

# Module 1, Lectures 1 and 2

Marcelo Bertalmío

## CONTENTS

<b>I</b>	<b>Light as color stimulus</b>	3
<b>II</b>	<b>Matching colors</b>	5
<b>III</b>	<b>The first standard color spaces</b>	8
III-A	Chromaticity diagrams . . . . .	8
<b>IV</b>	<b>Perceptual color spaces</b>	11
IV-A	Color constancy and the von Kries coefficient law . . . . .	11
IV-B	Perceptually uniform color spaces . . . . .	12
IV-C	Limitations of CIELUV and CIELAB . . . . .	13
<b>V</b>	<b>Color appearance</b>	15
<b>VI</b>	<b>Image processing pipeline</b>	17
<b>VII</b>	<b>Image sensors</b>	17
VII-A	Pixel classes . . . . .	17
VII-B	Sensor classes . . . . .	19
VII-C	Interlaced vs. progressive scanning . . . . .	19
VII-D	CCD types . . . . .	20
VII-E	CMOS types . . . . .	20
VII-F	Noise in image sensors . . . . .	20
VII-G	Capturing colors . . . . .	21
<b>VIII</b>	<b>Exposure control</b>	22
VIII-A	Exposure metering . . . . .	23
VIII-B	Control mechanisms . . . . .	23
VIII-C	Extension of dynamic range . . . . .	23
<b>IX</b>	<b>Focus control</b>	23
<b>X</b>	<b>White balance</b>	24
<b>XI</b>	<b>Color transformation</b>	27
XI-A	The colorimetric matrix . . . . .	27
XI-B	A note on color stabilization . . . . .	28
XI-C	Encoding the color values . . . . .	28
<b>XII</b>	<b>Gamma correction and quantization</b>	32
XII-A	The need for gamma correction . . . . .	32
XII-B	Transfer function and quantization . . . . .	34
XII-C	Color correction pipeline . . . . .	34
<b>XIII</b>	<b>Edge enhancement</b>	35
<b>XIV</b>	<b>Output formats</b>	36
XIV-A	Compression . . . . .	36
XIV-B	Recording in RAW . . . . .	37
<b>XV</b>	<b>Additional image processing</b>	37
XV-A	Lens spatial distortion correction . . . . .	37
XV-B	Lens shading correction . . . . .	37

XVI The order of the stages of the image processing pipeline	37
References	39

## I. LIGHT AS COLOR STIMULUS

We live immersed in electromagnetic fields, surrounded by radiation of natural origin or produced by artifacts made by humans. This radiation has a dual behavior of wave and particle, where the particles can be considered as packets of electromagnetic waves. Waves are characterized by their wavelength, the distance between two consecutive peaks. Of all the radiation that continuously reaches our eyes, we are only able to *see* (i.e. our retina photoreceptors are only sensitive to) electromagnetic radiation with wavelengths within the range of 380nm to 740nm (a nm or nanometer is one-billionth of a meter). We are not able to see radiation outside this band, such us ultraviolet radiation (wavelength of 10nm to 400nm) or FM radio (wavelengths near 1m). Therefore, light is defined as radiation with wavelengths within the *visible spectrum* of 380nm to 740nm. Figure 1 shows the full spectrum of radiation with a detail of the visible light spectrum. The sun emits full-spectrum radiation, including gamma rays and ultraviolet and infrared “light,” which of course have an effect on our bodies even if we are not able to see them.

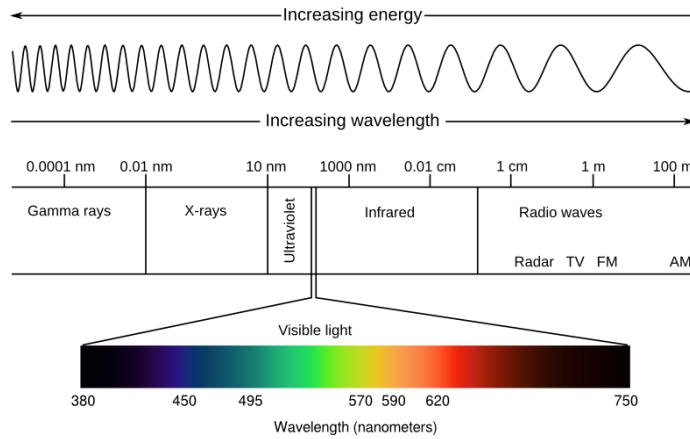


Fig. 1. Electromagnetic spectrum and visible light.

The frequency of a wave is its number of cycles per second; all electromagnetic waves travel at the same speed, the speed of light ( $c$  in the vacuum), therefore the wavelength of a wave is inversely proportional to its frequency. Shorter wavelengths imply higher frequencies and also higher energies. When we look at a *single isolated light*, if it has short wavelength we perceive it as blue, if it has middle-length wavelength we see it as green, and if it has long wavelength it appears to us as red. But we must stress that light in itself is not colored (there are no different kinds of photons), color is a perceptual quantity: for instance, the same light that appears red when isolated may appear yellow when it is surrounded by other lights. So the light stimulus at a given location in the retina is not enough to determine the color appearance it will produce; nonetheless, it must be characterized since it constitutes the input to our visual system and therefore what color appearance will depend on. Among other ways, light stimuli can be described by radiometry, which measures light in energy units and does not consider the properties of our visual system, and by colorimetry, which reduces the multi-valued radiometric spectrum of a light stimulus to three values describing the effect of the stimulus in the three types of cone receptors in the retina [64].

With a radiometric approach, the properties of a light emitting source are described by its power spectrum function  $I(\lambda)$ , the *irradiance*, which states for each wavelength  $\lambda$  the amount of power  $I$  the light has. The light absorption properties of a surface are described by its *reflectance*  $R(\lambda)$ , which for each wavelength  $\lambda$  states the percentage of photons that are reflected by the surface. When we see a surface, the light that is reflected by it and reaches our eyes is called *radiance* and its power spectrum  $E(\lambda)$  is the product of the spectrum functions for the incident light and the reflectance function of the surface:

$$E(\lambda) = I(\lambda) \times R(\lambda). \quad (1)$$

Figure 2 shows the irradiance functions of several light emitting sources, and Figure 3 shows the reflectances of some patches of different colors. From these figures and Equation 1 we can see that when we illuminate a red patch (Figure 3(a)) with sunlight (Figure 2(a)) we get from the patch a radiance  $E$  with its power concentrated in the longest wavelengths, which as we mentioned corresponds to our sensation of red.

The human retina has photoreceptor neurons, with colored pigments. These pigments have their particular photon absorption properties as a function of wavelength, and absorbed photons generate chemical reactions that produce electrical impulses, which are then processed at the retina and later at the visual cortex in the brain. The sensitivity of the pigments depends on the *luminance* of the light, which is a measure of the light's power, formally defined as intensity per unit area in a given direction. We have two types of photoreceptors:

- Rods, for low and mid-low luminances (at high luminances they are active but saturated). There are some 120 million of them.

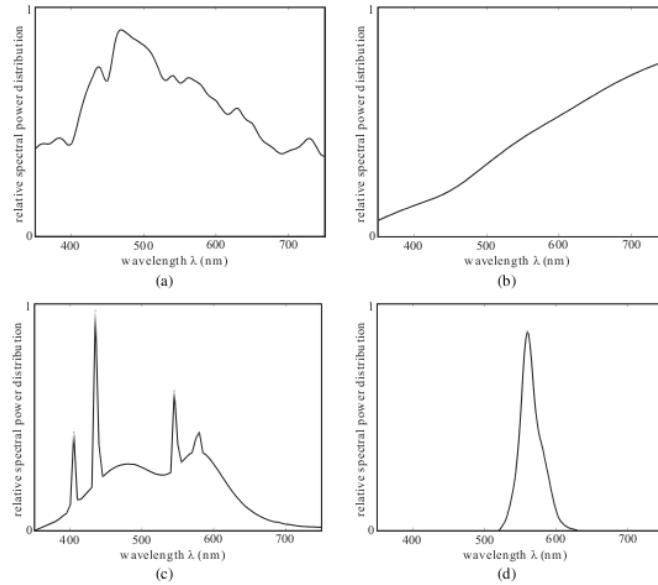


Fig. 2. Spectral power distribution of various common types of illuminations: (a) sunlight, (b) tungsten light, (c) fluorescent light, and (d) LED. Figure from [53].

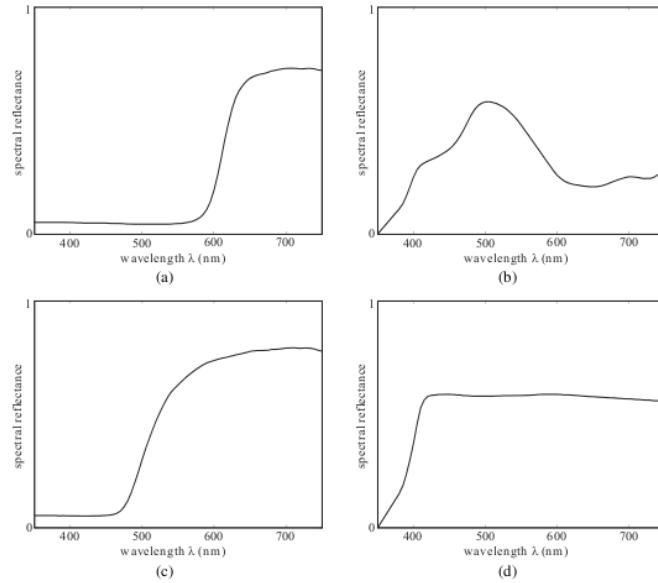


Fig. 3. Spectral reflectance of various colored patches: (a) red patch, (b) blue patch, (c) yellow patch, and (d) gray patch. Figure from [53].

- Cones, which have pigments that are 500 times *less* sensitive to light than the rods' pigment, rhodopsin. Therefore, cones work only with high luminances; at low luminances they are not active. There are some 6 million of them, most of them very densely concentrated at the fovea, the center of the retina.

There are three types of cones: S-cones, M-cones and L-cones, where the capital letters stand for “short,” “medium” and “long” wavelengths, respectively. Hubel [40] points out that three is the minimum number of types of cones that allow us not to confuse any monochromatic light with white light. People who lack one type of cone do perceive certain colors as gray. Frisby and Stone [38] mention that while there are several animal species with more than three types of color receptors, which can then tell apart different shades of color that we humans perceive as equal, this probably comes at the price of less visual acuity, for there are more cones to be accommodated in the same retinal area. Low luminance or *scotopic* vision, mediated only by rods, is therefore color-less. In a low-medium range of luminances, the so-called *mesopic* vision, both rods and cones are active, and this is what happens in a typical movie theatre [20]. In high-luminance or *photopic* vision cones are active and the rods are saturated. Each sort of cone photoreceptor has a spectral absorbance function describing its sensitivity to light as a function of wavelength:  $s(\lambda)$ ,  $m(\lambda)$ ,  $l(\lambda)$ . These curves were first determined experimentally by König in the late 19th

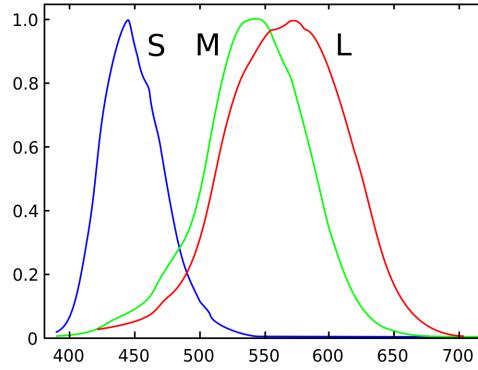


Fig. 4. Cone sensitivities (normalized). Figure from [9].

century. The sensitivity curves are quite broad, almost extending over the whole visible spectrum, but they are bell-shaped and they peak at distinct wavelengths: S-cones at 420nm, M-cones at 533nm and L-cones at 584nm; see Figure 4. These three *wavelength* values correspond to monochromatic blue, green and red light, respectively.

With the colorimetric approach, the sensation produced in the eye by radiance  $E(\lambda)$  (the stimulus of a light of power spectrum  $E(\lambda)$ ) is determined by a triplet of values, called the tristimulus values, given by the integral over the visible spectrum of the product of the radiance by each of the three cone sensitivity functions:

$$\begin{aligned} L &= \int_{380}^{740} l(\lambda)E(\lambda)d\lambda \\ M &= \int_{380}^{740} m(\lambda)E(\lambda)d\lambda \\ S &= \int_{380}^{740} s(\lambda)E(\lambda)d\lambda. \end{aligned} \quad (2)$$

## II. MATCHING COLORS

If two lights with different spectra  $E_1(\lambda)$  and  $E_2(\lambda)$  produce the same tristimulus vector  $(L, M, S)$ , then both lights are producing the same cone responses and (if viewed in isolation) we will see them as having the same color. Lights with different spectra that appear to have the same color are called metamer (lights with the same spectra always appear to have the same color and are called isomers [46]).

Expressing Equation 2 in discrete terms, it will be easy to see that each light has many metamer:

$$\begin{aligned} L &= \sum_{i=380}^{740} l(\lambda_i)E(\lambda_i) \\ M &= \sum_{i=380}^{740} m(\lambda_i)E(\lambda_i) \\ S &= \sum_{i=380}^{740} s(\lambda_i)E(\lambda_i), \end{aligned} \quad (3)$$

which, arranged into matrix form, becomes:

$$\begin{bmatrix} L \\ M \\ S \end{bmatrix} = \mathcal{S}\mathbf{E}, \quad (4)$$

where  $\mathcal{S}$  is a  $3 \times 361$  matrix whose rows are the (discrete) cone sensitivities, and  $\mathbf{E}$  is the radiance spectrum expressed as a column vector. Equation 4 is a (very) undetermined system of equations, with only three equations for the 361 unknowns of the radiance vector, and therefore for every light  $\mathbf{E}_1$  there will be many lights  $\mathbf{E}_2$  producing the same tristimulus  $(L, M, S)$ , i.e. many metamer.

From Equation 4 we can also derive the property of *trichromacy*, which is a fundamental property of human color vision: simply put, it means that we can generate any color by mixing three given colors, merely adjusting the amount of each. In

his excellent account of the origins of color science, Mollon [57] explains how this was a known fact and how it was already applied in the 18th century, for printing in full color using only three types of colored ink. Trichromacy is due to our having three types of cone photoreceptors in the retina, therefore we must remark that it is *not* a property of light but a property of our visual system. This was not known in the 18th century: light sensation was supposed to be transmitted directly along the nerves, so trichromacy was thought to be a physical characteristic of the light. Thomas Young was the first to explain, in 1801, that the variable associated with color in light is the wavelength and, since it varies continuously, the trichromacy must be imposed by the visual system and hence there must be three kind of receptors in the eye. He was also the first to realize that visible light is simply radiation within a certain waveband, and radiation with frequencies outside this range was not visible but could be felt as heat.

Following the approach in [72], we can take any three primaries of spectra  $\mathbf{P}_i, i=1,2,3$ , as long as they are colorimetrically independent, meaning that none of the three lights can be matched in color by a combination of the other two lights. Let  $\mathbf{P}$  be a  $N \times 3$  matrix whose columns are  $\mathbf{P}_i$ . Given a stimulus light of power spectrum  $\mathbf{E}$ , we compute the following three-element vector  $\mathbf{w}$ :

$$\begin{bmatrix} w_1 \\ w_2 \\ w_3 \end{bmatrix} = \mathbf{w} = (\mathcal{S}\mathbf{P})^{-1}\mathcal{S}\mathbf{E}. \quad (5)$$

Pre-multiplying each side of the equation by  $\mathcal{S}\mathbf{P}$ , we get:

$$\mathcal{S}\mathbf{P}\mathbf{w} = \mathcal{S}\mathbf{E}, \quad (6)$$

which, according to Equation 4, means that lights  $\mathbf{Pw}$  and  $\mathbf{E}$  are metamers. But light  $\mathbf{Pw}$  is just a linear combination of the primaries  $\mathbf{P}_i$  with weights  $w_i$ :

$$\mathbf{Pw} = (w_1\mathbf{P}_1 + w_2\mathbf{P}_2 + w_3\mathbf{P}_3). \quad (7)$$

Therefore, any light stimulus  $\mathbf{E}$  can be matched by a mixture of any three (colorimetrically independent) lights  $\mathbf{P}_i$  with the intensities adjusted by  $w_i$ , which is the property of trichromacy.

But, for any set  $\mathbf{P}$  of primaries, there are always some lights  $\mathbf{E}$  for which the weight vector  $\mathbf{w}$  has a negative component [72]. While mathematically this is not a problem, physically it makes no sense to have lights of negative intensity, so in these cases it is *not* possible to match  $\mathbf{E}$  with the primaries. What is physically realizable, though, is to match the mixture of  $\mathbf{E}$  and the primary of the negative weight with the other two primaries. For instance, if the weight vector is  $w_1 = -\alpha, w_2 = \beta, w_3 = \gamma$ , with  $\alpha, \beta, \gamma > 0$ , i.e. the first primary has negative weight, then from equations 6 and 7 we get:

$$\mathcal{S}(\beta\mathbf{P}_2 + \gamma\mathbf{P}_3) = \mathcal{S}(\alpha\mathbf{P}_1 + \mathbf{E}). \quad (8)$$

In this example, then, light  $\mathbf{E}$  cannot be matched by any mixture of  $\mathbf{P}_1, \mathbf{P}_2, \mathbf{P}_3$ , but if we add  $\alpha\mathbf{P}_1$  to  $\mathbf{E}$  we can match this light to  $\beta\mathbf{P}_2 + \gamma\mathbf{P}_3$ . This is precisely what Maxwell observed in his seminal color matching experiments, in 1855-1860, which served as the basis for all technological applications involving color acquisition and reproduction, from cameras to displays to printing. He presented an observer with a monochromatic test light of wavelength  $\lambda$ , and the observer had to match its color by varying the intensity of red, green and blue lights. In the cases when this wasn't possible, the red or the blue light were added to the test light so that the other two primaries were matched to the mixture of the original test light and the remaining primary. For each  $\lambda$  Maxwell recorded the weights  $w_1, w_2, w_3$ , which produced the color match. The weights are also a function of  $\lambda$ : the functions  $w_i(\lambda), i = 1, 2, 3$ , are called the color matching functions for the primaries. In 1861, in a lecture before the Royal Society of London, Maxwell created the first color photograph: he took three black and white photographs of the same object, in each occasion placing a different colored filter (red, green or blue) before the camera; then, each photograph was projected onto the same screen, using three projectors with the corresponding colored filters before them. The final image observed on the screen reproduced, although imperfectly, the colors of the photographed object [20]. In the late 1920's, W. David Wright (and, independently, John Guild) conducted the same experiment with a group of observers, asking them to color-match a given monochromatic light by varying the brightness of a set of red, green and blue monochromatic lights [77]. For each monochromatic test light of wavelength  $\lambda$ , the experiment recorded the (average, over all observers) amounts of each primary needed to match the test:  $\bar{r}(\lambda)$  for the red,  $\bar{g}(\lambda)$  for the green, and  $\bar{b}(\lambda)$  for the blue. These are the color matching functions for a standard observer.

We can see in Figure 5 that the  $\bar{r}(\lambda)$  function clearly has some negative values (the other two functions have negative values as well). As we explained above, these correspond to cases when the test color can't be matched as it is, unless a certain amount of primary color (red) is added to it, in which case the brightness for the green and blue lights can be adjusted so as to match the modified test light.

