



Munsell Color Science Laboratory Technical Report

The Science of Digitizing Two-Dimensional Works of Art for Color-Accurate Image Archives – Concepts Through Practice

Roy S. Berns

May, 2000

Abstract

A review of the human visual system, the CIE L^* , a^* , b^* color space and its use in evaluating color-image quality, and digital image capture is presented, the goal of which is to provide background information for imaging professionals involved in creating digital-image databases for museums, archives, and libraries. Following this review, an analysis was performed to determine the effects of bit depth, dynamic range, gamma correction, and color correction on the ability to estimate colorimetric data from R, G, B digital images with a minimum of error. The proper use of gray scale and color targets was also considered. Recommendations are presented for the direct digital image capture of two-dimensional works of art. Finally, a brief look into the future using spectral-imaging techniques is given.

Introduction

Digital-image databases have become ubiquitous with museums, archives, and libraries. (For simplicity, “museum” will be used to represent the many repositories.) Connect to their websites and a wealth of visual information is available. Museums are sharing their images to increase access such as the Art Museum Image Consortium, AMICO (see www.amico.org). It seems that each month, another digital imaging project is proposed or initiated.

The quality of the digital images can be quite varied and at times, inconsistent with the basic philosophy of these repositories, preserving mankind’s treasures for the future. Leaving aside the issue of the long-term storage of digital information, clearly a critical issue but beyond the scope of this publication, there are many reasons that quality might be poor. Some are technical in origin and some are caused by constraints imposed by the sheer magnitude of creating and maintaining a digital archive of many thousands, and in some cases, hundreds of thousands of images.

The problems are manifest almost immediately. Trying to define specifications for image capture, storage, retrieval, and display (both in soft and printed forms) with an eye on cost, time, and who will do the work is extremely challenging. Since the scope of this publication is limited to direct digital capture of art and will not address digitizing photographic reproductions, our eye must also be trained on conservation issues. The process of digitizing must not damage the object, which can occur all too easily by excessive handling and irradiation by high-intensity light sources. Clearly, it is desirable to digitize the art once and in a manner that facilitates a variety of derivatives created for a variety of applications, be it web-based, printed publication, scientific analysis, or any number of scholarly endeavors by art historians.

Thus the purpose of this publication is to provide sufficient background to aid in setting specifications for image capture systems. Having presented this background, it is possible to propose a methodology for digital image capture that minimizes the inherent limitations of many digital systems in use today. Finally, considerations will be given to the future. What are the optimal characteristics of a digital-imaging system designed to record two-dimensional works of art?

A Review of Digital Imaging

The Human-Visual System

The first step in understanding digital imaging is to understand about the human visual system. Four concepts will be defined: spectral sensitivity, opponent encoding, spatial resolution, and nonlinear response. This description is very simplified. For greater detail, but still at a simple level, see Berns.¹ For a thorough understanding of the human visual system, see Wandell.²

Incident light interacts with visual receptors, rods and cones. Following a chemical reaction, light energy is converted to a neural signal. There are three classes of cones (our color receptors), L, M, and S, named for their dominant sensitivity to long, medium, and short wavelengths of light. Each has a unique spectral sensitivity, shown in Figure 1. Spectral sensitivity defines a detector’s sensitivity as a function of wavelength. Because there are only three types of cones and their spectral sensitivities are broad, many different objects can produce the same cone responses, leading to the identical color. This is known as metamerism (pronounced “me-tam’-er-ism”) and explains why color reproduction is possible in which small numbers of colorants can reproduce our world, composed of thousands of colorants. A metameric match between foliage, illuminated by natural daylight, and a CRT color reproduction of the foliage is shown in Figure 2. Despite large differences in their spectral properties, these two stimuli match in color.

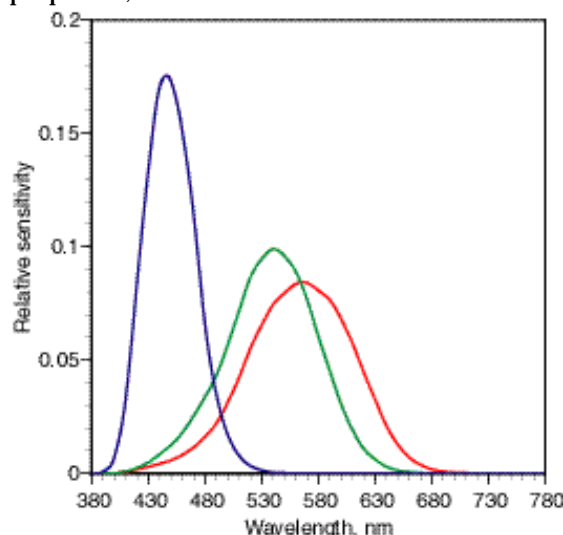


Figure 1. The spectral sensitivities of the human-visual system’s cones³ (normalized to equal area).

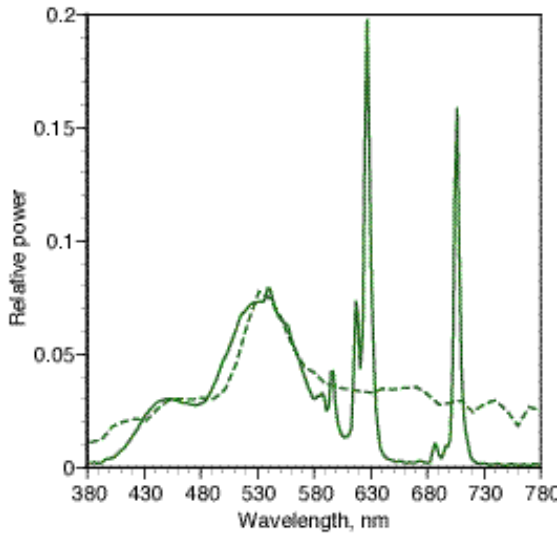


Figure 2. Spectral properties of a metameric pair formed by foliage, illuminated by natural daylight, and a CRT display.¹

The cones combine spatially and form opponent signals, white/black, red/green, and yellow/blue. During the late 1800's, Hering considered these six colors as elemental, shown in Figure 3. Lines connecting the various elemental colors indicate possible continuous perceptions. There are not lines connecting yellow and blue and red and green. It is not possible to have a color that is simultaneously reddish and greenish, for example.

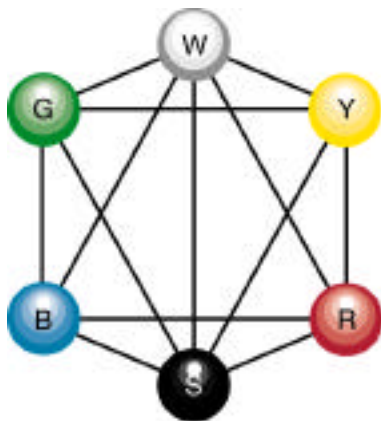


Figure 3. Six elemental colors postulated by Hering along with lines indicating possible perceptions (e.g., yellowish red).¹

Because there are different numbers of each cone type, when they combine to form opponent signals, there are different spatial resolutions among the opponent signals. Spatial resolution defines the resolving power of an imaging system and leads to the ability to

discern fine detail. The black/white signal has the highest spatial resolution, followed by the red/green signal. The yellow/blue signal has quite low spatial resolution. These differences in spatial resolution have been exploited in image compression such as JPEG.⁴ The concept of “visually lossless compression” originates with this property of the eye. Spatial resolution is reduced in the chromatic channels in a manner that results in images that look identical to their uncompressed counterparts when viewed at typical distances, hence the term “visually lossless.” However, information is still being discarded.

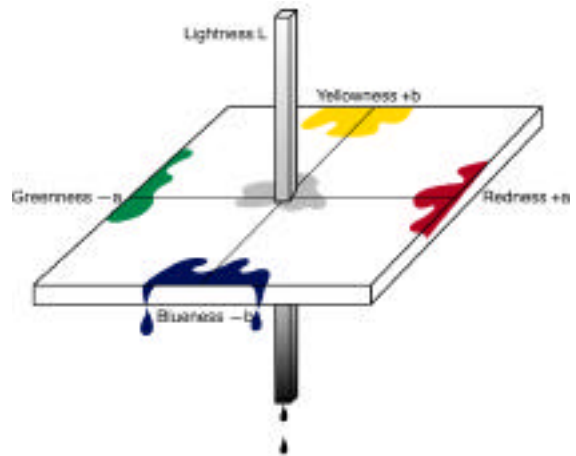


Figure 4. Conceptualized CIELAB color space.¹

The neural processing from cone receptor signals through signals interpreted by the brain and resulting in color names such as yellow, brown, and gray is exceedingly complex. Every year, vision scientists fill in more pieces of the puzzle. From a practical perspective, it is very useful to have a simple mathematical model that enables color perceptions to be estimated from light imaged onto our visual system. A model was derived by the International Commission on Illumination (CIE) in 1976, known as CIE L^* , a^* , b^* (pronounced “el-star” and so on) or its official abbreviation, CIELAB (pronounced “see-lab”).⁵ The coordinates, L^* , a^* , and b^* represent the perceptions of lightness, redness/greenness, and yellowness/blueness, respectively. CIELAB is considered a color space in which positions indicate an object’s color. This type of color space is diagrammed in Figure 4. The coordinates are calculated from the knowledge of an object’s spectral reflectance or transmittance properties, a light source interacting with the object, and the spectral

sensitivities of an observer. For a number of reasons, the L, M, and S spectral sensitivities, shown in Figure 1, are not used; rather, a set of curves mathematically related to spectral sensitivities, known as color-matching functions,^{1,5} are used instead.

The complex signal processing of the visual system results in a nonlinear (curved) relationship between light imaged onto the eye and color perceptions, shown in Figure 5 for lightness. This means that for dark colors, small changes in an object's reflectance or transmittance lead to large changes in lightness. For light colors, the opposite occurs: Large changes in an object's reflectance or transmittance lead to small changes in lightness. This curvature is known as a compressive function. This compression also occurs for the chromatic channels.

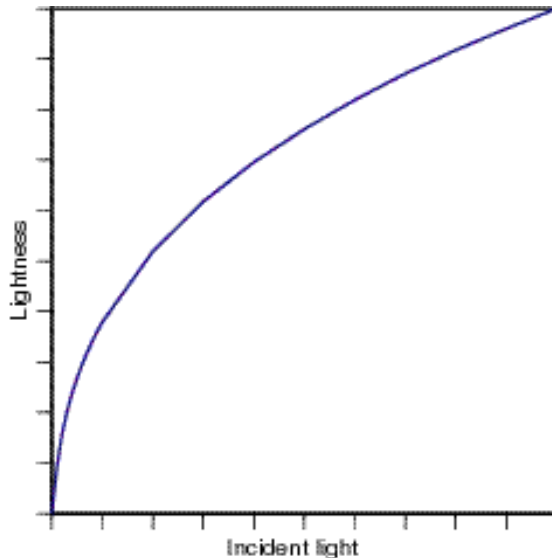


Figure 5. Nonlinear response between incident light and lightness. The incident light can have units of radiance, irradiance, luminance, illuminance, or simply power per unit area. (CIE L^* is used as a representation of lightness.⁵)

In summary, the human visual system has three cone types, each with unique spectral sensitivities that overlap greatly. They combine spatially forming opponent signals: white/black, red/green, and yellow/blue. Each channel has a different spatial resolution. The relationship between incident light and color perceptions is nonlinear. A simple model, CIELAB, can be used to calculate color perceptions from measurements of the object, its illumination, and the observer.

The Digital Camera

Digital cameras consist of three main components: input optics, sensor, and signal processing. The optical system is typical of conventional photography in which a scene is imaged on to a two-dimensional plane, much like the human visual system.⁶ Light interacts with a sensor, often a charged-couple device (CCD) that converts light energy to an electrical signal.⁷ The sensor may be one-dimensional — a single row of sensors — or two-dimensional — a grid of sensors. In one-dimensional systems, the row of sensors is scanned across the image plane. These are “scanbacks” and are similar to flat-bed scanners. Because scanning occurs, the objects being digitized are stationary. Through signal processing, the electrical signal is converted to a digital signal. An amplifier and analog-to-digital converter are used in the conversion process.

For color, the one-dimensional detector array is trebled, each array having either a red, green, or blue filter in front of the detector. In some systems, a color filter wheel is positioned between the optical system and a monochrome (i.e., grayscale) scanning array. The array scans across the image plane multiple times, once for each filter-wheel position. For two-dimensional arrays, detectors are filtered in a mosaic pattern. Similar to the human visual system, there is not the same number of red, green, and blue sensors. Image processing is used to interpolate missing data for each color plane. The design of the mosaic and accompanying image processing is optimized to minimize artifacts caused by sampling, known as aliasing. It is also possible to have three two-dimensional monochrome arrays, each filtered with either a red, green, or blue filter or a single array and filter wheel. These systems do not require the complex interpolation procedures.

For color accuracy, the most important characteristic of the digital camera is its spectral sensitivities. Ideally, they should closely resemble the human visual system's spectral sensitivities shown in Figure 1.⁷ For many scanbacks, this critical characteristic is seldom achieved, shown in Figure 6. It is seen that these

⁷ Strictly, a camera's spectral sensitivities should be linear transforms of the human visual system's spectral sensitivities.⁸ That is, through a linear transformation, the L, M, and S sensitivities are well estimated. Thus, CIE color-matching functions, which do not resemble L, M, and S sensitivities, meet this criterion. This is also known as the Luther or Luther-Ives condition.

sensitivities have very little overlap and the position of the red sensitivity is shifted considerably to longer wavelengths. It is important to note that these sensors were designed for flat-bed scanners. These devices are optimized for imaging photographic materials. These sensitivities are densitometric, rather than colorimetric. They are designed to record the amount of cyan, magenta, and yellow dye in photographic materials.⁹ These sensitivities are reasonably well suited for this purpose. If a photographic collection were digitized using this scanback, it is possible to transform the raw R, G, and B data to highly accurate L^* , a^* , b^* estimates, though the computations are complex.¹⁰ The author is not aware of this technique being used when digitizing photographic collections. For digitizing works of art, these spectral sensitivities will result in large errors, even with the addition of color management, demonstrated in a latter section.

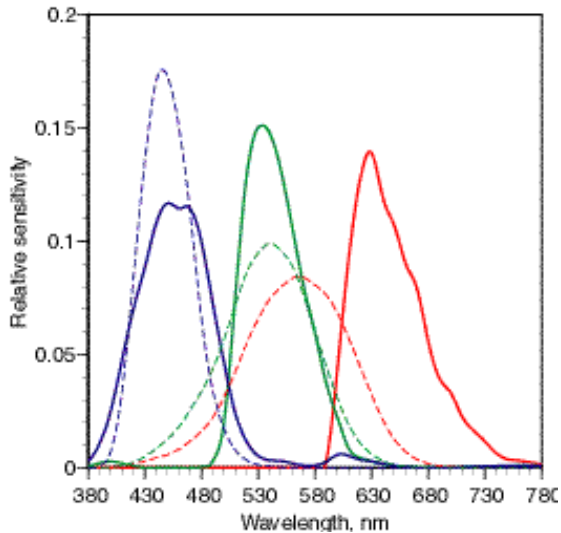


Figure 6. Spectral sensitivities of a typical scanback (solid lines), normalized to equal area. These sensitivities include the CCD spectral sensitivity, filter transmittances, and infrared radiation blocking filter. The human visual system's spectral sensitivities, shown in Figure 1, are also plotted (dashed lines).

