Imaging Systems

Joshua Marrah
Image Credits:

All images, unless otherwise noted here, were taken/created by Josh Marrah © 2010.

Used under the GNU Copyleft License:

Top p.21, all p.22, top p.24, top p.36.

Other photographs supplied by:

PhaseOne (p.21), Betterlight (p.24), Foveon (p.24), Epson (p.25,36), HowStuffWorks.com (p.36), HP (p.37)

Other:

P.28: Supplied by Professor Nitin Sampat, RIT

© 2010 by Josh Marrah
Designer, Writer, and Photographer
All Rights Reserved
Printed in association with Rochester Institute of Technology
Course: Imaging Systems
Professor Nitin Sampat
Focus stacking is a technique for acquiring the sharpest images with deepest depth of field possible.

Most lenses are sharpest around two or three stops down from their maximum aperture. When opened all the way they tend to get soft and when they are stopped all the way down they are diffraction limited, also causing them to become blurry. However, in macro photography using the lens at its prime aperture often does not grant a deep enough depth of field, since this decreases as the subject gets closer to the lens.

Focus stacking allows a user to have the best of both worlds, using the lens at its sharpest aperture while also acquiring as deep a depth of field as is required, thus avoiding softness due to lens diffraction.

In order to do this, the lens must first be evaluated in order to find its sharpest aperture setting. Once this is known, the lens is used at that aperture to take a series of photographs. Starting with the focus at one end (near or far) of the subject, a series of photographs is taken, slightly changing the point of focus each time until the whole range of focus required has been covered. The in focus areas need to overlap a little bit in order for later steps to work, so correct image capture is a necessity here.

Once the photographs have been taken, they then need to be aligned since the tripod may have moved and the scale of reproduction often changes along with focus. Once the images have all been aligned, they are cropped down to the same size and saved as new files. They are now ready to be run through the focus stacking software, which combines all the in focus areas from each image to form the final image.

Examples of software to do this with include: CombineZM, Helicon Focus, PhotoAcute Studio, Macnification, and Zerene Stacker. Some of these are free and some of them are not. The software that costs money usually is capable of doing all of the post-capture steps in one step.
Chapter 5: Focus Stacking

Focus Stacked image with images captured at F/11

Fundamentals

Input

Processing

Output

Special Topic: Focus Stacking

Table of Contents

9 Fundamentals

19 Input

27 Processing

33 Output

39 Special Topic: Focus Stacking
Chapter Five
Special Topic
Focus Stacking
Inkjet is another digital printing process. Instead of having all of these rollers, plates, and lasers it has a print head that hovers over the paper and spits little droplets of ink. Each print head has thousands of nozzles that it uses to eject ink onto the paper. On consumer grade inkjet printers, the print heads are replaced along with the ink reservoir when it is empty. On higher end printers though, the print head would be too expensive to replace regularly since it is normally large and has more nozzles which allow it to print faster with more colors.

Inkjet printers are very good for printing on papers that are highly textured since there are no rollers that need to be in direct contact with the paper. They are also used in high end photography since some have as many as twelve different inks, giving them a much larger color gamut.

However, inkjet printers are very slow at the moment since a print head must move back and forth across the page and the nozzles can only spit out very small drops of ink (which is good for sharpness but bad for speed). They also tend to be much more expensive per page than a digital laser printer.
What is an Image?

First off, in order to start a book about imaging systems, we need some basic understandings, fundamentals, with which to build the rest of our book upon.

First and foremost, what is an image?

An image is nothing more than a combination of spatial, tonal, and spectral information. A fourth resolution, temporal, is included in certain applications.

Along with that, what is an imaging system?

It is the whole pipeline, from start to finish. A source of image data, a way of processing that data, and then a final output.

By the end of this book, a reader should have an in-depth understanding of all of these steps and their effects.

Printing

An image has been made, it has been processed, and now something must be done with it. An image is made to be shown.

Traditionally an image would be printed at this stage, but how does a print get made and which printer should someone use?

The easiest and largest distinction for printers is the split between analog and digital.

An analog printer uses a plate that has been inked to transfer an image, whereas a digital printer simply does not.

Almost all books, magazines, newspapers, advertisements, and other large scale (run lengths) printing jobs are printed on analog offset presses. These are used for large repetitious jobs because they are the fastest printers available, but are also expensive to run, own, and print on. Since these printers use photomechanical plates, each page needs a new plate to be made and each plate is expensive. If the page also has 4 colors (CMYK) then a plate needs to be made for each color channel also. However, once these plates are made they will last a long time.

Once in the press, the plate is wrapped around a roller and evenly covered in ink. The ink that needs to be applied to the paper gets applied to a second roller called a “blanket” (the offset) and this applies the ink directly to the paper.

So, if a digital printer does not use plates, how does the information get transformed into ink?