From the color matching functions, the  $(R, G, B)$  tristimulus values for a light source with spectral distribution  $E(\lambda)$  can be computed as:

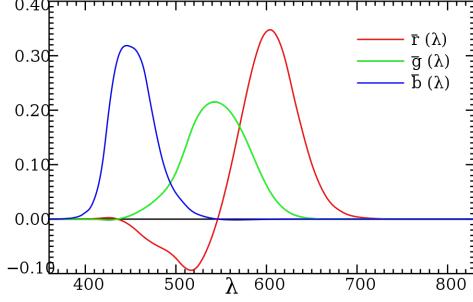


Fig. 5. Color matching functions. Figure from [7].

$$\begin{aligned} R &= \int_{380}^{740} \bar{r}(\lambda) E(\lambda) d\lambda \\ G &= \int_{380}^{740} \bar{g}(\lambda) E(\lambda) d\lambda \\ B &= \int_{380}^{740} \bar{b}(\lambda) E(\lambda) d\lambda, \end{aligned} \quad (9)$$

or, in discrete terms:

$$\begin{aligned} R &= \sum_{i=380}^{740} \bar{r}(\lambda_i) E(\lambda_i) \\ G &= \sum_{i=380}^{740} \bar{g}(\lambda_i) E(\lambda_i) \\ B &= \sum_{i=380}^{740} \bar{b}(\lambda_i) E(\lambda_i). \end{aligned} \quad (10)$$

It is easy to show [72] that, since every light stimulus can be expressed as a mixture of monochromatic stimuli  $E(\lambda_i)$  and these in turn can be color-matched to a linear combination of primaries with weights given by  $\bar{r}(\lambda_i), \bar{g}(\lambda_i)$  and  $\bar{b}(\lambda_i)$ , then it follows that the color matching functions are a linear combination of the cone sensitivity functions  $l(\lambda), m(\lambda), s(\lambda)$ . We will now prove this statement, following [72]; let  $\mathbf{e}_i$  be a monochromatic light of wavelength  $\lambda_i$ , i.e. the vector  $\mathbf{e}_i$  will be 1 at position  $i$  and 0 elsewhere. Then, from Equation 4:

$$\begin{bmatrix} L_i \\ M_i \\ S_i \end{bmatrix} = \mathcal{S}\mathbf{e}_i = \begin{bmatrix} l(\lambda_i) \\ m(\lambda_i) \\ s(\lambda_i) \end{bmatrix}, \quad (11)$$

and from Equation 6:

$$\begin{aligned} \mathcal{S}\mathbf{e}_i &= \mathcal{S}\mathbf{P}\mathbf{w} = \mathcal{S}(\bar{r}(\lambda_i)\mathbf{R} + \bar{g}(\lambda_i)\mathbf{G} + \bar{b}(\lambda_i)\mathbf{B}) = \\ &\quad \bar{r}(\lambda_i)\mathcal{S}\mathbf{R} + \bar{g}(\lambda_i)\mathcal{S}\mathbf{G} + \bar{b}(\lambda_i)\mathcal{S}\mathbf{B} = \\ &\quad \bar{r}(\lambda_i) \begin{bmatrix} L_R \\ M_R \\ S_R \end{bmatrix} + \bar{g}(\lambda_i) \begin{bmatrix} L_G \\ M_G \\ S_G \end{bmatrix} + \bar{b}(\lambda_i) \begin{bmatrix} L_B \\ M_B \\ S_B \end{bmatrix}. \end{aligned} \quad (12)$$

From Equations 11 and 12:

$$\begin{bmatrix} l(\lambda_i) \\ m(\lambda_i) \\ s(\lambda_i) \end{bmatrix} = \bar{r}(\lambda_i) \begin{bmatrix} L_R \\ M_R \\ S_R \end{bmatrix} + \bar{g}(\lambda_i) \begin{bmatrix} L_G \\ M_G \\ S_G \end{bmatrix} + \bar{b}(\lambda_i) \begin{bmatrix} L_B \\ M_B \\ S_B \end{bmatrix}. \quad (13)$$

By definition of  $\mathcal{S}$ :

$$\mathcal{S} = \begin{bmatrix} l(\lambda_{380}) & l(\lambda_{381}) & \dots & l(\lambda_{740}) \\ m(\lambda_{380}) & m(\lambda_{381}) & \dots & m(\lambda_{740}) \\ s(\lambda_{380}) & s(\lambda_{381}) & \dots & s(\lambda_{740}) \end{bmatrix} = \mathcal{M} \begin{bmatrix} \bar{r}(\lambda_{380}) & \bar{r}(\lambda_{381}) & \dots & \bar{r}(\lambda_{740}) \\ \bar{g}(\lambda_{380}) & \bar{g}(\lambda_{381}) & \dots & \bar{g}(\lambda_{740}) \\ \bar{b}(\lambda_{380}) & \bar{b}(\lambda_{381}) & \dots & \bar{b}(\lambda_{740}) \end{bmatrix}, \quad (14)$$

where

$$\mathcal{M} = \begin{bmatrix} L_R & L_G & L_B \\ M_R & M_G & M_B \\ S_R & S_G & S_B \end{bmatrix}. \quad (15)$$

Therefore, if lights  $E_1(\lambda)$  and  $E_2(\lambda)$  are metamers, then not only do they produce the same  $(L, M, S)$  tristimulus (as per definition of metamerism), they also produce the same  $(R, G, B)$  tristimulus. Furthermore, we can convert  $(R, G, B)$  to  $(L, M, S)$  by simple multiplication by a  $3 \times 3$  matrix  $\mathcal{M}$  whose  $i$ -th column ( $i=1,2,3$ ) is the  $(L, M, S)$  tristimulus value of the  $i$ -th primary, (e.g. the first column of  $\mathcal{M}$  is the  $(L, M, S)$  tristimulus of the red primary), and we can convert  $(L, M, S)$  into  $(R, G, B)$  by multiplication by the inverse of  $\mathcal{M}$ . The matrix  $\mathcal{M}$  is invertible because it is not singular, since the primaries must be colorimetrically independent, as pointed out before.

### III. THE FIRST STANDARD COLOR SPACES

From what has been said in the previous section, a color sensation can be described with three parameters; given a test color, we call its *tristimulus* values the amounts of three colors (primaries in some additive color model) that are needed to match that test color. If two single, isolated colored lights have different spectral distributions but the same tristimulus values, then they will be perceived as being of the same color. A *color space* is a method that associates colors with tristimulus values. Therefore, it is described by three primaries and their corresponding color matching functions, as seen in Equation 14.

In 1931, the International Commission on Illumination (or CIE, for its French name) amalgamated Wright and Guild's data [33] and proposed two sets of color matching functions for a standard observer, known as CIE RGB and CIE XYZ; this standard for colorimetry is still today one of the most used methods for specifying colors in the industry. The CIE RGB color matching functions are the functions  $\bar{r}(\lambda), \bar{g}(\lambda), \bar{b}(\lambda)$  mentioned earlier. The tristimulus values  $(R, G, B)$  for a light  $E(\lambda)$  are computed from these functions as stated in Equation 9.

For each wavelength  $\lambda$ , one of the three functions is negative. This posed a problem, since the calculators of the time were manually operated and hence errors were quite common in the computation of the tristimulus values [72]. That's why the CIE XYZ color matching functions  $\bar{x}(\lambda), \bar{y}(\lambda), \bar{z}(\lambda)$  were also introduced alongside the CIE RGB ones. From the CIE XYZ functions, the  $(X, Y, Z)$  tristimulus values for a light source with spectral distribution  $E(\lambda)$  can be computed as:

$$\begin{aligned} X &= \int_{380}^{740} \bar{x}(\lambda) E(\lambda) d\lambda \\ Y &= \int_{380}^{740} \bar{y}(\lambda) E(\lambda) d\lambda \\ Z &= \int_{380}^{740} \bar{z}(\lambda) E(\lambda) d\lambda. \end{aligned} \quad (16)$$

The color matching functions  $\bar{x}(\lambda), \bar{y}(\lambda), \bar{z}(\lambda)$  are obtained as a linear combination of  $\bar{r}(\lambda), \bar{g}(\lambda), \bar{b}(\lambda)$  by imposing certain criteria, chiefly among them:

- $\bar{x}(\lambda), \bar{y}(\lambda), \bar{z}(\lambda)$  must always be positive;
- $\bar{y}(\lambda)$  is identical to the standard luminosity function  $V(\lambda)$ , which is a dimension-less function describing the sensitivity to light as a function of wavelength; therefore,  $Y = \int \bar{y}(\lambda) E(\lambda) d\lambda$  would correspond to the luminance or perceived brightness of the color stimulus;
- $\bar{x}(\lambda), \bar{y}(\lambda), \bar{z}(\lambda)$  are normalized so that they produce equal tristimulus values  $X = Y = Z$  for a white light, i.e. a light with a uniform (flat) spectrum.

An important point we must stress is the following. In section II we saw that for any set of physically realizable primaries there were wavelengths  $\lambda$  for which the color matching values were negative. Since  $\bar{x}(\lambda), \bar{y}(\lambda), \bar{z}(\lambda)$  are always positive, this implies that their primaries can never be physically realizable. This is why the primaries for CIE XYZ are called *virtual primaries*.

#### A. Chromaticity diagrams

There are three quantities that describe our perception of a given color:

- its hue, what we normally refer to as “color” (yellow, red, and so on) and which depends on the wavelength values present in the light;
- its saturation, which refers to how “pure” it is as opposed to “how mixed with white” it is: for instance, the color red (as in blood-red) is more saturated than the color pink (red mixed with white); saturation depends on the spread of the light spectrum around its wavelength(s), a more concentrated spectrum corresponds to a more saturated color;
- its brightness, which expresses the intensity with which we perceive the color; it corresponds to the average of the power of the absorbed light.

It is usual to decouple brightness from the other quantities describing a color, and the pair hue-saturation is referred to as *chromaticity*: e.g. light blue and dark blue have the same chromaticity but different brightness.

Given their relationship to tristimulus values, color spaces are three-dimensional: each color can be represented as a point in a three-dimensional plot. See Figure 6 for a 3D representation of the XYZ color space.

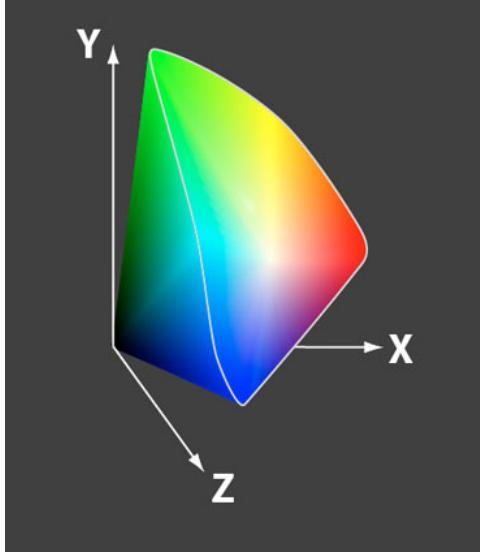


Fig. 6. CIEXYZ color space. Figure from [11].

But, as we just mentioned, when describing colors it is usual to decouple luminance from chromaticity. A light stimulus  $E_1(\lambda)$  and a scaled version of it  $E_2(\lambda) = kE_1(\lambda)$  will produce tristimulus  $(X_1, Y_1, Z_1)$  and  $(kX_1, kY_1, kZ_1)$ , respectively, which have the same chromaticity but different intensity. We now define the values  $x, y, z$ :

$$\begin{aligned} x &= \frac{X}{X + Y + Z} \\ y &= \frac{Y}{X + Y + Z} \\ z &= \frac{Z}{X + Y + Z}. \end{aligned} \quad (17)$$

It is easy to see that, for the abovementioned example of lights  $E_1$  and  $E_2 = kE_1$ , these values are identical:  $x_1 = x_2$ ,  $y_1 = y_2$ ,  $z_1 = z_2$ . This is why  $x, y, z$  are called the *chromaticity coordinates*, because they don't change if the light stimulus only varies its intensity.

By construction  $x + y + z = 1$ , so  $z = 1 - x - y$  and all the information of the chromaticity coordinates is contained in the pair  $(x, y)$ . Therefore, all the possible chromaticities can be represented in a 2D plane, the plane with axes  $x$  and  $y$ , and this is called the CIE  $xy$  chromaticity diagram; see Figure 7.

This tongue-shaped region represents all the different chromaticities that can be perceived by a standard observer; it can be seen as the result of performing this operation: slicing the  $XYZ$  volume with the plane  $X + Y + Z = 1$ , then projecting the resulting plane onto the  $XY$  plane. See Figure 8.

It is worth remarking that the triplet of values formed by chromaticity  $(x, y)$  and luminance  $Y$  perfectly describes a color, and from  $(x, y, Y)$  we can obtain  $(X, Y, Z)$ , and also  $(R, G, B)$ .

The upper boundary of the chromaticity diagram is a horseshoe-shaped curve corresponding to monochromatic colors: this curve is called the spectrum locus [72]. The lower boundary is the purple line, and corresponds to mixtures of lights from the extrema of the spectrum. The relationship between the spectrum  $\mathbf{E}$  of a light stimulus and its tristimulus values  $(X, Y, Z)$  is linear, given by the multiplication of the radiance by a matrix  $\mathcal{S}_{XYZ}$  whose rows are the color matching functions for the CIE XYZ color space:

$$\begin{bmatrix} X \\ Y \\ Z \end{bmatrix} = \mathcal{S}_{XYZ} \mathbf{E}. \quad (18)$$

The linear relationship stated in Equation 18, combined with the definition of chromaticity coordinates in Equation 17, tell us the following: if monochromatic lights  $E_1$  and  $E_2$  have coordinates  $(x_1, y_1)$  and  $(x_2, y_2)$  (that will lie on the spectrum locus

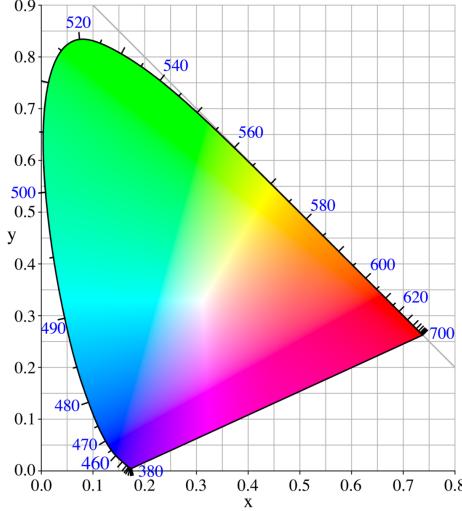


Fig. 7. CIE xy chromaticity diagram. Figure from [7].

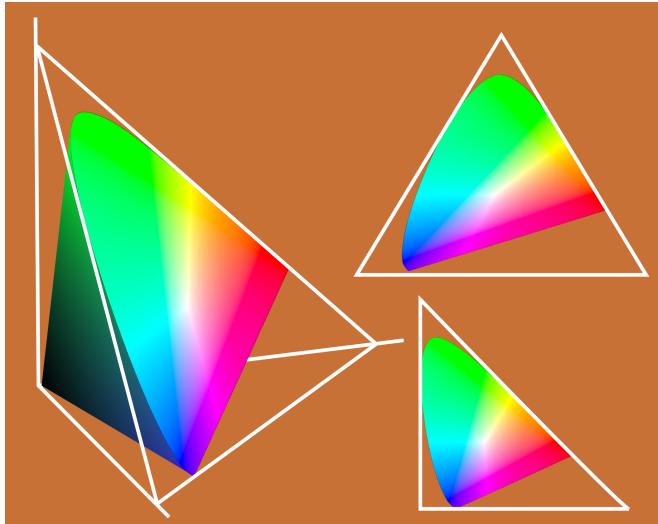


Fig. 8. Left: XYZ volume. Top right: after slicing volume with plane  $X + Y + Z = 1$ . Bottom right: after projecting plane onto  $XY$  plane.

because the lights are monochromatic), the mixture  $E_3 = E_1 + E_2$  will have coordinates  $(x_3, y_3)$  located in the segment joining  $(x_1, y_1)$  and  $(x_2, y_2)$ . Therefore, the tongue-shaped region delimited by the spectrum locus and the purple line represents all the possible chromaticities that we can perceive, as mentioned above.

Mollon [57] explains how in 1852 Hermann Helmholtz provided a formal explanation for the differences between additive color mixing (of lights) and subtractive color mixing (of colored materials): for example, when we mix blue and yellow pigments we obtain green, but when we mix blue and yellow light we obtain white. Helmholtz suggested that pigments were composed of particles that absorbed light of some wavelengths and reflected some others, and the color of the pigment mixture will correspond to the wavelengths that are not absorbed by either of the constituents' pigments. In the aforementioned example, the yellow pigment reflects yellow, red and green but absorbs blue and violet, whereas the blue pigment reflects blue, violet and green and absorbs yellow and red; therefore, the light reflected from the mixture will be green. Helmholtz also showed, after a theoretical work by Hermann Grassmann (1853) that each monochromatic light had a complementary, i.e. the mix of both lights yields white. Monochromatic lights with wavelengths in the range between red and yellow-green have monochromatic complementaries with wavelengths in the range between blue-green and violets. The complementary of green is not a monochromatic light but purple, a mixture of blue and red light from the two ends of the visible spectrum.

Perfect white (i.e. light with a completely uniform power spectrum) has coordinates  $x = y = \frac{1}{3}$ , so as we mix a monochromatic light with white, its chromaticity coordinates move inwards and the saturation of the colors is reduced. A pure monochromatic light has 100% saturation while white has 0% saturation. But in practice, white lights never have a completely flat spectrum. The CIE has defined a set of standard illuminants: A for incandescent light, B for sunlight, C for

average daylight, D for phases of daylight, E is the equal-energy illuminant, while illuminants F represent fluorescent lamps of various compositions [10]. The illuminants in the D series are defined simply by denoting the temperature in Kelvin degrees of the black-body radiator whose power spectrum is closer to that of the illuminant. A black-body radiator is an object that does not reflect light and emits radiation, and the power spectrum of this radiation is uniquely described by the temperature of the object. See Figure 9. For instance, CIE illuminant D65 corresponds to the phase of daylight with a power spectrum close to that of a black-body radiator at 6500K, and D50 to 5000K. These two are the most common illuminants used in colorimetry [72]. The chromaticity coordinates of these illuminants form a curve called *Planckian locus*, shown in Figure 10. This is why in photography it is common to express the tonality of an illuminant by its *color temperature*: a bluish white will have a high color temperature, whereas a reddish-white will have a lower color temperature.

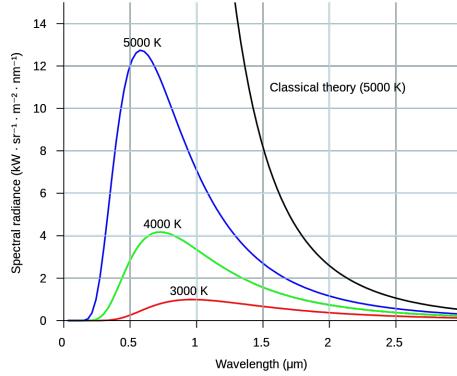
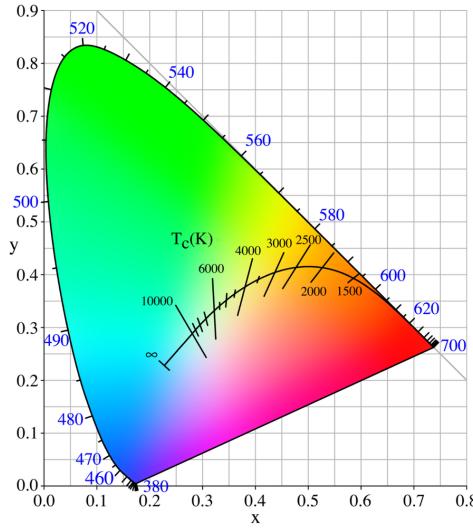


Fig. 9. Spectrum for black-body radiators is a function of their temperature. Figure from [3].



