Behavior Research Methods, Instruments, & Computers
1984, 16 (2), 88-95

SESSION II
INVITED ADDRESS

Geoffrey Loftus, *Presider*

Visual sensing by humans and computers

BRIAN A. WANDELL
Stanford University, Stanford, California

I review a research project whose goal is to develop algorithms to automatically assess object color, despite an unknown ambient illumination. The successful completion of the project will require (1) identifying the nature of the image information that can—in principle—permit separation of object from lighting effects in the sensed image, (2) devising computational methods for obtaining the desired information, and (3) selecting a mapping from the numerical information in images into useful color descriptors. Each of these issues is reviewed, and hypotheses are described.

This presentation provides an overview of a project that overlaps the fields of human and computer vision. I have been working on the problem of automatic color assessment, that is, color assessment by a computer.

My description of the specific problem varies a bit, depending upon whether I am facing an audience of computer scientists or one of psychologists. For this audience, which is sophisticated both in the use of computers and in psychology, I think it will be useful to describe the problem both ways.

Programming Problem

Suppose we are designing a mobile robot, that is, one that roams around a factory floor. The robot often comes to the same place in the room and tries to assess its position. The robot uses, in part, information about the color of various objects it expects to find on the factory floor. For the robot, assessing "color" means assigning one of a small list of potential descriptors, such as color names, to the different points in its field of view.

The robot, a tireless friend of the American Industrial Revolution can work at all times, under all kinds of conditions. Therefore, there are large lighting differences at the different times the robot positions itself. The sensed image may be input on a sunny or on a cloudy day, or at night under tungsten lighting, and sometimes, if the manager decides to save energy, under energy-efficient fluorescent lighting. The robot's image analysis program must be designed, therefore, to meet an important challenge: The method for assigning image descriptions of surface spectral reflectance must yield the same descriptors under all lighting conditions. The program is a success when the color names assigned to each region of a picture are invariant with respect to changes in the ambient lighting conditions.

This work was supported by Grant 2 RO1 EY03164 from the National Eye Institute and Grant NASA-NCC-2-44 from the National Aeronautics and Space Administration. I thank J. Farrell, L. Maloney, and M. Pavel for their comments. My mailing address is: Department of Psychology, Stanford University, Stanford, CA 94305.

Psychological Problem

From a psychologist's point of view, I have merely said that an algorithm must be created that will associate color descriptors to objects, and that the algorithm must exhibit color constancy.

An analysis of this problem can be divided into three parts. First, the information in the image must be analyzed. In part, this will help explain why color constancy is a difficult problem. It also provides a means of introducing the single current hypothesis available for describing where the information crucial for color constancy may be found in the image. Second, assuming that the hypothesis concerning the nature of the relevant information is correct, how might we extract that information from the image. Finally, once the relevant information has been extracted, we still must select a method for mapping the (large set of possible) numerical values into the (small set of possible) color descriptors. Different methods of achieving this color descriptor assignment will be described.

THE INFORMATION

Why is it that color constancy poses a challenge? Well, consider what happens to the image at the eye under two kinds of changes to the ambient illumination: a simple change in intensity, and then a change in the spectral character of the illumination.

Intensity Change

Let us consider the light reflected from a point x, at each wavelength λ. When the surface reflectance is defined as $R(\lambda,x)$, and the lighting reaching the point is $L(\lambda,x)$, we may simply multiply R and L at each wavelength and sum over wavelength. This provides an estimate of the total number of quanta, $N(x)$, from point x.

$$N(x) = \int L(\lambda,x)R(\lambda,x)d\lambda$$

Suppose there is a change in the lighting, by a factor I. From each point we have then simply multiplied the total number of quanta by the change in intensity. This changes the number of quanta arriving at our eyes by this same factor, so that $N(x)$ becomes $IN(x)$. It follows that, under a simple intensity change of the lighting, the ratio of the number of incident quanta at two nearby points, say x and y, will be preserved.

$$\frac{N(x)}{N(y)} = \frac{IN(x)}{IN(y)}$$

Various authors (e.g., Land, 1964) have suggested this ratio to assess the relative image reflectance at nearby points.

