
Automatic White Balancing in Digital Photography

Edmund Y. Lam and George S. K. Fung

10.1	Introduction	267
10.2	Human Visual System and Color Theory	268
10.2.1	Illumination	269
10.2.2	Object	270
10.2.3	Color Stimulus	271
10.2.4	Human Visual System	273
10.2.5	Color Matching	275
10.3	Challenges in Automatic White Balancing	278
10.4	Automatic White Balancing Algorithms	279
10.4.1	Gray World	280
10.4.2	White Patch	281
10.4.3	Iterative White Balancing	282
10.4.4	Illuminant Voting	284
10.4.5	Color by Correlation	286
10.4.6	Other Methods	287
10.5	Implementations and Quality Evaluations	288
10.6	Conclusion	292
	Acknowledgments	292
	References	292

10.1 Introduction

Color constancy is one of the most amazing features of the human visual system. When we look at objects under different illuminations, their colors stay relatively constant. This helps humans to identify objects conveniently. While the precise physiological mechanism is not fully known, it has been postulated that the eyes are responsible for capturing different wavelengths of the light reflected by an object, and the brain attempts to “discount” the contribution of the illumination so that the color perception matches more closely with the object reflectance, and therefore is mostly constant under different illuminations [1].

A similar behavior is highly desirable in digital still and video cameras. This is achieved via white balancing which is an image processing step employed in a digital camera imaging pipeline (detailed description of the camera imaging pipeline can be found in [Chapters 1](#) and [3](#)) to adjust the coloration of images captured under different illuminations [2], [3].

This is because the ambient light has a significant effect on the color stimulus. If the color temperature of a light source is low, the object being captured will appear reddish. An example is the domestic tungsten lamp, whose color temperature is around 3000 Kelvins (K). On the other hand, with a high color temperature light source, the object will appear bluish. This includes the typical daylight, with color temperature above 6000 K [4].

Various manual and automatic methods exist for white balancing. For the former, the camera manufacturer often has predefined settings for typical lighting conditions such as sunlight, cloudy, fluorescent, or incandescent. The user only needs to make the selection, and the camera will compute the adjustment automatically. Higher-end cameras, such as prosumer (professional-consumer) and single-lens reflex (SLR) digital cameras, would even allow the user to define his or her own white balance reference. Most amateur users, however, prefer automatic white balancing. The camera then needs to be able to dynamically detect the color temperature of the ambient light and compensate for its effects, or determine from the image content the necessary color correction due to the illumination. The automatic white balancing (AWB) algorithm employed in the camera imaging pipeline is thus critical to the color appearance of digital pictures. This chapter is devoted to a study of such algorithms commonly used in digital photography.

We organize this chapter as follows. In Section 10.2, we first briefly review the human visual system and the theory of color, which are necessary background materials for our discussion on AWB strategies in cameras. Certain terminologies would also be introduced that are commonly used in discussing color. This is followed by looking into the physical principles of color formation on an electronic sensor. The challenges that exist in digital photography are described in Section 10.3. Then, in Section 10.4, we describe a few representative AWB algorithms. Our goal is not to be encyclopedic, which is rather impossible considering the wide array of methods in existence and the proprietary nature of some of these schemes, but to be illustrative of the main principles behind the major approaches. In Section 10.5, experimental results of some of these representative algorithms are presented to evaluate and compare the efficacy of various techniques. Some concluding remarks are given in Section 10.6.

10.2 Human Visual System and Color Theory

It is instructive for us to begin the discussion on color with the physics of light and the physiology of the human visual system. The primary reason is that humans are typically the end-user and the judge of the images in our camera systems. A secondary reason is that the eye is itself a complex and beautifully made organ that acts as an *image capturing device* for our brain. In many cases, we model our camera system design on the natural design of our eyes.

Visible light occupies a small section of the electromagnetic spectrum, which we call the visible spectrum. In the seventeenth century, Sir Isaac Newton (1642–1727) was the first to demonstrate that when a white beam entered a prism, due to the law of refraction of light the exit beam would consist of shades of different colors. Further experimentation and

measurement show that different colors correspond to electromagnetic waves of different wavelengths, usually denoted with the symbol λ and measured in nanometers (nm). Visible spectrum roughly spans from $\lambda = 400\text{nm}$ to $\lambda = 700\text{nm}$.

Our eyes interpret different colors based on the wavelength of the electromagnetic wave. For example, at $\lambda \approx 400\text{nm}$ we have the sensation of blue; at $\lambda \approx 550\text{nm}$ we have the sensation of green; at $\lambda \approx 700\text{nm}$ we have the sensation of red. For a range of $\lambda < 400\text{nm}$, the region is called ultra-violet (UV), while for a range of $\lambda > 700\text{nm}$, we refer to it as infra-red (IR). UV and IR are gaining importance in imaging, particularly in medical imaging and remote sensing respectively, but our focus with digital camera systems is on the visible spectrum. Hence we will focus on the range approximately from $400\text{nm} < \lambda < 700\text{nm}$ hereafter.

10.2.1 Illumination

Imaging begins with the source of light called *illumination*. Virtually all illuminations consist of light with multiple wavelengths. (In fact, we have to go to extraordinary lengths to create lasers that are of a single wavelength or a very narrow bandwidth. Imaging under such circumstances is rare for digital photography and is mostly for scientific research purposes, and thus we do not take them into account here.) Each illumination then is described with a curve showing the strength of the electromagnetic radiations at different values of λ . If we normalize the curves of various illuminations, the result is a very useful description of the *spectral power distribution* of illumination as a function of wavelength. We can then compare the representative spectral power distributions of various common light sources. Note that we usually only describe the general characteristics of the illuminations; an actual measurement for sunlight, for example, would depend on the location, altitude, and atmospheric and weather conditions during the measurement.

As an example, Figure 10.1¹ shows the spectral power distribution of various common illumination sources. Figure 10.1a is the curve for typical sunlight, which is continuous (although not uniform) over the visible spectrum. Tungsten light, as shown in Figure 10.1b, also appears to be rather smooth. Later, we will see why sunlight tends to be perceived as bright yellowish-white, while tungsten usually gives us a sensation of yellowish hue. In contrast, the fluorescent lamp consists of sharp spikes in the spectral power distribution, as shown in Figure 10.1c. Finally, we see that with a light-emitting diode (LED), the spectral power distribution is also smooth, but is often limited to a narrow range of the visible spectrum as depicted in Figure 10.1d. LEDs with spectral power concentrating at the higher wavelength region are more common, and as a result we have mostly red LEDs.

In this chapter we denote the spectral power distribution of an illumination as $I(\lambda)$. This is an important quantity as we relate this with other color production factors to be explained next. In general, there are numerous possible curves $I(\lambda)$. It is also possible to express $I(\lambda)$

¹Data for some of the spectral power distributions and spectral reflectance can be obtained online from the Munsell Color Science Laboratory, Chester F. Carlson Center for Imaging Science at Rochester Institute of Technology.

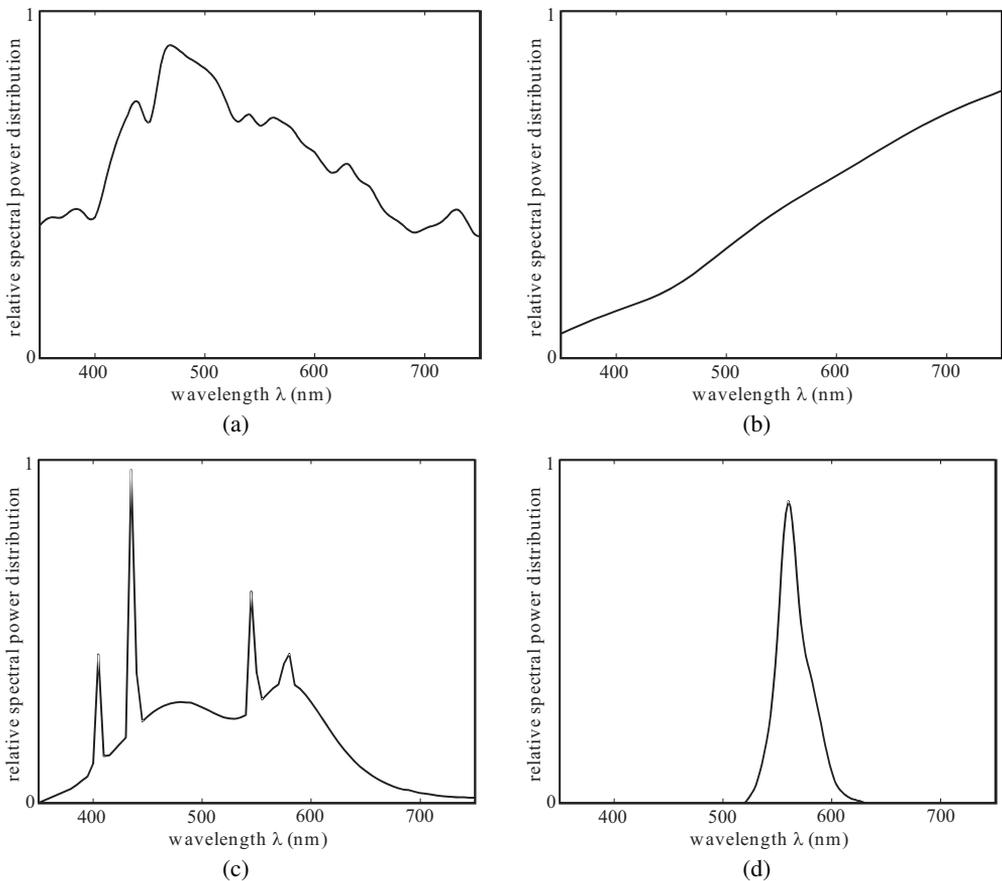


FIGURE 10.1

Spectral power distribution of various common types of illuminations: (a) sunlight, (b) tungsten light, (c) fluorescent light, and (d) light-emitting diode (LED).

as a linear combination of known basis functions $I_j(\lambda)$ with

$$I(\lambda) = \sum_{j=1}^m \alpha_j I_j(\lambda), \quad (10.1)$$

where, for example, three basis functions (corresponding to $m = 3$) are sufficient to represent standard daylight [5]. This property can be used in the design of AWB algorithms.