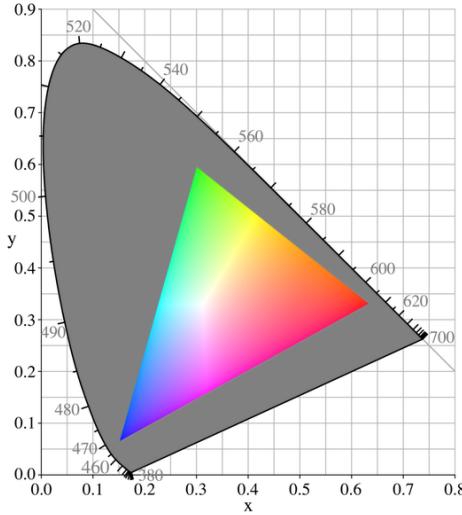


Fig. 11. Chromaticities of a CRT television set. Image from [6].

should perceive as white, whereas the same paper under an orange light will produce radiance with more power in the longer wavelengths, which in theory we should perceive as orange. But as we know this is *not* what happens, we perceive the paper as being white in both cases, a manifestation of what is called *color constancy*, which is our ability to perceive objects as having a constant color despite changes in the color of the illuminant. In the example above, the *stimulus* is different (white light in the former case, orange light in the latter) but the *perception* is the same. Color constancy was already known in the late 17th century but it was first formally addressed by Gaspard Monge in 1789 [57]. Monge conjectured that our perception of color does not depend solely on the characteristics of the light reaching our eye; we also consider the color of the illuminant that lights the scene and we are able to “discount it.” For instance, if an object under a red illuminant reflects white light to our eyes, we will perceive the object as being green, not white, because a white object would send red light. Or, in other words, if we receive white light and take away from it the red of the illuminant, we end up with green.

In 1905, Johannes von Kries formulated an explanation for color constancy that is known as von Kries’ coefficient law and which is still used to this day in digital cameras to perform white balance. This law states that the neural response of each type of cone is attenuated by a gain factor that depends on the ambient light [73]. In practice, von Kries’ law is applied by dividing each element of the tristimulus value by a constant depending on the scene conditions but not on the stimulus: typically, each element is divided by the corresponding element of the tristimulus value of the scene illuminant. Regardless of the original chromaticity of the illuminant, after applying the von Kries’ rule the chromaticity coordinates of the illuminant become  $x = y = z = \frac{1}{3}$ , which correspond to achromatic, white light. In other words, von Kries’ coefficient law is a very simple way to modify the chromaticity coordinates so that, in many situations, they correspond more closely to the *perception* of color.

#### B. Perceptually uniform color spaces

The CIE XYZ color space, and consequently the CIE xy chromaticity diagram, suffer from several limitations in terms of the perception of colors:

- the distance between two points in XYZ space or in the xy diagram is not proportional to the perceived difference between the colors corresponding to the points;
- a mixture of two lights in equal proportions will have chromaticity coordinates that do not lie exactly at the middle of the segment joining the chromaticities of the original two lights.

This can be observed in Figure 12, where equal-radius circles representing perceptual differences of the same magnitude are mapped into different-sized ellipses in the CIE xy diagram. Therefore we can say that, in terms of perception, the CIE XYZ space is not *uniform*.

But in color reproduction systems, perceptual uniformity is a very useful property because it allows us to define error tolerances, and therefore much work has been devoted to the developing of uniform color spaces. Research was carried out independently in two lines: finding a uniform lightness scale and devising a uniform chromaticity diagram for colors of constant lightness [72].

Lightness is the perceived level of light relative to light from a region that appears white, whereas brightness is the overall level of perceived light [73]. Experimentally it has been found that lightness is approximately proportional to the luminance raised to the power of  $\frac{1}{3}$  [78].

In 1976 the CIE introduced two new color spaces: CIE 1976  $L^*u^*v^*$  (abbreviated CIELUV) and CIE 1976  $L^*a^*b^*$  (abbreviated CIELAB). They are both designed to be perceptually uniform, using the same  $\frac{1}{3}$  power law for lightness  $L^*$  and different criteria for chromaticity. In CIELUV the chromaticity coordinates  $u^*, v^*$  are chosen so that just noticeably different colors are roughly equi-spaced. In CIELAB the chromaticity coordinates  $a^*, b^*$  are chosen so that the Euclidean distance between two points in CIELAB space is proportional to the perceptual difference between the colors corresponding to those points. Both CIELAB and CIELUV perform a normalization with respect to the tristimulus of a reference white, in what is a crude approximation to the color constancy property of the visual system; in CIELAB, it is directly based on the von Kries' coefficient law [72].

For CIELUV, the transformation from  $(X, Y, Z)$  to  $(L^*, u^*, v^*)$  coordinates is computed as

$$\begin{aligned} L^* &= \begin{cases} 116 \left( \frac{Y}{Y_n} \right)^{\frac{1}{3}} - 16, & \text{if } \frac{Y}{Y_n} > \left( \frac{6}{29} \right)^3 \\ \left( \frac{29}{3} \right)^3 \frac{Y}{Y_n}, & \text{otherwise} \end{cases} \\ u^* &= 13L^*(u' - u'_n) \\ v^* &= 13L^*(v' - v'_n) \\ u' &= \frac{4X}{X + 15Y + 3Z} \\ v' &= \frac{9Y}{X + 15Y + 3Z} \\ u'_n &= \frac{4X_n}{X_n + 15Y_n + 3Z_n} \\ v'_n &= \frac{9Y_n}{X_n + 15Y_n + 3Z_n}, \end{aligned} \quad (19)$$

where  $(X_n, Y_n, Z_n)$  are the tristimulus values of a reference white, typically called the “white point” and which is usually taken as the brightest stimulus in the field of view: again, following von Kries' approach for color constancy;  $u'_n$  and  $v'_n$  are the  $(u', v')$  chromaticity coordinates of the white point. Figure 13 compares the  $(u', v')$  chromaticity coordinates with the  $xy$  chromaticity diagram.

In CIELAB, the lightness coordinate  $L^*$  is the same as in CIELUV and the chromaticities are computed as

$$\begin{aligned} a^* &= 500 \left( f \left( \frac{X}{X_n} \right) - f \left( \frac{Y}{Y_n} \right) \right) \\ b^* &= 200 \left( f \left( \frac{Y}{Y_n} \right) - f \left( \frac{Z}{Z_n} \right) \right), \end{aligned} \quad (20)$$

where

$$f(x) = \begin{cases} x^{\frac{1}{3}}, & \text{if } x > \left( \frac{6}{29} \right)^3 \\ \frac{1}{3} \left( \frac{29}{6} \right)^2 x + \frac{4}{29}, & \text{otherwise,} \end{cases} \quad (21)$$

which is the same power-law function used for the computation of the lightness  $L^*$ .

In CIELAB the chromaticity coordinates  $(a^*, b^*)$  can be positive or negative:  $a^* > 0$  indicates redness,  $a^* < 0$  greenness,  $b^* > 0$  yellowness and  $b^* < 0$  blueness. For this reason it is often more convenient to express CIELAB colors in cylindrical coordinates  $L^*C^*h^*$ , where

- $C^* = \sqrt{a^{*2} + b^{*2}}$ , the radius from the origin, is the *chroma*, which can be defined as the degree of colorfulness with respect to a white color of the same brightness: decreasing  $C^*$  the colors become muted and approach gray;
- $h^* = \arctan \frac{b^*}{a^*}$ , the angle from the positive  $a^*$  axis, is the *hue*: a hue angle of  $h^* = 0^\circ$  corresponds to red,  $h^* = 60^\circ$  corresponds to yellow,  $h^* = 120^\circ$  corresponds to green, etc.

Analogous correlates for chroma and hue exist for CIELUV in cylindrical coordinates: in that case,  $C_{uv}^* = \sqrt{u^{*2} + v^{*2}}$  and  $h_{uv}^* = \arctan \frac{v^*}{u^*}$ . Figure 14 shows the CIELAB color space in both Cartesian and cylindrical coordinates.

### C. Limitations of CIELUV and CIELAB

We mentioned that in the CIE  $xy$  diagram, Euclidean distances between chromaticities do not correspond to perceptual differences between colors, e.g. the mixture of two colors in the same proportion should be located at the midpoint of the segment determined by these colors in the diagram, but generally this is not the case. And this motivated the research on uniform color spaces CIELAB and CIELUV, where if the Euclidean distance between a pair of colors is the same as the distance between another pair, then the corresponding differences in perception are also (roughly) equal. But this is just an approximation, and CIELUV and CIELAB are both only partially uniform.

In some parts of the color space (mainly around blue) CIELAB suffers from *cross-contamination* [61]: changing only one attribute (such as hue) produces changes in the perception of another attribute (such as saturation). This is a consequence of the

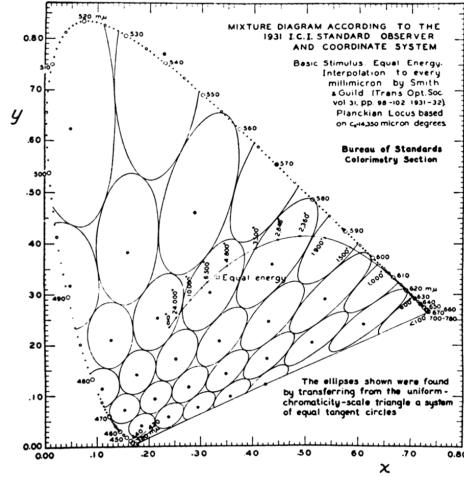


Fig. 12. Ellipses representing the chromaticities of circles of equal size and constant perceptual distances from their center points. Figure from [47].

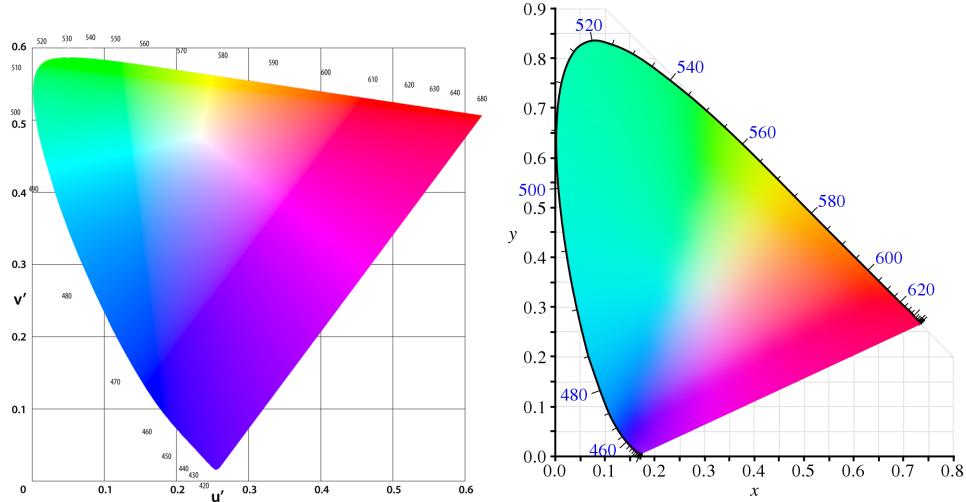


Fig. 13. Left:  $(u', v')$  chromaticity coordinates, from [5]. Right:  $(x, y)$  chromaticity coordinates from the CIE 1931 xy chromaticity diagram. Figure from [7].

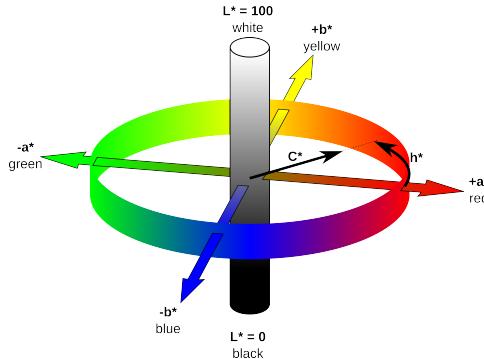


Fig. 14. CIELAB color space in both Cartesian and cylindrical coordinates.

deficiencies of the system with respect to the correlates for hue [72]: the correlate for hue is the angle  $\arctan \frac{b^*}{a^*}$ , and therefore constant hue should correspond to planes passing through the  $L^*$  axis, but what is observed experimentally are curved surfaces instead of planes. These surfaces depart more from the intended planes near the negative  $b^*$  axis, hence the problems around blue.

CIELUV also suffers from approximate perceptual uniformity and poor hue correlation in some regions; furthermore, the translational way of imposing white point normalization (i.e. computing differences  $u' - u'_n$  and  $v' - v'_n$ ) may create imaginary colors (falling outside the spectral locus) in some contexts. Shevell [73] explains that when the CIE introduced CIELUV and CIELAB in 1976 the organization recognized that it was difficult to choose one color space over the other because each worked better on different data sets, but more recently the general opinion seems to favor CIELAB, and CIELUV is no longer widely recommended [32].

Another, very important, limitation of both color spaces is that they are only useful for comparing stimuli under similar conditions of adaptation [72]: they don't consider any of the factors affecting color appearance in real-life conditions.

## V. COLOR APPEARANCE

In laboratory conditions, the three perceptual dimensions of hue, saturation and brightness characterize how we perceive the colors of single, isolated lights. But in real-life situations, lights are seldom isolated and our perception of the color of an object is not completely determined by the light coming from it; it is influenced by factors such as the ambient illumination, the light coming from other objects, or the current state of the neural pathways in eye and brain [73].

We have insisted upon the fact that color appearance is a perceptual, not physical, phenomenon, and have seen an example in the color constancy property of the human visual system: the color appearance of objects is in good correspondence with their reflectance, despite changes in the illuminant, i.e. the light reaching our eyes changes but our perception remains the same. Another example is that of color induction: objects that send the same light to our eyes are perceived as having different colors because of the influence of their surroundings. Figure 15 shows an example: the inner rings in both sets of concentric rings are identical, as the square in the middle shows, yet they appear to have very different hues.

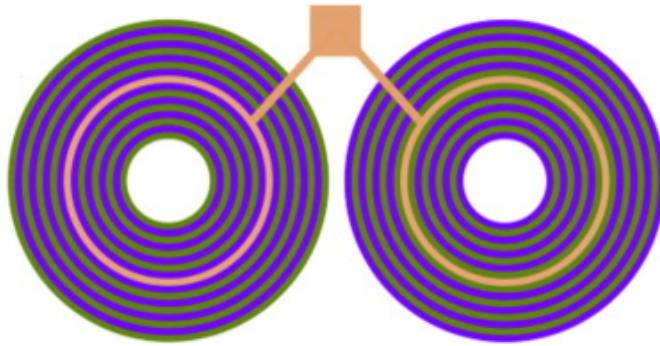


Fig. 15. Chromatic induction: color appearance depends on surroundings. The inner rings are identical, yet they appear to us as having different colors. Figure by P. Monnier [58].

Figure 16 shows the famous checker board illusion by Edward Adelson. Squares A and B are identical (as the image at the right shows) but in the left image they appear to have very different lightness.

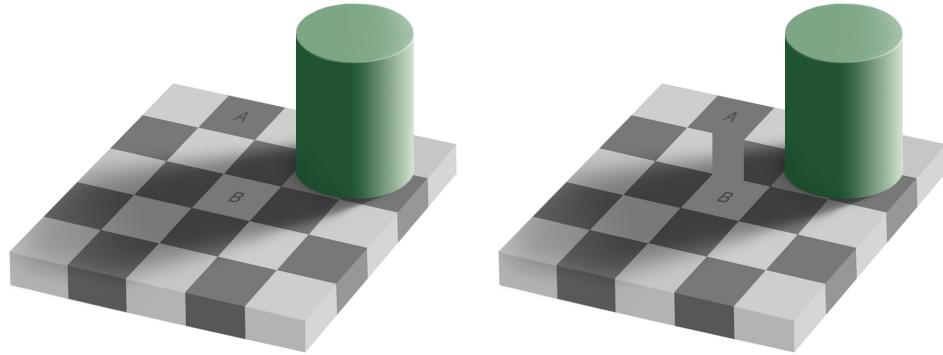


Fig. 16. The “checker shadow illusion” shows that lightness depends on context. Figure by Edward H. Adelson (1995), from [4].

Likewise, our perception of saturation and contrast is also dependent on ambient conditions: saturation decreases when the illumination decreases (the Hunt effect), contrast also decreases with diminishing illumination (the Stevens effect). See Figure 17.

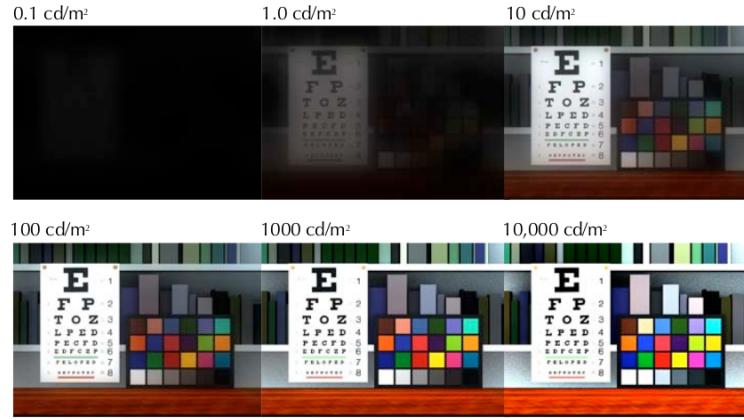


Fig. 17. Hunt and Stevens effects. Image from [31].

Neither the CIEXYZ color space nor CIELUV or CIELAB are useful to compare stimuli under different adaptation conditions, for this we don't need a color space but a *color appearance model*. For given colorimetry under specified reference viewing conditions, a color appearance model predicts the colorimetry required under the specified test viewing conditions for producing the same color appearance [72].

The CIE has proposed several color appearance models, the most recent of which is CIECAM02, published in 2002. The two major pieces of CIECAM02 are a chromatic adaptation transform and equations for computing correlates of perceptual attributes, such as brightness, lightness, chroma, saturation, colorfulness and hue [60]. The chromatic adaptation transform considers chromaticity variations in the adopted white point and was derived based on experimental data from corresponding colors data sets. It is followed by a non-linear compression stage, based on physiological data, before computing perceptual attributes correlates. The perceptual attribute correlates were derived considering data sets such as the Munsell Book of Color. CIECAM02 is constrained to be invertible in closed form and to take into account a sub-set of color appearance phenomena.

## VI. IMAGE PROCESSING PIPELINE

The image processing pipeline of a camera is the sequence of steps in which a digital image is formed, and both the ordering as well as the particular algorithm used in each stage are important [70], [50]. Though camera makers usually don't make public this information, digital cameras commonly perform the following operations, which we'll introduce in this chapter and then discuss in more detail in the remainder of the book: demosaicing, black level adjustment, white balance, color correction, gamma correction, noise reduction, contrast and color enhancement, and compression. At the end of the chapter we'll briefly discuss criteria for the ordering of these operations. Figure 18, taken from the survey by Ramanath et al. [70] shows the schematic of a pipeline.

