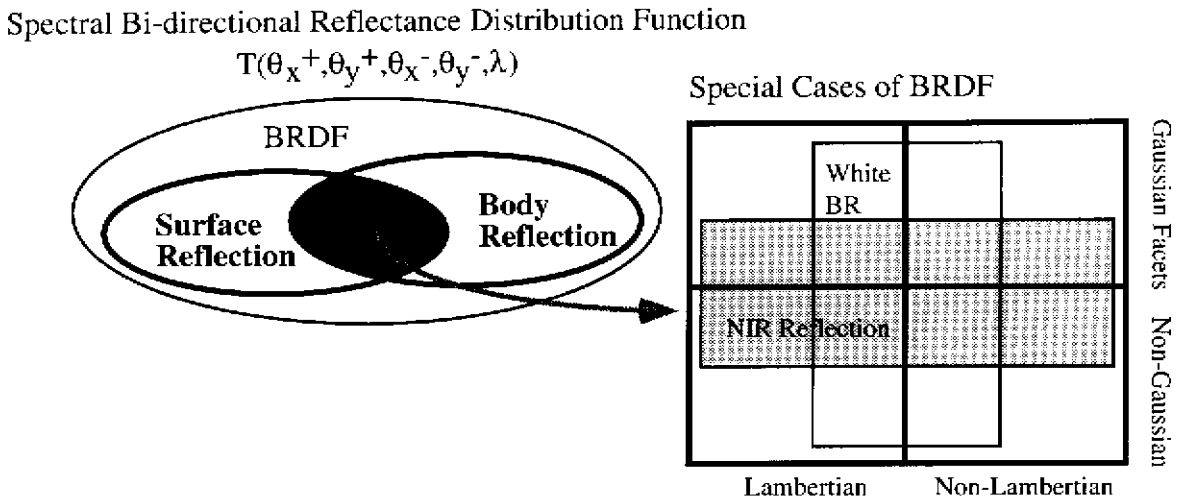


Figure 22. Taxonomy of the bi-directional reflectance distribution function

3.2.4. Taxonomy of the Transfer Function

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

This reduced transfer function still includes surfaces with arbitrary changes in the transfer function over (u, v). Such surface patches can have differing color and texture within their extent. Therefore, we further identify two nested subsets: transfer functions that are piecewise-uniform, and those that are completely uniform over the extent of the (u, v) parameters. The subset of uniform transfer functions, shown in Figure 22, can be specified by the reduced function $\Re(\theta^+, \varphi^+, \theta^-, \varphi^-, \lambda)$, as it is constant over all relevant values of u and v. This form of the transfer function is recognizable as the well-known *spectral bi-directional reflectance distribution* function [spectral BRDF] for a uniform surface [30].

Within this set are further interesting subspaces of the transfer function. Transfer functions with surface reflection or body reflection are two important overlapping subspaces. Their relationship within the BRDF and the interaction of the union of these subspaces is shown in Figure 22. Surface reflection, as noted previously takes place at the interface between an object and the surrounding air. 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 [17][18]. Materials whose surface reflection is approximately constant over the visible spectrum form a useful subset and are said to have *neutral interface reflection* (NIR) [25]. The surface reflection from an NIR material is assumed to be 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 [14].

Many materials displaying surface reflection are optically "rough." They possess microscopic surfaces with local surface normals that differ from the macroscopic shape, as shown in Figure 23. 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 sur-

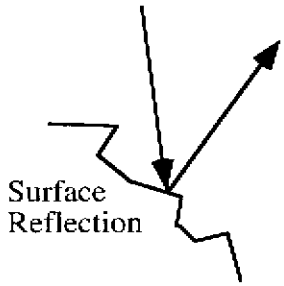


Figure 23. Microfacet surface reflection model.

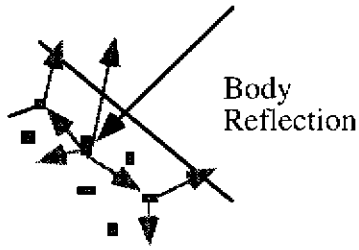


Figure 24. Body reflection model: transparent medium with pigment particles.

faces using a Gaussian distribution assumption for some surface characteristic [3][10][13][25][29][34][38]. These models fit into our taxonomy of transfer functions as shown in Figure 22.

A more complex form of reflection, body reflection, takes place when light penetrates a surface and interacts with colorant particles as shown in Figure 24. During this interaction, some of the wavelengths may be absorbed, coloring the reflection. The remaining wavelengths are re-emitted in random directions, striking other colorant particles, and some 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. The subset of transfer functions that display this independence in their body reflection are called Lambertian surfaces. These subset relationships are shown in Figure 22. The body reflection of the Lambertian subset is said to obey Lambert's Law, which states that the reflection is dependent upon the incoming light's intensity and cosine of the angle of incidence [16]. Improved models of body reflection are being researched [15][31][39].

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 [30][37]. Transfer functions within this overlapping region have been approximated by the *dichromatic reflection model* [33].

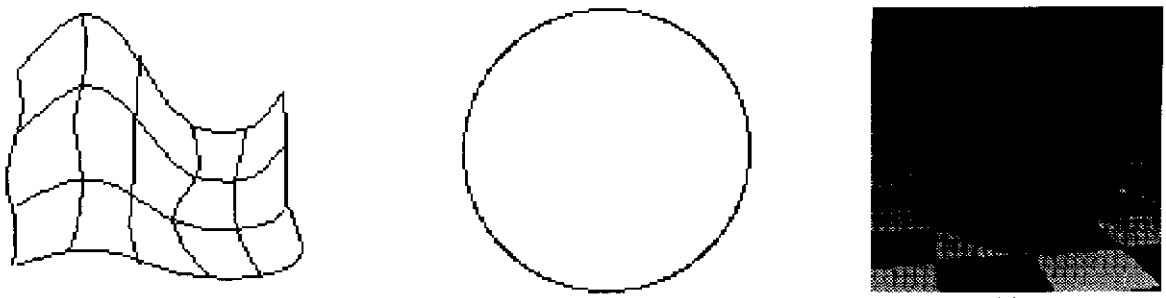
Metals also fall into the spectral BRDF category, although they only display surface reflection and have been modeled by the *unichromatic reflection model* [14]. Most models for rough specular surfaces apply directly to metals.

For the purposes of our proposed segmentation method, we initially consider objects whose transfer functions fall within the union of body reflection and surface reflection. Objects with these properties naturally divide into two categories: *metals* and *dielectrics*. 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. Illustrations of these two classes of the transfer function can be seen in Figure 12 and Figure 13.

3.3. Hypothesis Classification

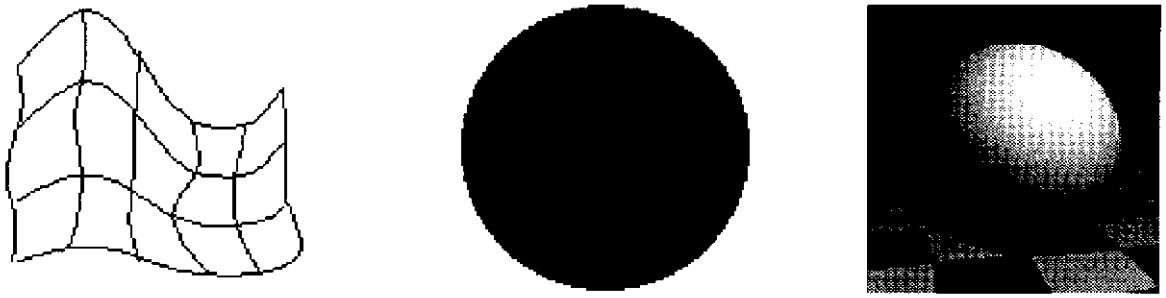
Based on the above taxonomies of S, L^+, and $\Re$, we now identify a simple, yet comprehensive set of hypotheses for explaining the color and brightness variation of a UCR. To accomplish this task, we first form a set of hypothesis classes based upon the forms previously developed for the individual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

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



(a) (b) (c)

Figure 25. (Plate 16) Fundamental hypothesis with body reflection as color source: (a) surface, (b) illumination environment, (c) transfer function.