10.2.2 Object

Illumination is one of the three main factors contributing to the sensation of color in our brains. The second major factor is the object. When the electromagnetic radiations from the illumination reach an object, they are partially absorbed, and partially reflected or transmitted (for transparent objects). For different items, the proportion of the reflection or transmission varies with wavelengths, but this is an inherent property of the object irrespective of the illumination that takes place. We can therefore characterize an object's *spectral reflectance* or *spectral transmittance* as a function of wavelength for comparison.

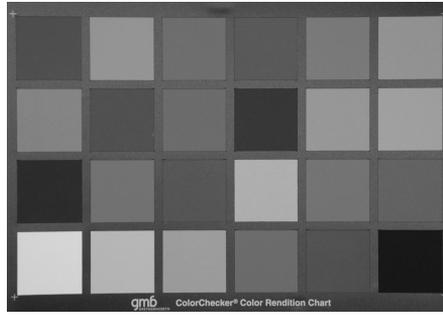


FIGURE 10.2 (See color insert.)

GretagMacbeth color rendition chart.

To illustrate, we plot the spectral reflectance corresponding to several typical object colors. These colors represent patches taken from the GretagMacbeth Color Checker (Figure 10.2) which is often used to test digital camera performances. The spectral reflectance plots are shown in Figure 10.3. For each plot, the y-axis denotes the fraction of light that is being reflected from the object. Figure 10.3a to Figure 10.3d show the spectral reflectance of a red, light blue, yellow, and gray patches of color, respectively. As expected, a red patch absorbs most of the greenish-blue frequencies and reflects most of the higher-wavelength light that gives the sensation of red. However, it should be noted that some residual lower-wavelength frequencies are also reflected, only that the amount is much smaller. We observe a similar behavior for the light blue and yellow patches as well. For the gray patch, the spectral reflectance is roughly a constant for the different wavelengths, causing the resulting gray sensation to be neutral in color.

We denote the spectral reflectance of an object with $R(\lambda)$. Similarly, we can also define the spectral transmittance of an object with $T(\lambda)$. There are also attempts to decompose the spectral reflectance into a summation of known basis functions $R_j(\lambda)$ such that

$$R(\lambda) = \sum_{j=1}^m \beta_j R_j(\lambda). \quad (10.2)$$

It has been shown that three basis functions (corresponding to $m = 3$) can accurately represent 433 Munsell-chips reflectance functions [6], and seven basis functions (corresponding to $m = 7$) are sufficient for a large number of natural objects [7].

10.2.3 Color Stimulus

Illumination and object reflectance or transmittance together determine the *color stimulus*. The spectral power distribution of the illumination governs how much energy is incident on the object at every wavelength. For a reflective object, the spectral reflectance dictates what fraction of that radiation is reflected and will arrive at the eye or the sensor, again at every wavelength. Similarly, for a transmissive object, the spectral transmittance determines the fraction of the radiation being transmitted through the object. Therefore, the spectral power distribution of an object is the *product* of the spectral power of the illumination and the spectral reflectance of the object. This is also called the color stimulus.

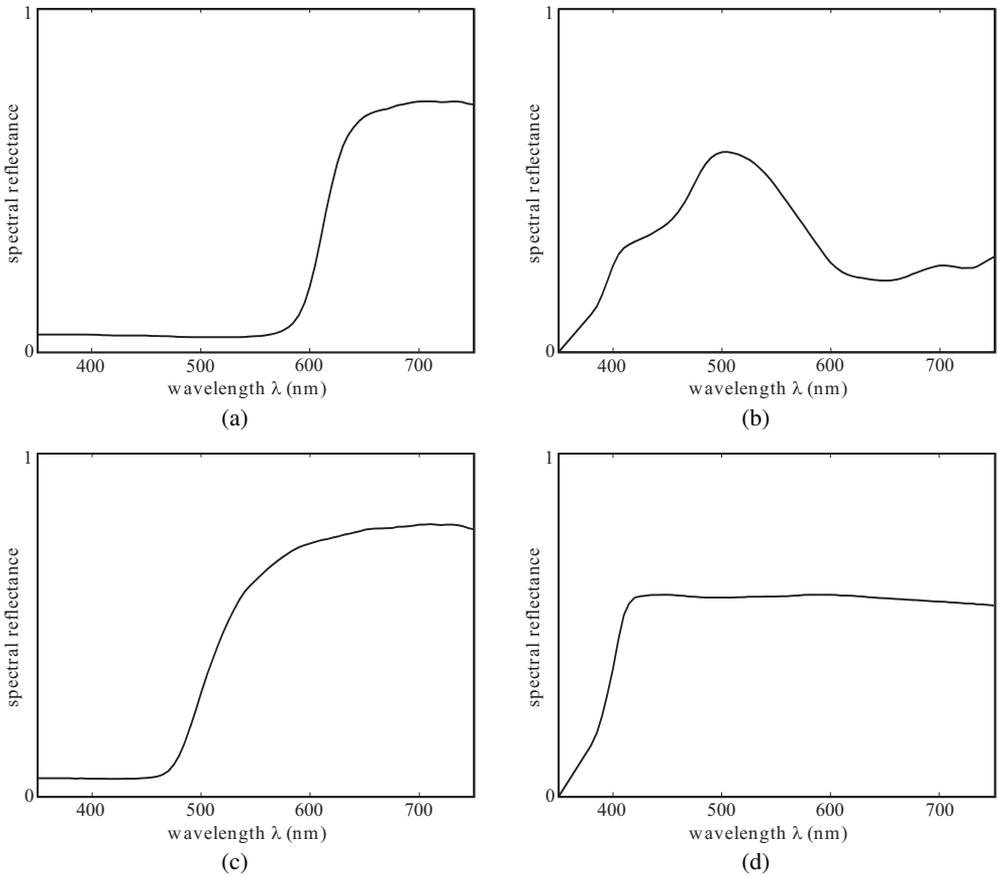


FIGURE 10.3

Spectral reflectance of various color patches: (a) red patch, (b) light blue patch, (c) yellow patch, (d) gray patch.

Mathematically, we denote the color stimulus with $S(\lambda)$. This is related to the illumination $I(\lambda)$ and object reflectance $R(\lambda)$ by [8]

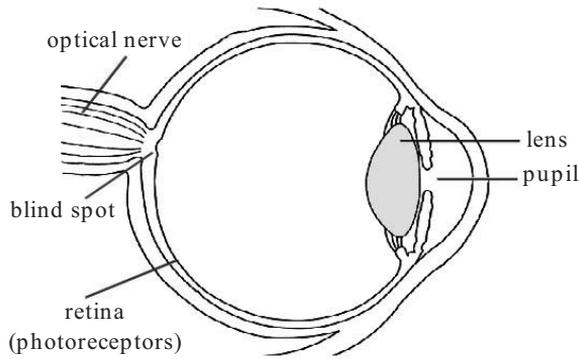
$$S(\lambda) = I(\lambda)R(\lambda). \tag{10.3}$$

If we use the basis decomposition in Equations 10.1 and 10.2, this becomes

$$S(\lambda) = \left(\sum_{j=1}^m \alpha_j I_j(\lambda) \right) \left(\sum_{k=1}^n \beta_k R_k(\lambda) \right) \tag{10.4}$$

$$= \sum_{j=1}^m \sum_{k=1}^n \alpha_j \beta_k I_j(\lambda) R_k(\lambda). \tag{10.5}$$

Although the above equations appear deceptively simple, they underscore an important fact in color science that we must emphasize here. The color stimulus depends on both the illumination and the object. Therefore, given any object, we can theoretically manipulate the illumination so that it produces any desired color stimulus! We will see shortly that the

**FIGURE 10.4**

An anatomy of a human eye.

color stimulus in turn contributes to our perception of color. One noteworthy corollary is that color is not an inherent feature of an object. A red object, for example, can be perceived as blue by a clever design of the illumination.

In other words, the illumination is as important as the object reflectance. This fact is very important for our camera system design. When we take a picture of the same object first under sunlight and then under fluorescent light, for example, the color stimuli vary significantly due to differences in illumination. We must adapt our camera to interpret the color stimuli differently, or otherwise the color of the photographs would look very different. This explains why AWB is so critical and challenging. However, before we can proceed on discussing AWB algorithms, we need to explore how our eyes interpret the color stimuli first. This brings us to the third major factor contributing to the color sensation: the physiology of our eyes.

10.2.4 Human Visual System

The color stimulus is a function of wavelength. If we had to faithfully reproduce the color stimulus, there is a lot of information to be stored. A key fact from color science is that there is no need to reproduce the entire spectral distribution in color reproduction. As a matter of fact, we often need only three values to specify a color, despite the obvious loss in information and possibility of ambiguity. The reason lies in our human visual system.

A basic anatomy of a human eye is shown in Figure 10.4. Visible light enters through the pupil and is refracted by the lens to create an image on the retina. The optical nerve transfers the image on the retina to the brain for interpretation. The retina is able to form an image because of tiny sensors called photoreceptors. For a normal person, there are two types of photoreceptors, cones and rods, which function very differently.

The rods are responsible for scotopic, or dim-light, vision. We have somewhere between 75 and 150 million rods in each of our eyes, and they are distributed all over the retina. Their chief aim is to give us an overall picture of the field of view of our eyes, rather than for color vision. When we enter a room that is rather dark, we may still be able to see objects even though they tend to be colorless. This is because under such illumination, only the rods are able to give us the images.