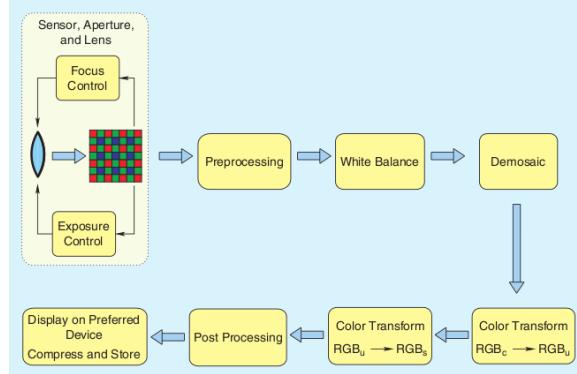


Fig. 18. Image processing pipeline. Figure from [70].

## VII. IMAGE SENSORS

The first stage is that of image acquisition, performed by the camera sensor(s). An image sensor is a semiconductor device that uses the photoelectric effect to convert photons into electrical signals [62]. It is formed by an array of cells (picture elements or pixels); at each of these locations, incident photons raise energy levels in the semiconductor material, letting loose electrons and creating electrical charge [74]. The relationship between number of photogenerated electrons and number of photons is linear, but the subsequent electronic processes (that convert electrons into amplified output voltage) may introduce nonlinearities. The proportion of photons absorbed decays with the wavelength of the incident light, and with wavelengths longer than 1100nm silicon behaves as a transparent material, as Figure 19 (from [62]) shows. But since the visible spectrum ends at 700nm, it is necessary to put in the optical path an infrared (IR) filter that prevents wavelengths higher than 700nm to reach the sensor. This can be achieved by coating an optical surface (the lens or some other optical component in the optical path) with layers of dielectric materials that reflect IR light, in what is known as a *hot mirror*; alternatively, the filter could absorb IR radiation, in which case there's the potential problem of overheating [17].

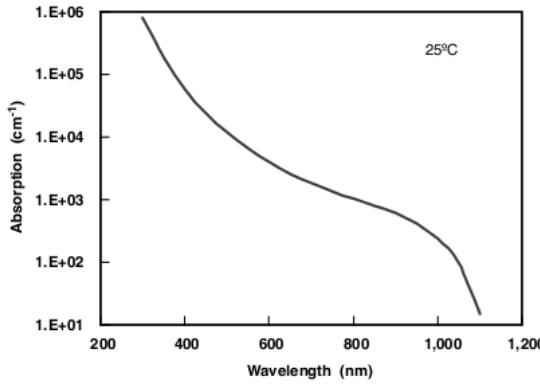


Fig. 19. Absorption coefficient of light in silicon. Figure from [62].

### A. Pixel classes

Pixels can be of two different classes:

- Photodiodes, which store the electrical charge around metal junctions created by implanting ions into the silicon.
- Photogates, which store the electrical charge in “potential wells” created by capacitors.

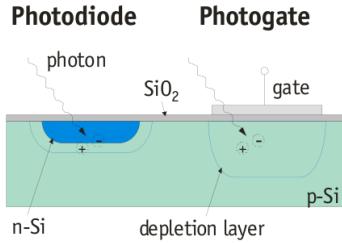


Fig. 20. Types of picture elements. Figure from [74].

A most important characteristic in a pixel is its fill factor, which is the portion of the pixel area that is photosensitive. The major advantage of photogates is their very high fill factor, of almost 100%, allowing them to convert more photons and generate a larger signal. As a downside, the polysilicon gate over the pixel reduces its sensitivity in the blue end of the spectrum [74]. Photodiodes, on the other hand, have better sensitivity but they are more complex and require the presence of opaque circuit elements (transfer gate, channel stop region to isolate pixels, shift register to move the charge) that considerably reduce the fill factor to 30-50% [62].

In order to overcome these limitations, photogates may use very thin polysilicon gates so as not to compromise sensitivity, and photodiodes may use microlenses to increase fill factor. A microlens array is placed on top of the photodiode array, and each microlens focuses incident light onto the photosensitive part of the pixel, effectively increasing fill factor to around 70%; see Figure 21. This approach noticeably enhances sensitivity, but it's not free of shortcomings. Its main disadvantage is that it produces shading in the image, because now the percentage of photons that are focused onto the photosensitive part of the pixel depends on the angle of incidence of the light; see Figure 22. Possible solutions to this issue include decreasing the distance between the microlenses and pixel surface, shifting the position of the microlenses depending on image location, and adding another layer of microlenses [62]. Furthermore, with large apertures the microlens array may cause vignetting, pixel crosstalk, light scattering and diffraction problems [43]. Some of these problems can be reduced by means of image processing after capture [74].

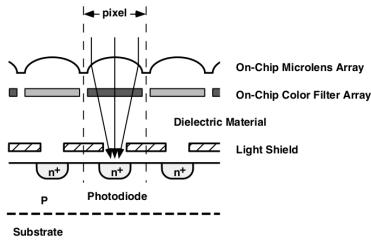


Fig. 21. Scheme of photodiode with microlens array. Figure from [62].

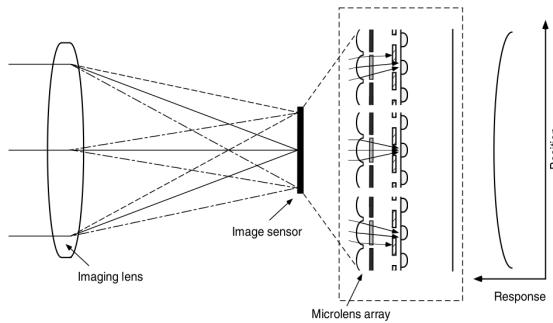


Fig. 22. Shading caused by microlens array. Figure from [62].

### B. Sensor classes

The electrical charge is accumulated while the sensor is being exposed to light, and then it must be converted into voltage or current through the scanning of the image array. There are two large families of sensor devices (both using photodiodes and photogates) and their differences arise from the way they perform this scanning:

- A Charged Coupled Device (CCD) transfers the charge vertically from pixel to pixel over the entire array, then it is transferred horizontally and converted into voltage at just one output amplifier. See Figure 23, left.
- CMOS (Complementary Metal Oxide Semiconductor) devices perform the charge-to-voltage conversion at each pixel location. See Figure 23, right.

These two approaches imply very different implementation strategies, as we will discuss.

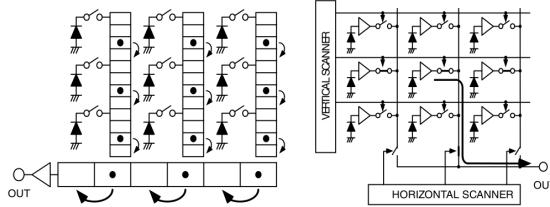


Fig. 23. Scanning the image array. Left: CCD transfer scheme. Right: CMOS X-Y address scheme. Figure from [62].

CCDs have a very high uniformity in their output, because they use a limited number of amplifiers; on the other hand, these amplifiers must be able to work at the very high bandwidth of cinema (several millions of pixels per second) under the limitation of amplifier noise. CMOS devices have at least one amplifier per pixel, which has advantages (these amplifiers can be simple, with low bandwidth and low power consumption, allowing for higher frame rates) and inconveniences (it's difficult to make millions of amplifiers in a uniform way, so CMOS devices suffer from higher fixed-pattern noise; amplifiers reduce the fill-factor, therefore microlens arrays are almost always used with CMOS) [74].

Most sensors output analog signals that are later digitized at a camera module. Given that in principle CMOS sensors are smaller and have lower consumption than their CCD counterparts, the analog-to-digital conversion can be implemented at the sensor along with some image processing functionality (the so-called “camera on a chip” approach) although this causes complexity and optimization problems which explains why most successful CMOS imagers don't perform analog to digital conversion on chip [74].

### C. Interlaced vs. progressive scanning

There are two modes of scanning: progressive and interlaced. In the progressive mode all lines in an image are read in sequential order, one after the other. In the interlaced mode the odd lines are read first, and then the even lines, thus creating two image *fields* for every image *frame*. Interlaced scanning has its origins in the early days of television, when it was devised as a way of achieving an acceptable compromise between frame-rate, image resolution and bandwidth requirements. It was incorporated into all major broadcasting standards, and because of backward-compatibility policies it still is very much in use, although definitely not in professional cinema. But for non-professional cinema as well as for legacy material there is the need of performing interlaced to progressive conversion, since current display technology (unlike, say, CRT TV sets) produces a progressive output.

Figure 24 shows examples of the same image acquired under these different modes. Notice how motion is quite problematic both for interlaced and raster progressive scanning. This latter case is the rolling-shutter problem of simple CMOS imagers, which we'll comment on later.



Fig. 24. Left: progressive scanning, raster acquisition (line by line). Middle: interlaced scanning. Right: progressive scanning, simultaneous acquisition. Figure from [63].

But for image sensors, interlaced scanning has the advantage of allowing us to average line values and hence increase sensitivity (at the expense of reducing vertical resolution) [79]. Also, in CCDs interlaced scanning requires simpler circuitry, permitting a larger resolution [63].

#### D. CCD types

There are three main types of layouts for CCDs [74]:

- Full frame CCDs use photogates and have the simplest sensor layout with the highest fill factor. A mechanical shutter is required to block the light during charge-transfer (otherwise the image would be smeared while the sensor is reading it line by line) but this is not a problem in cinematography since most cameras come equipped with a rotating shutter.
- Frame transfer CCDs also use photogates. They have a duplicated frame, the regular full frame plus a storage region to which the image is transferred (at high speed) and where it is read out, while at the same time the next image is being acquired. This layout improves on speed and reduces smear, at the price of duplicating silicon and increasing power dissipation.
- Interline transfer (ILT) CCDs use photodiodes which, as we mentioned, provide good sensitivity for the blue part of the spectrum at the price of reducing fill factor due to the opaque circuit elements (these elements allow for electronic shuttering, but in digital cinema this is not really an improvement since cameras already come with mechanical shutters), which is why most ILT CCDs use microlens arrays.

The charge transfer scanning procedure of CCDs makes them subject to blooming, which occurs when charge overflows a pixel (e.g. if there is a bright light source in the shot) and spills vertically; see Figure 25. In order to perform antiblooming, an overflow drain must be added.

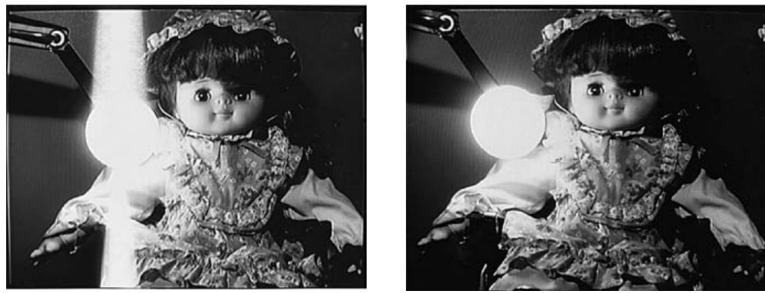


Fig. 25. Left: blooming. Right: CCD with overflow drain. Figure from [63].

#### E. CMOS types

In their most simple configuration, CMOS sensors have one transistor per (photodiode) pixel. This “passive pixels” architecture provides a good fill factor (though still much lower than with CCDs) at the price of a low signal to noise ratio. To reduce the noise, an amplifier is added to each pixel in the “active pixels” layout: as each amplifier is made with three transistors, this configuration is also referred to as 3T. Noise response is better than with passive pixels, but also lacking. To improve it, the “pinned photodiode” configuration requires an extra diode (so it’s also called 4T) but allows us to perform Correlated Double Sampling (CDS) [26]. CDS is a noise reduction technique consisting of sampling two images, one with the shutter closed and another after exposure, and subtracting the latter from the former, thus reducing *dark current* noise (we’ll discuss noise sources below). An extra transistor (5T) allows us to perform global shuttering, i.e. to cease exposure on all pixels simultaneously, thus avoiding image artefacts related to fast motion [74]. Without global shuttering, images with motion are distorted by the so-called *rolling shutter effect*: rows are read sequentially, and the time lag between rows may cause noticeable visual artifacts such as stretching of objects when motion is horizontal; see Figure 26.

Transistors are opaque, therefore most CMOS devices use microlens arrays. Microlenses with CMOS sensors pose worse problems than with CCDs: on top of the silicon surface is a stack of transistors and circuitry, which forces the microlens to be further away from the pixel thus compromising color response and introducing a strong dependence of sensitivity on the angle of incidence of the light. Also, more involved circuitry requires piling up more layers on top of the substrate which increases noise and optical cross-talk [26].

#### F. Noise in image sensors

The amount of noise present in the image signal is a very important characteristic of the sensor device. The dynamic range of the sensor is a magnitude that conveys its ability to capture detail both in bright and dark regions of the image simultaneously, and it is described using the ratio between the pixel’s largest charge capacity and the noise floor; it is expressed as a ratio (e.g. 4096:1), in decibels, or bits [74]. Noise can be classified as constant in time (fixed-pattern noise or FPN) or varying in time (temporal noise) [62].

FPN is caused mainly by dark current and shading, and being constant it can not be eliminated by Correlated Double Sampling (CDS). Dark current is a parasite current, not originated from photon conversion, that nevertheless is integrated

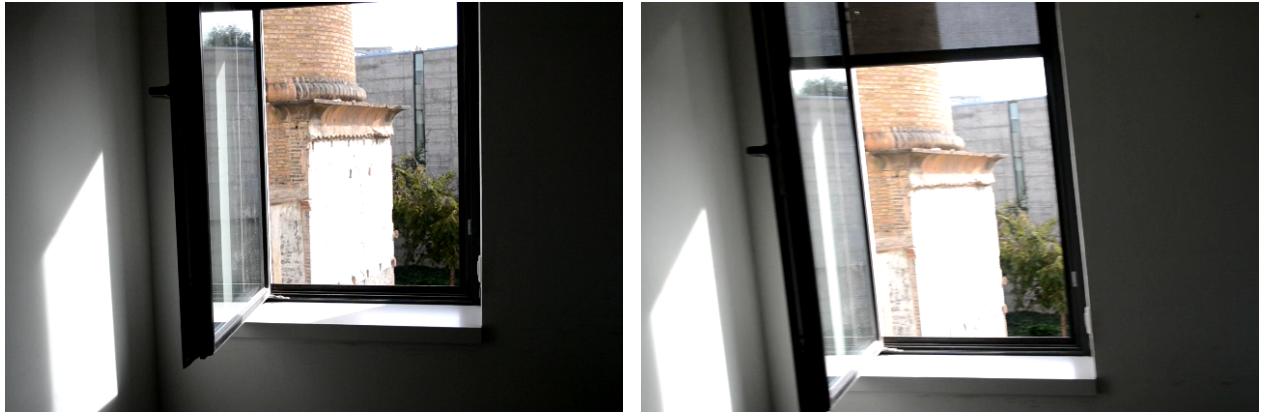


Fig. 26. The rolling shutter effect. Left: image with static camera. Right: image when camera moves.

as charge. This current may be originated inside a pixel, or in a transfer channel of a CCD sensor, or at an amplifier in a CMOS device. It increases with exposure time and with temperature, reducing dynamic range (by increasing the output level corresponding to “dark”). The borders of image arrays are composed of “optical black pixels,” which are never exposed so that they can be used to estimate dark current levels and therefore a proper black level for the image. Shading is a slow spatial variation in the brightness of the image, which may be originated by a local heat source (which in turn generates dark current), a microlens array (as discussed above), non-uniformity of electrical pulses in CCDs, or nonuniformity of bias and ground in CMOS sensors [62].

Temporal noise manifests in three kinds of noise: thermal, shot and flicker, both in CCD and CMOS sensors. Thermal noise is due to thermal motion of electrons inside a resistance. MOS transistors behave as a resistance during exposure, making thermal noise appear (this is called “reset noise” or “kTC noise”); readout electronics are also a thermal noise source [62]. Shot noise is caused by the discrete nature of incident photons and generated electrons. The arrival of photons follows a Poisson distribution which, if the number of particles is not high, exhibits fluctuations that translate into variations in brightness called photon shot noise. Dark current shot noise is the same type of phenomenon, in this case associated to the random generation of charge in the pixel. Flicker noise, also called 1/f noise or pink noise, is originated by the properties of the materials that make up amplifiers, and exhibits a spectral distribution that is inversely proportional to frequency (hence the 1/f name) as opposed to thermal and shot noise, which have a uniform spectral distribution (noise with a uniform spectral distribution is called “white” in analogy with white light, whereas 1/f noise is called “pink” because that’s the appearance of light with a 1/f distribution).

#### G. Capturing colors

As we have seen, image sensors work by transforming photons into electrical signals, but this is only a measure of the incident light intensity, not of its color. Photons don’t carry color information, since there is only one kind of photon; color is a property of the wavelengths of the light. In order to capture colors, two main configurations are in use [17]:

- Three-sensor systems. Incoming light is separated into three different color channels using a beam splitter (a type of prism) and there is one sensor devoted to each. See Figure 27 (left).
- Color Filter Arrays (CFA). In this one-sensor approach, the image array is covered by a mosaic color filter of three colors, making each pixel in the array capture one color channel only. The CFA is added to the sensor either by a pigment process or a dye process (pigment-based CFAs are more resistant to heat and light). See Figure 27 (right).

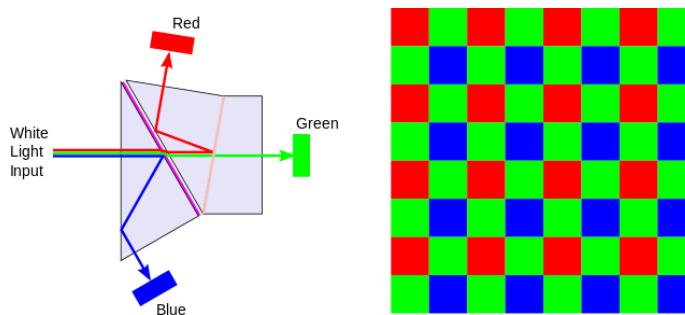


Fig. 27. Capturing color information. Left: a three sensor system. Right: Bayer CFA over a single sensor. Left image adapted from [8].

Practical problems abound with the three-sensor approach: sensors must be very carefully aligned to avoid chromatic aberration, the simple refraction of beamsplitters may not be enough for precise color separation, and the optical path through the prism increases both lateral and longitudinal aberration [74].

As for CFAs, color choices for the filter are primary colors (red, green and blue) and complementary colors (cyan, yellow and magenta); RGB CFAs are superior in terms of color separation, color reproduction and signal to noise ratio. By far the most popular CFA is the Bayer pattern, using an RGB configuration with two green pixels per red or blue pixel (with the rationale that the human visual system is more responsive to light of medium wavelengths) [62]. The downside is that, since for each pixel we only have color information for one channel, an interpolation process called demosaicing must be carried out to estimate the values of the other two channels. Furthermore, in the presence of aliasing the demosaicing algorithms may cause substantial visual artifacts, especially if the aliasing is not taking place in the three channels.

In order to avoid this, antialiasing filters are used, mainly based on one of these two techniques [17] illustrated in Figure 28:

- polarization: some pieces of birefringent material split an incoming light ray into several beams, each going to a different pixel;
- phase delay: the surface of an optical element is etched with a pattern, so that some light goes through more filter material, suffering a phase delay and interfering with the light going through less filter material, thus reducing higher frequencies.