(a) (b) (c)

Figure 26. (Plate 17) Fundamental hypothesis with illumination as color source: (a) surface, (b) illumination environment, (c) transfer function.

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

3.3.1. Fundamental Hypotheses

To account for the distribution of color between the elements of a given form, we identify a set of hypotheses that is simple, and yet provides a significant amount of explanatory power. This set, which we call the set of *fundamental hypotheses*, consists of those hypotheses in which the color of the region is due to only one of the possible color producing elements: the body reflection, the surface reflection, or the illumination environment. Figure 25 and Figure 26 illustrate two of the fundamental hypotheses for a region of the cup. In Figure 25, we see that the curved plastic is colored and the illumination is white. Figure 26 shows the illumination as the color source with white plastic. Both are equally possible explanations for the UCR.

Combining the broad classes for hypothesis elements with the requirement that the color of a pixel is due to either the transfer function or the illumination environment, but not both, creates a finite set of hypotheses that must be considered for an UCR. Given two material types, two shape classes, three illumination environments, and three possible color sources, we arrive at 36 possible hypotheses. As the body reflection cannot be the color source if there is no body reflection (metals), there are at most 30 fundamental hypotheses that explain the same UCR for non-polarized, opaque, non-fluorescent surfaces. Note this is true for *all* UCRs, no matter the shape, color, or brightness distribution.

Closer analysis of these 30 fundamental hypotheses shows that the six hypotheses corresponding to dielectrics whose color source is the surface reflection are highly unlikely, and probably do not conform to the single color source rule. These six hypotheses are unlikely because of the commonly used Neutral Interface Reflection assumption, which states that the spectrum of surface reflection of a dielectric is approximately uniform, or neutral in terms of its effect on the color of the reflection [25]. This assumption is based upon the observation that one of the manufacturing criteria for the medium of many common dielectrics is that it have a neutral, or uniform spectrum. This criterion ensures that the coloring will be imparted entirely by the pigment materials added during the manufacturing process. As most

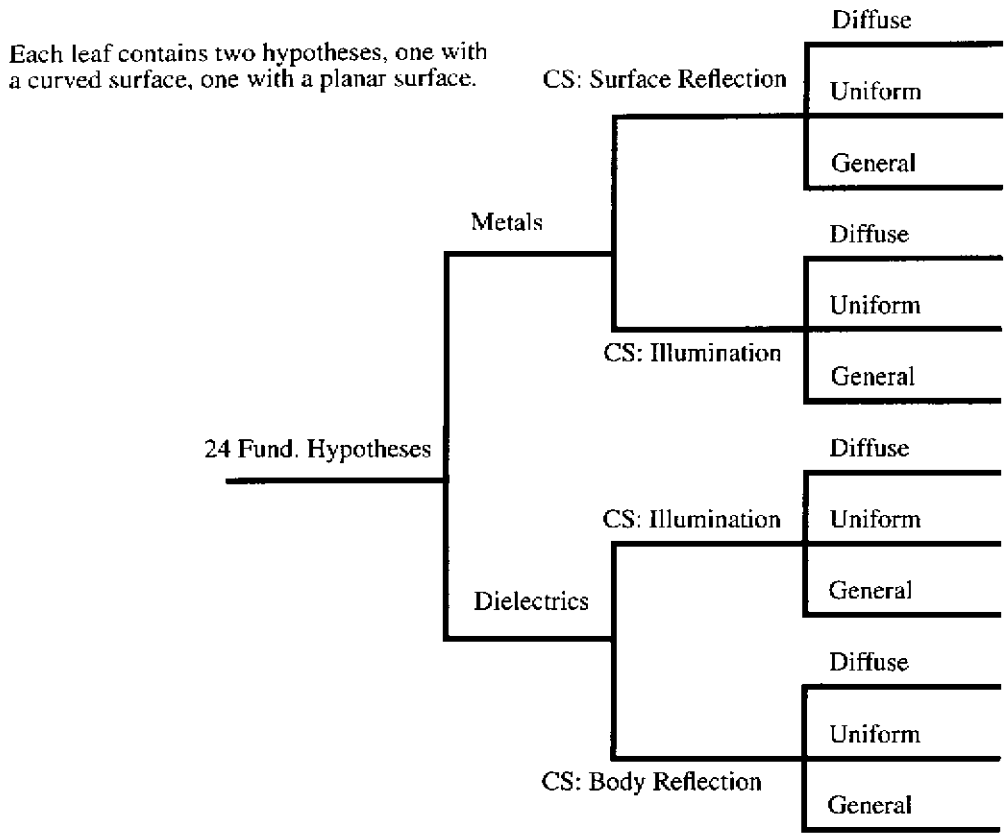


Figure 27. A Taxonomy of fundamental hypotheses.

of the dielectric surfaces we are concerned with are manufactured materials--paint, plastic, glass, ink, paper, cloth, ceramic--using the NIR assumption to prune the list of fundamental hypotheses does not significantly alter the explanatory power of this method. Furthermore, if the surface reflection does not admit all wavelengths evenly to the colorant particles that constitute the body reflection, then the body reflection will be colored as well, violating our single color source constraint. Pruning these six hypotheses results in 24 fundamental hypotheses of image formation.

3.3.2. Taxonomy of Fundamental Hypotheses

We can arrange these 24 hypotheses in a tree structure according to the material type, color source, illumination environment, and shape as shown in Figure 27. The first branching indicates the material, or general transfer function of the hypothesis and divides the 24 hypotheses into two subsets. The second branching indicates the color source of the hypothesis. As the body reflection cannot be a color source for a metal, and the surface reflection cannot be a color source for a dielectric, four subsets result from this branching. The third branching specifies the illumination environment of the hypothesis. With three possible illumination environments for each category of material and color source, this divides the 24 hypotheses into twelve subsets. Each of these subsets has two leaves, not shown in the tree, one representing a hypothesis with a curved surface, and one with a planar surface.

The resulting tree with its 24 leaves represents a taxonomy of fundamental, or the simplest hypotheses, classifying the different physical explanations for an image region. The true importance of this taxonomy is that it represents a finite set of simple, yet relatively comprehensive hypotheses for describing an appearance patch corresponding to a UCR. Therefore, we can postulate a hypothesis set, with a reasonably small number of hypotheses, for each UCR we identify in an image. This provides an initial segmentation and sets the stage for us to begin reasoning about and merging hypothesis regions.

4. Analysis and Merging of Hypotheses

In this section we further analyze the fundamental hypotheses and develop a set of tools for comparing and merging them. To illustrate these tools we simultaneously work through a simple example image of a lambertian sphere with a stripe in the middle, shown in Figure 28. The goal of this section is to develop the outline of a segmentation algorithm using reasoning about the physics underlying fundamental hypothesis regions.

4.1. Analysis of the Fundamental Hypotheses

The taxonomy shown in Figure 27 might be taken to suggest that all of the fundamental hypotheses are of equal value in explaining a scene. We do not believe this is the case for most images. As a deeper analysis shows, several of the fundamental hypotheses have little relative value in explaining image regions. This same analysis also shows that some mechanism will have to be proposed for the orderly development of more complex hypotheses to explain some common physical phenomena.

We begin with a structured analysis of each subtree of the taxonomy, considering in turn each of the four possible combinations of material and color source and the six associated hypotheses. The goal of this examination is to divide the 24 hypotheses into two groups, or tiers, corresponding to common and rare physical situations. Common hypotheses we specify as belonging to tier one, and rare hypotheses we place in tier two.