Figure 7 is a plot of the spectral sensitivities of typical digital cameras employing color CCD two-dimensional arrays. There is considerably more overlap than densitometric scanbacks. However, the red sensitivity is still shifted towards longer wavelengths. It must be noted that color accuracy is only one of a number of criteria when designing a digital camera. Low light sensitivity, image noise, resolution, read-out

speed, manufacturing costs, and so on all must be considered. Many of these design criteria are mutually exclusive; the final design is always a compromise.

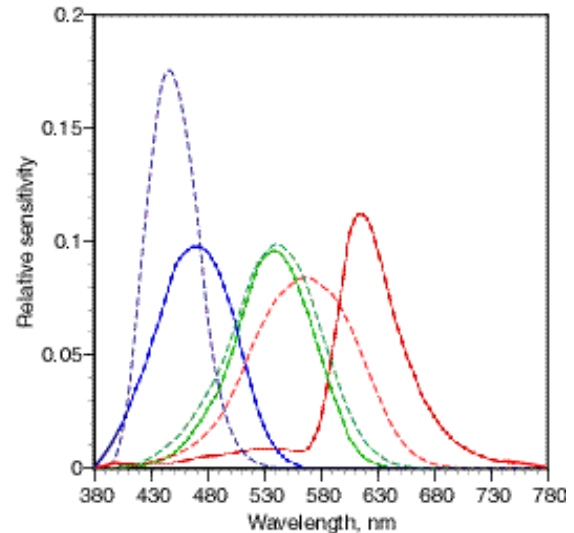


Figure 7. Spectral sensitivities of a typical color CCD two-dimensional array (solid lines), normalized to equal area. These sensitivities include the CCD spectral sensitivity, filter transmittances, and infrared radiation blocking filter. The human visual system's spectral sensitivities, shown in Figure 1, are also plotted (dashed lines).

Figure 8 is a plot of the spectral sensitivities of a monochrome scanner with a filter wheel. The key design criterion was color accuracy. The spectral sensitivities of this camera system are much closer to the human visual system than the other sensitivities shown in Figures 6 and 7. The red sensitivity has its peak response coincident with the eye's L sensitivity.

Finally, it is worthwhile showing the spectral sensitivity of a typical color transparency film,⁹ shown in Figure 9. Because of the inherent design of film in which sensitivities are stacked, limited choices for photosensitive materials, and a system that must simultaneously consider image capture and display, the spectral sensitivities do not closely resemble the human visual system. Furthermore, color photographic materials are designed to make pleasing reproductions, not necessarily accurate reproductions. For all of these reasons, photographic reproductions are not color accurate although with experience, it is possible through filtration using color-compensating filters to achieve global color matches. This is done routinely in museum photographic departments.

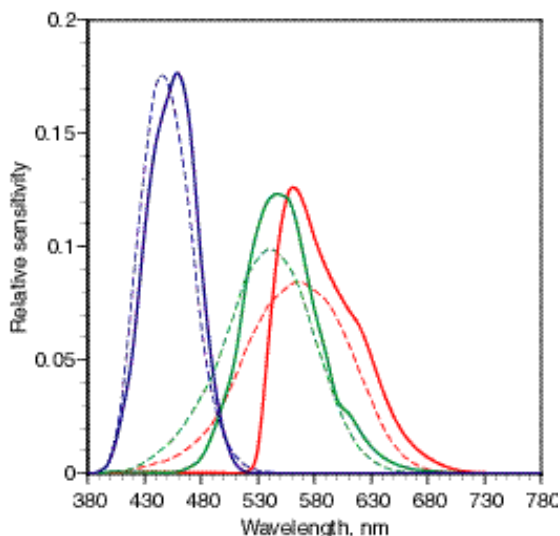


Figure 8. Spectral sensitivities of a monochrome scanback with three optimized color filters (solid lines), normalized to equal area. These sensitivities include the CCD spectral sensitivity, filter transmittances, and infrared radiation blocking filter. The human visual system's spectral sensitivities, shown in Figure 1, are also plotted (dashed lines).

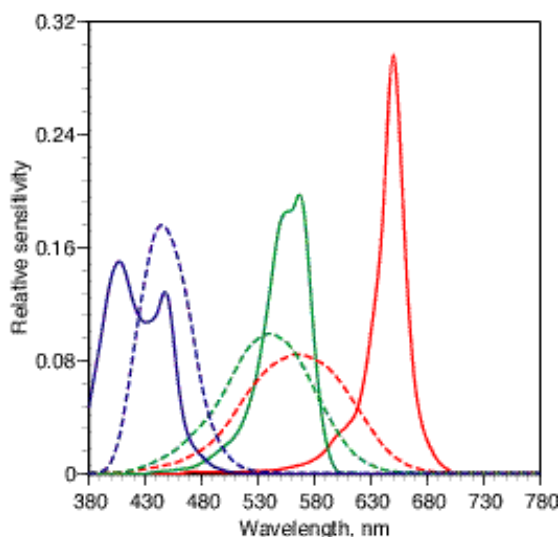


Figure 9. Spectral sensitivities of typical transparency film (solid lines), in this case Eastman Kodak Ektachrome 64 Tungsten. The human visual system's spectral sensitivities, shown in Figure 1, are also plotted (dashed lines).

CCD detectors have a linear response to incident light, shown in Figure 10. This is a dramatic difference to the nonlinearity shown in Figure 5. Depending on how the "raw" detector

signal is processed and the digital properties of the stored image, this difference can be inconsequential or result in large visual artifacts.

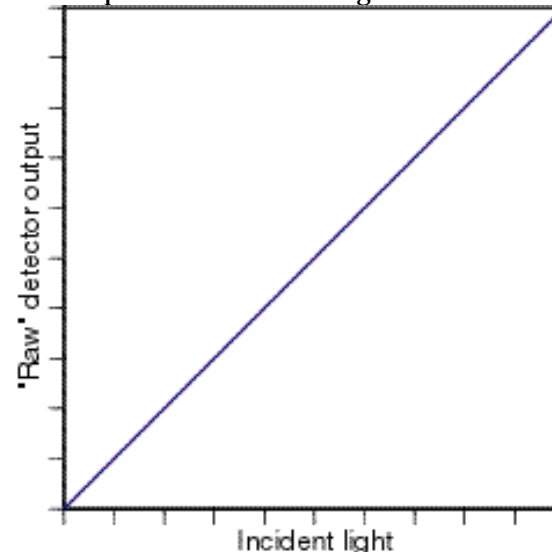


Figure 10. Relationship between incident light and raw detector signals for CCD arrays.

The raw detector signals, in the form of an analog signal (i.e., millivolts), is amplified and digitized using an analog-to-digital converter (ADC). The range of digital values depends on the number of bits in the ADC. Most commonly, 8, 12, or 16 bit ADC's are used. This means that there are 2^8 (256), 2^{12} (4096) or 2^{16} (65,536) number of gray levels for each color channel (also called "color levels" as well as "bit depth"). The greater the number of levels, the less the amount of error caused by the conversion from analog to digital signals. Visual artifacts caused by an insufficient number of gray levels are described and shown below in the next section.

Thus the camera has an inherent spatial and color resolution. The spatial resolution is determined in large part by the number of detectors. For example, high-resolution scanning systems commonly have one-dimensional arrays with 6000 detector elements that scan in 7000 or 8000 steps. Thus the camera has spatial resolution of 6000 X 7000 or 6000 X 8000 pixels ("picture elements"). This does not mean that a digital camera with more pixels is always better than one with less. Having many pixels combined with poor optics may have poorer image quality than fewer pixels with excellent optics. Ultimately, spatial resolution is determined for the entire camera system including the optics, sensor, and image processing (e.g., spatial interpolation) by imaging targets designed to quantify resolution

and mathematically analyzing the digital images of these targets.^{7, 11} Color resolution is determined by the number of bits in the ADC. For high-quality digital cameras, the ADC's have a minimum of 12 bits.

The Digital Image

A digital image is simply a two-dimensional array of numbers in which each number relates loosely to the amount of light reflected by an object. The image resolution defines the dimensions of the array. The concept of relating numbers to light reflection is easier to understand when a black and white (i.e., "monochrome" or "grayscale") image is considered. Essentially, dark areas have small numbers while light areas have large numbers. The most common bit depth for images is 8 bits per channel. Since digital values begin at 0, the numbers range from 0 to 255 (2⁸-1). A white is near 255 while a black is near 0.

There is a tendency to treat digital values as if they were perceptions: a perfect white is 255, a perfect black is 0, and a medium gray (e.g, Kodak Gray Card) is 128. (For color, all three channels have the identical digital values.) Although it is always true that a larger number corresponds to a lighter gray, the relationship between digital values and perception is rarely one to one. Relating digital values to perception requires system characterization using standard test targets.

This difficulty in relating numbers to perceptions becomes further exacerbated when an image is output and viewed. Simply taking a digital image and displaying it on different computer platforms and printing it out on different printers reveals the large range of renditions that result from the identical image file. Again, this problem is minimized through system characterization using standard test targets and measurement equipment. The origins of this problem stem from the inherent nonlinearity of many imaging devices.⁹ Conventional photography, broadcast television, computer-controlled CRT and LCD displays, and printing all have a nonlinear relationship between their input (e.g., exposure, voltage, digital counts, dot area) and their output, measured as reflectance, transmittance, or light energy.

A specific nonlinearity worth exploring is the nonlinearity of CRT displays, shown in Figure 11. By coincidence, this nonlinearity is similar to the human visual system, except inverted. This is an expansive function that is inherent to all vacuum tubes. During the early 20th century, a simple mathematical equation

was derived to relate input and output, shown as Eq. (1).^{12, 13} The normalized output was predicted by exponentiating the normalized input. The specific exponent is notated by the Greek symbol gamma, γ . The use of γ to notate an exponent stems from photographic science.¹⁴ However, a film's gamma has a different physical meaning than a display's gamma. Digital systems use gamma as shown in Eq. (1).[†]

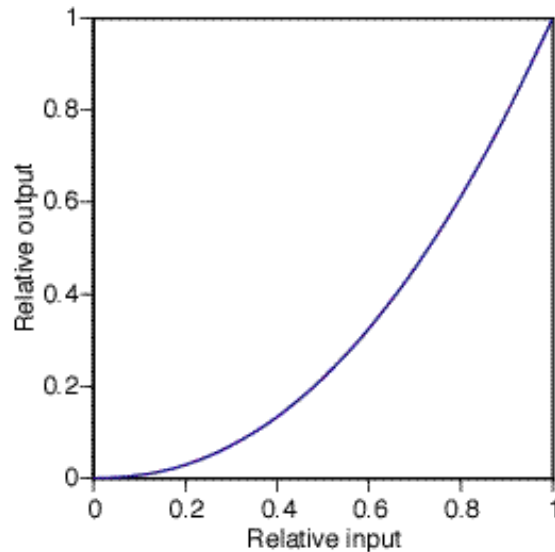


Figure 11. CRT nonlinearity. (This nonlinearity is inherent to all vacuum tubes.^{12, 13})

$$\frac{\text{output}}{\text{maximum output}} = \frac{\text{input}}{\text{maximum input}} \quad (1)$$

Because CRT displays have inherent nonlinearity, raw detector signals from digital cameras will lead to poor image quality unless the image is first "gamma corrected," that is, the inherent nonlinearity is appropriately compensated. The raw signals are exponentiated by the inverse of the display's gamma, shown in Eq. (2).

[†] This is somewhat of a simplification. For very small digital counts, the exponent is replaced with a slope term, for example, sRGB.¹⁵ Furthermore, this equation will only accurately model computer-controlled CRT displays if the monitor has its black level set perfectly: at 0 digital code value, there is not any emitted light while at 1 digital code value, light is emitted. Since this display set up is rarely achieved, more complex equations have been derived.^{16, 17}

$$\text{output} = (\text{max.output}) \frac{\text{rawinput}^\gamma}{\text{max.rawinput}^\gamma} \quad (2)$$

The net effect of gamma correcting the raw image and the display's inherent nonlinearity is an image that appears correct. That is, the tonal

properties appear reasonable. This is shown in Figure 12. Imposing a gamma function to a digital image is a common and useful procedure.

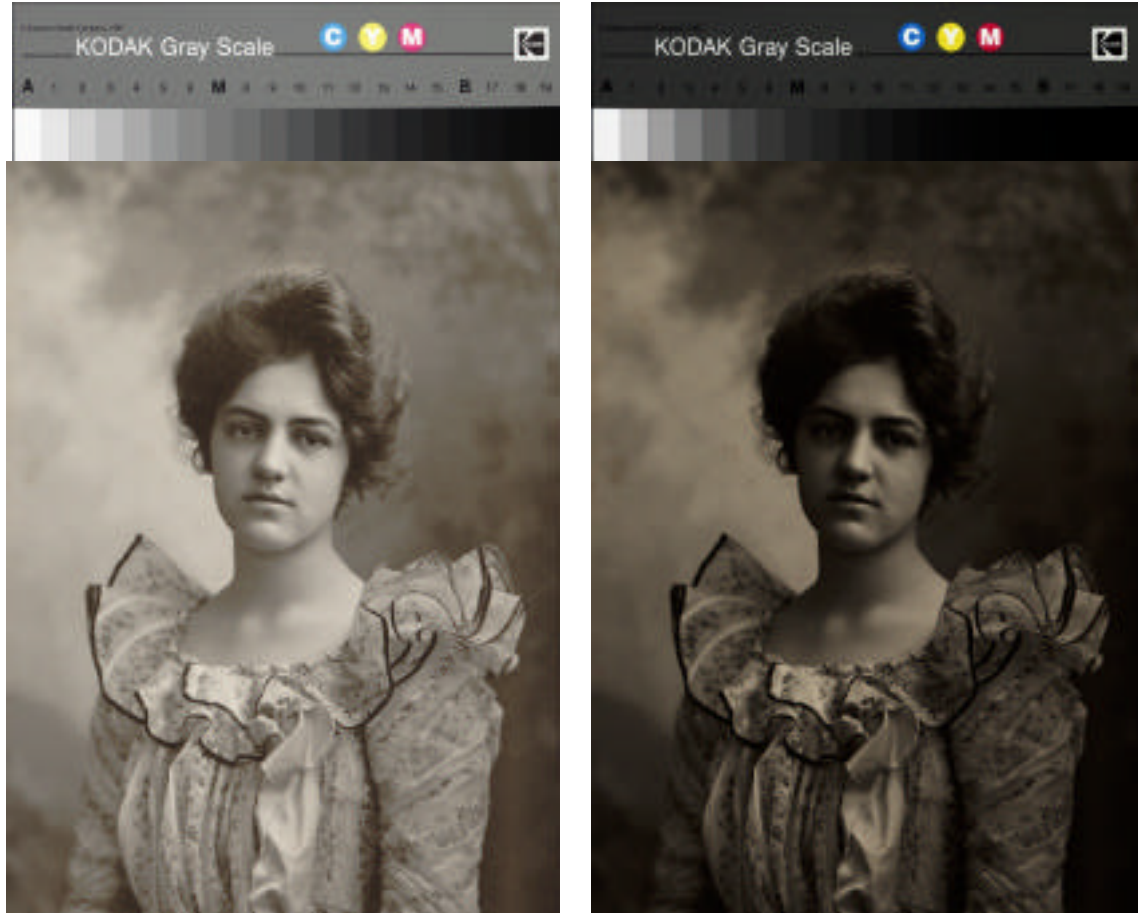


Figure 12. Photograph and gray scale imaged linearly and displayed with (top) and without (bottom) “gamma correction.” (Image courtesy of C. McCabe)

