

IMAGE UNDERSTANDING 1985-86

Edited by

Whitman Richards

Department of Brain and Cognitive Sciences

and

Shimon Ullman

Artificial Intelligence Laboratory

Massachusetts Institute of Technology

Ablex Publishing Corporation
Norwood, New Jersey 07648

chapter 2

Spectral categorization of materials*

**John M. Rubin
W.A. Richards**

Massachusetts Institute of Technology
Natural Computation Group
Cambridge, MA 02139

1.0 INTRODUCTION

The human visual system performs a remarkable feat. The pattern of light that reaches the eye from a scene is the result of a complex interaction among several factors: the quality of the illuminant, the geometry of the scene, and the properties of the materials composing the visible surfaces. Yet somehow these confounded factors are mostly separated in our perception. We see particular spatial arrangements of objects. These objects appear bounded by surfaces having properties—color and texture—roughly invariant over a range of conditions of geometry and illumination. To compute invariant descriptions of the material properties of surfaces is an important goal of any visual system. Such material descriptors are useful for object recognition and visual search.

It's commonplace to assume color vision has something to do with capturing the albedoes of surface materials.¹ But exactly what aspect of the albedo function would serve a visual system best? Consider the grandiose goal of recovering a material's albedo as a continuous function of wavelength. Not

* This chapter is a revised and abridged version of MIT AI Memo 764, entitled "Color vision: representing material categories" (Rubin & Richards, 1984). The research was done at the Department of Psychology and the Artificial Intelligence Laboratory of the Massachusetts Institute of Technology, and was supported by NSF and AFOSR under a combined grant for studies in Natural Computation, grant 79-23110-MCS, and by the AFOSR under an Image Understanding contract F49620-83-C-0135. John Rubin was supported by an NSF Graduate Fellowship, and by a pre-doctoral fellowship from the M.I.T. Center for Cognitive Science. The authors would like to thank T. Poggio, Nancy Kanwisher, Andrew Knapp, and the members of the Natural Computation group for their comments, and Bill Gilson for a meticulous reading of the manuscript.

¹ The albedo of a material is a function of wavelength $\rho(\lambda)$, with range (0, 1), that indicates what fraction of photons (emitted by some light source) at each wavelength will be reflected.

only is this goal impractical; it is counter to the aim of finding invariant descriptors. With such an over-zealous representation, unimportant variations in a surface would prevent its being recognized as a single region, a patch of one kind of stuff. The perception of the world would be shattered with spectral acuity too fine; one literally wouldn't be able to see the forest for the trees.

Here we seek a representation of material reflectance in which trivial surface variations can be overlooked in order to appreciate important similarities.² At the same time, the representation must allow some discrimination among different materials. Below we develop such a categorical color space, based on a theoretical solution to the problem of identifying material changes. A trichromatic system, it will be shown, yields a two-dimensional color space in which the axes will turn out to represent boundaries between different materials. The four quadrants of the two-dimensional space represent material categories.

2.0 SPECTRAL INFORMATION AT EDGES

When two image regions arise from different materials in the scene, the transition from one material to another will usually bring about an edge in the image. Thus we restrict our search for material changes to edges. How can we decide whether an edge is due to a material change?

An edge in the image will usually arise from a single event or state of affairs in the three-dimensional scene (Marr, 1982). The most common edge types are shadows, highlights, surface orientation discontinuities, and pigment density changes.³ Alternatively, an edge may be due to a material change, a discontinuity between two different kinds of stuff.⁴ How can a material change edge be distinguished from other types of edges? Rubin and Richards (1982) attempted to answer this question. Edges which arise from shadows, orientation changes and highlights are *lawful* in the sense that there are equations that describe how image intensities will change across these edges. By contrast, material changes are completely unpredictable; they are arbitrary changes, and as such, can only be inferred by ruling out, at a given edge, the possibility of any of the above lawful changes.

² We are not suggesting any spectral information be thrown away. We are merely exploring a single problem. Other problems may require detailed spectral information.

³ Surface orientation change and shadow can coincide at an edge, but this exception is unimportant to the arguments that follow. See Rubin and Richards (1982), footnote 16.

⁴ We consider materials to consist of some spectrally neutral embedding material (e.g., cellulose) impregnated with a single pigment (e.g., chlorophyll). A material change is a change in pigment type, or a change in both pigment and embedding material.

To infer material changes, we now face the awkward prospect of having to reject, one by one, each of the lawful changes. A method of rejecting all of those edges *en masse* would be desirable. Fortunately, there is a simple ordinal rule common to all the edges formed by lawful processes: if the intensity at one wavelength decreases across a lawful edge (shadows, highlights, and so on) then the intensity must also decrease at all other wavelengths taken across the same edge (Rubin & Richards, 1982). When the condition is violated, we say there is a "spectral crosspoint" across the edge. Spectral crosspoints imply material changes; a spectral crosspoint is illustrated in Figure 2a. The spectral crosspoint is not the only means of discovering material changes, however. We will show that a second and independent condition holds for each of the lawful processes—namely the preservation of ordinality of image intensity across wavelength. A violation of this condition implies a material change.

3.0 THE OPPOSITE SLOPE SIGN INFERENCE

3.1. The Lawful Processes

Figure 1a shows two image intensity graphs of the same shape. Intuitively, the two graphs, of similar shape, arise from measurements taken on either side of a "lawful" edge type. Figure 1b shows two graphs of different shape. None of the lawful edge types could have produced such a distortion, and intuitively it seems that a material change edge is the best explanation. We now must make explicit what we mean by "same shape" and then show that this definition of spectral shape remains invariant across edges created by shadows, changes in surface orientation, highlights or variations in pigment density—namely the lawful conditions we wish to reject as material changes.

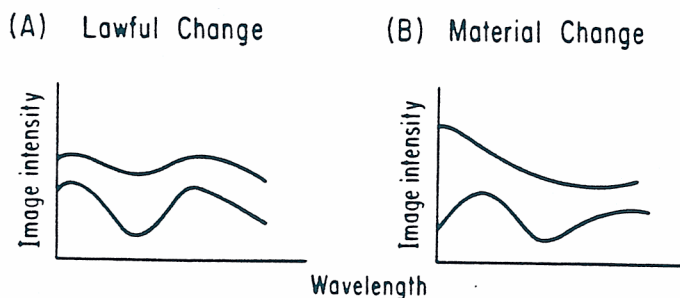


Figure 1. Graphs of image intensity versus wavelength. Each curve represents the image intensity measurable from one image region. A) Two graphs of same shape: a likely lawful change. B) Two graphs of different shape: a candidate for material change.

to
of
le
n-
s,
en
is
il
-
J
1

Definition: Two curves of intensity versus wavelength have the *same shape* if the ordinal relations of image intensity across wavelength are preserved.

More formally, if $I_X(\lambda)$ and $I_Y(\lambda)$ are image intensities as functions of wavelength measured on both sides, X and Y , of an edge, then $I_X(\lambda)$ and $I_Y(\lambda)$ have identical ordinality if, for all λ_1 and λ_2 , $I_X(\lambda_1) < I_X(\lambda_2)$ iff $I_Y(\lambda_1) < I_Y(\lambda_2)$. Note that two image intensity functions of identical ordinality will have local extrema at the same values of wavelength.

Given this ordinal definition of "same shape," Appendix I shows that the ordinality relationship is preserved across all edges arising from the lawful edge types, provided that the following two conditions hold:

Gray world condition: The average of all the different albedoes in the scene will be a spectrally flat "gray," so that the diffuse reflected light will have the same spectral character as the direct light.

Spectral normalization: The spectral samples of image intensity have been normalized with respect to the spectral content of the illuminant.