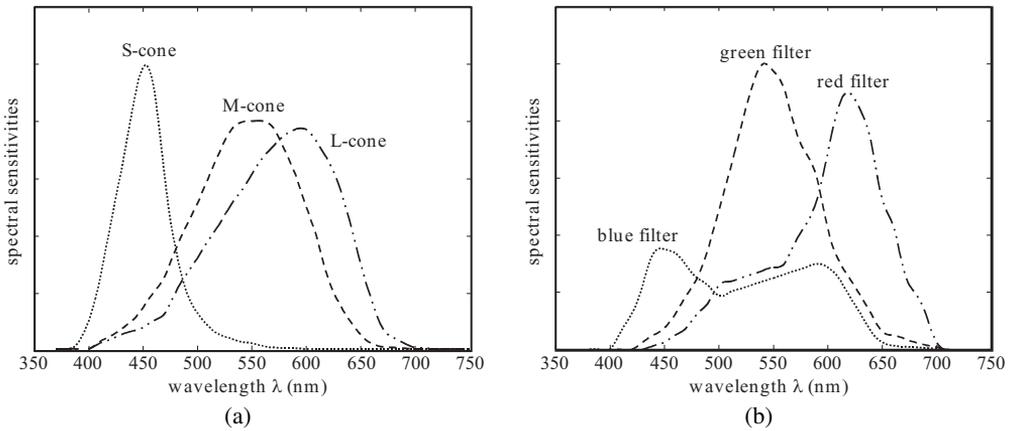


FIGURE 10.5

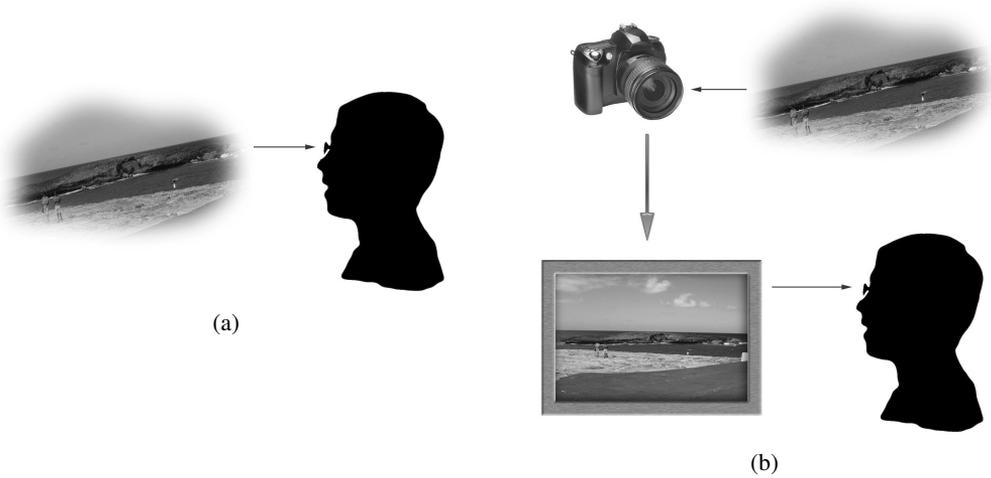
Spectral sensitivities of: (a) the three types of cones in a human eye, and (b) a typical digital camera.

The cones, on the other hand, are highly sensitive to color. There are far fewer cones in our eyes; an average person has about six to seven million only. They are also localized at a place called the fovea, rather than distributed all over the retina. They help us resolve fine details in images, and are responsible for photopic, or bright-light, vision. More importantly, there are three types of cones:

- *L-cones* which have peak sensitivity towards the long wavelength section of the visible spectrum,
- *M-cones* which have peak sensitivity towards the middle wavelength section of the visible spectrum, and
- *S-cones* which have peak sensitivity towards the short wavelength section of the visible spectrum.

These three types of cones together give us the sensation of color vision. When one or more of the cone types are defective, those people are said to possess what we collectively refer to as color deficiency or color blindness. Approximately one in twelve men has this condition to varying degrees, and this is more common in men than in women.

Figure 10.5a shows the spectral sensitivities of the three types of cones of the human eye. In subsequent discussions, we use $l(\lambda)$, $m(\lambda)$, and $s(\lambda)$ to denote the spectral sensitivity responses of the L-, M-, and S-cones respectively. For the sake of comparison, the curves have been normalized to equal area. It is interesting to observe that they do not cover disjoint sections of the visible spectrum, nor do they cover it entirely. In fact, the responses of L-cones and M-cones overlap significantly, and all three curves show low response to stimulus below around 400 nm and above around 650 nm. As we will see in the next section, camera designs mimic the responses of our human eyes. The sensor in the camera consists of three filters, typically red, green, and blue filters. The spectral sensitivities of a typical camera are shown in Figure 10.5b. We can observe that the peaks of these filters correspond to the peaks of the L-, M-, and S-cones of our eyes.

**FIGURE 10.6**

The aim of photography. Observing an object: (a) directly through a human eye, and (b) indirectly through a photograph.

When an object with stimulus $S(\lambda) = I(\lambda)R(\lambda)$ is observed, each of the three cones responds to the stimulus by summing up the reaction at all wavelengths. Therefore, three values are produced from the three cones, in accordance with the equations:

$$\begin{aligned}
 X &= \int_{400}^{700} l(\lambda)I(\lambda)R(\lambda) d\lambda \\
 Y &= \int_{400}^{700} m(\lambda)I(\lambda)R(\lambda) d\lambda. \\
 Z &= \int_{400}^{700} s(\lambda)I(\lambda)R(\lambda) d\lambda
 \end{aligned} \tag{10.6}$$

The triplet (X, Y, Z) is called trichromatic response. Despite its simplicity to describe color, it is estimated that humans are capable of resolving about 10 million color sensations!

An important consequence of trichromatic response is that in the digital camera, we only require three numbers at each pixel to capture the color information. We do not need to record the color stimulus at all wavelengths! In fact, this gives rise to a useful phenomenon: Even if we consider the spectral sensitivities to be known, for any given triplet (X, Y, Z) there could be an infinite number of possibilities for the color stimulus according to Equation 10.6. Two stimuli that produce the same trichromatic response are called metamers. The possibility of metamers is key to color photography.

10.2.5 Color Matching

The goal of photography is a bit different from the way our eyes perceive the color of an object. Consider the two scenarios depicted in Figure 10.6. In Figure 10.6a, our eyes observe a certain color object. In Figure 10.6b, our camera captures the object, and in turn

produces an image on screen or in a hardcopy. Our eyes then observe that object. Note that through the capturing device, because the spectral sensitivities of the camera differ from our eyes and the ink in the hardcopy differs in reflectance from the object, the print is not of identical color to the original object. The goal of color photography is to make the image appear as similar to the object as possible.

As such, our aim is *color matching*. Consider digital images shown on a screen, such as using cathode-ray tube (CRT), liquid crystal display (LCD), or even organic light-emitting device (OLED). In these cases, each pixel consists of three color patches called primaries, which are usually red, green, and blue. Color is formed from a linear combination of intensities from these primaries. Each primary is associated with a certain color stimulus, which we denote as $P_r(\lambda)$, $P_g(\lambda)$, and $P_b(\lambda)$ for the three colors.

It is not difficult to realize that we only need to perform color matching on monochromatic light sources. Any real stimulus, caused by any real illumination reflected or transmitted through any object, can be decomposed as a linear combination of these single-wavelength light sources. Mathematically, we assume that our light source has the color stimulus

$$S(\lambda) = \delta(\lambda - \lambda_0), \quad (10.7)$$

which indicates that it has unit strength at wavelength λ_0 and zero elsewhere. We assign scalar weights l_0 , m_0 , and s_0 , respectively, to the three primaries $P_r(\lambda)$, $P_g(\lambda)$, and $P_b(\lambda)$ in matching colors. Note further that the color matching must be performed with respect to an observer. It is common to define a standard observer, with a particular set of spectral sensitivities $l(\lambda)$, $m(\lambda)$, and $s(\lambda)$.

Equipped with all these parameters, we can now calculate the tristimulus value of directly observing the original stimulus with unit strength at wavelength λ_0 to be

$$\begin{aligned} X &= \int_{400}^{700} l(\lambda)S(\lambda) d\lambda = l(\lambda_0) \\ Y &= \int_{400}^{700} m(\lambda)S(\lambda) d\lambda = m(\lambda_0). \\ Z &= \int_{400}^{700} s(\lambda)S(\lambda) d\lambda = s(\lambda_0) \end{aligned} \quad (10.8)$$

The tristimulus value of indirect observation, through the three primaries of our display device, would be

$$\begin{aligned} \hat{X} &= \int_{400}^{700} l(\lambda) [l_0P_r(\lambda) + m_0P_g(\lambda) + s_0P_b(\lambda)] d\lambda \\ \hat{Y} &= \int_{400}^{700} m(\lambda) [l_0P_r(\lambda) + m_0P_g(\lambda) + s_0P_b(\lambda)] d\lambda. \\ \hat{Z} &= \int_{400}^{700} s(\lambda) [l_0P_r(\lambda) + m_0P_g(\lambda) + s_0P_b(\lambda)] d\lambda \end{aligned} \quad (10.9)$$

To match the color, we require only that the tristimulus values match, i.e., $X = \hat{X}$, $Y = \hat{Y}$, and $Z = \hat{Z}$. Equating Equations 10.8 and 10.9, we have

$$\underbrace{\begin{bmatrix} \int P_r(\lambda)l(\lambda) d\lambda & \int P_g(\lambda)l(\lambda) d\lambda & \int P_b(\lambda)l(\lambda) d\lambda \\ \int P_r(\lambda)m(\lambda) d\lambda & \int P_g(\lambda)m(\lambda) d\lambda & \int P_b(\lambda)m(\lambda) d\lambda \\ \int P_r(\lambda)s(\lambda) d\lambda & \int P_g(\lambda)s(\lambda) d\lambda & \int P_b(\lambda)s(\lambda) d\lambda \end{bmatrix}}_P \underbrace{\begin{bmatrix} l_0 \\ m_0 \\ s_0 \end{bmatrix}}_{\mathbf{v}_c} = \underbrace{\begin{bmatrix} l(\lambda_0) \\ m(\lambda_0) \\ s(\lambda_0) \end{bmatrix}}_{\mathbf{v}}. \quad (10.10)$$

This equation is fundamental to color science. Several remarks can be made for the matrix equation above:

1. The vector \mathbf{v}_c is the only unknown in the above equation. With three equations and three unknowns, the solution is unique provided that the matrix P is not singular.
2. The process can be repeated for different values of λ_0 . If we view l_0 above not as a scalar but as a value of the curve $x(\lambda)$ at $\lambda = \lambda_0$, by repeating the process at different wavelengths we can generate the entire curve of $x(\lambda)$. Similarly, we generate $y(\lambda)$ from m_0 and $z(\lambda)$ from s_0 . These curves are called color-matching functions.
3. The color-matching functions depend on the primaries ($R(\lambda)$, $G(\lambda)$, and $B(\lambda)$) and the observer ($l(\lambda)$, $m(\lambda)$, and $s(\lambda)$). Changing either, or both, of these quantities would result in new color-matching functions.
4. The vector \mathbf{v} is fixed for the same observer. In this case, the color-matching functions with different primaries are simply linear combinations of one another. Therefore, the two sets of color-matching can be described as $P_1\mathbf{v}_{c1} = \mathbf{v}$ and $P_2\mathbf{v}_{c2} = \mathbf{v}$, and thus

$$\mathbf{v}_{c1} = P_1^{-1}P_2\mathbf{v}_{c2}. \quad (10.11)$$

5. One particularly useful set of color-matching functions defined by the Commission Internationale de l'Éclairage (CIE), called the CIE Standard Colorimetric Observer color-matching functions, is shown in Figure 10.7. They are used in the calculation of the CIE tristimulus values X , Y , and Z , which quantify the trichromatic characteristics of color stimuli [8].