Spectral Shift

The world does not treat us kindly if we allow that a change in illumination can include a change in the wavelength distribution of the lighting, such as occurs under most changes of illumination. The reason for this is shown in the next equation, which represents the effects of an arbitrary change in the amount of energy in the lighting at each wavelength. If we denote a change in the ambient lighting near the point x by writing that $L(\lambda,x) \rightarrow L(\lambda,x) + \Delta(\lambda)$, the the effect on $N(x)$ is

$$N(x) \rightarrow N(x) + \int \Delta(\lambda)R(\lambda,x)\, d\lambda.$$

Under this sort of transformation, we cannot depend on the ratio of quanta coming from the two points being preserved.

$$\frac{N(x)}{N(y)} \rightarrow \frac{N(x) + \int \Delta(\lambda)R(\lambda,x)d\lambda}{N(y) + \int \Delta(\lambda)R(\lambda,y)d\lambda}$$

The exact effect of a lighting change depends on an interaction between the difference in the light distribution, $\Delta(\lambda)$, and the entire reflectance function, $R(\lambda,x)$. We cannot even be certain as to which term, $\int \Delta(\lambda)R(\lambda,x)d\lambda$ or $\int \Delta(\lambda)R(\lambda,y)d\lambda$, will be larger. Ratio theories, in particular, will not enable us to solve the problem.

IDENTIFYING THE INFORMATION

Since the values at pairs of adjacent points seem inadequate to capture the effects of change in the lighting, we must seek a solution based on a more complete analysis of the spatial distribution of the image. At present, there is only a single hypothesis in which the information may be found. The idea is implicit in the work of Land (1964) and Land and McCann (1971) and is given its clearest statement in an elegant paper by Horn (1974).

Consider the important case in which the lighting source is distant from the surfaces. In this case, there will be a significant difference between the spatial distribution of lighting effects and the spatial distribution of surface properties. When the lighting source is distant from the objects, the spatial variation in the retinal image that arises due to the lighting tends to be slow and gradual compared with the spatial variation in the retinal image that arises due to surface boundaries. Under these conditions, spatial[1] variations in the retinal image that occur due to surface variation will occur at higher frequencies than variations due to the ambient lighting. Thus, if the observer is gazing at a truly uniform surface, the shading across the surface—except for shadows—will slowly vary. When the objects contain edges or surface markings, generally called surface discontinuities, the rate of change over space tends to occur slightly more rapidly.

Computational Analysis

In addition to describing the problem in these terms, Horn (1974) suggested a simple computational approach to the problem that is exemplary in its clarity, though inadequate for practical application. Horn modeled the world as consisting of either simple edge discontinuities, as shown by the steep edges in Figure 1, or effects due to lighting, as shown by the gradual changes in surface intensity. In this simplified world, Horn pointed out that we can separate lighting effects from surface properties, as shown in Figure 1. First, the derivative of the image intensity is computed. For the restricted world, Horn assumed—following Land (1964)—the result will be large only near edges. One can then threshold the derivative, and reintegrate. This estimates the brightness up to a scaling parameter that is free from the integration.[2]

Statement in Terms of Filtering

The point of Horn's (1974) scheme, although it is phrased in terms of derivatives and integrals, is to selectively filter the image so that the low spatial frequency information (assumed to arise from variation in the ambient lighting) is removed, leaving only intensity variation due to properties of the surfaces.

The hypothesis as to the source of the information is simply this: Look for surface properties in the high

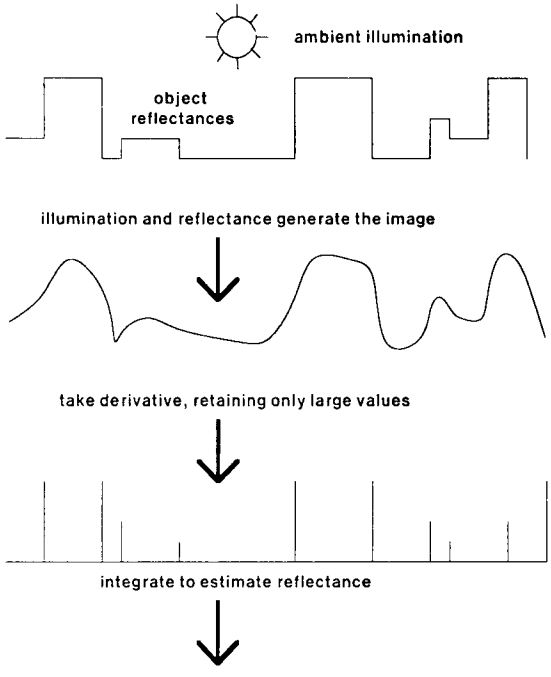

Figure 1. A flowchart indicating the principal steps in Horn's (1974) analysis of lightness constancy. The original scene is characterized by a remote illumination source and objects with idealized edges. The relatively rapid variations in the image are due, then, principally to the effects of the edges of the objects. By differentiating, and then thresholding the results, we are left with a function taking nonzero values at only the edges. The integral of this function is the algorithm's estimate of object reflectance.

spatial frequency component of the image. And, conversely, look for information about the ambient lighting in the low spatial frequency information.

HUMAN VISION

The significance of this hypothesis is that it provides a response to our first question: Information about surface reflectance should be recoverable from the relatively higher spatial frequency terms in the image.

