MAS.963: Computational Camera and Photography **Fall 2009**

Epsilon Photography-Improving Film-like Photography

Prof. Ramesh Raskar September 25, 2009

Epsilon Photography: Improving Film-like Photography (Guest Lecture by Ankit Mohan)

This session was a guest lecture by Ankit Mohan [3], covering the topic of Epsilon Photography. The lecture consisted on a swift review of film-like optics, then a description of prior work which lead to the field of Epsilon Photography such as Auto-Focus, Auto Exposure and Bayer Demosaicing which is the method used for color sensing in digital still cameras. Subsequently Epsilon Photography was explained, with multiple examples for Epsilon over time, sensors, pixels, and of course the combination of multiple axis. The vast amount of examples shown did not only describe the basics of Epsilon Photography but also the different methods that have been developed to reach a certain goal, which provided a broad overview of the current state of research and tools at hand. The talk was complemented by a lecture by Roarke Horstmeyer describing his work on a Single-shot Multi-domain Camera.

1.Review: Film-like Optics

To understand Epsilon Photography and Computational Photography in general we first have to fully understand how film-like optics in conventional photography work [8].

1.1. 'Pinhole' Model

The 'Pinhole' model consists of a scene on the left, a center of projection, and the 2D Sensor on the right which forms an image of the scene, a single ray of light from each point in the scene. (not shown here: the box which contains the sensor and has one ideally infinitely small hole). With a Pinhole camera everything is in focus, this is called an "infinite depth of field". In this model depiction the aperture size is infinitely small (allows only one ray of light to come through), thus everything is in focus.

Most objects around us are actually diffuse (technical term: lambertian), which means that if you have light on an object it reflects light in all directions. All rays that come out of a point on an object have roughly the same intensity. The opposite would be a specular object which reflects light primarily in one direction. So when you have a pinhole camera it looks pretty similar to how it looks to the eye (captures most of the information from the scene).

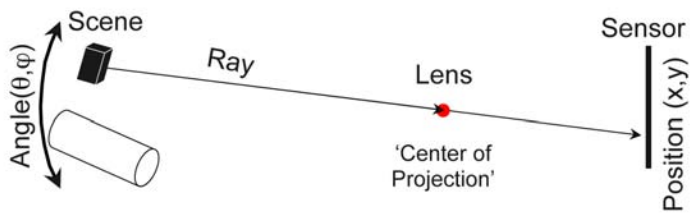

Figure 1: the 'Pinhole' model

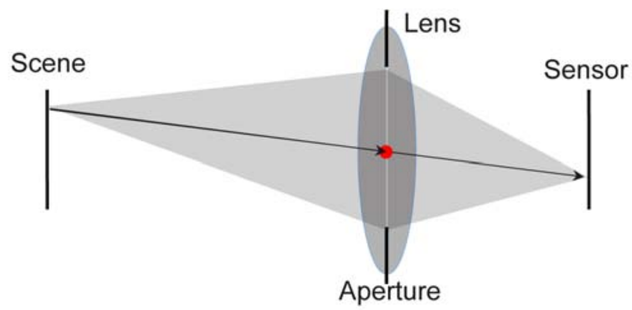

Figure 2: An idealized version of lens photography (this figure ignores diffraction)

1.2. Lens Photography

When a lens is added to this model, something slightly different happens: The lens integrates a bundle of rays over an angular extent. You have rays from one point in a scene, and the lens is able to capture a whole cone of rays and project them onto the 2D sensor. In this drawing all points that come from one point in the scene get focused on the sensor. This is how a modern camera works.

