

A Physical Approach to Color Image Understanding

Gudrun J. Klinker, Steven A. Shafer and Takeo Kanade

Abstract

In this paper, we present an approach to color image understanding that can be used to segment and analyze surfaces with color variations due to highlights and shading. The work is based on a theory - the Dichromatic Reflection Model - which describes the color of the reflected light as a mixture of light from surface reflection (highlights) and body reflection (object color). In the past, we have shown how the dichromatic theory can be used to separate a color image into two intrinsic reflection images: an image of just the highlights, and the original image with the highlights removed. At that time, the algorithm could only be applied to hand-segmented images. This paper shows how the same reflection model can be used to include color image segmentation into the image analysis. The result is a color image understanding system, capable of generating physical descriptions of the reflection processes occurring in the scene. Such descriptions include the intrinsic reflection images, an image segmentation, and symbolic information about the object and highlight colors. This line of research can lead to physics-based image understanding methods that are both more reliable and more useful than traditional methods.

1. Introduction

When humans are asked to describe a picture, they generally give a list of objects and their relative positions in the scene. A closer look at the image reveals that they have omitted a lot of detail in the description. Objects cast shadows, possibly upon other objects. The brightness on the objects varies, appearing much darker at object parts that are tilted away from the light than where surfaces face the light directly. Moreover, some objects have highlights; we may even see a mirror image of one object on another object. These optical effects are caused by various physical processes by which light interacts with matter. If a vision system is expected to generate descriptions similar in quality to the ones given by humans, the system it will have to discount the influence of these physical processes.

Physical processes in the scene have not been a strong point of interest in the traditional line of computer vision research. It has been common practice to divide the image understanding problem into two phases, a low-level segmentation or feature extraction phase and a higher-level reasoning phase in which the image features are related to object features described in object models of the scene [9, 17]. Within this line of research, image segmentation has been considered to be a statistical image processing problem with the major concern being to determine statistically significant changes of pixel values under the presence of noise - under the implicit assumption that such significant changes generally correspond to object boundaries in the scene. However, the generated edge or region images outlined not only material boundaries, but also shadows, highlights, and object edges along which the surface orientation changes abruptly. The segmentation was then passed to a higher-level phase which tried to combine regions across highlights or shadows either by matching image features with objects models [9] or by determining the physical cause of edges or regions [6, 28].

Recently, work in image understanding has started to use intrinsic models of physical processes in the scene to analyze intensity or color variations in the image [2, 6, 11, 14, 16, 18, 19, 28, 32, 35]. But such analyses have not been applied to image segmentation. This paper presents an approach to color image understanding that uses an intrinsic reflection model, called the Dichromatic Reflection Model [30], to generate an image segmentation, along with descriptions of object and highlight colors and intrinsic reflection images, showing how shading and highlight reflection vary in the image. The color image analysis system alternates between generating hypotheses about the scene from the image data and verifying whether the hypotheses fit the image. The hypotheses relate object color, shading, highlights and camera limitations to the shapes of color clusters in local image areas. They are used to segment the images and to separate them into two intrinsic images, one showing the scene without highlights, and the other one showing only the highlights. In this way, the algorithm is driven by a physical model of light reflection to incrementally identify local and global properties of the scene and to use them in interpreting and segmenting pixels in the images. The system adapts the image interpretation process to local scene characteristics and reacts differently to color and intensity changes at different places in the image. This line of research is aimed at developing physics-based low-level image analysis methods that are both more reliable and more useful than traditional approaches.

2. The Dichromatic Reflection Model

This section provides a short overview of Shafer's Dichromatic Reflection Model [30]. A more detailed description and discussion can be found in [20, 22, 30].

The Dichromatic Reflection Model describes the light, $L(\lambda, i, e, g)$, which is reflected from a point on a dielectric, non-uniform material as a mixture of the light $L_s(\lambda, i, e, g)$ reflected at the material surface and the light $L_b(\lambda, i, e, g)$ reflected from the material body (see Figure 1). The parameters i , e , and g describe the angles of the incident and emitted light and the phase angle; λ is the wavelength parameter. L_s is called the *surface reflection component*. It generally has approximately the same spectral power distribution as the illumination and appears as a highlight or as gloss on the object. L_b is called the *body reflection component*. It provides the characteristic object color and exhibits the properties of object shading.

$$L(\lambda, i, e, g) = L_s(\lambda, i, e, g) + L_b(\lambda, i, e, g) \quad (1)$$

Figure 1: Light reflection of dielectric materials

The model separates the spectral reflection properties of L_s and L_b from their geometric reflection properties, modeling them as a products of spectral power distributions, $c_s(\lambda)$ or $c_b(\lambda)$, and geometric scale factors, $m_s(i, e, g)$ or $m_b(i, e, g)$, which describe the intensity of the reflected light. Substituting these terms into equation (1), we obtain the Dichromatic Reflection Model equation:

$$L(\lambda, i, e, g) = m_s(i, e, g)c_s(\lambda) + m_b(i, e, g)c_b(\lambda) \quad (2)$$

The model thus describes the light that is reflected from an object point as a mixture of two distinct spectral power distributions, $c_s(\lambda)$ and $c_b(\lambda)$, each of which is scaled according to the geometric reflection properties of surface and body reflection. In the infinite-dimensional vector space of spectral power distributions (each wavelength defines an independent dimension in this vector space [15, 29]), the reflected light can be described as a linear combination of the two vectors $c_s(\lambda)$ and $c_b(\lambda)$.

3. Object Shape and Spectral Variation

The Dichromatic Reflection Model describes the spectral properties separately for every single pixel on an object. In itself, this description is not yet very helpful in computer vision since it describes the light at each object point by a set of four (so far unknown) factors. It does not provide a mechanism to uniquely determine these factors from the reflected light beam. This section explores the observation that the light reflected from all points on an object uses the same two spectral vectors and that these two factors are thus constant over an object. We will now show how object shape determines spectral variation on the object and how this spectral variation is related to the body and surface reflection vectors.

3.1. Basic Principles

An investigation of the geometrical properties of surface and body reflection reveals that the light mixtures form a dense spectral cluster in the dichromatic plane. The shape of this cluster is closely related to the shape of the object. Figure 2 shows a sketch of a shiny cylinder. The left part of the figure displays the magnitudes of the body and surface reflection components. The lighter curves show the loci

of constant body reflection, assuming perspective viewing and illumination geometry and Lambertian body reflection. The darker curves are the loci of constant surface reflection. Since $m_s(i,e,g)$ decreases sharply around the object point with maximal surface reflection, m_{smax} , these curves are shown only in a small object area. We call the points in this area *highlight points*. The remaining object points are *matte points*. The right part of the figure shows the corresponding spectral histogram in the dichromatic plane. As we will describe below, the object points form two linear clusters in the histogram.

Figure 2: The shape of the spectral cluster for a cylindrical object

Light reflection at **matte points** is primarily determined by the body reflection process. In principle, matte points may contain a some surface reflection, as the result of light diffusion at a rough material surface. However, the following analysis assumes that this surface reflection component can be neglected. The observed light at matte points then depends mainly on $c_b(\lambda)$, scaled by $m_b(i,e,g)$ according to the geometrical relationship between the local surface normal of the object and the viewing and illumination directions. Consequently, the matte points form a *matte line* in the dichromatic plane in the direction of the body reflection vector, $c_b(\lambda)$, as shown in the right part of Figure 2.

Highlight points exhibit both body reflection and surface reflection. However, since $m_s(i,e,g)$ is much more sensitive to a small change in the photometric angles than $m_b(i,e,g)$, the body reflection component is generally approximately constant in a highlight area, as displayed by the curve with label m_{bH} in Figure 2. Accordingly, the second term of the Dichromatic Reflection Model equation (2) has a constant value, $m_{bH}c_b(\lambda)$, and all spectral variation within the highlight comes from varying amounts of $m_s(i,e,g)$. The highlight points thus form a *highlight line* in the dichromatic plane in the direction of the surface reflection vector, $c_s(\lambda)$. The line departs from the matte line at position $m_{bH}c_b(\lambda)$, as shown in Figure 2. More precisely, the highlight cluster looks like a slim, skewed wedge because of the small variation of the body reflection component over the highlight.

The combined spectral cluster of matte and highlight points looks like a skewed T. The skewing angle of the T depends on the spectral difference between the body and surface reflection vectors while the position of the highlight line depends on the illumination geometry, as described below.

3.2. Relationship between Illumination Geometry and Spectral Histogram Shape

There exists a close relationship between the illumination geometry and the amounts of body and surface reflection on an object depend. This relationship influences the shape of the clusters in the spectral histogram. It describes constraints that an image analysis program can use when it analyzes spectral variation. We will now derive such a constraint, the *50%-heuristic*, for spherical objects. The heuristic states that, under reasonable assumptions, the highlight cluster starts in the upper 50 percent of the matte cluster. It will be used in section 5 for image segmentation to distinguish spectral changes between matte and highlight points on one object from spectral changes between neighboring objects.

The amount of body reflection under the highlight depends on the phase angle g between the illumination and viewing direction: If g is very small, i.e.: the camera is very close to the light source, the

incidence direction of the light at the highlight is close to the surface normal and the underlying amount of body reflection is very high. The highlight line then starts near the tip of the matte line, and the skewed T becomes a skewed L or a "dog-leg" [6, 7]. When the phase angle g increases, the highlight moves away from the area with maximal body reflection, and the amount of body reflection under the highlight decreases. The highlight line moves accordingly away from the tip of the matte line.

The following analysis investigates this relationship more precisely for the case of a spherical object with radius r , viewed and illuminated under perspective projection (see Figure 3). The analysis assumes that the sphere is illuminated by a point light source at some distance d from the object center. It further assumes that the camera is positioned at the same distance d from the object center and that an arbitrary phase angle g exists between the viewing and the illumination direction. Although these assumptions on the illumination geometry and the object shape are limiting, the following analysis provides general insight into the relationship between object shapes and the shapes of spectral clusters.