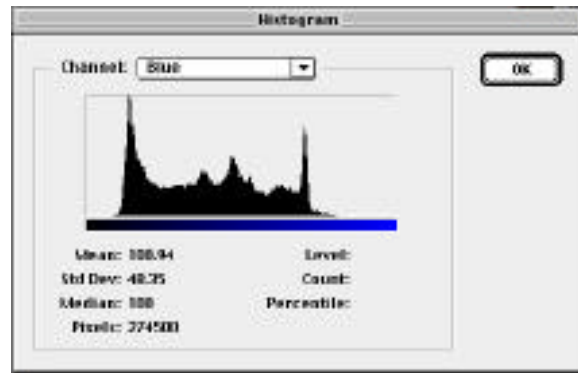
The vast majority of color images are 24-bit images, 8 bits for each of the red, green, and blue channels. Yet, for high-quality digital cameras, their analog-to-digital converters are 12 or 16 bits per channel. In order to have an 8-bit per channel image, the 4096 or 65,536 levels from the camera must be transformed to 256 levels. This is a reduction of information, and in similar fashion to spatial compression, this should be done in a manner that minimizes visual artifacts. The most common artifacts are banding in which smooth gradations become banded, or blocking in which details are lost,

shown in Figure 13. These are known as quantization errors. When a continuous signal (i.e., analog signal) becomes discrete (i.e., digital), it is quantized into a number of specific levels. The specific number of levels is determined by the bit depth. For all imaging applications, 2^{16} number of levels is a sufficient number of levels such that quantization errors are essentially eliminated. However, when converting to 8 bits per channel, quantization errors can be quite noticeable.

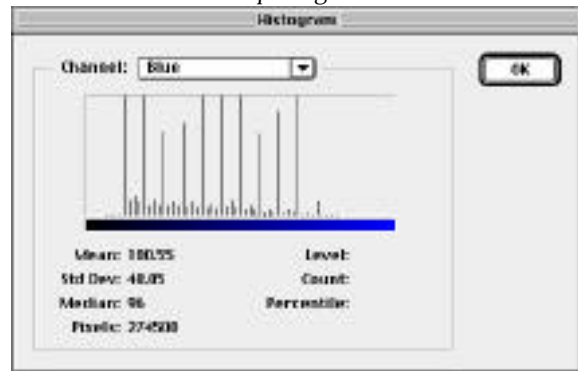


Figure 13. The effects of quantization on image quality, shown in the bottom figure. Notice that smooth gradations become banded and that fine detail in the subject's hair are lost.

A useful method to analyze quantization is by evaluating image histograms. The blue-channel histograms for the Figure 13 images are shown in Figure 14, created using the histogram analysis tool of Adobe Photoshop®. The height of a peak represents the number of pixels with a particular digital value. The far left represents 0 digital counts (black) and the far right represents 255 digital counts (blue). The left-hand image has a histogram in which each level between the minimum and maximum digital counts has a number of pixels. Conversely, the right-hand image is missing data, resulting in only a few levels. The visual banding is quantified via an image histogram. Similar histograms result for the red and green channels.



top image



bottom image

Figure 14. Histograms of the images shown in Figure 13.

Assessing Color Image Quality

Obviously, a digital capture system is capable of transforming an object into a digital image. The important question is whether the digital image is an accurate reproduction. Can the digital values be used to estimate the color of the object? Even though any "picture tells a thousand words," digital-image archives should

do more; they should facilitate scientific documentation and study of works of art. Analyzing color accuracy is most easily accomplished by imaging test targets along with the work of art and comparing standard values of the test target with its estimates. The standard values of interest are those that predict what observers see. CIELAB provides these required values.

Two test targets are used. The first is a gray scale, such as the Kodak Gray Scale (Catalog 152 7654). This target characterizes the relationship between digital values and lightness. It also characterizes whether the image is gray-balanced. That is, each channel should have matched digital values for the gray scale. Although it is common practice to include the Kodak Separation Guide (Catalog 152 7654) along with the gray scale, this target cannot be used to effectively evaluate color errors. Ideally, color targets should span the color range of imaged objects and have similar spectral properties (i.e., made using similar colorants). For photographic and printed materials, standard targets are available.^{18, 19} For art objects, targets are largely nonexistent. A target that is often used in broadcast television in order to evaluate the color accuracy of broadcast television cameras is the GretagMacbeth ColorChecker Color Rendition Chart (usually called the ColorChecker chart),²⁰ shown in Figure 15. It is produced using painted papers. Because it has a number of pigments, it is a convenient target to evaluate the accuracy of digitized paintings. Both targets can be cut up to make smaller targets as well as purchased in smaller forms so that resolution is not significantly reduced when including these targets along with the object undergoing digitization.

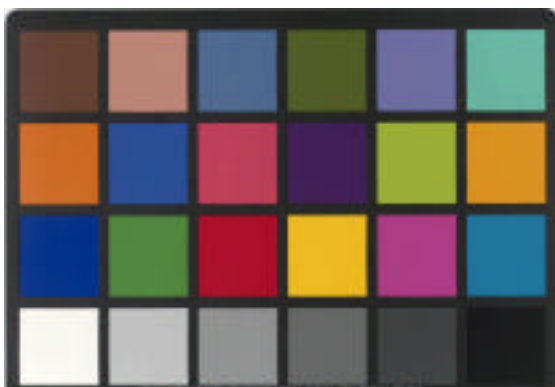


Figure 15. GretagMacbeth ColorChecker Color Rendition Chart.

Both targets are first measured using a spectrophotometer, usually having bidirectional

geometry, a geometry that minimizes specular reflection from contributing to the measured values.¹ From the spectral measurements, CIELAB values are calculated for a standard illuminant and observer, the specific standards dependent on the intent of the archive. The choice of illuminant and observer is complex. Should the object's color be defined by its appearance in an exhibition gallery, its appearance during conservation, or its appearance as envisioned by the artist? Exhibition galleries can have a range of illumination from bluish natural daylight (7500 K) to yellowish incandescent (2200 K).[‡] Conservators tend to use a combination of natural and artificial daylight, probably averaging to 6500 K. Depending on the artist, there can be a single or multiple choice of illumination. Notice that the choice depends on the work of art and not its color reproduction. Quite often, the image archive is defined by how the archive will be used. If the end product is a printed publication, CIE illuminant D50 and the 1931 standard observer are used.²¹ If the end product is web based, CIE illuminant D65 and the 1931 standard observer are used.¹⁵ These output-oriented images should be considered derivative images, created using principles of color management.^{1, 9, 22} For the analyses in this article, CIE illuminant D65 and the 1964 standard observer are used. In the author's experience, this combination leads to the best correlation between numerical and visual color quality. Furthermore, this corresponds to viewing objects in a natural-daylight lit studio with north-facing windows or typical conservation laboratories.

Assuming that the digital camera system is gray-balanced, the gray scale is used to ascertain whether the image is linear or nonlinear with respect to light input. Because CIELAB L^* is nonlinearly related to incident light as shown in Figure 5, luminance factor, instead is used. Luminance factor, also known as CIE tristimulus value Y , is linearly related to incident light and is a standard of light measurement.⁵ (It is also used in the calculation of L^* .) The average digital values for each patch from the two images of the Kodak gray scale shown in Figure 12 are plotted against luminance factor in Figure 16. As expected, the

[‡] 7500 K and 2200 K refers to correlated color temperature, the temperature of a blackbody radiator which generates a white light matching the color of the source of interest. Sources and displays are often defined in terms of their correlated color temperatures.

raw digital counts are linearly related to incident light. The “gamma corrected” digital values are nonlinearly related to incident light. This nonlinearity can compensate for typical display nonlinearities, such as that shown in Figure 11. This results in images that have reasonable tone reproduction, as demonstrated in Figure 12. (Because of various types of imaging noise, the curves are not perfectly smooth.)

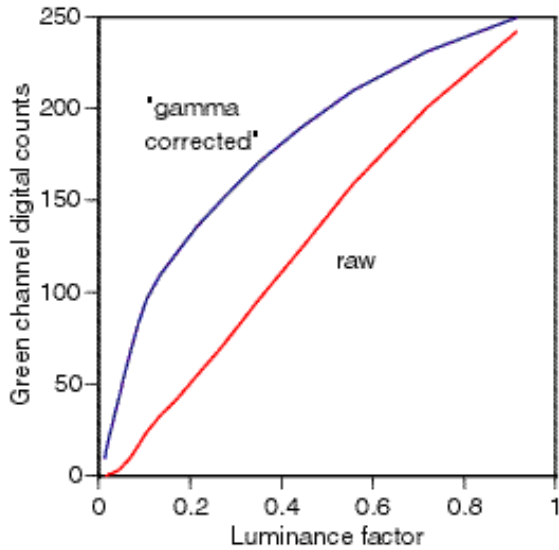


Figure 16. Luminance factor plotted against green channel digital counts for the Kodak gray scales from Figure 12. (Similar relationships were found for the red and blue channels.)

A scanback-type digital camera with spectral sensitivities similar to those shown in Figure 6 was used to digitize the Kodak Gray Scale and ColorChecker chart.²³ The raw signals were used to estimate L^* , a^* , and b^* coordinates. The general methodology for this estimation is described below. Both numerical and graphical analyses are performed. Graphically, “vector plots” are made as shown in Figure 17. The upper-most portion of Figure 17 is an a^*b^* projection. The end of the tail represents the measured coordinates while the head represents the estimated coordinates. Vectors pointing either towards or away from the origin (the center of the dashed lines at $a^* = 0$, $b^* = 0$) indicate chroma error. (Chroma is defined as the degree of departure of a color from a gray of the same lightness.¹) Vectors pointing away from the origin indicate the estimate is over-predicting chroma. Vectors pointing towards from the origin indicate the estimate is under-predicting chroma. Vectors pointing in directions other than towards or away from the origin indicate hue errors. The length of the

vector indicates the magnitude of the chromatic error. Ideally, the errors should be sufficiently small such that only the arrowhead is shown without a tail. There are some systematic trends: yellows have large chroma errors, blues have large hue errors, and reds and greens have relatively smaller errors. The bottom graph of Figure 17 shows an $L^*C^*_{ab}$ projection where C^*_{ab} represents chroma ($\sqrt{a^{*2} + b^{*2}}$). Vectors that are pointing upwards indicate that the estimation is too light; vectors pointing downward indicate that the estimation is too dark. Vectors parallel with the C^*_{ab} axis indicate accurate lightness estimation. There are two systematic trends. First, many of the chromatic samples’ lightnesses have been underestimated since the vectors are pointing downwards. Second, the lighter samples of the ColorChecker’s gray scale have been over estimated (vectors pointing upwards) and the darker neutrals have been under estimated (vectors pointing downwards).[§]

[§] Ordinarily, errors of this type for neutral samples indicate calibration errors in which there is a mismatch between actual and assumed photometric properties of the image capture system. In this case, these estimation errors were caused by spatial nonuniformities in illumination. Nonuniform illumination of the Kodak gray scale was interpreted as a slight nonlinear photometric response. This was compensated for during color management. Because the spatial nonuniformities varied across the image plane, this compensation resulted in systematic errors in different image locations.

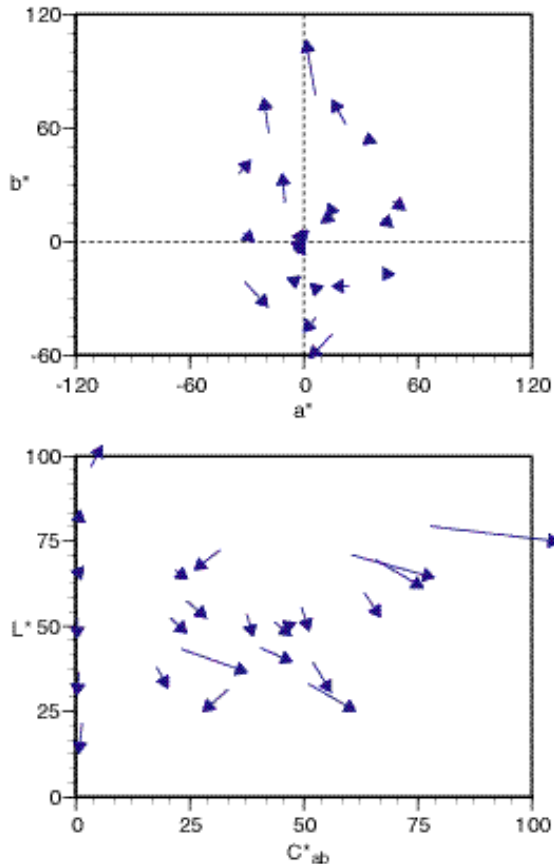


Figure 17. CIELAB vector plots in which the arrowhead defines the coordinates of the estimated values while the end of the arrow tail defines the coordinates of the measured values.

These differences are quantified numerically by calculating differences in color positions. CIELAB is a rectangular color space with L^* , a^* , and b^* axes (shown in Figure 4). Thus L^* , a^* , and b^* values are calculated where the Greek symbol (Δ) represents difference. L^* is calculated by subtracting the measured value from the estimated value (i.e., $L^* = L^*_{\text{estimated}} - L^*_{\text{measured}}$), and in similar fashion for the other coordinates. A positive value indicates that the estimated value has more of the particular quantity than the measured value. It is also useful to describe differences in lightness, chroma (C^*_{ab}), and hue (H^*_{ab}), rather than lightness, redness/greenness, and yellowness/blueness, particularly for chromatic samples. All of these values have been tabulated in Table 1. With experience, the numerical information can be quite useful. In many cases, the graphical information is much more intuitive.