Spectral normalization is a transformation of spectral sample values to what they would have been under white illumination. White light has the same photon flux at all wavelengths. A scheme for spectral normalization will be presented in Section 4.

3.2. The Opposite Slope Sign Operator

We now can proceed to test for "same shape" using the ordinality relation. If ordinality is violated across an edge, then we infer the edge does not arise from one of the "lawful" processes and hence must represent a material change (provided also, of course, that our grey world condition is not violated).⁵

What is the simplest way to seek violations of ordinality? A pair of spectral samples suffices. Let the image intensities on both sides of an edge be measured at wavelengths λ_1 and λ_2 . If image intensity at λ_1 is greater than that at λ_2 on one side of the edge, then the ordinality condition requires the same relationship hold on the other side. So if the two sides of the edge do not have greater intensity in the same spectral sample, ordinality is violated; the edge cannot be lawful. (Details are given in Appendix I.) This condition is called the *opposite slope sign condition*.⁶ Examples are shown in Figure 2a and 2b.

⁵ It is possible when the grey world assumption is wrong, material changes will be inferred from images. This is not entirely bad news; if human perception also goes awry when the grey world assumption is violated, then our theory will gain support as an account of biological visual systems.

⁶ The opposite slope sign condition is described here as existing statically, across an edge. It is a spatial comparison of spectral information. A comparison of spectral information in time is equivalent. Such a *temporal opposite slope sign condition* would work as follows: An eye could sweep across an edge, and the spectral information before and after the movement could be compared. Similarly, there is a temporal equivalent of the crosspoint. Consequences of these isomorphic computations in the temporal domain will be not explored here.

The "slope" of the opposite slope sign condition is the slope of the graph of intensity versus wavelength; it is the derivative of the spectral image intensity function, $dI/d\lambda$.

More formally, given two regions X and Y across an edge and intensity samples I taken at two wavelengths λ_1 and λ_2 , we have the following test for a material change:

Opposite Slope Sign Condition:

$$(I_{X,\lambda_1} - I_{X,\lambda_2})(I_{Y,\lambda_1} - I_{Y,\lambda_2}) < 0 \quad (1)$$

where $I_{X,\lambda_1} = I_X(\lambda_1)$. Condition (1) may be contrasted with the previously derived crosspoint condition (Rubin & Richards, 1982):

Spectral Crosspoint Condition:

$$(I_{X,\lambda_1} - I_{Y,\lambda_1})(I_{X,\lambda_2} - I_{Y,\lambda_2}) < 0 \quad (2)$$

Note that the spectral crosspoint and the opposite slope sign conditions are completely independent. Figure 2a shows the two occurring together. Each condition can arise alone, as shown in Figures 2b and 2c. Finally neither condition is necessary, as shown in Figure 2d.

Independence of Crosspoint and Opposite Slope Sign

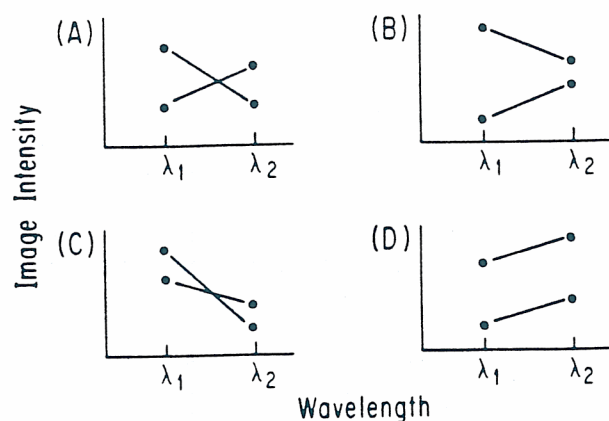


Figure 2. Graphs of image intensity (ordinate) versus wavelength (abscissa). Two wavelength samples, λ_1 and λ_2 , are shown. An image region yields two samples of intensity, one for each wavelength, and is represented by the line segment connecting the two sample values. (a) and (c) Two examples of the spectral crosspoint (Rubin & Richards, 1982). (a) and (b) Two examples of the opposite slope sign condition. This is the minimal configuration that shows different ordinalities. Note that the crosspoint and opposite slope sign condition are completely independent, since they can occur together (a), or each can occur alone (b and c), or neither can occur (d).

The two conditions are related by a kind of symmetry. The spectral crosspoint must make two comparisons across an edge (one for each wavelength), and combine them logically (both comparisons must work out in the correct way). The opposite slope sign condition must make two comparisons, one within each image region, and then combine them logically across the edge.

To summarize: the spectral crosspoint—our original means of finding material changes—has been augmented by a second and independent material change condition: opposite slope sign. The opposite slope sign condition is the key theoretical result on which we will base our spectral representation of material types. We choose opposite slope sign rather than the crosspoint, because the opposite slope sign condition tells us something about each of the two regions that produce it. Namely, one region has positive spectral slope, the other negative. By contrast, the spectral crosspoint cannot be decomposed into assertions about the two regions that produce it. In a crosspoint, spatial and spectral information are hopelessly intertwined. We do not cast aside the crosspoint, though; it is essential for spectral normalization.

4.0 SPECTRAL NORMALIZATION

For the opposite slope sign test to find material edges successfully, it is necessary for the measured spectral intensities to be normalized with respect to illuminant color.⁷ That is, these samples must be transformed to what they would have been under a white (spectrally flat) illuminant. Without this correction, the spectral skew of an illuminant may not only reduce the number of observed opposite slope sign pairs, but more seriously, may transform pairs having the same slope sign (under white light) into opposite slope sign pairs.

By contrast, the spectral crosspoint condition is insensitive to the spectral content of the illuminant, as can be seen by inspecting panels A and C of Figure 2. (See Rubin & Richards, 1982, for a more formal treatment.) We capitalize on this property of the crosspoint to devise a theory of spectral normalization.

Consider now a scene composed of a large number of randomly selected materials. For each image region (simple closed curves defined by edges), take two samples of intensity I_{λ_1} and I_{λ_2} at wavelengths λ_1 and λ_2 . Each region will be associated with a spectral slope sign, which is just the sign of the difference $I_{\lambda_1} - I_{\lambda_2}$. If the illuminant were white, we would expect to have roughly equal numbers of regions of positive spectral slope and regions

⁷A note on mechanism: Neural responses that encode luminance edges can only accurately represent a contrast of 1 to 2 log units. Daylight conditions, however, can range over 5 log units. Therefore some crude adaptation must precede our edge-based spectral normalization.

of negative spectral slope. This expectation is based on two assumptions. The first is that there is a random collection of materials in the scene. The second is that materials in the world are such that a random collection of them will be divided equally between positive and negative spectral slope.⁸

As suggested above, normalization requires a collection of image regions that arises from a random set of materials. What about using *all* image regions? The set of all image regions is not likely to represent a random collection of materials, because many materials will recur in several image regions. For example, if a cast shadow cuts across a single piece of material, that material will be twice represented, once for each side of the shadow edge. A second example arises with pigment density changes. In a forest scene, all leaves are composed of the same material (chlorophyll embedded in a cellulose base). A sensible normalization scheme would not take each leaf as a distinct patch of material; minor variations in pigment density from leaf to leaf ought to be ignored.

It seems clear, then, that not all image regions should participate in normalization. Perhaps a subset of image regions can be found that is more likely to represent a random collection of materials. The spectral crosspoint offers a means of finding such a random subset of regions. Suppose we considered only pairs of regions that have a spectral crosspoint on the edge between them. We would be guaranteed that each pair of regions would correspond to distinct materials. The pairs of different material regions found with the crosspoint will be the subset of image regions that will be used for normalization.

