J. Webster (ed.), *Wiley Encyclopedia of Electrical and Electronics Engineering*

IMAGE COLOR ANALYSIS

Color image analysis is a complex process. A color image is a multidimensional entity. In addition to the color attributes, an image is sampled in a two-dimensional plane having a quantized value to indicate the tone intensity in the third dimension. The smallest unit in the two-dimensional image plane is called the picture element (*pixel* or *pel*), constituted by the pixel size (width and height) and depth (the number of tone levels). However, the appearance of a color image is much more intriguing than a few physical measures; there are psychological and psychophysical attributes that cannot be measured quantitatively. Therefore, the foremost important factor in the image analysis, color or not, is human vision, because the human being is the ultimate judge of the image quality. Human vision provides the fundamental guidance to digital imaging design, analysis, and interpretation.

Basics of Image Color Analysis

Color images have at least two components, spatial patterns and colorimetric properties. Spatial patterns include inadequate sampling (aliasing, stair casing, etc.), the interaction of color separations (moiré and rosette), and false contouring caused by insufficient tone levels. Colorimetric properties involve color transformation, color matching, color difference, surround, adaption, and so on. The overall image quality is the integration of these factors as perceived by human eyes. Various theories and models are developed to address these problems and requirements. Sampling and quantization theories ensure proper resolution and tone level. Colorimetry provides quantitative measures to specify color stimuli.

Color mixing models and transformation techniques give the ability to move color images through various imaging devices. Color appearance models attempt to account for phenomena, such as chromatic adaptation, that cannot be explained by colorimetry. Device models deal with various device characteristics, such as contone rendering, halftoning, color quantization, and the device dot model for generating appearing color images. A thorough image color analysis should include the human visual model, color measurement, color mixing model, color space transform, color matching, device models, device characteristics, and the color appearance model.

Human Vision Theory. Human vision is initiated from the photoreceptors (rods and cones) in the retina of the eye. Rods and cones respond to difference levels of luminance; rods serve vision at low luminance levels, and cones at higher luminance levels. In the intermediate levels, both rods and cones function. There is only one type of rod receptor with a peak spectral responsivity at approximately 510 nm. There are three types of cone receptors [see Fig. 1 (Ref. 3)], referred to as L, M, and S cones for their sensitivity in the long, middle, and short wavelengths, respectively. The modern theory of color vision is based on trichromatic and opponent color theories, as shown in Fig. 1 (1). The incoming color signal is separated by three types of cones; the neurons of the retina encode the color into opponent signals. The sum of three cone types (L + M + S) produces an achromatic response, while the combinations of two types with the difference of a third type produce the red-green (L − M + S) and yellow-blue (L + M − S) opponent signals (Ref. 1, pp. 22–23).

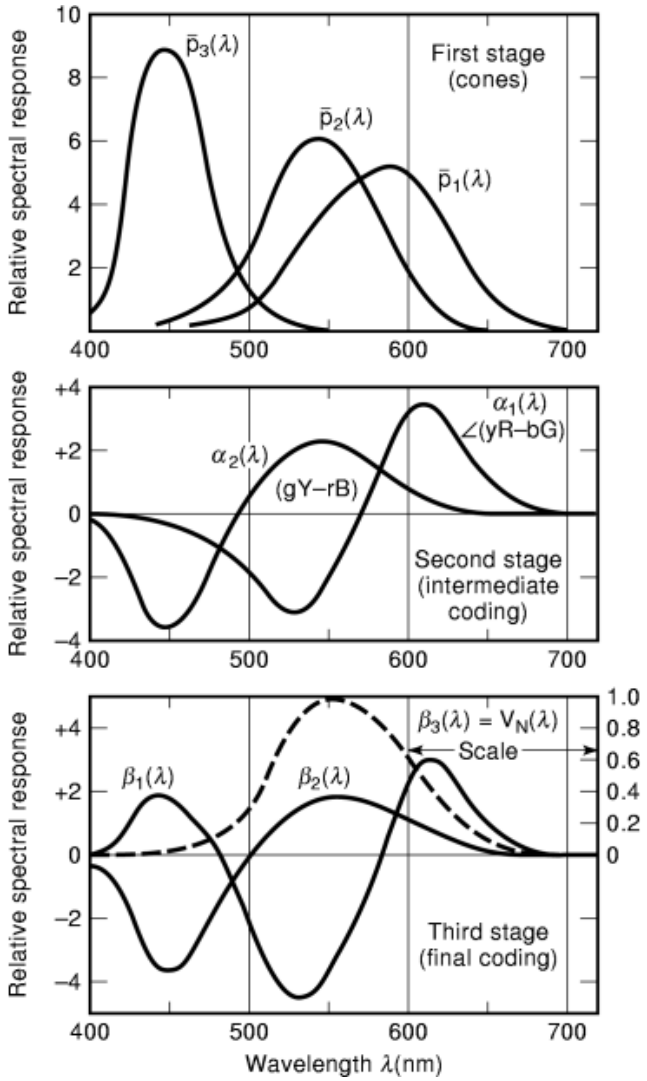

Fig. 1. Schematic illustration of the encoding of cone signals into opponent-color signals in the human visual system.

In addition to static human vision theory, dynamic adaptation is very important in color perception. It includes dark, light, and chromatic adaptations. Dark adaptation is the change in visual sensitivity that occurs when the illumination is decreased, such as when one walks into a darkened theater on a sunny afternoon. At first, the entire theater appears completely dark. But after a few minutes one is able to see objects in the theater, such as the aisles, seats, and people. This happens because the visual system is responding to the lack of illumination by becoming more sensitive and capable of producing a meaningful visual response at the lower illumination level. Light adaptation is a decrease in visual sensitivity upon increases in the overall level of illumination. For example, it occurs when one leaves a darkened theater and walks into a sunny street. In this case, the visual system must become less sensitive and adjust to the normal vision (Ref. 1, pp. 23–28). Chromatic adaptation is achieved through the largely independent adaptive gain control on each of the L,

M, and S cone responses. The gains of the three channels depend on the state of eye adaptation, which is determined by preexposed stimuli and the surround, but independent of the test stimulus (2). An example of chromatic adaptation is viewing a piece of white paper under daylight and then under incandescent light. The paper appears white under both illuminants, despite the fact that the energy reflected from the paper has changed from the blue shade to yellow shade. Chromatic adaptation is the single most important property of the human visual system with respect to understanding and modeling color appearance.

Human visual system (*HVS*) models are attempts to utilize human visual sensitivity and selectivity for modeling and improving the perceived image quality. The HVS is based on the psychophysical process that relates psychological phenomena (color, contrast, brightness, etc.) to physical phenomena (light intensity, spatial frequency, wavelength, etc.). It determines what physical conditions give rise to a particular psychological (perceptual, in this case) condition. A common approach is to study the relationship between stimulus quantities and the perception of the just noticeable difference (*JND*)—a procedure in which the observer is required to discriminate between color stimuli that evoke JNDs in visual sensation (3). To this end, Weber's and Fechner's laws play key roles in the basic concepts and formulations of psychophysical measures. Weber's law states that the JND of a given stimulus is proportional to the magnitude of the stimulus. Fechner built on the work of Weber to derive a mathematical formula, indicating that the perceived stimulus magnitude is proportional to the logarithm of the physical stimulus intensity. Weber fractions in the luminance, wavelength, color purity, and spatial frequency are the foundations of the color difference measure and image texture analysis. Another important foundation is the Stevens power law. Based on the study of over 30 different types of perceptions, Stevens hypothesized that the relationship between perceptual magnitude and stimulus intensity followed a power law with various exponents for different perceptions. The Stevens power law has been used to model many perceptual phenomena and can be found in fundamental aspects of color measurement, such as the relationship between CIEXYZ and CIELAB (Ref. 1, pp. 45–48).

Color Mixing. Many of us are familiar with the mixing of colors from childhood experience. But the proper mixing of colors is not a simple matter. There are many theories and models developed solely for interpreting and guiding the mixing of colors. The simplest color mixing model is the ideal block dye model. Roughly, the visible spectrum can be divided into three bands—the blue band from the end of the ultraviolet light to about 520 nm, the green band from about 520 nm to around 610 nm, and the red band from about 610 nm to the onset of the infrared radiation. Additive primaries of block dyes are assumed to have a perfect reflection at one band of the visible spectrum and a total absorption for the other two bands. Subtractive primaries of block dyes reflect perfectly in two bands and absorb completely in one. Additive and subtractive color mixings are different. Additive color mixing is applied to imaging systems that are light emitters, where the subtractive mixing is for imaging systems that combine unabsorbed light. We use ideal block dyes to illustrate the difference between additive and subtractive color mixings.

Additive color mixing is defined as a color stimulus for which the radiant power in any wavelength interval, small or large, in any part of the spectrum, is equal to the sum of the powers in the same interval of the constituents of the mixture, constituents that are assumed to be optically incoherent (Ref. 3, p. 118). An important characteristic of the additive system is that the object itself is a light emitter (a good example is a color television). Red, green, and blue (*RGB*) are additive primaries. For ideal block dyes, the mixture of two additive primaries produces a subtractive primary. The mixture of the red and green is yellow, the mixture of the green and blue is cyan, and the mixture of the red and blue is magenta. The mixture of all three additive primaries produces a spectrum of total reflection across the whole visible region, giving a white color. These eight colors define the size of all colors that can be rendered by the additive system. They form a cube with three additive primaries as the axes, having strength in the range of [0,1]. Thus, the full-strength primaries have coordinates of (1,0,0) for red, (0,1,0) for green, and (0,0,1) for blue. The origin of the cube (0,0,0), emitting no light, is black. For two-color mixtures, yellow locates at (1,1,0), cyan at (0,1,1), and magenta at (1,0,1), where the three-color mixture is white and locates at (1,1,1). This cube is called the additive color space, and the

volume of this cube is the size of the color gamut. Note that the additive mixing of ideal block dyes behaves like a logical OR operation; the output is 1 when any of the primaries is 1.

Subtractive color mixing is defined as color stimuli for which the radiant powers in the spectra are selectively absorbed by an object such that the remaining spectral radiant power is reflected or transmitted and then received by observers or measuring devices (4). The object itself is not a light emitter, but a light absorber. Cyan, magenta, and yellow (*cmy*) are subtractive primaries. This system is the complement of the additive system. When cyan and magenta block dyes are mixed, cyan absorbs the red reflection of the magenta and magenta absorbs the green reflection of the cyan. This leaves blue as the only nonabsorbing region. Similarly, the mixture of magenta and yellow gives red, and the mixture of cyan and yellow gives green. When all three subtractive primaries are mixed, all components of the incident light are absorbed to give a black color. Thus, the subtractive mixing of block dyes behaves like a logical AND operation; for a given region of the visible spectrum, the output is 1 if all participated primaries are 1 at that spectral region. Similar to the additive system, the subtractive primaries also form a cube with positions occupied by the complementary colors of the additive cube, where the complementary colors is a pair of colors that gives an achromatic hue (gray tone) such as yellow–blue, green–magenta, and cyan–red pairs. A good example of subtractive color mixing is color printing on paper.

Color Spaces and Transforms. A color image is acquired, displayed, and rendered by different imaging devices. For example, an image can be acquired by a flat-bed color scanner, displayed on a color monitor, and rendered by a color printer. Different imaging devices use different color spaces; well-known examples are RGB space for monitors and cmy (or cmyk) space for printers. There are two main types of the color space: device dependent and device independent. Color signals produced or utilized by imaging devices, such as scanners and printers, are device dependent; they are device specific, depending on the characteristics of the device, such as imaging technology, color description, color gamut, and stability. Thus, the same image rendered by different imaging devices may not look the same. In fact, the images usually look different. These problems can be minimized by using some device-independent color spaces. The Commission Internationale de l'Éclairage (*CIE*) developed a series of color spaces using colorimetry that are not dependent on the imaging devices. CIE color spaces provide an objective color measure and are being used as interchange standards for converting between device-dependent color spaces. A brief introduction will be given in the section titled "CIE Color Spaces."

