2.3 Image Acquisition

Once an image has formed on the surface of a sensor as an irradiance function, the information must be converted by the sensor into a digital image, which is then transmitted or stored. The steps involved in this process are the focus of this section.

2.3.1 Sampling and Quantization

Let $s(\lambda)$, where $0 \leq s(\lambda) \leq 1$, be the sensitivity of the sensor to a particular wavelength λ. Then the image pixel value $I(x, y)$ can be modeled as the integration of the irradiance function over the area of the pixel and over all wavelengths, after first multiplying by the sensitivity function:

$$I(x, y) = \varphi\left(\iiint E(x', y', \lambda')s(\lambda')\, dx'\, dy'\, d\lambda'\right) \tag{2.12}$$

where the primes indicate dummy variables. The integrals over wavelength and sensor position, in addition to an integration over time (which is not shown), perform the work of **sampling** to convert the continuous irradiance function into a discrete function defined only over the rectangular lattice of integer (x, y) coordinates. **Quantization** then assigns a discrete gray level to every pixel in order to represent its value in digital form. However, to avoid an artifact known as *false contouring*,[†] it is important not only that there is a sufficient number of gray levels but also that they are meaningfully spaced. This is accomplished by applying a nonlinear mapping known as **gamma compression**, described in detail below, prior to quantization. The function φ includes both gamma compression and quantization, along with any sensor artifacts like blooming or noise.

2.3.2 Gamma Compression

The basic idea of gamma compression is shown in Figure 2.29. The raw measurement of light obtained by the sensor is transformed by a nonlinear mapping before transmission, storage, and/or manipulation. This step transforms the linear, physical light intensity into a perceptually uniform quantity, so that the pixel values in a digital image are not (in most cases) directly proportional to the amount of light collected by the sensor.

To understand gamma compression, we need to go back in time to consider an important fact of a now-obsolete technology. Cathode ray tubes (CRTs), which were the prevailing display technology for three-quarters of a century, have the curious property that the intensity of the light displayed on the screen is nonlinearly related to the applied voltage. More specifically, the transfer function of a CRT display follows a power law, in which the displayed intensity L (representing radiance or luminance) is proportional to the voltage V raised to some power:

$$L = cV^{\gamma} + b \tag{2.13}$$

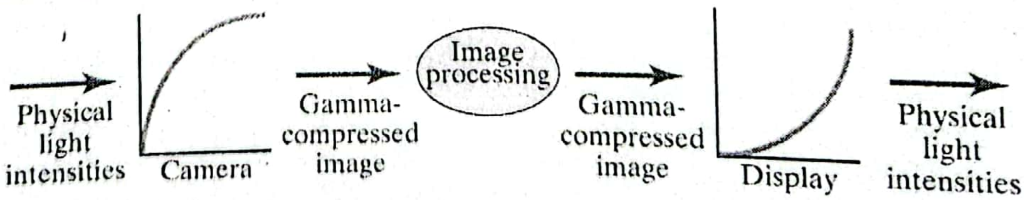

Figure 2.29 Linear light intensities are gamma compressed by the camera into perceptually uniform quantities, which are then gamma expanded by the display.

[†] Also known as *banding*, a form of *posterization*.

where γ is the exponent of the power function, and the constants b and c are the *blacklevel* and *contrast*, respectively, of the CRT display. If the monitor is adjusted properly so that its blacklevel is zero (i.e., the black pixels just barely emit light), then $b = 0$, leading to a simpler formulation:

$$L = cV^{\gamma} \tag{2.14}$$

Because of the widespread use of the Greek letter gamma (γ) for the exponent, this function is known as a **gamma function**. Figure 2.30 shows several plots of this function for different values of gamma, assuming $c = 1$ for simplicity. If $\gamma > 1$, the function is convex (curves upward) and is known as *gamma expansion*; if $\gamma < 1$, the function is concave (curves downward) and is known as *gamma compression*; if $\gamma = 1$, the function is linear.