An obvious question concerns the magnitude of error. Are these errors large or

small? Are they visible? Are they objectionable? Simply stated, these errors are large and objectionable, particularly since the goal is to estimate colorimetry from an image archive. One problem is the systematic nature of some of the errors shown in Figure 17. The second problem is the size of these errors. When all three dimensions are considered simultaneously, the length of the vector can be used as an error metric. In CIE terminology, this is called a total color difference and is notated by E^*_{ab} . However, CIELAB as a color-difference space is poorly correlated with visual judgments of color difference.¹ The last column in Table 1 lists a weighted color difference, E^*_{94} , also referred to as CIE94. This equation was designed for industries manufacturing colored products.²⁴ The weighting was optimized to improve correlation with visual tolerances. A CIE94 color difference of unity is slightly above a visual threshold. Thus, if pairs of colored patches were prepared with these differences in their CIELAB coordinates, the differences would all be obvious. This is simulated in Figure 18 in which the top half of each color simulates the estimated coordinates and the bottom half simulates the measured coordinates using a spectrophotometer. Clearly, this camera system is a poor analytical tool.

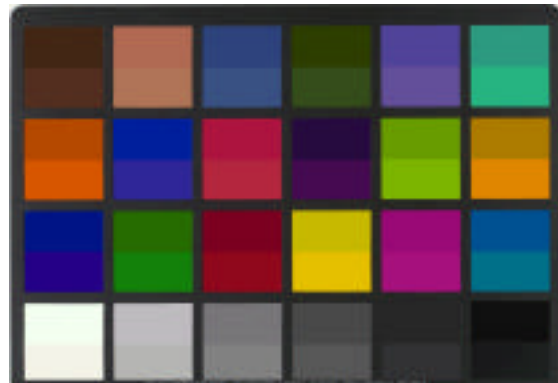


Figure 18. Simulation of color errors. The top half of each patch are the estimated values; the bottom half are the measured values. (Note that the estimated values for yellow are outside the color gamut of both printed and displayed media.)

Table 1. Colorimetric errors comparing measured and estimated values for a GretagMacbeth ColorChecker color rendition chart imaged with a scanback-type digital camera followed by color management.

Sample	L*	a*	b*	C* _{ab}	H* _{ab}	E* _{ab}	E* ₉₄
Dark Skin	-6.0	-0.3	3.6	2.5	-2.6	7.0	6.5
Light Skin	-2.6	4.1	-0.4	2.6	3.2	4.9	3.8
Blue Sky	-4.2	0.5	-3.9	3.7	-1.3	5.8	4.7
Foliage	-6.8	-1.4	14.9	14.5	4.0	16.5	10.3
Blue Flower	-4.9	-1.0	-4.7	4.3	2.1	6.8	5.5
Bluish Green	-5.8	5.2	-4.6	-5.6	-4.2	9.1	6.9
Orange	-7.0	-0.8	4.5	3.4	-3.1	8.4	7.2
Purplish Blue	-4.1	-6.6	-7.5	6.9	7.1	10.7	6.5
Moderate Red	-3.8	4.8	-4.3	3.7	5.3	7.5	5.1
Purple	-6.4	-8.8	-0.3	-5.4	7.0	10.9	8.2
Yellow Green	-6.5	-2.5	18.0	18.0	2.7	19.3	8.2
Orange Yellow	-8.2	-7.1	12.6	10.2	-10.2	16.7	10.0
Blue	-8.1	-11.7	-12.5	10.4	13.6	18.9	11.6
Green	-6.5	6.4	7.0	1.6	9.3	11.5	8.5
Red	-8.7	6.0	-4.2	4.0	6.1	11.3	9.4
Yellow	-4.6	-4.8	28.5	28.2	-6.1	29.3	8.3
Magenta	-3.3	-1.1	-1.8	-0.3	2.1	3.9	3.5
Cyan	-6.3	12.1	-12.7	1.3	-17.5	18.6	12.9
White	6.1	-3.9	1.1	2.4	-3.2	7.3	7.1
Neutral 8	1.5	0.6	-1.7	0.2	1.8	2.3	2.3
Neutral 6.5	-3.3	0.5	-1.2	1.1	0.7	3.5	3.5
Neutral 5	-5.7	0.3	-0.5	-0.2	-0.5	5.8	5.8
Neutral 3.5	-6.6	0.7	0.4	-0.5	-0.6	6.7	6.7
Black	-8.8	0.6	1.1	-0.8	-0.9	8.8	8.8
Average	-5.0	-0.4	1.3	4.4	0.6	10.5	7.1
Maximum						29.3	12.9

It is important to point out that CIE94 was designed to correlate with comparisons of the quality of colored patches, not colored images. As a rule of thumb, two pictorial images viewed side by side with systematic errors that result in an average E^*_{ab} of 2.5 or less are indistinguishable from one another.²⁵ If the errors are 5 E^*_{ab} on average, the accuracy is acceptable for high-quality color reproduction. However, the 2.5 and 5 E^*_{ab} rule of thumb applies to imaging systems where errors from both image input and output occur. Since this analysis is only considering image input, the errors should be even smaller. In the author's opinion, a well-designed image input device should result in average errors of less than

2 E^*_{ab} for the ColorChecker. Presently, there has been insufficient evaluation of CIE94's effectiveness for predicting the color quality of pictorial images and accordingly, recommended figures of merit cannot be given.

As a final analysis, a color-difference histogram should be evaluated, shown in Figure 19. Quite often, differences are not normally distributed and average and maximum errors can be misleading.

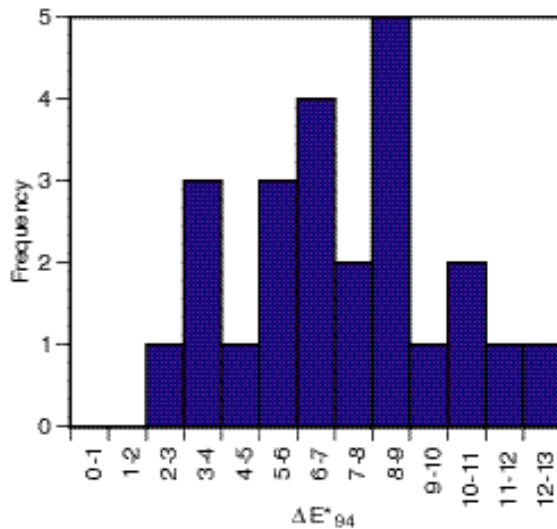


Figure 19. CIE94 histogram of the data given in Table 1.

Despite color management, these errors are large although not unexpected. It is not surprising given the large discrepancy between the camera and human spectral sensitivities, plotted in Figure 6. Although the camera does a fine job in recording red, green, and blue light, the human visual system does not “record” light in the same manner. No matter how complex the color management system, there will always be estimation errors.

Optimal Encoding – Lightness

From the review of digital imaging, minimizing quantization errors is critical in order to maximize an archive’s scientific value. This is achieved by the optimal encoding of the lightness properties of the work of art. The first step towards optimization is a review of the proper digital capture practices for two-dimensional works of art and documents, listed in Table 2.

Many of these practices are consistent with proper conventional photographic practices. It is worth stressing the importance of having sufficient illumination to keep exposure (scan) times short. With an increase in exposure time, noise accumulates.⁷ This is observed most easily by evaluating the blue channel in dark image areas. Noise appears as a random speckled pattern, shown in Figure 20. The blue channel is used because CCD detectors have their poorest sensitivity to the blue region of the visible spectrum and thus are most affected by noise accumulation.

Table 2. List of proper digital capture practices.

Object and image planes are parallel
Correct camera aperture for appropriate depth of field
Spatially uniform illumination across the object
Sufficient amount of illumination
Amount of ultraviolet and infrared radiation minimized (or eliminated by filtering)
Stray light minimized
Specular reflections minimized
Digital cropping is minimized
Calibration targets are included along with object
Exposure time appropriate to maximize dynamic range of raw digital data
Exposure times or amplification for each channel appropriate to yield gray balance



Figure 20. Incorrect exposure time or an insufficient amount of light results in excessive image noise, shown for the blue channel.

In conventional photography, many films are “very forgiving” if the image is either under- or over-exposed. Because most films are designed for photographing scenes with very large dynamic ranges, there is more exposure latitude when capturing two-dimensional works of art. CCD detectors have less exposure latitude; an incorrect exposure is “less forgiving.” If the exposure time is too long, the CCD array saturates causing blooming. A specular highlight smears in the direction of the one-dimensional array. Also, light areas are clipped, mapped to the same maximum value, resulting in a loss of highlight information. If the exposure time is too short, the raw data do not

encompass the full range of possible digital values. This results in quantization error.

A technique to evaluate the effects of quantization on image quality is to define an imaging system, assume that there is 1 bit of uncertainty, and evaluate the effect of this uncertainty on colorimetric error.²⁶ For this analysis, only L^* errors were considered based on analyzing the Kodak gray scale. If the raw camera data were 16 bit and the digital image encompassed the full possible dynamic range (0 - 65,535), there would not be any significant error as shown in Table 3. When the quantization is reduced to 12 bits, errors appear, largely for the darker samples. If the exposure times are too short, this is equivalent to reducing the bit depth. Thus the 10 and 8 bit data shown in Table 3 correspond to the effects of underexposure using a digital camera. The 10-

bit (0 - 1023) level of error could be observable in smoothly varying colors. At 8 bits (0 - 255), there are large errors, particularly for dark colors. Digital uncertainty from linear signals has a greater visual effect for dark colors. This is seen in the 8 bit data. The errors decrease with increasing lightness of the gray scale. This would result in large banding, especially in shadows and dark colors. Shadow detail would also be lost. Clearly, it is critical to set the exposure time properly.

The systematic trend, in which errors increase with decreasing lightness, is explained by the human visual system's nonlinear lightness response, shown in Figure 5. An error of one bit for a dark colors results in a larger visual difference in lightness than a one-bit error for a light color.

Table 3. The effects of quantization levels on lightness errors, ΔL^* .

Kodak gray scale	L^*	L^* for 16 bits	L^* for 12 bits	L^* for 10 bits	L^* for 8 bits
19	11.70	0.01	0.15	0.47	2.22
18	11.86	0.01	0.15	0.50	2.22
17	14.27	0.01	0.13	0.50	1.93
B	16.54	0.01	0.11	0.50	1.72
15	19.43	0.01	0.09	0.49	1.43
14	21.53	0.01	0.08	0.48	1.33
13	25.46	0.00	0.07	0.47	1.05
12	27.63	0.00	0.06	0.46	0.95
11	31.41	0.00	0.05	0.45	0.82
10	35.09	0.00	0.04	0.44	0.70
9	38.62	0.00	0.04	0.43	0.62
8	43.56	0.00	0.03	0.42	0.52
M	48.08	0.00	0.03	0.41	0.45
6	53.35	0.00	0.02	0.40	0.38
5	59.17	0.00	0.02	0.39	0.32
4	65.66	0.00	0.02	0.39	0.28
3	72.53	0.00	0.01	0.01	0.24
2	79.37	0.00	0.01	0.35	0.20
1	87.88	0.00	0.01	0.41	0.17
A	96.63	0.00	0.01	0.01	0.14
average		0.00	0.06	0.40	0.88
maximum		0.01	0.15	0.50	2.22

As Table 3 shows, 8-bit quantization results in large visual errors for linear signals. However, when 12 or 16 bit data are transformed to 8 bit per channel data, there isn't a restriction that the transformation be linear. It was noted above that by coincidence, the nonlinearities of CRT displays are nearly the exact inverse of the human visual system's nonlinearity. As a consequence, gamma-correcting signals as part of the bit-reduction

transformation has the "free" benefit of reducing visual errors. This is shown in Table 4 in which 12 bit signals encompassing the full dynamic range (0 - 4095) are gamma corrected [i.e., Eq. (2)], mapped to 8 bits, exponentiated by gamma, then transformed to lightness values. Mapping linear 12-bit raw signals nonlinearly onto 8-bit signals is clearly advantageous. The level of improvement is equivalent to increasing bit depth.

Table 4. The effects of gamma on lightness errors, ΔL^* .

Kodak gray scale	L*	L* for = 3	L* for_ =2.5	L* for = 2	L* for_ = 1	L* for scanned photography
19	11.7	0.00	0.00	0.01	2.22	1.43
18	11.86	0.03	0.04	0.08	2.22	1.43
17	14.27	0.17	0.2	0.27	1.93	1.31
B	16.54	0.23	0.26	0.34	1.72	1.13
15	19.43	0.29	0.31	0.38	1.43	1.01
14	21.53	0.31	0.33	0.39	1.33	0.92
13	25.46	0.35	0.36	0.40	1.05	0.76
12	27.63	0.36	0.37	0.41	0.95	0.69
11	31.41	0.38	0.38	0.40	0.82	0.59
10	35.09	0.39	0.39	0.40	0.7	0.49
9	38.62	0.40	0.39	0.39	0.62	0.43
8	43.56	0.41	0.39	0.38	0.52	0.34
M	48.08	0.42	0.39	0.37	0.45	0.27
6	53.35	0.42	0.39	0.36	0.38	0.21
5	59.17	0.43	0.38	0.35	0.32	0.17
4	65.66	0.43	0.38	0.34	0.28	0.18
3	72.53	0.43	0.38	0.32	0.24	0.24
2	79.37	0.43	0.37	0.31	0.20	0.32
1	87.88	0.43	0.37	0.30	0.17	0.40
A	96.63	0.44	0.36	0.29	0.14	0.46
average		0.34	0.32	0.32	0.88	0.64
maximum		0.44	0.39	0.41	2.22	1.43

The relationship between incident light, gamma-corrected signals, and normalized lightness (i.e., CIE $L^*/100$) are plotted in Figure 21. All these curves have similar shape; their differences are not significant in terms of how the encoding affects quantization errors. As a comparison, a Kodak gray scale was photographed using 4"x5" positive film in a museum photographic department, then digitized using a drum scanner with a linear response. The relationship between incident

light and raw 12-bit signals is plotted in Figure 22. This nonlinearity is typical of conventional photography, though quite different than what is usually shown as the nonlinear response of film. This is for two reasons: linear signals are plotted rather than logarithmic signals and when photographing artwork using controlled illumination, the full dynamic range of conventional film is not utilized. These data were used to perform a quantization error analysis in which the photographic nonlinearity was used rather than a gamma function in

transforming 12-bit to 8-bit per channel data, also shown in Table 4. The photographic nonlinearity has much larger error than any of the gamma-corrected signals. This means that using experiential understanding gained via conventional photography does not translate to optimal performance when applied to digital imaging.

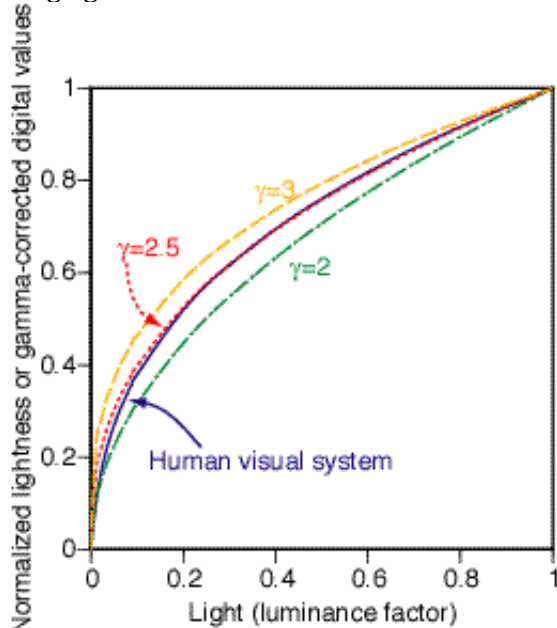


Figure 21. Relationship between light, expressed as luminance factor, and lightness (solid line) or gamma-corrected signals ($\gamma = 3$: dashed line; $\gamma = 2.5$: dotted line; $\gamma = 2$: dot-dashed line).