Because various color spaces are used in the color image process, there is a need to convert them. Color space transformation is essential to the transport of color information during image acquisition, display, and rendition. Converting a color specification from one space to another means finding the links of the mapping. Some transforms have a well-behaved linear relationship; others have a strong nonlinear relationship. Many techniques have been developed to provide sufficient accuracy for color space conversion under the constraints of cost, speed, design parameters, and implementation concerns. Generally, these techniques can be classified into four categories. The first one is based on theoretical color mixing models like the Neugebauer equations and Kubelka–Munk theories. With analytical models, one can predict the color specifications using relatively few measurements. The second method, polynomial regression, is based on the assumption that the correlation between color spaces can be approximated by a set of simultaneous equations. Once the equations are selected, a statistical regression is performed to a set of selected points with known color specifications in both source and destination spaces for deriving the coefficients of the equations. The third approach uses a three-dimensional lookup table with interpolation. A color space is divided into small cells; then the source and destination color specifications are experimentally determined for all lattice points. Nonlattice points are interpolated, using the nearest lattice points. The last category comprises approaches such as neural networks and fuzzy logic. Detailed descriptions and formulations of these techniques can be found in Ref. 4.

Color Appearance Phenomena. CIE colorimetry is based on color matching; it can predict color matches for an average observer, but it is not equipped to specify color appearance. Colorimetry is extremely useful if two stimuli are viewed with identical surround, background, size, illumination, surface characteristics,

and so on. If any of these constraints is violated, it is likely that the color match will no longer hold. In practical applications, the constraints for colorimetry are rarely met. The various phenomena that break the simple CIE tristimulus system are the topics of the color appearance model, such as simultaneous contrast, hue change with luminance level (Bezold–Brücke hue shift), hue change with colorimetric purity (Abney effect), colorfulness increase with luminance (Hunt effect), contrast increase with luminance (Stevens effect), color consistency, discounting of the illuminant, and object recognition (Ref. 1, pp. 133–157). A frequently used example to show the appearance phenomenon is a picture containing two identical gray patches, one surrounded by a black background and the other by a white background. The one with the black surround appears lighter. If we cover both backgrounds with a piece of white paper (or any uniform paper), with two square holes for gray patches to show through, then these two gray patches will look exactly the same. This simultaneous contrast also occurs in color images. This example illustrates the need for the color appearance model to account for the appearance phenomena because the color measurement will not be able to tell the difference. The key features of color appearance models are the transform of spectral power distribution to a 3-D signal by cones, chromatic adaptation, opponent color processing, nonlinear response functions, memory colors, and discounting of the illuminant. Color appearance modeling is still an actively researched area, and various models have been proposed. A brief review and an example will be given in the section titled "Color Appearance Model."

Device Characteristics. Color images, either printed on paper or displayed on a monitor, are formed by pixels that are confined in the digital grid. Generally, the digital grid is a square (sometimes rectangular, if the aspect ratio is not 1) ruling in accordance with the resolution of the imaging device; that is, if a printer resolution is 300 dots per inch (*dpi*), then the digital grid has 300 squares per inch. This discrete representation creates problems for the quality of color images in many ways. First, the quality of images is dependent on the sampling frequency, phase relationship, and threshold rule (Ref. 4, pp. 153–156). Then, it is dependent on the arrangement of dots in the digital grid. There are three types of dot placement techniques: dot-on-dot, dot-off-dot, and rotated dot (Ref. 4, pp. 244–247). Each technique has its own characteristics and concerns. A CRT display can be considered as a dot-off-dot technique, in which three RGB phosphor dots are closely spaced without overlapping in a triangular pattern. In printing, a frequently used technique is rotated dots, in which color dots are partially overlapped. The rotated dot is produced by the halftone algorithm. Many halftone algorithms are available to create various pixel arrangements for the purpose of simulating a gray sensation from bilevel imaging devices. At a closer look, the image texture is also dependent on the size and shape of the pixel. Even in the ideal case of the square pixel shape that exactly matches the digital grid, there are still image quality problems; the ideal square pixel will do well on the horizontal and vertical lines, providing sharp and crisp edges, but will do poorly on the angled lines and curves, showing jaggedness and discontinuity.

Most rendering devices, such as CRT and dot matrix, ink jet, and xerographic printers, have a circular pixel shape (sometimes an irregular round shape because of the ink spreading). A circular pixel will never be able to fit exactly the size and shape of the square grid. To compensate for these defects, the diameter of the round pixel is made $\sqrt{2}$ bigger than the digital grid period τ such that dots will touch in diagonal positions. This condition is the minimum requirement for a total area coverage by round pixels in a square grid. Usually, the pixel size is made larger than $\sqrt{2}\tau$ but less than 2τ to provide a higher degree of overlapping. These larger pixels eliminate the discontinuity and reduce the jaggedness even in the diagonal direction. But these larger pixels cause the rendered area per pixel to be more than the area of the digital grid. When adjacent dots overlap, the interaction in the overlapped pixel areas is not simply an additive one, but more resembles a logical OR (as if areas of the paper covered by ink are represented by a 1 and uncovered areas are represented by a 0). That is, parts of the paper that have already been darkened are not made significantly darker by an additional section of a dot being placed on top. The correction for the overlapping of the irregularly shaped and oversized pixels is an important part of the overall color imaging model; the formulation of the circular dot overlap model will be given in the section titled "Dot Overlap Model."

6 IMAGE COLOR ANALYSIS

In addition to the discrete nature of the digital grid, different imaging devices use different imaging technology. The most important rendering methods are the contone and halftone techniques. A contone device is capable of producing analog signals in both spatial and intensity domains, providing continuously changing shades. An example of the true contone imaging is the photographic print. Scanners, monitors, and dye diffusion printers with 8 bit depth (or more) can be considered to be contone devices if the spatial resolution is high enough. The additive color mixing of the contone devices (e.g., monitors) can be addressed by Grassmann's laws, in which the subtractive color mixing of contone devices (e.g., printers) can be interpreted by the Beer–Lambert–Bouguer law and Kubelka–Munk theory. These will be discussed in the section titled "Color Mixing Models."

The halftone process is perhaps the single most important factor in image reproduction for imaging devices with a limited tone level (usually bilevel). Simply stated, halftoning is a printing and display technique that trades area for depth by partitioning an image plane into small areas containing a certain number of pixels; it then modulates the tone density by modulating area to simulate the gray sensation for bilevel devices. This is possible because of limited human spatial contrast sensitivity. At normal viewing distance, persons with correct (or corrected) vision can resolve roughly 8 cycles/mm to 10 cycles/mm. The number of discernible levels decreases logarithmically as the spatial frequency increases. With a sufficiently high frequency, the gray sensation is achieved by integrating over areas at normal viewing distance. This binary representation of the contone image is an illusion; much image information has been lost after the halftoning. Thus, there is a strong possibility of creating image artifacts such as moiré and contouring. Moiré is caused by the interaction among halftone screens, and the contouring is due to an insufficient tone level in the halftone cell. There will be no artifacts if a continuous imaging device is used, such as photographic prints. Recently, printers have been developed that can output a limited number of tone levels (usually levels ≤ 16). The result of this capability creates the so-called multilevel halftoning, which combines the halftone screen with limited tone levels to improve color image quality. Halftoning is a rather complex technique; there are books and lengthy reviews on this topic. Readers interested in learning about this technique can refer to the (Ref. 4, pp. 208–260).

Most color displays are based on a frame buffer architecture, in which the image is stored in a video memory from which controllers constantly refresh the screen. Usually, images are recorded as full color images, where the color of each pixel is represented by tristimuli with respect to the display's primaries and quantized to 8 or 12 bits per channel. A 3-primary and 8-bit device, in theory, can generate $256 \times 256 \times 256 = 16{,}777{,}216$ colors. Often, the cost of high-speed video memory needed to support storage of these full color images on a high-resolution display is not justified. And the human eye is not able to resolve these many colors; the most recent estimate is about 2.28 million discernible colors (5). Therefore, many color display devices limit the number of colors that can be displayed simultaneously to 8, 12, or 16 bits of video memory, allowing 256, 4096, and 65,536 colors, respectively, for the cost consideration. A palettized image, which has only the colors contained in the palette, can be stored in the video memory and rapidly displayed using lookup tables. This color quantization is designed in two successive steps: (1) the selection of a palette and (2) the mapping of each pixel to a color in the palette. The problem of selecting a palette is a specific instance of the vector quantization. If the input color image has N distinct colors and the palette is to have K entries, the palette selection may be viewed as the process of dividing N colors into K clusters in 3-D color space and selecting a representative color for each cluster. Ideally, this many-to-one mapping should minimize perceived color difference (2). The color quantization can also be designed as a multilevel halftoning. For example, the mapping of a full color image to an 8-bit color palette (256 colors) is a multilevel halftoning with 8 levels of red (3 bits), 8 levels of green (3 bits), and 4 levels of blue (2 bits).

Color Description and Formulation

The following sections present mathematical formulations of the CIE color space, color transformation, human visual model, color appearance model, color mixing model, and dot overlap model. These models are developed to address the image color analysis and quality improvement.

CIE Color Spaces. CIE color spaces are defined in CIE Publication 15.2 (6). The most frequently used CIE color spaces are CIEXYZ, CIELUV, and CIELAB. CIEXYZ is visually nonuniform CIE color space; it is built on the spectroscopic foundation together with the human color matching ability. The measured reflectance spectrum of an object P is weighed by the spectra of the color matching functions, $\bar{x}$, $\bar{y}$, and $\bar{z}$, and the standard illuminant Q. Resulting spectra are integrated across the whole visible region to give the tristimulus values:

$$X = k \int P(\lambda)Q(\lambda)\bar{x}(\lambda)\, d\lambda \cong k \Sigma P(\lambda)Q(\lambda)\bar{x}(\lambda)\, \Delta\lambda \qquad (1)$$

$$Y = k \int P(\lambda)Q(\lambda)\bar{y}(\lambda)\, d\lambda \cong k \Sigma P(\lambda)Q(\lambda)\bar{y}(\lambda)\, \Delta\lambda \qquad (2)$$

$$Z = k \int P(\lambda)Q(\lambda)\bar{z}(\lambda)\, d\lambda \cong k \Sigma P(\lambda)Q(\lambda)\bar{z}(\lambda)\, \Delta\lambda \qquad (3)$$

$$k = 100/\Sigma Q(\lambda)\bar{y}(\lambda)\, \Delta\lambda \qquad (4)$$

where X, Y, and Z are tristimulus values of the object, and λ is the wavelength. In practice, the integrals are approximated by finite step summations.