Our normalization scheme works as follows. Recall that we expect the regions found by the crosspoint to represent a random collection of materials. So we expect roughly the same number of regions having positive spectral slope as negative. For the subset of image regions defined by the crosspoint, tally the number having positive spectral slope and the number having negative slope. If the numbers are approximately equal, our expectation has been met; we can infer that the illuminant is white (spectrally flat).⁹ Suppose to the contrary that the number of regions of positive spectral slope exceeds the number of negative-slope regions. Then we can infer that the illuminant is

⁸Our scheme can be easily modified to capture any fixed distribution of positive- and negative-slope materials in the world.

⁹Note there must be some crosspoints for normalization to proceed. If there are no crosspoints, there are no regions to consider. So although it is technically true that there are equal numbers of positive-slope regions and negative-slope regions (namely, zero), we do not want to infer the illuminant is white for two reasons. First, we have no information about any image region, and thus it seems imprudent to guess blindly that the light is white. Second, we have evidence that the scene consists of a single material since it has no crosspoints. Normalization would bring about material change assertions via the opposite slope sign condition, in contradiction to the evidence of uniformity from the crosspoint.

more intense at long wavelengths than at short. (Positive spectral slope means greater intensity in the longer wavelength sample.) Now multiplicatively scale one of the spectral samples. In the example here, we need to multiply all long wavelength samples by some number less than one. Exactly which number? The one that will fulfill our expectation of equal numbers of positive and negative spectral slope. That is, multiply all long wavelength samples by some number (less than one) such as that half of the regions under consideration will have greater intensity in the modified long wavelength sample than the short wavelength sample, and half, the reverse. For a large number of samples, the multiplicative constant of normalization can be calculated from the mean value of the spectral slopes of all regions participating in crosspoints. An algorithm for spectral normalization is presented in Appendix II.

This crosspoint normalization scheme has some useful properties. Each image region participates in normalization only to the extent that it participates in crosspoint edges. There is no strict relation between the size of a region and its potency in normalization.¹⁰ This is good for two reasons. First, the scheme is largely independent of the areas of image regions. This is desirable since we would not want visual systems to treat an image of a large blue thing and a small red thing differently from an image of a small blue thing and a large red thing. Second, the scheme is strictly independent of the length of an edge separating two regions; each crosspoint edge contributes two points regardless of its length.

It is worth comparing our crosspoint normalization with Land's latest normalization theory. Land's (1983) scheme involves comparing the image intensity of a target region with that of a few hundred *random* locations in the image. In such a theory, the larger an image region, the more random locations it will contain. Land's algorithm is therefore strongly dependent on the areas of image regions. Our procedure behaves differently; we expect no effect on normalization from the sizes of image regions (for a fixed pattern of connectivity of image regions), or from the lengths of image edge segments.

5.0 CHOOSING A REPRESENTATION

Assume now that the image has been normalized using the spectral crosspoint condition, as described in section 4. We next select a representation of spectral information based on that rule. In particular, *we seek a simple, convenient spectral representation of materials that is invariant under shadow, highlight, surface orientation change, and pigment density change.*

¹⁰ There is a loose relationship, though. Relatively large regions will tend to have more edge segments (see Appendix II). As the number of edge segments of a region increases, so on average will the number of crosspoints.

For any region in the image, intensity can be measured at a long wavelength and at a second, shorter wavelength. Call these two measurements of image intensity L and S , respectively, for each image region. Suppose we'd like to represent the spectral character of a region with a single number, namely some mapping of the pair (L, S) . Furthermore, we would like the mapping (L, S) to be invariant under the lawful changes. The recognition of material differences would be easy in such a representation. A single material in its different guises—fully lit, shadowed, having different densities of pigmentation, with different surface orientations—would map ideally to a single point. If there were such a mapping, then whenever two image regions mapped to distinct points, we would know they correspond to distinct materials.

The lawful edge types are unfortunately so diverse that there is no function giving us the desired mapping. No single, nontrivial, continuous function of (L, S) will be invariant under multiplicative (shadow), exponential (pigment density), and additive (highlight) changes. Material change, then, cannot be reduced to the problem of distinguishing two points in the range of some function.

The problem isn't hopeless, however, for there is a continuous function invariant under *some* of the lawful changes, namely the multiplicative ones (shadow and surface orientation change). Consider again the two image intensity samples S and L . The quotient L/S will have the identical value on both sides of a surface orientation change or a shadow edge. The simple quotient is not unique in remaining constant across an orientation edge. We will choose among three simple functions having this property:

$$\frac{L}{S} \quad \frac{L}{L+S} \quad \frac{L-S}{L+S} \quad (3)$$

How can we select among these candidates? The function L/S takes image regions into the unbounded interval $(0, \infty)$, while the other two functions take image intensities into closed intervals. ($L/(L+S)$ maps intensities into $[0, 1]$; $(L-S)/(L+S)$ maps into $[-1, 1]$.) The function L/S will be rejected, since any reasonable computational system will be better off using quantities that fall within a closed interval, rather than those that could be arbitrarily large. To choose between the two remaining candidate functions we consider the ease of discovering material changes in these two maps. In particular, how does the opposite slope sign condition appear in each of the candidate mappings?

Given two image regions X and Y , let F denote the function $L/(L+S)$ so that $F(X)$ and $F(Y)$ are the values of the function F of regions X and Y , respectively. Then for F , the opposite slope sign condition is expressed by $[\text{sign}(F(X) - 1/2) \neq \text{sign}(F(Y) - 1/2)]$. (The reason for this expression is that the function F takes on the value $1/2$ whenever $L = S$.)

Let G denote the function $(L-S)/(L+S)$, a common measure of contrast. This is a simple function that facilitates the computation of material change. The sign of G is the sign of the spectral slope of an image region. That is, $[sign(G(X)) \neq sign(G(Y))]$ emerges as the opposite slope (material change) condition.

We prefer the function G to the F for our representation. Whereas to determine material change with G requires only a sign check, with F , the system must maintain the constant $1/2$ and perform two subtractions. The particular choice of F or G , though, seems not to be critical for the goals we have in mind.

Figure 3 shows the interval $[-1,1]$, the range of the function G . Two image regions corresponding to lit and shadowed versions of the same material, or two different surface orientations, will, by design of G , be mapped to the same point. This is shown in Figure 3a. Two image regions of different pigment density have the same slope sign; hence, in the G map, the corresponding pair of points cannot straddle the zero. The same holds for a pair of points corresponding to a highlight and a neighboring matte region. The latter two edge types are shown in the G mapping in Figure 3b. If two image regions are mapped to points straddling the zero (Figure 3c), they arise from different materials.

How Edges Map into the Spectral Representation

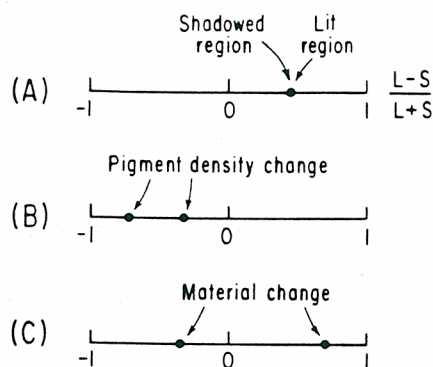


Figure 3. How various processes appear in the spectral representation implied by the mapping $(L-S)/(L+S)$, the range of which is $[-1,1]$. a) Two image regions differing only in surface orientation or shadow map to a single point. b) Two regions differing as matte and highlighted, or as two different degrees of pigmentation density, map to the same half of the range, i.e., they map to points having same-sign coordinates. c) Only two different materials can map to points straddling the zero, i.e., to points of different-sign coordinates.