These antialiasing elements are called optical low-pass filters (OLPF), and they provide the only effective way of reducing aliasing, which can't be handled adequately in post-production [16]. The use of OLPFs is essential in movies shot with DSLR cameras, where the sensor may have 10-20 megapixels from which the 2 megapixels HD video images are obtained simply by decimation, i.e. just by downsampling, not by averaging pixels with their neighbors [29]. Therefore, unless an OLPF is used, the final HD video is practically guaranteed to show aliasing.

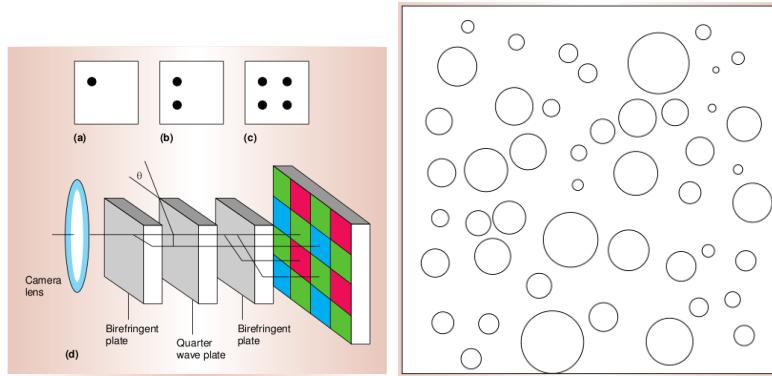


Fig. 28. Left: antialiasing based on polarization, (a) to (c) are the split beams, (d) shows the full light path. Right: phase delay antialiasing filter. Images from [17].

### VIII. EXPOSURE CONTROL

Exposure is the amount of light that is allowed to reach the sensor while capturing an image. In order to regulate it we may modify the aperture (varying the f-number) and/or the shutter speed (which defines the exposure time). In photography, an *exposure value* or EV denotes all combinations of relative aperture and shutter speed that provide the same exposure:

$$EV = \log_2 \left( \frac{F^2}{t} \right), \quad (22)$$

where  $F$  is the f-number and  $t$  is the exposure time (in seconds), and the quantity  $\frac{F^2}{t}$  is proportional to the exposure. For instance, if we double the exposure time but divide by two the aperture area, the exposure remains the same and so does the EV. If we just double the exposure time while keeping the relative aperture fixed, then the exposure is also doubled and the EV decreases in one unit (one “stop”). The optimum exposure for a given image is clearly image-dependent, e.g. more brightly lit scenes will require a larger EV.

Automatic exposure control is the set of processes by which the camera modifies aperture and shutter speed (and sometimes sensor gain) in order to expose each image frame correctly, i.e. avoiding overexposure of bright image regions and underexposure of dark regions. Cinema professionals prefer manual exposure control, mainly because automatic controls do not allow the cinematographer to choose freely the aesthetics of the image, but also because the automatic exposure may degrade the quality of the image, increasing noise when compensating for low light levels, modifying the depth of field when changing the aperture, or creating annoying oscillations in the average luminance in highly dynamic scenes [29].

As explained in [21], most auto-exposure algorithms work by taking a (temporary, not to be recorded) picture with a pre-determined exposure value  $EV_{pre}$ , computing a single brightness value  $B_{pre}$  from that picture, selecting an optimal brightness value  $B_{opt}$ , computing the optimal exposure value thus:

$$EV_{opt} = EV_{pre} + \log_2 \left( \frac{B_{pre}}{B_{opt}} \right), \quad (23)$$

and finally modifying aperture, shutter speed and gain so as to take an exposure-corrected picture with exposure value  $EV_{opt}$ .

#### A. Exposure metering

There are several strategies for metering [28]:

- Zone metering. A single output measurement is obtained as the weighted average of contributions from a number of zones, with higher weights for central areas. The most common approach is to compute the average luminance; but median, mode (of the histogram distribution of the luminance) or peak white measurements (to avoid clipping of very bright areas) can also be used.
- Matrix or multi-zone metering. Light intensity is measured in several points in the scene, and several factors are taken into account to derive the best exposure value: autofocus point, distance to subject, areas in focus or out of focus, colors, backlighting, and, very importantly, matchings with a pre-stored database of thousands of exposures. This metering system is mainly used with still cameras, as its complexity may make it unstable with dynamic scenes.
- Content-based metering. These systems try to expose optimally those regions in the image that are deemed most relevant, using measures such as contrast, focus, skin-tones, object-based detection and tracking, etc.

#### B. Control mechanisms

As mentioned above, there are three ways to control the exposure, by modifying the aperture, shutter speed and sensor gain.

The aperture control may be performed entirely at the lens (AC iris) with an integrated amplifier changing the iris area in order to keep constant the exposure; or the signal may come from outside the lens unit (DC iris). The latter is the preferred approach for high-end video; difficulties arise from the fact that the camera must provide signal and control mechanisms that should work with any DC iris lens. Modifying shutter speed and sensor gain provides stable and fast (next frame) exposure control; also, reducing exposure time decreases motion blur, which is something that can't be achieved via aperture control and is critical in rolling-shutter CMOS sensors. On the other hand, while integration time may change by a factor of 1000 at the most, apertures have a much wider range, and if the amount of light is large it is better to reduce the aperture so that less light reaches the sensor, thus avoiding deterioration of the dye materials and burn-in effect. But each option has its problems: short time exposures cause flicker, and small apertures increase depth of field and may not be too small in order to avoid diffraction problems. The output signal level is regulated also by the sensor gain. While increasing exposure time increases SNR, increasing the gain does not change SNR but it does increase the amplitude of the noise. Therefore, it is often preferred to leave gain control as a last option if controlling exposure time and aperture is not sufficient [28].

#### C. Extension of dynamic range

Scenes with a high dynamic range, with details both in bright and dark regions, may require a sensor with a dynamic range of 100dB, whereas a good CCD or CMOS sensor may be able to provide 74dB. As we mentioned above, reducing the noise floor is essential to increase the dynamic range of the signal in standard CCD and CMOS sensors. There are other alternatives, like using non-linear (log-shaped) response sensors, dual-pixel sensors (each pixel is split into two parts of different responsiveness), or exposure bracketing (long and short exposure frames are captured and then merged). But none of them really works, having color fidelity, sensitivity or ghosting problems, so HDR video remains an open problem [28].

## IX. FOCUS CONTROL

Automatic focus control (AF) is the process by which the distance between sensor and optical system is regulated, so that the image of the object of interest in the scene is formed on a plane that coincides with the surface of the sensor. Digital cameras may use different AF mechanisms (see [44] and references therein):

- infrared AF, where an infrared ray is used to estimate the distance to the object of interest;
- through-the-lens AF, where the sensor-lens distance is adapted until the signals from both the upper and the lower part of the lens are in phase [70];
- contrast detection AF, where the degree of focus is estimated as the amount of high-frequency information in the image (since blurry regions have lower frequency components).

Nevertheless, and as was the case with automatic exposure, cinema professionals absolutely avoid the autofocus: deciding what is in focus and what is not is a crucial artistic choice, and camera operators usually have an assistant (focus puller) who, as the shot progresses, manually changes the focus making sure it evolves in accordance with the director's intentions [29].

## X. WHITE BALANCE

An essential property of human vision is our ability to perceive as constant the color of an object despite variations in the illumination conditions. We see a white shirt as being the same white under daylight as under (yellowish) tungsten light or under (bluish) fluorescent light. The same happens if the color of the shirt is other than white; we don't see it changing under changing illuminants. But the light coming from the shirt and reaching our eyes does indeed change with the illumination, that is, different *sensations* (e.g. white, yellowish and bluish light) produce the same *perception* of color (white). Furthermore, in the same scene we may have two objects producing the same sensation (the light coming from them has the same intensity and spectral power distribution) but we perceive them as being of different colors. It is clear then that our visual system is doing more than measuring the light at each location in the image; therefore we need our cameras to mimic this behavior; otherwise, if we just used the triplets of colors as captured by the sensor, the colors would never appear right except under the exact lighting conditions used during the calibration of the camera [17].

This problem is usually considered with several simplifying assumptions: all objects in the scene are flat, matte, Lambertian surfaces (diffuse surfaces, which reflect light evenly in all directions) and uniformly illuminated [37], with a single illuminant. In this scenario, the triplet of RGB values captured at any given location of the sensor is:

$$\begin{aligned} R &= \int_{380}^{780} r(\lambda) I(\lambda) S(\lambda) d\lambda \\ G &= \int_{380}^{780} g(\lambda) I(\lambda) S(\lambda) d\lambda \\ B &= \int_{380}^{780} b(\lambda) I(\lambda) S(\lambda) d\lambda, \end{aligned} \quad (24)$$

where  $r(\lambda)$ ,  $g(\lambda)$  and  $b(\lambda)$  are the spectral sensitivities of the red, green and blue filters used by the camera,  $I(\lambda)$  is the power distribution of the illuminant, and  $S(\lambda)$  is the spectral reflectance of the object.

Experiments in color constancy [46] indicate that our perception of the color of an object is in many cases independent of the illuminant and matches pretty well the reflectance values of the object. Therefore, if the camera is to replicate this behavior, it must "discount the illuminant" from the observed RGB values. Another assumption used in practice is to treat the spectral sensitivity functions  $r(\lambda)$ ,  $g(\lambda)$  and  $b(\lambda)$  as if they were delta functions centered at the peak sensitivities  $\lambda_R$ ,  $\lambda_G$  and  $\lambda_B$ , respectively. This is *not* an accurate representation of these functions, which are very much spread over the visible spectrum as we see in Figure 29, but it is a useful simplification for the problem at hand, good enough for many contexts [36]. Under this hypothesis then, the equations 24 become:

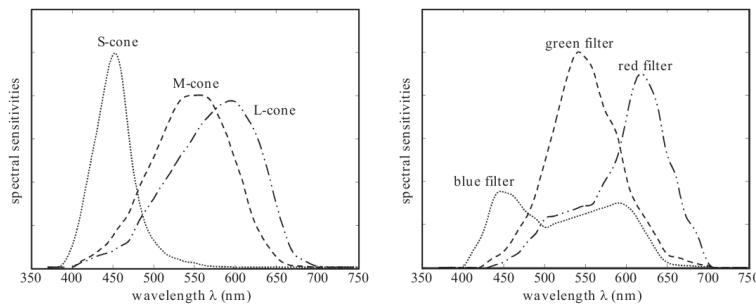


Fig. 29. Spectral sensitivities of: (a) the three types of cones in a human eye, and (b) a typical digital camera. Images from [53].

$$\begin{aligned} R &= I(\lambda_R) S(\lambda_R) \\ G &= I(\lambda_G) S(\lambda_G) \\ B &= I(\lambda_B) S(\lambda_B), \end{aligned} \quad (25)$$

or, in matrix form:

$$\begin{bmatrix} R \\ G \\ B \end{bmatrix} = \begin{bmatrix} I(\lambda_R) & 0 & 0 \\ 0 & I(\lambda_G) & 0 \\ 0 & 0 & I(\lambda_B) \end{bmatrix} \begin{bmatrix} S(\lambda_R) \\ S(\lambda_G) \\ S(\lambda_B) \end{bmatrix}. \quad (26)$$

Since the camera observes the values  $(R, G, B)$  but we want to have  $(S(\lambda_R), S(\lambda_G), S(\lambda_B))$ , we need to estimate the illuminant  $(I(\lambda_R), I(\lambda_G), I(\lambda_B))$ ; once we have the illuminant, the reflectances could be recovered with a simple division:

$$S(\lambda_R) = \frac{R}{I(\lambda_R)}, \quad S(\lambda_G) = \frac{G}{I(\lambda_G)}, \quad S(\lambda_B) = \frac{B}{I(\lambda_B)}. \quad (27)$$

The white balance process by which a color corrected triplet  $(R', G', B')$  is obtained from  $(R, G, B)$  can be written in matrix form thus:

$$\begin{bmatrix} R' \\ G' \\ B' \end{bmatrix} = \begin{bmatrix} \frac{1}{I(\lambda_R)} & 0 & 0 \\ 0 & \frac{1}{I(\lambda_G)} & 0 \\ 0 & 0 & \frac{1}{I(\lambda_B)} \end{bmatrix} \begin{bmatrix} R \\ G \\ B \end{bmatrix}. \quad (28)$$

Some cameras proceed this way although it is usually preferred to apply the diagonal model of Equation 26 not directly on the RGB values but on their correspondent cone tristimulus values, since color adaptation in the eye is very much dependent on the sensitivity of the cones [42]. We can convert *RGB* values into CIE tristimulus values *XYZ* with:

$$\begin{bmatrix} X \\ Y \\ Z \end{bmatrix} = A \begin{bmatrix} R \\ G \\ B \end{bmatrix}, \quad (29)$$

where  $A = [A_{ij}]$  is a  $3 \times 3$  matrix that is chosen for each camera so as to optimize the color reproduction of a given (small) set of colors considered important [42], under a single illuminant, and applied as it is for all the illuminants that can occur [23] (see next section for more details). The color correction procedure can then be written as:

$$\begin{bmatrix} R' \\ G' \\ B' \end{bmatrix} = A^{-1} \begin{bmatrix} \frac{1}{I_X} & 0 & 0 \\ 0 & \frac{1}{I_Y} & 0 \\ 0 & 0 & \frac{1}{I_Z} \end{bmatrix} A \begin{bmatrix} R \\ G \\ B \end{bmatrix}. \quad (30)$$

The white balance process consists then of two steps, illumination estimation and color correction. Ideally it should be performed not on-camera but afterwards, as offline post-processing: the rationale, as with demosaicking and denoising algorithms, is that with offline postprocessing we have much more freedom in terms of what we can do, not being limited by the constraints of on-camera signal processing (in terms of speed, algorithm complexity and so on). Offline we may use more sophisticated color constancy methods, which do not require all the simplifying assumptions enumerated before and which can therefore deal more accurately with realistic situations where surfaces are not perfectly diffuse but have specular reflections, where there is more than one illuminant or it is not uniform, etc. In fact, in [55] it is shown how even small departures from perfectly uniform illumination generate considerable deviations in appearance from reflectance, which goes to say that the diagonal models of white balance, based on discounting the illumination and equating appearance with reflectance, do not predict appearances in real life scenes. But offline white balance requires that the image date is stored as *raw*, i.e. as it comes from the sensor, and many cameras do not have this option, storing the image frames already in color corrected (and demosaicked, and compressed) form.

If offline white balance is not an option, then the next best possibility is that of manual illumination estimation: the operator points the camera towards a reference white, such as a simple sheet of paper, and the triplet of values recorded for this object (at, say, the center of the image) is used by the camera as  $(I(\lambda_R), I(\lambda_G), I(\lambda_B))$  to perform the color correction (because in a white object we have that  $(S(\lambda_R) = S(\lambda_G) = S(\lambda_B))$ , so the observed  $(R, G, B)$  should be equal to  $(I(\lambda_R), I(\lambda_G), I(\lambda_B))$ ). This method ensures that objects perceived as white at the scene will also appear white in the recorded images, and in general that all achromatic objects will appear gray. But cinematographers often find this effect too realistic, and use instead the manual white balance for artistic expression by fooling the camera, giving it as reference white an object with a certain color; in this way, by deliberately performing a wrong color correction, a certain artistic effect can be achieved [29].

Finally there is the option of automatic white balance (AWB), where the illumination estimation is automatically performed on-camera and the color correction (done as in Equation 28 or Equation 30, i.e. “discounting the illuminant”) is carried out on the raw data domain or just after color interpolation [71]. The most common approaches for illuminant estimation in AWB are:

- *Gray World*. This approach was formalized by Buchsbaum in 1980 [25], although the same technique was proposed by Judd forty years before [45], [48]. The main assumption here is that the colors present in the scene are sufficiently varied, in which case the average of reflectances is gray; in other words, that reflectances are uniformly distributed over the interval  $[0, 1]$  and hence their average is 0.5 for each waveband. Assuming also that the illuminant is uniform and averaging Equation 25 over the whole image, we get:

$$\begin{aligned}
\frac{1}{A} \sum_{x,y} R(x,y) &= I(\lambda_R) \frac{1}{A} \sum_{x,y} S(\lambda_R, x, y) = \frac{I(\lambda_R)}{2} = R_{average} \\
\frac{1}{A} \sum_{x,y} G(x,y) &= I(\lambda_G) \frac{1}{A} \sum_{x,y} S(\lambda_G, x, y) = \frac{I(\lambda_G)}{2} = G_{average} \\
\frac{1}{A} \sum_{x,y} B(x,y) &= I(\lambda_B) \frac{1}{A} \sum_{x,y} S(\lambda_B, x, y) = \frac{I(\lambda_B)}{2} = B_{average},
\end{aligned} \tag{31}$$

where  $A$  is the area of the image. In practice, one channel (usually the green one) is taken as reference and the other two are scaled to perform the color correction, since AWB is only concerned about the ratio of the color signals [53]:

$$\begin{aligned}
R'(x,y) &= \alpha R(x,y) \\
G'(x,y) &= G(x,y) \\
B'(x,y) &= \beta B(x,y),
\end{aligned} \tag{32}$$

where:

$$\alpha = \frac{G_{average}}{R_{average}}, \quad \beta = \frac{G_{average}}{B_{average}}. \tag{33}$$

The end result is that, after this correction, the average of the new image is gray:  $R'_{average} = G'_{average} = B'_{average}$ . Clearly, this method is not effective if the main hypothesis is violated, i.e. if the colors of the image are not sufficiently varied (e.g. when a large, monochromatic object takes up most of the image). To try to overcome this limitation, some approaches like [59] compute the color distribution of achromatic charts under a set of typical illuminants, then decompose the image into blocks and for each block analyze its color differences so as to determine if it's in the same region as that of a common illuminant; finally, all blocks are considered, along with the absolute scene light level, to determine the likely scene illuminant [65].

- *White patch.* This approach is based on the fact that the brightest object in a scene is perceived as white. The observation of this phenomenon is often attributed, incorrectly, to the Retinex theory of Land [54], but it has a long history that dates back at least to the works of Helmholtz [49], [46]. We start by computing the maximum value for each color channel:

$$\begin{aligned}
R_{max} &= \max_{x,y} R(x,y) \\
G_{max} &= \max_{x,y} G(x,y) \\
B_{max} &= \max_{x,y} B(x,y),
\end{aligned} \tag{34}$$

although it is usually preferred to lowpass the image first or to treat clusters of pixels, so as to avoid the problems caused by a few bright pixels that are outliers [53]. Next we compute the scaling factors:

$$\alpha = \frac{G_{max}}{R_{max}}, \quad \beta = \frac{G_{max}}{B_{max}}. \tag{35}$$

And finally, the correction is:

$$\begin{aligned}
R'(x,y) &= \alpha R(x,y) \\
G'(x,y) &= G(x,y) \\
B'(x,y) &= \beta B(x,y),
\end{aligned} \tag{36}$$

After the White Patch correction, the brightest point in the scene becomes achromatic:  $R'_{max} = G'_{max} = B'_{max}$ . Gray World and White Patch are very commonly used for AWB, sometimes combined, as reported in [30].

Once more we must point out that, as was the case with automatic exposure and automatic focus control, AWB is absolutely avoided by cinema professionals, both for technical and artistic reasons. Technically, AWB produces visible fluctuations in color in dynamic scenes, which is definitely something we don't want. Artistically, the choice of color palette of a shot is a fundamental aspect of the director's vision and therefore it is not something that one would like to automatize.