The three-dimensional nature of color suggests plotting the value of each tristimulus component along orthogonal axes. The result is called tristimulus space. The projection of this space to the two-dimensional X and Y plane yields the chromaticity diagram of Fig. 2. Mathematically, this can be expressed as

$$x = X/(X+Y+Z),\ y = Y/(X+Y+Z),\ z = Z/(X+Y+Z)$$

and

$$x + y + z = 1 \qquad (5)$$

where x, y, and z are the chromaticity coordinates. They are the normalization with respect to the sum of tristimulus values. The chromaticity coordinates represent the relative amounts of three stimuli X, Y, and Z required to obtain any color. However, they do not indicate the luminance of the resulting color. The luminance is incorporated into the Y value. Thus, a complete description of a color is given by the triplet (x, y, Y). The 1931 CIE chromaticity diagram is shown in Fig. 2; the boundary of the color locus is the plot of the color matching functions (the spectral colors). The chromaticity diagram provides much useful information, such as the dominant wavelength, complementary color, and purity of a color. The dominant wavelength for a color is obtained by extending the line connecting the color and the illuminant to the spectrum locus. For example, the dominant wavelength for color s_1 in Fig. 2 is 584 nm. The complement of a spectral color is on the opposite side of the line connecting the color and the illuminant used. For example, the complement of the color s_1 is 483 nm under the illuminant D_{65}. If the extended line for obtaining the dominant wavelength intersects the "purple line," the straight line that connects two extreme spectral colors (usually 380 nm and 770 nm), then the color will have no dominant wavelength in the visible spectrum. In this case, the dominant wavelength is specified by the complementary spectral color with a suffix c. The value is obtained by extending a line

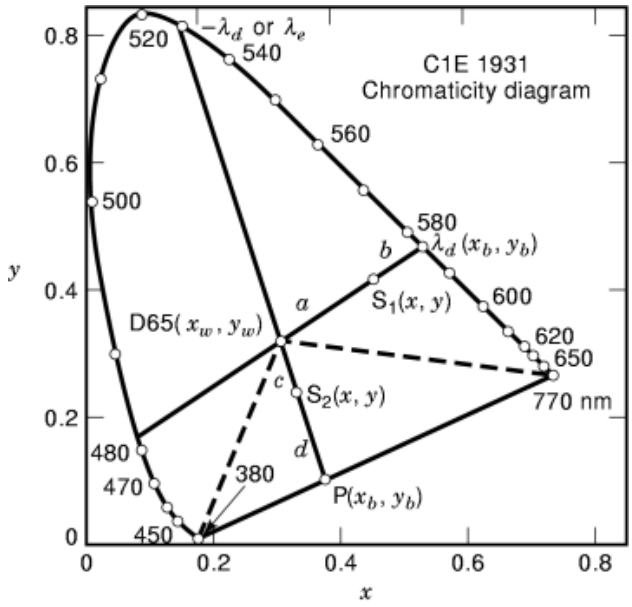

Fig. 2. CIE chromaticity diagram showing the dominant wavelength and purity.

backward through the illuminant to the spectrum locus. For example, the dominant wavelength for color s_2 in Fig. 2 is 530c nm. Pure spectral colors lie on the spectrum locus that indicates a fully saturated 100% purity. The illuminant represents a fully diluted color with a purity of 0%. The purity of intermediate colors is given by the ratio of two distances: the distance from the illuminant to the color and the distance from the illuminant through the color to the spectrum locus or the purple line. In Fig. 2, the purity of color s_1 is $a/(a + b)$ and s_2 is $c/(c + d)$ expressed as a percentage (7).

The CIE chromaticity coordinates of a mixture of two or more colors obey Grassmann's laws (Ref. 3, p. 118). The laws provide the proportionality and additivity to the color mixing:

(1) Symmetry law—If color stimulus A matches color stimulus B, then color stimulus B matches color stimulus A.

(2) Transitivity law—If A matches B and B matches C, then A matches C.

(3) Proportionality law—If A matches B, then αA matches αB, where α is any positive factor by which the radiant power of the color stimulus is increased or reduced while its relative spectral distribution is kept the same.

(4) Additivity law—If A matches B, C matches D, and (A + C) matches (B + D), then (A + D) matches (B + C), where (A + C), (B + D), (A + D), (B + C) denote, respectively, additive mixtures of A and C, B and D, A and D, and B and C.

Visually uniform color spaces of CIELAB and CIELUV are derived from nonlinear transforms of CIEXYZ; they describe colors using opponent type axes relative to a given absolute white point reference. The CIELUV is a transform of the 1976 UCS chromaticity coordinate u', v', Y.

$$L^* = 116\, f(Y/Y_n) - 16 \qquad (6)$$

$$u^* = 13L^*(u' - u'_n) \tag{7}$$

$$v^* = 13L^*(v' - v'_n) \tag{8}$$

where

$$u' = 4X/(X + 15Y + 3Z) \tag{9}$$

$$v' = 9Y/(X + 15Y + 3Z) \tag{10}$$

CIELAB is the transform of 1931 CIEXYZ:

$$a^* = 500[f(X/X_n) - f(Y/Y_n)] \tag{11}$$

$$b^* = 200[f(Y/Y_n) - f(Z/Z_n)] \tag{12}$$

and

$$\begin{aligned} f(t) &= t^{1/3} \quad 1 \geq t > 0.008856 \\ &= 7.787t + (16/116) \quad 0 \leq t \leq 0.008856 \end{aligned} \tag{13}$$

where L$*$ is the lightness and X_n, Y_n, Z_n are tristimulus values of the illuminant. It is sometimes desirable to identify the components of a color difference as correlates of perceived hue, colorfulness, and lightness. The relative colorfulness, or chroma C^*_{ab}, and hue angle h_{ab} are defined as

$$C^*_{ab} = (a^{*2} + b^{*2})^{1/2} \tag{14}$$

$$h_{ab} = \tan^{-1}(b^*/a^*) \tag{15}$$

The color difference is defined as the Euclidean distance in the three-dimensional space. For CIELAB space, the color difference ΔE^*_{ab} is

$$\Delta E^*_{ab} = (\Delta L^{*2} + \Delta a^{*2} + \Delta b^{*2})^{1/2} \tag{16}$$

The just noticeable color difference is approximately 0.5 to 1.0 ΔE^*_{ab} unit.

Device-Dependent Color Transform. Again, the simplest model for converting from RGB to cmy is the block dye model that a subtractive primary is the total reflectance minus the reflectance of corresponding additive primary. Mathematically, this is expressed as

$$c = 1 - R, \quad m = 1 - G, \quad y = 1 - B \tag{17}$$

where the total reflectance is normalized to 1. This model is simple with minimum computation cost, but it is grossly inadequate. This is because none of the additive and subtractive primaries come close to the block dye

spectra; there are substantial unwanted absorptions in cyan and magenta primaries. To minimize this problem and improve the color fidelity, a linear transform is applied to input *RGB*:

$$R' = k_{11}R + k_{12}G + k_{13}B \tag{18}$$

$$G' = k_{21}R + k_{22}G + k_{23}B \tag{19}$$

$$B' = k_{31}R + k_{32}G + k_{33}B \tag{20}$$

Then the corrected additive primary, $R'G'B'$, is removed from the total reflectance to obtain the corresponding subtractive color.

$$c = 1 - R', \quad m = 1 - G', \quad y = 1 - B' \tag{21}$$

The coefficients k_{ij} of the transfer matrix are determined in such a way that the resulting *cmy* have a closer match to the original or desired output. Generally, the diagonal elements have values near 1 and off-diagonal elements have small values, serving as the adjustment to fit the measured results. However, this correction is still not good enough because it is a well-known fact that the RGB to cmy conversion is not linear. A better way is to use three nonlinear functions or lookup tables for transforming *RGB* to $R'G'B'$. More realistic color mixing models are given in the section titled "Color Mixing Models."

Color Gamut Mapping. One major problem in the color transform is the color gamut mismatch, where the sizes of the source and destination color spaces do not match. This means that some colors in the source space are outside the destination space and cannot be faithfully reproduced. Conversely, some colors in the destination space are outside the source space and will never be used. In this case, the color gamut mapping is needed to bring the out-of-gamut colors within the gamut of the destination color space. The color gamut mapping is a technique that acts on a color space to transform a color from one position to another for the purpose of enhancing the appearance or preserving the appearance of an image. A special case is the tone reproduction curve, which maps input lightness values to output device values. Conventionally, color gamut mapping is applied to all components of a color space. It is one of the difficult problems in the color reproduction; many approaches have been developed. In general, they consist two components—a directional strategy and a mapping algorithm. The directional strategy decides which color space to perform the color mapping, determines the reference white point, computes the gamut surface, and decides what color channel to hold constant and what channel to compress first or to do them simultaneously. The mapping algorithm provides the means for transforming input colors to the surface or inside of the output color space. Frequently used mapping algorithms are clipping, linear compression, and nonlinear compression. Clipping is characterized by the mapping of all out-of-gamut colors to the surface of the output color space, and no change is made to input colors that are inside the output gamut. This will map multiple colors into the same point that may lose fine details and cause shading artifacts in some images. The linear compression can be expressed as

$$L^*_{\mathrm{o}} = L^*_{\mathrm{o,min}} + [(L^*_{\mathrm{o,max}} - L^*_{\mathrm{o,min}})/(L^*_{\mathrm{i,max}} - L^*_{\mathrm{i,min}})]L^*_{\mathrm{i}} \tag{22}$$

where L^*_{i} and L^*_{o} are the lightness for the input and output, respectively, the subscript max indicates the maximum value, and the subscript min indicates the minimum value. A similar expression can be obtained for chroma at a constant hue angle. Nonlinear compression, in principle, can have any functions. In practice, the midtone is usually retained, while the shadow and highlight regions are nonlinearly mapped by using

a tangent to a linear function near the gray axis, then approaching a horizontal line near the maximum output saturation. This is designed to minimize the loss of details that occur with clipping while retaining the advantage of reproducing most of the common gamut colors accurately. Linear and nonlinear mappings maintain the shading but can cause an objectionable decrease in saturation.

Color gamut mapping is most effective in color spaces in which the luminance and chrominance are independent to each other, such as CIELAB and CIELUV. Popular directional strategies for CIELAB or CIELUV are (1) sequential $L*$ and $C*$ mapping at constant h, (2) simultaneous $L*$ and $C*$ mapping to a fixed anchor point at constant h, (3) simultaneous $L*$, $C*$, and h mapping to minimize the color difference between two gamuts, and (4) image-dependent gamut mapping. More information on color gamut mapping can be found on pp. 22–28 of Ref. 4.

Human Visual Responses

Human vision response is not uniform with respect to spatial frequency, luminance level, object orientation, visible spectrum wavelength, and so on. The behavior of the human visual system can be studied by employing the Weber and Fechner laws.

Weber and Fechner Laws. Weber proposed a general law of sensory thresholds that the JND between one stimulus and another is a constant fraction of the first. This fraction ω, called the Weber fraction, is constant within any one sense modality for a given set of observing conditions (Ref. 3, pp. 489–490). The size of the JND, measured in physical quantities (e.g., radiance or luminance) for a given attribute, depends on the magnitude of the stimuli involved. Generally, the greater the magnitude of the stimuli, the greater the size of the JND. The mathematical formulation of this empirical fact is given in Eq. (23):

$$\Delta L/(L + L_0) = \omega \tag{23}$$

where ΔL is the increment that must be added to the magnitude L of the given stimulus to be just noticeable. The quantity L_0 is a constant that can be viewed as the internal noise.

Fechner reasoned that the JNDs must represent a change in sensation measure Φ. Thus, he postulated that JNDs are equal increments $\Delta\Phi$ in sensation magnitude at all stimulus magnitudes:

$$\Delta\Phi = \omega'[\Delta L/(L + L_0)] \tag{24}$$

where ω' is a constant specifying an appropriate unit of the sensation-magnitude increment. By treating $\Delta\Phi$ and ΔL as differentials, $d\Phi$ and dL, respectively, Eq. (24) can be integrated:

$$\Phi(L) = \int d\Phi = \omega' \int dL/(L + L_0) = \mu + \nu \log(L + L_0) \tag{25}$$

Equation (25) relates the sensation measure $\Phi(L)$ by a logarithmic function of the stimulus L, measured in terms of a physical unit. The quantities μ and ν are constants. The relation expressed by Eq. (25) is known as the Fechner law Ref. 3, pp. 489–490. The implication of Eq. (25) is that a perceptual magnitude may be determined by summing JNDs. The Fechner law of Eq. (25) suggests that quantization levels should be spaced logarithmically in reflectance (i.e., uniformly in density domain).

