4.1 The Digital Camera

A digital camera (or a digicam) is a camera where light is captured by electronic sensors instead of on photographic film. As a result, the images produced by such a camera are digital; they consist of pixels. A digital camera has the basic components of a film camera, most importantly a lens (often with zoom capability), an autofocus mechanism, a diaphragm, a shutter mechanism, and a timer. These components are used to admit the correct amount of light to the light sensing medium, which is an array of electronic sensors. Thus, the chief differences between digital and film cameras are (1) the light sensing medium and (2) the extensive use of electronics, which allows for a screen and a memory card. Digital cameras have the following advantages over film cameras:

- The image is displayed immediately (on a small screen in the back of the camera) after being captured. This enables the photographer to identify, delete, and retake bad pictures.

- Hundreds (often even thousands) of digital images can be stored in the camera on a small memory card that is reusable. Any image stored in the card can be displayed in the camera, transferred to a computer, and deleted. In contrast, a typical consumer film cartridge has room for only 24 exposures and it cannot be reused.

- Most digital cameras can record video (with sound) in addition to still images.

- In addition to the zoom produced by lens movement (optical zoom), a digital camera often features digital zoom, produced inside the camera by software that scales the image. New pixels are computed by interpolating neighboring pixels.

- Higher-quality digital cameras can perform simple image editing operations inside the camera.

- In addition to the consumer market, specialized digital cameras are made for and used in PDAs, notebook computers, cellular telephones, security devices, and telescopes.

The following is a list of the main types of consumer digital cameras:

- Compact cameras. Those are aimed at the casual photographer, the so-called point-and-shoot market. They are small and lightweight (the smallest ones are designated subcompacts and are also very thin). In order to achieve these goals and also reduce costs, compact cameras sacrifice picture resolution and quality, have restricted video and zoom capabilities, and also eliminate most or all advanced features. The pictures are stored only in the lossy JPEG format and the built-in flash is small and weak. At the time of writing (early 2010), compact cameras have a capacity of 8–10 megapixel, but these numbers grow steadily. Higher-quality compact cameras offer features such as larger screens (currently 2.5 in, 2.7 in, and 3 in), wider optical zoom range (currently up to 18 or even 24), image stabilization (sensors that compensate for camera shake by

(1) shifting elements within the lens, (2) shifting the sensor array, and (3) adjusting the amount of light), and a wide ISO range (a feature that mimics the sensitivity of film).

An important, but often misunderstood feature of compact cameras is the small physical size of the sensor array. It is typically only 6 mm on the diagonal, which makes for very small (actually, microscopic) individual sensors. A small sensor simply does not have enough surface area to collect much light, which is why compact cameras do not perform well under low-light situations. On the other hand, low light implies more depth-of-field (DOF) at a given aperture (Section 4.1.6).

- Hybrid cameras. These have recently been developed for those who want more capabilities and options than are offered by compact cameras, but don't want the size, weight, and price of a DSLR. Hybrid cameras are smaller and lighter than a typical DSLR, they have smaller lenses, and they lack the pentaprism mirror system that defines an SLR. As a result, such a camera does not have an optical viewfinder. Instead, it uses an electronic viewfinder from the sensor array. This feature implies that a hybrid camera can shoot video, which gives it an edge over a DSLR.

- Digital single lens reflex cameras (DSLRs). The SLR camera design has been around for decades, so it was natural for camera makers to adopt it to the realm of the digital. A DSLR i bulkier and more expensive than a compact camera, but has features to compensate for this. These cameras have wider optical zoom, interchangeable lenses, and a large sensor array, typically 18–36 mm on the diagonal. A large sensor array implies large sensors, not necessarily more sensors. As a result, a sensor can gather more light, which gives the DSLR better performance in low-light situations, but also reduces the depth-of-field (DOF) at a given aperture (Section 4.1.6).

  A minor point is the distinctive clack sound made by a DSLR when it takes a picture. This is caused by the mechanical movement of the mirror which is flipped out of the way and then back in.

In addition to these three types, there are other classes of digital cameras, such as bridge cameras, live preview cameras, and line-scan cameras. The latter type contains a single row, instead of a two-dimensional array, of pixel sensors. The camera scans a line on the object, waits for the object to move a bit, and repeats. The pixel data generated by those line scans is sent to a computer where a two-dimensional image is created. This type of camera is suitable for industrial purposes, where products constantly move on a conveyor belt and have to be scanned and automatically checked for defects.

4.1.1 History of Digital Cameras

As early as the 1960s, researchers developed and patented arrays of electronic (then referred to as solid-state) light sensors. Those were very bulky and were used for special applications such as tracking spacecraft. The first practical digital camera was built in 1975 at Eastman Kodak by Steven Sasson. The sensor array in this camera consisted of 10,000 CCD devices (developed by Fairchild semiconductors two years earlier). The camera was big and heavy, and the (black and white) images were saved on a cassette tape. The first image was taken in december, 1975. A picture of this camera is available at [msnbc.camera 09].