1.3. Depth of Field and Aperture

Unlike a pinhole camera, when working with a lens you have only one plane in the scene which gets imaged on the sensor correctly, and this is defined by the focal length and the distance between lens and sensor. So if you have a finite aperture lens you have a finite depth of field. A lens can only focus on precisely one distance, but the sharpness decreases gradually so for a finite range behind and in front of the focused plane we still perceive it as in focus. When the bundle of rays is not perfectly focused on one point on the 2d sensor

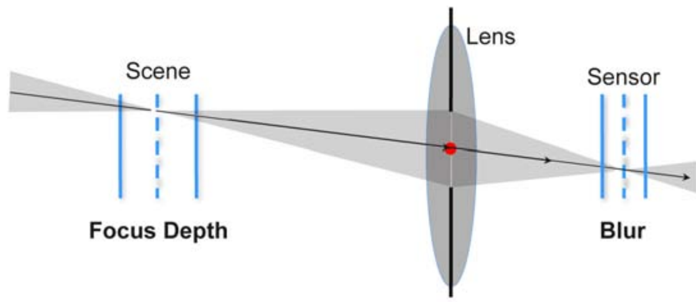

Figure 3: A smaller aperture creats a deep DOF

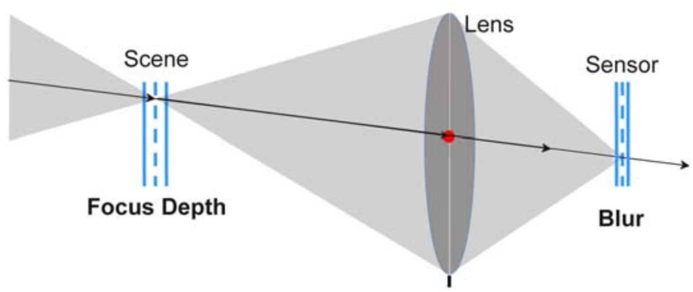

Figure 4: A larger aperture creates a shallow DOF

it forms a round disc. If that disc is equal or smaller in size than a pixel it is still perceived as "in focus".

The size of the depth of field depends on the size of the aperture: The smaller the aperture, the "deeper" is the depth of field, the larger the aperture, the more shallow it is. This can be easily described when looking at the drawings of the pinhole camera and two different aperture sizes. With an ideal pinhole camera only one ray is perfectly depicted. With a slightly larger aperture you get a focus depth, but because the angular cone of light is only slightly larger there is a wide range of depth of field. If you have a really large aperture the depth of field is smaller, as the angular cone is bigger. Depth of field is used by photographers depending on the application. For example in a portrait you want to focus on the face or the eyes, whereas in landscape photography you might want to have everything as focused as possible. Most modern cameras give the option to set the aperture size, which defines the depth of field.

2. Towards Epsilon Photography

Even before the emergence of Computational Photography there have been years research and development to improve the different aspects of traditional photography. Thus it makes

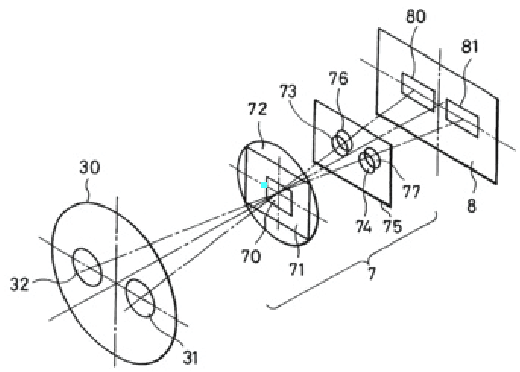

Figure 5: Original Diagram from the Auto Focus Patent by Minolta

sense for us to learn about Auto Focus, Auto Exposure and the way color sensing works in digital cameras.

2.1. Auto Focus

As you only have one plane that is actually in focus you would want to be able to select which plane is in focus, and most modern cameras do this now using Auto Focus. There are different techniques for Auto Focus. The most common one today is the phase-based auto focus (Minolta). In this method a rangefinder is used, with a baseline corresponding to the diameter of the lens [4]. The rangefinder uses two samples, from two corners of a lens, and compares the light intensity of both, so you could almost say that it is a stereoscopic process. If both images have a very similar light intensity the object is in focus. If not, there is a phase mismatch, and by observing the phase-mismatch one can determine in which direction, and how far, to move the lens. Thus this method is a fast single-shot method which is currently used in most digital SLR cameras. Unfortunately this specific hardware is necessary for this method. In most compact digital cameras another technique called the contrast-based auto focus is used. This process simply looks at one image, and compares it to another while moving the lens back and forth. The image with the highest contrast is assumed to be in focus. This means the process basically searches through multiple images until it finds the maximum contrast, and thus this is not a single-shot method and is slower. On the other hand this process does not require dedicated hardware.