CRT displays have a typical value of $\gamma_d \approx 2.2$, where the subscript indicates that this is the gamma of the display. To counter this effect and to simplify the electronics, video engineers decided many years ago that the voltage inside a display should be proportional not to the intensity of light being displayed but rather to the intensity raised to the power of γ_c, where $\gamma_c \approx 1/\gamma_d$. Cameras were therefore designed to encode the image according to $V = L_i^{\gamma_c}$, where L_i is the incoming light intensity, while displays produced light according to $L = V^{\gamma_d} = L_i^{\gamma_c\gamma_d}$. Images encoded in such a way are said to be *gamma compressed*, and if the gammas are inverses of each other $(\gamma_d = 1/\gamma_c)$ then they cancel each other $(L = L_i)$ so that the intensity displayed is the same as the intensity captured.

In practice, while the exponents used by the camera and display are nearly inverses of each other, they are not exactly so. In fact, when the image is expected to be viewed in lighting conditions different from those under which it was captured, the compression exponent γ_c is intentionally designed so that the product $\gamma_c\gamma_d$ is not 1. The reason for this choice is a perceptual phenomenon known as **simultaneous contrast** in which the human visual system's ability to discern contrast decreases in dark surroundings, as depicted in Figure 2.31. As a result, if a scene is captured in a bright outdoor setting but the resulting image (or movie) is viewed in a dim room or dark theater, it will lack contrast if displayed at the same intensity level. For this reason, various television and movie industry standards specify the *viewing gamma* (the product of the camera and display gammas) to be between 1.0 and 1.2, so that the viewing experience is subjectively correct even though it is not necessarily mathematically correct. Today, video cards typically also have a lookup table (LUT) that provides an additional adjustment, and the viewing gamma is defined to take this into account as well.

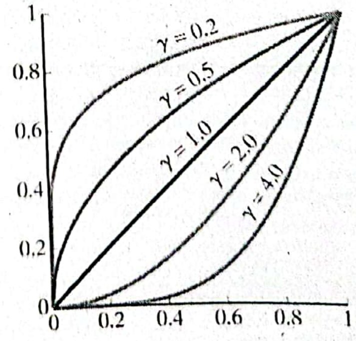

Figure 2.30 Gamma function with different values of γ.

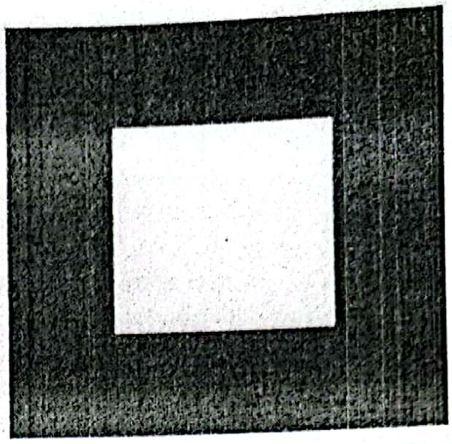

Figure 2.31 Simultaneous contrast. The pixels inside the middle squares have the same luminance, but the pixels on the right appear brighter due to its surroundings. Therefore, if an image is displayed at the correct luminance in a dimmer environment than the one in which it was captured, it will appear to be lacking in contrast.

In addition to this curious property of CRT displays, there is another more foundational reason that television engineers chose to introduce gamma compression many decades ago. Due to an amazing coincidence, the compression of light intensity according to $\gamma_c \approx 1/\gamma_d$ closely models the way in which the human visual system perceives light. In other words, although the doubling of the amount of light produces a *physical intensity* equal to twice the original, the *perceived intensity* is not increased linearly but rather nonlinearly according to a function quite similar to $L_i^{\gamma_c}$. Gamma compression therefore transforms a linear light intensity into a nonlinear quantity that is perceptually uniform. As a result, additive noise introduced in the transmission of a gamma-compressed analog video signal has minimal impact on visual perception, because a constant amount of additive noise affects the perceived intensity by an equal amount regardless of the overall signal value. Without gamma compression, however, additive noise in dark regions would produce a much more noticeable (and objectionable) effect on the viewing experience than noise added to bright regions. In the digital age this phenomenon still applies, particularly in the case of lossy compression which introduces additive noise to the image. JPEG and MPEG compression, for example, should always be performed on the nonlinear, perceptually uniform gamma-compressed signal rather than on the original linear signal, in order to minimize unacceptable artifacts. For the same reason, gamma compression prior to quantization results in a more effective use of the finite number of digital codes available. It is important therefore to view gamma compression not simply as an unfortunate relic necessary to maintain backward compatibility with now-obsolete CRT displays, but rather as an essential part of the image digitization and transmission process based upon timeless characteristics of human visual perception.