It took until 1981 for the first handheld digital camera, the Sony Mavica (Magnetic Video Camera) to make its debut (years later, Sony developed another Mavica that is completely different). The original Mavica had a sensor array, but it did not convert the electrical charges on the sensors to numbers. Instead, they were translated into analog electrical signals (similar to television signals) that were written on a 2-inch magnetic floppy disk. As a result, images produced by this camera had noticeable scan lines and were similar to television images. Other analog cameras were developed in the 1980s, but their costs and poor image quality made them unsuitable for consumer applications. They were useful only for special applications such as newspaper and television reporting (the images could be sent on telephone lines, and their resolution was similar to that of newspaper graphics).

It seems that the first fully digital camera was the model DS-1P, made by Fuji in 1988. It was impractical and it probably was never sold commercially. In 1990, Dycam Inc. made and sold its mode 1, a digital camera based on a 376×240 CCD sensor array. This camera produced grayscale (256 levels) images, stored up to 32 pictures internally in a 1 MB memory. The pictures could later be transferred to a computer. Other camera makers soon followed, with the result that size, weight, and price dropped, while resolution and number of colors increased steadily.

The adoption of the JPEG and MPEG compression standards in 1988 and the development of small, inexpensive LCD displays also helped to accelerate the development of digital cameras in the 1990s.

The first digital camera that also took videos made its debut in 1995 and the first megapixel cameras appeared in 1997. The first DSLR, the Nikon D1 (2.74 megapixel), was released in 1999. It was too expensive for casual users, but was affordable by professional photographers, especially since they could use the same Nikon lenses they already owned.

4.1.2 Camera Resolution

The term resolution is usually understood to mean the total number of light sensors, but should really refer to the width and height of the sensor array. The number of sensors (or pixels) in digital cameras has grown from 2–3 Mpixel in the late 1990s to around 8–12 Mpixel today (early 2010, although expensive cameras may have up to 60 Mpixels). The sensors are arranged in a rectangular array and the dimensions of the array (measured in sensors) determine both the total number of pixels and the aspect ratio (width over height). Instead of looking only the total number of sensors, a potential camera user should consider the dimensions of the sensor array. Examples of current sensor array dimensions are $2{,}012 \times 1{,}324$ (a total of 2.74 Mpixel and an aspect ratio of 3:2), $3{,}072 \times 2{,}048$ (a total of 6.3 Mpixel and an aspect ratio of 3:2), and $3{,}648 \times 2{,}736$ (a total of 10 Mpixel and an aspect ratio of 4:3).

Ten million sounds like a large number, but images are two-dimensional, which is why doubling the number of pixels of an image does not double the size of the image. Given n^2 pixels, they form a square image of n units on a side. Doubling the number of pixels to $2n^2$ increases each side of the square image to $\sqrt{2n^2} = \sqrt{2}n \approx 1.4n$, approximately 40% bigger. Thus, once we spread the pixels over rows and columns, there may not be enough of them to cover a single printed page. Here is what a total of 10 Mpixel means for printing. Suppose we want to print a 10 Mpixel image at the reasonable resolution of 300 dpi (300 dots per inch on the paper). Each square inch of paper will have $300 \times 300 = 90{,}000$ dots, so our 10 mega pixels can cover only 111 square inches, or an area of approximately 12.4×9 inches. This is a little more than the area of a standard American letter-size page.

If the same 10-megapixel image has to be printed on a poster-size sheet of paper, say 2×3 feet, then two approaches suggest themselves.

- Reduce the printing resolution. An area of 2×3 feet equals 864 square inches, so 10 million pixels provide 11,574 pixels per square inch, for a printing resolution of $\sqrt{11{,}574} \approx 108$ dpi. This sounds low and is certainly low for printing text, but images have noise (i.e., the eye may not notice when several pixels, or even many pixels, have the wrong colors), and experience shows that images printed at such a low resolution do not appear degraded and do not feature ragged edges or other negative effects of low resolution.
- Photograph the original image in several overlapping parts, and then use software to stitch these parts into a single, large image with enough pixels to be printed at a reasonable resolution on a large sheet of paper. This approach is employed by photographers and artists who produce large panoramas.

It is also important to bear in mind that the resolution is only one of the factors that affect the quality of a camera, the other factors being the quality of the lens, the physical size of a sensor, the filter array, and the demosaicing algorithm used by the camera (Section 4.1.5). A small sensor simply does not have the surface area to receive many photons and does not have the volume to hold much electrical charge, so when the output of a small sensor is digitized, it always results in a small number.

4.1.3 Light Sensors

An image sensor is a device that converts light energy to electrical energy. More specifically, it converts the energy of the photons impinging on it to electrical charge. Currently (early 2010), image sensors are either charge-coupled devices (CCD) or a complementary metal-oxide-semiconductor (CMOS) devices. These devices operate differently, but the final output is digital; the electrical charges are converted to numbers (normally 12-bit integers). A more detailed discussion of these devices is outside the scope of this book (in fact, of most books), and here they are simply referred to as light sensors, image sensors, or CCDs.

Virtually all current consumer digital cameras, including DSLRs, are of the single-shot type. In this type, there is a single sensor array with a Bayer filter mosaic (Section 4.1.5). A variation of this type employs three sensor arrays, one for each color component, that are exposed simultaneously via a beam splitter.