## XI. COLOR TRANSFORMATION

### A. The colorimetric matrix

At first glance, it would seem that for the camera to accurately capture colors matching our perception, the color triplets obtained by the camera sensor(s) should correspond to the cone responses of the human visual system. But this is never the case, as Figure 29 shows, and since the spectral sensitivities of sensor and cones are different, the responses must also be different. There are several reasons for this difference, like the fact that it is difficult to tune the spectral response of the pigments or dyes of the CFAs, and that having spectral sensitivities with a large amount of overlap (as in the responses of medium and long wavelength cones) would not be practical for the sensor from a signal-to-noise point of view [41]. But while emulating cone responses is not practical for image capture, it is essential in the subsequent processing of the image signal: the stimulus the scene would have produced in the human visual system must be estimated as accurately as possible [41]. This is why we must be able to transform the  $(R, G, B)$  values of the sensor into  $(X, Y, Z)$  tristimulus values, i.e. go from  $RGB$  into CIE  $XYZ$ , which we recall is a perceptually-based color space that uses the color-matching functions  $\bar{x}, \bar{y}, \bar{z}$  of a standard observer:

$$\begin{aligned} X &= \int_{380}^{780} \bar{x}(\lambda)L(\lambda)d\lambda \\ Y &= \int_{380}^{780} \bar{y}(\lambda)L(\lambda)d\lambda \\ Z &= \int_{380}^{780} \bar{z}(\lambda)L(\lambda)d\lambda, \end{aligned} \quad (37)$$

where  $L$  is the irradiance and the color-matching functions are a linear transformation of the cone sensitivities [42].

We can transform  $(R, G, B)$  into  $(X, Y, Z)$  by imposing the Luther-Ives condition [72]: that the sensor response curves are a linear combination of the color matching functions. We recall from Equation 24 that:

$$\begin{aligned} R &= \int_{380}^{780} r(\lambda)L(\lambda)d\lambda \\ G &= \int_{380}^{780} g(\lambda)L(\lambda)d\lambda \\ B &= \int_{380}^{780} b(\lambda)L(\lambda)d\lambda, \end{aligned} \quad (38)$$

where  $r, g, b$  are the camera spectral response functions for the red, green and blue channels, so the Luther-Ives condition can be stated as:

$$\begin{bmatrix} r(\lambda) \\ g(\lambda) \\ b(\lambda) \end{bmatrix} = \begin{bmatrix} b_{11} & b_{12} & b_{13} \\ b_{21} & b_{22} & b_{23} \\ b_{31} & b_{32} & b_{33} \end{bmatrix} \begin{bmatrix} \bar{x}(\lambda) \\ \bar{y}(\lambda) \\ \bar{z}(\lambda) \end{bmatrix}, \quad (39)$$

or, alternatively as:

$$\begin{bmatrix} \bar{x}(\lambda) \\ \bar{y}(\lambda) \\ \bar{z}(\lambda) \end{bmatrix} = \begin{bmatrix} a_{11} & a_{12} & a_{13} \\ a_{21} & a_{22} & a_{23} \\ a_{31} & a_{32} & a_{33} \end{bmatrix} \begin{bmatrix} r(\lambda) \\ g(\lambda) \\ b(\lambda) \end{bmatrix}. \quad (40)$$

As we just commented above, manufacturing processes and the properties of the materials used make it difficult to adjust at will the sensor response curves, and the Luther-Ives condition is usually not met in practice [42], but despite this fact a three-channel camera with three arbitrary sensor response curves is able to estimate the tristimulus values of an object as long as the object's spectral reflections are always composed of three principal components and they don't change steeply with respect to wavelength [42]. This implies that with a linear transformation we can go from the observed  $(R, G, B)$  triplet to its corresponding  $(X, Y, Z)$  tristimulus value, plugging Equations 40 into Equations 37:

$$\begin{bmatrix} X \\ Y \\ Z \end{bmatrix} = \begin{bmatrix} a_{11} & a_{12} & a_{13} \\ a_{21} & a_{22} & a_{23} \\ a_{31} & a_{32} & a_{33} \end{bmatrix} \begin{bmatrix} R \\ G \\ B \end{bmatrix} = A \begin{bmatrix} R \\ G \\ B \end{bmatrix}. \quad (41)$$

For each different triplet  $(R, G, B)$  we could have a different (and optimal) matrix  $A$ , but this would definitely be something very unpractical. A single colorimetric matrix  $A$  to be applied to all  $(R, G, B)$  colors can be computed in the following way [42]:

- 1) Build a set of  $n$  test patches of representative or important colors.
  - 2) Under controlled conditions, with a known illuminant (e.g. D65), measure the tristimulus values of the patches with a tristimulus colorimeter obtaining  $(X_i, Y_i, Z_i)$ ,  $1 \leq i \leq n$ .
  - 3) Under the same conditions, use the camera to measure the  $(R, G, B)$  values of the patches, obtaining  $(R_i, G_i, B_i)$ ,  $1 \leq i \leq n$ .
  - 4) A colorimetric matrix  $A$  gives an estimated tristimulus value  $(\hat{X}_i, \hat{Y}_i, \hat{Z}_i)$  from  $(R_i, G_i, B_i)$ .  $A$  is computed so as to minimize the total visual color difference  $J$ , which is a weighted sum of the color differences  $\Delta E$  (computed for instance in a CIE uniform color space) between the target tristimulus  $(X_i, Y_i, Z_i)$  and its estimate  $(\hat{X}_i, \hat{Y}_i, \hat{Z}_i)$ , for each patch  $i$ ,  $1 \leq i \leq n$ :
- $$J = \sum_{i=1}^n w_i \Delta E(X_i, Y_i, Z_i, \hat{X}_i, \hat{Y}_i, \hat{Z}_i),$$
- where  $w_i$  are the weights for the different patches.  $A$  can be obtained through least squares minimization, for instance.

In [70] it is noted that finding  $A$  by minimizing  $J$  as defined just above has the problem that the white point is not preserved, i.e. white in  $RGB$  is not mapped to white in the CIE  $XYZ$  color space; an additional term can be added to  $J$  in order to prevent this [35], and more accurate and robust techniques have also been proposed [24].

The colorimetric matrix is a function of the scene illuminant, so ideally a different matrix should be used for each different scene illuminant the camera is working with [65]. The above process finds the best colorimetric matrix (in terms of minimal error) for the calibration patch set under a given illuminant, and many cameras (most consumer models) use only this one matrix [24]. Some cameras come with several pre-set matrices computed under different illuminations. For instance, using the matrix for fluorescent lighting removes a noticeable green cast that would otherwise be present if we used a matrix computed with a standard illuminant like D65 or D50; other pre-sets may correspond for instance to a “film look” (with de-saturated colors), or may give a very vivid color palette. These pre-set matrices can also be adjusted manually so as to achieve a certain image look, since changing the colorimetric matrix affects hue and saturation (the white point is preserved, though, and color matrix adjustment must not be confused with white balance.)

Broadcast cameras were the first to incorporate the possibility of modifying the colorimetric matrix, so that multiple cameras in live broadcasts could be color matched and no color jumps appeared when switching from one camera to another [19].

### B. A note on color stabilization

We expect two pictures of the same scene, taken under the same illumination, to be consistent in terms of color. But if we have used different cameras to take the pictures, or just a single camera with automatic white balance (AWB) and/or automatic exposure (AE) correction, then the most common situation is that there are objects in the scene for which the color appearance is different in the two shots. Exactly the same happens in video: shots from two different cameras don't match in terms of color, and single-camera video with AE and/or AWB on will show noticeable changes in color whenever the camera motion makes the background luminance change [34].

This is problematic in many contexts. With a single camera, the only way to ensure that all pictures of the same scene are color consistent would be to save images in the RAW format, or to use the same set of manually fixed parameters for all the shots. These are not common choices for amateur users, but professional users face the same challenges: the most popular DSLR cameras for shooting HD video don't have the option of recording in RAW [29]; and while in cinema the exposure and color balance values are always kept constant for the duration of a take (i.e. AE and AWB are never used), the shooting conditions may require us to change these values from shot to shot. With different cameras the problem is aggravated, because using the same parameter values in all cameras is not enough to guarantee the stability of color across shots [56]. In many professional situations several cameras are used at the same time (e.g. large photo shoots, many mid-scale and most large-scale cinema productions), and in some cases the multi-camera set-up is required, not optional (TV broadcasts, 3D cinema).

The end result is that color discrepancies are unavoidable. We can expect most amateur video and stills to exhibit unpleasing color fluctuations, and in the TV and film industries much care and work is devoted to removing these color changes, both in production and post-production. For instance, TV broadcasts employ devices called Camera Control Units (CCU), operated by a technician called a Video Controller or Technical Director (TD): while each camera operator controls most of his/her camera functions such as framing or focus, the TD controls the color balance and shutter speed of a set of cameras so as to ensure color consistency across them, so for instance a 20-camera broadcast may have five CCU operators, each controlling four cameras [75]. In cinema and professional video, both for single and multi-camera shoots, color consistency is a key part of a post-production stage called color grading, and it's performed by seasoned technicians who manually tune parameters so as to match colors between frames [66]. In professional 3D movie production it is crucial that the twin cameras used in a stereo set-up perform in exactly the same way [56]; if there are discrepancies they are typically fixed in post-production by color-matching one view to the other, called the master view, which is taken as reference.

### C. Encoding the color values

Now that we have the  $(X, Y, Z)$  tristimulus values, which are an estimate of the stimulus the scene would have produced in the human visual system, we must convert them to a standard  $RGB$  color space so that the image can be displayed by a

device like a TV monitor, a computer screen or a digital projector. Within a given color imaging system, the encoding provides a link between the input obtained by the camera and the output of the display [39]. We can't just use the original  $(R, G, B)$  values and omit the conversion to  $(X, Y, Z)$ , because the output device has its own *RGB* color space which is not related to and does not match the color space of the camera. We'll briefly review these concepts in the following paragraphs.

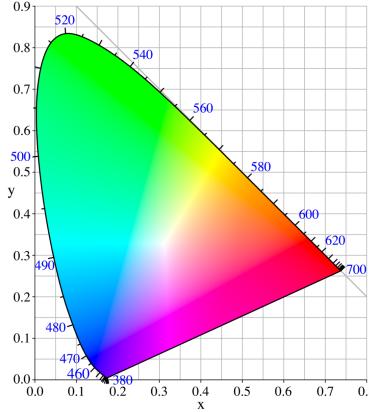


Fig. 30. The color gamut of the visible spectrum in CIE  $xy$  coordinates. Image from [7].

Recall that in CIE  $xy$  coordinates, monochromatic colors (corresponding to light of a single wavelength) lie on the boundary of the horseshoe-shaped region which corresponds to the visible spectrum. Colors that are mixtures, i.e. corresponding to light with a power spectrum that covers different wavelengths, lie inside this region. As a color becomes more mixed and its spectrum spreads, its  $xy$  location in the diagram moves inward, with achromatic colors in the center; see Figure 30.

In a trichromatic device like a TV display or a digital projector, colors are created as a linear combination of the three primaries of the device, the three particular primary colors it uses. Therefore, if we want to represent in CIE  $xy$  coordinates the color gamut of a device (the set of colors it can reproduce), it will be the triangle with vertices in the points  $(x_R, y_R), (x_G, y_G), (x_B, y_B)$  which are the  $xy$  coordinates of the *R*, *G* and *B* primaries. All colors reproducible by the device will lie inside this triangle, because they are a linear combination of  $(x_R, y_R), (x_G, y_G)$  and  $(x_B, y_B)$ . These primaries are (except for very special hardware) not monochromatic, because the physical characteristics of the materials with which these devices are manufactured make them have spectral sensitivities that are rather spread out. Therefore the primaries lie not on the boundary but inside the visible spectrum region, and the implication is clear: there are many colors that we could see but that the device isn't capable of reproducing. For instance, Figure 31 shows the color gamut of a CRT television set, where the primaries are given by the spectral characteristics of the light emitted by the red, green and blue phosphors used in CRT's.

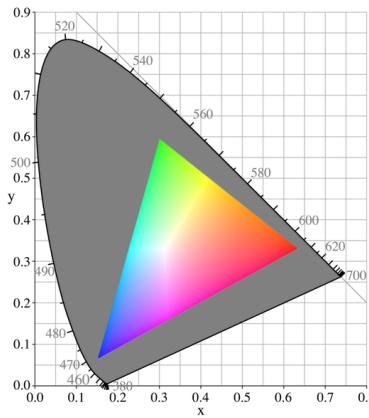


Fig. 31. The color gamut of a CRT television set. Image from [6].

In short, the primaries define the color space of the device. Figure 32 compares the color gamut of a specific model of digital cinema camera with that of print film, the DCI P3 standard for a digital projector (labeled DCI) and the BT.709 standard for HDTV.

The very wide gamut of the camera shown in Figure 32 is due to the dyes used in the CFA, while the same camera maker uses for another of its models a different set of dyes so that its color gamut is BT.709 [1], much smaller. An obvious question

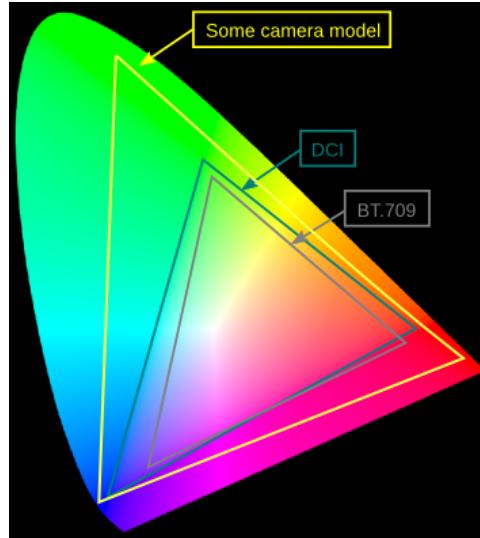


Fig. 32. The color gamut of a digital cinema camera model and of some standards.

then would be: why doesn't the maker use the wide-gamut CFA for all its camera models? The answer is that depending on the intended use and market of the camera, too much color space from the sensor may not be a good thing. If the camera will be used in broadcasting to be screened on regular monitors (which adhere to the BT.709 standard), then a wide-gamut camera is wasting processing and information space on colors that will never be reproduced; on the other hand, for digital cinema work the goal is not to have the camera be the limiting factor so that current and future display mediums will best be served by the cameras capabilities, in which case the widest gamut is desirable [1].

Furthermore, for a given intended output (e.g. HDTV) there are also differences in gamut for different technologies. For instance, Figure 33 shows the gamuts of TV monitors based on CRT, LCD and plasma technology.

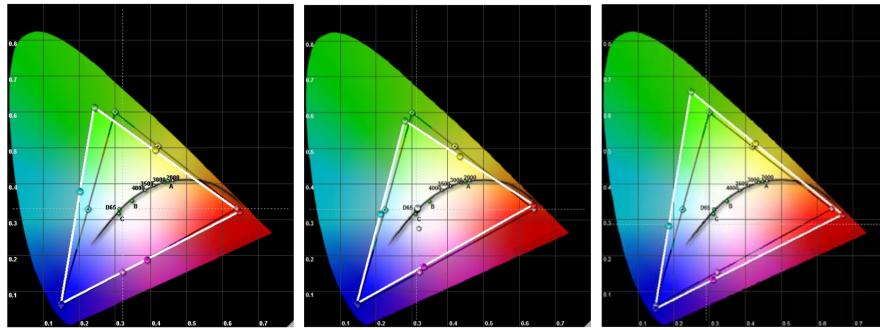


Fig. 33. The color gamut (as white triangles) of different TV display technologies, with reference to the BT.709 color space (black triangle). From left to right: CRT, LCD, plasma. Images by P.H. Putman [69].

This brings us back to the issue of how to encode the color information. Given that the primaries define the gamut and that they vary greatly among devices, it is essential that the camera uses a standard set of primaries for color encoding, so that each display device can adapt later on to these standard primaries. This is the usual arrangement in color encoding for all imaging systems, splitting the color processing into two parts, one for the input and another for the output so that each has its own associated transform, *to* and *from* a previously agreed upon color encoding specification [39].

The main possibilities then for primary sets are three, corresponding to the following gamuts:

- BT.709 (ITU-R Recommendation BT.709, also known as Rec 709). Agreed upon in 1990, it defines the HDTV standard. For technical reasons (compatibility and noise reduction) it uses primaries very close to the phosphor primaries of CRTs, and since newer display technologies are not based on CRTs it is now recognized that color correction at the receiver should compensate for the difference [72]. This is the standard used in broadcast and cable TV, DVD and Blu-ray discs.
- DCI P3 (defined in the Digital Cinema System Specification [14] and in SMPTE 431-2). This is a color gamut for digital cinema projectors, agreed upon in 2006. It is based on the gamut of Xenon lamps which are commonly used in digital projectors, approximating (and in some cases surpassing) the color gamut of film, and therefore it is wider than BT.709. Professional movies today are typically mastered in DCI P3 [68]. Figure 34 compares DCI P3 with BT.709.

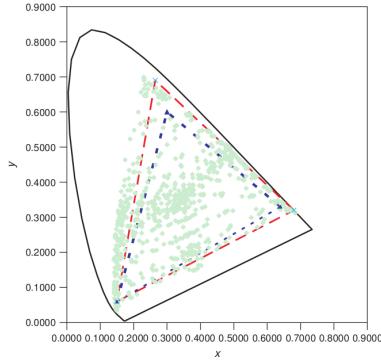


Fig. 34. Comparison of color gamut of film (pastel green dots), DCI P3 (dashed red line), BT.709 (dashed blue line). Image from [51].

- DCI  $X'Y'Z'$  (defined in the Digital Cinema System Specification [14]). Digital Cinema Initiative (DCI) is a consortium established by major motion picture studios, formed to develop a standard architecture for digital cinema systems. After considering several options for the primaries, such as a wider-gamut RGB space that encompassed all film colors or a parametric RGB encoding with the primaries of the projector as metadata, it was pointed out that rather than argue over which set of wide gamut primaries to use, DCI could just adopt the widest gamut set: the CIE  $XYZ$  primaries [51]. These primaries are the basis of a widely used international standard and met all DCI requirements, specifically providing no limits on future improvements since they enclose all visible colors. Since these primaries fall outside the visible spectrum they are not real but virtual primaries. Several movies have been encoded and released in this format [51]. Figure 35 compares DCI  $X'Y'Z'$  with DCI P3.

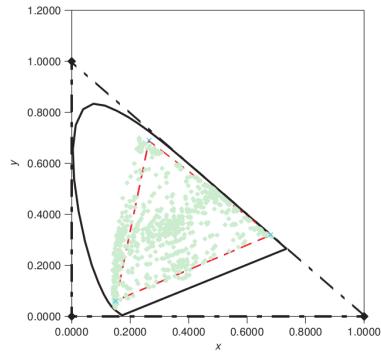


Fig. 35. Comparison of color gamut of film (pastel green dots), DCI P3 (dashed red line), DCI  $X'Y'Z'$  (dashed black line). Image from [51].