To see this connection between gamma compression and human visual perception, recall from our discussion on the human visual system that two luminances[7] can be discerned if their difference is at least 1% (approximately). That is, the **contrast threshold** of the human visual system is approximately $\frac{\Delta L}{L} = 1\% = 0.01$. Also recall that for image reproduction purposes the range of luminances is about 100 to 1. Now suppose we were to digitize these luminances in equally spaced intervals, in increments of 0.01, from 1 to 100, resulting in 9901 digital codes representing the luminance values of 1.00, 1.01, 1.02, . . . , 99.98, 99.99, 100.00. The drawback of this approach would be that, while the consecutive codes at the lower end of the scale, say 1 and 1.01, are discernible, consecutive codes at the higher end of the scale, say 99.99 and 100.00, are not discernible at all. The reason for this is that $(100.00 - 99.99)/100.00 = 0.01/100 = 0.0001 = 0.01\%$, which is much less than 1%, so that many codes at the higher end of the scale would be

[7] Luminance, which is radiance multiplied by the sensitivity of the sensor, is discussed in more detail later.

wasted. On the other hand, if we were to digitize luminances in equally spaced intervals in increments of 1.00, the digital codes of 1, 2, 3, ..., 98, 99, 100 would yield a barely discernible difference at the higher end but unacceptably large differences at the lower end, leading to objectionable false contouring. This is because $(2-1)/1 = 1.0 = 100\%$, which is much greater than 1%.

A more effective use of the digital codes occurs when consecutive codes correspond to relative luminance differences of approximately 1%. If we let v be the gray level and L be the luminance, then this is expressed as

$$0.01\Delta v = \frac{\Delta L}{L} \tag{2.15}$$

or

$$\frac{\Delta v}{\Delta L} = \frac{1}{0.01L} \tag{2.16}$$

where $\Delta v = 1$ is understood to be the difference between two consecutive gray levels, and ΔL is the difference between the corresponding luminances. If we let φ be the function that maps luminances to gray levels, i.e., $v = \varphi(L)$, the derivative of this function is $d\varphi/dL \approx \Delta v/\Delta L$. Therefore, the function f is the integral of the above expression, or

$$\varphi(L) = \int \frac{1}{0.01L} = 100 \log L \tag{2.17}$$

where log is the natural logarithm. This expression tells us that the desired nonlinear function that maps linear intensity to a perceptually uniform value is logarithmic. Table 2.1 compares this nonlinear coding with the two linear coding attempts just described. Linear coding requires approximately $100/0.01 = 10000$ gray levels, or 14 bits, to cover the 100:1 range with an increment of 0.01, while an increment of 1.0 requires $100/1 = 100$ gray levels, or 7 bits. In contrast, if the codes are spaced nonlinearly according to a ratio of 1.01, then only $(\log 100)/(\log 1.01) = 463$ gray levels, or 9 bits, are needed. The common 8-bit format, which owes its popularity to the widespread practice of grouping 8 bits into a byte in a digital computer, is sufficient for about a 50:1 ratio, roughly equivalent to traditional broadcast-quality television.

The assumption of a constant 1% threshold in Equation (2.15) is known as **Weber's law**. Although Weber's law is a good model of the transfer function of some cortical cells, it is not

linear 14-bit codes			linear 7-bit codes			logarithmic 9-bit codes		
gray level	L	$\Delta L/L$	gray level	L	$\Delta L/L$	gray level	L	$\Delta L/L$
00000000000000	1.00	1.00%	00000	1.00	100.00%	000000000	1.00	1.00%
00000000000001	1.01	1.00%	00001	2.00	50.00%	000000001	1.01	1.00%
00000000000010	1.02,	1.00%	00010	3.00	33.33%	000000010	1.02	1.00%
⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮
10011010101010	99.98	0.00%	1100001	98.00	1.02%	111001100	98.01	1.00%
10011010101011	99.99	0.00%	1100010	99.00	1.01%	111001101	99.00	1.00%
10011010101100	100.00	0.00%	1100011	100.00	1.00%	111001110	100.00	1.00%