Cameras designed for special applications, such as shooting stationary subjects, can be of the multi-shot type, where the sensor array is exposed several times in the same shot. This type cannot be used with moving subjects. One way to implement this type of camera is to have three filters and place a different filter in front of the array during each exposure. Another option is to employ a single array with a Bayer filter and to physically move it, along the focus plane inside the camera, for each exposure in order to obtain a large number of pixels. In principle, it is possible to combine these versions of multi-shot; to expose a single array three rimes with different filters, and then move it and again expose it three times.

4.1.4 Gamma Correction

Since their inception, in the late 1830s and for many years afterwards, cameras were based on film. Even today (early 2010), most digital camera users have used, or have at least seen, film cameras. Therefore, a discussion of digital cameras should mention the most important differences between them and film cameras. The obvious difference is the use of solid-state sensors instead of film to capture the image, but a more basic difference stems from the fact that human perceptions are nonlinear. Here is what this means.

Imagine listening to a whisper (a sound intensity measured at about 20 dB) and immediately afterwards turning on a noisy appliance (such as a vacuum cleaner or a lawn blower) with a sound level of 120 dB. The difference in sound intensity may be a factor of 10,000, but the ear perceives the appliance noise as only about 9–10 times louder than the whisper. The amplitude response of the ear is nonlinear, and the same is true for other human senses, most importantly, weight and vision. When we wake up in a dark room and then walk into bright sunshine, the change in brightness may again be a factor of around 10,000, something that would overwhelm the brain, but the eye and brain perceive it only as factor of 9 or 10. Our senses protect us by their nonlinear responses, but are as a result unreliable as measuring instruments.

It has been shown experimentally that the nonlinear nature of human perception is logarithmic and is expressed by an elegant relation, known as Weber's law

$$dp = k\frac{dS}{S},$$

where dS is the change in a stimulus S, dp is the perceived change, and k is a constant whose value depends on the particular physical units used to measure the stimulus. Integrating this expression yields $p = k\log_e S$, implying that the perceived stimulus is proportional to the natural logarithm of the actual stimulus.

Thus, human vision is nonlinear, but so is film! When a scene with dark and bright areas is captured on film, the difference between the dark and bright areas on the film is less than the actual, physical difference. The sensors in a digital camera, on the other hand, respond linearly. The output I of the sensors should therefore be transformed to a value O that resembles the actual intensities that would be perceived by film or by the eye.

The electrical charge collected by a sensor is converted to an integer. This is normally a 12-bit integer in the range 0 to 4,095 (enough to express 4,096 levels of gray). A value of 0 indicates black (no photons sensed by the sensor) whereas 4,095 indicates white (the largest number of photons counted by any sensor). The middle value 2,047 corresponds to 50% gray. The eye can certainly sense black and white, so these two values should not be affected by the transform, but what about the values in between? The discussion above shows that these values should become darker, i.e., they should be decreased.

In order to understand the transform (which is referred to as gamma correction), we assume that the output I is a real number in the interval $[0, 1]$ where 0 is black and 1 is white. This is a reasonable assumption because we can simply convert the 12-bit integer $bb \ldots b$ to the real number $0.bb \ldots b$ and such a number is in the interval $[0, 1 - 2^{-12}]$.

The transform $I \rightarrow O$ should be a nonlinear but simple function that decreases all values of I, except 0 and 1, nonlinearly The simplest such transform has the form

$$O = I^{\gamma},$$

where γ (gamma) is greater than 1 and its precise value is selected experimentally.

This transform leaves the two values 0 and 1 unchanged, and decreases small (dark) values of I less than large (bright) values. As an example of this nonlinearity, consider $\gamma = 2.2$. Increasing I from 0.1 to 0.2 with this gamma, decreases O by 0.0226816, while increasing I from 0.8 to 0.9 decreases O by 0.181045. Figure 4.1 illustrates this transform. The top part shows a linear variation of grayscale from black to white and the bottom part shows how the values are darkened nonlinearly with a gamma value of 2.2. The Mathematica code is also listed, for readers who would like to experiment with this type of transform.

```
gamma = 2.2;
Show[Graphics[
 Table[{GrayLevel[x], Rectangle[{x,0},{x+.01,0.1}]},{x,0,1,0.01}]],
 Graphics[{GrayLevel[0], Rectangle[{1,0},{1.001,0.1}]}]]
Show[Graphics[
 Table[{GrayLevel[x^gamma], Rectangle[{x,0},{x+.01,0.1}]},{x,0,1,0.01}]],
 Graphics[{GrayLevel[0], Rectangle[{1,0},{1.001,0.1}]}]]
```

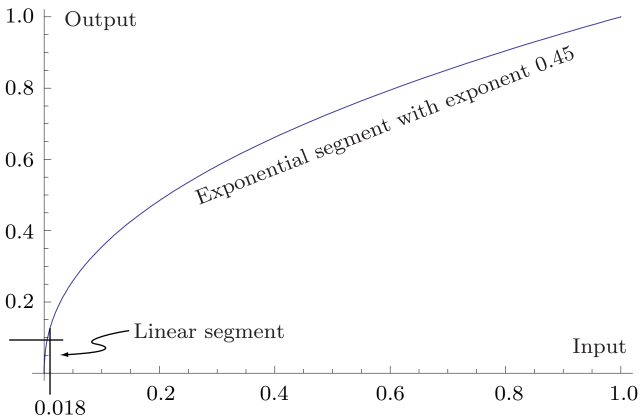
Figure 4.1: The Gamma Transform.