In order to provide this hypothesis a fair test, we must devise a sensible procedure for separating relatively "high" spatial frequency components from relatively "lower" components. The image, based on the higher spatial frequency terms, should be generated to capture the surface properties. I will suggest later that the second image, derived from the lower spatial frequencies, should be generated to capture the ambient lighting properties and serve the function of regulating system sensitivity.

This type of analysis of computer images was suggested by Barrow and Tenenbaum (1978), who referred to the images for representing different aspects of a scene as intrinsic images.

How to Perform the Separation

How should we perform the separation of images? There has been considerable excitement in the human vision literature over the last 15 years or so about the concept, suggested by Campbell and Robson (1968), that the human visual system is composed of many primitive detectors, each of which is sensitive to a fairly restricted range of spatial frequencies. Such a sensor is called a bandpass filter.

The physiological basis for these spatial bandpass filters is thought to originate, but not end, in the spatial sensitivity characteristics of the ganglion cells of the retina, illustrated in Figure 2. Ganglion cell responses are the final signal emerging from the retina, their axons giving rise to the optic nerve. In mammals, these cells respond to light with a spatial organization originally described by Kuffler (1953); this is now widely referred to as center-surround organization. This means that the discharge rate of the neurons is increased when a spot falls on a central region and decreased when light falls on the surrounding region. This type of cell has an on-center and off-surround. If the polarity of the responses is reversed, the cell is an off-center and on-surround cell. Figure 2 also illustrates why a center-surround cell responds more strongly to spatial sine waves at a central frequency than to spatial frequencies that are higher or lower. These cells act as bandpass filters. The spatial frequency whose peak region just covers the receptive field center, and whose troughs fall in the surround, is an excellent stimulus for such a receptive field. If the frequency is too low, then the peak region will cover the entire cell, and, due to the center-surround antagonism, the net response will be rather poor. Similarly,

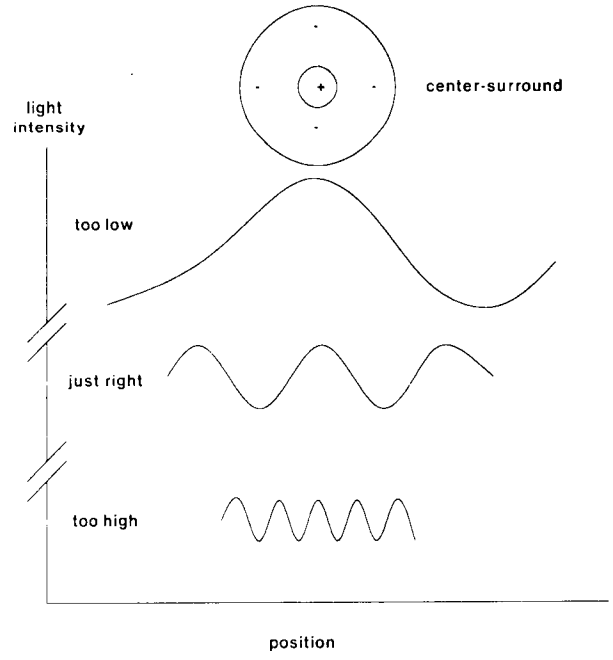

Figure 2. The center-surround structure of a ganglion cell means that the cell may be thought of as a frequency selective filter. The sensitivity of the receptive field causes it to be relatively insensitive to low spatial frequency gratings and high spatial frequency gratings. The optimal grating will have a half-cycle matched in size to the cell's central region.

if the frequency is too high, little response will be generated from either the center or the surround. The actual best frequency for any individual cell is determined by the size of the center and surround, larger cells being tuned primarily to lower spatial frequencies.

Receptive Field Size Changes Center Frequency

When there is a set of such cells whose receptive fields overlap on some region of the retinal surface, the response of each cell is an estimate of the amount of energy in a bandpass region of the spatial frequency domain. The responses of these elementary units of different size are estimates of the spatial frequency energy near the peak frequency of the particular unit, in a local region of the retinal surface.

In principle, it would be possible to construct a set of retinal ganglion cells of different center-surround sizes, with overlapping receptive fields centered at one position in the retina.[3] Peichl and Wassle (1979), however, quantitatively analyzed the size distribution of retinal ganglion cells in the cat and demonstrated that the size coverage of any particular morphological type of retinal ganglion cell shows rather little spatial variation at a single retinal location. It becomes likely, therefore, that if such multiple bandpass filtering exists in the visual system, it is not present until later, at probably cortical levels (see, e.g., Movshon, Thompson, & Tolhurst, 1978a, 1978b, 1978c).