To summarize, we sought a function of spectral information invariant over the lawful changes. That goal being impossible, we chose $(L-S)/(L+S)$ for two reasons. First, it is invariant across shadows and surface orientation changes. Second, finding material changes with the opposite slope sign condition is easy. The range of the function can be divided into two parts, $(-1,0)$ and $(0,1)$. Materials with albedoes of positive spectral slope sign will map into the positive half of the range, and negative-sloping albedoes to the negative part of the range.¹¹

Finally, it's worth reiterating why we built our spectral representation around the opposite slope sign condition, and the spectral crosspoint. Spectral slope sign is an invariant property of a material's albedo function.¹² The opposite slope sign condition can be decomposed into separate meaningful statements about properties of two image regions: The slope sign of one region is positive, and that of the other, negative. We know something about each region. The crosspoint, by contrast, hopelessly confounds spatial and spectral information. Higher goals of color vision involve describing the properties of individual image regions, and cannot be reached by the crosspoint alone.

6.0 TRICHROMACY: FINDING MORE MATERIAL CHANGES

Suppose we add a third spectral sample, call it M , to our original S and L samples. Adding a third spectral sample will allow the detection of new kinds of material changes.¹³ However, more importantly, the number of basic material categories will be increased from two to six.

In the two-wavelength-sample material representation, an image region is encoded essentially by the rank order of the spectral samples, or equivalently by the sign of the slope of the line segment connecting the samples. Thus, given two wavelength measurements, there are two types of material—negative slope and positive slope. With three wavelength samples, an image region is associated with three slope signs—a slope between each pair of samples (SM , ML , SL). There are six possible rank orderings of the measure-

¹¹ Many continuous maps share the same invariance. We selected our map on the basis of *algorithmic* considerations. The particular choice is independent of the *theory* of finding material change edges.

¹² Since a material is defined as a kind of stuff, a single material can have different albedoes as pigment density changes. What stays constant over these changes in density of pigment is spectral slope sign.

¹³ The additional number of material changes detected with each new spectral sample will drop sharply after the third sample. The reason is that the albedoes of natural objects (in the visible range) are typically slow-changing functions of wavelength (Krinov, 1971; Snodderly, 1979). Cohen (1964) showed that three carefully chosen functions of wavelength captured over 99% of the albedo functions of Munsell chips.

ments ($3! = 6$), and thus six possible basic material types. Any two regions that produce distinct rank orderings of the wavelength samples will bring about one or more opposite slope signs. Any two such regions must therefore be distinct materials.

As a first step in constructing the trichromatic material representation, we combine slope information from two of the three pairs of samples. Arbitrarily, we begin with SM and LM , combining the information in a two-dimensional space as shown in Figure 4a. Image regions are mapped to points in the square $[-1,1] \times [-1,1]$, and a pair of points separated by an axis (or both axes) correspond to two regions of different material, just as did a pair of points straddling the zero in Figure 3c. Any pair of points in a single quadrant may arise from a single material. This is the sense in which quadrants represent material categories. Without yet considering comparisons between S and L samples, we already have a categorical representation in Figure 4a, in which in each quadrant corresponds to a material category.

Let's now examine the third pairing of samples, S and L . What condition holding between a pair of points in the preliminary representation of Figure 4a corresponds to the opposite slope sign condition between S and L ? It is easily shown that if a pair of points straddles the line of unit slope, the points

Trichromatic Representations

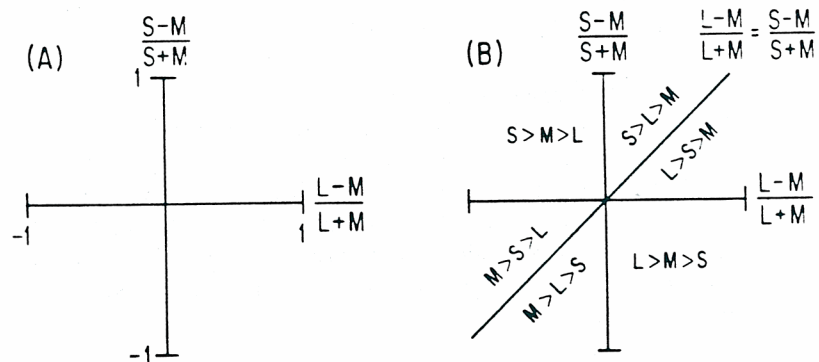


Figure 4. Steps in the construction of the trichromatic material representation. (a) Two axes comparing L and M , and S and M samples, are joined orthogonally. Each quadrant is a material category. Points in different quadrants correspond to distinct materials. Points within one quadrant may belong to the same material; they are considered equivalent in this representation. (b) The line of unit slope in the figure above represents the comparison between S and L samples. Adding the unit slope line divides the color space into six regions or "hexants." Points in different hexants arise from different materials. Note the hexants do not have equal areas.

arise from materials with opposite (S and L) slopes.¹⁴ Furthermore, not just the sign, but the continuous value $(L-S)/(L+S)$ of the L to S comparison is contained implicitly in the representation defined by ordered pairs $(S-M)/(S+M)$, $(L-M)/(L+M)$ that Figure 4a illustrates.¹⁵

The unit slope line in the $SM-LM$ space therefore has special significance, and is added to the representation as a third material change axis in Figure 4b. A pair of points lying across any of the three axes will correspond to a material type, or equivalently, to a rank ordering of the three samples. The particular rank ordering associated with each "hexant" is shown in Figure 4b. Note the hexants of Figure 4b do not have equal areas. The original pair of axes can be joined in a skew fashion to allocate more or less area to the different material categories.

To summarize, image intensities are measured at S , M and L , normalized according to the crosspoint normalization of section 4, and mapped to $(L-M)/(L+M)$, $(S-M)/(S+M)$ in a rectangular coordinate system, initially creating four basic material types. A further subdivision into six types can arise by using the line of unit slope as a third axis, dividing the region $[-1,1]^2$ into six regions, each corresponding to a different material type. Points in different hexants arise from different materials, whereas points common to one hexant may arise from lawful edge events occurring on a single material.

7.0 RELATION TO PSYCHOPHYSICS: THE UNIQUE PRIMARIES¹⁶

Our spectral representation of material types is so far only an abstract model of biological color vision. In our theory, certain terms are left undefined. For example, neither the "spectral samples" of the theory, nor the psychological correlate of "materials" are specified. How then can we assess its relevance? One simple linking assumption will guide the interpretation of our theory: of the traditional psychological color variables, hue, saturation, and lightness, it is hue that encodes material type. (Saturation and lightness likely encode other material properties.)

¹⁴ The line of unit slope is given by $(S-M)/(S+M) = (L-M)/(L+M)$. This is equivalent to $(S-M)(L+M) = (S+M)(L-M)$, or $L = S$. Points above this unit slope line correspond to $L > S$, points below to $S > L$.

¹⁵ Given the values $(S-M)/(S+M)$, $(L-M)/(L+M)$, we can compute the value of $(L-S)/(L+S)$. Let $Q = (S-M)/(S+M)$ and $R = (L-M)/(L+M)$. Then $(L-S)/(L+S) = (Q-R)/(Q+R)$.

¹⁶ We relate our two material-change operators to Land's observations in Rubin and Richards (1984). Also in that reference we also show that the operators correspond to two distinct types of double-opponent units.

One hundred years ago, Ewald Hering (1878, 1920, 1964) offered a simple model for categorical color perception, based upon the notion of "opponent processes." He observed that "redness and greenness, or yellowness and blueness are never simultaneously evident in any color, but rather appear to be mutually exclusive." *Reddish* and *greenish* are mutually exclusive hue categories, and if hue is encoding material properties, then the two categories will partition materials. See Figure 5a. Similarly, *bluish* and *yellowish* will partition materials. See Figure 5b. These two sets of mutually exclusive hue pairs divide the color space into four regions, as in Figure 5c, just as did our trichromatic color space (Figure 4a).