Figure 3: Assumed illumination geometry

Distance d determines the size of the cone that illuminates the sphere (described by γ), as well as the range of phase angles $[-g_{max}, g_{max}]$ under which the camera sees a part of the illuminated area:

$$\gamma = \arctan \frac{r}{\sqrt{d^2 - r^2}} \quad (3)$$

$$g_{max} = 180^\circ - 2\gamma = 2 \arctan \left(\frac{\sqrt{d^2 - r^2}}{r} \right) \quad (4)$$

Figure 4 describes how the highlight and its underlying body reflection component m_{bH} are related to phase angle g . Since camera and light source are positioned at the same distance d from the object center, the maximal surface reflection component occurs where the surface normal, \mathbf{n}_H , bisects the phase angle. Under perspective projection, the illumination angle i_H at the highlight depends on the position of the highlight which, in turn, depends on g . Using the law of sines and the law of cosines, i_H can be described as

$$i_H = \arcsin \frac{d \sin(g/2)}{\sqrt{d^2 + r^2 - 2dr \cos(g/2)}} \quad (5)$$

Assuming Lambertian body reflection, $m_b(i, e, g) = \cos i$, the underlying body reflection component, m_{bH} , at the highlight is then given by

$$m_{bH} = \cos(i_H) = \sqrt{1 - \frac{d^2 \sin^2(g/2)}{d^2 + r^2 - 2dr \cos(g/2)}} \quad (6)$$

Figure 4: Photometric angles at highlight

According to equations 5 and 6, m_{bH} approaches 1 when g goes to 0° , confirming the informal description at the beginning of this section that the spectral cluster looks like a skewed L when camera and light source are close. When g approaches g_{max} , on the other hand, m_{bH} decreases to 0. When relating m_{bH} to the length of the matte color cluster, we need to take into account that the point with globally maximal body reflection, m_{bMax} , is not always visible from the camera. If it is occluded, the matte line in the spectral histogram does not extend entirely to $m_{bMax} c_b(\lambda)$ but only to the point representing the brightest visible point, $m_{bMaxVisible} c_b(\lambda)$. Figure 5 displays the illumination geometry for the brightest matte point that is visible from the camera. The surface normal, $\mathbf{n}_{MaxVisible}$, at this object point is determined by the angle α , which is given as

$$\alpha = \max(g - (90^\circ - \gamma), 0) = \begin{cases} g - (g_{max}/2) & \text{if } g \geq g_{max}/2 \\ 0 & \text{otherwise} \end{cases} \quad (7)$$

Figure 5: Brightest matte point that is visible from the camera

Following the derivation of equations 5 and 6, the angle between the illumination vector and the surface normal at the local maximum, as well as the amount of body reflection at that point are given by equations 8 and 9:

$$i_{MaxVisible} = \arcsin \frac{d \sin \alpha}{\sqrt{d^2 + r^2 - 2dr \cos \alpha}} \quad (8)$$

$$m_{bMaxVisible} = \cos(i_{MaxVisible}) = \sqrt{1 - \frac{d^2 \sin^2 \alpha}{d^2 + r^2 - 2dr \cos \alpha}} \quad (9)$$

The starting point of the highlight line on the matte line can now be described relative to the length of the matte line. The ratio q of body reflection at the highlight in relation to the maximally visible amount of body reflection is expressed as:

$$q = \frac{m_{bH}}{m_{bMaxVisible}} \quad (10)$$

Figure 6 shows how q varies as a function of g and d . Each curve in the figure describes for a different fixed d how q depends on g . In all curves, q approaches 0.5 as g approaches g_{max} . Note, however, that q is not defined for $g = g_{max}$ where $i_{bMaxVisible} = i_{bH} = 90^\circ$ and thus $m_{bMaxVisible} = m_{bH} = 0$. Distance d is varied between the curves. For small values of d , the curves exhibit a minimum in the middle of the curve. With increasing d , this minimum becomes less pronounced and q becomes a monotonically decreasing function with an inflection point. In the extreme, when $d = \infty$, the curve for q assumes the shape of a cosine function for phase angles g under which the globally brightest point is visible ($g \leq 90^\circ$), and the shape of a cosecant function, if the brightest point is invisible. A detailed derivation of the formula is given in [20].

$$q_\infty(g) = \begin{cases} \cos(g/2) & \text{if } g \leq 90^\circ \\ 1/(2 \sin(g/2)) & \text{if } g \geq 90^\circ \end{cases} \quad (11)$$

Figure 6: Starting point q of the highlight line as a function of phase angle g and distance d

The figure indicates that if d is a moderate or large multiple of r ($d \geq 5r$), q does not drop below 0.5. The body reflection component under the highlight then is always at least half as large as the body reflection component at the brightest visible matte point. Under such illumination geometries, the highlight line always starts in the upper 50% of the matte line. Since laboratory set-ups for scenes with several objects in the field of view generally use camera and illumination distances that are larger than the fivefold object size, we use the above finding as the *50%-heuristic* in the segmentation algorithm in section 5.

4. A Camera Model

To account for camera limitations in real images, we combine the Dichromatic Reflection Model with a camera model. This section describes briefly how some characteristics of CCD-cameras influence the pixel values in real images. A more detailed description, including color pictures with color clusters from real images, can be found in [20, 22].

- **Spectral Integration:**

Instead of using the infinite-dimensional vector space of spectral color distributions, research in computer color vision generally analyzes a three-dimensional *color space*. This color space comes from color images obtained by taking three pictures of the scene, through a red, a green and a blue color filter. The process transforming the spectrum of an incoming light beam into a color triple $\mathbf{C} = [R, G, B]$ is called *spectral integration*. It is a linear transformation from the infinite-dimensional vector space to the three-dimensional color space [8, 29]. The Dichromatic Reflection Model is then describable as a function of two three-dimensional vectors, $\mathbf{C}_s = [R_s, G_s, B_s]$ and $\mathbf{C}_b = [R_b, G_b, B_b]$ which span a dichromatic plane in the three-dimensional color space:

$$\mathbf{C} = m_s \mathbf{C}_s + m_b \mathbf{C}_b$$

(12)

- **Color Clipping:**

Real cameras have only a limited dynamic range to sense the brightness of the incoming light. This restricts our analysis of light reflection to a *color cube*. If the incoming light is too bright at some pixel position, the camera cannot sense and represent it adequately and the light is *clipped* in one or more color bands.

- **Blooming:**

In CCD-cameras, too much incident light at a pixel may completely saturate the sensor element at this position. This causes *blooming* in the camera as a result of which adjacent pixels increase their values. We call such neighboring pixels *bloomed color pixels*.

- **Color Balancing:**

CCD-Cameras are generally much less sensitive to blue light than to red light. To provide an equal scaling on the three color axes in the color space, we rescale the pixel data separately in the color bands by changing the aperture. We refer to this procedure as *aperture color balancing*. We also use a total IR suppressor in front of our camera to eliminate the CCD-camera's sensitivity to infrared light.

- **Gamma Correction:**

The color pixels depend on the camera response to incident light flux. Due to *gamma-correction*, the camera output is generally related by an inverse power law to the

incident flux. It introduces curvature into the color space, distorting the linear properties of the Dichromatic Reflection Model. We linearize the color data, reversing the gamma-correction process.

- **Chromatic Aberration:**

Since the refraction index of optical glass is a function of wavelength, picture taking suffers from *chromatic aberration*: The focal length, f , of the camera increases with wavelength and, for a fixed focal length, only one color band of an image can be perfectly in focus. The other two image bands show slightly magnified or shrunk images of the scene. Chromatic aberration has a strong influence on color changes in the outer image areas. Since the color bands do not increase or decrease simultaneously, pixels in the rising and falling parts of the profiles generally exhibit color changes that have little relationship to the physical processes in the scene. This effect depends on the quality of the camera lens. It appears quite strongly in some of our images [20]. Since we have not yet been able to eliminate this effect, it limits the performance of our algorithm.

So far, we have discussed how optical and camera properties influence the image function. We will now present an image understanding system that demonstrates how physical knowledge can be used for interpreting images. The system consists of two major components which will be presented in the next sections: a color image segmentation algorithm and a method to split color pixels into their body and surface reflection components.

5. Color Image Segmentation

The goal of segmentation is to identify objects in an image, as delimited by material boundaries. Because most current color segmentation methods are based on a very simple interpretation of color changes in an image, they generally segment images not only along material boundaries but also along other lines exhibiting color or intensity variations, such as highlight and shadow boundaries, or internal object edges with significant shading changes. The Dichromatic Reflection Model provides a more sophisticated interpretation scheme relating the physics of light reflection to color changes in the image and in the color space. We use the Dichromatic Reflection Model in a segmentation algorithm to distinguish color changes at material boundaries from color changes due to shading changes or highlights.

5.1. Using the Dichromatic Reflection Model in Color Image Understanding

The Dichromatic Reflection Model and the sensor model describe processes that occur in the scene and in the camera. The previous sections have projected the influence of these processes from scene properties onto image properties. We will now invert this line of reasoning and design an algorithm which concludes from the color variations in the image which processes have taken part in the image formation process. Unfortunately, the interpretation process is locally ambiguous because the influence of any optical process causes local color variation. The algorithm needs to accumulate local color variations over extended image areas to determine distinguishing characteristics of the processes, such as the T-shape of a color cluster. But how can the optimal extent of an area be determined? There seems to be a circular problem of, on one hand, needing a prior segmentation to relate color variation to physical processes and, on the other hand, needing an understanding of the physical processes to provide a good

segmentation.

Conceptually, there are two different ways to approach this problem. The algorithm may start out with very large image areas and subsequently shrink or split them until the regions correspond to objects in the scene. Alternatively, it may start with small local areas and then merge or grow them. Within the framework of the Dichromatic Reflection Model, the first method identifies object areas by projecting the entire image into the color space and then distinguishing between several skewed T's in the color space. It may encounter problems when several objects with very similar colors exist in the scene, such that color clusters from different objects overlap. This problem is illustrated in Figure 2. The figure displays the color histogram of the image in Figure 1, showing a scene of eight plastic objects under white light. The various clusters overlap significantly. The second method starts out with small image areas and merges or grows them, including neighboring pixels or areas that fit the same skewed T. Since local color variation on flat or dark objects may be similar in magnitude to camera noise, the local approach has to know how to distinguish camera noise from systematic color variation. This paper presents an algorithm that uses this local approach to color image understanding.