Breaking up the Image

The center-surround mechanism of the retina and subsequent cortical mechanisms provide a plausible means for performing a local frequency analysis of the image. What remains to be specified is how, over each part of the image, we may determine which filter responses should be used to generate the high spatial frequency reflectance image, and which should be used to generate the low spatial frequency ambient illumination image. This certainly cannot be done by simply choosing a fixed value. The reason for this is shown in Figure 3. As an observer moves toward an object, the high spatial frequencies get mapped into lower spatial frequencies. Using a fixed value for the frequency separation would have the effect of failing to be invariant as objects changed their distance with respect to the observer. Any filtering scheme must be flexible enough to adjust what is meant by high and low spatial frequencies as a function of the contents of the image.

Hierarchical Filtering Scheme

What is required is a flexible method for determining the frequencies at which it is appropriate to separate the images into surface reflectance and ambient lighting images. Ideally, the method should be based upon information extracted from the images themselves. One way to analyze this problem is to consider the response of the different bandpass filters in the human visual system

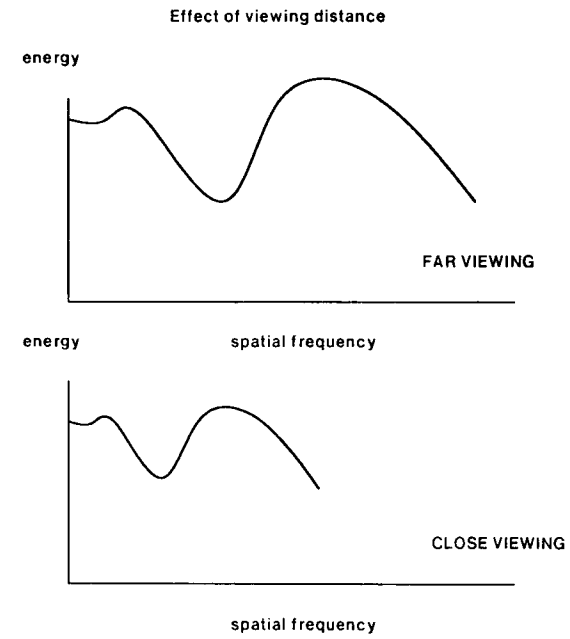

Figure 3. When an object is seen from different viewing distances, to a first approximation its spatial frequency energy distribution is changed by merely scaling the frequency axis. This effect is illustrated in the two panels.

in response to an image that has a clear separation of high and low spatial frequencies.

Figure 4 depicts a case in which there is a clear segregation of the high and low spatial frequency regions. This represents a best-case analysis, useful in the description of the algorithm, but not always true of images.[4] The case drawn here has a notch that identifies the natural place to divide the image to obtain surface and lighting information.

The way to define the two separate images is as follows. First, we evaluate the response of the smallest receptive fields through the largest, thereby effectively calculating the spatial frequency content of the image at some location. Next, we search for a minimum in the value of the response of the cells. This defines the lower limit of the region of spatial frequencies that we will use to estimate surface reflectance. Notice that this method, based on looking for the notch, flexibly adapts itself and maintains the image grouping at the notch, independently of the viewing distance.[5]

Forming the Image

Having thus identified the set of bandpass filters that are carrying the information, we must now find some method for pooling the information across the separate images in order to form a single reflectance image. The task of pooling the information to form some judgment concerning image properties is a difficulty that must be resolved for any system whose initial encoding is based upon separate bandpass encodings. Some of the interesting recent proposals about how

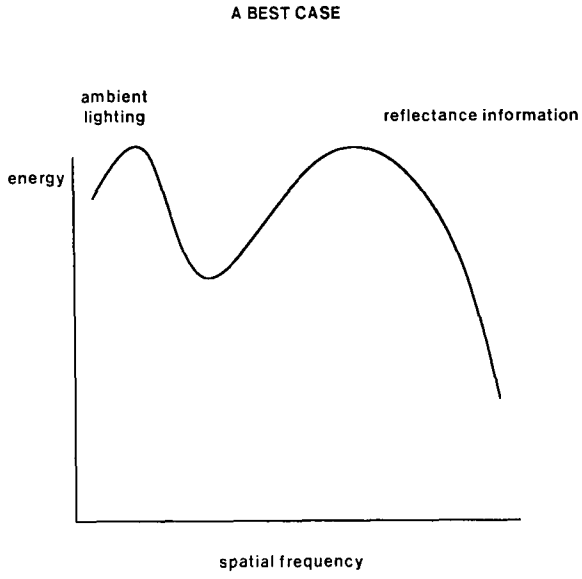

Figure 4. In some regions of some images, we find clear notches in the spatial frequency content. The notch location can be estimated from the responses of filters with center-surround organization, differing only in their size. We use the high frequency information to estimate properties of surface reflectance and the low frequency information to estimate properties of the ambient illumination.

the information should be combined require considerable computation and additional theory (Witkin, 1983).