Compact film cameras used ultrasonic or IR based auto focus systems, but these are not very efficient or accurate and are not really being used today anymore. Another technique is a rangefinder camera, which is a separate unit from the imaging camera. With the previous methods the focus was determined by information coming through the lens. With this system you have a separate unit which can determine the focus. This is usually done manually. Auto focus is interesting for us because there has been a wide range of prior engineering and inventive work to achieve this goal, even before the field of Computational Photography came into being.

Image removed due to copyright restrictions.

Figure 6: A diagram of the placement of Nikon's matrix-metering scheme

2.2. Auto Exposure

Another problem to solve is exposure, or in other words: How to map the whole light spectrum perceptible by human vision (from daylight to a lightbulb). It ranges from 10-6 to 10+8 cd/m2 but we have only 0-255, so 8 or maybe 12 bits to actually capture this wide range of information. So we have to deal with the limits of dynamic range by deciding which exposure to use on a camera. Exposure is defined by the three parameters: Shutter speed, Aperture size and light sensitivity of the sensor or the film (ISO). For a long time the photographer had to set these parameters manually in relation to each other. In the mid-eighties Nikon has changed this drastically by introducing an Auto-Exposure system. The system, called the Nikon matrix-metering scheme, consists of an array of lightmeters in five different zones that are placed in top of the pentaprism. Through mirrors the light reaches the sensors before it reaches the viewfinder and exposure settings can be determined dynamically [5]. Both Auto-Focus and Auto-Exposure lead to the full Auto mode of contemporary cameras.

2.3. Color sensing in digital cameras

Most digital cameras have a Bayer filter. This is an array of color filters placed in front of the imaging sensor, for the three colors RGB (two green for every red and blue). Thus the resulting image consists of pixels which alternate between color representations of the three colors. A process called Demosaicing is an algorithm that generates a high-resolution image in color from this information. There is another method, Foveon sensing, where they do the same thing in depth, resulting in a true full resolution color image. Considering the whole electromagnetic spectrum, only a very small range is visible to the human eye, which is between 400-700nm. We are imaging this range with just the 3 color channels RGB only because the photoreceptors in the human eye work in a similar way. But actually there could be many more channels even within that range.

2.4 Chromaticity Diagram and color spaces

The chromaticity diagram depicts all color perceptible to the human eye. The edges define pure colors, and if you mix between two colors within this spectrum all intermediate colors

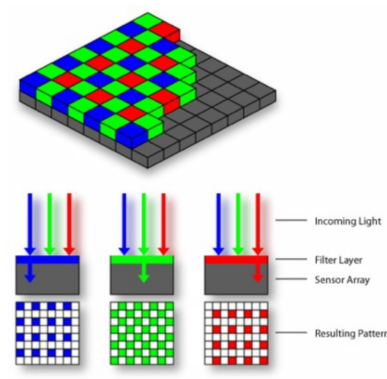

Figure 7: The bayer filter is placed in front of the 2D sensor

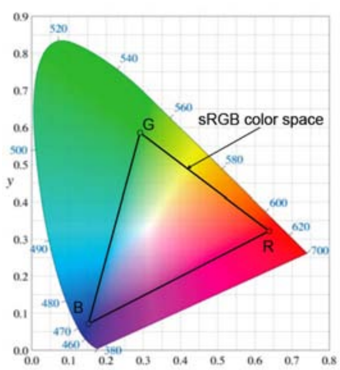

Figure 8: The CIE 1931 Chromaticity Diagram

sit on a straight line connectin both points. This diagram is used to show color spaces of devices such as photo sensors, displays or projectors, I.e. the sRGB color space [7].