Laser printers work very similarly to analog offset presses but instead of the plate roller, they use a laser and a Photo Imaging Cylinder (PIP Drum). The laser scans across the drum as it spins, giving certain areas a charge. This charge attracts toner, which is then transferred to a blanket roller or applied directly to the paper itself.

This process is great because it means that an expensive plate does not need to be made every time someone clicks ‘print’ on their computer. Each page can be different without extensive setup time and costs. However, this is a relatively slow process since the drum needs to be cleaned, neutralized, and re-exposed for every page and every separation. An analog press has separate plates for each color, digital presses must reuse the photoreceptive roller for each color. This means that a sheet of paper could go around the blanket roller as many as four times before it allows the next page to start being printed.
Quick Glossary of Photographic Terms

**Exposure:** Allowing the correct amount of light to hit a sensing device such that there will be enough signal received to record data, but not so much that the sensor is overwhelmed.

**Shutter Speed:** The length of time a shutter exposes a sensing device.

**Aperture:** The opening in a lens that adjusts to let in differing amounts of light. This also changes the depth of field. The bigger the f/number the smaller the opening.

**ISO/ASA:** This describes the sensitivity of the sensing device to light. The higher the number, the higher the sensitivity.

**Stop:** Term used to describe the halving or doubling of light. Halving the shutter speed is a change of 1 stop and lets in half as much light.

**Exposure Latitude:** Also called “dynamic range”, this describes how separate the highlights and shadows can be before details fail to be recorded. The higher this is, the more room there is for error.

**Contrast Ratio:** The difference between the highlights and shadows in an image.
Chapter Four

Output
Spatial Resolution (x,y)

Spatial resolution is the horizontal and vertical resolution of the imaging device. For example, digital cameras are reported as having a certain number of pixels. A PhaseOne p21+ has 18 megapixels (18,000,000). Which is its spatial resolution, 4904px by 3678px.

Each of these pixels does nothing more than measure light. Made out of silicon, it records how much light gets absorbed into it, and then reads out a certain voltage for that level. Putting together all of these pixels into a grid pattern will give you a series of light values that you can then represent as a grayscale image.

Spatial resolution mainly affects how sharp or detailed a image is going to be. The higher the pixel count, the sharper the image (assuming all things equal).

How much spatial resolution one needs is highly dependant on their usage. If a photograph is taken of a white piece of paper, than one pixel will report the same information that 6 million will. Now, if one photographs an extremely detailed piece of work, such as a whole page from the Book of Kells, then more pixels would be desired.
Tonal Resolution (z)

Tonal resolution, also known as dynamic range, is the range of tones and differing amounts of light a single pixel can record, from brightest to darkest. The higher this is, greater the exposure latitude. It also increases the users margin for error in an exposure if the scene has less contrast range than the sensor can record.

The greater the tonal resolution the smoother gradients will be and subtler changes in luminance level can be recorded. In general, human eyes can discern about 128 distinct levels of grey. This is the equivalent of a 7 bit tonal resolution depth ($2^7$).

Most consumer cameras record images at 12 bits per channel and then convert the images to 8 bits/channel, which gives them some buffer room for noise and calculations. This adds up to a total of 24 bits/pixel for a RGB image.

Some professional cameras will capture and record images at 16 bits, which allows for greater accuracy when calculating color and gradients, and also makes the files twice as large.
Unsharp Mask

Unsharp mask is one method for sharpening an image. Developed in the film darkroom, the process is fairly complicated and there were dedicated unsharp mask technicians, who got paid to know how to do this one darkroom technique. Now, things are much easier with software applications that will do it with one click. However, when this filter is opened, the user is presented with options and unless they understand what these parameters mean, the filter may not work very effectively. If used incorrectly, it may end up increasing noise in solid areas - sharpening an image where there is nothing to sharpen.

The overall concept of how this works is actually pretty simple. Start with an image containing edges that need to be enhanced, unsharp mask just isolates the edges from the rest of the image, and then adds them back to the original. This increases contrast in that area, and makes it appear sharper.

Breaking it down into steps, first start with an image. Then the image gets blurred slightly - effectively getting rid of the hard edges. Now, this blurred image is subtracted from the original. This has the effect of leaving only the edges in a photograph since the smooth areas are what got removed leaving what will be used for a mask. So now this mask is added back to the original image, and the edges are enhanced.

So in Adobe Photoshop’s unsharp mask filter dialog, a user is presented with three sliders. The first is amount, which changes how much this mask affects the final image. Second is the radius slider which is the radius of the blur used to make the mask. The last slider, which is threshold, changes where the mask is applied. This is added because it allows the user to avoid applying a sharpening mask to an area of smooth gradients, which would just increase noise in those areas.

An example starting point for an unsharp mask may be an amount of 100%, a radius of 1 px, and a threshold of 3 levels (this is only an example and each image will require its own settings), for my sample image this provides a moderate level of sharpening.