For the present analysis, the procedure for combining the multiple bandpass versions of the image for the purposes of constructing a reflectance image is quite simple. We may use the response generated by the center of the smallest center-surround unit and the response generated by the surround of the largest center-surround unit that define the frequency band of interest. The filter characteristics formed from the difference of this center and surround are matched to the spatial frequency region that we have identified. Effectively, then, we are using the response of the multiple bandpass filters to form a new, single bandpass filter whose bandwidth is matched to the estimated spatial frequency range. With this flexible filtering scheme, we can postpone to a fairly late stage in the image analysis the center and the surround that we wish to choose in representing the spectral reflectance information in the image.

Analysis of the Properties of This Flexible Filtering Scheme

The algorithm I described is analytically rather complex, but it has some rather simple features. First, despite the apparent complexity and flexibility of the algorithm, the generated response of the entire system is nearly that of a linear filter. Specifically, if the input to the algorithm is a sinusoidal spatial frequency grating, the spatial output of the filters is also a sinusoidal output distribution. Further, multiplying the amplitude of the input grating multiplies the pattern of filter responses.

When the principle of superposition (Bracewell, 1978) is tested directly,[6] we find a particularly simple deviation. The mixture of two sine-wave intensity patterns yields an output that is a mixture of two sine-wave patterns in the filter response. The amplitude of the two responses, however, is not the same as that which occurs when the sine-wave patterns are presented to the filters separately. For example, a 2-cycle and a 9-cycle grating were separately used as test stimuli to the analysis program. Then the sum of the 2 cycle and 9 cycle were input. The program output is confined to a response with energy only at 2-cycle and 9-cycle responses, but the relative amplitude of the 2-cycle component is significantly reduced by the presence of the 9-cycle component. This is expected, since the algorithm is set to discover the notch in the spatial frequency energy plot below 9 cycles and above 2 cycles. After the notch has been found, a center and a surround region that enhance the 9-cycle and reduce the 2-cycle-per-degree component are selected. In this example, the stimulus is particularly simple because the local estimates of spatial frequency are constant across the image. When the local estimates differ, the selection of center and surround sizes will vary across the image, leading to yet more complicated behavior.

Role of the Lower Spatial Frequencies

The algorithm provides a general description of the method for extracting two intrinsic images from a single real image: the high frequency estimate of surface reflectance and the low frequency estimate of the ambient illumination. Although I cannot devote much time here to discussing the role of the lower spatial frequency image, its proper utilization is important in the development. The lower spatial frequency image provides an estimate of the ambient lighting image. The human visual system clearly uses ambient lighting information to govern dramatic changes in visual sensitivity. We plan to use the information in the low frequency image to control processes of gain adjustment in estimating lightness and color, in this way performing a function that is directly analogous to light and dark adaptation in the human visual system.

Methods for Assigning Color Descriptors

How may we assign color descriptors to the numerical values that are estimated from the higher spatial frequency components (appropriately corrected in gain by the data in the lower spatial frequency image)? It is useful to consider why one might want to assign such descriptors.

The numerical data at a point in the scene consist of the values in the higher spatial frequency images derived

from the red, green, and blue channels. In a typical system, each of the pixel values takes on one of 2^8 possible values, roughly commensurate with the range of discriminable contrasts available to the human visual system under a fixed state of adaptation. This implies that the numerical description for the three channels at each point can take on any of 2^{24}, or approximately 16 million, values, too many values to provide a comprehensible number of descriptors for human programmers and analysts. So, we must find a method of assigning the numerical values of the filtered outputs to a smaller set of color names, or, more generally, color descriptors.

A Relativistic Method

One simple scheme showing how we might jointly represent the effects of the red and green images on a pixel-by-pixel[7] basis is shown in Figure 5. This plot shows all the pairs of values that may occur at a corresponding pixel in the two colored images.

This first method to reduce the number of values to be assigned is to use the sensitivity limitations of the system. From the value of the ambient illumination image, we will determine a sensitivity for each of the color coordinate directions, just as light adaptation determines how sensitive a human observer is to lights with different wavelength distributions. Because the ambient illumination is characterized by the low spatial frequency terms, the changes in sensitivity will vary slowly across the image, compared with the changes