3. Epsilon Photography

Epsilon Photography is the process of capturing multiple photos, each with slightly different camera parameters. These can be settings for exposure, the color spectrum, focus, the camera position or the scene illumation. Changing the parameters can occur over time (taking multiple images consecutively), over multiple sensors, over multiple pixels and of course a combination of multiple axis (I.e. time and sensors). Epsilon Photography was coined by Prof. Ramesh Raskar [1], and the goal is to improve the way film-like cameras work today. Thus it forms the low-level vision within the field of Compuational Photography, alongside Coded Photography and Essence Photography.

3.1. Epsilon over time (bracketing)

This is one of the most common form of Epsilon Photography and it simply means taking multiple images over time and combining them to one. Actually even one of the first color images was produced with a bracketing process: At the turn of the century Prokudin-Gorskii made not one, but three portrait photos of each person, each time using a different color filter (RGB). These were later recombined using three overlayed projections. The color wheel within DLP projectors can be seen as epsilon projection, as it rapidly turns and the projector rapidly projects the three different colors, and this happens so fast that we perceive it as one image.

HDR (High Dynamic Range) capture

Even when you have the correct exposure you can not capture everything the scene contains, because it can contain a too big range of brightness difference for the camera to capture, so the contrast ratio would just be too big. One way around this problem is to take multiple images with varying exposure settings and combining them to one image. There is research on the best way to combine the images together, how to display it and also how to compress it back to an 8 bit image.

3.2. Epsilon over sensors

Epsilon Photography over sensors is usually done in two ways: Either multiple cameras or multiple sensors within one camera. One example for the latter is the 3CCD imaging system, usually used in current high quality video cameras. The system consists of a prism that, depending on the index of defraction, divides the light into the three colors (RGB) and guides it to three different sensors. All three sensors are optically within the same distance to the scene and all three images get taken simultaneously. Afterwards they get recombined to one image again. McGuire et al, 2005 applied the same principle but using eight different cameras, all optically in the same distance to the scene. This method allowed each of the eight sensors to have different settings, such as focus, exposure, etc. Another way to do this is using camera arrays. The difference is of course that you also have not only an epsilon variation over the different sensors, but also an epsilon over the camera position.

3.3. Epsilon over pixels

Here the process is defined by different pixels actually capturing different information. Bayer Mosaicing as explained above is of course an example for this. In 2001Schechner and Nayar proposed a principle called Generalized Mosaicing, which elaborates on the technique currently used for panorama photography, where the camera is moving or changing it's positions and the adjacent pixels are stitched together to form a great picture. But in Generalized Mosaicing a one dimensional filter is placed in front of the whole sensor (for example gradually varying polarizations from left to right, a color spectrum or a varying exposure from left to right) and the filter is not moving. After the camera has completed the motion, the complete set defined by the filter has been captured with each pixel, and the pixels can be recombined to one image at will.

Single-shot Multidomain Camera-Roarke Horstmeyer

Roarke Horstmeyer presented his work on a single-shot Multidomain camera (Horstmeyer, Euliss, Athale, The MITRE Corp, Levoy, Stanford University, 2009). The main approach is to build a camera that can capture many different aspects, such as the polarization, multispectral information and a high dynamic range with one single shot imaging process. This is achieved by combining a pinhole array in front of the 2D sensor (similar to a Light-Field camera) with an array of the imaging pupil plane of the lens. This approach is different from Bayer Mosaicing because now each pinhole creates an image through the filter array.

Multiple experimental results were shown, such as using 6 different filters (RGB and 3 different polarizations), 16 filters (6 colors, IR and polarizations) and using a color spectrum as filter which resulted in 25 spectral channels per pixel. The advantage is that all these properties are imaged at the same time, and the filters themselves are easily interchangeable. The disadvantage with this method is decrease in spatial resolution, as it is defined by the numbers of pinholes in the array divided by the number of filters.