There are several different gamma functions used in the imaging industry. These can be either *de facto* or legitimate national or international standards. Choices include 1.8, the system gamma of Apple computer systems; 2.2, the inherent nonlinearity of CRT displays when modeled by Eq. (1) and recently incorporated into the sRGB display encoding;¹⁵ 3.0, the nonlinearity incorporated into CIELAB; 2.3, an empirical fit to CIE L^* used in the RLAB color-appearance model;^{27*} and 2.4, the sRGB exponent when an offset is included. The effect

* The CIE L^* equation, $L^* = 116(Y/Y_n)^{1/3} - 16$ where Y is luminance factor of a stimulus and Y_n is the luminance factor of the white object-color stimulus, has an offset term. If the offset is removed and the following equation is fit: $L^*/100 = (Y/Y_n)^{1/\gamma}$, a gamma of 2.3 results when the independent data are uniformly sampled by L^* (i.e., $Y/Y_n = 0.00, 0.01, 0.03, 0.06, 0.11 \dots$) and 2.4 when the independent data are uniformly sampled by Y/Y_n (i.e., $Y/Y_n = 0.00, 0.10, 0.20, 0.30, 0.40 \dots$).

of these particular gamma values on quantization errors is shown in Table 5, calculated in identical fashion to Table 4. For the three highest gamma values, the effects are nearly identical. For gamma of 1.8, the error increases slightly. The differences between these various gamma values, however, are not statistically significant. Therefore, the particular gamma correction is not critical.

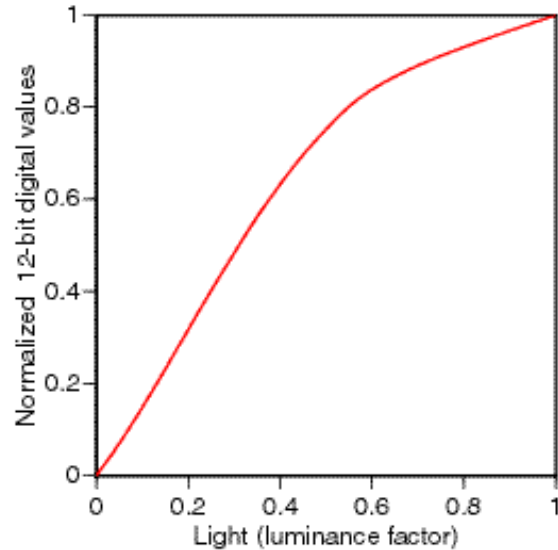


Figure 22. Relationship between light, expressed as luminance factor, and normalized 12-bit raw signals for a digitized photographic positive transparency.

Table 5. The effects of gamma on lightness errors, ΔL^* .

Kodak gray scale	L^*	L^* for = 2.4 (exp of 0.41)	L^* for_ =2.3 (exp of 0.43)	L^* for = 2.2 (exp of 0.45)	L^* for = 1.8 (exp of 0.56)
19	11.7	0.00	0.00	0.00	0.03
18	11.86	0.04	0.05	0.06	0.11
17	14.27	0.21	0.22	0.23	0.33
B	16.54	0.27	0.28	0.30	0.40
15	19.43	0.32	0.33	0.34	0.43
14	21.53	0.34	0.35	0.36	0.43
13	25.46	0.36	0.37	0.38	0.44
12	27.63	0.37	0.38	0.38	0.43
11	31.41	0.38	0.39	0.39	0.42
10	35.09	0.39	0.39	0.39	0.41
9	38.62	0.39	0.39	0.39	0.40
8	43.56	0.39	0.39	0.38	0.38
M	48.08	0.39	0.38	0.38	0.37
6	53.35	0.38	0.38	0.37	0.35
5	59.17	0.38	0.37	0.36	0.34
4	65.66	0.38	0.37	0.36	0.32
3	72.53	0.37	0.36	0.35	0.30
2	79.37	0.37	0.35	0.34	0.29
1	87.88	0.36	0.34	0.33	0.27
A	96.63	0.35	0.34	0.32	0.26
average		0.32	0.32	0.32	0.34
maximum		0.39	0.39	0.39	0.44

The above analyses lead to the following recommendations. First, the digital range of the raw data should be maximized, but without clipping. For 12-bit data, the maximum digital values should not exceed about 4000, likely specular highlights. For sample A of the Kodak Gray Scale, the average digital value should be around 3900. For 16-bit data, the maximum digital values should not exceed about 64,000 and sample A of the Kodak Gray Scale should have an average digital value of about 63,000. When writing images in standardized formats such as TIFF, the raw data should be directly saved as 16-bit TIFF. For 12-bit data, the digital values are premultiplied by 2^4 (i.e., bit shifted) so that the images can be viewed in Photoshop without adjustment. (This does not affect quantization errors.) If this premultiplication (bit shifting) is not performed the displayed image will be very dark. For 8-bit TIFF, the raw data must be rescaled and gamma corrected. The maximum digital value should not exceed

about 250. For sample A of the Kodak Gray Scale, the average digital value should be around 245. The minimum values for all bit depths should approach 0. However, for a variety of computational and image noise reasons, the values for sample 19 of the Kodak Gray scale will be about 5, 80, or 1000 for 8-, 12-, and 16-bit data, respectively. The key is ensuring that the available range of digital values is maximized during image acquisition. That is, it is critical to achieve proper exposure. If software is available to display raw data histograms, this can provide the needed information. Proper exposure should not be verified by looking at the digitized image. Since the display is 8 bits per channel, camera software will have user-controlled adjustments to impose a transfer function, sometimes called a "process curve", to convert from raw to display data. (The same curve will be used to generate the 8 bit TIFF file.) Although the image may look reasonable, quantization error is still being introduced.

Optimal Encoding - Color

Dealing with color is significantly more complex than optimally encoding lightness. As described above, most digital cameras in use today have inappropriate spectral sensitivities. If the sensor was optimized for photographic materials, shown in Figure 6, or for consumer applications, shown in Figure 7, they are not accurate color measurements instruments. For scientific imaging, the camera should be thought of as an imaging colorimeter or spectrometer. Several systems have been designed as imaging colorimeters: the IBM Pro 3000 system,^{28 - 30} the VASARI system,^{31 - 32} and the MARC system.³³ Systems designed as imaging spectrometers are still at the research and development stage, summarized in references 34 and 35.

An experiment was performed to evaluate to colorimetric potential of four imaging systems, three direct digital cameras and scanned conventional photography. Camera A has spectral sensitivities closely related to the human visual system, similar to those shown in Figure 8. It consists of a monochrome sensor and three color filters. Camera B has spectral sensitivities similar to those shown in Figure 7. It consists of a two-dimensional color CCD array. Camera C has spectral sensitivities similar to those shown in Figure 6. It is a scanback employing a flat-bed scanner-type trilinear color-filter array. An 8" x 10" view camera along with tungsten-balanced chrome 64 ASA film was used for conventional photography. Color-compensation filters were used to achieve a visual color match between neutral areas of the test target and its reproduction. The transparency was digitized using a flat-bed scanner with spectral sensitivities similar to Figure 6. The illumination for all four systems was tungsten halogen with a correlated color temperature around 3200 K. Image capture was performed adhering to the practices outlined in Table 2. All four digital systems have 12-bit analog to digital converters. The raw signals were recorded as 16 bit TIFF. There was not any signal processing such as gamma correction imposed. Thus, these images were "linear RGB." However, because of differences in each channel's light sensitivity, the spectral characteristics of tungsten illumination, and stray light, the range of digital values were quite different for each channel and for each imaging system.

Three targets were imaged: the Kodak Gray Scale, the GretagMachbeth ColorChecker Color Rendition Chart, and a custom target of 68 artist oil paints, shown in Figure 23. Each paint

was mixed with titanium white in order to maximize the spectral "fingerprint" of a given colorant. The Kodak Gray Scale and ColorChecker chart were used to derive a transformation from R, G, and B to L*, a*, and b* values. The paint target was used to evaluate the accuracy of the transformation. The transformation can be thought of as employing color management principles.

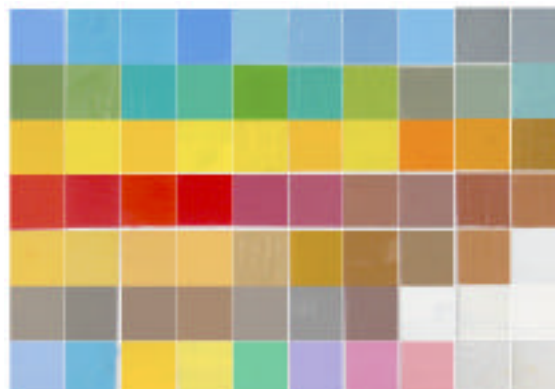


Figure 23. Artist oil paint target.

The first step was to develop a transform that linearized the raw data with respect to CIELAB. The general equation is shown as follows:

$$L^* = k_{g,2} k_{g,1} \frac{d}{65,535} + (1 - k_{g,1}) + k_o \quad (3)$$

where L* is the measured L* values of the Kodak Gray Scale, d is either the average raw R, G, or B digital values of each gray-scale sample, and the remaining terms are model coefficients. This particular equation enables amplification, properties of the ADC, stray light, and differences in measurement geometry between the spectrophotometer used to measure the Gray scale and the digital capture system to be taken into account. In most cases, not all of the model terms were statistically significant and the equation was reduced in complexity accordingly. This equation can be thought of as a more analytical form of the exponential model described by Eq. (1). Nonlinear optimization was used to estimate the model parameters for each channel and for each camera. The camera signals, linearized with respect to L*, are referred to as R*, G*, and B*. Equation (3) also facilitates gray balance. In the case of digitized conventional photography, a different approach was taken to linearize the raw signals because the nonlinearity shown in Figure 22 can not be

well fit using Eq. (3). A fifth-order polynomial was used, Eq. (4):

$$L^* = 100 + \frac{d}{65,535} + \frac{d^2}{65,535} + \frac{d^3}{65,535} + \frac{d^4}{65,535} + \frac{d^5}{65,535} \quad (4)$$

where $a_0 - a_5$ are model coefficients estimated by linear optimization.

The average R^* , G^* , and B^* values of each color patch of the ColorChecker were used to develop a linear transformation, Eq. (5):

$$\begin{aligned} \hat{L}^* &= a_{1,1}R^* + a_{1,2}G^* + a_{1,3}B^* \\ \hat{a}^* &= a_{2,1}R^* + a_{2,2}G^* + a_{2,3}B^* \\ \hat{b}^* &= a_{3,1}R^* + a_{3,2}G^* + a_{3,3}B^* \end{aligned} \quad (5)$$

where each term is a model coefficient. Hats (“^”) are shown over the CIELAB values because these are estimated values. The model coefficients were constrained via Eq. (6)

$$\begin{aligned} a_{1,1} + a_{1,2} + a_{1,3} &= 100 \\ a_{2,1} + a_{2,2} + a_{2,3} &= 0 \\ a_{3,1} + a_{3,2} + a_{3,3} &= 0 \end{aligned} \quad (6)$$

The purpose of Eq. (6) was to maintain gray balance. The model parameters for each camera were estimated using least-squares linear optimization. A numerical example of this approach to color management is presented in reference 1. This is referred to as a colorimetric transformation. Although polynomial expansions of Eq. (5) are often used for color management, these often do not result in improved performance when independent data are evaluated.

Before evaluating the independent data, the oil paint target, it is useful to evaluate the modeling performance of the Kodak Gray Scale and ColorChecker chart. The average CIE94 color difference between measured and estimated values for the Gray Scale varied between 0.5 for camera A and 1.0 for camera B. The estimation performance for the ColorChecker chart is shown in Figures 24 and 25 and in Table 6. As expected, the closer the camera’s spectral sensitivities are to the human visual system, the better the colorimetric

estimation. The causes for the large errors for camera C were discussed above. The errors for camera A and B are random indicating that this simple transformation is performing appropriately. It was surprising that the performance for camera B was not better. The MARC system, having similar spectral sensitivities, has a reported estimation error of $2.5 E^*_{ab}$ for the ColorChecker [33] whereas for this analysis, the average was $6 E^*_{ab}$. Further analyses revealed that the colorimetric performance was very sensitive to the illumination system’s spectral power distribution, spatial uniformity, and illuminance, as well as setting the optimal exposure time. Prefiltering the source in order to balance the three channels and improve its daylight characteristics, numerical flat-fielding in order to achieve spatial uniformity, signal averaging, and more careful attention to exposure reduced the average estimation error to $3.5 E^*_{ab}$.^{††} All of the camera systems’ colorimetric estimation accuracies were highly dependent on image capture characteristics underscoring the importance of proper capture procedures listed in Table 2. The colorimetric accuracy of digitized film was intermediate between cameras B and C, consistent with film’s spectral sensitivities shown in Figure 9.

^{††} The Marc I camera employed a Sony ICX021AK sensor whereas camera B employed a Sony ICX085AK sensor. The latter sensor has improved light sensitivity but its spectral sensitivities are poorer approximations to the human visual system. Thus, the reduction in estimation accuracy from 2.5 to $3.5 E^*_{ab}$ is consistent with these differences.