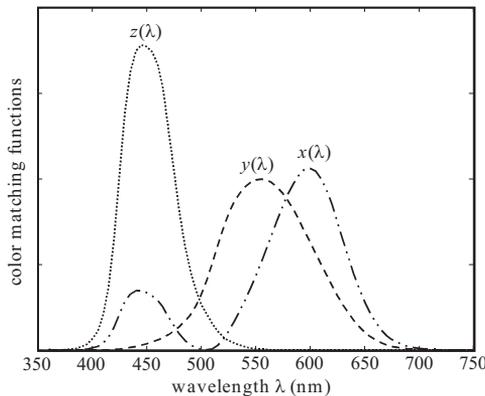


FIGURE 10.7

The CIE standard colorimetric observer color-matching functions.

10.3 Challenges in Automatic White Balancing

In the previous section, we have discussed how the color stimulus is equally dependent on the illumination and the object reflectance. Mathematically, we would thus expect the color stimulus of the same object to be different under different lighting conditions. However, our experience seems to the contrary: the same object appears to be of the same color even under different illuminations. This is known as *color constancy*. It is also known that the human visual system corrects for the prevailing scene illumination [9], [10]. However, for digital cameras, this is a challenging engineering problem.

Digital cameras nowadays use a single-image sensor, with a mosaic of color filters on top of each photodetector. For details refer to [Chapters 1](#) and [5](#). These filters can be fabricated as a photoresist layer mixed with the red, green, or blue dyes [11], with spectral sensitivities such as those shown in [Figure 10.5b](#).

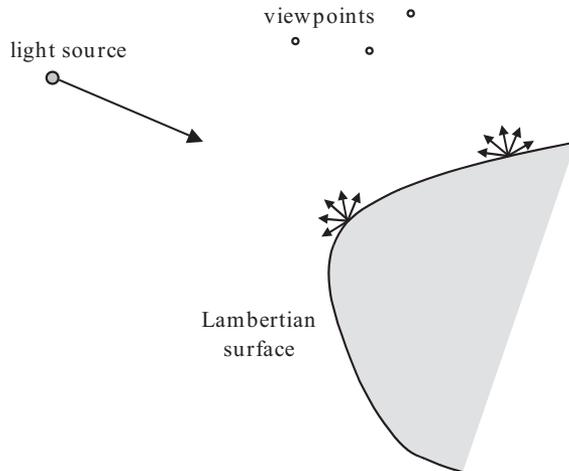
When an object with stimulus $S(\lambda) = I(\lambda)R(\lambda)$ is observed, each filter responds to the stimulus by summing up the reaction at all wavelengths. Therefore, three values are produced from the three filters, in accordance with the equation

$$\begin{aligned} R_{\text{sensor}} &= \int_{400}^{700} r(\lambda)I(\lambda)R(\lambda) d\lambda \\ G_{\text{sensor}} &= \int_{400}^{700} g(\lambda)I(\lambda)R(\lambda) d\lambda, \\ B_{\text{sensor}} &= \int_{400}^{700} b(\lambda)I(\lambda)R(\lambda) d\lambda \end{aligned} \quad (10.12)$$

where $r(\lambda)$, $g(\lambda)$, and $b(\lambda)$ refer to the spectral sensitivities of the sensors under the red, green, and blue filters respectively. This equation is essentially identical to Equation 10.6 except that we are now concerned with the spectral sensitivities of the camera sensors rather than our cone responses in our eyes. We can now state our AWB goal as follows: we seek to minimize the effect of $I(\lambda)$ and ensure that R_{sensor} , G_{sensor} , and B_{sensor} correlate with the object reflectance $R(\lambda)$ only [12].

The solution, however, involves dealing with an underdetermined set of equations. We can count the number of variables and equations as follows. For simplicity, assume for the moment that our sensors do not rely on demosaicking to recover the full RGB image. Hence, for an image of size $n \times n$, we have n^2 pixels and therefore $3n^2$ captured values. From these known values, we want to estimate parameters for the n^2 pixels together with the illuminant. Assume we discretize Equation 10.12 above so that

$$\begin{aligned} R_{\text{sensor}} &= \sum_{j=1}^m r(\lambda_j)I(\lambda_j)R(\lambda_j) \Delta\lambda \\ G_{\text{sensor}} &= \sum_{j=1}^m g(\lambda_j)I(\lambda_j)R(\lambda_j) \Delta\lambda. \\ B_{\text{sensor}} &= \sum_{j=1}^m b(\lambda_j)I(\lambda_j)R(\lambda_j) \Delta\lambda \end{aligned} \quad (10.13)$$

**FIGURE 10.8**

The Lambertian reflection model.

We are then using m sample points to represent the integral. Thus, for each pixel we want to derive $R(\lambda_j)$ for m values, hence there are a total of mn^2 unknowns. In addition, we have m unknowns for the illuminant. Thus, comparing $3n^2$ known values with $mn^2 + m = m(n^2 + 1)$ unknowns, it is clear that we do not have sufficient equations [12].

Moreover, we also need to note that Equations 10.12 and 10.13 above correspond to a simplified two-dimensional world in which all objects are flat, matte, Lambertian surfaces, and uniformly illuminated [12]. A Lambertian, or diffuse, surface assumes that light energy reaching a surface is reflected evenly in all directions [13], as shown in Figure 10.8. Thus, a planar patch appears to be of uniform brightness for all visible viewpoints. This occurs when the surface is rough enough relative to the wavelength of the light. Otherwise, considerations such as flare will substantially complicate the problem further.

To deal with the underconstrained nature of this problem, we often make additional assumptions about the world. Many of the AWB techniques to be mentioned in the following section rely on particular assumptions. For instance, the gray world method, as the name implies, considers that the average intensity of the scene is gray. The white patch method assumes there are always some white pixels in the image. Different assumptions thus lead to different implementations, and the efficacy of various AWB algorithms can be judged from how well the actual scenes satisfy the prior assumptions.

10.4 Automatic White Balancing Algorithms

For the above discussions, it can be seen that ideally AWB techniques require information about the camera being used, and possibly are based on assumptions about the statistical properties of the expected illumination and spectral reflectance [14]. In practice, many AWB algorithms follow a two-stage process:

1. *Illuminant estimation* — This may be done explicitly, often choosing from a known set of possible illuminants, or implicitly with assumptions about the effect of such illuminants.
2. *Image color correction* — This generates a new image as if it had been taken under a standard illuminant. The correction is often achieved through an independent gain regulation of the three color signals. This is known as the Von Kries hypothesis [15]. Commonly, it is achieved by adjusting the intensities of red and blue only, as AWB is concerned about the ratio of the three color signals.

Below we discuss several representative algorithms. We present them independently, but we should also note that combination techniques exist (e.g., Reference [14]) where multiple algorithms are run simultaneously, and a consensus decision is required afterwards to select the best results.

10.4.1 Gray World

The first method incorporates the gray world assumption, which argues that the average reflectance of a scene is achromatic. In other words, the mean of the red (R_{sensor}), green (G_{sensor}), and blue (B_{sensor}) channels in a given scene should be roughly equal. This method has its root in film photography, where for the negatives the average is biased towards dark regions of the scene, which tend to be neutral [4]. Algorithmically, as stated above we can adjust a gain factor to two of the channels so that both their means are now equal to the reference channel, which is often taken to be green.

We denote a full-color image of size $n \times n$ as $RGB_{\text{sensor}}(x, y)$, where x and y denote the indices of the pixel position. The individual red, green, and blue color components are then $R_{\text{sensor}}(x, y)$, $G_{\text{sensor}}(x, y)$, and $B_{\text{sensor}}(x, y)$, respectively. We compute

$$\begin{aligned} R_{\text{avg}} &= \frac{1}{n^2} \sum_{x=1}^n \sum_{y=1}^n R_{\text{sensor}}(x, y) \\ G_{\text{avg}} &= \frac{1}{n^2} \sum_{x=1}^n \sum_{y=1}^n G_{\text{sensor}}(x, y). \\ B_{\text{avg}} &= \frac{1}{n^2} \sum_{x=1}^n \sum_{y=1}^n B_{\text{sensor}}(x, y) \end{aligned} \quad (10.14)$$

If the three values are identical, the image already satisfies the gray world assumption and no further adjustment is necessary. In general, they may not be. We then compute the gain for the red and blue channels as $\hat{\alpha}$ and $\hat{\beta}$, where

$$\hat{\alpha} = \frac{G_{\text{avg}}}{R_{\text{avg}}} \quad \text{and} \quad \hat{\beta} = \frac{G_{\text{avg}}}{B_{\text{avg}}}. \quad (10.15)$$

The corrected image is formed with $\hat{R}_{\text{sensor}}(x, y)$, $\hat{G}_{\text{sensor}}(x, y)$, and $\hat{B}_{\text{sensor}}(x, y)$, where

$$\begin{aligned} \hat{R}_{\text{sensor}}(x, y) &= \hat{\alpha} R_{\text{sensor}}(x, y) \\ \hat{G}_{\text{sensor}}(x, y) &= G_{\text{sensor}}(x, y). \\ \hat{B}_{\text{sensor}}(x, y) &= \hat{\beta} B_{\text{sensor}}(x, y) \end{aligned} \quad (10.16)$$

Often, this image has sufficient intensity range for all the channels. In the event that the highest intensity of the three channels is significantly below the maximum allowable value, we can scale all three channels by the same amount so that the average intensity is still preserved. This gray world method is quite effective in practice, except in situations where a certain color may dominate, such as a blue hue for the sky, or when an object with a substantial amount of a certain color occupies the majority of the view.