3.4. Epsilon over multiple axis

LCD Adaptive Light Attenuator

Another method to create an HDR image over time is to place a LCD in front of the sensor. There is an intial capture, the image is analyzed, and when there are certain pixels are overexposed the LCD is adjusted. In the next capture the LCD darkens in those areas, to compensate for the brightness. Thus the sensor still has only a low dynamic range but in combination with the LCD you can create a high dynamic range image. The resolution for

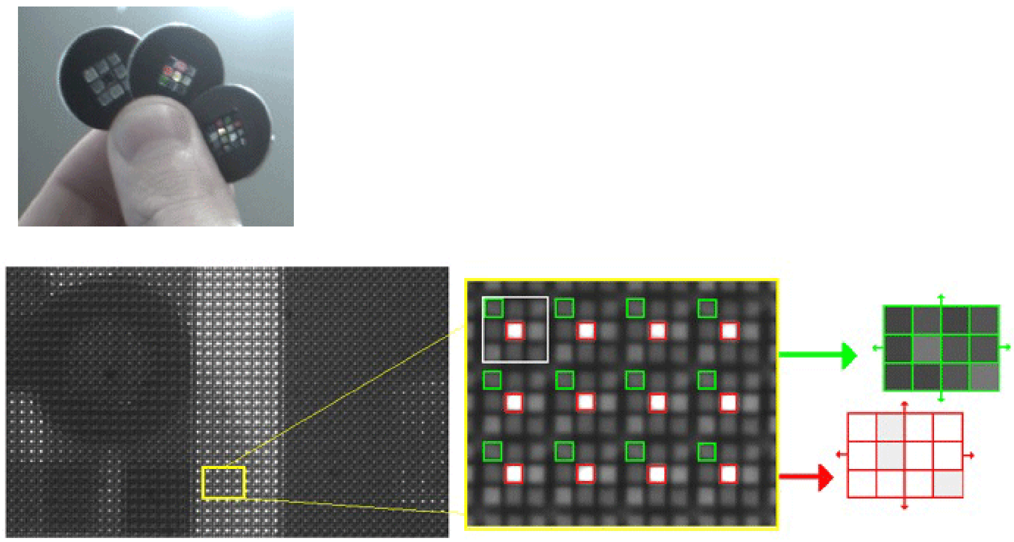

Figure 9: Each pinhole creates an image of all filters, one filter per pixel

the LCD is not required to be as high as the actual sensor. A similar method was used for a high dynamic range display [Seetzen, Heidrich, et al, 2004]. Here also a LCD is placed in front of a projector.

Extending depth of field

There are many applications where you want a very deep depth of field. Traditionally this is only possible with a very small aperture, which results in very little light coming into the camera. Multiple methods have been proposed, especially in the field of microscopy, to extend the depth of field while at the same time keeping the aperture reasonably large. One of these is called Focal Stacks, where multiple photos with different focal planes are taken and combined algorithmically to create an image that has a deeper depth of field [Fusion, Agrawala et al., Digital Photomontage, Siggraph 2004]. Another way of extending the depth of field is using a light field camera (not explained here), which allows the extraction of the focal stack. This again can be recombined to create an image with a larger DOF.

Decreasing depth of field

Another goal people try to achieve is to decrease the depth of field, for example for compact digital cameras, as lenses in these cameras, and thus the aperture size is very small. Large digital SLR cameras with larger lenses can create images with a shallow depth of field (object in foreground sharp, background out of focus) naturally. This is a hard problem because it is hard to find out from one image which part is in focus or not, and of course the focus decreases gradually going away from the plane of focus. Bae and Durand (2007)

Vaish, V., et al. "Using Plane + Parallax for Calibrating Dense Camera Arrays." *Proceedings of CVPR 2004*. Courtesy of IEEE. Used with permission. © 2004 IEEE.

Figure 10: Left: Photo from the same position with a normal camera. Right: Focused on plane behind the bush, everything else blurs out of focus

proposed an image processing technique (not epsilon phtography) that tries to estimate a 3D "defocus map", depicting the different depths within this scene. By combining this map with the image using a spatially varying blur filter an image with shallow depth can be generated.