Our claim that Hering's color quadrants correspond to our material categories is predictive: we expect that shadows, surface orientation changes, and pigment density changes would only rarely cause perceived hue to change from *reddish* to *greenish* (or vice versa), or from *yellowish* to *bluish* (or vice versa). (However, as noted in Appendix I, highlights could be troublesome.)

The fact that there are four hue categories supports the idea that trichromatic human vision uses two opposite slope sign checks, as in Figure 4a, but not the third, as shown in Figure 4b. (Goethe [1808], however, proposed a theory of color perception based on six hue categories, which might correspond to the use of all three opposite slope sign checks.) Evidence from infants (Bornstein, Kessen, & Weiskopf, 1976) supports Hering's theory of four hue categories as independent of language and culture. Pigeons also have categorical color perception (Wright & Cumming, 1971), suggesting the computational scheme that we propose here is fundamental to color vision across species.

Opponent Color Theory

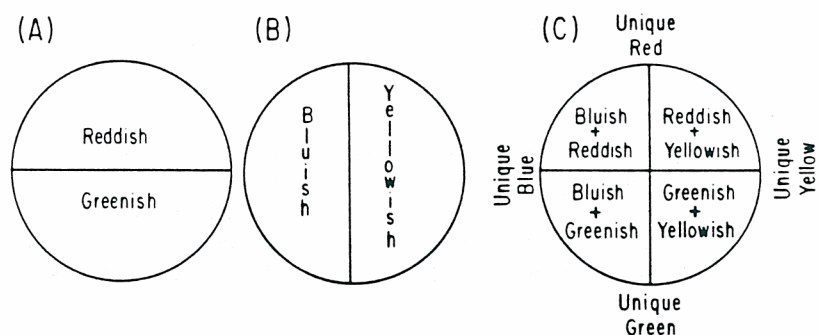


Figure 5. Hering's notion of opponent color processes. (a) All colors are either reddish or greenish, but never both. (b) All colors are either bluish or yellowish, but never both. (c) The two pairs of mutually exclusive colors divide the color circle into four quadrants, similar to the trichromatic representation that we develop in Figure 6a.

Hering's notion of opponent color processes implies four special hues. They are indicated in Figure 5c. These hues, which Hering called the unique psychological primaries, are the boundaries that separate color categories. Primary red is that hue among the reddish hues that separates the yellowish from the bluish; primary blue is that hue among the bluish that splits the reddish from the greenish; and so on. These primary colors are unstable in the sense that any deviation from them involves a change of color categories. Hering's psychological primaries correspond to the axes of our trichromatic representation (Figure 4a).

Just why these primaries have their particular locations in the spectrum is an interesting evolutionary question not addressed here. One possibility is that a creature's material boundaries are positioned in some way as to make the greatest number of discriminations among materials encountered in its environment.¹⁷ Interesting work has been done along these lines. Snodderly (1979) attempted to relate the color vision of New World monkeys to the spectral characteristics of their jungle habitat. Levine and MacNichol (1982) and McFarland and Munz (1975) linked the photopigment characteristics of fishes to the spectral character of light in their environments.

In sum, our spectral representation of material categories is a two-dimensional space in which each quadrant represents a material type, and the axes represent the boundaries between categories. Image regions that map to different quadrants necessarily arise from distinct materials; image regions that map to the same quadrant may arise from a single material. Supposing that hue encodes material type, Hering's observation about human color vision makes sense: hues are divided into four fundamental "material" categories by the mutually exclusive red-green and blue-yellow pairs.

8.0 SUMMARY

Our theory of color vision presents two types of operators—the spectral crosspoint for normalization and the opposite slope sign—which suffice in most cases to normalize for the illuminant and to categorize the albedoes in the scene. Our scheme should differentiate between the common natural pigments (chlorophylls, xanthophylls and flavanoids, for example), but not

¹⁷ Material boundaries can be changed in two ways. The wavelength at which a photopigment captures the greatest percentage of photons can be altered, or new "channels" can be created by combining photopigments. One sort of combination of two spectral samples S and L is a rotation; that is, new coordinates $(S \cos \theta - L \sin \theta, S \sin \theta + L \cos \theta)$ can be created for some angle of rotation θ . The original and rotated coordinate systems will not always agree about whether two image regions satisfy the opposite slope sign condition. That is, the two spectral coordinate systems differing only by a rotation will make different material distinctions. An angle θ can therefore be selected to maximize the number of material changes detected.

between the density variations of any one of these pigments. The theory does not address this latter problem—namely how we interpret the fine changes in the grain of a piece of teakwood. A quantitative color vision system, of greater complexity than the qualitative computations described here, is needed for such fine discriminations. However, categorical color vision does allow coarse but rapid and reliable judgments about materials.

APPENDIX I: LAWFUL PROCESSES

This appendix shows that image edges that arise from (1) change in surface orientation, (2) pigment density variations, (3) shadows and (4) highlights all preserve the ordinal relations of image intensities across that edge, and hence cannot cause the opposite slope sign condition.

A1.1 Surface Orientation Change

Let X and Y be regions on either side of an edge due solely to a surface orientation discontinuity. Then the image intensities (as functions of wavelength) $I_X(\lambda)$ and $I_Y(\lambda)$, measured in X and Y , respectively, are related multiplicatively. That is, $I_X(\lambda) = \alpha I_Y(\lambda)$ for some constant α (Rubin & Richards, 1982; Horn & Sjöberg, 1979). Two functions differing only by a multiplicative constant have identical ordinality.

A1.2 Pigment Density Variation

Suppose X and Y are two regions on a planar piece of a single material that differ only in pigment density. Then if the albedo (as a function of wavelength) of region X is $\rho(\lambda)$, the albedo of Y can be approximated¹⁸ by $\rho^b(\lambda)$, where b is a constant related to pigment density (Rubin & Richards, 1982; Wyszecki & Stiles, 1967).

The light measured from regions X and Y is the product of the albedoes of X and Y with the radiant intensity of the illuminant. Since X and Y are assumed coplanar (recall that pigment density change is stipulated as the *sole* cause of the edge), and the illumination is the same for both, then any difference between measured intensities from the two regions will be due to a difference in the albedo functions. But the albedo functions are related by an exponential constant, and two functions so related will have identical ordinality. Therefore, image intensities across a pigment density change will have identical ordinality. [Examples of this relation for natural pigments can be seen in Krinov (1971), Francis & Clydesdale (1975) or Snodderly (1979).]

¹⁸ This exponential relation presumes that the embedding material is spectrally neutral. If the embedding layer reflects different wavelengths unequally, then change in pigment density has a more complex description. In particular, pigment density changes can mimic material changes.

A1.3 Shadow

Illumination generally consists of two components. *Direct* light comes from a source, which is usually localized. *Diffuse* light is source light that has been reflected off other surfaces (Goral, Torrance, Greenberg, & Battaille, 1984), and is roughly global. Now consider an edge separating a lit region from a shaded one. Both lit and shaded regions reflect diffuse illumination toward the viewer. The lit region, in addition, reflects a direct source. If $I_{lit}(\lambda)$ and $I_{shad}(\lambda)$ are image intensities (as functions of wavelength λ) from lit and shaded regions, respectively, then:

$$I_{shad}(\lambda) = E_{diffuse}(\lambda)\rho(\lambda) \quad (4)$$

$$I_{lit}(\lambda) = [E_{diffuse}(\lambda) + E_{direct}(\lambda)]\rho(\lambda)$$

where $E_{diffuse}(\lambda)$ and $E_{direct}(\lambda)$ are the diffuse and direct components of illumination, and $\rho(\lambda)$ characterizes the albedo of the material.