Color Figure 1: Scene with eight plastic objects

Color Figure 2: Color histogram of the scene with eight plastic objects

Our segmentation method alternates between generating optical hypotheses from local image data and verifying whether the hypotheses fit the image (see Figure 7). The hypotheses relate object color, shading, highlights and camera limitations to the shapes of color clusters in local image areas. The algorithm looks in a bottom-up process for color clusters from local image areas that exhibit the characteristic features of the body and/or surface reflection processes. When it finds a "promising" cluster in an image area, it generates a hypothesis that describes the object color and/or highlight color in the image area and determines the shading and highlight components of every pixel in the area. The algorithm then applies the new hypothesis to the image, using a region-growing approach to determine the precise extent of the image area to which the hypothesis applies. This step verifies the applicability of the hypothesis. The physical knowledge embedded in the new hypothesis can be used to split every pixel in the new image area into its intrinsic body and surface reflection components, as will be shown in section 6. The resulting intrinsic images and the hypotheses together instantiate the general concepts of shading and highlights of the Dichromatic Reflection Model, describing the specific reflection processes that occur in this part of the scene.

Figure 7: Using an Intrinsic Model for Image Understanding

The physical knowledge from all hypotheses instantiated so far can be used to incrementally adapt the image analysis steps to the already established knowledge about the scene. Our algorithm performs its generate-and-test analysis in several stages, each of which is related to a particular aspect of the Dichromatic Reflection Model or the sensor model. It starts out with the simplest and most common aspects and then uses the established knowledge to address the more complicated aspects. In this fashion, the algorithm performs the following steps:

1. Compute an initial, rough description of color variation in local image areas.
2. Generate hypotheses on matte color clusters. Exploit these hypotheses for image segmentation.
3. Extend the matte hypotheses into hypotheses on skewed T's in dichromatic planes. Resegment the image, exploiting these hypotheses.
4. Analyze the effects of blooming and color clipping.
5. Exploit the hypotheses to split the pixels into their reflection components.

Each of these steps will be described in detail below.

This control structure exploits a physical model of light reflection to incrementally identify local and global properties of the scene, such as object and illumination colors. It uses these properties to interpret the pixels in the images. By using this control structure, the algorithm can adapt its image interpretation process to local scene characteristics and react differently to color and intensity changes at different places in the image.

5.2. Generating Initial Estimates for Color Clusters

To start the image analysis process, our algorithm generates some initial estimates about the color variations in the image. These initial estimates are too coarse to be useful as hypotheses without further considerations. But they form the basis for the formulation of hypotheses about linear and planar color variation in the next sections.

5.2.1. A Classification Scheme for Local Color Variation

To estimate and classify local color variation in the image, the algorithm divides the image into small, non-overlapping windows of fixed size. It projects the color pixels from one window at a time into the color space and finds the principal components of the color distributions, as indicated by the eigenvectors and eigenvalues of the covariance matrix of the cluster [1, 10]. The eigenvalues and eigenvectors determine the orientation and extent of the ellipsoid that optimally fits the data. For convenience, the eigenvalues are sorted by decreasing size such that a reference to the first eigenvalue corresponds to the largest one, the second eigenvalue to the medium one and the third eigenvalue to the smallest one.

Principal component analysis is not a new concept in color image segmentation. Gershon and Ohta used it to determine optimal color feature axes for a histogram-based, recursive region-splitting approach. They expected the major principal component to indicate color changes between objects - a color feature along which they could best separate objects from one another. Such use of principal component analysis implicitly assumes that the significant color changes in an image occur at the object boundaries while color variation within a single object area is minimal. This is a special case of the dichromatic framework which we will develop below: the assumption holds only for matte clusters from dark or flat objects for which the color clusters from single objects form points or small spheres in the color space - no linear or even planar color variation on an object is accounted for. We will now present a more general investigation of the relationship between the principal components of the color clusters and physical

scene properties.

The shape of the ellipsoids provides information which relates local color variation to physical interpretations. We distinguish between eight classes of ellipsoids, for each of which exist a few physical interpretations. The classification is based on the number of eigenvalues that are approximately zero, within the limit of pixel value noise, σ_0 , in the image. The decision for each eigenvalue is based on a χ^2 -test with (n^2-1) parameters, where n is the window size. The algorithm then classifies each color cluster according to how many eigenvalues are determined to be significantly greater than zero. A summary of the classification is given in Table 1.

Table 1: Summary of Color Cluster Interpretations

- In zero-dimensional (*pointlike*) clusters, all three eigenvalues of the window are very small. No significant color variation exists in such a window and the body and surface reflection components are nearly constant. This is the case, if the window lies on a very flat object such that the photometric angles are nearly constant within the window. Alternatively, a pointlike cluster can originate from a matte dark object, independently of the surface curvature of such object. In the extreme, a perfectly black object does not exhibit any shading variation.
- One-dimensional (*linear*) clusters are clusters for which only the first eigenvalue is significantly larger than the estimated camera noise. Pixels in such a window may come from a matte object area forming part of a matte cluster. They can also come from the interior of a highlight area such that they form part of a highlight cluster. As a third possibility, the window may overlap the matte object areas of two neighboring objects that are very dark or flat. Such a window consists of two pointlike color clusters or of one dark pointlike cluster and one linear cluster which together fit well into one linear ellipsoid.
- Two-dimensional (*planar*) clusters have large first and second eigenvalues. The local color data fits a plane in the color cube. Such clusters occur at windows that cover some matte and some highlight pixels of one object. In this case, the third eigenvector of the color cluster determines the normal to the dichromatic plane of the object. But planar clusters also arise in windows that overlay matte pixels from two neighboring regions. Such windows have planar clusters because the matte clusters of all objects converge at dark pixels such that matte clusters are pairwise coplanar.
- In three-dimensional (*volumetric*) clusters, all three eigenvalues are large. Such color clusters may arise in the middle of highlights where color clipping and blooming significantly increase the noise in the pixel measurements. Volumetric color clusters also occur along material boundaries when three or more objects in different colors share a window or when a window overlays matte pixels of one object and matte and highlight pixels of another object.

Figure 3 shows the classification of the color clusters from Figure 1, using windows that each contain 10 x 10 pixels. Pointlike clusters are displayed in yellow, linear clusters in red, planar clusters in green, and volumetric clusters in blue. The image shows that the classifications relate in the expected way to scene properties: most matte object areas are covered by linear windows, while windows at material boundaries and at highlights are planar or volumetric.

Color Figure 3: Color cluster classification for initial image areas, scene with eight plastic objects

5.2.2. Determining Compatible Color Variation

If the window size is chosen appropriately, it is small compared to the sizes of the objects in the scene. Most windows lie then inside object areas and only few windows overlay material boundaries. There will generally be many windows within a single object area. All linear clusters of such windows indicate the direction of the same matte or highlight cluster and all planar clusters support the same hypothesis about a dichromatic plane. This subsection describes how our algorithm merges such windows to gather increased support for a small number of estimates about color clusters in the image.

The algorithm merges neighboring windows that have similar color characteristics. It proceeds in row-major order, testing for all pairwise combinations of neighboring areas whether the two can be merged. In order not to merge windows across material boundaries, it only merges windows, if both of them, as well as the resulting larger window, have the same classification. It uses the χ^2 -test described above to classify the larger window. Accordingly, it combines windows with pointlike clusters into larger pointlike windows; it merges linear windows into larger linear windows; and it merges planar windows into larger planar windows. It does not merge volumetric color clusters since there is no constraint on the resulting cluster. This process continues until no more areas can be merged. The results are initial hypotheses about the positions and orientations of pointlike, linear, and planar clusters in color space and their respective approximate extents in the image.

Figure 4 presents the results of merging neighboring windows of the same class from Figure 1. The image shows a few large image areas. Such image areas correspond roughly to the center of matte object areas. They are generally surrounded by many smaller areas close to material boundaries or highlights. Only very few planar windows in highlight areas have been merged, probably due to the increased spread of highlight pixels from color clipping and blooming.

Color Figure 4: Initial grouping into approximate image areas, scene with eight plastic objects

5.3. Linear Hypotheses

In its subsequent steps, our algorithm uses the initial estimates of color clusters to generate and exploit hypotheses about the existing color variation in the image. The above discussion and Figure 4 show that the initial estimates are rather coarse indications of local color variation: there are far more image areas than objects and, due to the coarse initial window width, the area boundaries do not coincide with the material boundaries. We now need to combine neighboring areas at a much finer granularity, resegmenting the image by considering every pixel on an individual basis. The algorithm reanalyzes and combines the estimates of local color variation by selecting suitable initial hypotheses for local color variation and using them to resegment the image. The hypotheses relate the color clusters to properties of the Dichromatic Reflection Model, describing color pixels as matte or highlight pixels or as pixels showing camera problems.

5.3.1. Generating Linear Hypotheses

The algorithm starts by choosing large image areas with linear color clusters as linear hypotheses. Large linear clusters generally correspond to the internal parts of large matte object areas. Such large areas are less influenced by estimation errors in local color variation than small image areas along material boundaries and in highlight centers with increased camera noise from color clipping and blooming. Linear color clusters are also easier to interpret than planar, pointlike or volumetric color clusters because their first eigenvector describes the orientation of major color variation in the cluster; information which can be related to a physical interpretation. The eigenvectors of planar clusters (e.g.: in areas close to highlights) are less useful because the directions of the first and second eigenvectors are linear combinations of the directions of several linear clusters (e.g.: of a matte and a highlight cluster). Their direction does not describe the branches of the skewed T of a dichromatic cluster and they cannot be related to the physical properties of body and surface reflection.