Given the chromaticity coordinates  $(x_R, y_R), (x_G, y_G), (x_B, y_B)$  of the primaries of an RGB system, and also the  $(X_w, Y_w, Z_w)$  value of its white point, the camera converts the  $XYZ$  tristimulus values computed in the previous section to output  $RGB$  values (compliant with BT.709 or DCI P3, for example) by multiplying by a  $3 \times 3$  matrix [13]:

$$\begin{bmatrix} R \\ G \\ B \end{bmatrix} = M^{-1} \begin{bmatrix} X \\ Y \\ Z \end{bmatrix}, \quad (42)$$

where:

$$M = \begin{bmatrix} S_r X_r & S_g X_g & S_b X_b \\ S_r Y_r & S_g Y_g & S_b Y_b \\ S_r Z_r & S_g Z_g & S_b Z_b \end{bmatrix} \quad (43)$$

$$\begin{aligned} X_r &= \frac{x_r}{y_r} & ; X_g &= \frac{x_g}{y_g} & ; X_b &= \frac{x_b}{y_b} \\ Y_r &= Y_g = Y_b = 1 & & & & \\ Z_r &= \frac{1 - x_r - y_r}{y_r} & ; Z_g &= \frac{1 - x_g - y_g}{y_g} & ; Z_b &= \frac{1 - x_b - y_b}{y_b} \end{aligned} \quad (44)$$

$$\begin{bmatrix} S_r \\ S_g \\ S_b \end{bmatrix} = \begin{bmatrix} X_r & X_g & X_b \\ Y_r & Y_g & Y_b \\ Z_r & Z_g & Z_b \end{bmatrix} \begin{bmatrix} X_w \\ Y_w \\ Z_w \end{bmatrix}. \quad (45)$$

If any of the obtained  $(R, G, B)$  values fall outside the  $[0, 1]$  interval they are normally *clipped* by the camera; this imposes on the color encoded images the color gamut determined by the primaries  $(x_R, y_R), (x_G, y_G), (x_B, y_B)$ , and it's a lossy operation which can't be undone. Therefore it is preferred that the camera performs color encoding in the widest gamut it is capable of, minimizing information loss. Then after post-production this gamut can be adapted to different output gamuts in a color correction stage performed by a skilled colorist, who is sensitive to the creative intent of the movie and who is under the supervision of the director and/or cinematographer [68]; this last stage of color correction, just prior to movie distribution, usually involves 3D LUTs [20]. In the academic literature the problem of adapting color gamuts is called *gamut mapping* [61].

## XII. GAMMA CORRECTION AND QUANTIZATION

### A. The need for gamma correction

At the onset of broadcast television it was observed that CRTs produce luminance as a non-linear, *power* function of the device's voltage input:  $L = \alpha V^\gamma$ , where  $L$  is the luminance,  $V$  the voltage and  $\gamma$  the exponent of this power function, which has a value of around 2.5 ( $\alpha$  is just a proportionality coefficient). But the luminance signal captured by the camera is linearly proportional to light intensity. Therefore, for correct luminance reproduction on a TV set, the camera's luminance signal must be non-linearly scaled with a power function with the inverse exponent of the CRT's power function, thus both non-linearities cancel each other out and the luminance of the CRT can be a faithful (scaled) representation of the luminance reaching the camera:  $V = \beta L'^{\frac{1}{\gamma}}$ , where  $L'$  is the camera's luminance ( $\beta$  is just a proportionality coefficient). This process is called *gamma correction*. See Figure 36.

At that time it was also very well known that humans have a perceptual response to luminance that is also non-linear: perceived lightness is roughly the 0.42 power of physical luminance [67]; see Figure 37. This means that differences in the dark parts of an image are more noticeable than differences *of the same amount* on bright parts of the same image. Given that over-the-air analog transmission introduced noise in the TV signal, a simple linear transmission of the luminance values captured by the camera would make this noise much more apparent in the darkest regions of the image. Therefore, non-linearly scaling the camera's luminance with a power function of exponent 0.42 makes this noise less perceptible.

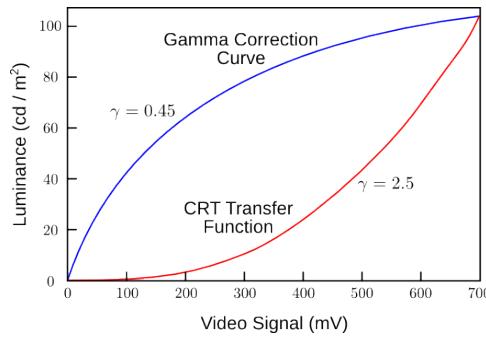


Fig. 36. Gamma correction on a CRT.

The amazing coincidence, as Charles Poynton puts it in [67], is that the CRT voltage-to-luminance function is very nearly the inverse of the perceptual luminance-to-lightness relationship, i.e. the lightness perception curve of Figure 37 is very similar to the power function of exponent  $\frac{1}{\gamma}$  that compensates for the CRTs nonlinearity (see Figure 36), a fact already recognized in 1939 [68]. In the early days of TV, then, gamma correction was essential for two tasks: compensating for the CRT nonlinear response, and reducing noise through perceptual encoding, which implies that even if the CRT response had been linear the gamma correction process would have had to be performed in the same way.

But why is gamma correction still used today, when CRTs have become obsolete and air transmission noise is no longer a major concern with the advent of digital TV?

The main reason is that since digital signals have a limited number of bits to code each pixel value, if we quantize a gamma corrected signal the quantization intervals are wider at higher luminance values, where changes are less perceptible. In other words, gamma correction allows us to use more bits at the darkest regions, where we are more sensitive to differences, and less bits at the brightest regions, where we are less sensitive to differences; this is what we mean by perceptual coding. For a fixed number of bits, then, perceptual coding permits us to maximize image quality (or, conversely, without perceptual coding we would have to use more bits to represent images with the same perceptual quality).

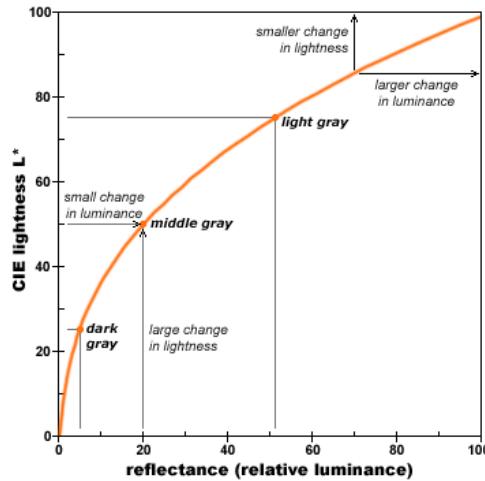


Fig. 37. Lightness perception as a function of luminance is approximately a power function of exponent 0.42. Image from [15].

There is another, very important reason, which is the issue of color appearance. As explained in [68], we can't just aim for the reproduced image to have values proportional to those in the original scene, because the appearance of the images is modified by the environment in which they are seen by the viewer; this environment is usually different from that of the original scene, typical displays have lower luminance and contrast than typical scenes, and their surrounds are darker. The differences in environment produce three main effects that must be compensated [68]:

- The Hunt effect: colorfulness decreases as illumination decreases. If an image is taken in daylight and linearly displayed in a dim environment, the image will look as if it was captured at twilight. See Figure 38.
- The Stevens effect: contrast decreases as illumination decreases, i.e. dark colors look lighter and light colors look darker. See Figure 38.

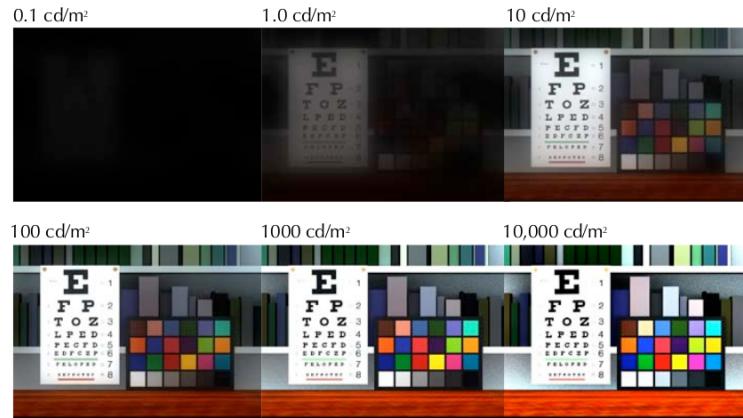


Fig. 38. Hunt and Stevens effects. Image from [31].

- The simultaneous contrast effect: the center squares in Figure 39 have the same shade of gray, but the one surrounded by dark seems lighter than the other center square.



Fig. 39. The simultaneous contrast effect.

Experience shows that all three appearance effects can be ameliorated at the same time by imposing a modest end-to-end power function: rather than encoding with a power  $\frac{1}{\gamma}$  and decoding with  $\gamma$ , different gamma values are used for encoding ( $\gamma_E$ ) and decoding ( $\gamma_D$ ) so that the net effect is a power function with an exponent slightly greater than one,  $\frac{\gamma_D}{\gamma_E} \simeq 1.2$  [68].

For the aforementioned reasons, cameras normally apply gamma correction, which is performed after the color correction stage (some high-end digital cameras allow recording of a linear output, without gamma correction). The actual value of gamma that is used depends on the intended viewing conditions. If using the BT.709 color encoding, intended for HDTV viewing on low contrast displays in dim conditions, the encoding gamma is equivalent to  $0.45 = \frac{1}{2.2}$ . If using DCI P3 or DCI  $X'Y'Z'$  color encoding, intended for viewing in a cinema with dark surround and high contrast, the encoding gamma is  $\frac{1}{2.6}$ . In the BT.709 case the end-to-end coefficient of the power function is 1.2, while in the digital cinema case it is 1.5 [68].

It must be noted though that cameras also use power-laws to enhance contrast and achieve pleasant-looking images, aside from the gamma-correction inherent to the standard [52]. The end result is that the actual gamma value of any recorded shot may not be the gamma value of the standard the shot has been recorded in; if this value is needed (e.g. to perform linearization previous to color correction) it needs to be estimated.

### B. Transfer function and quantization

In order to implement gamma correction a transfer function is used. For instance, in BT.709, each R, G and B component is processed with the following transfer function, obtaining a gamma-corrected value  $R'$ ,  $G'$  or  $B'$ :

$$V' = \begin{cases} 4.5L, & 0 \leq L \leq 0.018 \\ 1.099L^{0.45} - 0.099, & 0.018 \leq L \leq 1 \end{cases} \quad (46)$$

where  $L$  denotes R, G or B and  $V'$  the corresponding gamma-corrected value. Notice how this is not a pure power function, it has a linear segment near black, with a limited slope. The reason for this is practical: the slope of a pure power function whose exponent is less than one is infinite at zero, therefore a transfer function that was a pure power function would have infinite gain near black [67]. Poynton [67] points out that because of the linear segment introduced for low values, the overall transfer function is very similar to a square root ( $\gamma_E \simeq 0.5$ ), hence it is not accurate to describe BT.709 as having  $\gamma_E = 0.45$ .

The values in Equation 46 are in the range  $[0, 1]$ . For quantization in 8 bits (the most common case), they are scaled by 219, offset by 16 and rounded to nearest integer; values below 16 and above 235 are reserved [67] and provide range for filter over and under shoots. Digital video standards were developed as an addition to analog broadcast standards and were required to be backward-compatible with existing analog video equipment, hence providing ample headroom for analog variations was a prudent practice to avoid clipping artifacts [51]. For quantization in 10 bits, scaling and offset values are multiplied by 4. Digital still cameras encode pictures in sRGB, which utilizes the full range 0 – 255.

For digital cinema, the DCI specification [14] states that the transfer function for the  $XYZ$  tristimulus values must be calculated with a normalizing constant of 52.37 (corresponding to an absolute luminance of  $52.37cd/m^2$ ) and a gamma value of 2.6:

$$CV_{X'} = \text{round} \left( 4095 \times \left( \frac{X}{52.37} \right)^{\frac{1}{2.6}} \right) \quad (47)$$

$$CV_{Y'} = \text{round} \left( 4095 \times \left( \frac{Y}{52.37} \right)^{\frac{1}{2.6}} \right) \quad (48)$$

$$CV_{Z'} = \text{round} \left( 4095 \times \left( \frac{Z}{52.37} \right)^{\frac{1}{2.6}} \right) \quad (49)$$

The DCI specification establishes quantization in 12 bits per component, following the results from the experiments in [27]. These experiments were performed in cinema viewing conditions and demonstrated that 12 is the minimum number of bits necessary for observers not to perceive quantization artifacts. Unlike with BT.709, the full range is used and there are no illegal code values in digital cinema mastering and distribution, since they are both purely digital processes that therefore don't require us to provide headroom for analog variations [51]. Nevertheless, when digital systems use the SMPTE 372 dual-link high definition serial interface (HD-SDI) they are subject to its reserved code values, which are 0 – 15 and 4080 – 4095: values below 15 or above 4080 are clipped, although in practice this is visually unnoticeable [51].

### C. Color correction pipeline

We can summarize all the above elements of the color processing chain as follows [23]:

$$\begin{bmatrix} R \\ G \\ B \end{bmatrix}_{out} = \left( \alpha \begin{bmatrix} c_{11} & c_{12} & c_{13} \\ c_{21} & c_{22} & c_{23} \\ c_{31} & c_{32} & c_{33} \end{bmatrix} \begin{bmatrix} r_{AWB} & 0 & 0 \\ 0 & g_{AWB} & 0 \\ 0 & 0 & b_{AWB} \end{bmatrix} \begin{bmatrix} R \\ G \\ B \end{bmatrix}_{in} \right)^\gamma \quad (50)$$

where  $RGB_{in}$  is the camera raw triplet, to which a diagonal white balance matrix is applied, followed by the matrix  $[c_{ij}]$  which cascades the colorimetric matrix of Equation 41 with the color encoding matrix of Equation 42, a gain factor  $\alpha$  and finally a power function of exponent  $\gamma$  is applied.  $RGB_{out}$  is the output of this color correction pipeline but not the actual triplet value recorded by the camera, because as we shall see there still remain some image processing operations in the full camera pipeline, e.g. edge enhancement and video compression, which will alter the final values.

In [65] it is pointed out that the  $RGB_{in}$  camera raw triplet is usually not the original sensor signal but a corrected version of it, where the original nonlinear camera exposures have been linearized through a LUT; thus, we can assume that all  $RGB_{in}$  triplets are actually proportional to the exposures. The final gamma correction power function is also implemented with a LUT, and since the color transformation matrices can be cascaded into a single 3x3 matrix, the whole color processing pipeline can be expressed as a LUT-matrix-LUT sequence [65].

### XIII. EDGE ENHANCEMENT

The optical properties of the lens system determine its MTF, which expresses contrast as a function of spatial frequency. The lens system produces image blurring, making the MTF decrease faster, but so do the optical anti-aliasing filter and the sensor aperture. The result is that the sharpness of the image is reduced, because perceived sharpness can be approximately quantified as the area under the MTF curve squared, the “equivalent line number” proposed in 1948 by Otto Schade [12]. In order to compensate for this loss of contrast, many cameras incorporate an edge enhancement process, also called sharpening or “unsharp masking.” Because of perceptual considerations already discussed, the amount of sharpening to be performed depends on the viewing conditions (mainly screen size and viewing distance, but also the luminance of the surround). Because these are variable, digital camera edge enhancement is usually optimized for a selected and conservative viewing condition [18]. As with other automated processes mentioned before in this chapter (e.g. AWB), cinematographers prefer that the camera doesn’t perform edge enhancement and instead prefer to boost contrast in post-production where there is much more control over the process and a wide range of possible algorithms to choose from.

Basic edge enhancement is linear: an edge map  $E$  is computed from the image  $I$ , scaled and added back to  $I$  to obtain the sharpened image  $I'$  [18]:

$$I' = I + kE. \quad (51)$$

A common way to compute the edge map is through an “unsharp mask”: the image  $I$  is blurred by convolution with a Gaussian  $g$ , then subtracted from the original  $I$  to obtain  $E$  [18]:

$$E = I - g * I. \quad (52)$$

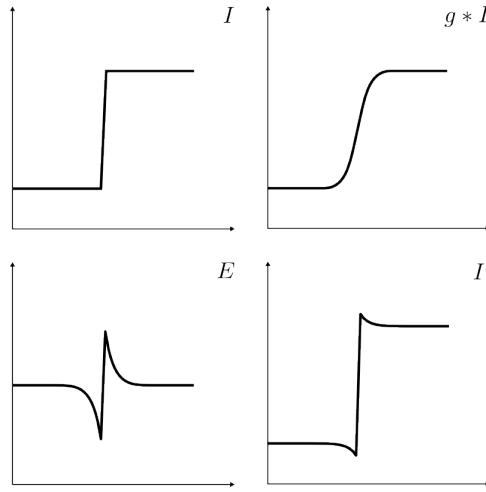


Fig. 40. Linear edge enhancement (unsharp masking). Top left: edge. Top right: after Gaussian blur. Bottom left: edge map. Bottom right: sharpened result.

The term “unsharp masking” comes from film photography and originally referred to the use of an unsharp (blurred) positive film mask made from the negative: when this mask was contact printed in register with the negative, the contact print would

accentuate the edges present in the image, resulting in an image with a crisper and sharper look [2]. We can see that this process, translated into digital terms, is exactly the same as what Equations 51 and 52 are doing. Figure 40 shows, on the top left, an edge plot from the original image ( $I$ ); on the top right, the image after blurring ( $g * I$ ); on the bottom left, the edge map ( $E$ ); and on the bottom right, the sharpened result ( $I'$ ). This figure shows how the resulting edge has been enhanced, but it is also clear that if the scale parameter  $k$  in Equation 51 is too large the enhancement's over and undershoots may be perceived as halos, which is definitely something to be avoided. Better results can be obtained with nonlinear edge enhancement. It consists of applying to the edge map  $E$  a nonlinear function such as the one depicted in Figure 41, for instance through a simple LUT operation; this soft-thresholding operation reduces noise by eliminating edge enhancement for small values, and reduces halos by limiting large edge values [18].

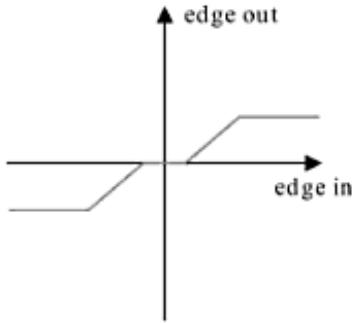


Fig. 41. Nonlinear edge enhancement function. Figure from [18].

Though it would seem preferable to perform edge enhancement in a color space that is linear with exposure, since many optical degradations such as optical blur occur in linear space, experience shows that it is actually better to do the sharpening on a color space that is uniform with visual perception, which is why in-camera edge enhancement is typically performed after gamma correction [65]. Also, edge enhancement is often applied to the luma channel (the Y channel of the YCbCr space) instead of the RGB color channels, so as to avoid amplification of colored edges created by chromatic aberration or other artifacts earlier in the image processing pipeline [18].

#### XIV. OUTPUT FORMATS

Recording images of 2880x1620 pixels at 30 frames per second and with 10 bits of color information per channel requires a data rate of approximately 4 gigabits per second. This figure is beyond the capabilities of todays' cameras, which must therefore resort to compression before recording in-camera; a few models, like the ARRI Alexa, allow for uncompressed recording but this must be performed off-camera and requires special recording equipment.

##### A. Compression

The compression may be lossless, in which images can be recovered without any information loss, or lossy, with information loss increasing with the compression rate.

Lossless compression is based on Huffman coding and run-length encoding and is normally used for raw data, as in the CinemaDNG format, based on the TIFF/EP format with which raw images are recorded in several models of still cameras.

For lossy compression there are many alternatives. For starters, the color coding can be in RGB or YCbCr. In YCbCr there are different possibilities of color subsampling, noted thus:

- 4:4:4, no subsampling. This is the preferred choice for digital cinema.
- 4:2:2, reduction of the chroma components by a factor of 2 in the horizontal direction. Common in HDTV cameras for broadcast.
- 4:2:0, reduction of the chroma components by a factor of 2 in both the horizontal and vertical directions. Common in reflex still cameras capable of shooting HD video [29].