Weber's and Fechner's laws have been used to form scales of color differences. Quantitative studies address thresholds for a single attribute of color (e.g., brightness, hue, and colorfulness). Important ones are discussed next.

The Weber fraction $\Delta L/L$ as the function of the luminance indicates that the Weber fraction is not constant over the range of luminances studied. Because the threshold Weber fraction is the reciprocal of sensitivity, the sensitivity drops quickly as the lumminance becomes weak.

Similar determinations have been made of the threshold difference in wavelength necessary to see a hue difference. Results indicate that the sensitivity to hue is much lower at both ends of the visible spectrum. Approximately, the human visual system can distinguish between colors with dominant wavelengths differing by about 1 nm in the blue–yellow region, but near the spectrum extremes a 10 nm separation is required (8). The threshold measurements of the purity show that the purity threshold varies considerably with wavelength. A very marked minimum occurs at about 570 nm; the number of JND steps increases rapidly on either side of this wavelength (9).

The visual contrast sensitivity with respect to the spatial frequency is known as the contrast sensitivity function (*CSF*). The CSF is very important in color image analysis. The image contrast is the ratio of the local intensity to the average image intensity (10). Visual contrast sensitivity describes the signal properties of the visual system near threshold contrast. For sinusoidal gratings, contrast C is defined by the Michelson contrast as (11)

$$C = (L_{\max} - L_{\min})/(L_{\max} + L_{\min}) = \Lambda_{\mathrm{L}}/L \qquad (26)$$

where $L_{\max}$ and $L_{\min}$ are the maximum and minimum luminances, respectively, Λ_{L} is the luminance amplitude, and L is the average luminance. Contrast sensitivity is the reciprocal of the contrast threshold.

A CSF describes contrast sensitivity for sinusoidal gratings as a function of spatial frequency expressed in cycles per degree of the visual angle. Typical human contrast sensitivity curves are given in Fig. 3 (Ref. 3); the horizontal axis is spatial frequency and the vertical axis is contrast sensitivity—namely, $\log(1/C) = -\log C$, where C is the contrast of the pattern at the detection threshold. The CSF of Fig. 3 shows two features. First, there is a fall-off in sensitivity as the spatial frequency of the test pattern increases, indicating that the visual pathways are insensitive to high-spatial-frequency targets. In another words, the human vision has a lowpass filtering characteristic. Second, there is no improvement of sensitivity at low spatial frequencies; there is even a loss of contrast sensitivity at lower spatial frequencies for luminance CSF. This behavior is dependent on the background intensity; it is more pronounced at higher background. Under scotopic conditions of low illumination, the luminance CSF curve is lowpass and peaks near 1 cpd. On intense photopic backgrounds, CSF curves are bandpass and peaks near 8 cpd. Chromatic CSFs show a lowpass characteristic and have much lower cutoff frequencies than the luminance channel (see Fig. 3), where the blue–yellow opponent color pair has the lowest cutoff frequency. This means that the human visual system is much more sensitive to fine spatial changes in luminance than it is to fine spatial changes in chrominance. Images with spatial degradations in the luminance will often be perceived as blurry or not sharp, while similar degradations in the chrominance will usually not be noticed. This property has been utilized in practice for transmitting color signals in that few bits and lower resolution are used for chrominance channels to reduce the bandwidth (12,13). It is also being used in building the color HVS. For the oblique effect, the visual acuity is better for gratings oriented at 0° and 90° and least sensitive at 45°. This property has been used in printing for designing halftone screens (14). Moreover, the CSF has been used to estimate the resolution and tone level of imaging devices. An upper limit of the estimate is about 400 dpi with a tone step of 200 levels. In many situations, much lower resolutions and tone levels still give good image quality. The high-frequency cutoff and exponential relationship with spatial frequency form the basis for various of HVS models.

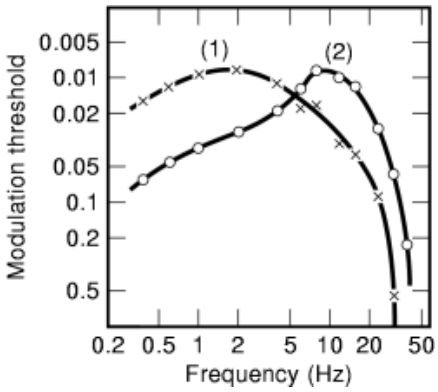

Fig. 3. Spatial contrast sensitivity functions for luminance and chromatic contrast.

Color Appearance Model

A color appearance model is any model that includes predictors of at least the relative color appearance attributes of lightness, chroma, and hue. To have reasonable predictors of these attributes, the model must include at least some form of a chromatic-adaptation transform. All chromatic-adaptation models have their roots in the von Kries hypothesis. The modern interpretation of the von Kries model is that each cone has an independent gain control, as expressed in Eq. (27):

$$\begin{bmatrix} L_a \\ M_a \\ S_a \end{bmatrix} = \begin{bmatrix} 1/L_{max} & 0 & 0 \\ 0 & 1/M_{max} & 0 \\ 0 & 0 & 1/S_{max} \end{bmatrix} \begin{bmatrix} L \\ M \\ S \end{bmatrix} \qquad (27)$$

where L, M, and S are the initial cone responses; L_a, M_a, and S_a are the cone signals after adaptation; and L_{max}, M_{max}, and S_{max} are the cone responses for the scene white or maximum stimulus. This definition enables the uniform color spaces CIELAB and CIELUV to be considered as color appearance models (Ref. 1, p. 217). More complex models that include predictors of brightness, colorfulness, and various luminance-dependent effects, mentioned in the color appearance phenomena, are Nayatani (15), Hunt (16), RLAB (17), ATD (18), and LLAB (19) models. Nayatani's and Hunt's models, evolved from many decades of studies, are capable of predicting the full array of color appearance parameters. They are complete and complex and are not suitable to discuss in this short article. ATD, developed by Guth, is a different type of model than others. It is aimed at describing color vision. RLAB and LLAB are similar in structure. They are the extension of CIELAB and are relatively simple. We present a simple version of the CIE 1997 Interim Color Appearance Model CIECAM97s; the descriptions of other models can be found in Refs. 15 to 16,17,18,19 and in the book *Color Appearance Models* by Fairchild (1).

The formulation of CIECAM97s builds upon the work of many researchers in the field of the color appearance (Ref. 1, pp. 384–389). It begins with a conversion from CIEXYZ to spectrally sharpened cone

responses RGB via the following matrix multiplication for both sample and white point:

$$[R\,G\,B]^{\mathrm{T}} = \boldsymbol{M}_{\mathrm{B}}[X/Y \quad Y/Y \quad Z/Y]^{\mathrm{T}}$$

$$\boldsymbol{M}_{\mathrm{B}} = \begin{bmatrix} 0.8951 & 0.2664 & -0.1614 \\ -0.7502 & 1.7135 & 0.0367 \\ 0.0389 & -0.0685 & 1.0296 \end{bmatrix} \tag{28}$$

The next step is the chromatic-adaptation transform, which is a modified von Kries adaptation with an exponential nonlinearity added to the short-wavelength blue channel:

$$R_{\mathrm{c}} = [\varepsilon(1.0/R_{\mathrm{w}}) + 1 - \varepsilon]\,R \tag{29}$$

$$G_{\mathrm{c}} = [\varepsilon(1.0/G_{\mathrm{w}}) + 1 - \varepsilon]\,G \tag{30}$$

$$B_{\mathrm{c}} = [\varepsilon(1.0/B_{\mathrm{w}}^{\eta}) + 1 - \varepsilon]\,|B|^{\eta} \tag{31}$$

$$\eta = (B_{\mathrm{w}}/1.0)^{0.0834} \tag{32}$$

$$\varepsilon = F - F/[1.0 + 2L_{\mathrm{A}}^{1/4} + L_{\mathrm{A}}^{2}/300] \tag{33}$$

where the ε factor is used to specify the degree of adaptation. The value of ε is set equal to 1.0 for complete adaptation, 0.0 for no adaptation. It takes no intermediate values for various degrees of incomplete chromatic adaptation. If B happens to be negative, then B_{c} is also set to be negative. Similar transformations are made for the source white point. Various factors, including a background induction factor g, the background and chromatic brightness induction factors, N_{bb} and N_{cb}, and the base exponential nonlinearity z, are computed for later calculations.

$$k = 1/(5L_{\mathrm{A}} + 1) \tag{34}$$

$$F_{\mathrm{L}} = 0.2k^{4}(5L_{\mathrm{A}}) + 0.1(1 - k^{4})^{2}(5L_{\mathrm{A}})^{1/3} \tag{35}$$

$$g = Y_{\mathrm{b}}/Y_{\mathrm{w}} \tag{36}$$

$$N_{\mathrm{bb}} = N_{\mathrm{cb}} = 0.725(1/g)^{0.2} \tag{37}$$

$$z = 1 + F_{\mathrm{L}}g^{1/2} \tag{38}$$

The postadaptation signals for both sample and source white are transformed from the sharpened cone responses to the Hunt–Pointer–Estevez cone responses; then nonlinear response compressions are performed by using Eqs. (39)to (42).

$$[R' \quad G' \quad B']^{\mathrm{T}} = M_{\mathrm{H}} M_{\mathrm{B}}^{-1} [R_{\mathrm{c}} Y \quad G_{\mathrm{c}} Y \quad B_{\mathrm{c}} Y]^{\mathrm{T}} \tag{39}$$

$$M_{\mathrm{H}} = \begin{bmatrix} 0.38971 & 0.68898 & -0.07868 \\ -0.22981 & 1.18340 & 0.04641 \\ 0.0 & 0.0 & 1.00 \end{bmatrix}$$

$$R'_{\mathrm{a}} = [40(F_{\mathrm{L}} R'/100)^{0.73}]/[(F_{\mathrm{L}} R'/100)^{0.73} + 2] + 1 \tag{40}$$

$$G'_{\mathrm{a}} = [40(F_{\mathrm{L}} G'/100)^{0.73}]/[(F_{\mathrm{L}} G'/100)^{0.73} + 2] + 1 \tag{41}$$

$$B'_{\mathrm{a}} = [40(F_{\mathrm{L}} B'/100)^{0.73}]/[(F_{\mathrm{L}} B'/100)^{0.73} + 2] + 1 \tag{42}$$

The appearance correlates, a, b, and hue angle, h, are calculated as follows:

$$a = R'_{\mathrm{a}} - 12G'_{\mathrm{a}}/11 + B'_{\mathrm{a}}/11 \tag{43}$$

$$b = (1/9)(R'_{\mathrm{a}} + G'_{\mathrm{a}} - 2B'_{\mathrm{a}}) \tag{44}$$

$$h = \tan^{-1}(b/a) \tag{45}$$

The achromatic response Ä is calculated by using Eq. (46) for both sample and white. The lightness J is calculated from the achromatic signals of the sample and white using Eq. (47):

$$\ddot{A} = [2R'_{\mathrm{a}} + G'_{\mathrm{a}} + (1/20)B'_{\mathrm{a}} - 2.05]N_{\mathrm{bb}} \tag{46}$$

$$J = 100\,(\ddot{A}/\ddot{A}_{\mathrm{w}})^{\sigma Z} \tag{47}$$

where σ is the constant for the impact of surround; $\sigma = 0.69$ for average surround, 0.59 for dim surround, 0.525 for dark surround, and 0.41 for cut-sheet transparencies. Other appearance correlates, such as brightness, saturation, chroma, and colorfulness, can also be calculated.