The concept of gamma and gamma correction are important in all areas of optical electronics, not just in digital cameras. It has to be applied to camcorders, CRT monitors, LCD monitors, light detectors, and other devices. Gamma correction is needed because many components of imaging and optical electronics systems respond nonlinearly and their response has to be converted to linear. For many years, the CRT was the primary component in computer monitors and televisions, so its performance has been carefully measured and is known in detail. A CRT is driven by a low-voltage video signal and it generates luminance on its screen. It turns out that the luminance (the CRT output) is a nonlinear function of this voltage (the CRT input).

The original NTSC video standard specified a gamma correction function with an exponent of $1/2.2 \approx 0.45$. For practical reasons, this ideal function has been changed and a new standard was proposed and approved as [SMPTE-170M] standard. It defines the two-part function (Figure 4.2)

$$\text{if } (V_{in} < 0.018),\ V_{out} = 4.5V_{in},$$
$$\text{if } (V_{in} \geq 0.018),\ V_{out} = 1.099V_{in}^{0.45} - 0.099,$$

where V_{in} and V_{out} are in the range $[0,1]$]. This is interpreted as follows: For low values of V_{in} (up to 1.8% of the maximum), the output is a linear function of V_{in} with a slope of 4.5. For higher values of V_{in}, the output is a power function with an exponent of 0.45. At $V_{in} = 0.018$, the two functions have the same value 0.081.




```
gamma = 0.45;
Plot[x^gamma, {x, 0, 1}]
```

Figure 4.2: NTSC Gamma Correction Curve.

4.1.5 Raw image format

We start with an analogy. The stone age of photography started in 1839 with Louis Daguerre. His photographic technique, known today as Daguerreotype, created the photograph as a one-of-a-kind image that was not reproducible. At about the same time, Fox Talbot revealed his photographic technique, which was based on a negative, and therefore allowed for easy reproduction of a photograph. Later eras of photography saw the development, among others, of color film and transparencies.

Consider the difference between shooting pictures with a negative and shooting with transparencies. In the former case, the negative has to be developed and a positive is then printed from it. This adds a step to the overall photo production, but also allows for processing of the image in the laboratory. While transferring the negative to the positive, a skilled photographer could create effects such as blurring, zooming, and variations of brightness and contrast. In the case of transparencies, the original film is already positive. The film has to be developed, but little lab processing can be done to improve or vary the resulting image.

Today, in the era of digital cameras and images, raw format and JPEG became the modern analogy of negative and transparency film. When an image is saved in a camera (and is later output) in raw format, it has to be processed before it can be printed or viewed. This adds a step to the overall image production, but also gives an experienced user a chance to process the image in a computer in useful ways. (The only camera settings that cannot be changed by software in this case are the ISO speed, the shutter speed, and the aperture.) On the other hand, if a camera compresses and converts each image to JPEG as it is being taken, and immediately discards the raw image data, opportunities for later processing are reduced, because much image information disappears in the lossy JPEG compression. Because of this analogy, raw image files are sometimes referred to as negatives and the process of converting such a file into a viewable/printabe format is called developing.

In principle, a raw image file should contain the dimensions of the image, the number of bits per pixel, a code for the color space used (RGB, CMY, or others), and the three color values for each pixel. In practice, such a file also contains metadata that is generated by the camera for each image. Examples of metadata are the date and time of shooting, the camera model and serial number, the shutter speed and aperture, the focal length, and whether the flash fired in taking the image. This type of metadata is also referred to as EXIF (exchangeable image format). Another important type of metadata is the color filter configuration of the light sensor.

Currently (early 2010) there are many formats of raw data, developed by makers of digital cameras. Most of these are proprietary. Such a format may not even be raw and may include lossless compression. It is also known or suspected that some raw formats are even encrypted, to prevent an occasional user from processing the image data. Some names of raw image formats are `.3fr` (Hasselblad), `.arw`, `.srf`, `.sr2` (Sony), `.bay` (Casio), `.crw`, `.cr2` (Canon), `.dcs`, `.dcr`, `.drf`, `.k25`, `.kdc`, `.tif` (Kodak), `.dng` (Adobe), `.erf` (Epson), `.fff` (Imacon), `.mef` (Mamiya), `.mos` (Leaf), `.mrw` (Minolta), `.nef`, `.nrw` (Nikon), `.orf` (Olympus), `.ptx`, `.pef` (Pentax), `.pxn` (Logitech), `.r3d` (Red), `.raf` (Fuji), `.raw`, `.rw2` (Panasonic), `.raw`, `.rwl`, `.dng` (Leica), `.rwz` (Rawzor), and `.x3f` (Sigma).