By inspection of equations (4), it is clear that ordinality can be violated in the case of shadow. That is, a false target is possible. The visual world, fortunately, offers certain regularities. There is usually some close relation between diffuse and direct illumination. This is not surprising, since diffuse light results from diverse, random reflections of the direct light from a variety of materials in the scene. An assumption will be made that this is usually the case: a visual system can presume that diffuse light has the same spectral character as the direct light. That is, $E_{diffuse}(\lambda) = kE_{direct}(\lambda)$, for some constant k . This we call the "grey world" assumption (see Section 3.1), because it is implied by the statement that all the albedoes of a scene will average to grey. Anecdotal data support the grey world assumption. Hailman (1979) measured spectral irradiance functions in a pine woods in a sunny area and in nearby shade. The functions are strikingly similar in shape, and are shown in Figure 6.

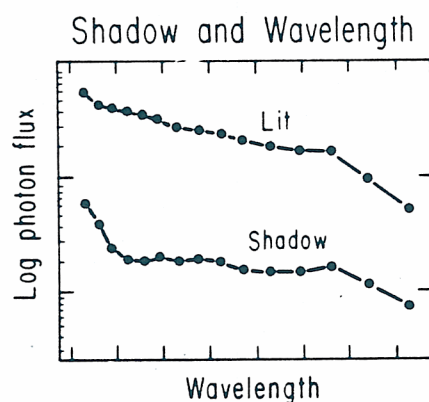


Figure 6. Measurements of the spectral irradiance functions of direct sunlight and nearby shade in a Florida pine woods, adapted from Hailman (1979), Figure 74a. On the ordinate is the logarithm of photon flux. The abscissa shows wavelength.

Invoking the grey world assumption, equations (4) become:

$$I_{shad}(\lambda) = kE_{direct}(\lambda)\rho(\lambda) \quad (5)$$

$$I_{lit}(\lambda) = (1 + k)E_{direct}(\lambda)\rho(\lambda)$$

Note that the lit and shaded regions now give rise to multiplicatively related image intensity functions. Ordinality will therefore be preserved.

A1.4 Highlights

The analysis of highlights is slightly more complex. The following equations (Rubin & Richards, 1982, equations 14a) express the image intensities to be found in a highlight and neighboring matte region.

$$I_{matte}(\lambda) = (E_{diffuse}(\lambda) + E_{direct}(\lambda))\rho(\lambda) \quad (6)$$

$$I_{highlight}(\lambda) = \delta E_{direct}(\lambda) + (1 - \delta)[E_{diffuse}(\lambda) + E_{direct}(\lambda)]\rho(\lambda)$$

where $I_{matte}(\lambda)$ and $I_{highlight}(\lambda)$ are the image intensities (as functions of wavelength) in matte and highlight regions, and $\delta \in (0,1)$ is a constant that indicates to what extent the surface is mirrorlike ($\delta = 1$ describes a perfect mirror). (See Richards, Rubin, & Hoffman, 1982, for a more extended treatment.)

The equations express the fact that both highlighted and matte regions reflect both direct and diffuse light. In addition, the highlight, acting as a partial mirror, reflects the direct light.

Applying the grey world assumption, equations (6) become:

$$I_{matte}(\lambda) = (1 + k)E_{direct}(\lambda)\rho(\lambda) \quad (7)$$

$$I_{highlight}(\lambda) = \delta E_{direct}(\lambda) + (1 - \delta)(1 + k)E_{direct}(\lambda)\rho(\lambda)$$

which reduces to

$$I_{matte}(\lambda) = (1 + k)E_{direct}(\lambda)\rho(\lambda) \quad (8)$$

$$I_{highlight}(\lambda) = E_{direct}(\lambda)[\delta + (1 - \delta)(1 + k)\rho(\lambda)]$$

By inspecting equations (8), it can be seen that highlights can produce a spurious violation in ordinality. Assume now that the image has been normalized with respect to the color of the illuminant. Normalization is any scheme that allows recovery of the spectral character of the illuminant. (Such a computation is presented in section 4.) Normalization is equivalent to a transformation of the image intensities to what they would have been had the

illuminant been white; it allows us to set $E_{direct}(\lambda) = \beta$, where β is some constant.

Both equations (8) can now be rewritten substituting (β) for $E_{direct}(\lambda)$, yielding

$$I_{matte}(\lambda) = \beta(1 + k)\rho(\lambda) \quad (9)$$

$$I_{highlight}(\lambda) = \beta[\delta + (1 - \delta)(1 + k)\rho(\lambda)]$$

With the two assumptions of grey world and spectral normalization, highlights will not produce violations in ordinality. This can be seen in equations (9), where the image intensity function of the highlighted region is simply related to the image intensity function of the neighboring matte region. The intensity in the matte region is multiplied by a constant $(1 - \delta)$, and then a constant function ($I(\lambda) = \delta\beta$) is added. These two operations preserve ordinality; hence no opposite slopes will arise given our assumptions.

APPENDIX II: ALGORITHM FOR SPECTRAL NORMALIZATION AND MATERIAL CATEGORIZATION

Given a full-color image of a scene lit by an unknown illuminant, and a way of finding edges and regions, regions can be assigned to one of a small number of material categories. Regions in different categories are made of different materials. An algorithm for categorizing materials is sketched below.¹⁹ The first step is to correct for colored illumination; the second is to categorize.

A2.1 In the Beginning

The original full-color image can be viewed through three spectral filters, yielding three distinct maps of image intensity, say R , G , and B . See Figure 7a. These three maps of image intensity we call "spectral images." The number of filters, or their spectral characteristics, should not be important. All that matters is that the filters yield independent measurements.

A2.2 Spectral Normalization

First, apply an edge operator to the image. The particular edge operator should not be crucial. Assume the edge operator produces a closed set of

¹⁹ This algorithm makes the grey-world assumption. If one prefers not to make this assumption, then normalize the dark (shadowed) regions separately from the bright (fully-lit) regions.