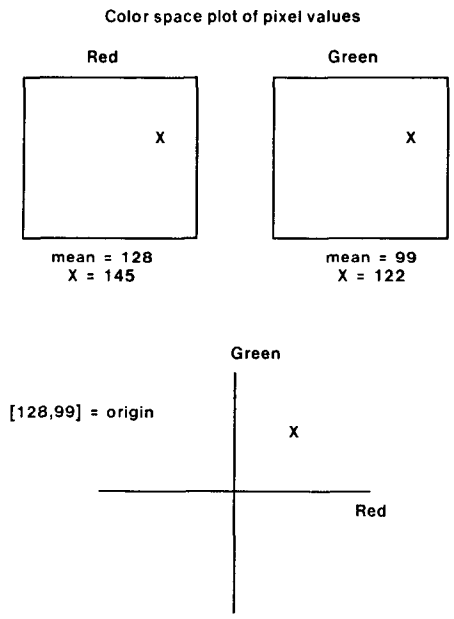

Figure 5. A single pixel, plotted in the red and green images at X, will have different intensity values in each color image. I have indicated on the top half of the image the mean value of the separate images and the value taken on at the pixel location itself. The bottom half of the diagram represents the position of the pixels in color coordinates. The origin of the graph is at the mean values of the red and green images. The position of the single pixel is indicated by the X.

in estimated surface reflectance. Suppose that, for two pixels near one another, we can use a common description of the sensitivities.

Having a sensitivity limit reduces the number of possible color descriptors in several ways. The first is illustrated in the ellipse at the center of Figure 5. For each color direction measured from the lighting conditions—the average value of the lighting is plotted at the origin of the graph—we can determine the minimum amount of deviation that is required before we assign a different name to the two pixels. The elliptical region at the center represents a contour of least deviation from the ambient illumination before a light can generate a detectable signal. Nearby pixels whose red and green values fall within this central elliptical region will not be discriminated simply because none generates a detectable response.

Consider now the case of determining whether two pixels, both outside this "invisible" region, should be assigned different color descriptors. The problem is illustrated in Figure 6. One strategy is to use the full resolution of the system when making discrimination judgments. This amounts to suggesting that if the two lights are separated by a vector difference that can be resolved, then they should be assigned different descriptors. (Notice that the required length of the vector difference varies with the vector direction.) According to this hypothesis, therefore, the only limit on discrimination between visible lights is the limit imposed by the sensitivity derived from the ambient illumination image.

In terms of system performance, this classification method implies that two lights will be discriminable precisely when the difference between them is detectable. The discriminability judgments are predicted by the vector difference between the color representations of the two pixels. If the difference is large enough, in the sense that the difference between them exceeds the limits imposed by the gain adjustment, then we assign their responses to different color descriptors. Otherwise, the pixels are assigned a common descriptor.

The color assignment scheme is relativistic in the sense that we answer only the question of whether two pixels take on the same or different descriptors relative to one another. We do not assign a descriptor absolutely. This kind of assignment scheme would lead naturally to pixel assignments of the form "A is relatively more red and less bright than B." It would not lead naturally to the assignment "A is red."

An Absolute Classification Method

An alternative method of color assignment is to partition the set of possible color responses into different classification regions, as illustrated in Figure 7. Any pixel falling in region "R" would be assigned the color descriptor red, and any pixel in the color region "G" would be assigned the descriptor green.

For this kind of a color assignment scheme, the role of the ambient illumination image is to define the place-

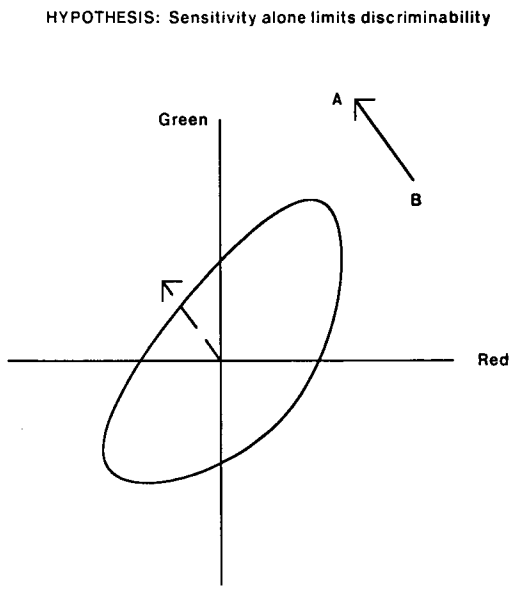

Figure 6. A color space plot of two pixels, A and B, in an image. The (roughly) elliptical region at the center of the graph indicates the sensitivity limits derived from the low frequency information in the image. The positions of the two pixels are determined from the high frequency red and green images. If the full resolution of the system is applied throughout color space, then A and B will be discriminable, and therefore will be assigned different color descriptors. The discriminability limit is illustrated by the dashed vector near the origin, which is simply the vector difference between A and B displaced for comparison with the sensitivity region. Since this vector difference exceeds the limit, the pixels are discriminable.

ment of the lines determining the classification of pixels. For example, when the ambient illumination image contains a large response in the red image, the classification region for assigning the classifier red must include only those pixels with a considerably more than average red component. As the ambient illumination varies, the position of the classifcation regions must shift.