There are a number of extensions to this method that can deal with such situations. One example is given in Reference [16]. In this method, one defines a region in the $R_{\text{avg}} - G_{\text{avg}}$ versus $B_{\text{avg}} - G_{\text{avg}}$ plane. If the computed $\{R_{\text{avg}}, G_{\text{avg}}, B_{\text{avg}}\}$ falls within the region, the scene is considered good enough and AWB adjustments using Equation 10.16 will not be performed.

10.4.2 White Patch

The second method is based on the Retinex theory² of visual color constancy, which argues that perceived white is associated with the maximum cone signals [18]. This is also known as the white world assumption [19]. This is because the brightest point in an image is often due to reflectance of a glossy surface, which tends to reflect the actual color of the light source [20]. The white balancing scheme then attempts to equalize the maximum value of the three channels to produce a white patch. To avoid disturbances to the calculation caused by a few bright pixels, one can treat clusters of pixels or lowpass the image [4]. To implement this, we compute

$$\begin{aligned} R_{\max} &= \max_{x,y} R_{\text{sensor}}(x,y) \\ G_{\max} &= \max_{x,y} G_{\text{sensor}}(x,y). \\ B_{\max} &= \max_{x,y} B_{\text{sensor}}(x,y) \end{aligned} \quad (10.17)$$

If G_{\max} is too small we can scale the green intensities up first, otherwise we keep the green channel unchanged. We define the gain for the red and blue channels as $\tilde{\alpha}$ and $\tilde{\beta}$, where

$$\tilde{\alpha} = \frac{G_{\max}}{R_{\max}} \quad \text{and} \quad \tilde{\beta} = \frac{G_{\max}}{B_{\max}}. \quad (10.18)$$

The corrected image is formed with $\tilde{R}_{\text{sensor}}(x,y)$, $\tilde{G}_{\text{sensor}}(x,y)$, and $\tilde{B}_{\text{sensor}}(x,y)$, where

$$\begin{aligned} \tilde{R}_{\text{sensor}}(x,y) &= \tilde{\alpha} R_{\text{sensor}}(x,y) \\ \tilde{G}_{\text{sensor}}(x,y) &= G_{\text{sensor}}(x,y). \\ \tilde{B}_{\text{sensor}}(x,y) &= \tilde{\beta} B_{\text{sensor}}(x,y) \end{aligned} \quad (10.19)$$

Gray world and white patch methods have their respective strengths. It is conceivable that satisfying the conditions in both methods would result in even better images. But we first need to make the following remarks:

²Retinex, which comes from the words *retina* and *cortex*, was coined to suggest that both the eye and the brain are involved in visual color constancy [17].

- For most images, the two methods produce different results. In other words, the corrected image can rarely satisfy both the gray world assumption and the Retinex theory.
- Equations 10.16 and 10.19 are both linear adjustments to the pixel intensities. Furthermore, there is also a fixed point in the mappings: for pixels with zero intensity, the two mappings would not affect their values. Evidently, it is rarely possible to achieve the requirements of both gray world assumption and Retinex theory with a linear technique.

Instead, a simple adjustment with a quadratic mapping of intensities was described in Reference [21]. Let the change to the red channel be

$$\check{R}_{\text{sensor}}(x, y) = \mu R_{\text{sensor}}^2(x, y) + \nu R_{\text{sensor}}(x, y), \quad (10.20)$$

where μ and ν are parameters to be found. The adjustment to the blue channel is computed analogously. To satisfy the gray world assumption, we require that

$$\sum_{x=1}^n \sum_{y=1}^n \check{R}_{\text{sensor}}(x, y) = n^2 G_{\text{avg}}, \quad (10.21)$$

and therefore,

$$\mu \sum_{x=1}^n \sum_{y=1}^n R_{\text{sensor}}^2(x, y) + \nu \sum_{x=1}^n \sum_{y=1}^n R_{\text{sensor}}(x, y) = n^2 G_{\text{avg}}. \quad (10.22)$$

Simultaneously, to satisfy the Retinex assumption to produce a white patch, we need

$$\max_{x, y} \check{R}_{\text{sensor}}(x, y) = G_{\text{max}}, \quad (10.23)$$

and therefore, if we assume that $R_{\text{sensor}}(x, y)$ takes on integer values between 0 and 255, and that μ and ν are positive numbers,

$$\mu \max_{x, y} R_{\text{sensor}}^2(x, y) + \nu \max_{x, y} R_{\text{sensor}}(x, y) = G_{\text{max}}. \quad (10.24)$$

Equations 10.22 and 10.24 together form two equations in two unknowns. We can represent them in a matrix form

$$\begin{bmatrix} \sum_{x=1}^n \sum_{y=1}^n R_{\text{sensor}}^2(x, y) & \sum_{x=1}^n \sum_{y=1}^n R_{\text{sensor}}(x, y) \\ \max_{x, y} R_{\text{sensor}}^2(x, y) & \max_{x, y} R_{\text{sensor}}(x, y) \end{bmatrix} \begin{bmatrix} \mu \\ \nu \end{bmatrix} = \begin{bmatrix} n^2 G_{\text{avg}} \\ G_{\text{max}} \end{bmatrix}. \quad (10.25)$$

This can be solved analytically for μ and ν using Cramer's rule.

10.4.3 Iterative White Balancing

The gray world method and white patch method described above are global techniques in that all pixels are involved in the computation. A drawback is that both may be susceptible to statistical anomalies. For the former, the method will give incorrect results if the scene

is heavily biased towards certain color cast, such as an outdoor scene of an ocean and the sky is typically rich in blue. For the white patch method, if a few pixels in the image have very large red, green, or blue values, they end up dominating the calculations.

In contrast, we have algorithms that pre-select a subset of pixels fulfilling certain *a priori* criteria, and the necessary color correction is derived from these pixels, although the adjustment is performed on all pixels subsequently. We can, for instance, perform an iterative white balancing technique as follows by extracting certain white points. We first convert the RGB values to YUV, a color space commonly used in video signals such as the PAL format, given by the following formula:

$$\begin{bmatrix} Y \\ U \\ V \end{bmatrix} = \begin{bmatrix} 0.299 & 0.587 & 0.114 \\ -0.147 & -0.289 & 0.436 \\ 0.615 & -0.515 & -0.100 \end{bmatrix} \begin{bmatrix} R_{\text{sensor}} \\ G_{\text{sensor}} \\ B_{\text{sensor}} \end{bmatrix}. \quad (10.26)$$

An ideal white point is when $R_{\text{sensor}} = G_{\text{sensor}} = B_{\text{sensor}} = 255$, which when put to the equation above makes $Y = 255$ and $U = V = 0$. Relaxing this condition a bit, we extract the pixels as white points if they satisfy the condition [22]

$$\begin{aligned} Y &> \xi \\ |U| &< \rho, \\ |V| &< \tau \end{aligned} \quad (10.27)$$

or an alternative criterion defined as [23]:

$$Y - |U| - |V| > \zeta, \quad (10.28)$$

where ξ , ρ , τ , and ζ are some pre-defined constants. While such a local method can avoid the scene being dominated by statistical anomalies, there are also situations that this would fail such as when there is no white object in the scene.

Another refinement is to look at gray points, which form a superset of the white points and therefore are more abundant in a typical scene. Reference [24] proposes selecting these points by the formula

$$\frac{|U| + |V|}{Y} < \eta, \quad (10.29)$$

where η is a positive threshold value much less than 1. The rationale is that if the light source is biased, say, to have a stronger red component, we can represent the captured red component \tilde{R} as

$$\tilde{R} = (1 + \kappa_R)R, \quad (10.30)$$

where R is the true red component if captured in a canonical light source, and κ_R denotes the percentage increase. This gives rise to a set of Y , U , and V where

$$\begin{bmatrix} Y \\ U \\ V \end{bmatrix} = \begin{bmatrix} 0.299 & 0.587 & 0.114 \\ -0.147 & -0.289 & 0.436 \\ 0.615 & -0.515 & -0.100 \end{bmatrix} \begin{bmatrix} (1 + \kappa_R)R \\ G \\ B \end{bmatrix}. \quad (10.31)$$

Note that for a canonical illumination, we have $R = G = B$ for a gray point, and therefore

$$\begin{bmatrix} Y \\ U \\ V \end{bmatrix} = \begin{bmatrix} 0.299 & 0.587 & 0.114 \\ -0.147 & -0.289 & 0.436 \\ 0.615 & -0.515 & -0.100 \end{bmatrix} \begin{bmatrix} (1 + \kappa_R)R \\ R \\ R \end{bmatrix} \quad (10.32)$$

$$= \begin{bmatrix} 1 + 0.299\kappa_R \\ -0.147\kappa_R \\ 0.615\kappa_R \end{bmatrix} R. \quad (10.33)$$

Putting the above in Equation 10.29, we get

$$\frac{|U| + |V|}{Y} = \frac{0.147\kappa_R + 0.615\kappa_R}{1 + 0.299\kappa_R} \quad (10.34)$$

$$= \frac{0.762\kappa_R}{1 + 0.299\kappa_R} < v. \quad (10.35)$$

This value is close to zero if κ_R is small. Similar results can be derived if the color cast is in green or blue.

After selecting these gray points, we compute their average U and V values as \hat{U}_{avg} and \hat{V}_{avg} . An iterative procedure is then employed to adjust them both to zero. At the j th iteration, we compute

$$\phi_j = \max(|\hat{U}_{\text{avg}}|, |\hat{V}_{\text{avg}}|). \quad (10.36)$$

If this equals to \hat{U}_{avg} , implying that the color is biased towards blue, we adjust the gain of the blue channel. Otherwise, the color is biased towards red, and the gain of the red channel is adjusted. The amount of adjustment used in Reference [24] is empirical and determined by trial and error. This changes \hat{U}_{avg} and \hat{V}_{avg} for the next iteration, and Equation 10.36 is computed again until satisfactory results are obtained.