5.3.2. Exploiting Linear Hypotheses

The merging algorithm of the previous section has used a bottom-up approach to extract information about the scene from the image. In a top-down step, we now use a selected hypothesis to locally resegment the image. The linear hypothesis provides a model of what color variation to expect on the object part on which the image area lies. The mean value and the first eigenvector describe the position and orientation of a linear color cluster, while the second and third eigenvalues determine the extent of the color cluster perpendicular to the major direction of variation. According to the Dichromatic Reflection Model, color variation along the major axis can be attributed to a physical property of the scene, e.g.: a changing amount of body or surface reflection or a material boundary. At this stage, we attribute color variation perpendicular to the first eigenvector to random noise. Accordingly, we model the linear color cluster as a cylinder (see Figure 8). The radius depends on the estimated camera noise. Our algorithm excludes dark color pixels from its color analysis because all matte clusters merge near the dark corner of the color cube. By this *dark heuristic*, the cylinder is delimited at its dark end by a sphere which is centered at the black corner.

We use the color cylinder of the current hypothesis to locally resegment the image. The algorithm selects a start pixel from the image area associated with the color cluster. This pixel must have a color that is contained within the color cylinder. The algorithm then grows a four-connected region from this starting point, recursively examining the four neighbors of pixels on the fringe of the region, and including them if their color lies within the color cylinder. The result is an image area of pixels that are consistent with the current linear hypothesis. The boundary of this new region may be very different from the image area that was initially associated with the linear hypothesis. Due to the coarse initial window size, the initial image area may have contained pixels from neighboring objects or highlight pixels with some amount of surface reflection. Such pixels are excluded from the new region. The initial window may also have contained dark pixels in very shaded object areas or - as is the case in Figure 1 - pixels from a dark background. Due to the dark heuristic, such pixels are also excluded from the new region. On the other hand, the initial segmentation may have had several neighboring image areas on the same matte object part that could not be merged because the characteristics of the color clusters were biased by pixels that

came from other objects. Many pixels in such neighboring image areas may lie on the current matte object part and be consistent with the current linear hypothesis. They are included into the new image area. As a consequence, the new region can be significantly different from the initial image area.

Since there is generally more than one object in the scene, the algorithm iterates the above steps for each large image area with a linear cluster, selecting the areas by decreasing size. It stops when the next selected area is too small. Since all matte clusters converge at dark pixels, there exists a potential conflict between neighboring matte areas. The dark heuristic eliminates the most difficult cases; yet, the cylinders of neighboring clusters may still intersect beyond the selected dark threshold. This depends on the cylinder radius and on the angle between the two cylinders. Neighboring objects with very similar colors have conflicts even at fairly bright colors. Our algorithm assigns pixels with a color conflict to the cluster with the closest axis, as shown in Figure 8. This is called the *proximity heuristic*.

Figure 8: Proximity heuristic to resolve a color conflict between the color clusters of two neighboring matte object areas

In principle, the resulting "linear" regions may be related in any of several different ways to the physical processes in the scene. As discussed in section 5.2.1, a linear color cluster may be a matte cluster or a highlight cluster or even a combination of two clusters across a material boundary. However, linear color clusters from highlights and across material boundaries are generally much smaller than clusters from matte object parts. Since this section considers only linear hypotheses of large image clusters, the following sections assume that all newly grown regions correspond to matte linear clusters.

Figure 5 shows the results of selecting and applying linear hypotheses to the image with the eight plastic objects. The region boundaries outline the matte object parts in the scene, with the material boundaries being well observed. The highlight areas on the objects have not yet been analyzed. This sometimes divides the matte pixels on an object into several matte areas, as shown on the green donut in the lower right part of the image.

Color Figure 5: Linear segmentation, scene with eight plastic objects

5.4. Planar Hypotheses

The Dichromatic Reflection Model states that color variations on a single object lie in a plane in the color space. The linear hypotheses and the linear segmentation generated so far provide a good starting point to describe color variation in an image. But they do not account for the effects of surface reflection. The hypotheses will now be extended into planar hypotheses that describe dichromatic planes and skewed T's.

5.4.1. Generating Planar Hypotheses

In principle, information on color variation in a dichromatic plane may be obtained from inspecting the characteristics of the initially determined planar color clusters. However, such information is generally not very reliable, due to the small size of most planar image areas. Furthermore, the eigenvectors of such

planar clusters do not describe the skewed-T structure of the color cluster because the directions of the first and second eigenvectors are linear combinations of the directions of several linear clusters (e.g.: of a matte and a highlight cluster). Their direction does not describe the branches of the skewed T of a dichromatic cluster and they cannot be related to the physical properties of body and surface reflection. Instead of using such planar color information, our algorithm uses the existing linear hypotheses as a basis describing already one branch of skewed T's. It now determines the orientations and positions of the second branch of the skewed T's.

In this process, the algorithm considers all neighbors of a linear region as prospective highlight candidates. Some of such neighboring regions may be matte object areas of neighboring objects while others may be a part of a highlight on the current object. Such types of regions must be distinguished from one another such that the algorithm does not wrongly combine color information from two neighboring matte image areas into a planar hypothesis. To distinguish neighboring matte regions from highlight regions, Gershon has tested whether two clusters are nearly parallel or whether they intersect, suspecting that parallel clusters are neighboring matte clusters [7]. We replace this test by the more precise 50% heuristic of section 3: the color clusters of a matte and a highlight region on the same object intersect in the upper 50 percent of the matte cluster, while color clusters from neighboring matte regions converge at dark pixels. Accordingly, our algorithm tests whether the two clusters form a skewed T and meet in the upper 50 percent of the matte cluster. For this purpose, it searches for the brightest matte point in the color cluster to determine the length of the matte line. However, the matte cylinder may contain some highlight points from the foothill of the highlight cluster, as shown in Figure 9. Such highlight points may be brighter than the brightest matte point. To distinguish them from matte points, we exploit the observation that highlight clusters always grow inwards into the color cube, due to the additive nature of body and surface reflection. Accordingly, the algorithm chooses the brightest matte point only from pixels with color values on the outside of the matte line.

Figure 9: Finding the brightest matte pixel

Once the length of the matte line is determined, the algorithm computes the intersection points of the current matte cluster and of the neighboring prospective highlight cluster. It uses their color means and first eigenvectors to describe a matte and a prospective highlight line and to determine the two points on the two lines that are closest to one another. Comparing the distance to the camera noise, algorithm decides whether the clusters meet in a skewed T. Similarly, the algorithm tests whether the clusters intersect in the upper 50% of the matte cluster. We also check whether the neighboring cluster has a positive direction - a test related to the additive properties of body and surface reflection. If this is not the case, the window of the highlight candidate probably overlays a material boundary, partially covering pixels from the current matte region just outside its color cylinder and partially covering pixels from a neighboring matte region. Candidates that fail any of these tests are discarded.

After all these tests, there may still exist several highlight candidates because the highlight on the object may consist of a series of small windows that had not been merged. Since many such highlight areas are very small and since they may contain some clipped and bloomed pixels, the orientations of

their color clusters may vary significantly. To select a good representative of the entire highlight, the algorithm averages the intersection points of all highlight candidates, weighted by the number of pixels in the regions. It then selects the highlight region whose intersection point is closest to the average.

Since the surface reflection color of dielectrics is generally very similar to the illumination color, all highlight clusters are parallel to one another and to the illumination color vector. As a consequence, all dichromatic planes intersect along one line, which is the illumination vector. The algorithm uses this constraint to further reduce the error in estimating the orientations of the highlight clusters. It computes the average direction of all highlight clusters to combine all highlight hypotheses into a single hypothesis on the illumination color vector. The algorithm then generates planar hypotheses, combining the existing linear hypotheses with the direction of the illumination vector, as well as with information on the starting points of the respective highlight clusters. Such hypotheses describe the skewed-T shapes of the color clusters in terms of the positions and orientations of the matte and highlight clusters. The cross product of the illumination vector and the first eigenvector of a matte cluster determine the normal to the dichromatic plane of a color cluster. The position of the dichromatic plane in the color space is given by the color mean.

5.4.2. Exploiting Planar Hypotheses

Starting from the previously generated linear segmentation, the algorithm now resegments the image to account for the effects of surface reflection on objects. Applying the planar hypotheses one at a time, it proceeds iteratively until no more unprocessed planar hypotheses for large image areas exist.

The chosen planar hypothesis describes the position and orientation of a dichromatic plane in the color space, as well as the shape of the color cluster within the plane. To account for camera noise, the algorithm extends the plane into a slice, as shown in Figure 10.

Figure 10: A planar slice

When the algorithm uses the chosen planar slice to locally resegment the image, it starts, in principle, from the selected matte region and expands it until no more pixels at the region boundaries fall into the planar slice. This planar region growing method is augmented with special provisions to handle coplanar color clusters from neighboring objects. Such coplanar color clusters occur when the illumination vector lies in the plane spanned by the matte vectors of two neighboring objects. The color clusters of such objects lie in the same dichromatic plane and cannot be distinguished in a simple planar region growing approach. The resulting segmentation would generally be quite counterintuitive since the matte object colors of objects with coplanar clusters may be very different and even complementary. For example, a red body reflection vector $(255,0,0)$ and a white illumination vector $(255,255,255)$ span a dichromatic plane with normal $(0, -1/\sqrt{2}, 1/\sqrt{2})$. This plane contains all body reflection vectors (r,g,g) for which the g -component and the b -component are identical. In particular, the plane contains the vector $(0,255,255)$ which describes a cyan object color, as shown in Figure 11. In this case, a simple region growing method would not be able to distinguish a red object from a cyan object.

Figure 11: Coplanar color clusters

To avoid such segmentation problems, the algorithm exploits the previously gathered knowledge about existing matte color clusters. When the planar region growing process encounters pixels from a previously grown matte region other than the starting region, it only continues growing if the pixel lies within the matte color cylinder of the starting region. It thus applies the unrestricted planar growing criterion only to pixels that have not been previously recognized as matte pixels, while it falls back to the linear region growing method when matte pixels are concerned. This reflects the observation that if several matte areas exist in an object area, separated by highlight regions, all such matte areas form a single matte cluster. The algorithm also applies the proximity heuristic described in 5.3.2 to resolve ambiguities for color pixels at the intersection of dichromatic planes.