Color Mixing Models

Roughly speaking, two types of color mixing theories exist. The first is formulated for the halftone process, such as the Neugebauer equations, Murray–Davies equation, and Yule–Nielsen model. The other type is based on the light absorption, such as the Beer–Lambert–Bouguer law and Kubelka–Munk theories, for subtractive color mixing.

Neugebauer Equations. In mixing three subtractive primaries, Neugebauer recognized that there are eight dominant colors—namely, white, cyan, magenta, yellow, red, green, blue, and black—for constituting any color halftone prints. A given color is perceived as the integration of these eight Neugebauer dominant colors. The incident light reflected by one of the eight colors is equal to the reflectance of that color multiplied by its area coverage. The total reflectance is the sum of all eight colors weighed by the corresponding area coverages (20,21). Therefore, it is based on broadband color mixing. An example of a three-primary Neugebauer equation is

$$P = A_w P_w + A_c P_c + A_m P_m + A_y P_y + A_r P_r + A_g P_g + A_b P_b + A_{3p} P_{3p} \quad (48)$$

where P is the total reflectance of the R, G, or B primary. The spectral powers P_w, P_c, P_m, P_y, P_r, P_g, P_b, and P_{3p} are the reflectance of the paper, cyan, magenta, yellow, red, green, blue, and black, respectively, measured with a given primary. Neugebauer employed Demichel's dot overlap model for the area of each dominant color. The areas are computed from the halftone dot areas of three primaries, a_c, a_m, and a_y, to give the areas of all eight dominant colors A_w, A_c, A_m, A_y, A_r, A_g, A_b, and A_{3p} (Ref. 4, pp. 34–42).

$$\begin{aligned} A_w &= (1-a_c)(1-a_m)(1-a_y) \\ A_c &= a_c(1-a_m)(1-a_y) \\ A_m &= a_m(1-a_c)(1-a_y) \\ A_y &= a_y(1-a_c)(1-a_m) \\ A_r &= a_m a_y(1-a_c) \\ A_g &= a_c a_y(1-a_m) \\ A_b &= a_c a_m(1-a_y) \\ A_{3p} &= a_c a_m a_y \end{aligned} \quad (49)$$

Neugebauer's model provides a general description of the halftone color mixing; it predicts the resulting red, green, and blue reflectances or tristimulus values *XYZ* from given dot areas in the print. In practical applications, one wants the *cmy* (or *cmyk*) dot areas for a given color. This requires solving the inverted Neugebauer equations; that is, specifing the dot areas of individual inks required to produce the desired *RGB* or *XYZ* values. The inversion is not trivial because the Neugebauer equations are nonlinear. The exact analytical solutions to the Neugebauer equations have not yet solved. Numerical solutions by computer have been reported (22,23). Pollak has worked out a partial solution by making assumptions and approximations (24,25,26).

The three-primary expression of Eq. (48) can readily be expanded to four primaries by employing the four-primary fractional area expressions given by Hardy and Wurzburg (27):

$$\begin{aligned} P = {} & A'_w P_w + A'_c P_c + A'_m P_m + A'_y P_y + A'_r P_r + A'_g P_g + A'_b P_b \\ & + A'_{3p} P_{3p} + A'_k P_k + A'_{ck} P_{ck} + A'_{mk} P_{mk} + A'_{yk} P_{yk} + A'_{rk} P_{rk} \\ & + A'_{gk} P_{gk} + A'_{bk} P_{bk} + A'_{3pk} P_{3pk} \end{aligned} \quad (50)$$

where

$$
\begin{aligned}
A'_{\mathrm{w}} &= (1-a_{\mathrm{c}})(1-a_{\mathrm{m}})(1-a_{\mathrm{y}})(1-a_{\mathrm{k}}) \\
A'_{\mathrm{c}} &= a_{\mathrm{c}}(1-a_{\mathrm{m}})(1-a_{\mathrm{y}})(1-a_{\mathrm{k}}) \\
A'_{\mathrm{m}} &= a_{\mathrm{m}}(1-a_{\mathrm{c}})(1-a_{\mathrm{y}})(1-a_{\mathrm{k}}) \\
A'_{\mathrm{y}} &= a_{\mathrm{y}}(1-a_{\mathrm{c}})(1-a_{\mathrm{m}})(1-a_{\mathrm{k}}) \\
A'_{\mathrm{r}} &= a_{\mathrm{m}}a_{\mathrm{y}}(1-a_{\mathrm{c}})(1-a_{\mathrm{k}}) \\
A'_{\mathrm{g}} &= a_{\mathrm{c}}a_{\mathrm{y}}(1-a_{\mathrm{m}})(1-a_{\mathrm{k}}) \\
A'_{\mathrm{b}} &= a_{\mathrm{c}}a_{\mathrm{m}}(1-a_{\mathrm{y}})(1-a_{\mathrm{k}}) \\
A'_{3\mathrm{p}} &= a_{\mathrm{c}}a_{\mathrm{m}}a_{\mathrm{y}}(1-a_{\mathrm{k}}) \\
A'_{\mathrm{k}} &= a_{\mathrm{k}}(1-a_{\mathrm{c}})(1-a_{\mathrm{m}})(1-a_{\mathrm{y}}) \\
A'_{\mathrm{ck}} &= a_{\mathrm{c}}a_{\mathrm{k}}(1-a_{\mathrm{m}})(1-a_{\mathrm{y}}) \\
A'_{\mathrm{mk}} &= a_{\mathrm{m}}a_{\mathrm{k}}(1-a_{\mathrm{c}})(1-a_{\mathrm{y}}) \\
A'_{\mathrm{yk}} &= a_{\mathrm{y}}a_{\mathrm{k}}(1-a_{\mathrm{c}})(1-a_{\mathrm{m}}) \\
A'_{\mathrm{rk}} &= a_{\mathrm{m}}a_{\mathrm{y}}a_{\mathrm{k}}(1-a_{\mathrm{c}}) \\
A'_{\mathrm{gk}} &= a_{\mathrm{c}}a_{\mathrm{y}}a_{\mathrm{k}}(1-a_{\mathrm{m}}) \\
A'_{\mathrm{bk}} &= a_{\mathrm{c}}a_{\mathrm{m}}a_{\mathrm{k}}(1-a_{\mathrm{y}}) \\
A'_{3\mathrm{pk}} &= a_{\mathrm{c}}a_{\mathrm{m}}a_{\mathrm{y}}a_{\mathrm{k}}
\end{aligned}
\tag{51}
$$

For four-primary system, the analytic inversion is even more formidable because there are four unknowns (a_{c}, a_{m}, a_{y}, and a_{k}) and only three measurable quantities (either *RGB* or *XYZ*). The practical approach is to constrain on the black printer and then seek solutions on *cmy* that combine with black to produce a target color.

Murray–Davies Equation. The Murray–Davies equation derives the reflectance via the absorption of halftone dots. In a unit area, if the solid ink reflectance is P_{s}, then the absorption by halftone dots is $(1 - P_{\mathrm{s}})$ weighed by the dot area coverage A. The reflectance P of a unit halftone area is the white reflectance, which is 1 or the reflectance of the medium P_{w}, subtracting the dot absorptance (28).

$$
P = 1 - A(1 - P_{\mathrm{s}}) \quad \text{or} \quad P = P_{\mathrm{w}} - A(P_{\mathrm{w}} - P_{\mathrm{s}}) \tag{52}
$$

Equation (52) is usually expressed in the density domain by using the density-reflectance relationship of $D = -\log P$ to give

$$
D = -\log[1 - A(1 - 10^{-D_{\mathrm{s}}})] \tag{53}
$$

The Murray–Davies equation is often used to determine the area coverage by measuring the reflectance of the solid and halftone step wedges.

Yule–Nielsen Model. From the study of the halftone process, Yule and Nielsen pointed out that light does not emerge from the paper at the point where it entered. They estimated that between one-fourth and one-half of the light that enters through a white area will emerge through a colored area, and vice versa. Based on this observation, Yule and Nielsen took light penetration and scattering into consideration; a fraction of light, $A\,(1 - T_{\mathrm{i}})$, is absorbed by the ink film on its way into the substrate, where T_{i} is the transmittance of the ink film. After passing through the ink, the remaining light, $1 - A\,(1 - T_{\mathrm{i}})$, in the substrate is scattered and

reflected. It is attenuated by a factor P_w when the light reaches the air-ink-paper interface. On its way out, this fraction of light is absorbed again by the ink film; the second absorption introduces a power of 2 to the equation. Finally, the light is corrected by the surface reflection r_s at the air-ink interface (29). Summing all these effects together, we obtain

$$P = r_s + P_w(1 - r_s)[1 - A(1 - T_i)]^2 \qquad (54)$$

The Yule–Nielsen model is a simplified version of a complex phenomenon involving light, colorants, and substrate. For example, the model does not take into account the facts that, in real systems, the light is not completely diffused by the paper and is internally reflected many times within the substrate and ink film. To compensate for these deficiencies, Yule and Nielsen made the square power of Eq. (54) into a variable, n, for fitting the data. With this modification, Eq. (54) becomes

$$P = r_s + P_w(1 - r_s)[1 - A(1 - T_i)]^n \qquad (55)$$

Assuming that the surface reflection r_s can be neglected and the reflectance is measured relative to the substrate, we obtain the expression

$$D = -n \log[1 - A(1 - 10^{-D_s/n})] \qquad (56)$$

in the form of the optical density D. Note the similarity of Eq. (56) to the Murray–Davies equation [Eq. (53)]. Equation (56) is very popular and is often used to determine the ink area coverage or dot gain. Generally, the effective transmittance, $T_i(\lambda)$, of mixed inks is derived from a nonlinear combination of the primary ink transmittances measured in solid area coverage.

Clapper–Yule Model. After formulating the Yule–Nielson model, Yule then worked with Clapper to develop an accurate account of the halftone process from a theoretical analysis of multiple scattering, internal reflections, and ink transmissions (30,31). In this model, the light is being reflected many times from the surface, within the ink and substrate, and by the background. The total reflected light is the sum of light fractions that emerge after each internal reflection cycle.

$$P = k_s + \{\rho_e(1 - r_s)\rho_r(1 - A + AT_i)^2\}/ \{1 - (1 - \rho_e)\rho_r(1 - A + AT_i^2)\} \qquad (57)$$

where

k_s = the specular component of the surface reflection
ρ_r = a fraction of light that is reflected at the bottom of the substrate
ρ_e = a fraction of light that is emerged at the top of the substrate

Halftone color mixing models of the Murray–Davies, Yule–Nielsen, and Clapper–Yule show a progressively more thorough consideration of the halftone printing process with respect to light absorption, transmission, and scattering induced by the ink-paper interaction.