TABLE 2.1 Logarithmic coding is a more efficient use of the available bits than linear coding because it results in successive codes that differ by the contrast threshold of 1% across the entire range of luminances. In contrast, linear 14-bit coding waste bits in bright regions where successive gray levels look identical, and linear 7-bit coding produces objectionable artifacts in dark regions.

an accurate model of human visual perception over all luminances.* As it turns out, a more accurate mapping between physical light intensity and perceived light intensity is obtained with a power-law function, known as **Stevens' power law**:

$$\varphi(L) = cL^{\gamma_h} \tag{2.18}$$

where $\gamma_h \approx 0.5$. Here we see the amazing coincidence that this gamma of the human visual system γ_h is nearly the same as the inverse of the CRT display gamma, γ_d, because $1/2.2 \approx 0.45$. Therefore, the gamma compression of a camera produces nearly the same mapping as that of the human visual system, which justifies our saying that gamma-compressed signals are perceptually encoded. Note that the power-law function with $\gamma < 1$ performs a similar operation to that of a logarithm function, since they both have similar concave shapes. While the linear quantity L is referred to as the *luminance* as we saw earlier, the nonlinear quantity that captures human perception on a uniform scale is known as **lightness**.

One drawback of the gamma compression function $\varphi(L) = cL^{\gamma}$ is that its slope is infinite at $L = 0$, leading to high amplification of noise in dark regions of the image. To overcome this problem, it is common practice to modify the function by specifying a linear section for values below some threshold τ:

$$\varphi(L) = \begin{cases} mL & \text{if } L \leq \tau \\ (1+\epsilon)L^{\gamma} - \epsilon & \text{otherwise} \end{cases} \tag{2.19}$$

where the slope m and offset ϵ are set to ensure that the value and first derivative of the two sections of the function match at the point $L = \tau$:

$$m = \frac{\gamma\tau^{\gamma-1}}{\tau^{\gamma}(\gamma - 1) + 1} \tag{2.20}$$

$$\epsilon = \frac{1}{\tau^{\gamma}(\gamma - 1) + 1} - 1 \tag{2.21}$$

The nonlinear transfer function obtained by modifying gamma compression is uniquely specified by the parameters γ and τ. There are two widely used standards that offer slightly different variations of gamma compression by choosing different values for these two parameters. **Rec. 709**,† the standard for high-definition television (HDTV) that was first approved in 1990, uses $\gamma = 0.45 \approx 1/2.222$ and $\tau = 0.018$, leading to $m = 4.5$ and $\epsilon = 0.099$:

$$\varphi_{709}(L) = \begin{cases} 4.5L & \text{if } 0 \leq L < 0.018 \\ 1.099L^{0.45} - 0.099 & \text{if } 0.018 \leq L < 1 \end{cases} \tag{2.22}$$

where the intensity L has been normalized to be in the range of 0 to 1.‡ Six years after the approval of Rec. 709, **sRGB** was developed to standardize the RGB color space used for still images for display on computer monitors and printers. sRGB uses $\gamma = 1/2.4 \approx 0.417$ and $\tau \approx 0.0031308$, leading to $m = 12.92$ and $\epsilon = 0.055$:

$$\varphi_{sRGB}(L) = \begin{cases} 12.92L & \text{if } 0 \leq L \leq 0.0031308 \\ 1.055L^{(1/2.4)} - 0.055 & \text{if } 0.0031308 < L \leq 1 \end{cases} \tag{2.23}$$

* In fact, if b is the bit depth, it is easy to see from Table 2.1 that the ratio of the highest luminance to the lowest non-zero luminance is $(1.01)^n$, where $n = 2^b - 1$. For $b = 9$ bits, this yields $(1.01)^{511} \approx 162$, which is a reasonable number. But for $b = 16$, the logarithmic model yields $(1.01)^{65535} \approx 10^{283}$, which is more than the number of atoms in the universe.

† Formally known as ITU-R Recommendation BT.709.