Figure 6 displays the results of segmenting the scene using the generated planar hypotheses. In comparison to the linear segmentation in Figure 5, the segmented image areas have grown into the highlight areas. As a result, the two matte image areas on the green donut - previously separated by a highlight - are now united. Due to camera limitations, not all pixels in the centers of the highlights are yet integrated into the object areas. This will be discussed and remedied in the next subsection.

Color Figure 6: Planar segmentation, scene with eight plastic objects

5.5. Accounting for Camera Limitations

Unfortunately, real images generally do not fully comply with the Dichromatic Reflection Model. Among other things, color clipping and blooming may significantly distort the color of pixels in and around highlights. As a result, the color pixels in the centers of highlights generally do not fall into the planar slice defined for the dichromatic plane of an object area. The planar segmentation excludes such pixels, as can be observed in Figure 6.

Since color information is so unreliable for these pixels, our algorithm does not use it. Instead, it uses a geometric heuristic to include distorted pixels into the region. Since pixels with distorted colors generally occur in the middle of the highlight areas, the algorithm starts from highlight pixels and expands the planar regions into areas that are next to highlight pixels and contain very bright pixels (i.e: brighter than the intersection point between the matte and highlight cluster). This heuristic successfully includes highlight centers with clipped colors into the object areas, while leaving other bright image areas without neighboring highlight pixels untouched.

Figure 7 displays the results of segmenting the scene while using the generated planar hypotheses and accounting for camera limitations. Nearly all pixels in the highlight centers have now been integrated into the segmented regions, and the image segments correspond quite well to the objects in the scene. A few pixels on the highlights have been excluded, due to the heuristic of only integrating very bright pixels. Any image processing method for filling holes in image segments, such as a region expansion operation followed by region shrinking [27], should be able to include these pixels.

Color Figure 7: Final segmentation, scene with eight plastic objects

5.6. Optical Effects Beyond the Scope of the Segmentation Algorithm

Figure 7 shows that some pixels at object boundaries have not been included into the object areas. These exclusions are related to effects in the scene that are not modeled in the Dichromatic Reflection Model, such as shadow casting and inter-reflection between objects.

- A small area in the lower right part of the yellow donut was not included into the large image segment covering the donut. The color of these pixels has been significantly altered by inter-reflection from the orange cup which reflected a part of its body color onto the donut, influencing the reflected body color of the yellow donut.
- A small area at the upper left edge of the yellow donut has wrongly been assigned to the red donut above it. In the original image, this area is covered by a shadow, resulting in very dark pixel values. Since the color clusters of the two donuts are already merged at these color values, the assignment was based on the distance from the two cylinder axes, which happened to be smaller for the red donut. Inter-reflection may also have influenced the color values, biasing them towards the red body reflection vector.
- There is a mirror image of the yellow cup on the side of the orange cup, resulting in yellow surface reflection. Such yellow surface reflection is added to the orange body reflection - which, in addition, has also been influenced by the light reflected from the yellow cup. As a consequence, the segmentation does not include these pixels into the region representing the orange cup.
- Similarly, inter-reflection between the orange and green cup causes a few pixels on the uppermost planar surface of the polyhedral part of the green cup to be excluded from the region representing the green cup.
- Finally, some yellow light from the yellow donut underneath the blue donut illuminates the lower side of the blue donut and influences the surface reflection components of the blue pixels. The pixel values at the lower boundary of the blue donut thus have a more yellowish color than the rest of the donut. This case of inter-reflection was too subtle to be detected by the algorithm under the chosen width of the planar slice. Figures 5 and 6 show that such yellow pixels have been separated from the blue ones in the linear segmentation step but were merged during the planar segmentation step because the yellow inter-reflection cluster was close to the dichromatic plane of the blue donut under white light.

There seems to be a good mapping between pixel areas in which our algorithm 'fails' and object areas with optical properties outside the scope of the model. Such results are quite different from 'traditional' segmentation results for which the algorithm's perform correctly on some instances of an optical phenomenon (e.g. highlights) but incorrectly on others (as shown in section 7.2.).

6. Separating Pixels into their Reflection Components

As one application of the above method to analyze and segment color images, we can now use the gathered information about the color clusters to split every color pixel into its two reflection components. We thus generate two intrinsic images of the scene, one showing the objects as if they were completely matte, and the other showing only the highlights.

We have previously reported a method to detect and remove highlights from hand-segmented images [21, 22]. That method projected the pixels from a selected image area into the color space and fitted a

skewed T to the entire color cluster, thus determining the body and surface reflection vectors of the area. Since our new segmentation algorithm already provides this information as a result of its analysis of local color variations, we can now skip this step. However, due to possible estimation errors in the segmentation process and due to camera problems such as blooming and chromatic aberration, the vectors may not yet fit perfectly. In order to obtain a more precise fit to the data, we retest every pixel in the segmented area and label it as a matte or highlight pixel, depending on whether it is closer to the matte line or to the highlight line. We then refit the matte and highlight line to the matte and highlight pixels by determining the first eigenvectors and the color means of the clusters.

The orientations of the corresponding reflection vectors are shown in Table 2. In principle, the length of the reflection vectors, \mathbf{C}_b and \mathbf{C}_s , depends on the brightness of the object such that a white and a grey reflection vector have the same orientation but different lengths. However, it is impossible to determine these lengths from the color clusters since the length of a cluster depends both on the magnitude of the reflection curve of the material and on the geometric influence from the photometric angles: a strongly curved, dark object may produce a color cluster of the same length as a brighter, but relatively flat object. For this reason, the vectors in Table 2, $\bar{\mathbf{C}}_b$ and $\bar{\mathbf{C}}_s$, are unit vectors, describing only the orientations of the reflection vectors. The table also provides the average surface reflection vector and an independent estimate of the illumination color which has been obtained by taking a color picture of a grey chart under the given illumination. The surface reflection vectors and their average vector are generally close illumination vector. This demonstrates that the surface reflection components of dielectric materials have the same color as the illumination and that the surface reflection vectors from real images can be used to determine the illumination color - information which is useful for color constancy algorithms [4, 23, 24].

Table 2: Body and surface reflection vectors of the eight plastic objects under white light

Our algorithm then uses $\bar{\mathbf{C}}_b$ and $\bar{\mathbf{C}}_s$ of every segmented region to split the pixels into their reflection components. Since the physically correct length of $\bar{\mathbf{C}}_b$ and $\bar{\mathbf{C}}_s$ is not known, the algorithm can determine the reflection components only up to a scale factor a that depends on the brightness of the material:

$$\mathbf{C}(x,y) = m_s \mathbf{C}_s + m_b \mathbf{C}_b = \tilde{m}_s \bar{\mathbf{C}}_s + \tilde{m}_b \bar{\mathbf{C}}_b \quad (13)$$

with

$$\begin{aligned} \tilde{m}_s &= m_s \cdot a, & \|\bar{\mathbf{C}}_s\| &= 1 = \frac{\|\mathbf{C}_s\|}{a} \\ \tilde{m}_b &= m_b \cdot a, & \|\bar{\mathbf{C}}_b\| &= 1 = \frac{\|\mathbf{C}_b\|}{a} \end{aligned}$$

To split the pixels into their reflection components, we use $\bar{\mathbf{C}}_b$, $\bar{\mathbf{C}}_s$ and their cross product, $\bar{\mathbf{C}}_b \times \bar{\mathbf{C}}_s$, to define a new (not necessarily orthogonal) coordinate system in the color cube. This coordinate system describes every color in the cube in terms of the amounts of body reflection \tilde{m}_b , surface reflection \tilde{m}_s and noise ε . There exists an affine transformation, and thus a linear transformation matrix T , which transforms any color vector $[R,G,B]^T$ into a vector $[\tilde{m}_b, \tilde{m}_s, \varepsilon]^T$. After computing T from $\bar{\mathbf{C}}_b$ and $\bar{\mathbf{C}}_s$, we can thus

transform every color pixel in the image area into its constituent body and surface reflection components, \tilde{m}_b and \tilde{m}_s . By selecting the \tilde{m}_b -components of all pixels, we generate the *body reflection image* of the region. By selecting the \tilde{m}_s -components, we generate the corresponding *surface reflection image*. The ϵ -components provide the *intrinsic noise image*.

Figure 12: Decomposing a color pixel into its constituent body and surface reflection components

We cannot apply this method immediately to bloomed and clipped pixels since their color is altered by the sensing process in the camera. In order to restore the physically correct color of such pixels, we exploit the observation that in many cases, clipping and blooming occurs only in one or two color bands. The pixels may thus have correct data in the other color bands. We assume that the smallest of the three values of a color pixel comes from a color band without clipping and blooming. We then replace the clipped or bloomed pixel by a pixel on the matte or highlight line that has the same value in the undistorted band (see [22] for a detailed description).

Figures 8 - 10 display the resulting intrinsic images of the scene with the eight plastic objects. The images demonstrate that we are able to determine the body and surface reflection components of the various objects reasonably well. When evaluated qualitatively, the body reflection image provides smooth shading across most highlight areas. We expect that this image may therefore be a useful tool to determine object shapes from shading information [14]. There exist thin dark rings in the body reflection image around the highlight on the green donut. This error is due to chromatic aberration in the camera [20] which currently limits the performance of our algorithm.

Color Figure 8: Intrinsic body reflection image of the scene with eight plastic objects

Color Figure 9: Intrinsic surface reflection image of the scene with eight plastic objects

Color Figure 10: Intrinsic noise image of the scene with eight plastic objects

The surface reflection image exhibits gradual changes between areas with no surface reflection and areas with very high amounts of surface reflection, demonstrating that surface reflection is a function of the photometric angles. The surface reflection images may be useful input for methods that determine object shapes from highlights [12]. A careful inspection of the surface reflection image reveals that the surface reflection component also increases at the material boundaries. This effect is related to aliasing occurring at pixels that integrate light from two neighboring objects. The colors of such pixels are a linear combination of two matte object colors. Depending on the orientations of the dichromatic planes, they may be included in either of the two object areas in the image, or they may be left unassigned. If they are included into one of the object areas, they generally do not lie close to the matte line, thus resulting in higher surface reflection components.