Beer–Lambert–Bouguer Law. The Beer–Lambert–Bouguer law is based on light absorption; thus it is a subtractive color mixing theory. The Beer–Lambert–Bouguer law relates the light intensity to the quantity of the absorbant based on the proportionality and additivity of the colorant absorptivity (Ref. 4, pp. 47–48). The intensity, I, of a monochromatic light attenuated by the absorption of the medium is proportional to its

intensity:

$$dI(\lambda)/dw = -K(\lambda)wqI(\lambda) \tag{58}$$

where

$K(\lambda)$ = absorption coefficient of the absorbent at the wavelength λ
w = the length of light path traversing through the absorbent
q = the concentration of the absorbent

Integrating this differential equation over the entire thickness of the medium gives

$$P(\lambda) = I(\lambda)/I_o(\lambda) = \exp[-K(\lambda)wq]$$

or

$$\log[P(\lambda)] = -\xi(\lambda) = -(1/2.303)K(\lambda)wq \tag{59}$$

where

$I_o(\lambda)$ = a monochromatic light intensity before passing through an absorbent
$I(\lambda)$ = a light intensity after passing through an absorbent
$\xi(\lambda)$ = absorbance (also referred to as the optical density) of the absorbent

Assuming that the path length is a constant, and individual pixels have the same thickness, we can group the absorption coefficient and path length together as a compounded coefficient, K_d.

$$K_d(\lambda) = (1/2.303)K(\lambda)w \tag{60}$$

For mixed color films, the absorbance, $\xi_m(\lambda)$, is obtained by applying the proportionality and additivity rules.

$$\xi_m(\lambda) = K_{d1}(\lambda)q_1 + K_{d2}(\lambda)q_2 + \cdots + K_{dn}(\lambda)q_n \quad i = 1, 2, \ldots, n \tag{61}$$

where

K_{di} = the absorption coefficient of ith primary ink
q_i = the concentration of ith primary ink

Kubelka–Munk Theory. Kubelka and Munk formulated a theory based on two light channels traveling in opposite directions for modeling the translucent and opaque media. The light is being absorbed and scattered in only two directions, up and down. A background is presented at the bottom of the medium to provide the upward light reflection. The derivation of Kubelka–Munk formula can be found in many publications (4,32,33,34). The general expression of the Kubelka–Munk equation is

$$P(\lambda) = (1 - P_g(\lambda)[\hat{a}(\lambda) - \hat{e}(\lambda)\coth[\hat{e}(\lambda)S(\lambda)w]])/ \{\hat{a}(\lambda) - P_g(\lambda) + \hat{e}(\lambda)\coth[\hat{e}(\lambda)S(\lambda)w]\} \tag{62}$$

where

$$P(\lambda) = \text{the reflectance of the film}$$
$$P_g(\lambda) = \text{the reflectance of the background}$$
$$\hat{a}(\lambda) = 1 + K(\lambda)/S(\lambda) \tag{63}$$

$$\hat{e}(\lambda) = [\hat{a}(\lambda)^2 - 1]^{1/2} = \{[K(\lambda)/S(\lambda)]^2 + 2[K(\lambda)/S(\lambda)]\}^{1/2} \tag{64}$$

$$\coth[\hat{e}(\lambda)S(\lambda)w] = \{\exp[\hat{e}(\lambda)S(\lambda)w] + \exp[-\hat{e}(\lambda)S(\lambda)w]\}/ \{\exp[\hat{e}(\lambda)S(\lambda)w] - \exp[-\hat{e}(\lambda)S(\lambda)w]\}$$
$$K(\lambda) = \text{the absorption coefficient}$$
$$S(\lambda) = \text{the scattering coefficient} \tag{65}$$

This expression is called the two-constant Kubelka–Munk equation, because the two constants, K and S, are determined separately. It indicates that the reflectance of a translucent film is a function of the absorption coefficient, the scattering coefficient, the film thickness, and the reflectance of the background. Equation (62) is the basic form of the Kubelka–Munk equation and the foundation of various Kubelka–Munk formulas (33). To employ Eq. (62), one needs to know P_g and w in addition to the K and S constants.

Circular Dot Overlap Model

The circular dot overlap model is not the only dot overlap model. In fact, the Neugebauer equations utilize the Demichel's dot overlap model that is obtained from the joint probability for the overlapped area. Demichel's model is used to account for the color mixing, while the circular dot overlap model is used to correct tone reproduction due to oversized pixels. The circular dot model assumes an ideal round shape with equal size and a complete and uniform absorptance within the dot boundary. Pixel overlapped areas are physically computed by utilizing geometric relationships. The resulting density is a logical OR operation.

First, the size of an overlapped area is dependent on the pixel position and pixel size. The vertical (or horizontal) overlapping is different from the diagonal overlapping. For $\sqrt{2}\tau$ size pixels, there is no overlap in the diagonal position. The overlap in the vertical or horizontal direction can be calculated by using the geometric relationship; for a square grid, the overlapped area is enclosed in a square with the size of r^2, where r is the radius of the dot.

$$A_v = \pi r^2/4 + \pi r^2/4 - r^2 = (\pi - 2)\tau^2/4 \tag{66}$$

For dot size greater than $\sqrt{2}\tau$, one can derive relationships such as the diagonally overlapped area A_d, the vertically (or horizontally) overlapped area A_v, and the excess area A_e over the digital square pixel.

$$A_d = 2r^2 \sin^{-1}\{[1 - \tau^2/(2r^2)]^{1/2}\} - \tau(2r^2 - \tau^2)^{1/2} \tag{67}$$

$$A_v = 2r^2 \cos^{-1}[\tau/(2r)] - (\tau/2)(4r^2 - \tau^2)^{1/2} \tag{68}$$

$$A_e = \pi r^2 - \tau^2 \tag{69}$$

Monochromatic Dot Overlap Model. Pappas and coworkers expended the circular dot model to account for the amount of dot overlap at each pixel in terms of three area parameters α, β, and γ, as shown in Fig. 4 (35).

$$\begin{aligned} p'(i, j) &= P[W(i, j)] = 1 \quad \text{if } p(i, j) = 1 \\ p'(i, j) &= P[W(i, j)] = \kappa_1\alpha + \kappa_2\beta - \kappa_3\gamma \quad \text{if } p(i, j) = 0 \end{aligned} \tag{70}$$

where the window $W(i, j)$ consists of the pixel of interest $p(i, j)$ and its eight neighbors, κ_1 is the number of horizontally and vertically neighboring dots that are black, κ_2 is the number of diagonally neighboring dots that are black and not adjacent to any horizontally or vertically neighboring black dot, and κ_3 is the number of pairs of neighboring black dots in which one is a horizontal neighbor and the other is a vertical neighbor. These area parameters α, β, and γ are related to the overlapped areas A_e, A_d, and A_v as follows:

$$A_e/\tau^2 = 4\alpha + 4\beta = (\pi r^2 - \tau^2)/\tau^2 \tag{71}$$

$$\begin{aligned} A_d/\tau^2 = {} & 2\beta + 2\gamma = 2(r^2/\tau^2)\sin^{-1}\{[1 - \tau^2/(2r^2)]^{1/2}\} \\ & - (2r^2 - \tau^2)^{1/2}/\tau \end{aligned} \tag{72}$$

$$A_v/\tau^2 = 2\alpha + 4\beta = 2(r^2/\tau^2)\cos^{-1}[\tau/(2r)] - (4r^2 - \tau^2)^{1/2}/(2\tau) \tag{73}$$

The α, β, and γ can be solved from this set of simultaneous equations. The solutions are as follows:

$$\alpha = \pi\rho^2/4 + (2\rho^2 - 1)^{1/2}/4 - (\rho^2/2)\cos^{-1}[1/(\sqrt{2}\rho)] - 1/2 \tag{74}$$

$$\beta = -\pi\rho^2/8 - (2\rho^2 - 1)^{1/2}/4 - (\rho^2/2)\cos^{-1}[1/(\sqrt{2}\rho)] + 1/4 \tag{75}$$

$$\gamma = (\rho^2/2)\sin^{-1}[(\rho^2 - 1)/\rho^2]^{1/2} - (\rho^2 - 1)^{1/2}/2 - \beta \tag{76}$$

where $\rho = \sqrt{2}\, r/\tau$, the ratio of the actual dot radius to the minimum overlap radius $\tau/\sqrt{2}$. Since the dot diameter is between $\sqrt{2}\,\tau$ and 2τ, the ratio ρ is between 1 and $\sqrt{2}$ such that the black dots are large enough to cover a $\tau \times \tau$ square, but not too large to cover a white pixel, separating two black pixels in vertical or horizontal position.

The main application of this dot overlap model is in the digital halftoning. Halftone patterns for the different levels of the halftone dot are analyzed with this circular dot model to predict the actual reflectance of the output patterns (35). Threshold levels of the halftone dot are then adjusted so that the reflectance of the dot patterns matches the input reflectance values. This correction produces patterns with fewer isolated pixels, since the calculation gives a more accurate estimate of the actual reflectance achieved on the paper.

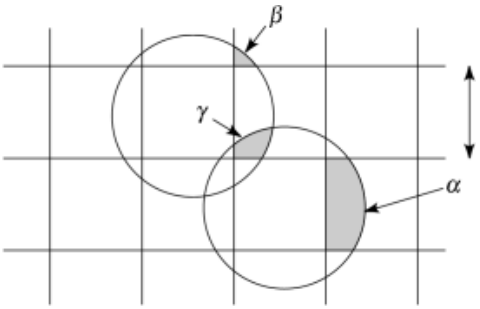

Fig. 4. Circular dot overlap model with the definition of three overlapped areas, α, β, and γ.

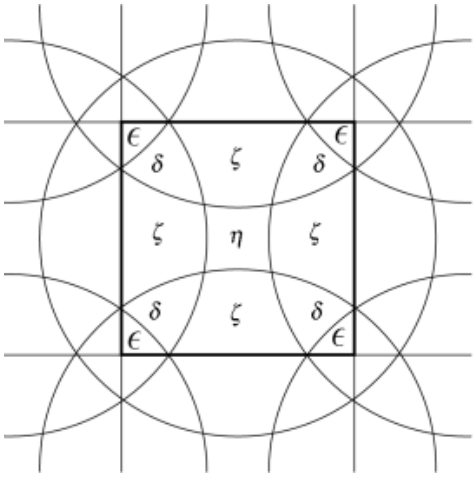

Fig. 5. Circular dot overlap model for color images and the definition of four overlapped segments, ε, δ, ζ, and η.

Color Dot Overlap Model. Pappas expanded the monochromatic dot overlap model to color (36). The color dot overlap model accounts for overlap between neighboring dots of the same and different colors. The resulting color is the weighted average of segmented colors, with the weight being the area of the segment. The overlapping color segments are shown in Fig. 5; the color of each segment depends on the absorption properties of the overlapped inks. For three-color systems such as *RGB* and *cmy*, each segment can have one of the possible combinations of primary colors (white, cyan, magenta, yellow, red, green, blue, and black). For four-color systems, there are $2^4 = 16$ dominant colors. This approach is basically the Neugebauer equations using geometrically computed areas instead of the Demichel's dot overlap model. The area of each segment can be calculated in terms of area parameters α, β, and γ given in Fig. 4.

$$\epsilon = \beta,\ \delta = \gamma - \beta, \quad \zeta = \alpha - 2\gamma, \quad \eta = 1 - 4\alpha + 4\gamma \qquad (77)$$

In Fig. 5, there are four ε, four δ, four ζ, and one η; these multiple areas can have different colors. This makes the computation of color in each segment much more complex.

Measurement-Based Tone Correction. The circular dot overlap model is an ideal case. It provides a good first-order approximation for the behavior of printers. In reality, it does not adequately account for all of printer distortions. Significant discrepancies often exist between the predictions of the model and measured

values (35). In view of this problem, a traditional approach to the correction of printer distortions is the measurement-based techniques.