A variety of mechanisms for placing these classification regions have been studied (see, e.g., Duda & Hart, 1973). The method is quite different from the measurement method used in the first scheme. For example, in this method, color discrimination is not well predicted by the vector difference between two pixels. Instead, discrimination must always be predicted relative to the position in color space relative to the classification boundaries.

In the last several years, I have studied human abilities to perform such color discriminations (Wandell, 1984). Surprisingly, the results provide examples for both methods. In discrimination of lights that vary only slowly over time, or of lights that differ only in hue but not in brightness, the relativistic procedure is well confirmed by the data. There is no evidence for the existence of any classification procedure for discrimination of lights. In discrimination of lights that vary more rapidly over time (say, at about 6 Hz of flicker), however, or of lights that fluctuate in both brightness and hue, there is striking evidence for a classification procedure in discrimination of lights. The classification appears to follow the distinction between brightness and hue, in the sense that the discrimination judgments appear to be based on whether one light flickered mainly in brightness and the other in hue. An open question is which discrimination mechanism is used under steady viewing conditions in which eye movements dominate the temporal character of the color difference stimulus.

The conclusion from these studies is that we must conduct computer experiments that compare these methods for color assignment. Presumably, depending upon viewing conditions—conditions that remain to be completely identified—one method will provide better color constancy than the other.

CONCLUSIONS

My hope is that I have managed to convey the general approach that we are taking in building this system. The approach rests on several assumptions. First, we are searching for clues that distinguish surface properties from ambient lighting properties by separating out the relatively high spatial frequency content of the image. This is a hypothesis that we share with Horn (1974), Land (1964), and all others working in the area. Our use of the low spatial frequency image is restricted primarily to regulating the local sensitivity in the colored images, and possibly defining the location of

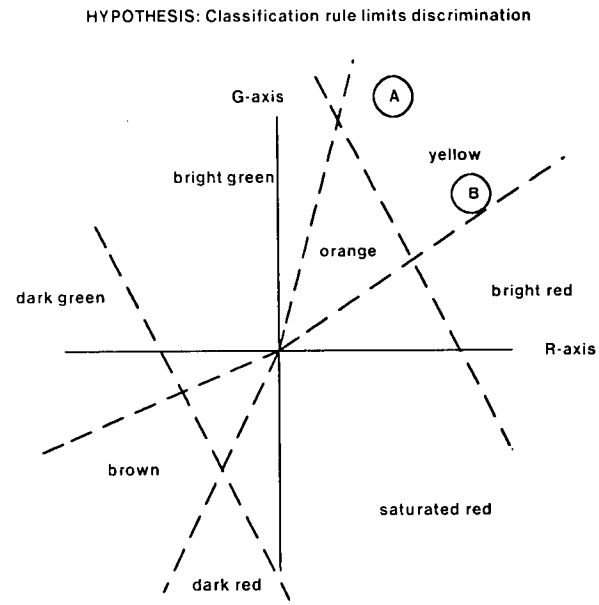

Figure 7. A classification method for assigning color descriptors to pixel values. In this case, the discriminability of pixels A and B is determined by whether or not they fall in different regions of the color space. Suggestive color descriptors have been placed in some of these regions, which are enclosed by dashed lines. The positions of the dashed lines are derived from the low frequency image, and the positions of pixels A and B are determined from the high frequency image. Notice that, in this case, pixels A and B would *not* be discriminable.

categorical boundaries for color descriptor assignment. The two images are meant to capture different aspects of a single image, much in the spirit of intrinsic images (Barrow & Tenenbaum, 1978).

Second, we have implemented a new, flexible filtering algorithm for automatically adjusting the criterion for the "the relatively high" versus "relatively low" spatial frequency content in a local region of an image. This algorithm is somewhat different from the alternative suggestion that calculations should be based upon multiple, narrow, bandpass copies of the image (Marr, 1982). The usefulness of the algorithm will depend upon the statistics of natural images. In particular, the algorithm will be successful only if, over a broad range of images, we regularly find nonmonotonic spatial frequency responses.

Third, we have been using human data to explore the mechanisms that are appropriate for assigning the image values in the relatively higher spatial frequency color images to color descriptors. This aspect of the study is intimately connected to perceptual experiments inquiring about the rules used by human subjects in discriminating colored lights (Wandell, 1984). These experiments have led to the recent discovery that under some viewing conditions (the Gaussian temporal waveform) a simple vector difference between the color image values is predictive of discriminability, whereas under other conditions (the 6-Hz Gabor temporal waveform) the vector difference is not predictive. Instead, a classification scheme based on absolute position in color space provides a better characterization.