10.4.4 Illuminant Voting

The three methods discussed above all make assumptions about the effects of illumination and adjust the pixel intensities directly. In principle, such methods attempt to adjust the intensity values of an image so that they appear “normal,” but there is no guarantee that the resulting image is indeed possible under any illuminant! In other words, we may have created an image that is not physically realizable with any lighting condition on the particular object. On the other hand, there are also various techniques that aim at recovering the illuminant explicitly from the observed images. One such example is the illuminant voting technique [25]. After identifying the illuminant, the correction to any alternative lighting condition will ensure that the resulting image is realizable.

This illuminant voting method is based on the idea of Hough transform. This is a well known technique in image processing, especially in pattern detection, and can be illustrated as follows. Suppose we would like to detect a straight line in an image. This line can be represented as

$$\rho = x \cos \theta + y \sin \theta \quad (10.37)$$

in the x - y plane, where ρ is the distance from the origin and θ is the angle of the line. The Hough transform of this line is then the point (ρ, θ) in a new parameter space. We

can think of the Hough transform as mapping a line to a point, but we can also think of it as mapping a point to a line, because a point in the original x - y plane can relate to a set of (ρ_j, θ_j) , where j is the index for the element, as long as they satisfy Equation 10.37 above. In theory there is an infinite number of (ρ_j, θ_j) , but in practice they are quantized and therefore there is only a finite number of elements. Thus, we can imagine each point casts one vote to each member of the element. When we have multiple points, the (ρ_j, θ_j) with the most number of votes denotes the strongest presence of a line. In implementation, we commonly pick θ_j first and then solve for ρ_j and eventually count the votes, before moving on to a new value of θ_j . Details of the Hough transform can be found in many image processing textbooks (e.g., Reference [26]).

In a similar manner, we rely on the observed data to vote for the most likely illuminant. This requires modelling of the illuminant and reflectance by using low order linear combinations, as described in Equations 10.1 and 10.2. Putting them to Equation 10.12, we have

$$\begin{bmatrix} R_{\text{sensor}} \\ G_{\text{sensor}} \\ B_{\text{sensor}} \end{bmatrix} = \begin{bmatrix} \int_{400}^{700} r(\lambda) \sum_{j=1}^m \sum_{k=1}^n \alpha_j \beta_k I_j(\lambda) R_k(\lambda) d\lambda \\ \int_{400}^{700} g(\lambda) \sum_{j=1}^m \sum_{k=1}^n \alpha_j \beta_k I_j(\lambda) R_k(\lambda) d\lambda \\ \int_{400}^{700} b(\lambda) \sum_{j=1}^m \sum_{k=1}^n \alpha_j \beta_k I_j(\lambda) R_k(\lambda) d\lambda \end{bmatrix} \quad (10.38)$$

$$= \left(\sum_{k=1}^n \beta_k M_k \right) \alpha, \quad (10.39)$$

where the j th column of M_k , denoted as $(M_k)_j$, equals

$$(M_k)_j = \begin{bmatrix} \int_{400}^{700} r(\lambda) I_j(\lambda) R_k(\lambda) d\lambda \\ \int_{400}^{700} g(\lambda) I_j(\lambda) R_k(\lambda) d\lambda \\ \int_{400}^{700} b(\lambda) I_j(\lambda) R_k(\lambda) d\lambda \end{bmatrix} \quad (10.40)$$

and $\alpha = [\alpha_1, \alpha_2, \dots, \alpha_m]^T$. Thus, it is clear that the equation above is linear in α . In fact, it is bilinear in α and β , where $\beta = [\beta_1, \beta_2, \dots, \beta_n]^T$, because the above equation can also be written by interchanging illumination and spectral reflectance [27].

Given this bilinearity, we can use the observed pixel data to vote for the set of illuminant and reflectance parameters in a way similar to the Hough transform. The procedure consists of the following steps:

1. *Selection of the reflectance parameters* — We pick a set of β_j , which determine the object reflectance given the basis functions $R_j(\lambda)$ in Equation 10.2. Since there could be many possible object spectral reflectances in the scene, we have to go through the three-step procedure many times, each with a different set of β_j .
2. *Determination of illumination parameters* — We solve for α_j using Equation 10.39 above. Note that this is an inverse problem. Provided that $(\sum_{k=1}^n \beta_k M_k)$ is not singular, a solution can be found. However, it should also be noted that if the matrix is ill-conditioned [28], [29], the solution can be very sensitive to noise. To deal with this problem, typically we retain only the cases where the system matrix is well-conditioned.

3. *Casting of vote* — A vote is cast for the α obtained. As with most implementations of Hough transform, this is quantized so that similar values are grouped together. Otherwise, there will be too many singleton votes.

After repeating the procedure for different object reflectances, the one with the most votes is deemed the illuminant.

10.4.5 Color by Correlation

The fundamental premise of the color by correlation method is that although there are numerous possible spectral power distributions such as those in [Figure 10.1](#), for instance, different hours of the day and different days would present different spectral power distributions of sunlight, there are only a small selection of substantially different illuminants (e.g., sunlight, fluorescent light, tungsten light, etc.). Some of these are modes or illumination conditions used in semi-automatic white balancing, where the user selects the particular mode and the camera performs white balancing accordingly. Similar to the previous method, the goal of AWB is achieved through illuminant identification, but the difference between the two is that the current method seeks not just a simple answer of the illumination function $I(\lambda)$, but a set of possible illuminants together with their likelihoods. Thus, not only does it determine the most likely illuminant, but it also computes the likelihood of all other illuminants so that the *error margin* of the subsequent choices is also known.

A prerequisite for this method is that we need to know the range and distribution of image colors that can be recorded by the camera under a set of possible lights [12]. We can then correlate the observed image with these distributions and identify the closest one as the most likely illuminant. More precisely, assume that there are k possible illuminants altogether. Instead of working with three sensory responses, we deal with only the chromaticity, where we can compute the chromaticities (c_1, c_2) as

$$c_1 = \frac{R_{\text{sensor}}}{G_{\text{sensor}}} \quad \text{and} \quad c_2 = \frac{B_{\text{sensor}}}{G_{\text{sensor}}}. \quad (10.41)$$

In practice, Reference [12] advocates the equation

$$c_1 = \left(\frac{R_{\text{sensor}}}{G_{\text{sensor}}} \right)^{\frac{1}{3}} \quad \text{and} \quad c_2 = \left(\frac{B_{\text{sensor}}}{G_{\text{sensor}}} \right)^{\frac{1}{3}} \quad (10.42)$$

which leads to chromaticities that are more uniformly distributed.

We partition the space of all chromaticities into $N \times N$ bins. The task is now to determine the possible (c_1, c_2) under each illuminant. There are a few possibilities:

1. The empirical way is to take the camera and capture a wide range of objects with various surface reflectances under each illuminant. We can then obtain the gamut of colors which the camera records under each lighting condition. This approach however can be rather cumbersome when there are many possible illuminants, and some may not be easily obtained at will (e.g., a bright sunlight illumination when it happens that the experimenters are experiencing rainy days!).

2. We can generate the chromaticities using Equations 10.12 and 10.42. This requires us to know the spectral response characteristics of the camera (such as in Figure 10.5b), the spectral power distribution of each illuminant, and the surface reflectance of a range of objects. In addition, we can take the convex hull of these chromaticities to form the gamut, setting all entries inside the gamut to be 1 and those outside to be 0.
3. We can further refine the scheme above by assigning the probability of each chromaticity value as the entries. This is computed empirically from the relative frequencies of occurrence estimated from the number of surfaces falling in each bin of the discretized chromaticity space.

We record the information above in an $N^2 \times k$ correlation matrix C , where each column (denoted as $(C)_j$ for the j th column) corresponds to a possible illuminant. Its row entry is the likelihood of observing that chromaticity under the particular illumination.

Now for a given image, we correlate the above information with that present in the image. We transform the pixel intensities to chromaticity values using the same formula as above, such as Equation 10.42. We then form a vector ω of length N^2 , where the j th element ω_j is one if the corresponding chromaticity value is present in the image, and zero otherwise. We can then compute the most appropriate illuminant \hat{j} by the formula

$$\hat{j} = \arg \max_j \langle \omega, (C)_j \rangle \quad (10.43)$$

where $\langle \cdot, \cdot \rangle$ denotes inner product. Another way to view this is that if we compute

$$\chi = \omega^T C \quad (10.44)$$

then the vector χ is a row vector of length k , where each value suggests the likelihood of the illuminant. Thus, in a single operation we can find not only the most likely illuminant but also the error margin of the others.

In summary, the color by correlation method entails the following three-step process:

1. *Preprocessing step* — Information about the interaction between image colors and illuminants is coded. This is considered the prior information about the illuminants.
2. *Correlation step* — This prior information is correlated with the information that is present in a particular image. In other words, the colors in an image determine the likelihood of each possible illuminant.
3. *Recovery step* — These likelihoods are used to recover an estimate of the scene illuminant.

10.4.6 Other Methods

The above discussion of AWB techniques is by no means exhaustive. Other promising techniques include the gamut mapping algorithm using coefficient rule (CRULE) [30], color in perspective [31], Bayesian formulation [10], neural networks [32], adaptive gains [33], [34], and combined strategies [14]. We refer readers to these original papers for further descriptions of their methods.