We begin with the tree corresponding to colored dielectrics under white illumination. These six hypotheses are grouped into three pairs according to the illumination environment. Clearly, curved and planar dielectrics under uniform lighting form a large subset of objects in a typical scene. Scenes that can be modeled by these two hypotheses include paper, plastic, and painted objects under one or more light sources of approximately equivalent brightness. Sunlight can also be frequently approximated by a uniform source when considering dielectrics because its effect on dielectric surfaces usually overwhelms any illumination from other directions. Likewise, curved and planar dielectrics under diffuse lighting are often used as a model for surfaces in shadow, where no light source is directly incident on the surface [10].

Curved and planar dielectrics under general function white lighting 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 of the real world. In the case of dielectrics, however, uniform and diffuse lighting models are probably sufficient for most surfaces. The major 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. Practically, this means that the exitant light energy field due to a strong incident light source from a single direction can overshadow the additional exitant light due to the incident light from all other directions. In scenes where there are one or most light sources incident on an object, therefore, we propose that most lighting conditions can be modeled as a set of uniform brightness white sources, which falls under the uniform illumination category. This analysis is strengthened by the fact that the 'general function' illumination in this case must still be uniform spectrum--black, white, or grey--at each point on the hemisphere because the color source is the body reflection. This makes the uniform illumination category an even better approximation to the general function category because only the geometry is approximated, rather than the spectral characteristics of the illumination. Because of this analysis, we propose that the branch of the taxonomy with colored dielectrics under white illumination has four common hypotheses which belong in tier one--those with diffuse and uniform illumination environments--and two rare hypotheses which belong in tier two--those with the general function illumination environment.

The next branch corresponds to white dielectrics under colored illumination. In common scenes we suggest that situations corresponding to these hypotheses are rare. 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

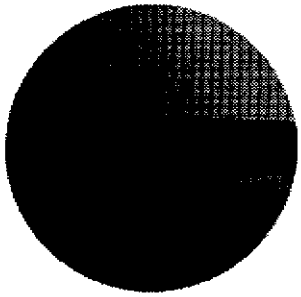


Figure 28. (Plate 18) Three-color lambertian sphere.

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 the curved and planar hypotheses with general function illumination be placed in tier one, and the other four hypotheses in tier two.

White metals under colored illumination form the next branch of the taxonomy. Unlike dielectrics, incident light from almost all directions is significant in the appearance of a metal appearance patch. This can be seen in Figure 1, where inter-reflected light that is dim relative to the global light source still has an effect on the appearance of the metal objects. For this reason, the hypotheses with general function 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 general function illumination belong to the first tier, and the other four hypotheses--diffuse and uniform illumination--belong to the second tier.

The final branch of hypotheses is the colored metals under white illumination. As with grey metals, the hypotheses with general function illumination are the most common models for colored metal objects. Unfortunately, because of the single color source constraint, this general function illumination cannot also be colored, severely restricting the set of objects these hypotheses can model. In fact, we propose that uniform illumination is sufficient to model any surfaces that would correspond to colored metal under white illumination. Diffuse illumination, as with grey metals, we believe is rare. From this analysis, the two hypotheses with uniform illumination belong in tier one; the other four belong in tier two.

It is the analysis of colored metals that most clearly demonstrates the need for a method to incorporate more complex hypotheses into the reasoning process. Our definition of "fundamental hypotheses" stipulating a single color source for a UCR will not be adequate to explain many images of colored metals, because their appearance will also depend on colored interreflection from nearby objects. The other area where more complex hypotheses are needed is for interreflection between colored objects, especially dielectrics. In the example we work through herein, these problems do not arise. However, we will ultimately need a mechanism for infusing more complex hypotheses--for example, red metals under colored illumination--in order to achieve the generality we desire in this segmentation method.

The overall result of this analysis is that there are ten common fundamental hypotheses in tier one, and fourteen less common or rare fundamental hypotheses in tier two. Figure 30 through Figure 39 illustrate the ten fundamental hypotheses in tier one.

In our example segmentation, we consider only the fundamental hypotheses in tier one. As shown in Figure 29, the two-color sphere divides into three UCRs: top, middle, and bottom. To each region we can attach the list of ten hypotheses from tier one, forming three hypothesis sets of ten hypotheses each.

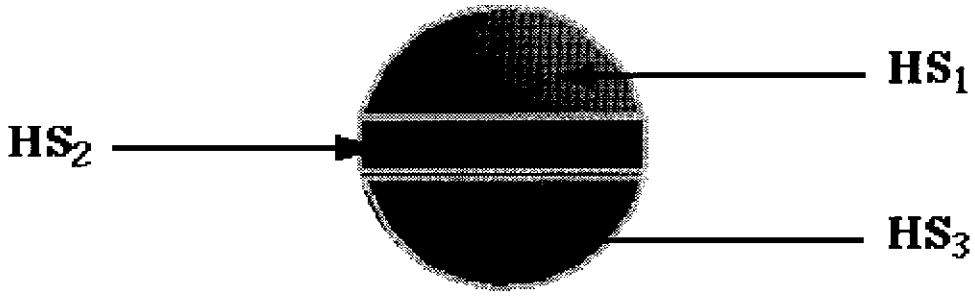


Figure 29. (Plate 19) Three hypothesis sets in the example image.

4.2. Merging Hypothesis Regions

Each of the UCRs in our example image has a hypothesis sets with 10 fundamental hypotheses explaining its physics of formation. We seek to agglomerate small regions into big ones in order to search for coherence between regions. Our basic method is to take two adjacent hypothesis sets $HS_1 = \langle P_1, H_{11}, H_{12}, \ldots \rangle$, and $HS_2 = \langle P_2, H_{21}, H_{22}, \ldots \rangle$, and form a new hypothesis set $HS_3 = \langle P_2 \cup P_2, H_{31}, \ldots \rangle$, in which the hypotheses H_{3i} are created by merging compatible hypotheses H_{1j} and H_{2k}.

A bulldozer approach would consider all possible combinations of the fundamental hypotheses, resulting in 10^3 aggregate hypotheses. But are there really 1000 plausible explanations for this combination of three regions? Such a merging method is not only unreasonable, but also too expensive to use even on simple images because of the exponential explosion of the number of hypotheses. The interaction between hypothesis regions and the nature of the physical explanations must provide a guide or constraint to limit this explosion.

Fortunately, the goal of the segmentation process provides a partial solution. The mergers in which we are interested during segmentation involve coherence in the general variables: material type, shape, color source, and illumination environment. When two hypotheses match in several or all of these four variables, but differ in color or other subfeature, it makes sense to combine them into a single region. It does not make sense to combine two hypotheses that propose different materials at this stage of the image analysis. Nor does it make sense to combine a hypothesis proposing the surface reflection as the color source with a neighboring hypothesis that proposes the body reflection as the color source. While such a merger may make sense on a more abstract scale--consider a watch with a painted face and metal watchband--it does not make sense in a low-level segmentation. Likewise, at this level of segmentation we propose that two hypotheses of differing shape should not be merged.

On the other hand, it is possible that two hypotheses with differing illumination environments should be combined. A common example of this is an object partly in shadow. One hypothesis for the surface not in shadow could have uniform illumination, while one hypothesis for the region in shadow may be diffuse illumination. Combining these two hypotheses is desirable if the surface shapes and material match. The resulting hypothesis would have a general function illumination environment, albeit with recognizable structure.

The constraints requiring that mergeable hypothesis pairs must have the same material type and color source sharply curtail the number of resulting explanations. The chart in Figure 40 shows all possible combinations of the fundamental hypotheses of two regions for the ten hypotheses in tier one. As it shows, twelve hypotheses result from merging the two hypothesis sets containing the ten hypotheses from tier one. The explicit rules we use to obtain these twelve hypotheses are:

- Hypotheses of differing materials should not be merged.
- Hypotheses of differing color sources should not be merged.
- Hypotheses of differing shape should not be merged.