Synthetic Aperture Photography

This method is more generally applicable but can also be used to create a shallow depth of field. The concept is simple: Instead of having one large aperture lens you can simulate it by having many small lenses and combining the resulting multiple images, generated by a larger ray bundle, into one. In their approach, Vaish et al. [10] used this technique for focus adjustment: Their system, consisting of a camera array, was able focus on a plane behind a bush, which in a normal photo was in front of the plane of interest. Because the DOF was so shallow the bush blurred out of focus, only the plane of interest was in sharp focus and thus became visible. Uncalibrated synthetic aperture photography (Kusumoto, Hiura, Sato, 2009) is a similar technique, but instead of a rigged camera array, here multiple photos from one camera are taken from multiple locations, and recombined using image processing to create a defocused photo with a shallow DOF. Obviously this technique only works if the scene is not changing.

Image Destabilization

Proposed by Mohan, Lanman et al. [11] , here the idea is to solve the same problem more optically. Lens and sensor are moved synchronously on one axis during the exposure. If the concept is considered using the Pinhole model, the center of projection is moving, thus the points that get exposed on the 2D sensor are correlate with that motion (producing a blur). The interesting aspect exploited here is that the distance this projection of one point travels varies depending on the distance from center of projection to point in scene, so it is possible to calculate that distance. Using this technique you can simulate a larger aperture lens and it can fit into a smaller camera body-but of course the hardware to create this physical motion is necessary.

Increasing Spatial Resolution-Superresolution

One important goal of Epsilon Photography is to increase the spatial resolution. The most common way of doing this is called Superresolution. Proposed by Kopf et al in 2007. The main principle is having a camera at a fixed position and moving the sensor by a fraction of the pixel size while continuously taking images. Because the sensor is in different locations for each exposure, and the change in location is only by a fraction of the pixel size, a different part of the scene is depicted with each image. These images can be combined by finding correspondence points and stitching to create a higher resolution image. This method can create a 3.5 giga pixel image, and of course you get a higher field of view. One issue with this method is precision, the minimal motion is usually less than a micron and it is used in high-end medium format cameras. Some of the challenges with this are high dynamic range issues, you receive many images with many different brightnesses. Also it makes it more difficult to find correspondence points. All these issues have to be addressed using image processing. Another way of creating superresolution is by using an array of cameras, which then is not epsilon over time but over camera location.

Increasing Temporal Resolution-High Speed video

Nowadays you can buy expensive cameras that can record video with up to 1000 frames per second. But instead of having one expensive camera to do this one can use multiple cameras that each take images at a normal framerate but are connected by synched triggers. Thus an array of cameras can simulate a higher framerate than each individual one can provide.

References

[1] Ramesh Raskar *http://web.media.mit.edu/ raskar/*

[2] Ramesh Raskar *http://web.media.mit.edu/ raskar/photo/*

[3] Ankit Mohan *http://web.media.mit.edu/ ankit/*

[4] ıthttp://en.wikipedia.org/wiki/Autofocus

[5] *http://en.wikipedia.org/wiki/Autoexposure*

[6] *http://en.wikipedia.org/wiki/Electromagnetic*$_s$*pectrum*

[7] *http://en.wikipedia.org/wiki/CIE*$_1$*931*$_c$*olor*$_s$*pace*

[8] Goodman,J.W. , 1968 *An Introduction to Fourier Optics*

[9] Aseem Agarwala, Mira Dontcheva, Maneesh Agrawala, Steven Drucker, Alex Colburn, Brian Curless, David Salesin, Michael Cohen *Interactive digital photomontage*, in SIGGRAPH 2004 Papers

[10] Vaibhav Vaish, Marc Levoy *Synthetic aperture imaging using dense camera arrays*, January 2007, Stanford University

[11] Ankit Mohan, Douglas Lanman, Shinsaku Hiura, Ramesh Raskar *Image Destabilization: Programmable Defocus using Lens and Sensor Motion*, in proceedings of IEEE ICCP 2009