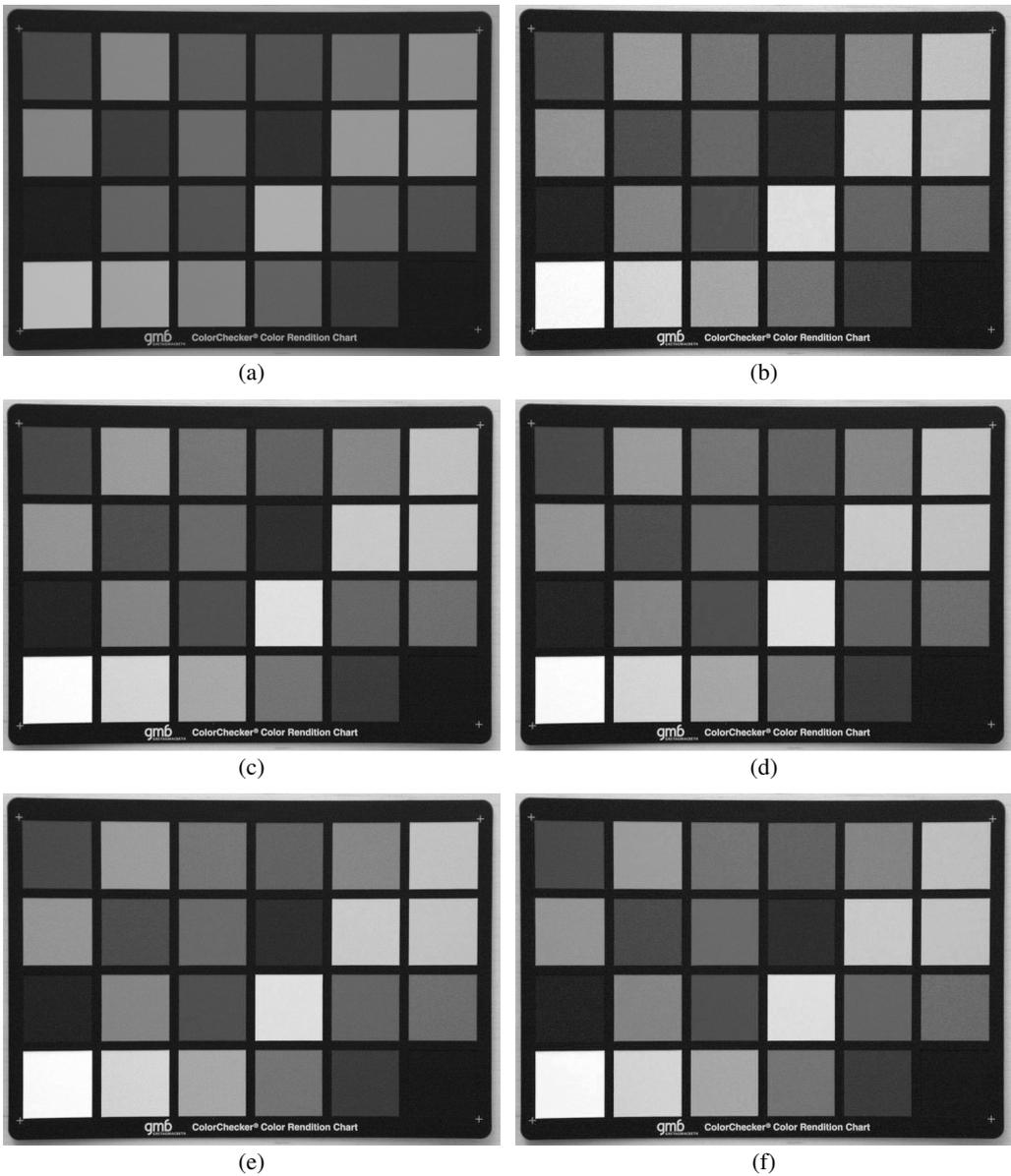


FIGURE 10.9 (See [color insert](#).)

AWB methods for the Macbeth color chart: (a) original image, (b) gray world, (c) white patch, (d) iterative white balancing, (e) illuminant voting, and (f) color by correlation.

10.5 Implementations and Quality Evaluations

In this section, we consider the performance of the above methods using a few test images. Our aim is not to extensively compare the various methods, which can often be found in the respective original papers and others written specifically for such a purpose [35], [36], but to give readers some general ideas of the performance of these techniques. It is also

known to be difficult to objectively evaluate image quality. With synthetic data, we can generate an *ideal* image with the desirable illumination, and a *test* image captured with another illumination but corrected with one of the AWB algorithms. If the former has intensities $\{R_{\text{ideal}}, G_{\text{ideal}}, B_{\text{ideal}}\}$ in the three channels and the latter has intensities $\{R_{\text{test}}, G_{\text{test}}, B_{\text{test}}\}$, we can compute the mean square error (MSE) between these images using the formula

$$\text{MSE} = \sum_{x=1}^n \sum_{y=1}^n \left[(R_{\text{ideal}} - R_{\text{test}})^2 + (G_{\text{ideal}} - G_{\text{test}})^2 + (B_{\text{ideal}} - B_{\text{test}})^2 \right], \quad (10.45)$$

where each of the quantities above has the argument (x, y) . A large MSE means that the ideal and test images are dissimilar, and suggests that the AWB algorithm may not be working well.

Unfortunately, MSE is commonly not a good metric for two reasons [37]. First, with real data we may not have the ideal image that we can compare with the test image. Even if we do, a second problem is that MSE does not correspond to the human perception of images. For instance, if we scale the intensities of the test image or shift it spatially by a small amount, the effects may be rather negligible perceptually, but mathematically the MSE can be significantly increased. There are other possibilities to compare the color differences objectively, such as using S-CIELAB [38] and CIEDE2000 [39]. Below, however, we mainly evaluate three sets of images subjectively to give readers a flavor of the AWB algorithms.

This first one is shown in Figure 10.9. In Figure 10.9a, the original image is seen to have a reddish-orange hue. The performance of the AWB algorithms is shown in Figure 10.9b to Figure 10.9f. In this particular case, the results of these algorithms are all quite satisfactory, and resemble each other. This can be attributed to the nature of this color chart object. For example, because many colors are present and there is no bias towards, for example, red, green, or blue, the assumption that the average intensity should be gray is quite agreeable. In a similar way, the object has a white patch that should correspond to the maximum intensity of the red, green, and blue channels. After the correction, the white patch is clearly evident.

Next, we consider the performance on another set of images shown in Figure 10.10. As in the previous example, Figure 10.10a is the original image. In this case, a bluish-green cast is visible in the image. Note that the actual image should be binary with black and white only. After the correction on the original image, the gray world technique shown in Figure 10.10b and the iterative white balancing scheme in Figure 10.10d both result in images that contain shades of gray only. However, the white background is rendered somewhat grayish in both cases. This can be attributed to the fact that both algorithms aim at reducing the chromatic components of the image, but do not have any mechanism that favors white to gray. On the other hand, the white patch technique in Figure 10.10c and the color by correlation method in Figure 10.10f are better in forcing the background to appear white. Note that in the white patch algorithm, we have filtered out the isolated bright pixels as mentioned in Section 10.4.2. Finally, the illuminant voting algorithm in Figure 10.10e seems to have over-compensated for the bluish cast and the resulting corrected image now contains mild shades of yellow.

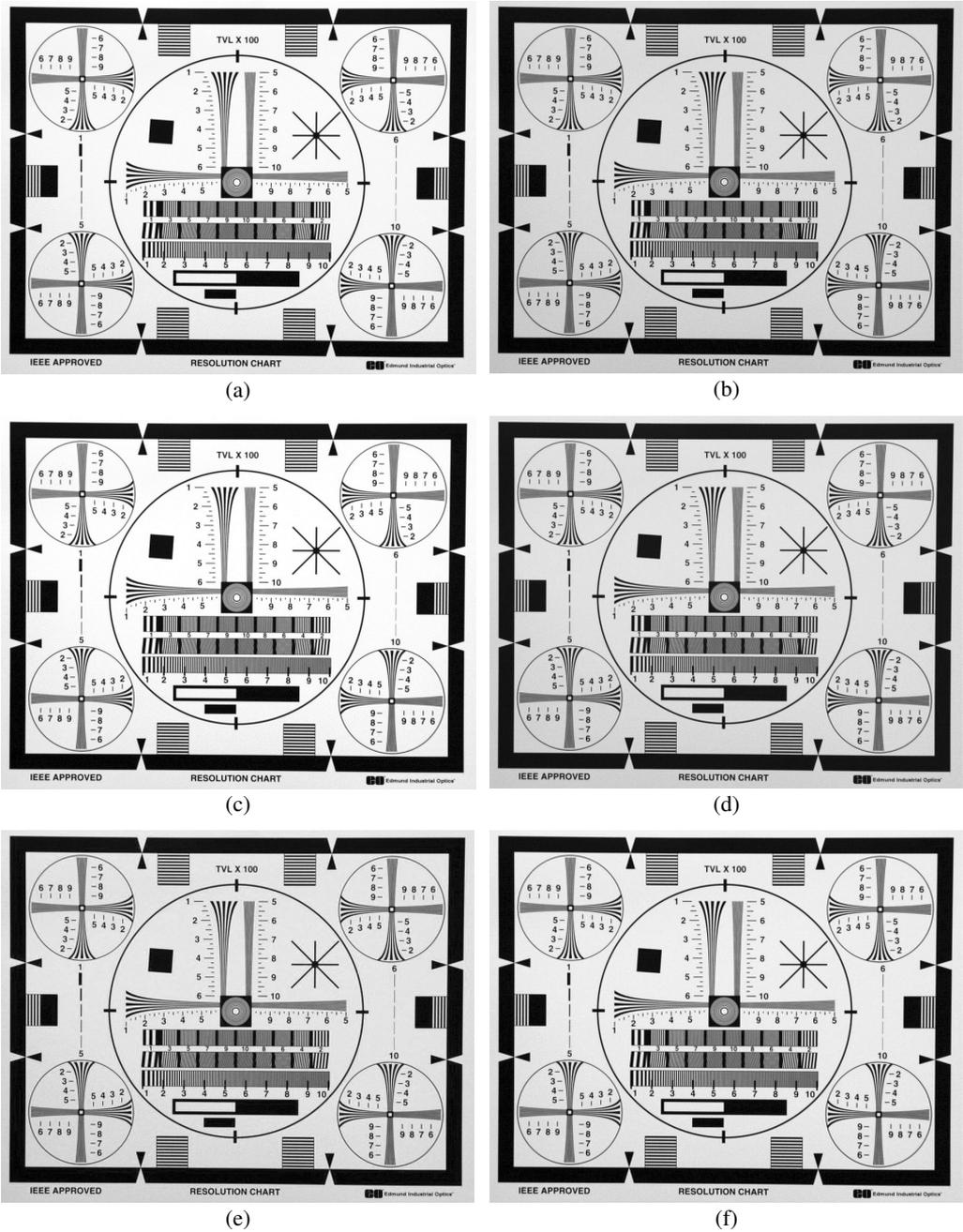


FIGURE 10.10 (See color insert.)