Raw image format has the following advantages over JPEG images:

- A typical consumer digital camera may have a few settings—such as cloudy, snow, beach, fluorescent, tungsten lights—to adjust exposure for the available lighting. When a raw image file is processed, the exposure can be modified to any desired values.

- A raw image file makes it possible to change the white balance to the correct value *after* the picture has been taken. The term "white balance" refers to the process of removing or changing wrong colors (or modifying the color temperature), such that white objects will be white in the final image.

- Depending on how the camera creates the JPEG file, a raw file format may provide considerably more dynamic range than a JPEG file. The term dynamic range refers to the range of light to dark that can be captured by a camera before becoming completely white or completely black.

- Each color component in a raw file is normally represented in 12 bits, as opposed to 8 bits in a JPEG image file. The larger number of bits makes it possible to correct minor exposure errors and adjust color tones when the raw image is processed outside the camera.

- JPEG compression of an image starts by changing the color space to luminance-chrominance. Compression is lossy, which is why trying to change the color space after such a file is decompressed leads to significant loss of visual information. A raw image file, in contrast, allows for quick and lossless transformations of the color space.

- JPEG files are small, but excessive JPEG compression results in annoying compression artifacts.

JPEG files, on the other hand, have the following useful features:

- The file is smaller.

- A beginner or an amateur photographer does not have to spend time processing the image files. The pictures are stored in the camera in their final form and can easily be examined, deleted if necessary, or transferred to a computer for printing and storage.

- A JPEG file can easily be exchanged between users because JPEG is a compression standard. Raw formats, on the other hand, are often proprietary.

- If the camera settings are correct, the resulting JPEG file will be as good as a raw file.

In a discussion of raw versus JPEG formats, it is important to explain how the light sensor array inside a camera is organized, what data it captures, and how the raw data is prepared. Most digital cameras have a rectangular array, called a mosaic sensor or color filter array (CFA), of CCD or CMOS sensors, each of which contributes a pixel to the final image. Light of many different wavelengths (corresponding to different colors) falls on each sensor, but the sensor counts only the total number of photons that impinge on it; it does not identify their frequencies (i.e, colors). During exposure, each sensor accumulates electrical charge that is proportional to the intensity (but not the color) of the light it has sensed. Thus, the sensor array generates a grayscale image.

Once this is grasped, it is not hard to figure out how to obtain color data from the sensors. Simply cover each sensor with a filter that lets only one color through. When we look at the world through rose-tinted spectacles, everything looks rosy, because only rose color reaches our eyes. Thus, sensors covered with a red filter output a grayscale value proportional to the red component of the light that strikes them.

> Rose-colored glasses are never made in bifocals. Nobody wants to read the small print in dreams.
>
> —Ann Landers

A slight problem arises because a color space is three dimensional but the sensor array is rectangular. It is easier to partition a rectangular array into groups of four sensors than into groups of three, but such partitioning can be done and Figure 4.3 illustrated two ways of doing so. The configuration in part (a) of the figure is very common and is called a Bayer pattern color filter [Bayer 76]. Some cameras may filter

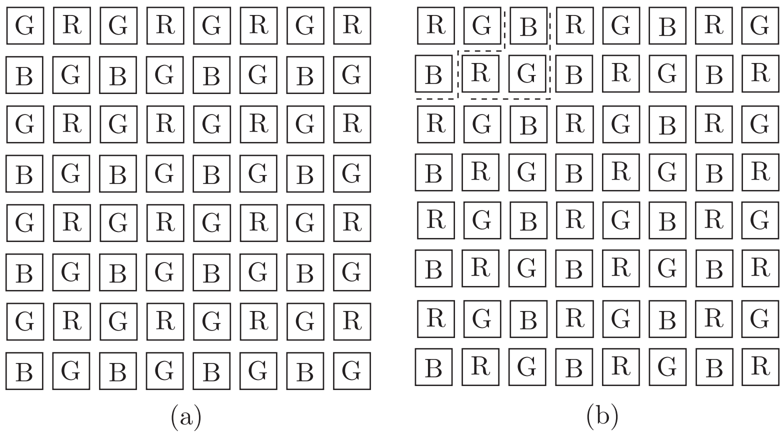


Figure 4.3: Color Filter Arrays.

four colors simply because it is easier to partition the array in groups of four sensors each.

Notice that half the sensors in the Bayer pattern are covered with green. This is because the eye is more sensitive to green than to red or blue (Figure). In the pattern of Figure 4.3b there are about the same number of sensors for each color (if the width w of the sensor array is divisible by 3, there are exactly $w/3$ sensors for each color, and the same is true for the height of the array).