I can also use unsharp mask to increase local contrast, by using a very large pixel radius and a small amount setting. This however increases overall local contrast instead of just edge contrasts, but is less likely to give a visible halo effect around objects.

Color Resolution

Color resolution defines the spectrum of light that a sensor can capture, often represented as a color gamut for the sensor. This gamut is dependant on the color and saturation of the filters used and the ability for an output device to reproduce the colors also. The bigger the gamut is of the capture device, typically the better, even if it can not be reproduced by an output device it still helps in color calculations.

Typically digital cameras capture light in three channels, red green and blue. Some special application cameras will have an extended range of sensitivity, into other ranges of the color spectrum.

Other color channels are used in different applications. Red, green, and blue are the primary colors of the additive color system. They’re used when working with light, because added together they give you white.

When working with ink on paper, or reflected light, then the subtractive color system is used. Consisting of cyan, magenta, and yellow these colors are the opposite of red green and blue. This system also uses a 4th channel, black, to increase the maximum density possible.
Temporal Resolution

Temporal resolution is the 4\textsuperscript{th} dimension of imaging, adding time. Still images have no temporal resolution other than the time captured during the time the shutter was open. For video television in the United States, this is locked at 24 or 29.97 frames per second.

File size of a photograph

Is calculated by multiplying all the different resolutions together. So, if a sensor is 1024px by 1024px, has 8 bits of tonal resolution, and has three channels (red, green, and blue) and takes a still image.

We take the spatial resolution and multiply it together to find the total amount of pixels, which is 1,048,576 pixels. We then multiply it by the bit depth per pixel (8), which leaves us with 25,165,824 bits which is finally multiplied by the number of channels (3). We have a final bit-count of 75,497,472 bits.

Converting this to a number that is a little more relatable, we have a file size of 9 megabytes.

If we were shooting video, we would then multiply this by the frame-rate/second. If the video is 45 seconds long, we have 30 frames/second for 45 seconds, multiplying the file size 1350 times.

File Formats

All of this data needs to be stored somewhere to be useful in the future. A file format is simply a container for all of this data. Different file types have different standard ways of storing and organizing that data within the container, and different tags that can be used to describe the data inside.

Typical file types for images include:
- JPG
- TIFF/TIF
- RAW - CR2, NEF, DNG, etc.
- PSD
- BMP
- PNG

Compression

Some of these file types try to compress your data to make the file size smaller.

JPG files use JPEG compression, which is a lossy compression algorithm, meaning the file sizes are much smaller, but at the cost of image quality since some data is thrown out.

Other files, such as TIFF and PSD files will losslessly compress files. This compression doesn't make it as small as JPEG compression, but it also doesn't lose any data that it then has to try and make up later.

In order to complete the very first step, we must look at the sensor itself. Unfortunately, in the manufacturing process dust and other imperfections can get introduced into the silicon of the sensor itself. Every sensor has these defects to differing degrees, so the first thing that must be done once the sensor reads out the data is called Pixel Defect Correction (PDC). This replaces any information that may be missing due to defects. The other step that needs to be done at this point is called Photo Response Non-Uniformity Correction (PRNU). This corrects for any differences in individual pixels sensitivities.

If a digital camera is outputting a “RAW” file, this is all that is done to the data in-camera. So, a RAW file is still processed to a certain point, but leaves much of it for later.

After this point, all the processing steps are still preformed, in the same order. It doesn't matter if your working with a RAW workflow or if you just output jpg files from your camera. The difference is that a RAW file's processing happens when a user opens the file on their computer and they always have that original information to refer back too. With a jpg file the following steps are all performed within the camera and are permanently combined with the original data.

After PDC and PRNU, the image needs to be Neutral Corrected. Commonly referred to as white balancing, this simply makes neutral greys look neutral, compensating for different light sources. Next is the CFA interpolation that was discussed in the previous chapter. This fills in the gaps in our color information due to the color filter array over the sensor. Then we have the Colorimetric Transform and Gamma Correction.

All of these steps happen before the image is even displayed on a monitor and rarely does the end user have any control over any of this.

The last four steps of our imaging pipeline are where programs like Adobe Photoshop actually operate.

These steps include exposure correction, tone correction, sharpening, and lens correction. In other words brightness, color and sharpening.

Once all of these steps are done the file is then saved as one of many options of file formats.
Imaging Pipeline

So, how does this information that comes off the sensor actually end up as an image on a screen and what happens to it before that point?

With film, we knew all the steps. First, you expose the film. Then you develop and fix the latent image. This left a usable strip of film that could either be enlarged in a very similar process, or scanned.