‡ Rec. 2020, used for UHDTV, uses the same transfer function as Rec. 709. However, in 12-bit mode, the precision is increased to $\tau = 0.0181$ and $\epsilon = 0.0993$.

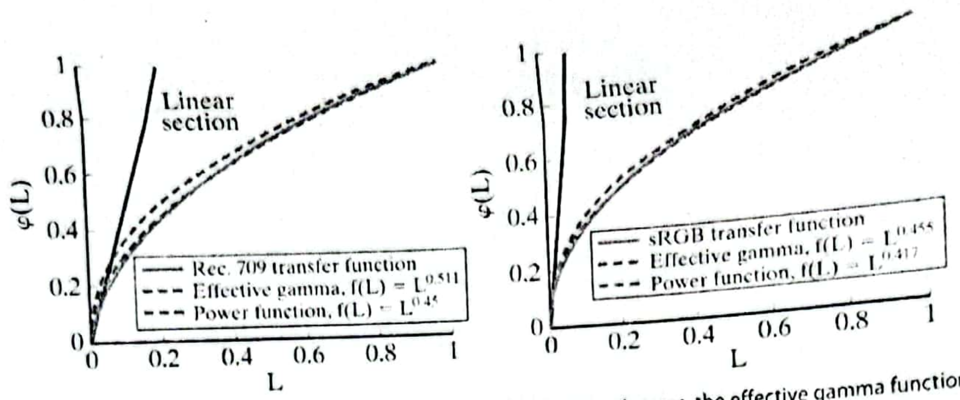

Figure 2.32 Rec. 709 gamma function (left), and sRGB gamma function (right). In both cases, the effective gamma function closely follows the modified gamma function, while the power function (using the same exponent as the modified gamma function but without the linear segment) is noticeably different. The dashed black line indicates the linear section, which is valid only for $L \leq \tau$.

as shown in Figure 2.32. Either way, the resulting nonlinear quantity is known as the **grayscale value** (or simply *gray level*), and since the nonlinear transfer function is designed to correspond closely to the power function of Steven's law, generally speaking *lightness* and *value* can be thought of as synonyms. Note that in the case of a color image, Equation (2.23) is applied to the color channels separately, with L replaced by R, G, and B after normalizing to the range of 0 to 1.

It is important to note that modifying the gamma function by inserting a linear section changes the **effective gamma** of the function. The effective gamma is defined as the exponent of the gamma function (*without* the linear section) that best fits the curve (*with* the linear section). As a result, although the exponent of the Rec. 709 transfer function is 0.45, its effective gamma is closer to 0.511. Similarly, although the exponent of sRGB is approximately 0.417, its effective gamma is closer to 0.455. In fact, both of these standards were designed by first specifying the desired effective gamma, then determining the exponent that best approximates that function. Because Rec. 709 is intended for viewing in dim environments, it was designed for a 1.125 viewing gamma, which is achieved using an effective camera gamma of $1/1.955555 \approx 0.511$, since $2.2/1.955555 = 1.125$, assuming a CRT gamma of 2.2. Empirically, this effective gamma of 0.511 is achieved pretty well using an exponent of $1/2.222222 = 0.45$. sRGB, on the other hand, was designed for a viewing gamma of 1.0, because it is intended for typical office environments. With a CRT gamma of 2.2, this yields $1/2.2 = 0.454545$ effective camera gamma, which empirically is achieved with an exponent of $1/2.4 = 0.416666$.

Having presented the concept of gamma compression in some detail, it is only appropriate to caution the reader that not all cameras perform gamma-compression. That is, some high-end cameras offer the possibility of storing the raw non-gamma compressed image, using a large number of bits per pixel in order to prevent false contouring. Raw images are useful for some computer graphics work, as well as for measuring the actual radiance of the scene. Nevertheless, unless you have good reason to believe otherwise, you should always assume that an image has been gamma compressed, especially with 8-bit-per-pixel-per-channel images.

2.3.3 CCD and CMOS Sensors

The light that falls onto the image plane is sampled by the sensor to produce values. These days, nearly all cameras are digital. The two most common digital sensors are CCD (charge-coupled device) or CMOS (complementary metal-oxide semiconductor). Each consists of a