The excitement in building a system like this is something that I am sure many of you share. As we psychologists develop new links with computer scientists and become more skilled at using the new tools, our theories and insights will improve. This society is a fine example of how the interaction between computer science and psychology can blossom into useful basic and applied research. I am glad to have been given the opportunity to describe my excitement as these opportunities unfold to us all.

REFERENCES

Barrow, H. G., & Tenenbaum, J. M. (1978). Recovering intrinsic scene characteristics from images. In A. R. Hanson & E. M. Riseman (Eds.), *Computer vision systems.* New York: Academic Press.

Bracewell, R. (1978). *The Fourier transform and its applications.* New York: McGraw-Hill.

Campbell, F. W., & Robson, J. G. (1968). Application of Fourier analysis to the visibility of gratings. *Journal of Physiology, London,* **197**, 551-566.

Duda, R. O., & Hart, P. E. (1973). *Pattern classification and scene analysis.* New York: Wiley.

Horn, B. K. P. (1974). Determining lightness from an image. *Computer Graphics and Image Processing,* **3**, 277-299.

Kuffler, S. W. (1953). Discharge patterns and functional organization of mammalian retina. *Journal of Neurophysiology,* **16**, 37-68.

Land, E. H. (1964). The retinex. *American Scientist,* **52**, 247-264.

Land, E. H., & McCann, J. J. (1971). Lightness and retinex theory. *Journal of the Optical Society of America,* **61**, 1-11.

Marr, D. (1982). *Vision.* San Francisco: Freeman.

Marr, D., & Hildreth, E. C. (1980). Theory of edge detection. *Proceedings of the Royal Society of London, Series B,* **207**, 187-217.

Movshon, J. A., Thompson, I. D., & Tolhurst, D. J. (1978a). Receptive field organization of complex cells in the cat's striate cortex. *Journal of Physiology,* **283**, 79-99.

Movshon, J. A., Thompson, I. D., & Tolhurst, D. J. (1978b). Spatial and temporal contrast sensitivity of neurones in areas 17 and 18 of the cat's visual cortex. *Journal of Physiology,* **283**, 101-120.

Movshon, J. A., Thompson, I. D., & Tolhurst, D. J. (1978c). Spatial summation in the receptive fields of simple cells in the cat's striate cortex. *Journal of Physiology,* **283**, 53-77.

Peichl, L., & Wassle, H. (1979). Size, scatter, and coverage of ganglion cell receptive field centers in the cat retina. *Journal of Physiology, London,* **291**, 117-141.

Wandell, B. A. (1984). *Color measurement and color categorization* (Tech. Rep. No. CR-TBD). Moffett Field, Mountain View, CA: National Aeronautics and Space Administration.

Witkin, A. (1983). Scale-space filtering. *Proceedings of the International Joint Conference on Artificial Intelligence,* **2**, 1019-1022.

NOTES

1. Due to eye movements, there is likely to be a correlation between the effects seen at high spatial frequencies and those seen at high temporal frequencies.

2. To simplify the presentation, I omitted discussion of an important preprocessing stage in which the logarithm of the image is first computed. This step is technically necessary, but its inclusion requires additional formal arguments which do not affect the basic hypothesis and clutter the argument. The details of this calculation will be presented elsewhere.

3. Marr (1982), for example, suggested that this retinal organization was used to generate multiple bandpass versions of the image. Since the information in each image is (approximately) band limited, the images can be (approximately) described by the zero crossings of the image. Marr and his colleagues (e.g., Marr & Hildreth, 1980) have suggested that these are the primary sources of information for defining object boundaries.

4. A more complete analysis of the algorithm for actual images requires a specification of what to do when one finds no notch or, alternatively, several notches. Currently, we are doing a statistical analysis of a large set of images to determine where and when the notches are found. We are also testing various procedures, including leaving the value at locations without notches unspecified, using extrapolation methods between notched frequency values, and fixing the Gaussian sizes in the neighborhood of a notch.

5. For computer vision applications, it is not necessary, of course, to use filters that are differences of Gaussians. Despite their theoretical significance, these types of filters have two disadvantages. First, they require more computer computation than filters based on mixtures of positive and negative unit (i.e., boxcar) weightings. Second, they are not as restricted in their frequency selectivity as filters based on Gabor functions (the product of a Gaussian and a sinusoid). The broader tuning of the difference of Gaussians means that small local notches in the image will not be detected in the pattern of filter responses.

6. The principle of superposition is the defining characteristic of a linear system. It states that when stimulus $s_1(t)$ causes response $r_1(t)$ and stimulus $s_2(t)$ causes $r_2(t)$, then stimulus $s_1(t) + s_2(t)$ must cause response $r_1(t) + r_2(t)$.

7. The word pixel is a shortening of the phrase "picture element" and refers to a point in an image.