The intrinsic noise image displays the distance ϵ of every pixel in an image area from the associated dichromatic plane, scaled relative to the width of the planar slices and of the cylinders. Grey pixels represent color pixels that lie on the dichromatic plane. Increased noise values are displayed as

increasingly bright or dark pixels, depending on whether the color pixels lie on the front side or on the back side of the dichromatic plane. Very bright or dark areas in the noise image indicate places where the Dichromatic Reflection Model is not satisfactory in explaining color variation as a mere combination of body and surface reflection with added random camera noise. Such areas generally occur at material boundaries where light from two neighboring objects is integrated into the pixels values. The colors of such pixels are a linear combination of the two matte object colors of the neighboring objects. Higher noise values also occur in the centers of the highlights at clipped and bloomed pixels and at internal object edges where the surface orientation changes abruptly. Chromatic aberration causes increased noise values at the highlight on the green donut in the lower right corner [20]. Finally, large noise values exist at places of inter-reflection between objects, such as on the yellow donut, between the orange and yellow cup and on the right half of the green cup.

To illustrate the changing amounts of body and surface reflection more quantitatively, Figure 11 displays the reflection profiles for a row of pixels in the intrinsic body and surface reflection images of the plastic scene in Figure 1. It also shows the amounts of noise, ϵ , at these pixels. The selected row runs across the yellow donut, the orange cup and the yellow cup. The body reflection profile shows that the highlights have been removed reasonably well from the objects. Some error was introduced in the body reflection components of restored clipped and bloomed pixels. The surface reflection profile consists of a series of spikes at the positions of the highlights on the objects. Each spike has the form of a peak which falls off to the sides, demonstrating the quantitative nature of the highlight analysis. A few small peaks exist at the object boundaries. These are pixels on material boundaries where light from more than one object is mixed into a single pixels value. Since such pixels generally do not lie close to the matte line, they generate a disproportionately high surface reflection component.

Color Figure 11: Profiles of the intrinsic reflection images along row 226, plastic scene

7. Discussion

The previous sections have presented an approach to color image understanding that exploits the Dichromatic Reflection Model. The description of the algorithms was accompanied by a few exemplary results that demonstrated how the algorithm performs on real images. This section now presents more results and discusses strengths, limitations and extensions of the approach.

7.1. Further Results

We have tested our algorithm on a series of color images of scenes with dielectric objects under varying illumination conditions. The upper left quarters of Figures 12 - 14 show the images of three plastic cups under yellow, white and pink light.

Color Figure 12: Color image segmentation and reflection analysis, scene with three plastic cups under yellow light

Color Figure 13: Color image segmentation and reflection analysis, scene with three plastic cups under white light

Color Figure 14: Color image segmentation and reflection analysis, scene with three plastic cups under pink light

The upper right quarters of the figures display the segmentations that were generated by our algorithm. They demonstrate that our segmentation method performs well on a variety of dielectric materials under varying illumination conditions. In all images, the segmentations outline the material boundaries of the objects very well. Because the algorithm models matte and highlight color variations as physical reflection processes, it ignores color changes along highlight boundaries and it also tolerates shading changes on the objects.

The lower quarters of Figures 12 - 14 show the intrinsic reflection images that were generated using the body and surface reflection vectors of Tables 3 - 5. The intrinsic images demonstrate that our algorithm detects and removes highlights well on a variety of dielectric materials under varying illumination colors. When evaluated qualitatively, the body reflection vectors have reasonable orientations and the body reflection images generally describe smooth shading across the highlight area. The surface reflection images in the lower right quarters show that the highlights from the original images have been detected well. The surface reflection images exhibit gradual changes between areas with no surface reflection and areas with very high amounts of surface reflection, demonstrating that surface reflection is a function of the photometric angles. The surface reflection vectors in Tables 3 - 5 are generally reasonably close to the estimated illumination colors. The reflection images may be a useful tool to determine object shapes from shading or from highlights while the surface reflection vectors can be used in color constancy algorithms as an indication of the illumination color.

Table 3: Body and surface reflection vectors of the three plastic cups under yellow light

Table 4: Body and surface reflection vectors of the three plastic cups under white light

Table 5: Body and surface reflection vectors of the three plastic cups under pink light

7.2. Comparison with a Traditional Color Image Segmentation Method

Figure 15 shows the results of applying a traditional color segmentation method to Figure 1. It has been obtained by using Phoenix [31], a segmentation program that recursively splits color images into smaller regions until the regions are homogeneous. To decide how to split regions, Phoenix uses a set of user-defined color features, such as R, G, B, intensity, hue, saturation, etc., each of which is encoded as a separate image band. When a region is considered for further splitting, Phoenix generates for every feature a histogram of all pixels in the region and looks for distinctive valleys in the histograms. It then splits the region according to the feature with the most prominent valley. If it cannot find a significant valley in any feature histogram, it decides that the region is homogeneous. Phoenix is one of a class of 'Ohlander-style' color image segmentation methods which all recursively split the image into segments by seeking for valleys in histograms of some color features. Extensions to the original method [25] centered around determining the color feature axes dynamically through principal component analysis [6, 26] rather

than by user input. We compare our results with Phoenix results since this system is available to us.

Color Figure 15: Region segmentation by Phoenix,
scene with eight plastic objects

Figure 15 was generated by running Phoenix on three features: intensity, hue and saturation (IHS). The IHS color features are often used to describe human color perception. We chose this color basis, because our sample scene is composed of objects with relatively few, quite distinct hues, and a comparison on this basis seemed more favorable to Phoenix than a comparison on the RGB-basis. Some of the highlights in Figure 15 are separated from the surrounding matte object parts while other highlights are integrated with the matte areas. In addition, the matte areas are sometimes split into dark and bright areas, as can be seen on the right half of the green cup, as well as on the dark red donut and on the green donut. It is not easy to predict how Phoenix will perform on a given image because its decisions are not directly related to the influences of physical processes. Phoenix uses a pre-defined, fixed set of color features. It measures color variation only in relationship to these features and segments image areas using only one feature at a time. Projecting the color information from a three-dimensional color space to a one-dimensional feature space, Phoenix loses most of the information about the three-dimensional structure of the color variation.

In comparison to Phoenix, our color cylinders and dichromatic planes capture color variations in a much more flexible way. The color cylinders and planes of the linear and planar hypotheses are determined by analyzing color clusters in their three-dimensional structure. Their position and orientation is independent of a chosen coordinate system in the color space, as well as of a user-defined set of features. Consequently, linear and planar hypotheses can be expected to capture color variation on objects much better and to generate better segmentations of the objects, possibly even under ambient light [34]. This is evident when Figure 15 is compared with Figures 5, 6, and 7. Since the linear, planar and final segmentations account for shading variation on matte object areas, for highlight reflection and finally for camera limitations, they include all matte and highlight pixels, as well as bloomed and clipped pixels into a single object area.

7.3. Control Parameters

We use a few control parameters and simplifying heuristics when we analyze color images. They are related to camera limitations, as well as to the limitations of color information and to the basic assumptions of the Dichromatic Reflection Model. This subsection shows the influence of the parameter-settings on the performance of our algorithm while the following subsections discuss the influence of the heuristics and the principle limitations of the dichromatic theory.

The control parameters, as well as the parameter values that were chosen to generate the results above, are listed in Table 6. It seems that most of the parameters are related in quite an intuitive way to the camera limitations and to the reflection model and that they are thus relatively easy to set.

Table 6: Control parameters of the algorithm

- **Clipping Threshold and Noise Density:**

The clipping threshold and the noise density parameter are related to the limited dynamic range of cameras. The density parameter expresses a heuristic to discard bloomed pixels (see section 4 and [22]).

- **Camera Noise:**

σ_0 describes the estimated camera noise of the camera. We estimate it by repeatedly taking a black-and-white picture of a white board, measuring the intensity of the center pixel and then computing its statistics. σ_0 corresponds to twice the value of the standard deviation σ of these measurements such that the integral of the Gaussian curve in the range $[\mu - \sigma_0, \mu + \sigma_0]$ describes about 96% of all measurements.

- **Minimal Intensity:**

To avoid conflicts in dark image areas where matte clusters overlap, dark pixels are excluded from the color analysis by setting a threshold for minimal intensity. If this threshold is too low, color clusters from neighboring objects overlap beyond the threshold and bleeding may occur between the image areas. If, on the other hand, the threshold is unnecessarily high, color clusters of dark objects may have lost most of their matte cluster such that their body reflection vectors can no longer be estimated reliably.

- **Cylinder Width:**

Due to inter-reflection and other optical effects in real scenes for which the Dichromatic Reflection Model does not account, the matte color clusters generally are not perfectly cylindrical. For such reasons and also because of possible estimation errors in the direction of the color vector, our algorithm needs to use fairly thick cylinders and planar slices to include most of the pixels on an object into the segmented area. The drawback of unnecessarily thick cylinders and planar slices is a loss of sensitivity to the color properties of the image.

- **Initial Window Size:**

The algorithm also depends on the window size in the initialization step of the algorithm. If the windows are too large, many of them overlap several objects, and many local color clusters are classified to be volumetric. If, on the other hand, the windows are too small, color variation on a relatively flat or dark object cannot be detected and the color clusters of such windows are classified as pointlike clusters. The initialization step then does not provide sufficient information to the subsequent steps of generating and applying linear or planar hypotheses.

- **Minimal Area Size:**

In order to first concentrate on matte object areas - which are easier to analyze - our algorithm currently inspects only color clusters of image areas beyond a minimal number of pixels. It assumes that linear color clusters of such image areas represent matte branches of skewed T's and it subsequently tries to find the related highlight branches. In principle, however, linear color clusters can also exist in highlight areas or at material boundaries. The size of such highlight areas with linear color properties depends on the surface curvature of the object. Thus, if the selected threshold for the minimal area size is too low, highlight areas of flat objects are selected as candidates for matte branches of skewed T's. If, on the other hand, the threshold is too high, small objects in the scene cannot be detected.