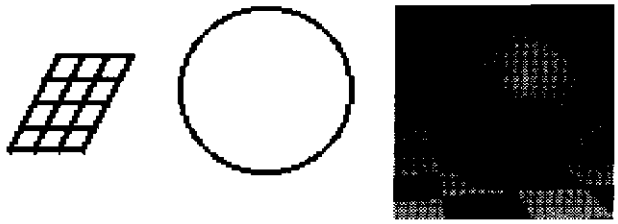


Figure 30. (Plate 20) Hypothesis 1: planar--diffuse ill.--colored dielectric.

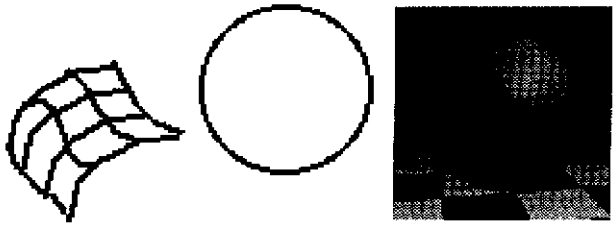


Figure 31. (Plate 21) Hypothesis 2: curved--diffuse ill.--colored dielectric.

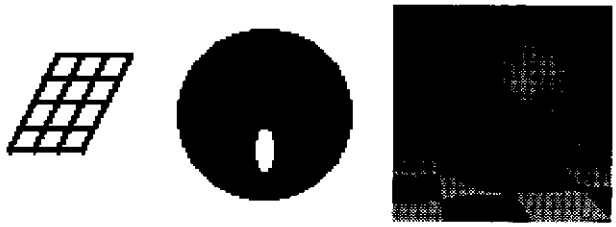


Figure 32. (Plate 22) Hypothesis 3: planar--uniform ill.--colored dielectric.

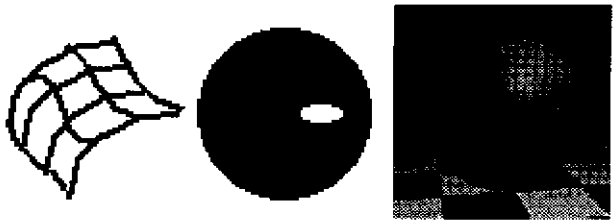


Figure 33. (Plate 23) Hypothesis 4: curved--uniform ill.--colored dielectric.

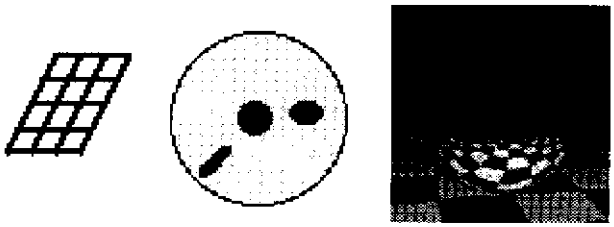


Figure 34. (Plate 24) Hypothesis 5: planar--general ill.--grey metal.

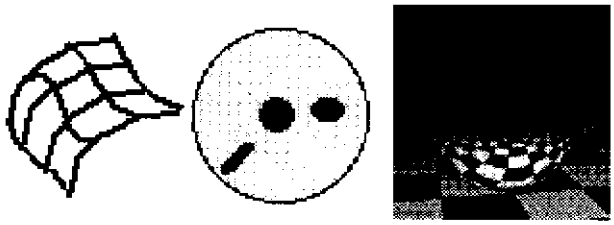


Figure 35. (Plate 25) Hypothesis 6: curved--general ill.--grey metal.

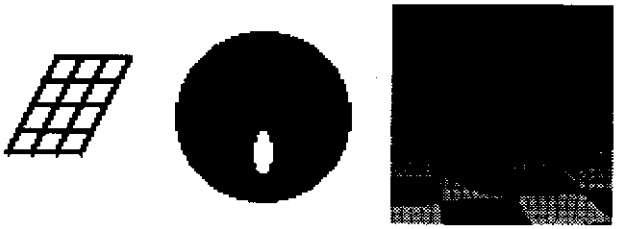


Figure 36. (Plate 26) Hypothesis 7: planar--uniform ill.--colored metal.

Figure 37. (Plate 27) Hypothesis 8: curved--uniform ill.--colored metal.

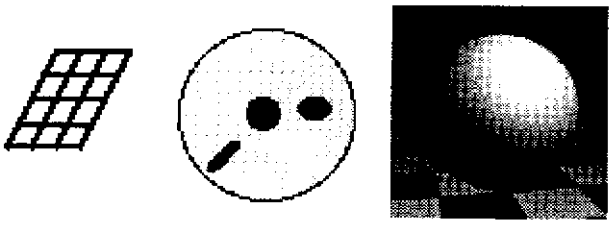


Figure 38. (Plate 28) Hypothesis 9: planar--general ill.--grey or white dielectric

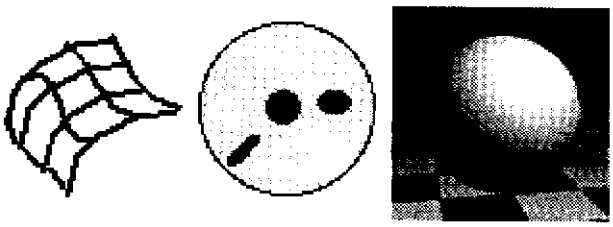


Figure 39. (Plate 29) Hypothesis 10: curved--general ill.--grey or white dielectric.

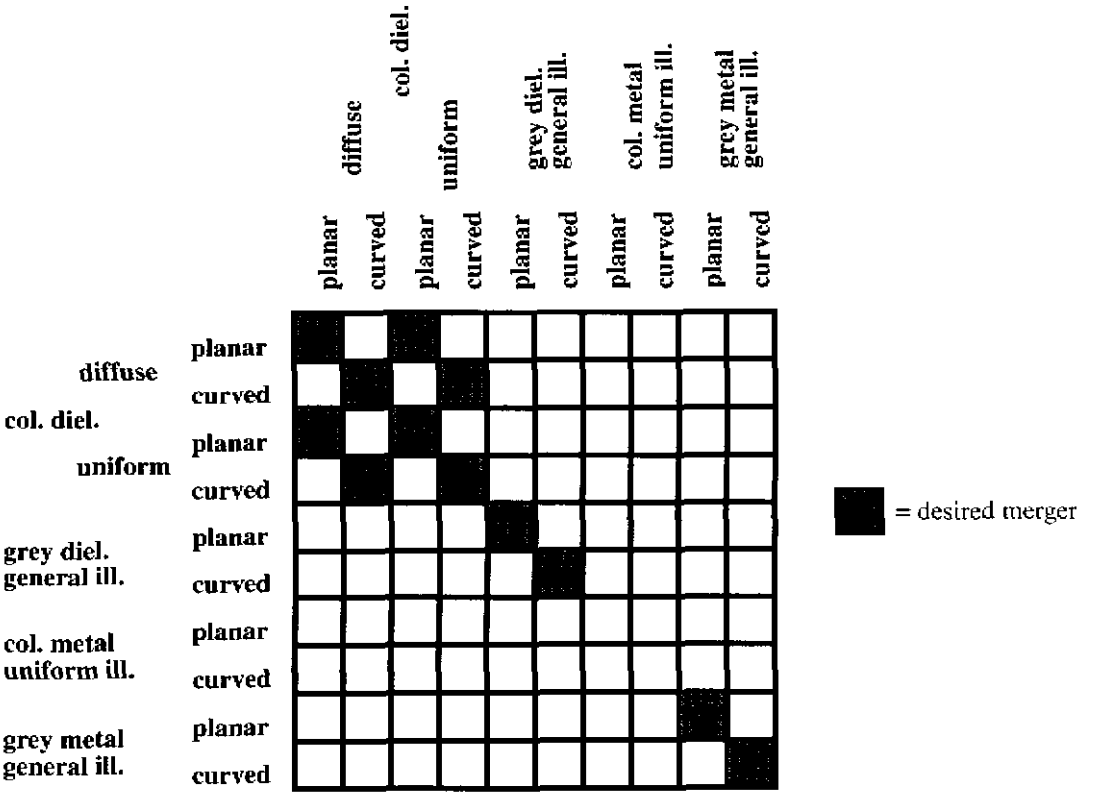


Figure 40. Possible mergers of the ten 'best' fundamental hypotheses for two regions of differing color. The grey squares indicate the desirable mergers.