Basically, the measurement-based technique is used to obtain the relationship between the input tone level and measured output reflectance or optical density for a specific halftone algorithm, printer, and substrate combination. The experiment procedure consists of making prints having patches that are halftoned by the algorithm of interest. Each patch corresponds to an input tone level. The number of selected tone levels should be sufficient to cover the whole toner dynamic range with fine step size, but it is not necessary to print all available input levels, such as 256 levels for an 8-bit system, because most printers and measuring instruments are not able to resolve such a fine step. The reflectance or density values of these patches is measured to give the tone-response curve of the halftone algorithm used in conjunction with the printing device. Correcting for the printer distortions consists of inverting this curve to form a tone-correction curve. This tone-correction curve is applied to the image data prior to halftoning and printing as a precompensation.

Device-Independent Color Imaging

In a modern office environment, various imaging devices, such as scanners, personal computers, workstations, copiers, and printers, are connected to this open environment. For example, a system may have several color printers using different printing technologies, such as dot matrix, ink jet, and xerography. Moreover, it does not exclude devices of the same technology from different manufactures. In view of these differences, the complexity of an office information system can be extremely high. The main concern in the color image analysis of the electronic information at the system level is the color consistency; the appearance of a document should look alike when the image moves across various devices and goes through many color transforms. However, the degree of the color consistency depends on the applications; a desktop printing for casual users may differ significantly from a short-run printing of high-quality commercial brochures.

Hunt pointed out that there are several levels of the color matching: spectral, colorimetric, exact, equivalent, corresponding, and preferred color matchings (37). Spectral color reproduction matches the spectral reflectance curves of the original and reproduced colors. At this level, the original and reproduction look the same under any illuminant; there is no metamerism. A pair of stimuli is metamers if they match under one illuminant but mismatch under a different illuminant. The colorimetric reproduction is a metameric match that is characterized by the original and the reproduction colors having the same CIE chromaticities and relative luminances. A reproduction of a color in an image is exact if its chromaticity, relative luminance, and absolute luminance are the same as those in the original scene. The equivalent color reproduction requires that the chromaticities, relative luminance, and absolute luminances of colors have the same appearance as the colors in the original scene when seen in the image viewing conditions. The corresponding reproduction matches the chromaticities and relative luminances of the color such that they have the same appearance as the colors in the original would have had if they had been illuminated to produce the same average absolute luminance level. The preferred color reproduction is defined as reproductions in which the colors depart from equality of appearance to those in the original, either absolutely or relative to white, in order to give a more pleasing result to the viewer. Ideally, the appearance of a reproduction should be equal to the appearance of the original or to the customer's preference. In many cases, however, designers and implementers of the color management systems do not know what users really want. For an open environment with such a diverse user demography, it is very difficult to satisfy everybody. Usually, this demand is partially met by an interactive system, adjustable knobs, or some kind of soft proofing system. This difficulty is compounded by the fact that modeling the human visual system is still an area of active research and debate.

Even color consistency based on colorimetric equivalence at the system level is not a simple task because each device has its unique characteristics. They are different in imaging technology, color encoding, color description, color gamut, stability, and applications. Because not all scanners use the same lamp and filters

and not all printers use the same inks, the same *RGB* image file will look different when it is displayed on different monitors. Likewise, the same *cmyk* image file will look different when it is rendered by different printers. Moreover, there are even bigger difference in color appearance when the image is moved from one medium to another, such as from monitor to print. Because of these differences, efforts have been made to establish a color management system (*CMS*) that will provide color consistency. An important event was the formation of the International Color Consortium (*ICC*) for promoting the usage of color images by increasing the interoperability among various applications (image processing, desktop publishing, etc.), different computer platforms, and operating systems. This organization established the architectures and standards for color transform via ICC profiles and color-management modules (*CMM*). A profile contains the data for performing color transform and a CMM is an engine for actually processing the image through profiles. The central theme in color consistency at the system level is the requirement of a device-independent color description and properly characterized and calibrated color imaging devices.

A proper color characterization requires image test standards, techniques, tools, instruments, and controlled viewing conditions. Main features to consider for color consistency are image structure, white point equivalence, color gamut, color transformation, measurement geometry, media difference, printer stability, and registration (Ref. 4, pp. 261–271).

Image structure refers to the way an image is rendered by a given device. The primary factor affecting image texture is the quantization; it is particularly apparent if the output is a bilevel device. Scanners and monitors that produce image elements in 8 bit depth can be considered as contone devices. For printers, the dye sublimation printer is capable of printing contone; other types of printers, such as xerographic and ink jet printers, have to rely on halftoning to simulate the contone appearance. For bilevel devices, halftoning perhaps is the single most important factor in determining the image structure. It affects the tone reproduction curves, contouring, banding, graininess, moiré, sharpness, and resolution of fine details. Because of these dependencies, color characterization has to be done for each halftone screen. A new characterization is needed when the halftone screen is changed.

Various white points are used for different devices; the white point for the scanner is very different from the one used in the CRT monitor. Moreover, the viewing condition of a hard copy is usually different from those used in the scanning and displaying. The operator or the color management system must convert the chromaticity of the white point in the original to that of the reproduction. From the definition of the tristimulus values, we know that the exact conversion requires the substitution of one illuminant spectrum for another, weighs by the color matching function and object spectrum and then integrates the whole visible range. This illuminant replacement process is not a linear correspondence. The change of viewing conditions has been addressed by the chromatic-adaptation models (Ref. 1, pp. 191–214).

The color gamut indicates the range of colors an imaging device can render; it is the volume enclosed by the most saturated primary colors and their two-color mixtures in a 3D color space. The gamut size and shape are governed mainly by the colorants, printing technology, and media. Good colorants can be found for use in each technology, but they tend to vary from technology to technology. As a result, different color devices have different gamut sizes such that a color monitor is different from a color printer and a xerographic printer is different from an ink jet printer. Color gamut mismatches of input, display, and output devices are the most difficult problem in maintaining color consistency. Numerous approaches for gamut mapping have been proposed; however, this is still an active research area. Color correction for ink-substrate combinations is only a partial solution. The relationship between the original and reproduction should be taken into account. In fact, the optimal relationship is not usually known; Hunt suggests having six different color matching levels (37). In most cases, the optimum relationship is a compromise. The compromise depends on the user's preference, on the colors in the original (e.g., monitor), and on those available in the reproduction (e.g., printer). The optimum transform for photographic transparency originals that contain mainly light colors (the high-key image) is different from the optimum for an original that contains mainly dark colors (the low-key image).

Graphic images probably require a different transform than scanned photographs (38). Color gamut mapping and preferred color transform are a part of the color appearance task.

Most colorimeters and spectrophotometers have measuring geometries that are deliberately chosen either to eliminate or totally include the first surface reflections. Typical viewing conditions, having light reflection, transmission, and scattering, lie somewhere between these two extremes. Colors that look alike in a practical viewing condition may measure differently, and vice versa. This phenomenon is strongly affected by the surface characteristics of the images and is most significant for dark colors. A practical solution to this problem is being proposed by using telecolorimetry that puts the measuring device in the same position as the observer (39).

As mentioned previously, CIE systems have limited capability to account for changes in appearance that arise from a change of the white point or substrate. Under different media, the isomeric samples may look different. Even within a given medium type, such as paper substrate, the difference can come from the whiteness of the paper, fluorescence, gloss, surface smoothness, and so on. This problem is also part of the color appearance model.

Color transformation is the central part of the color consistency at the system level. This is because different devices use different color descriptions. Various transformation methods, such as the color mixing model, regression, three-dimensional interpolation, artificial neural network, and fuzzy logic, have been developed. Major tradeoffs among these techniques are conversion accuracy, processing speed, and computation cost.

Color printers are known for their instability; it is a rather common experience that the first print may be quite different from the hundredth one. A study indicates that dye sublimation, ink jet, and thermal transfer techniques are more stable than lithography (40). It is an encouraging trend that new printing technologies are more stable than the conventional methods. A part of the device stability problems, such as drifting, may be corrected by the device calibration to bring the device back to the nominal state.

Color printing requires three (*cmy*) or four (*cmyk*) separations. They are printed by multiple passes through the printing mechanism or by multiple separated operations in one pass. These operations require mechanical control of the pixel registration. The registration is very important to the color image quality. But it is more a mechanical problem than an image processing problem. Thus, it is beyond the scope of this article.

As stated at the beginning of this article, the color image has two components: spatial patterns and colorimetric properties. Factors affecting the color matching are the color transform, color gamut mapping, viewing conditions, and media. Factors affecting the image structure are the sampling, tone level, compression, and halftoning. The surface characteristics are primarily affected by the media; various substrates may be encountered during image transmission and reproduction. These factors are not independent, they may interact to cause the appearance change. For examples, inadequate sampling and tone level not only give poor image structures but may also cause color shifts; a different halftone screen may cause tone changes and color shifts; an improper use of colors may create textures; images with sharp edges appear higher in saturation and have more contrast than images with soft edges (38); and finer screen rulings will give higher resolution and cleaner highlight colors (41).

With these complexicities and difficulties, a systematic formulated color imaging model that incorporates the human visual model, color mixing model, color appearance model, and dot overlap model is needed to improve and analyze color images.

Human Visual Model. The CSF plays an important role in determining image resolution, image quality improvement, halftoning, and compression. Therefore, many formulas of CFS have been derived from various experiment results. Here we introduce a popular one proposed by Mannos and Sakrison (42).

$$\begin{aligned} V(f_x, f_y) &= a(b + cf_r)\exp[-(cf_r)^d] \quad \text{if } f_r > f_{\max} \\ V(f_x, f_y) &= 1.0 \quad \text{otherwise} \end{aligned} \tag{78}$$

where the constants a, b, c, and d are calculated by regression fits of horizontal and vertical threshold modulation data to be 2.2, 0.192, 0.114, and 1.1, respectively. The parameter f_r is the normalized radial spatial frequency in cycles/degree of visual subtended angle, and f_{max} is the peak frequency of the bandpass CSF curve. The second expression in Eq. (78) changes the bandpass CSF to a lowpass CSF. To account for angular variations in the human visual function sensitivity, the normalized radial spatial frequency is computed from the actual radial spatial frequency using an angular-dependent scale function (43,44,45):

$$f_r = f/\Omega(\theta) \tag{79}$$

where

$$f = (f_x^2 + f_y^2)^{1/2} \tag{80}$$

and $\Omega(\theta)$ is given by

$$\Omega(\theta) = [(1-\phi)/2]\cos(4\theta) + [(1+\phi)/2] \tag{81}$$

where ϕ is a symmetry parameter with a value of 0.7 chosen from empirical data, and

$$\theta = \tan^{-1}(f_y/f_x) \tag{82}$$

This angular normalization produces a 70% bandwidth reduction at 45°. The last step is to convert from cycles/degree to cycles/mm on the document for further spatial imaging processing:

$$f_r = (\pi/180)[f_i/\sin^{-1}[1/(1+d_v^2)^{1/2}]] \quad i = 1, 2, 3, \ldots, N \tag{83}$$

where d_v is the viewing distance in millimeters, f_i is the discrete sampling frequency in cycles per millimeter given by

$$f_i = (i-1)/(N\tau) \tag{84}$$

and τ is the sample spacing on the document.

Color Visual Models. In general, color visual models are extensions of luminance models by utilizing the human visual difference in the luminance and chrominance. We can separate the luminance and chrominance channels by using the opponent color description. The luminance part of the model is the same as before. For the chrominance part, we use the fact that the contrast sensitivity to spatial variations in chrominance falls off faster than the luminance with respect to spatial frequency (see Fig. 3) to find a new set of constants for the chromatic CSF. Both responses are lowpass; however, the luminance response is reduced at 45°, 135°, 225°, and 315° for the orientation sensitivity. This will place more luminance error along diagonals in the frequency domain. The chrominance response has a narrower bandwidth than the luminance response. Using the narrower chrominance response as opposed to identical responses for both luminance and chrominance will allow more lower-frequency chromatic error, which will not be perceived by human viewers. We can further allow the adjustment of the relative weight between the luminance and chrominance responses. This is achieved by multiplying the luminance response with a weighting factor. As the weighting factor increases, more error will be forced into the chrominance components.