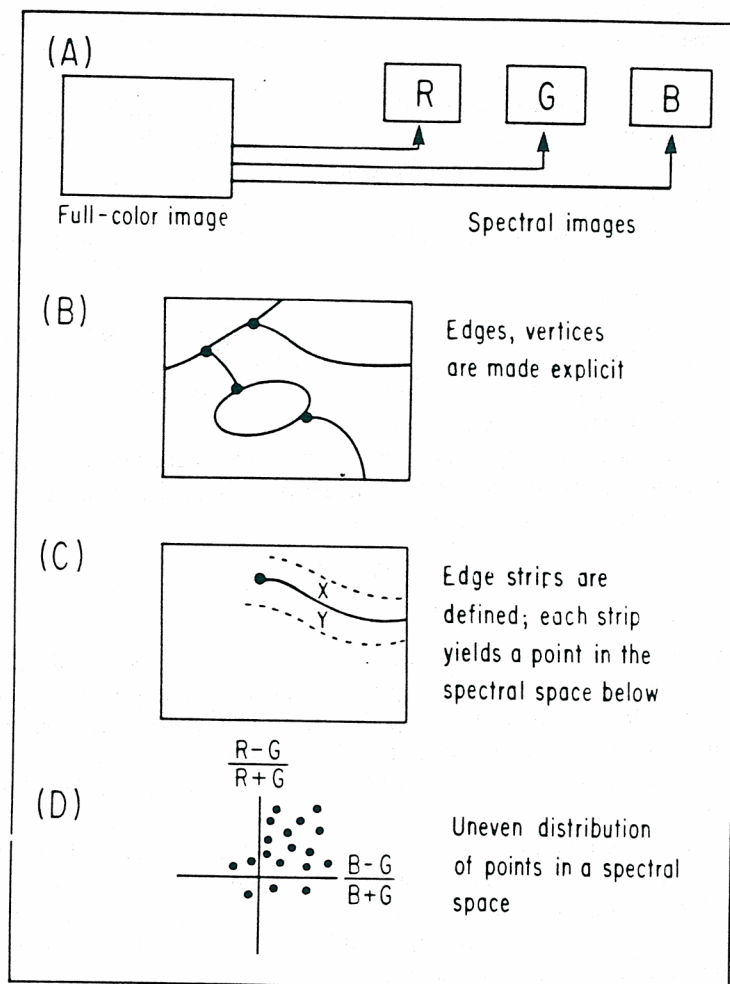


Figure 7. The full-color image is run through three spectral filters, *R*, *G*, and *B*. (b) Edge segments have been found and made explicit. This image shows five edge segments. Vertices have been found, and are here marked with large black dots. (c) On either side of one of the edges, narrow strips *X* and *Y* are defined. No edge segments should be in the strips. Intensity averages will be taken in the three spectral images in both of the strips, yielding six measurements. This is done for each edge segment in the image. (d) Measurements taken from strips about each edge map to points in a spectral space defined by axes as labeled. Normalization consists of multiplying *R* and *B* values by factors such that equal numbers of points will be found in each quadrant.

edges.²⁰ Next, each *edge segment* must be made explicit. (See Figure 7b.) This involves finding vertices, since two adjacent vertices delimit an edge segment. (A role for vertices in normalization has been suggested by Lettvin [1967] and Linden [1974].)

We next iterate through the list of edge segments. For each edge segment, two narrow strips must be defined, one on each side. Call the strips *X* and *Y*. (Here again vertices are important since the strips must be free of edges.) See Figure 7c.

Average the intensity values of each of the spectral images *R*, *G*, and *B* in both the *X* and *Y* strips. The output of this step is six values, R_X , R_Y , G_X , G_Y , B_X , and B_Y .

For each edge segment, check for two types of crosspoint, *RG*, and *BG*.²¹ (The conditions are $(R_X - R_Y)(G_X - G_Y) < 0$ and $(G_X - G_Y) \times (B_X - B_Y) < 0$, respectively.) Note the possibility of a third crosspoint involving the *R* and *B* samples.

Suppose an image has *n* crosspoint edge segments. For each crosspoint, record spectral information about the two abutting strips. In particular, store two color contrast values per region:

$$\frac{R_i - G_i}{R_i + G_i}, \frac{B_i - G_i}{B_i + G_i}, i = 1, \dots, 2n \quad (10)$$

where *i* is an index ranging over the *2n* edge strips defined around *n* crosspoints. This particular form of ratio is useful because its value must lie in the closed interval $[-1, 1]$. The spectral information recorded can be considered as *2n* points in a two-dimensional spectral space (with axes of $(R-G)/(R+G)$ and $(B-G)/(B+G)$) shown in Figure 7d. (See also Figure 4a.)

Let *U* be the number of points in the upper half-plane of the spectral space (Figure 7d), and *L* be the number of points in the left half-plane. Under a white illuminant, we'd expect a random assortment of materials to yield $U \approx L \approx n$; that is, points should be roughly equally distributed among the quadrants of the spectral space.

If the *2n* points are not divided equally among the quadrants of the spectral space, we must seek normalization constants α and β that satisfy the following criterion:

$$\text{MEDIAN} \left[\frac{\alpha R_i - G_i}{\alpha R_i + G_i} \right]_{i=1, \dots, 2n} = \text{MEDIAN} \left[\frac{\beta B_i - G_i}{\beta B_i + G_i} \right]_{i=1, \dots, 2n} = 0 \quad (11)$$

²⁰ If algorithm for edge detection does not produce closed edges, then regions must somehow be identified using edge fragments.

²¹ The *R* and *G* samples can yield crosspoints, and independently, so can the *B* and *G* samples. The *G* sample could just as easily be taken as the photopic luminosity function.

For a large enough number of image regions, we can take

$$\alpha = \frac{1 - \bar{C}_{RG}}{1 + \bar{C}_{RG}} \quad \beta = \frac{1 - \bar{C}_{BG}}{1 + \bar{C}_{BG}} \quad (12)$$

where \bar{C}_{RG} and \bar{C}_{BG} are means of the sets of measurements (12):

$$\bar{C}_{RG} = \frac{1}{2n} \sum_{i=1}^{2n} \frac{R_i - G_i}{R_i + G_i} \quad \bar{C}_{BG} = \frac{1}{2n} \sum_{i=1}^{2n} \frac{B_i - G_i}{B_i + G_i} \quad (13)$$

The values of α and β in (12) will provide a correct normalization (i.e., normalization criterion (13) will hold) given some simple statistical conditions.²²

The correctness of the normalization constants α and β can easily be checked by verifying that criterion (13) holds. If not, the values of α and β can be adjusted incrementally in an iterative procedure. The entire normalization algorithm is shown as a flowchart in Figure 8.

Once correct values of the normalization constants are returned by the algorithm, the three spectral images R , G and B can be transformed into a set of normalized spectral images. All values in the R image are multiplied by α , yielding R^* . (The asterisk superscript denotes normalized intensity; see section 7.2.2.) Similarly, $B^* = \beta B$. Spectral image G is unchanged: $G = G^*$. See Figure 9a.

A2.3 Spectral Categories

Suppose that when closed edge segments were found that image regions were made explicit. For each region i , measure the average values of the normalized spectral images, yielding the triplet (R_i^*, G_i^*, B_i^*) . A triplet of numbers yields one obvious pair of ordinal relations:

$$(R_i^*, G_i^*, B_i^*) \mapsto (\text{sign}_{RG}, \text{sign}_{BG})_i \quad (12)$$

where sign_{RG} is "+" if $G_i^* > R_i^*$, and "-" otherwise.

Each region can therefore be assigned to one of four material categories: $(+, +)$, $(-, +)$, $(-, -)$, $(+, -)$. This is shown in Figure 9b. The regions that are in different categories are composed of distinct materials.

Note that a third ordinal relation is sometimes independent, the $R^* - B^*$ comparison. If this relation is included, six spectral categories obtain.

²² There should be at least 12 independent crosspoint edges, and the mean and median of the set of measurements $\{(R_i - G_i)/(R_i + G_i)\}$ must approach the same value as $i \rightarrow \infty$, and similarly for the set of measurements $\{(B_i - G_i)/(B_i + G_i)\}$. (See Siegel, 1956.)