Table 6 shows that most of the parameter settings were fairly constant for all images. Slight adjustments had to be made in a few cases. The changes were triggered by the changing illumination color in the scene and the resulting change in contrast in the color images. A detailed case study showing color pictures of different parameter settings is given in [20]. It may be possible to automatically select - or even to locally adapt - the values of some of the control parameters, such as the window size, the cylinder width and the area size. We expect that such learning mechanisms fit well into the generate-and-test control structure of our algorithm and into our strategy of starting with the simplest image areas

and then progressing into areas where a more complex (or less conservative) analysis is required.

7.4. Simplifying Heuristics

Our algorithm relies on a few simplifying heuristics. They are all related to the problem of determining the structure of a color cluster. Our algorithm currently works with a built-in, very rigid model of what color cluster shapes to expect, assuming that clusters always consist of two linear subclusters that form a skewed T. We now discuss these heuristics.

- **Large Linear Clusters are Matte Clusters:**

As discussed above, the algorithm assumes that linear hypotheses from large image areas describe matte pixels on an object. This heuristic depends on the curvature on the objects in the scene, as well as on their distance from the camera.

- **Color Clusters have Only One Highlight Branch:**

When the algorithm selects a surface reflection vector from a list of highlight candidates, it implicitly assumes that the clusters of all highlight candidates are part of the same highlight cluster. This may not be the case for convex objects. The two planar faces in Figure 13 each reflect a highlight into the camera. The incidence angle between the illumination ray and the surface normal is larger at point P_1 than at point P_2 . Accordingly, the body reflection component under the highlight at P_1 is smaller than the body reflection component under the highlight at P_2 and the two highlight generate two separate highlight clusters in the dichromatic plane.

- **50%-Heuristic:**

The 50%-heuristic also makes limiting assumptions on the scene. It has only been evaluated so far for spherical objects that were illuminated and viewed from the same distance. Different conditions may exist on cylindrical, elliptical or planar objects. The heuristic does not apply to convex objects since, on such objects, the brightest visible matte point may lie on one surface while the highlight is on another surface with different illumination conditions. The highlight may then start at an arbitrary place on the matte line.

Figure 13: Cluster shapes for concave objects

7.5. Limitations of the Dichromatic Theory and Extensions

The Dichromatic Theory imposes some limitations on our approach. Since we attribute any linear color variation to the changing illumination geometry of a single material, we are unable to find material boundaries between objects with collinear matte clusters. We will need a geometrical analysis, linking intensity gradients to object shape, to distinguish between such objects. The same will be needed to analyze dark image areas which are currently excluded because their color information is too unreliable.

The model also makes simplifying assumptions about the illumination conditions and the materials in the scene. A color cluster from an object in an unconstrained scene will generally not be a skewed T composed of linear subclusters because the illumination color may vary on different parts on the object surface, and the reflection properties of the object may also change, due to illumination changes and to pigment variations in the material body. The necessary extensions to the model will be the subject of future work.

Furthermore, our method to split color pixels into their reflection components (but not the segmentation) relies on a characteristic color change between the matte object color and the highlight color. There

needs to be a certain angle between the orientations of the body and surface reflection vectors of an object. How big the angle needs to be depends on the cylinder width. If the matte and highlight clusters are approximately collinear, we cannot separate the reflection components. This is the case, if an object has the same color as the light source or if the object is a uniform reflector, such as grey objects or pale pastels. We also have problems when one of the two linear clusters does not exist or is very small for an object. The matte cluster is missing, if the viewed object is very dark, or if the scene is illuminated with a narrow-band illuminant that does not overlap with the wavelengths at which the material reflects light. Matte clusters also do not exist for metallic objects. On the other hand, the highlight cluster may be missing, if an object does not reflect a highlight into the camera, due to its position in the scene and the illumination geometry. As a third case, we need to consider objects with very rough surfaces such that every pixel in the image area has both a significant body and surface reflection component. The color cluster may then fill out the entire dichromatic plane. A common special case of this are so-called "matte" or "Lambertian" materials - as opposed to glossy materials - which reflect a constant amount of surface reflection in every direction and thus never exhibit a highlight in the common sense of the word. The corresponding color clusters are linear clusters, translated from the origin of the color space according to the constant surface reflection component. Our current method is not capable of distinguishing between all these cases. In combination with exploiting previously determined scene properties, such as the illumination color, we will need to analyze the intensity gradients along the linear axes and relate them to the properties of m_s and m_b , as described in a geometrical model of light reflection.

The Dichromatic Reflection Model also makes several simplifying assumptions which restrict the illumination conditions and the materials in the scene. Because of these assumptions the model can currently only be used to understand scenes taken under very restricted conditions. The results of using the reflection model in analyzing real images, such as the image in Figure 1, have shown that even images of carefully selected scenes often contain areas in which the model does not apply (see Figure 10). To provide a broader applicability of the method, the model needs to be extended to account for more physical processes, especially for more general illumination conditions. Inter-reflection and shadow-casting between objects exist in nearly every moderately complicated scene. Furthermore, many realistic scenes will contain more than a single, point-like light source. There will generally be some amount of ambient light and possibly even several light sources that shed differently colored light onto the scene. A first step in the direction of modeling more complicated illumination conditions has been suggested by Shafer and others [3, 30], adding a constant color term to the reflection model to account for ambient light. Tong and Funt have presented promising first results on analyzing real images with ambient light [34]. Their results indicate that the extended Dichromatic Model is appropriate for describing color variations under direct and ambient light. We expect our segmentation algorithm to be applicable to images taken under such illumination conditions. Since ambient light adds a constant amount of light to each pixel in an object area, it translates the entire color cluster away from the origin of the color space - without any change to the shape of the cluster. The principal component analysis captures the effect of translational color transformations in the mean color value of the cluster. Our segmentation algorithm uses both the color mean and the principal directions for positioning color

cylinders and color planes in the cube and thus should work as well under ambient light as it does without. However, the intrinsic reflection images are now presenting more complicated body or interface reflection processes since they are influenced by two types of illumination. The ambient light (a constant) may have to be subtracted out before a shape-from-shading or a shape-from-highlights method can be applied.

8. Conclusions

In this paper, we have demonstrated that it is possible to analyze and segment real color images by using a physics-based color reflection model. Our model accounts for highlight reflection and matte shading, as well as for some characteristics of cameras. By developing a physical description of color variation in color images, we have developed a method to automatically segment an image while generating hypotheses about the scene. We then use the knowledge we have gained to separate highlight reflection from matte object reflection. The resulting intrinsic reflection images have a simpler relationship to the illumination geometry than the original image and may thus improve the results of many other computer vision algorithms, such as motion analysis, stereo vision, and shape from shading or highlights [5, 12, 14, 33]. Since the surface reflection component of dielectric materials generally has the same color as the illumination, we can also determine the illumination color from the intrinsic surface reflection image, information which is needed by color constancy algorithms [4, 13, 23, 24].

The key points leading to the success of this work are our modeling of highlights as a linear combination of both body and surface reflection and our modeling of the camera properties. With few exceptions [4, 6, 23, 28], previous work on image segmentation and highlight detection has assumed that the color of highlight pixels is completely unrelated to the object color. This assumption would result in two unconnected clusters in the color space: one line or ellipsoid representing the object color and one point or sphere representing the highlight color. Our model and our color histograms demonstrate that, in real scenes, a transition area exists on the objects from purely matte areas to the spot that is generally considered to be the highlight. This transition area determines the characteristic shapes of the color clusters which is the information that we use to distinguish highlight boundaries from material boundaries and to detect and remove highlights. This view of highlights should open the way for quantitative shape-from-gloss analysis, as opposed to binary methods based on thresholding intensity.

By modeling the camera properties, we are able to obtain high quality color images (through color balancing and spectral linearization) in which most pixels maintain the linear properties of light reflection, as described in the Dichromatic Reflection Model. We can also detect most distorted color pixels in an image and thus generate an intrinsic error image which then guides our algorithm to separate only undistorted color pixels into their reflection components. We expect that the intrinsic error image will be similarly useful in guiding other computer vision algorithms, such as shape from shading. It may also enable us to automatically control the camera aperture so that we can obtain color images with minimal clipping and blooming.

Our hypothesis-based approach towards image analysis may provide a new paradigm for low-level

image understanding. Our method gains its strength from using an intrinsic model of physical processes that occur in the scene. The result are intrinsic images and hypotheses which are closely related in their interpretation to the intrinsic model, being instantiations of concepts formulated in the model. Our system alternates between a bottom-up step which generates hypotheses and a top-down step which applies the hypotheses to the images. Our analysis thus consists of many small, complete interpretation cycles that combine bottom-up processing with feed-back in top-down processing. This approach stands in contrast to traditional image segmentation methods which do not relate their analysis to intrinsic models and that also generally have a strictly bottom-up control structure. We feel that many low-level image understanding methods such as shape-from-x methods, stereo and motion analysis may be viewed and approached under this paradigm. We hope to extend our approach into a more complete low-level image analysis system which combines color analysis with a geometrical analysis of the scene, exploiting the body and surface reflection images. Along these lines, we may generate hypotheses about object shapes and about the object materials (see [13]). The highlight image may also provide strong evidence for the position of the light source.

Although the current method has only been applied in a laboratory setting, its success shows the value of modeling the physical nature of the visual environment. Our work and the work of others in this area may lead to methods that will free computer vision from its current dependence on statistical signal-based methods for low-level image analysis.