Many human visual models have been proposed that attempt to capture the central features of human perception. The simplest HVS includes just a visual filter that implements one of the CSFs mentioned previously. Better approaches include a module in front of the filter to account for nonlinearity like the Weber's law. The filtered signal is pooled together in a single channel of information. This structure is called a single-channel model. Because the digital image quality is a very complex phenomenon, it requires inputs of all types of distortions. In view of the image complexity, multiple-channel approaches are developed to include various inputs. This is achieved by putting a bank of filters before the nonlinearity in a systematic way; each filter addresses one aspect of the whole image quality. This approach is called a multichannel model.

S-CIELAB. Zhang and Wandell extended CIELAB to account for the spatial as well as color errors in reproduction of the digital color image (46). They called this new model as spatial-CIELAB or S-CIELAB. The design goal is to apply a spatial filtering to the color image in a small-field or fine-patterned area, but revert to the conventional CIELAB in a large uniform area. The procedure for computing S-CIELAB is as follows:

(1) Input image data are transformed into an opponent-colors space. This color transform converts the input image, specified in terms of the CIEXYZ tristimulus values, into three opponent-colors planes that represent luminance, red-green, and blue-yellow components.
(2) Each opponent-colors plane is convolved with a two-dimensional kernel whose shape is determined by the visual spatial sensitivity to that color dimension; the area under each of these kernels integrates to one. A lowpass filtering is used to simulate the spatial blurring by the human visual system. The computation is based on the concept of pattern-color separability. The parameters for the color transform and spatial filters were estimated from psychophysical measurements.
(3) The filtered representation is converted back to CIEXYZ space, then to CIELAB representation. The resulting representation includes both spatial filtering and CIELAB processing.
(4) The difference between the S-CIELAB representations of the original and its reproduction is the measure of the reproduction error. The difference is expressed by a quantity ΔE_s, which is computed precisely as ΔE^*_{ab} in conventional CIELAB.

The S-CIELAB reflects both spatial and color sensitivity. This model is a color-texture metric and a digital imaging model. This metric has been applied to printed halftone images. Results indicate that S-CIELAB correlates with perceptual data better than standard CIELAB (47). This metric can also be used to improve the multilevel halftone images (48).

Color Imaging Process. A complete color imaging process that incorporates various visual and physical models may be described as follows: An image with a specific sampling rate, quantization level, and color description is created by the input device that uses (1) the human visual model to determine the sampling and quantization, and (2) the device calibration to ensure that the gain, offset, and tone reproduction curves are properly set. The setting of tone reproduction curves may employ the dot overlap model, for example. This input device assigns device-dependent color coordinates (e.g., *RGB*) to the image elements. The image is fed to the color transform engine for converting to colorimetric coordinates (e.g., CIELAB). This engine performs the colorimetric characterization of the input device; it implements one or more of color space transform techniques (e.g., color mixing models, 3-D lookup tables). This engine can be a part of the input device, an attachment, or in the host computer.

The second step is to apply a chromatic-adaptation and/or color appearance model to the colorimetric data, along with additional information on the viewing conditions of the original image (if any), in order to transform the image data into appearance attributes such as lightness, hue, and chroma. These appearance coordinates, which have accounted for the influences of the particular device and the viewing conditions, are called viewing-conditions-independent space (Ref. 1, p. 346). At this point, the image is represented purely by its original appearance. After this step, the image may go through various processing, such as spatial scaling,

edge enhancement, resolution conversion, editing, or compression. If the output device is selected, the image should be transformed to a format suitable for the output device. If the output is a bilevel device, the image should be halftoned. A good halftone algorithm takes the human visual model, dot overlap model, and color mixing model into account. Then the color gamut mapping is performed to meet the output color gamut. Now the image is in its final form with respect to the appearance that is to be reproduced. Next, the process must be reversed. The viewing conditions for the output image, along with the final image-appearance data, are used in an inverted color appearance model to transform from the viewing-conditions-independing space to a device-independent color space (e.g., CIEXYZ). These values, together with the colorimetric characterization of the output device, are used for transforming to the device coordinates (e.g., *cmyk*) for producing the desired output image.

BIBLIOGRAPHY

1. M. D. Fairchild *Color Appearance Models*, Reading, MA: Addison-Wesley, 1998, p. 24.
2. G. Sharma H. J. Trussell Digital color imaging, *IEEE Trans. Image. Proc.*, **6**: 901–932, 1997.
3. G. Wyszecki W. S. Stiles *Color Science: Concepts and Methods, Quantitative Data and Formulae*, 2nd ed., New York: Wiley, 1982, pp. 489–490.
4. H. R. Kang *Color Technology for Electronic Imaging Devices*, Bellingham, WA: SPIE Optical Engineering Press, 1997, pp. 3–6.
5. M. R. Pointer G. G. Attridge The number of discernible colors, *Color Res. Appl.*, **23**: 52–54, 1998.
6. CIE, *Colorimetry*, Publication No. 15.2. Paris: Bureau Central de la CIE, 1971.
7. D. B. Judd G. Wyszecki *Color in Business, Science, and Industry*, 2nd ed., New York: Wiley, 1967, p. 138.
8. D. F. Rogers *Procedural Elements for Computer Graphics*, New York: McGraw-Hill, 1985, p. 389.
9. C. J. Bartleson Colorimetry, in F. Grum and C. J. Bartleson, *Optical Radiation Measurements*, Vol. 2, *Color Measurement*, New York: Academic Press, 1980, pp. 33–148.
10. B. A. Wandell *Foundations of Vision*, Sunderland, MA: Sinauer Associates, 1995, pp. 106–108.
11. A. A. Michelson *Studies in Optics*, Chicago: Univ. Chicago Press, 1927.
12. V. K. Zworykin G. A. Morton *Television*, 2nd ed., New York: Wiley, 1954.
13. F. Mazda, ed. *Electronic Engineers Reference Book*, 5th ed., London: Butterworths, 1983.
14. R. Ulichney *Digital Halftoning*, Cambridge, MA: MIT Press, 1987, pp. 79–84.
15. Y. Nayatani *et al.* Lightness dependency of chroma scales of a nonlinear color-appearance model and its latest formulation, *Color Res. Appl.*, **20**: 156–167, 1995.
16. R. W. G. Hunt *The Reproduction of Color in Photography, Printing and Television*, 5th ed., England: Fountain Press, 1995, Chap. 31.
17. M. D. Fairchild Refinement of the RLAB color space, *Color Res. Appl.*, **21**: 338–346, 1996.
18. S. Guth Further applications of the ATD model for color vision, IS&T/SPIE's Symposium on Electronic Imaging: Science & Technology, *Proc. SPIE*, **2414**: 12–26, 1995.
19. M. R. Luo M-C. Lo W-G. Kuo The LLAB (l:c) color model, *Color Res. Appl.*, **21**: 412–429, 1996.
20. H. E. J. Neugebauer Die theoretischen Grundlagen des Mehrfarbendruckes, *Z. wiss. Photogr.*, **36**: 73–89, 1937.
21. J. A. C. Yule *Principles of Color Reproduction*, New York: Wiley, 1967, Chap. 10.
22. I. Pobboravsky M. Pearson Computation of dot areas required to match a colorimetrically specified color using the modified, *Proc. TAGA*, **24**: 65–77, 1972.
23. R. Holub W. Kearsley Color to colorant conversions in a colorimetric separation system, SPIE Vol. 1184, *Neugebauer Memorial Seminar Color Reproduction*, 1989, pp. 24–35.
24. F. Pollak The relationship between the densities and dot size of halftone multicolor images, *J. Photogr. Sci.*, **3**: 112–116, 1955.
25. F. Pollak Masking for halftone, *J. Photogr. Sci.*, **3**: 180–188, 1955.
26. F. Pollak New thoughts on halftone color masking, *Penrose Annu.*, **50**: 106–110, 1956.
27. A. C. Hardy F. L. Wurzburg Color correction in color printing, *J. Opt. Soc. Amer.*, **38**: 300–307, 1948.
28. A. Murray *J. Franklin Inst.*, **221**: 721, 1936.

29. J. A. C. Yule W. J. Nielsen The penetration of light into paper and its effect on halftone reproduction, *Proc. Tech. Assn. Graphics Arts*, **4**: 65–75, 1951.
30. F. R. Clapper J. A. C. Yule The effect of multiple internal reflections on the densities of half-tone prints on paper, *J. Opt. Soc. Amer.*, **43**: 600–603, 1953.
31. F. R. Clapper J. A. C. Yule Reproduction of color with halftone images, *Proc. Tech. Assn. Graphic Arts*, 1–12, 1955.
32. P. Kubelka New contributions to the optics of intensely light-scattering materials. Part I, *J. Opt. Soc. Am.*, **38**: 448–457, 1948.
33. D. B. Judd G. Wyszecki *Color in Business, Science, and Industry*, 3rd ed., New York: Wiley, 1975, pp. 397–461.
34. E. Allen Colorant Formulation and Shading, in F. Grum and C. J. Bartleson, *Optical Radiation Measurements*, Vol. 2, New York: Academic Press, 1980, pp. 305–315.
35. T. N. Pappas D. L. Neuhoff Model-based halftoning, *Proc. SPIE*, **1453**: 244–255, 1991.
36. T. N. Pappas Model-based halftoning of color images, *IEEE Trans. Image Proc.*, **6**: 1014–1024, 1997.
37. R. W. G. Hunt *The Reproduction of Color in Photography, Printing and Television*, 4th ed., England: Fountain Press, 1987, pp. 177–197.
38. W. L. Rhodes Digital imaging: Problems and standards, *Proc. SID*, **30**: 191–195, 1989.
39. T. Johnson Device independent colour—Is it real? *TAGA Proc.*, 81–113, 1992.
40. S. B. Bolte A perspective on non-impact printing in color, SPIE Vol. 1670, *Color Hard Copy and Graphic Arts*, 1992, pp. 2–11.
41. G. G. Field The systems approach to other reproduction—A critique, *Proc. TAGA*, 1–17, 1984.
42. J. L. Mannos D. J. Sakrison The effects of a visual fidelity criterion on the encoding of images, *IEEE Trans. Info.Theory*, **IT-20**: 525–536, 1974.
43. T. Mitsa J. R. Alford Single-channel versus multiple-channel visual models for the formulation of image quality measures in digital halftoning, *IS&T NIP10*, 1994, pp. 385–387.
44. K. E. Spaulding R. L. Miller J. Schildkraut Methods for generating blue-noise dither matrices for digital halftoning, *JEI*, **6**: 208–230, 1997.
45. J. Sullivan L. Ray R. Miller Design of minimum visual modulation halftone patterns, *IEEE Trans. Syst. Man. Cybern.*, **21**: 33–38, 1990.
46. X. Zhang B. A. Wandell A spatial extension of CIELAB for digital color image reproduction, *SID Int. Symp., Dig. Tech. Papers*, 1996, pp. 731–734.
47. X. Zhang, et al. Color image quality metric S-CIELAB and its application on halftone texture visibility. *IEEE Compcon.* 1997, pp. 44–48.
48. X. Zhang J. E. Farrell B. A. Wandell Applications of a spatial extension to CIELAB, *Proc. SPIE/IS&T Symposium on Electronic Imaging*, 3025: 154–157, 1997.

HENRY R. KANG
Peerless System Corporation