AWB methods for the resolution chart: (a) original image, (b) gray world, (c) white patch, (d) iterative white balancing, (e) illuminant voting, and (f) color by correlation.

The third set of experimental results is given in Figure 10.11. This is an image of a bookshelf taken under fluorescent light but with incorrect white balance setting in the camera. We can observe that the white patch method in Figure 10.11c seems to perform the best in this case, whereas the gray world method in Figure 10.11b and the iterative white



FIGURE 10.11 (See color insert.)

AWB methods for the *bookshelf* image: (a) original image, (b) gray world, (c) white patch, (d) iterative white balancing, (e) illuminant voting, and (f) color by correlation.

balancing scheme in Figure 10.11d again appear to produce an image that is slightly grayish. Both the illuminant voting and color by correlation methods, in Figure 10.11e and Figure 10.11f respectively, perform white balancing insufficiently. This result seems to agree with the comment made in Reference [36] that algorithms taking advantage of the chromaticity statistics seem to perform worse than expected.

10.6 Conclusion

In this chapter we considered the issue of automatic white balancing (AWB) in digital photography. We discussed the nature of the problem, and various algorithms that could be implemented to achieve AWB, including gray world, white patch, iterative white balancing, illuminant voting, and color by correlation. Nevertheless, we should note that color constancy — the root problem of AWB — is still recognized as a difficult problem that has not been solved satisfactorily, or even understood well enough how humans and some other animals possess this fascinating quality. Even the state-of-the-art algorithms are not nearly as good as our own color constancy [10] or in some cases, sufficient for machine vision tasks such as object recognition [40]. Evidently, there is much room for research both in understanding color constancy in human visual systems and the possibility of adapting it or using some other methods to achieve AWB in digital camera systems design.

Acknowledgments

This work is supported in part by Grant 10204548 from the University Research Committee of the University of Hong Kong and by the Research Grants Council of the Hong Kong Special Administrative Region, China under Project HKU 7143/05E. Dong Liang of the University of Hong Kong put much effort in the implementation of the various AWB algorithms. Experience gained when the principal author was involved with the Stanford programmable digital camera project also helped shape the content of this chapter.

References

- [1] E. Land, “Recent advances in retinex theory and some implications for cortical computations: Color vision and the natural images,” *Proceedings of the National Academy of Science*, vol. 80, no. 16, pp. 5163–5169, August 1983.
- [2] E.Y. Lam, “Image restoration in digital photography,” *IEEE Transactions on Consumer Electronics*, vol. 49, no. 2, pp. 269–274, May 2003.
- [3] R. Lukac, “Single-sensor imaging in consumer digital cameras: A survey of recent advances and future directions,” *Journal of Real-Time Image Processing*, vol. 1, no. 1, pp. 45–52, October 2003.
- [4] R. Hunt, *The Reproduction of Colour*. Chichester, West Sussex, UK: John Wiley & Sons, 2004.
- [5] D.B. Judd, D.L. MacAdam, and G. Wyszecki, “Spectral distribution of typical daylight as a function of correlated color temperature,” *Journal of the Optical Society of America*, vol. 54, no. 8, pp. 1031–1040, 1964.

- [6] J. Cohen, "Dependency of the spectral reflectance curves of the Munsell color chips," *Psychoneurological Science*, vol. 1, no. 12, pp. 369–370, 1964.
- [7] M.J. Vrhel, R. Gershon, and L.S. Iwan, "Measurement and analysis of object reflectance spectra," *Color Research and Application*, vol. 19, no. 1, pp. 4–9, February 1994.
- [8] E. Giorgianni and T. Madden, *Digital Color Management*. Reading, MA: Addison Wesley, 1998.
- [9] L. Arend and A. Reeves, "Simultaneous color constancy," *Journal of the Optical Society of America A*, vol. 3, no. 10, pp. 1743–1751, October 1986.
- [10] D.H. Brainard, W.A. Brunt, and J.M. Speigle, "Color constancy in the nearly natural image. I. Asymmetric matches," *Journal of the Optical Society of America A*, vol. 14, no. 9, pp. 2091–2110, September 1997.
- [11] D. Qian, J. Toker, and S. Bencuya, "An automatic light spectrum compensation method for CCD white balance measurement," *IEEE Transactions on Consumer Electronics*, vol. 43, no. 2, pp. 216–220, May 1997.
- [12] G.D. Finlayson, S.D. Hordley, and P.M. Hubel, "Color by correlation: A simple, unifying framework for color constancy," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 23, no. 11, pp. 1209–1221, November 2001.
- [13] L.G. Shapiro and G.C. Stockman, *Computer Vision*. Upper Saddle River, NJ: Prentice Hall, 2001.
- [14] S. Bianco, F. Gasparini, and R. Schettini, "Combining strategies for white balance," in *Proceedings of the SPIE, Digital Photography III*, San Jose, CA, USA, January 2007, vol. 6502, id. 65020D.
- [15] M. Fairchild, *Color Appearance Models*. Chichester, West Sussex, UK: John Wiley & Sons, 2005.
- [16] Y. Kim, H.S. Lee, and A.W. Morales, "A video camera system with enhanced zoom tracking and auto white balance," *IEEE Transactions on Consumer Electronics*, vol. 48, no. 3, pp. 428–434, August 2002.
- [17] E. Land, "The Retinex," *American Scientist*, vol. 52, no. 2, pp. 247–264, 1964.
- [18] E. Land and J. McCann, "Lightness and Retinex theory," *Journal of the Optical Society of America*, vol. 61, no. 1, pp. 1–11, 1971.
- [19] N. Kehtarnavaz, H. Oh, and Y. Yoo, "Development and real-time implementation of auto white balancing scoring algorithm," *Real-Time Imaging*, vol. 8, no. 5, pp. 379–386, October 2002.
- [20] C. Weng, H. Chen, and C. Fuh, "A novel automatic white balance method for digital still cameras," in *Proceedings of the IEEE International Symposium on Circuits and Systems*, Kobe, Japan, May 2005, vol. 4, pp. 3801–3804.
- [21] E.Y. Lam, "Combining gray world and Retinex theory for automatic white balance in digital photography," in *Proceedings of the International Symposium on Consumer Electronics*, Macau, China, June 2005, pp. 134–139.
- [22] N. Nakano, R. Nishimura, H. Sai, A. Nishizawa, and H. Komtsu, "Digital still camera system for megapixel CCD," *IEEE Transactions on Consumer Electronics*, vol. 44, no. 3, pp. 460–466, August 1998.
- [23] R.Z. Zhou, J. He, and Z.L. Hong, "Adaptive algorithm of auto white balance for digital camera," *Journal of Computer-Aided Design and Computer Graphics*, vol. 17, no. 3, pp. 529–533, March 2005.
- [24] J. Huo, Y. Chang, J. Wang, and X. Wei, "Robust automatic white balance algorithm using gray color points in images," *IEEE Transactions on Consumer Electronics*, vol. 52, no. 2, pp. 541–546, May 2006.

- [25] G. Sapiro, "Color and illuminant voting," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 21, no. 11, pp. 1210–1215, November 1999.
- [26] R. Gonzalez and R. Woods, *Digital Image Processing*. Upper Saddle River, New Jersey: Prentice Hall, 2002.
- [27] G. Sapiro, "Bilinear voting," in *Proceedings of the Sixth International Conference on Computer Vision*, Bombay, India, November 1998, pp. 178–183.
- [28] E.Y. Lam and J.W. Goodman, "Iterative statistical approach to blind image deconvolution," *Journal of the Optical Society of America A*, vol. 17, no. 7, pp. 1177–1184, July 2000.
- [29] G. Golub and C. Van Loan, *Matrix Computations*. Baltimore, Maryland: Johns Hopkins University Press, 1996.
- [30] D.A. Forsyth, "A novel algorithm for color constancy," *International Journal of Computer Vision*, vol. 5, no. 1, pp. 5–36, August 1990.
- [31] G.D. Finlayson, "Color in perspective," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 18, no. 10, pp. 1034–1038, October 1996.
- [32] B.V. Funt, V. Cardei, and K. Barnard, "Learning color constancy," in *Proceedings of the Fourth Color Imaging Conference*, Scottsdale, AZ, USA, November 1996, pp. 58–60.
- [33] R. Lukac, "Refined automatic white balancing," *Electronic Letters*, vol. 43, no. 8, pp. 445–446, April 2007.
- [34] R. Lukac, "New framework for automatic white balancing of digital camera images," *Signal Processing*, vol. 88, no. 3, pp. 582–593, March 2008.
- [35] K. Barnard, V. Cardei, and B. Funt, "A comparison of color constancy algorithms - Part I: Methodology and experiments with synthesized data," *IEEE Transactions on Image Processing*, vol. 11, no. 9, pp. 972–983, September 2002.
- [36] K. Barnard, B. Funt, L. Martin, and A. Coath, "A comparison of color constancy algorithms - Part II: Experiments with image data," *IEEE Transactions on Image Processing*, vol. 11, no. 9, pp. 985–996, September 2002.
- [37] E.Y. Lam, "Robust minimization of lighting variation for real-time defect detection," *Real-Time Imaging*, vol. 10, no. 6, pp. 365–370, December 2004.
- [38] X. Zhang and B. Wandell, "Color image fidelity metrics evaluated using image distortion maps," *Signal Processing*, vol. 70, no. 3, pp. 201–214, November 1998.
- [39] G. Johnson and M. Fairchild, "A top down description of S-CIELAB and CIEDE2000," *Color Research and Application*, vol. 28, no. 6, pp. 425–435, December 2003.
- [40] B. Funt, K. Barnard, and L. Martin, "Is machine colour constancy good enough?," in *Proceedings of the Fifth European Conference on Computer Vision*, Freiburg, Germany, June 1998, pp. 455–459.