NORMALIZATION ALGORITHM

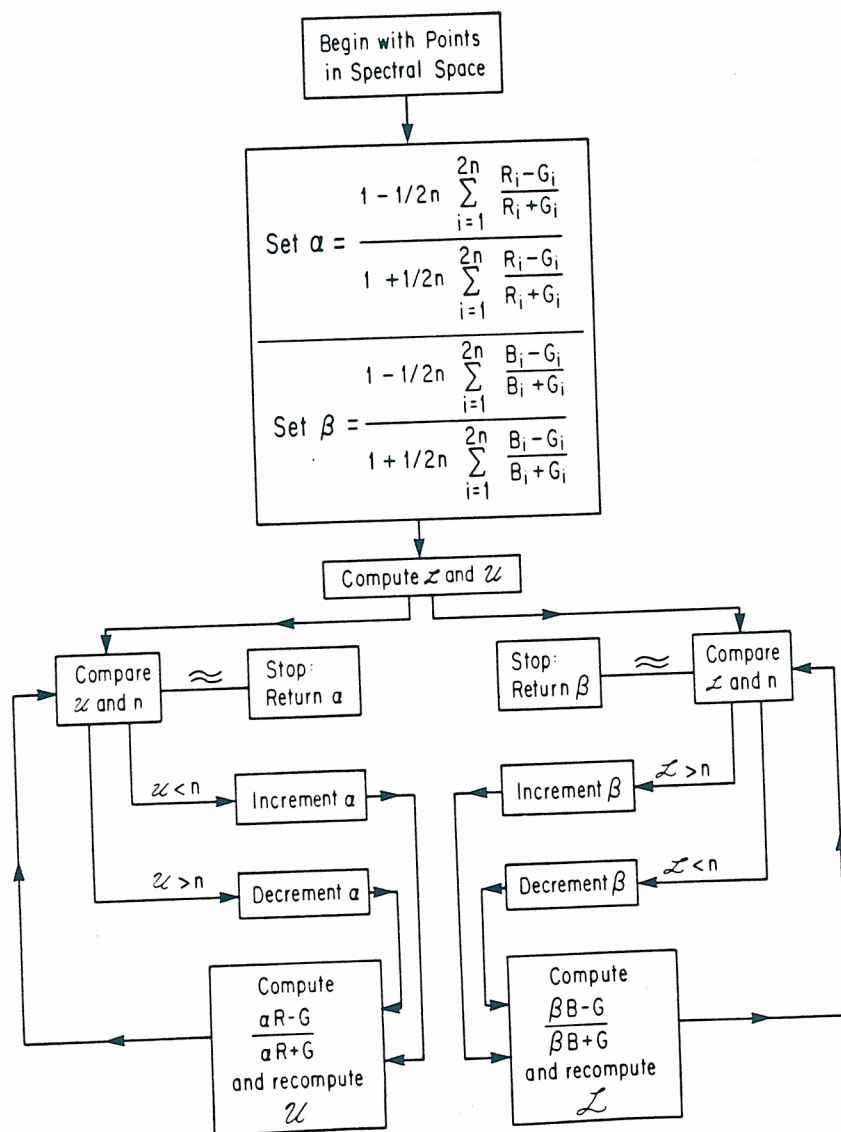


Figure 8. Normalization Flowchart. Begin with points scattered in spectral space, and end with a pair of multiplicative normalization coefficients, α to balance the R image with respect to G , and β to balance the B image with respect to G .

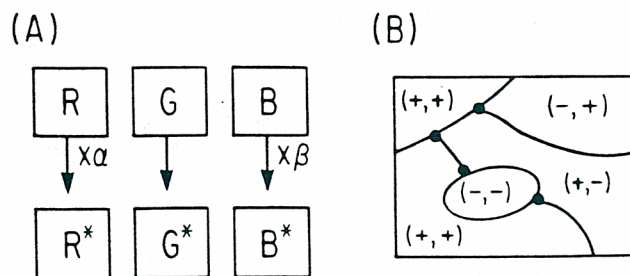


Figure 9. (a) The three spectral images R , G , and B are normalized using the multiplicative constants produced by the procedure shown in Figure 8. The normalized spectral intensity maps are R^* , G^* , and B^* . (b) The regions of the image sketched in Figure 7b labeled with material categories. Each region is assigned one of four possible ordinal doublets.

Finally, note that while the algorithm described here is categorical, continuous information has not been lost; it is still available for more refined purposes. For each region i , the continuous-valued coordinates

$$\left[\frac{R_i - G_i}{R_i + G_i}, \frac{B_i - G_i}{B_i + G_i} \right] \quad (13)$$

should be useful.

REFERENCES

- Bornstein, M.H., Kessen, W., & Weiskopf, S. (1976). Color vision and hue categorization in young human infants. *Journal of Experimental Psychology: Human Perception and Performance*, 2, 115-129.
- Cohen, J. (1964). Dependency of the spectral reflectance curves of the Munsell chips. *Psychonomic Science*, 1, 369-370.
- Francis, F.J., & Clydesdale, F.M. (1975). *Food colorimetry: Theory and applications*. Westport, CT: AVI.
- Goethe, J.B. von. (1808). *Zur farbenlehre, didaktischer teil*. (See R. Matthaehi, Ed., *Goethe's color theory*, 1970, Van Nostrand Reinhold, New York.
- Goral, C.M., Torrance, K.E., Greenberg, D.P., & Battaille, B. (1984). Modelling the interaction of light between diffuse surfaces. *Computer Graphics*, 18(3), 213-222.
- Hailman, J.P. (1979). Environmental light and conspicuous colors. In E.H. Burt, Jr. (Ed.), *The behavioral significance of color* (pp. 289-357). New York: Garland Press.
- Hering, E. (1878). *Zur lehre vom lichtsimm*. Vienna: Carl Gerald's Sohn.
- Hering, E. (1920). *Grundzuge der lehre vom lichtsimm*. Berlin: Julius Springer.
- Hering, E. (1964). *Outline of a theory of the light sense*. L.M. Hurvich and D. Jameson, trans.). Cambridge, MA: Harvard University Press.
- Horn, B.K.P., & Sjoberg, R.W. (1979). Calculating the reflectance map. *Applied Optics*, 18, 1770-1779.

- Krino, E.L. (1971). *Spectral reflectance properties of natural formations*. National Research Council of Canada, Technical Translation 439.
- Land, E.H. (1983). Recent advances in retinex theory and some implications for cortical computations: color vision and the natural image. *Proceedings of the National Academic of Science, USA*, 80, 5163-5169.
- Lettvin, J.Y. (1967). *The colors of colored things* (M.I.T. Research Laboratory of Electronics Quarterly Progress Report #87, 193-229). Cambridge, MA: Massachusetts Institute of Technology.
- Levine, J.S., & MacNichol, E.F., Jr. (1982). Color vision in fishes. *Scientific American*, 246(2), 140-149.
- Linden, L.L. (1974). *Evolutionary constraints on the dimensions of color space*. Masters thesis, Massachusetts Institute of Technology Department of Electrical Engineering.
- Marr, D. (1982). *Vision: A computational investigation into the human representation and processing of visual information*. San Francisco: W.H. Freeman.
- McFarland, W.N., & Munz, F.W. (1975). Part III: The evolution of photopic visual pigments in fishes. *Vision Research*, 15, 1071-1080.
- Richards, W.A., Rubin, J.M., & Hoffman, D.D. (1982). Equation counting and the interpretation of sensing data. *Perception*, 11, 557-576.
- Rubin, J.M., & Richards, W.A. (1982). Color vision and image intensities: When are changes material? *Biological Cybernetics*, 45, 215-226.
- Rubin, J.M., & Richards, W.A. (1984). *Color vision representing material categories* (A.I. Memo 764, M.I.T. Artificial Intelligence Lab). Cambridge, MA: MIT.
- Siegel, S. (1956). *Nonparametric statistics*. New York: McGraw Hill.
- Snodderly, D.M. (1979). Visual discriminations encountered in food foraging by a neotropical primate: Implications for the evolution of color vision. In E.H. Burt, Jr. (Ed.), *The Behavioral Significance of Color* (pp. 238-285). New York: Garland Press.
- Wright, A.A., & Cumming, W.W. (1971). Color-naming functions for the pigeon. *Journal of the Experimental Analysis of Behavior*, 15, 7-11.
- Wyszecki, G., & Stiles, W.S. (1967). *Color science: Concepts and methods, quantitative data and formulas*. New York: Wiley.