After an exposure, the electrical charges in the sensors are converted to numbers (grayscale values) that are written on the raw image file. The file has to be processed later for viewing or printing, a process known as “raw conversion.” The first step of raw conversion is to prepare three complete arrays of values for the three color components. Figure 4.4 illustrates this step. Only 25% of the original sensors produce red data, only 25% produces blue data, and only 50% of the sensors produce green data. Each empty position in the three arrays has to be filled up by interpolation from nearby pixels. The figure illustrates the simplest interpolation method. In part (a) of the figure (the red and blue components), each empty position labeled 2 in the top row is set to the average of its two nearest neighbors, while the leftmost position, labeled 1, is set equal to its only neighbor. The third row from the top is interpolated in the same way, and the second row is then computed as the average of its two neighbor rows. In part (b) of the figure (the green color component), each empty position has three or four near neighbors, except two positions (labeled 2) in opposite corners. Such an interpolation is known as demosaicing (or demosaicking), because the three pixel arrays resemble mosaics.

More sophisticated demosaicing methods are possible. Such a method may compute a value for an empty position as a weighted sum of eight (or more) positions, with larger weights assigned to nearby neighbors. However, including many neighbors in an interpolation may lead to blurring or even complete disappearance of small details. Imagine a detail that occupies a small group of 3×3 pixels centered on the X of

Figure 4.4b. It makes sense to compute a value for pixel X by interpolating its four nearest G neighbors, but if we also include in this interpolation the four G neighbors shown in gray (which are located outside the detail), all the visual information of the detail may be lost.

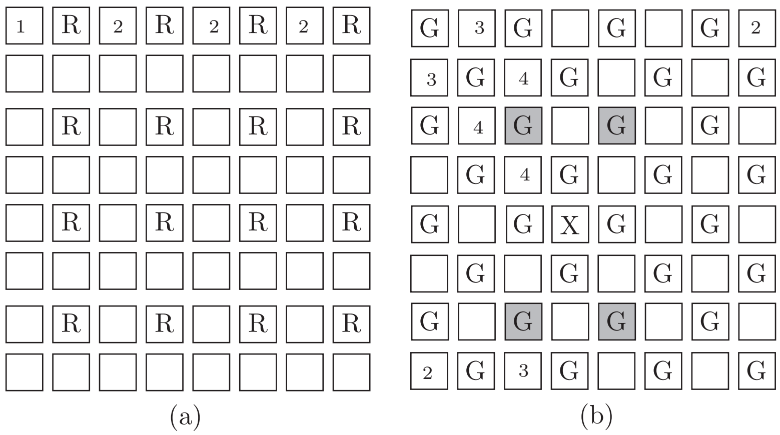


Figure 4.4: Empty Positions in Color Arrays.

Demosaicing by pixel grouping. This is an example of a fast, efficient algorithm that employs interpolation in an original way, depending on relations between neighboring pixels. This simple algorithm, by Chuan-kai Lin [cklin 03], is based on the observation that a continuous-tone image (an image of a natural scene, as opposed to an image of artificial objects) often contains groups of strongly-correlated pixels. Thus, given an empty position X in a Bayer grid, we can best compute a value for it by identifying those neighbors of X that are most similar to it. This principle is applied by the algorithm to the green positions. The red and blue empty positions are computed by simple interpolations based on hue transitions.

We use the following position numbering in a sample 5×5 Bayer grid:

R1	G2	R3	G4	R5
G6	B7	G8	B9	G10
R11	G12	R13	G14	R15
G16	B17	G18	B19	G20
R21	G22	R23	G24	R25

The algorithm computes values for the empty positions in three parts as follows:

Part I. Interpolate the green values in the red or blue positions in two steps.

Step 1. Every blue position (and most red positions) have four green immediate neighbors. In the few red positions that have only two or three immediate green neighbors, we use simple interpolation. For all other red and blue positions, we first compute four differences (or gradients). For position R13, for example, the four gradients are:

$$\begin{aligned}
\Delta N &= 2|R3 - R13| + |G8 - G18|,\\
\Delta E &= 2|R13 - R15| + |G12 - G14|,\\
\Delta W &= 2|R11 - R13| + |G12 - G14|,\\
\Delta S &= 2|R13 - R23| + |G8 - G18|.
\end{aligned}$$

Thus, gradient ΔN expresses the amount of color correlation in the north (up) direction about R13, and similarly for the other three gradients.

Step 2. Select the smallest gradient and compute a value for G13 as a weighted sum of four positions as follows

$$\mathrm{G13} = \begin{cases}
\Delta N \text{ is minimum}, & (3\mathrm{G8} + \mathrm{R13} + \mathrm{G18} - \mathrm{R3})/4,\\
\Delta E \text{ is minimum}, & (3\mathrm{G14} + \mathrm{R13} + \mathrm{G12} - \mathrm{R15})/4,\\
\Delta W \text{ is minimum}, & (3\mathrm{G12} + \mathrm{R13} + \mathrm{G14} - \mathrm{R11})/4,\\
\Delta S \text{ is minimum}, & (3\mathrm{G18} + \mathrm{R13} + \mathrm{G8} - \mathrm{R23})/4.
\end{cases}$$

Part II. Interpolate the blue and red values in the green positions. As an example, we compute B8 and R8 at position G8.