- "Colored metal" hypotheses of differing chromaticity and similar illumination should not be merged.
- If the hypotheses differ in their chromaticity and the illumination is the color source, then hypotheses with diffuse illumination environments should not be merged.

The reasoning behind the first three rules should be clear; we do not want to propose abstract relationships between image regions at this low-level stage of segmentation. The fourth rule results from the fact that the surface reflection, or material properties of the surface, determine the color of "red metal" hypotheses. Therefore, if two of these hypothesis regions differ in color but have the same illumination environment, they must be different materials. As they are different materials, they 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 are constant over the illumination hemisphere. Now consider two adjacent appearance patches with the illumination as the color source. If the illumination is diffuse, and the adjacent patches are at less than a 180° angle, there will be overlap between the illumination environments of the two patches. If the two patches are differing colors, 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.

Returning to our example, merging the top and middle regions, and the middle and bottom regions, we obtain twelve possible hypotheses for the merger of each pair. As the hypotheses for the middle region can be matched for each pair,

there are, in fact, twenty resulting hypotheses for the entire sphere. These twenty hypotheses are listed in Table 1.

Table 1. Final set of hypotheses for the example image

Hypothesis	Top Region	Middle Region	Bottom Region
1 Tier 1	Diel/CS=BR/Uni./Curved	Diel/CS=BR/Uni./Curved	Diel/CS=BR/Uni./Curved
2 Tier 2	Diel/CS=BR/Dif./Curved	Diel/CS=BR/Dif./Curved	Diel/CS=BR/Dif./Curved
3	Diel/CS=BR/Uni/Planar	Diel/CS=BR/Uni/Planar	Diel/CS=BR/Uni/Planar
4	Diel/CS=BR/Dif./Planar	Diel/CS=BR/Dif./Planar	Diel/CS=BR/Dif./Planar
5	Metal/CS=IL/gf/Curved	Metal/CS=IL/gf/Curved	Metal/CS=IL/gf/Curved
6	Metal/CS=IL/gf/Planar	Metal/CS=IL/gf/Planar	Metal/CS=IL/gf/Planar
7 Tier 3	Diel/CS=IL/gf/Curved	Diel/CS=IL/gf/Curved	Diel/CS=IL/gf/Curved
8	Diel/CS=IL/gf/Planar	Diel/CS=IL/gf/Planar	Diel/CS=IL/gf/Planar
9 Tier 4	Diel/CS=BR/Uni./Curved	Diel/CS=BR/Uni./Curved	Diel/CS=BR/Diff/Curved
10	Diel/CS=BR/Uni./Curved	Diel/CS=BR/Dif./Curved	Diel/CS=BR/Uni./Curved
11	Diel/CS=BR/Uni./Curved	Diel/CS=BR/Dif./Curved	Diel/CS=BR/Dif./Curved
12	Diel/CS=BR/Dif./Curved	Diel/CS=BR/Uni./Curved	Diel/CS=BR/Uni./Curved
13	Diel/CS=BR/Dif./Curved	Diel/CS=BR/Uni./Curved	Diel/CS=BR/Diff/Curved
14	Diel/CS=BR/Dif./Curved	Diel/CS=BR/Dif./Curved	Diel/CS=BR/Uni./Curved
15	Diel/CS=BR/Uni./Planar	Diel/CS=BR/Uni./Planar	Diel/CS=BR/Diff/Planar
16	Diel/CS=BR/Uni./Planar	Diel/CS=BR/Dif./Planar	Diel/CS=BR/Uni./Planar
17	Diel/CS=BR/Uni./Planar	Diel/CS=BR/Dif./Planar	Diel/CS=BR/Dif./Planar
18	Diel/CS=BR/Dif./Planar	Diel/CS=BR/Uni./Planar	Diel/CS=BR/Uni./Planar
19	Diel/CS=BR/Dif./Planar	Diel/CS=BR/Uni./Planar	Diel/CS=BR/Diff/Planar
20	Diel/CS=BR/Dif./Planar	Diel/CS=BR/Dif./Planar	Diel/CS=BR/Uni./Planar

4.3. Ranking Hypotheses

At this point in the segmentation, we use our postulate that the simplest explanation is the best explanation for a hypothesis set. As noted previously, because each broad hypothesis can provide a good approximation to the data, to implement the MDL principle we rank order the hypotheses into classes, or tiers, according to their relative simplicity in explaining the combined image regions. It is important to realize this stage of the segmentation process is dependent upon the specific region being described. For example, a region of uniform pixel values, can easily be described by a region of homogeneous color under a diffuse illumination environment. Regions such as those in our example, however, require a surface of non-homogeneous color if the illumination environment is diffuse. By using vision tools such as isobrightness contours, color histograms [21], and normalized color [14], as well as reasoning about the possible realizations given specific regions and hypotheses, we believe it is possible to rank-order the resulting merged hypotheses.

We finish our example by rank-ordering the final twenty hypotheses for the example image. We realize that we are

using some human reasoning in this process, but it is the first step towards developing a more rigorous, computable process.

Clearly, the simplest explanation for this scene is hypothesis 1, proposing that each region is a colored dielectric under uniform illumination. This hypothesis, the first hypothesis in Table 1, belongs by itself in the first tier. We propose this hypothesis as the simplest explanation because a realization exists where each element of the hypotheses for three sub-regions is both homogeneous and simple. Because of the homogeneity and simplicity, we can specify the scene with a small number of parameters and recreate it exactly. These parameters include the body reflection color of each region of the sphere, the radius of the sphere, the position of the sphere, the position of the light source, the color of the light source, and the parameters of the roughness model--e.g. Cook-Torrance. For no other hypothesis in Table 1 is there as compact a realization.

Tier two contains hypotheses 9 and 16, corresponding to a colored planar dielectric under white uniform or diffuse illumination, hypothesis 8, corresponding to a painted ball under diffuse illumination, and hypotheses 17 and 18 specifying that the image is planar or curved white metal reflecting colored light. The first two hypotheses propose that we are looking at an image of a picture. The grey metal hypotheses propose that we are looking at the reflection of an object in a mirror. We place these hypotheses in tier two because each of them puts all of the complexity for the image into a single element of the description, and the other elements of the hypotheses are simple and homogeneous over all three regions. Furthermore, each of these hypotheses for the image are realizable without the use of strange light sources or careful setup.

In tier three we place hypotheses 19 and 20, corresponding to white dielectrics under colored illumination. One realization of these hypotheses is a white object on which the scene is being projected. Note that the use of either active optical elements--a lens in a projector--or careful positioning and screening of the light sources may be necessary to recreate these hypotheses in a lab. Nevertheless, the remaining elements of the hypothesis are simple and homogeneous for all three sub-regions, which differentiates these hypotheses from the remaining ones in tier four.

Tier four contains the remaining 12 hypotheses, each of which is some combination of colored dielectrics under uniform or diffuse illumination. In none of these hypotheses are all of the regions homogeneous in their simple elements. This differentiates them from the first three tiers and make their physical realization more complex, or 'weird.'

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

5. Conclusions

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

We have not presented an implementation based upon our analysis. Implementation of even subsections of this method will be a large undertaking. The work by Breton *et al.* [5], for example, demonstrates the type of reasoning and algorithms necessary for each of the fundamental hypotheses. Their method, which analyzes large sets of surface shapes and lighting positions for a Lambertian surface, fits completely within a single fundamental hypothesis. As demonstrated by the taxonomy of fundamental hypotheses, there are some areas in which very little research has been undertaken to date. Describing this segmentation method in a computable fashion, will require integrating numerous techniques from diverse areas of physics based vision.