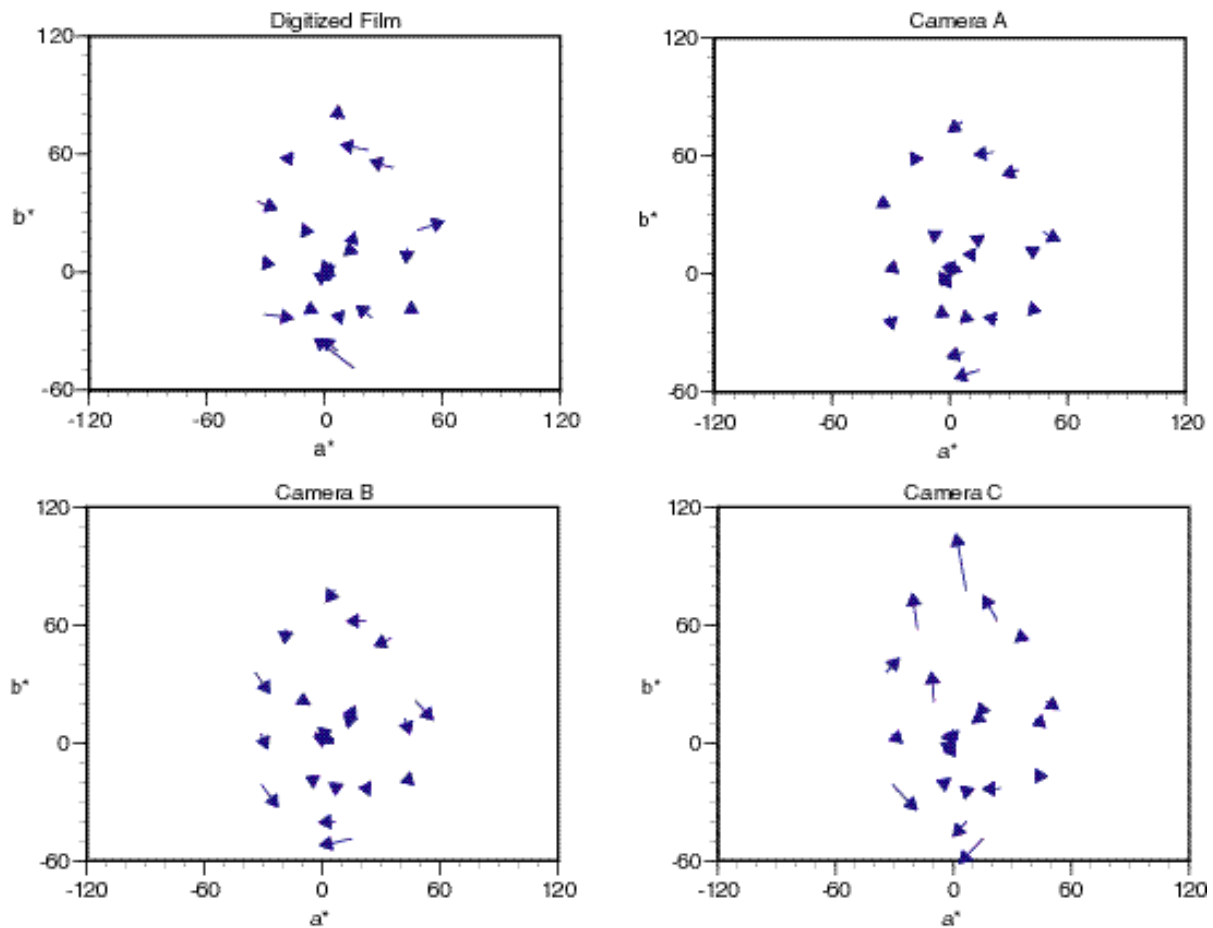


Figure 24. CIELAB a^*b^* projection vector plots showing estimation errors for the ColorChecker chart.

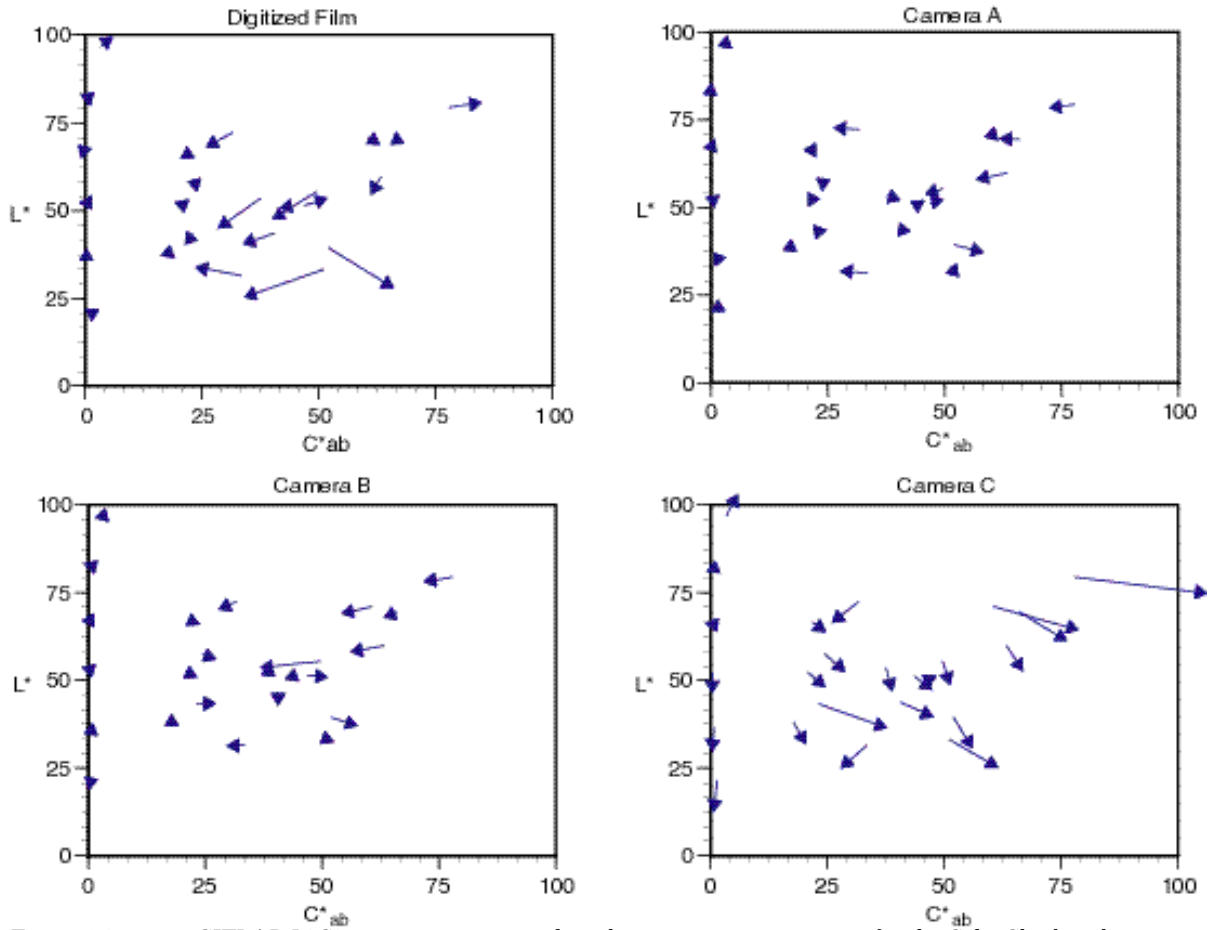


Figure 25. CIELAB $L^*C^*_{ab}$ projection vector plots showing estimation errors for the ColorChecker chart.

Table 6. CIE94 color differences between measured and estimated colorimetric coordinates for the ColorChecker chart.

Sample	Digitized Film	Camera A	Camera B	Camera C
Dark Skin	2.0	0.4	2.7	6.5
Light Skin	3.3	2.0	4.1	3.8
Blue Sky	2.8	1.9	1.9	4.7
Foliage	3.9	0.9	2.3	10.3
Blue Flower	3.3	1.4	2.0	5.5
Bluish Green	4.8	3.2	5.7	6.9
Orange	7.9	3.6	3.6	7.2
Purplish Blue	6.0	5.0	5.0	6.5
Moderate Red	4.4	2.6	5.1	5.1
Purple	4.6	3.5	2.9	8.2
Yellow Green	2.6	1.3	3.0	8.2
Orange Yellow	7.1	4.7	5.0	10.0
Blue	13.9	7.9	9.2	11.6
Green	8.8	2.0	4.6	8.5
Red	12.1	4.9	7.3	9.4
Yellow	2.1	3.5	2.7	8.3
Magenta	2.6	1.1	1.7	3.5
Cyan	11.8	4.4	9.1	12.9
White	3.4	3.4	1.7	7.1
Neutral 8	1.9	1.9	2.3	2.3
Neutral 6.5	1.5	2.0	2.3	3.5
Neutral 5	1.9	2.2	1.3	5.8
Neutral 3.5	2.1	0.5	1.0	6.7
Black	3.2	2.1	0.4	8.8
Average	4.9	2.8	3.6	7.1
Maximum	13.9	7.9	9.2	12.9

The estimation accuracy for the artist oil paint target is shown in Figures 26 – 28 and in Table 7. The trends for the direct digital capture systems were the same as the ColorChecker accuracy: Camera A had the least estimation error while camera C had the greatest estimation error. The digitized photographic system had the poorest estimation accuracy of the four imaging systems. For each imaging system, there were systematic errors; that is, certain color regions were not accurately estimated. This is observed by evaluating the error vectors. For example, the $+a^*+b^*$ and $-a^*-b^*$ quadrants had vectors pointed in similar directions for camera C. Lightness was underestimated for the scanned photographic system. Ideally, error

vectors should be largely random and small, such as those shown for the ColorChecker in Figure 24. Because CIELAB is not visually uniform, caution should be taken when evaluating error-vector plots. The same length vector corresponds to a different visual difference depending on the chroma of the measured color. Thus, the overall accuracy is judged by looking at the histograms plotted in Figure 28. Clearly, camera A has the highest estimation accuracy. One important property is to minimize the maximum errors. Camera A's largest error is $6.8 E^*_{94}$. The errors form a tight grouping. As a comparison, the digitized film has errors up to $12 E^*_{94}$ and the grouping is broadly dispersed.

The differences in estimation accuracy between the ColorChecker and the artist paint target underscores the importance of independent verification. Achieving high

estimation accuracy for a characterization target is much simpler than achieving high accuracy when using the system for imaging works of art.

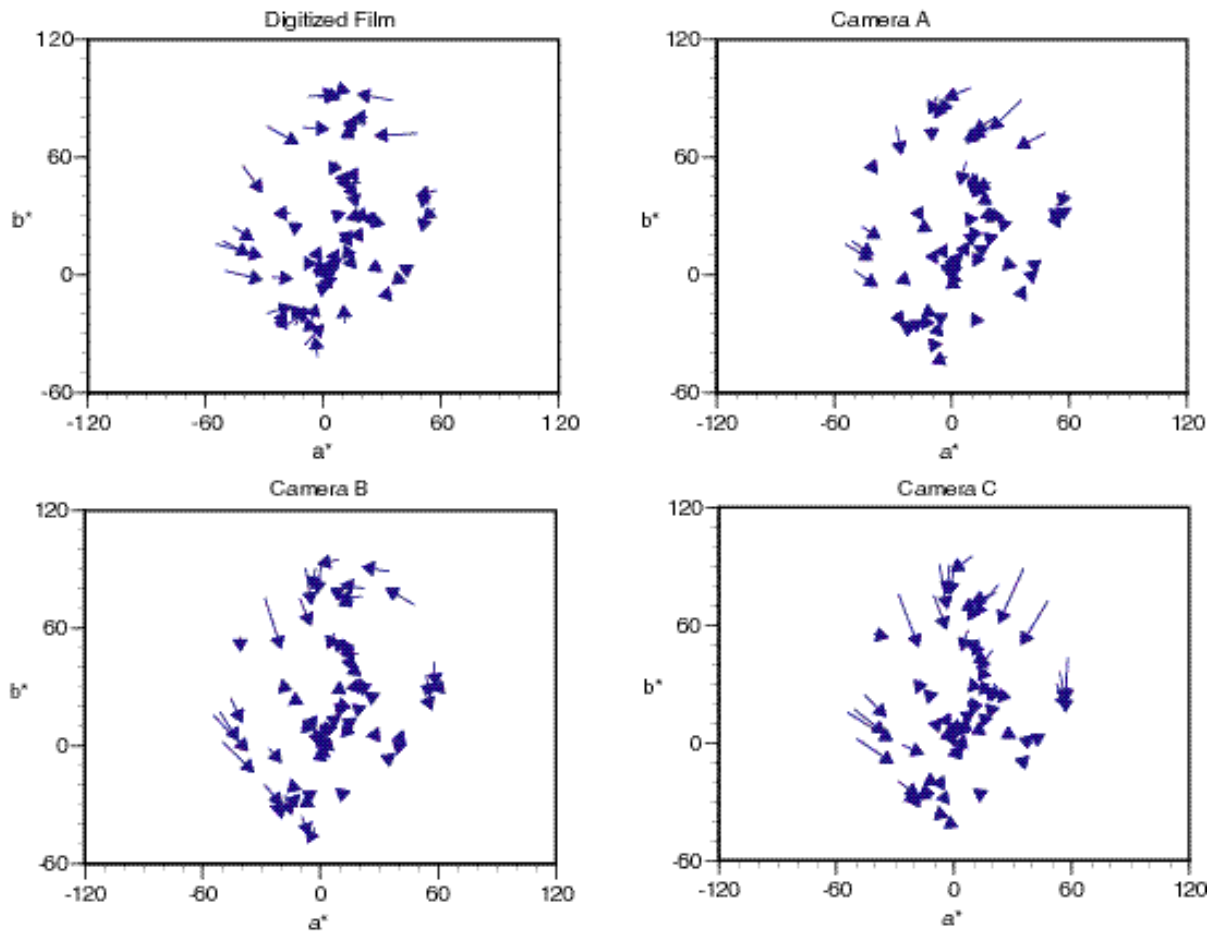


Figure 26. CIELAB a^*b^* projection vector plots showing estimation errors for the artist oil target.

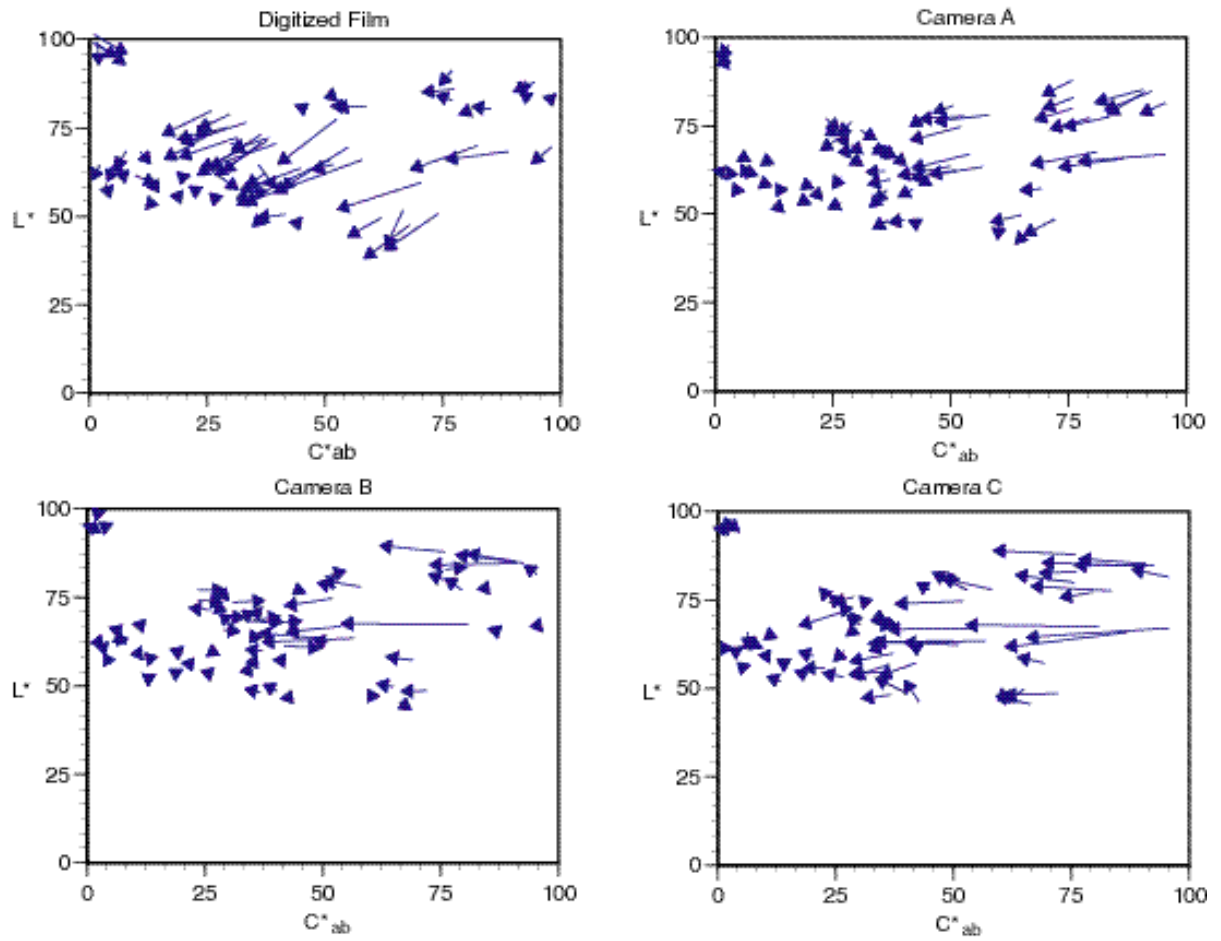


Figure 27. CIELAB $L^*C^*_{ab}$ projection vector plots showing estimation errors for the artist oil target.