Next, there is the possibility of using intra-frame coding or predictive coding.

When each frame is compressed independently from the others, treating them as standalone images, then all frames are said to be of type I for “intra-coded,” coded by intra-frame techniques. This is the approach used in several formats such as ProRes, DNxHD, and REDCODE.

Movies have a very high time redundancy, so if we want to increase the compression rate we may want to code frame differences instead of the frames themselves: for slowly varying sequences, the differences among motion-compensated images will be small and hence can be coded using less bits. This sort of coding, where only “new” information is processed, is called

predictive coding. Frames encoding (motion-compensated) differences with another frame are called type P for “predictive-coded,” or B for “bi-directionally predictive-coded.” The image sequence is partitioned into groups of pictures, GOPs, which may last from 1 to 15 pictures or more. Each GOP is independent from the others: all P and B frames in a GOP code motion-compensated differences with other I and P frames in the same GOP. A GOP always starts with an I frame, so if a codec only uses one-frame GOPs it is in fact coding all frames with intra-coding and all frames are independent. Videos in which GOPs have more than one picture are called Long-GOP. Very popular formats like MPEG-1, MPEG-2 and MPEG-4/AVC H.264 use predictive coding, i.e. their frames may be of type I, P or B. If a video in one of these formats uses Long-GOP it can't be edited directly due to the presence of prediction-coded frames P or B; it must be converted to a format suitable for editing in a process called “rushes upgrade,” going for instance from H.264 to ProRes 422 [29].

Finally, there's the actual lossy compression technique used, which can be based either on the DCT (discrete cosine transform) or on the wavelet transform. The DCT is the basis for MPEG-1, MPEG-2, MPEG-4/AVC H.264, DNxHD and ProRes (and for the JPEG format for still images). The wavelet transform is the basis for REDCODE and DCI  $X'Y'Z'$  [14], both of which use only I frames coded with variants of JPEG2000 (a wavelet-based compression standard for still images). Thresholding the transform coefficients is a typical denoising technique known as *coring* [70]: coefficients below the threshold are assumed to correspond to noise and set to zero. This is a basic denoising technique.

### B. Recording in RAW

Some cameras allow for recording in a raw image format, where the data is stored as it comes from the sensor, without the usual in-camera image processing described in the precedent sections: no demosaicking, color balance, color encoding, sharpening, denoising, etc. Without those operations the image is not ready for display, which is what the term *raw* implies. But RAW images present many clear advantages, allowing for the best image quality in digital cinema:

- The original image signal, as captured by the sensor, is recorded, preserving the native 12 bit or 14 bit depth without quantization.
- The original color gamut is also preserved and therefore the output color space can be chosen at will.
- With a RAW image the cinematographer has total freedom in the particular choice of image processing algorithms to be applied in postproduction: for instance, he/she may choose the best demosaicking technique, a custom non-local white balance algorithm, an image-based gamut mapping method, a state of the art non-local denoising algorithm, etc. None of these options are available in current digital cinema cameras, not even high-end professional models.
- Potentially problematic in-camera image processing, such as denoising and sharpening, is avoided.
- Compression artifacts are avoided as well: raw image formats are uncompressed (like ArriRaw), or use lossless compression (like CinemaDNG), or high quality lossy compression (like REDCODE).
- The dynamic range is better preserved than in formats where images are already tone-mapped, gamma-corrected and quantized.

## XV. ADDITIONAL IMAGE PROCESSING

### A. Lens spatial distortion correction

Optical aberrations such as pincushion and barrel distortions can be corrected through in-camera image processing. The distortion amount must be estimated for a given lens, and approximated by a function of the image height from the lens center [71]. See Figure 42.

### B. Lens shading correction

Many lens systems produce a vignetting effect on the image, causing the periphery to be darker than its center. The compensation algorithm consists then of amplifying each pixel value depending on the pixel's distance to the image center, and it can be easily implemented with a LUT operation [71]; see Figure 43. Some camera models have tens of precomputed correction LUTs for different lens systems and even allow to manually add some more [29].

## XVI. THE ORDER OF THE STAGES OF THE IMAGE PROCESSING PIPELINE

The stages discussed in the preceding sections appear in different order in different pipelines. In [50], Kao et al. propose an image processing pipeline and justify the ordering of its stages based on several considerations:

- 1) White balance should be performed early in the pipeline, because later operations depend on a correct relationship/ratio among channels.
- 2) Demosaicking should also precede other image processing operations where tri-stimulus values are needed.
- 3) Noise is produced in RGB space, so it's there where it should be addressed, not in YCbCr, where many denoising algorithms work (they do so because the human visual system is more sensitive to noise in the luminance channel than in the chroma ones). Also, noise reduction should come before color correction, because this latter process may enlarge

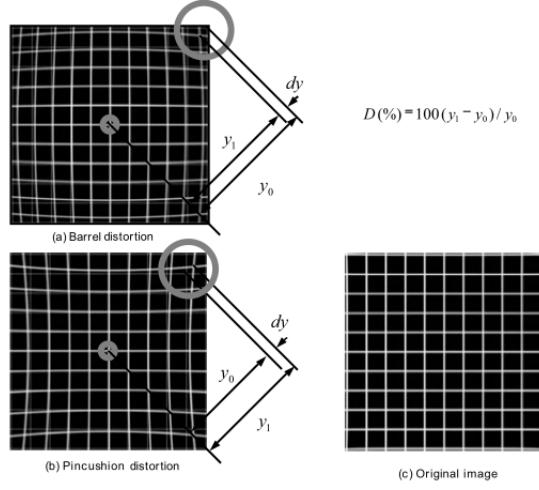


Fig. 42. Lens spatial distortion. Figure from [71].

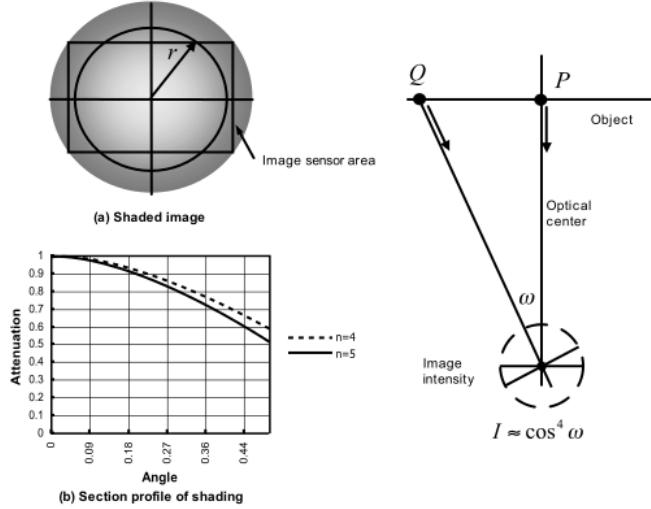


Fig. 43. Lens shading distortion. Figure from [71].

random noise. And it would be best if the denoising procedure would take into account an estimation of image edges so as not to blur them and decrease contrast.

- 4) Gamma correction should come last in the pipeline because of the non-linear nature of this operation, which makes it difficult to preserve chroma values afterwards.

On the other hand, Weerasinghe et al. [76] propose a pipeline where color correction precedes demosaicking, justifying this choice with an increase in SNR of the output. In fact, in [22] Bazhyna points out that several algorithms for image processing before demosaicking have recently been introduced that deal with digital zooming, denoising, deblurring and sharpening, among other tasks.

## REFERENCES

- [1] <http://blog.abelcine.com/2012/11/01/sonys-pmw-f5-and-f55-defining-cfa/>.
- [2] [http://dpanswers.com/content/tech\\_defects.php](http://dpanswers.com/content/tech_defects.php).
- [3] [http://en.wikipedia.org/wiki/Black-body\\_radiation](http://en.wikipedia.org/wiki/Black-body_radiation).
- [4] [http://en.wikipedia.org/wiki/Checker\\_shadow\\_illusion](http://en.wikipedia.org/wiki/Checker_shadow_illusion).
- [5] <http://en.wikipedia.org/wiki/CIELUV>.
- [6] [http://en.wikipedia.org/wiki/Color\\_gamut](http://en.wikipedia.org/wiki/Color_gamut).
- [7] [http://en.wikipedia.org/wiki/Color\\_matching\\_function#Color\\_matching\\_functions](http://en.wikipedia.org/wiki/Color_matching_function#Color_matching_functions).
- [8] [http://en.wikipedia.org/wiki/Dichroic\\_prism](http://en.wikipedia.org/wiki/Dichroic_prism).
- [9] [http://en.wikipedia.org/wiki/Spectral\\_sensitivity](http://en.wikipedia.org/wiki/Spectral_sensitivity).
- [10] [http://en.wikipedia.org/wiki/Standard\\_illuminant](http://en.wikipedia.org/wiki/Standard_illuminant).
- [11] <http://pages.videotron.com/graxx/CIE2.html>.
- [12] <http://www.adamwilt.com/TechDiff's/Sharpness.html>.
- [13] [http://www.brucelindbloom.com/index.html?Eqn\\_RGB\\_XYZ\\_Matrix.html](http://www.brucelindbloom.com/index.html?Eqn_RGB_XYZ_Matrix.html).
- [14] <http://www.dcimovies.com/specification/index.html>.
- [15] <http://www.handprint.com/HP/WCL/color2.html>.
- [16] <http://www.red.com/learn/red-101/resolution-aliasing-motion-capture>.
- [17] J. Adams, K. Parulski, and K. Spaulding. Color processing in digital cameras. *Micro, IEEE*, 18(6):20–30, 1998.
- [18] J.E. Adams Jr, A.T. Deever, E.O. Morales, and B.H. Pillman. Perceptually based image processing algorithm design. *Perceptual Digital Imaging: Methods and Applications*, 6:123, 2012.
- [19] G. Adcock. Charting your camera. *Creative COW Magazine*, November 2011.
- [20] D. Bankston. The color-space conundrum. *American Cinematographer*, 86(4), 2005.
- [21] S. Battiato, G. Messina, and A. Castorina. *Single-Sensor Imaging: Methods and Applications for Digital Cameras*, chapter Exposure correction for imaging devices: an overview, pages 323–349. CRC Press, 2008.
- [22] A. Bazhyna. Image compression in digital cameras. *Tampereen teknillinen yliopisto, Julkaisu-Tampere University of Technology*, 2009.
- [23] S. Bianco, A. Bruna, F. Naccari, and R. Schettini. Color space transformations for digital photography exploiting information about the illuminant estimation process. *Journal of the Optical Society of America A*, 29(3):374–384, 2012.
- [24] S. Bianco, F. Gasparini, A. Russo, and R. Schettini. A new method for RGB to XYZ transformation based on pattern search optimization. *IEEE Transactions on Consumer Electronics*, 53(3):1020–1028, 2007.
- [25] G. Buchsbaum. A spatial processor model for object colour perception. *Journal of the Franklin Institute*, 310:337–350, 1980.
- [26] J. Coghill. Digital imaging technology 101. Technical report, DALSA, 2003.
- [27] M. Cowan, G. Kennel, T. Maier, and B. Walker. Contrast sensitivity experiment to determine the bit depth for digital cinema. *SMPTE motion imaging journal*, 113(9):281–292, 2004.
- [28] S. Cvetkovic, H. Jellema, et al. Automatic level control for video cameras towards HDR techniques. *Journal on Image and Video Processing*, 2010:13, 2010.
- [29] S. Devaud. *Turner en vidéo HD avec les reflex Canon: EOS 5D Mark II, EOS 7D, EOS 1D Mark IV*. Eyrolles, 2010.
- [30] M. Ebner. *Color constancy*. Wiley, 2007.
- [31] M.D. Fairchild. Color appearance models: CIECAM02 and beyond. In *tutorial notes, IS&T/SID 12th Color Imaging Conference*, 2004.
- [32] M.D. Fairchild. *Color appearance models*. J. Wiley, 2005.
- [33] H.S. Fairman, M.H. Brill, and H. Hemmendinger. How the CIE 1931 color-matching functions were derived from Wright-Guild data. *Color Research and Application*, 22(1):11–23, 1997.
- [34] Z. Farbman and D. Lischinski. Tonal stabilization of video. *ACM Transactions on Graphics (TOG)*, 30(4):89, 2011.
- [35] G. Finlayson and M.S. Drew. Constrained least-squares regression in color spaces. *Journal of Electronic Imaging*, 6(4):484–493, 1997.
- [36] G. Finlayson, M.S. Drew, and B.V. Funt. Diagonal transforms suffice for color constancy. In *Proceedings of Fourth International Conference on Computer Vision*, pages 164–171. IEEE, 1993.
- [37] G. 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*, 23(11):1209–1221, 2001.
- [38] J.P. Frisby and J.V. Stone. *Seeing: the computational approach to biological vision*. The MIT Press, 2010.
- [39] E.J. Giorgianni, T.E. Madden, and K.E. Spaulding. Color management for digital imaging systems. *Digital color imaging handbook*, 2003.
- [40] D.H. Hubel. *Eye, Brain, and Vision*. Scientific American Library, 1995.
- [41] P.M. Hubel, J. Holm, G.D. Finlayson, M.S. Drew, et al. Matrix calculations for digital photography. In *The Fifth Color Imaging Conference: Color, Science, Systems and Applications*, 1997.
- [42] P. Hung. *Image Sensors and Signal Processing for Digital Still Cameras*, chapter Color theory and its application to digital still cameras, pages 205–221. CRC, 2005.
- [43] J. Janesick. Dueling detectors. CMOS or CCD? *SPIE OE magazine*, 41:30–33, 2002.
- [44] J. Jeon, J. Lee, and J. Paik. Robust focus measure for unsupervised auto-focusing based on optimum discrete cosine transform coefficients. *IEEE Transactions on Consumer Electronics*, 57(1):1–5, 2011.
- [45] D.B. Judd. Hue, saturation, and lightness of surface colors with chromatic illumination. *Journal of the Optical Society of America*, 30(1):2–32, 1940.
- [46] D.B. Judd. *Contributions to color science*, volume 545, chapter Color appearance, pages 539–564. National Bureau of Standards (USA), 1979.
- [47] D.B. Judd. *Contributions to color science*, volume 545, chapter Estimation of chromaticity differences and nearest color temperature on the standard 1931 ICI colorimetric coordinate system, pages 207–212. National Bureau of Standards (USA), 1979.
- [48] D.B. Judd. *Contributions to color science*, volume 545, chapter The unsolved problem of color perception, pages 516–522. National Bureau of Standards (USA), 1979.
- [49] D.B. Judd. *Contributions to color science*, volume 545, chapter Appraisal of Land’s work on two-primary color projections, pages 471–486. National Bureau of Standards (USA), 1979.
- [50] W.C. Kao, S.H. Wang, L.Y. Chen, and S.Y. Lin. Design considerations of color image processing pipeline for digital cameras. *IEEE Transactions on Consumer Electronics*, 52(4):1144–1152, 2006.
- [51] G. Kennel. *Color and Mastering for Digital Cinema (Digital Cinema Industry Handbook Series)*. Taylor and Francis US, 2006.
- [52] S. Kim, H. Lin, Z. Lu, S. Susstrunk, S. Lin, and M. Brown. A new in-camera imaging model for color computer vision and its application. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 2012.
- [53] E.Y. Lam and G.S.K. Fung. *Single-sensor imaging: Methods and applications for digital cameras*, chapter Automatic white balancing in digital photography, pages 267–294. CRC Press, 2008.
- [54] E.H. Land. The Retinex theory of color vision. *Scientific American*, 237:108–128, 1977.
- [55] J. McCann, C. Parraman, and A. Rizzi. Reflectance, illumination and edges. *Proceedings Color and Imaging Conference (CIC)*, 201(17):2–7, 2009.
- [56] B. Mendiburu. *3D movie making: stereoscopic digital cinema from script to screen*. Focal Press, 2009.
- [57] J.D. Mollon. *The science of color*, chapter The origins of modern color science, pages 1–39. Optical Society of America Oxford, 2003.

- [58] P. Monnier. Standard definitions of chromatic induction fail to describe induction with s-cone patterned backgrounds. *Vision Research*, 48(27):2708 – 2714, 2008.
- [59] A. Morimura, K. Uomori, Y. Kitamura, A. Fujioka, J. Harada, S. Iwamura, and M. Hirota. A digital video camera system. *IEEE Transactions on Consumer Electronics*, 36(4):3866–3876, 1990.
- [60] N. Moroney, M.D. Fairchild, R. Hunt, C. Li, M.R. Luo, and T. Newman. The CIECAM02 color appearance model. In *Proceedings of Color and Imaging Conference*. Society for Imaging Science and Technology, 2002.
- [61] J. Morovič. *Color gamut mapping*. John Wiley & Sons, 2008.
- [62] J. Nakamura. *Image sensors and signal processing for digital still cameras [Chapter 3]*. CRC, 2005.
- [63] J. Nakamura. *Image sensors and signal processing for digital still cameras [Chapter 4]*. CRC, 2005.
- [64] O. Packer and D.R. Williams. *The Science of Color*, chapter Light, the retinal image, and photoreceptors, pages 41–102. Elsevier Science Ltd, 2003.
- [65] K. Parulski and K. Spaulding. *Digital Color Imaging Handbook*, chapter Color image processing for digital cameras, pages 727–757. Boca Raton, FL: CRC Press, 2003.
- [66] F. Pitie, A.C. Kokaram, and R. Dahyot. Automated colour grading using colour distribution transfer. *Computer Vision and Image Understanding*, 107(1):123–137, 2007.
- [67] C. Poynton. *Digital Video and HD: Algorithms and Interfaces*. Morgan Kaufmann, 2003.
- [68] C. Poynton. Wide gamut and wild gamut: xvYCC for HD. Poynton’s Vector 8 (online), 2010.
- [69] P.H. Putman. What will replace the CRT for professional video monitors? In *SMPTE Conferences*. Society of Motion Picture and Television Engineers, 2009.
- [70] R. Ramanath, W.E. Snyder, Y. Yoo, and M.S. Drew. Color image processing pipeline. *Signal Processing Magazine, IEEE*, 22(1):34–43, 2005.
- [71] K. Sato. Image-processing algorithms. *Image sensors and signal processing for digital still cameras*, ed. J. Nakamura, pages 223–253, 2005.
- [72] G. Sharma. *Digital Color Imaging Handbook*, chapter Color fundamentals for digital imaging. CRC Press, 2003.
- [73] S.K. Shevell. *The Science of Color*, chapter Color appearance, pages 149–190. Elsevier Science Ltd, 2003.
- [74] A. Theuwissen. Image sensor architectures for digital cinematography. Technical report, DALSA, 2005.
- [75] WavelengthMedia. CCU Operations. Technical report, <http://www.mediacollege.com/video/production/camera-control/>, 2012.
- [76] C. Weerasinghe, W. Li, I. Kharitonenko, M. Nilsson, and S. Twelves. Novel color processing architecture for digital cameras with CMOS image sensors. *IEEE Transactions on Consumer Electronics*, 51(4):1092–1098, 2005.
- [77] W.D. Wright. A re-determination of the trichromatic coefficients of the spectral colours. *Transactions of the Optical Society*, 30(4):141, 1929.
- [78] G. Wyszecki and W. S. Stiles. *Color science: Concepts and methods, quantitative data and formulas*. John Wiley & Sons, 1982.
- [79] Y. Xu, A.J. Mierop, and A.J.P. Theuwissen. Charge domain interlace scan implementation in a CMOS image sensor. *Sensors Journal, IEEE*, 11(11):2621–2627, 2011.