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

The analysis of Rissanen [32] in his discussion of the selection of model classes also provides a methodology for our analysis, especially our selection of fundamental hypotheses, or model classes. The process of choosing the fundamental hypotheses is one of selecting an initial set of model families with which to analyze the image regions. Rissanen argues there is no algorithmic method for undertaking this task, and that human intuition is indispensable. What we have provided herein is a structured analysis of the segmentation problem that suggests a relatively small, justifiable set of models for the physics of a scene.

The potential of an integrated segmentation system based on a general model of the physics of image formation is tremendous. Because it would rely upon physical models, a proposed segmentation becomes not just a set of regions, but a physical explanation for every pixel in the image as well as how those explanations relate in the 3-D world with regard to shape, transfer function, and illumination. By considering multiple hypotheses for image regions, it should be able to provide multiple segmentations of the entire image, reflecting in a structured manner the ambiguity that is present in the mapping from an image to the real world. Finally, because the physical models are general enough to capture virtually any illumination environment, transfer function, or surface shape, this segmentation method has the potential to work on a wide range of images without prior knowledge of the scene.

Acknowledgements:

We would like to thank Glenn Healey, Lawrence Wolff, and John Krumm for their fruitful and insightful discussions of the issues raised in this paper.

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

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

[3] P. Beckmann, and A. Spizzochino, *The Scattering of Electromagnetic Waves from Rough Surfaces*, Norwood, Artech House, 1987.

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

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

[6] C. R. Brice and C. L. Fenema, "Scene analysis using regions," *Artificial Intelligence* 1, 205-226, 1970.

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

[8] M. H. Brill, "Photometric models in multispectral machine vision," in *Physics-Based Vision: Principles and Practice, Color*, ed. G. Healey, S. Shafer, and L. Wolff, Boston, Jones & Bartlett Publishers, 1992.

[9] M. Cohen and D. Greenberg, "The hemi-cube: a radiosity solution for complex environments," *Computer Graphics Proc. of SIGGRAPH-85*, pp.31-40, 1985.

[10] R. L. Cook and K. E. Torrance, "A Reflectance Model for Computer Graphics," *Computer Graphics* 15 (3), pp.307-316, 1981.

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

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

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

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

[15] P. Hanrahan and W. Krueger, "Reflection from Layered Surfaces due to Subsurface Scattering," *Computer Graphics Proc. of SIGGRAPH-93*, pp165-174, 1993.

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

[17] R. S. Hunter, *The Measurement of Appearance*, John Wiley and Sons, New York, 1975.

[18] D. B. Judd and G. Wyszecki, *Color in Business, Science, and Industry*, 3rd ed., John Wiley and Sons, New York, 1975.

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

[20] G. J. Klinker, S. A. Shafer and T. Kanade, "Using a Color Reflection Model to Separate Highlights from Object Color," in *Proceedings of International Conference on Computer Vision*, IEEE, New York, pp. 145-150, June 1987.

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

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

[23] Y. G. Leclerc, "Constructing Simple Stable Descriptions for Image Partitioning," *International Journal of Computer Vision*, 3, 73-102, 1989.

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

[25] H.-C. Lee, E. J. Breneman, and C. P. Schulte, "Modeling light reflection for color computer vision," *IEEE Trans. on Pattern Analysis and Machine Intelligence* PAMI-12(4),pp.402-409, April 1990.

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

[27] A. Leonardis, A. Gupta, and R. Bajcsy, "Segmentation as the Search for the Best Description of the Image in Terms of Primitives," in *Proceedings of International Conference on Computer Vision*, IEEE, pp.121-125, 1990.

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

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

[30] 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.

[31] M. Oren and S. K. Nayar, *Generalization of the Lambertian Model and Implications for Machine Vision*, Columbia University, Computer Science TR CUCS-057-92.

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

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

[34] R. Stone and S. Shafer, "The Determination of Surface Roughness from Reflected Step Edges," Submitted to JOSA, 1993.

[35] J. M. Tenenbaum, and H. G. Barrow, *Experiments in Interpretation-Guided Segmentation*, Artificial Intelligence Center, Stanford Research Institute Technical Note 123, March 1976.

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

[37] S. Tominaga and B. A. Wandell, "Standard surface-reflectance model and illuminant estimation," *Journal of the Optical Society of America* A 6(4), pp.576-584, April 1989.

[38] K. Torrance and E. Sparrow, "Theory for Off-Specular Reflection from Roughened Surfaces," in *Journal of the Optical society of America*, 57, pp.1105-1114, 1967.

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

[40] L. B. Wolff and T. E. Boult, "Polarization/Radiometric Based Material Classification," in *Proceedings of the International Conference on Computer Vision and Pattern Recognition*, pp.387-395, IEEE, San Diego, CA, 1989.

[41] Y. Yakimovsky and J. Feldman, "A semantics-based decision theory region analyzer," in *Proceedings 3rd International Joint Conference on Artificial Intelligence*, 1973, p. 580-588.

Plate 1. A complex scene composed of numerous materials, textures, and shapes.

Plate 2. An object, a mirror image of the object, and a picture of the object.

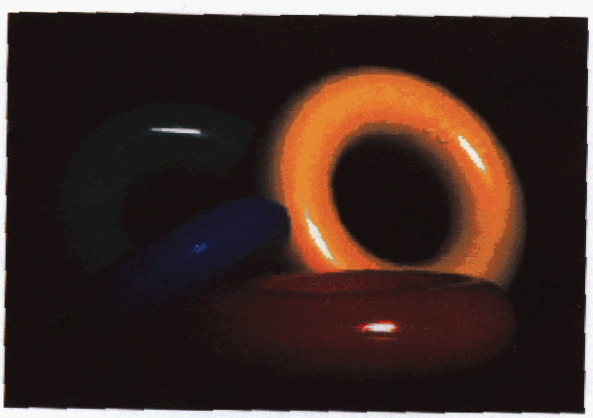


Plate 3. Image of uniformly colored inhomogeneous dielectrics.

Plate 4. Image of a single multi-colored object.

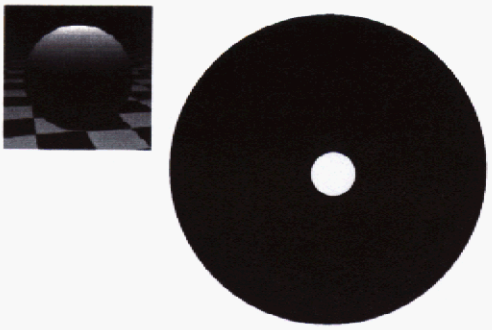


Plate 5. Illumination environment for inset image: orthogonal mapping of a white light source directly overhead.

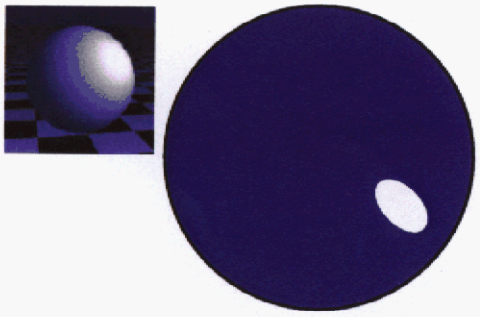


Plate 6. Blue ambient light with a white circular source to the right and behind.

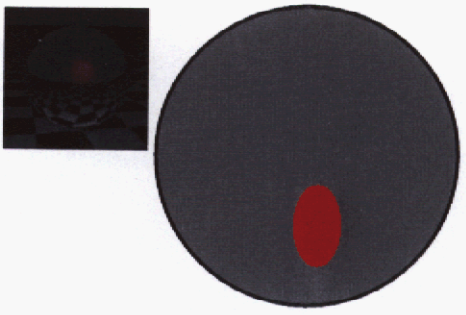


Plate 7. Grey ambient light with red light reflected off another object.