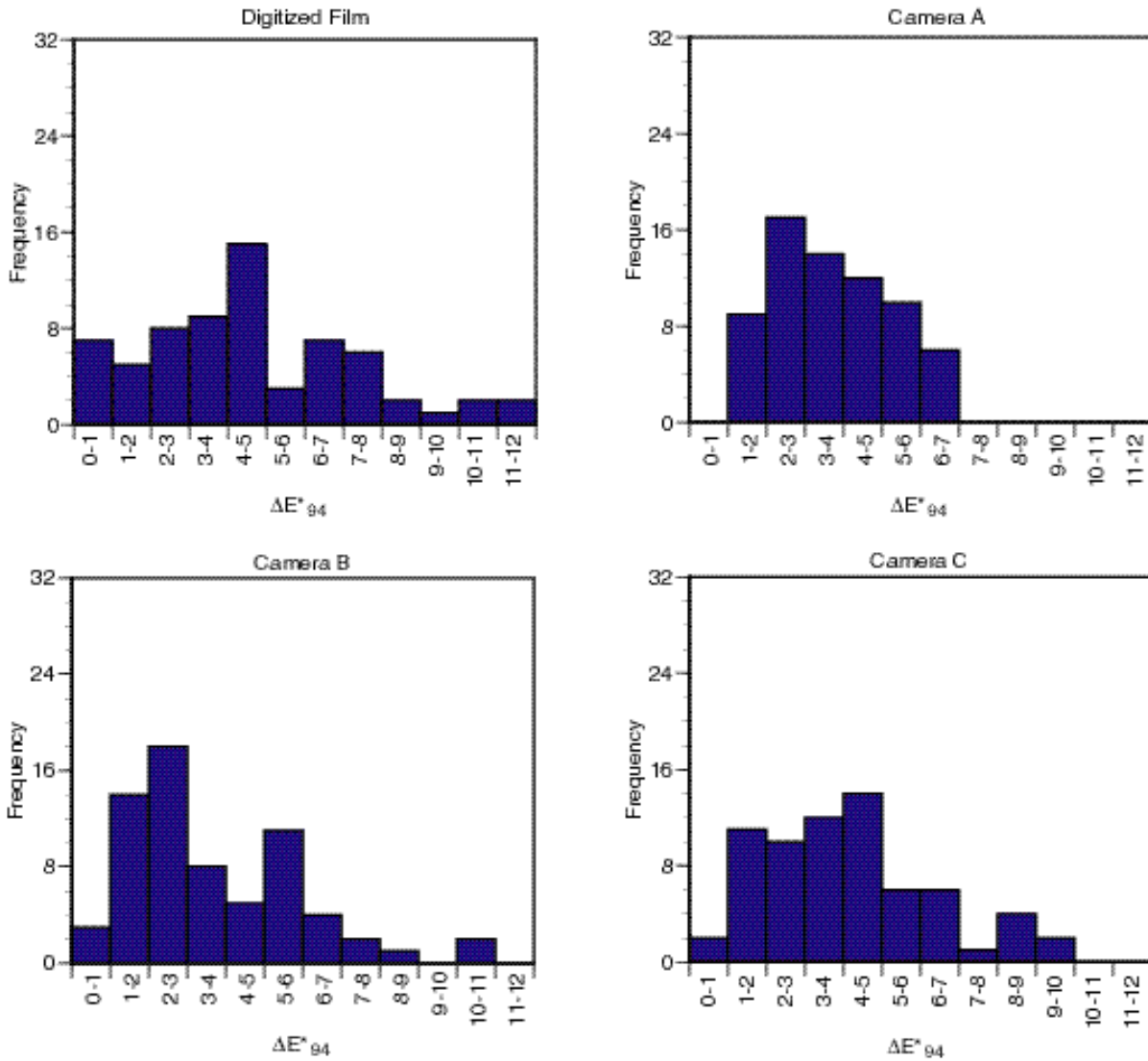


Figure 28. CIE94 color-difference histogram showing estimation errors for the artist oil target.

Table 7. CIE94 color differences between measured and estimated colorimetric coordinates for the artist oil paint chart.

Sample	Digitized Film	Camera A	Camera B	Camera C
Cobalt blue	7.7	4.2	4.7	3.9
Manganese blue	7.8	3.1	5.2	3.8
Manganese blue	6.3	2.9	5.6	4.7
Ultramarine blue	5.2	5.2	3.3	1.9
Prussian blue	4.6	2.5	2.7	2.0
Cerulean blue	5.3	2.8	3.4	2.9
Indanthrone blue	4.0	3.6	2.4	1.6
Phthalocyanine blue	6.6	3.2	5.9	4.8
Slate gray	1.6	1.0	1.4	3.0
Graphite gray	3.0	1.5	2.0	4.6
Sap green, mixture	3.9	3.2	2.9	3.3
Sap green, permanent	6.4	1.6	5.2	4.0
Phthalocyanine green	12.0	6.4	10.7	9.7
Phthalocyanine green, yellow	11.2	5.7	7.6	6.9
Permanent green light	6.6	1.2	2.6	3.1
Permanent green deep	11.8	6.0	10.0	8.9
Cadmium green	7.6	4.7	5.8	6.1
Terre verte- Gamblin	3.0	2.1	1.0	1.5
Terre verte - Williamsburg	1.3	2.3	1.2	2.4
Viridian	7.1	4.9	6.2	7.2
Cadmium yellow deep	4.9	5.5	6.0	5.4
cadmium yellow light	3.4	6.3	3.7	3.4
Cadmium yellow medium	4.5	4.6	2.9	4.4
Cadmium yellow pale	4.0	5.1	3.1	3.2
Basic yellow medium	0.7	6.1	4.8	4.9
Hansa yellow deep	4.7	5.4	5.2	5.5
Hansa yellow light	6.4	6.8	3.7	4.6
Cadmium orange	8.2	5.3	7.3	6.7
Indian yellow	7.6	5.8	5.7	6.3
Transparent earth yellow	2.9	2.8	1.6	4.5
Cadmium vermilion red light	9.3	4.2	5.3	8.3
Cadmium red	4.8	5.8	6.5	6.9
Cadmium red light	8.7	5.2	5.2	9.2
Napthol red, yellow shade	7.9	4.7	5.0	8.7
Alizarin crimson	2.9	1.0	2.4	6.1
Alizarin permanent	3.4	1.6	2.8	5.4
Burnt sienna, Gamblin	0.9	2.4	2.0	3.1
Burnt sienna, Williamsburg	0.7	2.0	2.3	2.2
Transparent earth red	1.3	2.8	2.3	2.6
Transparent earath orange	2.2	3.4	1.4	4.9
Naples yellow, hue	4.4	3.4	4.7	4.2
Naples yellow, blockX	4.0	3.8	3.1	4.5
Naples yellow	3.7	3.3	1.6	2.6

Naples Yellow Italian	2.7	3.0	1.9	5.5
Stil de grain	4.1	1.8	3.1	3.1
Yellow ochre	6.5	3.4	4.6	4.1
Raw sienna	2.9	3.0	2.3	3.0
Yellow ochre burnt	2.2	1.6	1.5	2.3
Italian earth	3.0	2.3	1.4	4.3
Permalba white	2.3	4.7	1.8	2.1
Raw umber, Gamblin	2.4	2.3	3.3	1.2
Raw umber, Williamsburg	1.8	2.2	2.5	0.7
Burnt umber	0.7	2.6	2.6	1.4
Asphaltum	0.8	2.6	1.8	2.2
Transparent brown	0.9	2.7	0.9	1.8
Van Dyke brown	1.9	2.2	0.7	1.2
Brown madder alizarin	0.8	1.3	1.7	2.3
Titanium white	4.1	4.0	1.4	1.3
Replacement Flake white	4.9	4.6	1.0	1.6
Alky white	4.9	4.5	2.9	3.5
Blue radiant	5.6	3.6	2.9	0.6
Turquoise radiant	10.4	6.0	8.7	8.2
Yellow radiant	4.0	5.6	5.9	5.9
Lemon radiant	6.0	4.5	4.0	4.6
Green radiant	10.6	4.9	6.1	5.4
Violet radiant	4.5	3.6	2.1	1.6
Magenta radiant	3.1	2.8	0.9	1.5
Red radiant	3.1	3.9	2.5	3.0
Average	4.7	3.7	3.6	4.1
Maximum	12.0	6.8	10.7	9.7

As an estimation problem, residual errors result, the magnitude of the errors dependent on the inherent properties of the camera and also on the method of estimation. A linear method has been described; nonlinear methods may result in improved performance. This is an evolving area of research. Furthermore, with each successive estimation, such as occurs when transforming between various encoding schemes, errors accumulate. Therefore, the recommended approach to color encoding is to retain the raw data. If a standardized format is required, 16 bit TIFF should be used. If 8 bit data must be stored, gamma correction between 1.8 and 2.4 should first be applied. Color management should not be applied to the raw data. Rather, color management should be applied for various derivative images. The necessary information to implement reasonable color management stems from the inclusion of imaged calibration targets

and their spectral and colorimetric data as metadata.

The Future

As conventional color measurement developed during the 20th century, there was a transition from predominantly colorimeters to predominantly spectrophotometers. That is, trichromatic instruments were replaced with spectral instruments. The principal advantages were numerous. The spectral data enabled an object's color to be calculated for any illuminant and observer of interest. Thus, the quality of a color match could be evaluated under any number of illuminating and viewing conditions. Second, the spectral data enabled a variety of analytical analyses such as colorant identification, instrumental-based color matching, and colorant strength.¹ Finally, and perhaps most importantly, the presence or absence of metamerism could be readily observed. As an illustrative example, Pablo

Picasso's *The Tragedy* was restored in 1937 and 1997. During 1937, paint losses were filled and inpainted. This is shown in Figure 29; the inpainted areas appear purplish when photographed. Spectral measurements of untreated and inpainted areas reveal that Picasso used prussian blue while the conservator used ultramarine blue, shown in Figure 30. Unfortunately, most of the 1937 inpainting could not be removed during 1997; hence the colorant mismatch persists. Because the spectral sensitivity of the red sensitive layer of film is shifted towards longer wavelengths

compared with the human visual system, shown in Figure 9, the ultramarine was reproduced as a grayish purple rather than a grayish blue. Spectral measurements would have alerted the conservator that a metameric match was produced. This specific problem with blue colorants has been described by Staniforth.³⁴ Clearly, an imaging spectrometer would be a useful analytical tool. Furthermore, the lack of colorimetric accuracy, causing the large color shift in this example, and the need to standardize an illuminant and observer, as described above, would be eliminated.



Figure 29. *The Tragedy*, Pablo Picasso 1903, oil on wood, 1.053 x .690 m (41 7/16 x 27 3/16 in.) Chester Dale Collection, National Gallery of Art, Washington: Photographed using Kodak Ektachrome 64T and tungsten illumination.

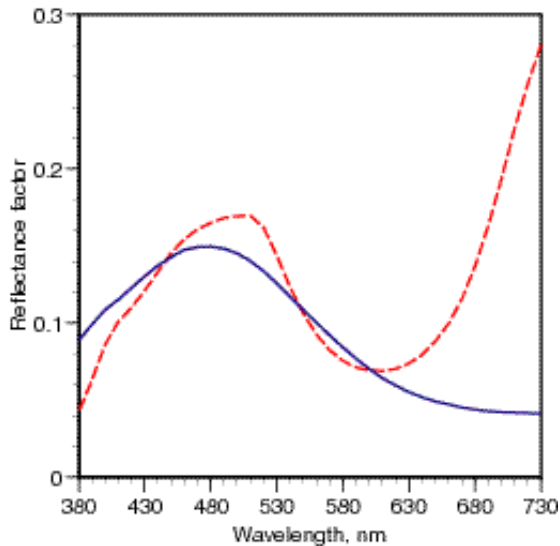


Figure 30. Spectral reflectance factor measurements of *The Tragedy* (solid line) and 1937 inpainting (dashed line).

Spectral-based imaging of artwork is currently at a research stage, summarized in references 35 and 36. Using camera A, a novel approach to spectral imaging was evaluated in which a second image, taken by filtering the camera lens with a light green filter (Wratten 38) was combined computationally to estimate the spectral reflectance of each sample.^{37 - 39} The spectral and camera data of the Kodak Gray Scale and ColorChecker chart were used to derive the necessary transforms. Following spectral estimation, CIELAB coordinates were calculated in the usual way. The ColorChecker was estimated to an average accuracy of $0.3 E^*_{94}$. The artist's paint target had an average accuracy of $2.0 E^*_{94}$, nearly a twofold improvement in colorimetric accuracy.