$$\begin{aligned}
\mathrm{B8} &= \mathrm{HueTransit}(G7, G8, G9, B7, B9),\\
R8 &= \mathrm{HueTransit}(G3, G8, G13, R3, R13),
\end{aligned}$$

where function HueTransit is defined as

```
function HueTransit(i3, i2, i3, v1, v3)=
 if(i1<i2<i3 or i1>i2>i3)
  then return v1+(v3-v1)(i2-i1)/(i3-i1)
  else return (v1+v3)/2+(2i2-i1-i3)/4
```

Part III. Interpolate the blue and red values in the red and blue positions. As an example, we compute B13 at position R13.

$$\begin{aligned}
&\Delta ne = |B9 - B17| + |R5 - R13| + |R13 - R21| + |G9 - G13| + |G13 - G17|,\\
&\Delta nw = |B7 - B19| + |R1 - R13| + |R13 - R25| + |G7 - G13| + |G13 - G19|.\\
&\text{if}(\Delta ne \le \Delta nw)\\
&\text{then } B13 = \mathrm{HueTransit}(G9, G13, G17, B9, B17)\\
&\text{else } B13 = \mathrm{HueTransit}(G7, G13, G19, B7, B19)
\end{aligned}$$

This algorithm is fast because it employs only addition, subtraction, few multiplications, and absolute value. The divisions by 2 and by 4 can be done by right shifts.

(End of algorithm.)

If the camera supports a raw image format, the raw conversion is done in a computer, normally with proprietary software. Such software is either supplied by the camera manufacturer or is implemented (as in the case of Adobe photoshop) by a software maker. Raw conversion starts with demosaicing, but may also include steps for the following types of processing:

- White balance.

- Colorimetric interpretation. The visual sensation of color is very personal. If you prepare a list of shades of red and ask people to choose the "real," or "best" shade, there may never be complete agreement. Similarly, filters installed in digital cameras differ in the precise shade of red (and any other color) that they transmit. Sophisticated raw conversion done in software may allow the user to correct each color component individually until all the colors of the final image are satisfactory.

- Gamma correction. This is discussed in Section 4.1.4.

- Noise reduction, antialiasing, and sharpening. Demosaicing is based on interpolation, so it necessarily results in a certain amount of blurring. A sharp edge in an image may be lost because of the interpolation, so a raw converter should include a sharpening algorithm. Antialiasing is discussed in Section .

Now we get to JPEG. A camera that stores and outputs its images in JPEG, includes a raw converter and a JPEG compressor. Each exposure is followed by a blank period of a second or so where the camera is busy converting the image, compressing it, and storing the resulting JPEG data on its memory card. The raw converter is built into the camera and generally cannot be modified (although in principle, the raw converter may be stored in the camera as firmware, and may be updated from time to time). Most cameras permit the user to specify parameters such as the ISO value, the final image size, amount of loss in compression, several light conditions (cloudy, seaside, fluorescent, tungsten light, nighttime), and aperture. Higher-quality consumer cameras may also offer user-controlled settings for color space, sharpening, contrast, and perhaps others. Obviously, the average user generally leaves these parameters at their default values and simply deletes and retakes any bad images. However, patience, attention to detail, and willingness to experiment with camera settings can work miracles and result in excellent images even when taken under unfavorable conditions. Thus, even the most occasional user is advised to read the camera's manual and experiment with all its settings, because once a bad picture has been taken and converted to JPEG, there is precious little that can be done to improve it.

In addition, most current digital cameras convert the electrical charge in a sensor to a 12-bit number, and raw image files save all 12 bits. A typical JPEG compressor built into a camera, starts by discarding the four least-significant bits of a raw value and retaining only the eight most-significant bits, thereby losing visual data even before compression starts.

Figure 4.5 lists the main steps taken in a camera to produce either a raw or a JPEG image file.

4.1.6 Appendix: Depth-of-Field

When a beam of light hits an object, it is absorbed, reflected, or refracted. It may even be partly absorbed, partly reflected, and partly refracted. Mirrors are useful because

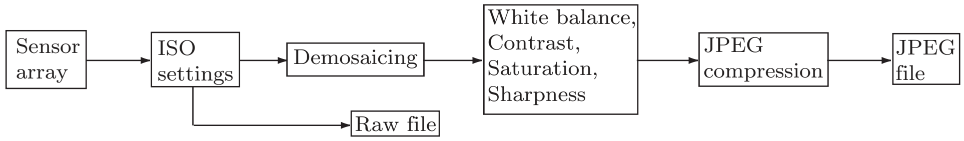


Figure 4.5: Producing JPEG and Raw Image files.

they reflect light, but the magic of lenses is based on light refraction. The light "bends" when it moves from air to glass and bends back when it exits from glass back to air. Refraction happens because the speed of light depends on the density of the medium it travels through.

Figure 4.6a shows how light rays that are perpendicular to a lens are bent because of refraction and converge to the focus at F, but the use of lenses in a digital camera is in focusing an entire image on the sensor array in the focus plane. The problem is that every point on the subject that is being photographed emits light in all directions. To get a sharp image on the sensors, all the rays emitted from a certain point x on the subject have to bent so that they hit the sensors at the same point y. Figure 4.6b shows a subject to the left of a lens and how three rays leaving point x are bent differently by the lens and end up at point y on the focus plane f. We say that the subject is focused at plane f. If the subject is moved away from the camera, it will be focused in another plane, closer to the focus F. This is why focusing in old cameras was done by moving the lens back and forth, thereby changing the distance between the lens and the film. If the object is moved all the way to infinity, its focus plane becomes the plane containing the focus point F.