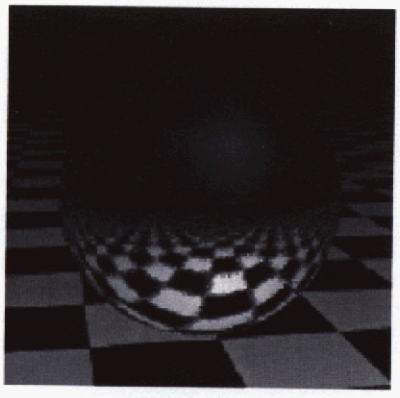


Plate 8. Illustration of the transfer function for a slightly rough metal object.

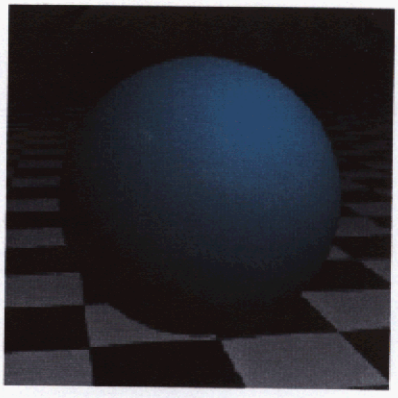


Plate 9. Illustration of the transfer function for a slightly rough plastic object.

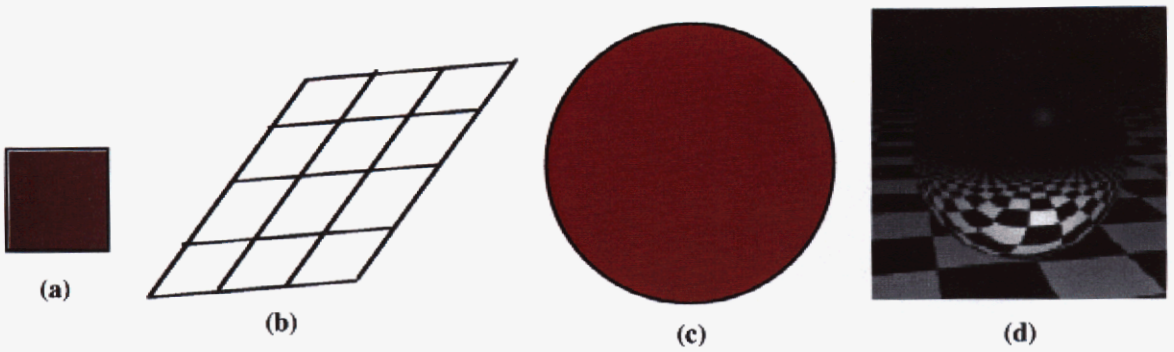


Plate 10. Illustration of a metal hypothesis: (a) actual region, (b) wire frame surface representation (planar), (c) illumination environment (diffuse), (d) transfer function (metal).

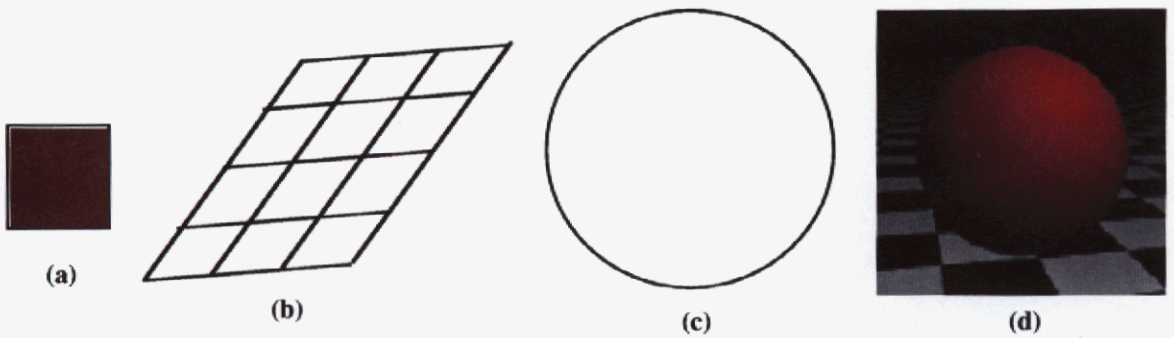


Plate 11. Illustration of a dielectric hypothesis: (a) actual region, (b) wire frame surface representation (planar), (c) illumination environment (diffuse), (d) transfer function (dielectric).

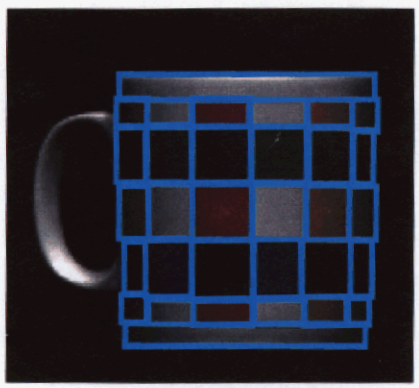


Plate 12. Mug divided into idealized uniform chromaticity regions.

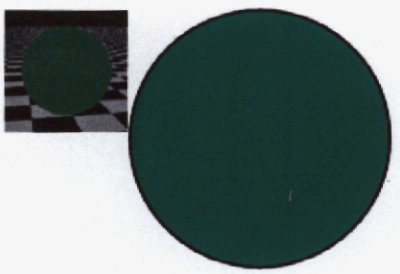


Plate 13. Diffuse illumination environment.

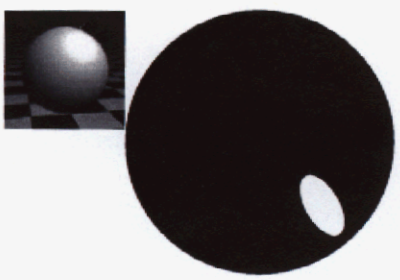


Plate 14. Uniform illumination environment.

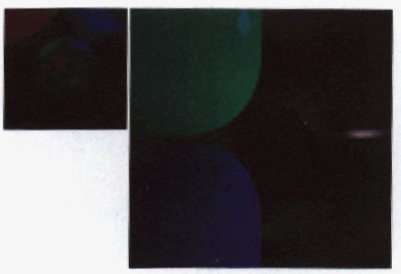
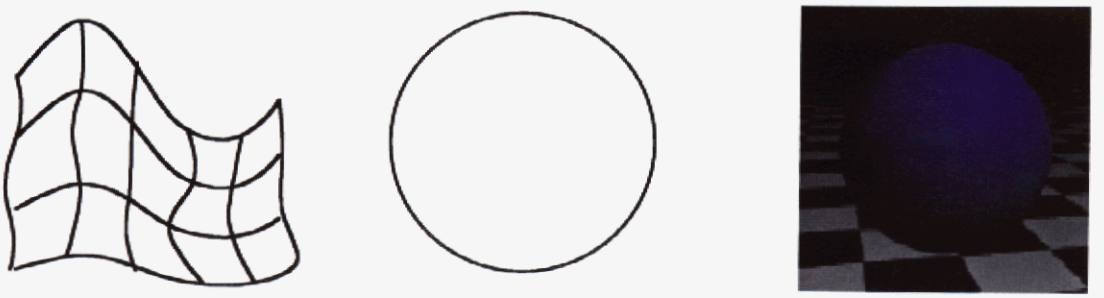
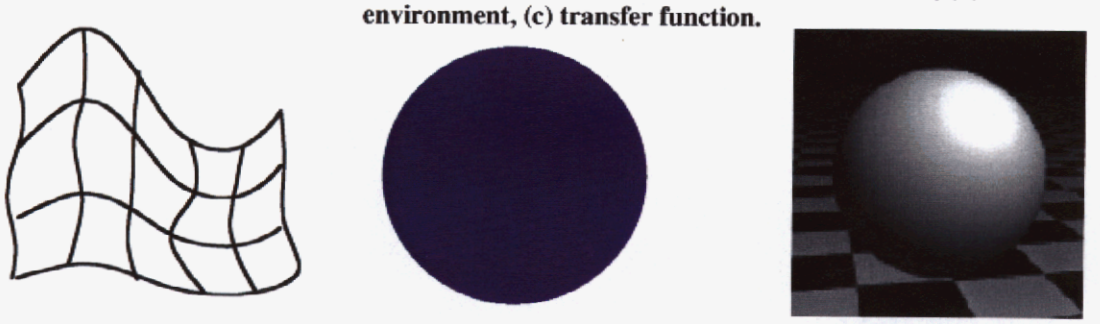


Plate 15. General Illumination Environment.