Footnotes

Affiliation of authors:

Computer Science Department
Carnegie Mellon University
Pittsburgh, PA 15213

Gudrun Klinker's current address:

Cambridge Research Laboratory, Digital Equipment Corporation, One Kendall Square, Bldg. 700,
Cambridge, MA 02139

Acknowledgement of financial support:

This research was supported by the Jet Propulsion Laboratory, California Institute of Technology, under NASA contract 957989, by the National Science Foundation under Grant DCR-8419990 and by the Defense Advanced Research Projects Agency (DOD), ARPA Order No. 4976, monitored by the Air Force Avionics Laboratory (Air Force Wright Aeronautical Laboratories, Aeronautical Systems Division (AFSC), United States Air Force, Wright-Patterson AFB, OHIO 45433-6543) under contract F33615-87-C-1499 (Arpa Order # 4976, Amendment 20). Any opinions, findings, and conclusions or recommendations expressed in this publication are those of the author and do not necessarily reflect the views of the Jet Propulsion Laboratory, the National Science Foundation, the Defense Advanced Research Projects Agency, or the US Government.

References

- [1] D.H. Ballard, and C.M. Brown.
Computer Vision.
Prentice Hall, Inc., Englewood Cliffs, NJ 07632, 1982.
- [2] H.G. Barrow, and J.M Tenenbaum.
Recovering Intrinsic Scene Characteristics from Images.
Computer Vision Systems.
Academic Press, New York, 1978, pages 3-26.
- [3] R.L. Cook and K.E. Torrance.
A Reflectance Model for Computer Graphics.
ACM Transactions on Graphics 1(1):7-24, January, 1982.
Also published in *Computer Graphics* 15(3), SIGGRAPH'81.
- [4] M. D'Zmura and P. Lennie.
Mechanisms of color constancy.
Journal of the Optical Society of America A (JOSA-A) 3(10):1662-1672, October, 1986.
- [5] L. Dreschler and H.-H. Nagel.
Volumetric Model and 3D Trajectory of a Moving Car Derived from Monocular TV Frame Sequences of a Street Scene.
Computer Graphics and Image Processing 20:199-228, 1982.
- [6] R. Gershon.
The Use of Color in Computational Vision.
PhD thesis, Department of Computer Science, University of Toronto, 1987.
- [7] R. Gershon, A.D. Jepson, and J.K. Tsotsos.
Highlight Identification Using Chromatic Information.
In *Proceedings of the first International Conference on Computer Vision (ICCV)*, pages 161-171.
IEEE, London, June, 1987.
- [8] H. Grassmann.
On the Theory of Compound Colors.
Phil. Mag. , April, 1854.
- [9] A.R. Hanson and E.M. Riseman.
VISIONS: A Computer System for Interpreting Scenes.
In A.R. Hanson and E.M. Riseman (editor), *Computer Vision Systems*, pages 303-333. Academic Press, New York, 1978.
- [10] H.H. Harman.
Modern Factor Analysis, second edition.
University of Chicago Press, Chicago and London, 1967.
- [11] G. Healey and T.O. Binford.
Local Shape from Specularity.
In L.S. Bauman (editor), *DARPA-Image Understanding (IUS) workshop*, pages 874-887. Los Angeles, CA, February, 1987.
- [12] G. Healey and T.O. Binford.
Local Shape from Specularity.
In *Proceedings of the First International Conference on Computer Vision (ICCV)*, pages 151-161.
IEEE, London, June, 1987.

- [13] G. Healey and T.O. Binford.
The Role and Use of Color in a General Vision System.
In L.S. Bauman (editor), *DARPA-Image Understanding (IUS) workshop*, pages 599-613. Los Angeles, CA, February, 1987.
- [14] B.K.P. Horn.
Understanding Image Intensities.
Artificial Intelligence 8(11):201-231, 1977.
- [15] B.K.P. Horn.
Exact Reproduction of Colored Images.
Computer Vision, Graphics, and Image Processing (CVGIP) 26:135-167, 1984.
- [16] B.K.P. Horn and B.G. Schunk.
Determining Optical Flow.
Artificial Intelligence 17:185-203, 1981.
- [17] T. Kanade.
Region Segmentation: Signal vs. Semantics.
In *Proceedings of the 4th International Joint Conference in Pattern Recognition (IJCPR)*, pages 95-105. IEEE, Kyoto, Japan, November, 1978.
- [18] J.R. Kender.
Shape from Texture.
PhD thesis, Computer Science Department, Carnegie-Mellon University, November, 1980.
Appeared also as technical report CMU-CS-81-102, Computer Science Department, Carnegie-Mellon University, Pittsburgh PA.
- [19] J.R. Kender, and E.M. Smith.
Shape from Darkness: Deriving Surface Information from Dynamic Shadows.
In *Proceedings of the First International Conference on Computer Vision (ICCV)*, pages 539-546. IEEE, London, June, 1987.
- [20] G.J. Klinker.
A Physical Approach to Color Image Understanding.
PhD thesis, Computer Science Department, Carnegie-Mellon University, May, 1988.
Available as technical report CMU-CS-88-161.
- [21] G.J. Klinker, S.A. Shafer, and T. Kanade.
Using a Color Reflection Model to Separate Highlights from Object Color.
In *Proceedings of the First International Conference on Computer Vision (ICCV)*, pages 145-150. IEEE, London, June, 1987.
- [22] G.J. Klinker, S.A. Shafer, and T. Kanade.
The Measurement of Highlights in Color Images.
International Journal on Computer Vision (IJCV) 2(1):7-32, 1988.
- [23] H.-C. Lee.
Method for computing the scene-illuminant chromaticity from specular highlights.
Journal of the Optical Society of America A (JOSA-A) 3(10):1694-1699, October, 1986.
- [24] L.T. Maloney and B.A. Wandell.
Color constancy: a method for recovering surface spectral reflectance.
Journal of the Optical Society of America A (JOSA-A) 3(1):29-33, January, 1986.
- [25] R. Ohlander, K. Price, and D.R. Reddy.
Picture Segmentation Using a Recursive Region Splitting Method.
Computer Graphics and Image Processing 8:313-333, 1978.

- [26] Y. Ohta, T. Kanade, and T. Sakai.
Color Information for Region Segmentation.
Computer Graphics and Image Processing 13:222-231, 1980.
- [27] A. Rosenfeld and A.C. Kak.
Digital Picture Processing.
Academic Press, New York, 1982.
2. edition.
- [28] J.M. Rubin and W.A. Richards.
Color Vision and Image Intensities: When are Changes Material ?
Biological Cybernetics 45:215-226, 1982.
- [29] S.A. Shafer.
Describing Light Mixtures Through Linear Algebra.
Journal of the Optical Society of America (JOSA) 72(2):299-300, February, 1982.
- [30] S.A. Shafer.
Using Color to Separate Reflection Components.
COLOR research and application 10(4):210-218, Winter, 1985.
Also available as technical report TR-136, Computer Science Department, University of Rochester, NY, April 1984.
- [31] S.A. Shafer and T. Kanade.
Recursive Region Segmentation by Analysis of Histograms.
In *Proceedings of the International Conference on Acoustics, Speech, and Signal Processing*,
pages 1166-1171. IEEE, Paris, France, May, 1982.
- [32] S.A. Shafer and T. Kanade.
Using Shadows in Finding Surface Orientations.
Computer Vision, Graphics, and Image Processing (CVGIP) 22:145-176, 1983.
- [33] C.E. Thorpe.
FIDO: Vision and Navigation for a Robot Rover.
PhD thesis, Computer Science Department, Carnegie-Mellon University, December, 1984.
Available as technical report CMU-CS-84-168.
- [34] F. Tong and B.V. Funt.
Specularity Removal for Shape from Shading.
In *proc. of the conference Vision Interface, Edmonton, Alberta, Canada*. 1988.
- [35] S. Ullman.
The Interpretation of Structure From Motion.
AI Memo 476, MIT AI Laboratory, Cambridge, Mass., October, 1976.

Table of Contents

1. Introduction	1
2. The Dichromatic Reflection Model	2
3. Object Shape and Spectral Variation	2
3.1. Basic Principles	2
3.2. Relationship between Illumination Geometry and Spectral Histogram Shape	3
4. A Camera Model	6
5. Color Image Segmentation	7
5.1. Using the Dichromatic Reflection Model in Color Image Understanding	7
5.2. Generating Initial Estimates for Color Clusters	9
5.2.1. A Classification Scheme for Local Color Variation	9
5.2.2. Determining Compatible Color Variation	11
5.3. Linear Hypotheses	11
5.3.1. Generating Linear Hypotheses	12
5.3.2. Exploiting Linear Hypotheses	12
5.4. Planar Hypotheses	13
5.4.1. Generating Planar Hypotheses	13
5.4.2. Exploiting Planar Hypotheses	15
5.5. Accounting for Camera Limitations	16
5.6. Optical Effects Beyond the Scope of the Segmentation Algorithm	17
6. Separating Pixels into their Reflection Components	17
7. Discussion	20
7.1. Further Results	20
7.2. Comparison with a Traditional Color Image Segmentation Method	21
7.3. Control Parameters	22
7.4. Simplifying Heuristics	24
7.5. Limitations of the Dichromatic Theory and Extensions	24
8. Conclusions	26

List of Figures

Figure 1: Light reflection of dielectric materials	2
Figure 2: The shape of the spectral cluster for a cylindrical object	3
Figure 3: Assumed illumination geometry	4
Figure 4: Photometric angles at highlight	4
Figure 5: Brightest matte point that is visible from the camera	5
Figure 6: Starting point q of the highlight line as a function of phase angle g and distance d	6
Figure 7: Using an Intrinsic Model for Image Understanding	8
Figure 8: Proximity heuristic to resolve a color conflict between the color clusters of two neighboring matte object areas	13
Figure 9: Finding the brightest matte pixel	14
Figure 10: A planar slice	15
Figure 11: Coplanar color clusters	16
Figure 12: Decomposing a color pixel into its constituent body and surface reflection components	19
Figure 13: Cluster shapes for concave objects	24

List of Tables

Table 1: Summary of Color Cluster Interpretations	10
Table 2: Body and surface reflection vectors of the eight plastic objects under white light	18
Table 3: Body and surface reflection vectors of the three plastic cups under yellow light	21
Table 4: Body and surface reflection vectors of the three plastic cups under white light	21
Table 5: Body and surface reflection vectors of the three plastic cups under pink light	21
Table 6: Control parameters of the algorithm	22