One of the very interesting aspects of this spectral estimation approach is that spectra are generated from a conventional digital color camera. The spectra can be used for a variety of color reproduction and conservation applications such as multi-ink printing^{40 - 41} and colorant identification.⁴² Using only spectral information from the ColorChecker chart, the estimated spectral reflectance factor data for four blue colorants from the artist's paint target are plotted in Figure 31. The fits are fairly typical of spectral-estimation techniques using broadband sensors: the estimates tend to have greater spectral selectivity. The overall shapes are reasonably predicted. A statistical method of colorant identification based on reflectance spectrophotometry⁴² was tested using these estimated spectra. The following colorants formed the database of possible blue colorants: cobalt, ultramarine, manganese, prussian, phthalocyanine, cerulean, and indanthrone. The database was formed from the artist's paint target. The cobalt and ultramarine blue spectra were correctly identified. Manganese blue was incorrectly identified as phthalocyanine blue. This was due to the secondary peak at 650 nm in the estimated spectrum. The prussian blue sample was incorrectly identified as manganese blue. The spectral estimation was repeated, except that spectral information from the painted target was used in place of the ColorChecker. The estimated spectra are also plotted in Figure 29. In all cases the spectral fits are improved. The colorant identification was repeated. Only the prussian blue sample was incorrectly identified, again as manganese blue. Given the known similarity in spectral properties between prussian and manganese blues,³⁴ these results were very encouraging.

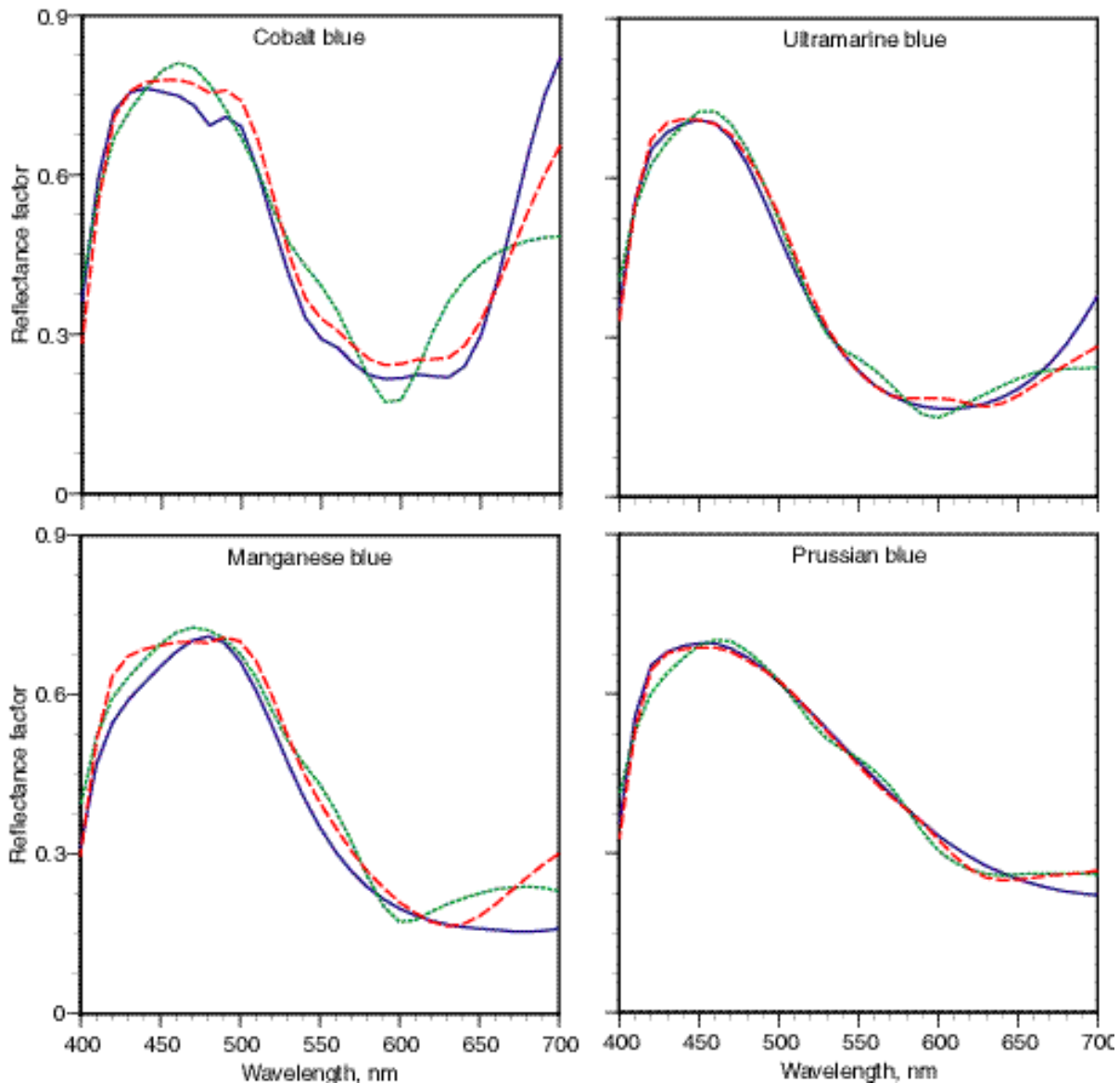


Figure 29. Spectral estimation of four samples from the artist oil target shown in Figure 23: measured, solid line, estimated from spectral data of ColorChecker, dotted line, estimated from spectral data of target, dashed line.

Conclusions

A digital image of a work of art should facilitate a variety of applications including web-based images, color reproduction, scientific study, art historical and other scholarly studies, and most importantly, an accurate recording of the physical properties of the work of art. As described, achieving an accurate recording is difficult and requires specialized hardware, optimal imaging practices, and specialized software.

Two issues need to be emphasized. The first is that scanback sensors are very common

among the museum community. Since the number of pixels is often the dominant criterion when specifying a imaging system, scanbacks fare quite well. However, the majority of tri-filter scanbacks are densitometric rather than colorimetric. As described above, this results in very poor color accuracy. There are two solutions: redesign the filter array or use a monochrome sensor and filter-wheel assembly. In general, greater emphasis needs to be placed on a sensor's spectral sensitivity. For accurate color imaging, sensors need to be closely related to the human visual system.

The second issue to be emphasized is that visual techniques have not been included in this article. This was deliberate and not an omission. Visual matching between objects and their CRT representation is very complex and is highly dependent on viewing conditions,^{9, 27} cognitive aspects given the dissimilar nature of light-emitting and light-reflecting stimuli,²⁷ the observer,² and monitor calibration.^{15 - 17} Furthermore, color management must be used, an evolving area of color technology.^{1, 9, 22} Transforming the raw camera signals to a derivative image such as an sRGB image should be a computational exercise, not a visual adjustment. The calculations are straightforward (a numerical example is given in reference 1) and eliminate the natural tendency to boost contrast and color saturation.

For those engaged or soon to be engaged in creating a digital archive, the recommendations described above should be seriously considered. At the very least, standardized (actual or *de facto*) gray scale and color targets must be included in each image.

References

1. R. S. Berns, *Billmeyer and Saltzman's Principles of Color Technology*, 3rd ed., John Wiley & Sons, New York, 2000.
2. B. A. Wandell, *Foundations of Vision*, Sinauer Assoc. Inc., Sunderland MA, 1995.
3. A. Stockman, D. I. A. MacLeod, and N. E. Johnson, 'Spectral sensitivities of the human cones', *J. Opt. Soc. Am. A* **10**, 2491-2521 (1993).
4. ISO/IEC DIS 10918-4, 'Information technology -- Digital compression and coding of continuous-tone still images: Registration of JPEG profiles, SPIFF profiles SPIFF tags, SPIFF colour spaces, APPN markers, SPIFF compression types and Registration Authorities (REGAUT)', International Organization for Standardization, Geneva, Switzerland.
5. CIE No. 15.2, *Colorimetry*, 2nd ed., Commission Internationale de l'Éclairage, Vienna, Austria (1986).
6. D. S. Falk, D. R. Brill, and D. G. Stork, *Seeing the Light: Optics in Nature, Photography, Color, Vision, and Holography*, John Wiley & Sons, New York, 1986.
7. G. Holst, *CCD Arrays, Cameras, and Displays*, 2nd ed, Intl. Soc. For Optical Eng., Bellingham, WA, 1998.
8. H. E. Ives, 'The transformation of color-mixture equations from one system to another', *J. Franklin Inst.* **180**, 673-701 (1915). Reproduced in D. L. MacAdam, Ed., 'Selected papers on colorimetry – fundamentals', SPIE Milestone Series, Vol. MS 77, Intl. Soc. For Optical Eng., Bellingham WA, 1993, pp. 57-71.
9. E. J. Giorgianni and T. E. Madden, *Digital Color Management Encoding Solutions*, Addison Wesley, Reading, MA, 1998.
10. R. S. Berns and M. J. Shyu, 'Colorimetric characterization of a desktop drum scanner using a spectral model', *J. Electronic Imag.* **4**, 360-372 (1995).
11. http://www.pima.net/standards/iso/tc42/wg18/kp_sfr_measure.htm from the Photographic and Imaging Manufacturing Association.
12. C.D. Child, 'Discharge from hot CaO', *Phys. Rev.* **32**, 492-511 (1911).
13. I. Langmuir, 'The effect of space charge and residual gases on thermionic current in high vacuum', *Phys. Rev.* **2**, 450-486 (1913).
14. T. H. James and G. C. Higgins, *Fundamentals of Photographic Theory*, Morgan & Morgan Inc., Hastings-on-Hudson, NY, 1968.

The image metadata must contain information about the targets and details of the image-capture system.

Acknowledgements

This publication was written while the author was a Senior Fellow in Conservation Science at the National Gallery of Art, Washington; the Gallery's financial support is sincerely appreciated. The author gratefully acknowledges the assistance of Janet Bliburg, Lorene Emerson, and Lyle Peterzellin collecting digital and conventional images, and Ross Merrill in producing the artist oil paint chart, all of the National Gallery of Art, Washington. Finally, the following imaging professionals were very helpful in reviewing manuscripts and providing the benefits of their extensive experiences: Connie McCabe, National Gallery of Art; Edward Giorgianni, Eastman Kodak Company; Mike Collette, Better Light; John Stokes, Stokes Imaging; Michael Stokes, Microsoft Corporation; and Francisco Imai, Munsell Color Science Laboratory.

15. IEC 61966-2-1, Colour measurement and management in multimedia systems and equipment - Part 2-1: Colour management - Default RGB colour space - sRGB, International Electrotechnical Commission, Geneva, Switzerland (1999).
16. R. S. Berns and N. Katoh, 'The digital to radiometric transfer function for computer controlled CRT displays', Proc. CIE Expert Symposium '97 Colour Standards for Image Technology 34-37 (1997).
17. T. Deguchi, N. Katoh, and R. S. Berns, 'Clarification of 'gamma' and the accurate characterization of CRT monitors', Proc. SID **30**, 786-789 (1999).
18. D. Q. McDowell, 'Summary of IT8/SC4 color activities', Proc. SPIE **1913**, 229-235 (1993).
19. D. Q. McDowell, 'Graphic arts standards update - 1995', Proc. SPIE **2413**, 323-332 (1995).
20. C. S. McCamy, H. Marcus, and J. G. Davidson, 'A color-rendition chart', J. Appl. Phot. Eng. **2**, 95-99 (1976).
21. ISO 13655, Graphic technology -- Spectral measurement and colorimetric computation for graphic arts images, International Organization for Standardization, Geneva, Switzerland.
22. ICC.1: 1998-09, 'File format for color profiles International Color Consortium', www.color.org (1998).
23. This imaging system is being used at the National Gallery of Art, Washington for the following imaging project: <http://www.nga.gov/feature/stieglitz/asmmain.htm>.
24. CIE No. 116, Industrial colour-difference evaluation, Commission Internationale de l'Éclairage, Vienna, Austria (1995).
25. M. Stokes, M. D. Fairchild, and R. S. Berns, 'Colorimetrically quantified tolerances for pictorial images', Proc. Tech. Assoc. Graphic Arts part 2, 757-778 (1992).
26. P. D. Burns and R. S. Berns, 'Quantization in multispectral color image acquisition', Proc. 7th IS&T/SID Color Imaging Conference, 32-35 (1999).
27. M. D. Fairchild, *Color Appearance Models*, Addison-Wesley, Reading, MA, 1998.
28. F. C. Mintzer, L. E. Boyle, A. N. Cazes, B. S. Christian, S. C. Cox, F. P. Giordano, H. M. Gladney, J. C. Lee, M. L. Kelmanson, A. C. Lirani, K. A. Magerlein, A. M. B. Pavani, and F. Schiattarella, 'Towards on-line worldwide access to Vatican library materials', *IBM J. Research and Development*, **40**, 139-162 (1996). Also available as <http://www.almaden.ibm.com/journal/rd/mintz/mintzer.html>.
29. F. Mintzer, 'Developing Digital Libraries of Cultural Content for Internet Access', *IEEE Communications* **37**(1), 72-78 (1999).
30. F. P. Giordano, G. W. Braudaway, J. Christensen, J. Lee, and F. Mintzer, 'Evolution of a high-quality digital imaging system' Proc. SPIE **3650**, 110-118 (1999).
31. K. Martinez, J. Cupitt, and D. Saunders, 'High resolution colorimetric imaging of paintings', Proc. SPIE **1901**, 25-36 (1993).
32. D. Saunders, J. Cupitt, 'Image processing at the National Gallery: The VASARI project', *National Gallery Tech. Bull.* **14**, 72-86 (1993).
33. J. Cupitt, K. Martinez, and D. Saunders, 'A methodology for art reproduction in colour: the Marc project', *Computers and the History of Art*, **6**(2), 1-19 (1996).
34. S. Staniforth, 'Retouching and colour matching: The restorer and metamerism', *Studies in Conservation* **30**, 101-111 (1985).
35. L. MacDonald and M. R. Luo, Eds., *Colour Imaging: Vision and Technology*, John Wiley & Sons, Chichester, 1999.
36. *Proceedings Intl. Symp. Multispectral Imaging and Color Reproduction for Digital Archives*, Chiba University, Soc. Multispectral Imag. Japan, 1999.
37. F. H. Imai and R. S. Berns, 'High-resolution multi-spectral image archives - A hybrid approach', *Proceedings IS&T/SID Sixth Color Imaging Conference*, 224-227 (1998).
38. R. S. Berns and F. H. Imai, 'Spectral estimation using trichromatic digital cameras', Proc. Intl. Sym. Multispectral Imaging and Color Reproduction for Digital Archives, Chiba University, Soc. Multispectral Imag. Japan, 42-49 (1999).
39. F. H. Imai and R. S. Berns, 'A comparative analysis of spectral reflectance reconstruction in various spaces using a trichromatic camera system', Proc. IS&T/SID Seventh Color Imaging Conference, 21-25 (1999).
40. R. S. Berns, 'Challenges for colour science in multimedia imaging systems', in L. MacDonald, and M. R. Luo, Eds., *Colour Imaging: Vision and Technology*, John Wiley & Sons, England, 1999, pp. 99-127.

41. R.S. Berns, F.H. Imai, P. D. Burns, and D.Y. Tzeng, 'Multi-spectral-based color reproduction research at the Munsell Color Science Laboratory', *Proc. SPIE* **3409**, 14-25 (1999).
42. R. S. Berns, J. Krueger, and M. Swicklik, 'Colorant selection for inpainting using visible reflectance spectrophotometry', *Studies in Conservation*, submitted 2000.