(a) (b) (c)

Plate 16. Fundamental hypothesis with body reflection as color source: (a) surface, (b) illumination environment, (c) transfer function.



(a) (b) (c)

Plate 17. Fundamental hypothesis with illumination as color source: (a) surface, (b) illumination environment, (c) transfer function.

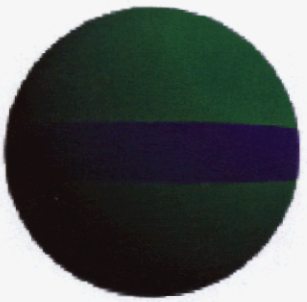


Plate 18. Three-color Lambertian sphere.

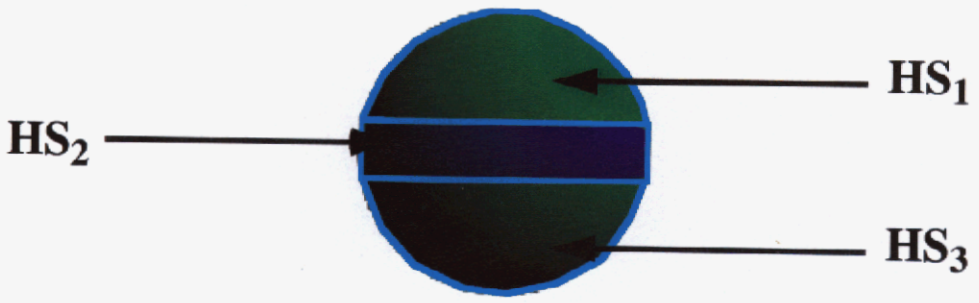


Plate 19. Three hypothesis sets in the example image.

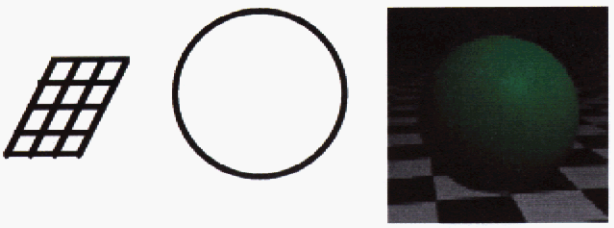


Plate 20. Hypothesis 1: planar--diffuse ill.--colored dielectric

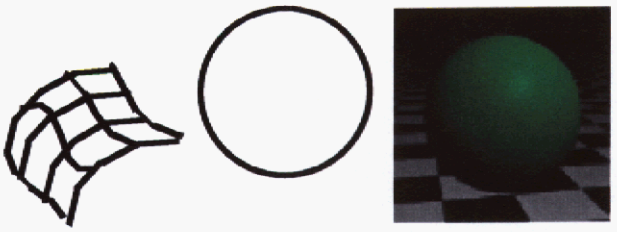


Plate 21. Hypothesis 2: curved--diffuse ill.--colored dielectric

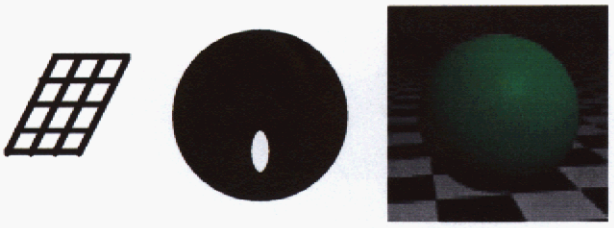


Plate 22. Hypothesis 3: planar--uniform ill.--colored dielectric

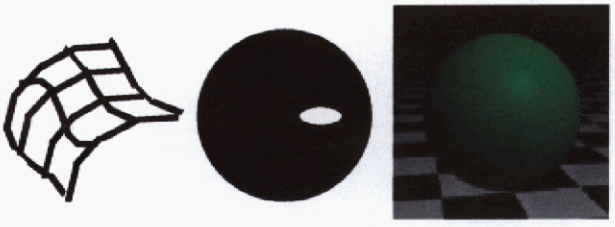


Plate 23. Hypothesis 4: curved--uniform ill.--colored dielectric

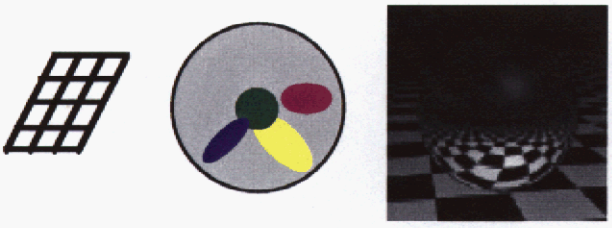


Plate 24. Hypothesis 5: planar--general ill.--grey metal

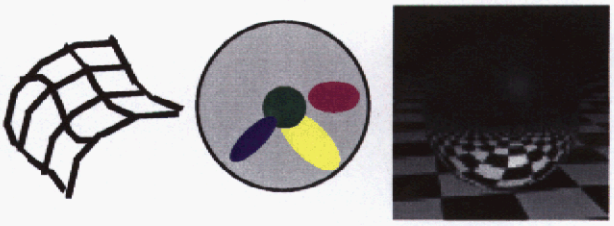


Plate 25. Hypothesis 6: curved--general ill.--grey metal

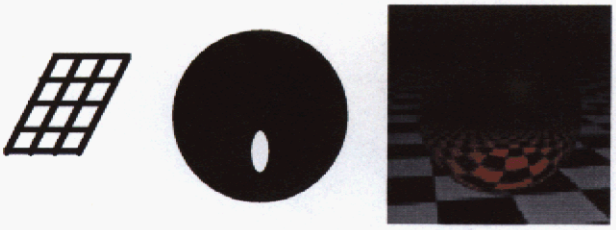


Plate 26. Hypothesis 7: planar--uniform ill.--colored metal

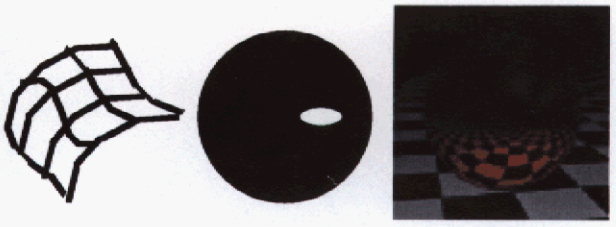


Plate 27. Hypothesis 8: curved--uniform ill.--colored metal

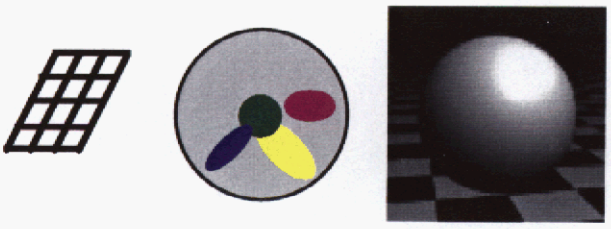


Plate 28. Hypothesis 9: planar--general ill.--grey or white dielectric

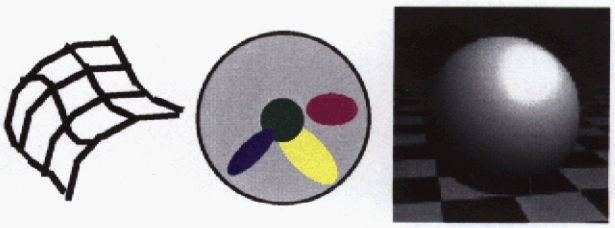


Plate 29. Hypothesis 10: curved--general ill.--grey or white dielectric