With a digital process, things are a little bit more complicated. So on this page is a graphical diagram of the pipeline to help see how everything feeds into each other.
Chapter Three

Processing
Chapter Two

Input
Scanners

Scanners also work in a very similar manner to a camera. The only difference being that instead of having uncontrolled lighting situations and exposures, a scanner just packs a studio into a box, including the sensor and the light source. Older, larger digital scanners would actually use a single sensor, and a lens to focus the light from the platen (scanner glass) onto the sensor. Newer thinner ones use sensors that are called Contact Image Sensors (CIS). This is nothing more than a series of small long sensors placed next to each other that act as one unit, and this whole sensor array scans across the platen without a lens. Since there is no lens, the sensor must be as close to in contact with the subject as possible in order to get an in focus image, hence the name.

There is another kind of scanner, called a drum scanner that is mainly used for scanning film. It is very large, can be very expensive, and rather slow but provides the highest quality scans possible.
Color Alternatives

Some other applications however require different methods for acquiring this color information. Many dedicated video cameras will have three sensors instead of one. Using a special group of prisms to separate the light channels, you have a sensor dedicated to red green and blue. This allows for very accurate color information and less light lost from color filters making it better in low light. However having three sensors means three times the amount of information to read out at once, so these often have lower spatial resolution and aligning the three images sometimes proves difficult. Sensor layouts like these are often seen in higher end dedicated video cameras.

Another option is the Direct Image Sensor. Developed by Foveon, this sensor relies on the depth that different wavelengths of light penetrate silicon. Using this, the sensor is capable of capturing RGB data at every pixel.

A third option, while being very expensive and slow, is to use a scanning back. Essentially just a scanner placed in the back of a 4x5 view camera, and it scans the projected image out of the camera. This is one of the highest resolution options available, yet exposure times get very long since the sensor must make an exposure at every line.

To start off, we must have an input device. An input device is really simply anything that will capture light and deliver it to us in a usable format. This could be any type of film, scanners, or cameras. Considering the decline in use of analog (film) devices, our input would ideally be in the digital form just for simple ease of use. Therefore, if film is required or requested then there are normally multiple input steps. First, there is the original exposure onto the film and its subsequent chemical processing and then it will be scanned in a flatbed or drum scanner.

Sensors

At their core, almost all digital devices use the same kind of technology. Referred to as a charged coupled device (CCD) or complimentary metal-oxide semiconductor (CMOS) sensor, these two different kinds of sensors have more commonalities than differences.

For an analogy, think of the sensor as a grid pattern of buckets and light is rain for the buckets. As it rains, the bucket collects water and when it stops raining the amount of rain collected by each bucket is recorded. This is how a sensor works too. It is a grid pattern of pixels that collect photons as the exposure is made. Once the
exposure is over, the pixels report how much light hit them, and this is then displayed as that pixel's brightness. In order to get an image, all of the pixels' brightness are combined together and displayed in a two dimensional grid.

The main difference between CCD and CMOS sensors is how the pixels report how much light hit them. On a CCD sensor, each number is passed along just like a bucket brigade from one pixel to the next until they've all been read-out. With a CMOS sensor, every individual pixel has a direct connection to the output.

As with all things, compromises are made in each. CCD sensors are normally considered to have lower “noise” in an image, however they also require significantly more energy than a CMOS sensor. CCD sensors are also more costly to manufacture since they need a special manufacturing plant. CMOS sensors require less electricity, but are often noisier and are also less sensitive to light, however they are much cheaper to produce since most computer chip manufacturing plants are already to produce CMOS chips.

**Color - CFA**

So, if a digital sensor simply records how much light hits each pixel, how then do we get color with a digital camera?

Computers make smart guesses and give us colors.

For a longer answer, our bucket analogy must be expanded. What if each bucket was restricted to only recording one specific type of precipitation, and we want to record rain, hail and snow. So one-third of the buckets would record rain, one-third would record hail, and the last third would record snow. Now, if those buckets are arranged in a certain way, a computer algorithm could guess reasonably well how much hail and snow should have fallen in a rain bucket based off the rain and hail buckets around it.

This is exactly what a sensor does. Each pixel can only record one light level and we want to record at least three, red green and blue. So, half of the pixels are green since humans are most sensitive to green light, and then blue and red cover a quarter of the pixels each. This pattern is known as the Bayer Pattern.

These filters are then laid out in such a way that will allow a computer to calculate the missing values. This filter array over the sensor is known as a Color Filter Array (CFA) and the process of guessing the two missing values for each pixel is called CFA interpolation.

This is how almost all input systems work, such as cameras and scanners. Cameras simply place a shutter and lens in front of the sensor and a scanner is nothing more than a studio in a box. Where a camera records light that changes, a scanner just places light sources alongside the sensor to illuminate the scanned object.

Using CFA patterns and guessing works pretty well, however two-thirds of an image is made-up color information. This is a pretty major drawback, but it is fast and cheap.