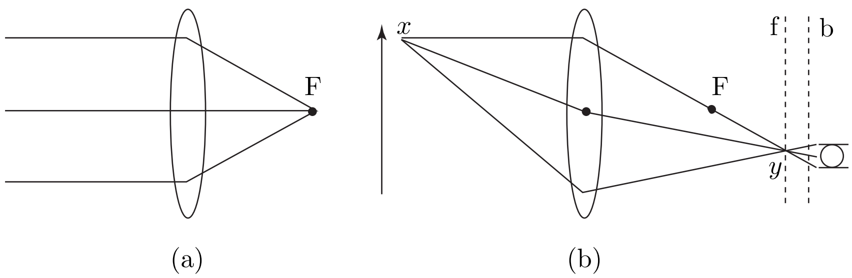


Figure 4.6: Refraction.

Now consider plane b in Figure 4.6b. The three rays from the subject diverge and hit this plane at three different (but nearby) points. A three-dimensional diagram showing more rays from point x would show that they form a small circle on plane b, the so-called *circle of confusion* (COF). It is now clear that if the subject is located at the precise distance the lens is focused, every point on it will focus to a point on the focus plane. When the subject moves out of focus, the points on the focus plane become circles. The farther out of focus the subject is, the larger these circles become.

(Circle of confusion: A group of photographers sitting around trying to understand Depth of Field.)

This is in principle. In practice, it is humans who look at photographs, and the human eye has limited resolving power (it has evolved to help our ancestors hunt saber tooth tigers, not bacteria). When we look at a small enough circle, we see it as a point, which is why subjects that are in principle out of focus may still look sharp in a photograph. The result is that an image appears to be sharp (in focus) over a range of distances and this range is termed the *depth of field* (DOF).

> In theory, there is no difference between theory and practice. But, in practice, there is.
>
> —Jan L. A. van de Snepscheut

We can therefore define the depth of field as the length of the interval in front of and behind a focused subject in which the photographed image appears sharp.

To measure the DOF of a lens, we first have to decide on the diameter of the COF. Different diameters yield different DOFs.

The resolving power of the eye depends on the person and on age. It varies widely, but on average we can use one minute of arc as a representative figure. This means that at a normal reading distance of 20 inches, the smallest detail a person with perfect eyesight can see (under ideal conditions) is about 1/16 (or 0.1667) mm. Two dots placed closer than this next to each other will appear as one dot. Obviously, the depth of field depends on what we consider blurred. A person who tolerates larger circles of confusion will claim that his camera has a great depth of field, while someone less lenient may find that the same camera produces a smaller depth of field. Lens manufacturers often write the depth of field on the lens, and for 35 mm film cameras this specification was based on the following argument:

In a 35 mm camera, 35 millimeters is the size of the diagonal of the negative, so the width of the negative is about 24 mm or one inch. To enlarge such a negative to a 5×7 print, the enlarging factor is 5. If we want the circles of confusion to be at most 0.1667 mm after the enlargement, they have to be at most $0.1667/5 \approx 0.0333$ mm before the enlargement. This was the COF size that 35 mm lens manufacturers used when measuring the depth of field of new lenses.

In principle, the depth of field depends on the following factors: aperture size, focus distance, lens focal length, sensor size, sensor array organization, and the final print size.

Of these factors, the easiest for the user to vary is the aperture size, so how does the depth of field depend on the lens size? The circles of confusion are formed by light that passes through the lens, so less light implies less confusion in the circles, and therefore greater depth of field. This is an intuitive explanation which is easy to illustrate with a camera. An old camera, where it is easy to vary the aperture, is best. Look at a scene that includes objects at different distances, close the diaphragm gradually and you'll see the scene sharper. You'll also see it darker, but you can compensate for that by increasing the exposure time.

Thus, a small lens results in a greater depth of field, but the application of this phenomenon is limited, because of the effects of diffraction. Some light is always diffracted at the edge of the lens, and for a small lens, this light becomes a significant percentage of the total light, resulting in a poor-quality image.

A detailed (but hard to obtain) reference on DOF is [Baker 85].

References

Baker, Alfred A. (1985) *Applied Depth of Field*, Boston and London, Focal Press (Elsevier).

Bayer (1976) is US patent 3,971,065 (20 July 1976), Bryce E. Bayer, *Color imaging array*. This is available in PDF from
`http://www.google.com/patents?id=Q_o7AAAAEBAJ&dq=3,971,065`.

cklin (2003) is `http://web.cecs.pdx.edu/~cklin/demosaic/`.

msnbc.camera (2009) is `http://www.msnbc.msn.com/id/9261340/`.

[SMPTE-170M] Society of Motion Picture and Television Engineers (1994), “Television—Composite Analog Video Signal—NTSC for Studio Applications,” SMPTE-170M. This is available from `http://www.smpte.org/standards`.