Multi-Aperture Photography

Paul Green[1] Wenyang Sun[2] Wojciech Matusik[2] Frédo Durand[1]

[1]Massachusetts Institute of Technology
Computer Science and Artificial Intelligence Laboratory

[2]Mitsubishi Electric Research Laboratory

Abstract

The emergent field of computational photography is proving that, by coupling generalized imaging optics with software processing, the quality and flexibility of imaging systems can be increased. In this paper, we capture and manipulate multiple images of a scene taken with different aperture settings (f-numbers). We design and implement a prototype optical system and associated algorithms to capture four images of the scene in a single exposure, each taken with a different aperture setting. Our system can be used with commercially available DSLR cameras and photographic lenses without modification to either. We leverage the fact that defocus blur is a function of scene depth and $f/\#$ to estimate a depth map. We demonstrate several applications of our multi-aperture camera, such as post-exposure editing of the depth-of-field, including extrapolation beyond the physical limits of the lens, synthetic refocusing, and depth-guided deconvolution.

Keywords: Computational Imaging, Optics, Image Processing

Figure 1: *Photographs of our prototype optical system to capture multi-aperture images. The system is designed as an extension to a standard DSLR camera. The lower right image shows a close-up of the central mirror used to split the aperture into multiple paths.*

1 Introduction

The new field of computational photography is fundamentally rethinking the way we capture images. In traditional photography, standard optics directly form the final image. In contrast, computational approaches replace traditional lenses with generalized optics that form a *coded* image onto the sensor which cannot be visualized directly but, through *computation*, can yield higher quality or enable flexible post-processing and the extraction of extra information. Some designs use multiple sensors to extract more information about the visual environment. Other approaches exploit the rapid growth in sensor spatial resolution, which is usually superfluous for most users, and instead use it to capture more information about a scene. These new techniques open exciting possibilities, and in particular give us the ability to modify image formation parameters *after* the image has been taken.

In this work, we focus on one of the central aspects of optical imaging: the effects of a finite aperture. Compared to pinhole optics, lenses achieve much higher light efficiency at the cost of integrating over a finite aperture. The choice of the size of the aperture (or $f/\#$) is a critical parameter of image capture, in particular because it controls the depth of field or range of distances that are sharp in the final image. Depending on the type of photography, more or less depth of field can be desirable. In portraits, for example, shallow depth of field is desirable and requires a wide physical aperture. Unfortunately, many users do not have access to wide aperture cameras because of cost, in the case of SLRs, and limited physical sensor size, in the case of compact cameras. Photographs of a few characters are even more challenging because the aperture diameter should be large enough to blur the background but small enough to keep all subjects in focus. In summary, aperture size is a critical imaging parameter, and the ability to change it during post-processing and to extend it beyond the physical capabilities of a lens is highly desirable.

Design goals We have designed an imaging architecture that simultaneously captures multiple images with different aperture sizes using an unmodified single-sensor camera. We have developed a prototype optical system that can be placed between the camera and an unmodified lens to split the aperture into four concentric rings and form four images of half resolution onto the camera sensor.

We designed our optical system to meet four goals:

- Sample the 1D parameter space of aperture size and avoid higher-dimensional data such as full light fields.
- Limit the loss of image resolution, in practice to a factor of 2×2.
- Design modular optics that can be easily removed in order to capture standard photographs.
- Avoid using beam splitters that cause excessive light loss (e.g., [McGuire et al. 2007]).

One advantage of our design is that the images we capture can be added directly, to form the image corresponding to various apertures and does not require non-linear processing and image analysis. More advanced post-processes can be performed using depth from defocus.

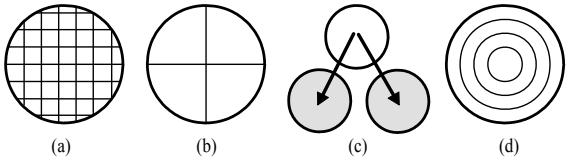


Figure 2: *(a-c) A pictorial representation of previous methods of splitting the aperture. (a) The standard Light field camera design [Adelson and Wang 1992; Ng 2005; Georgeiv et al. 2006]. (b) The splitting used by Aggarwal and Ahuja [2004] for high dynamic range imaging. (c) beam splitters [Mcguire et al. 2007], and (d) our decomposition.*

2 Related Work

We focus on work related to modifying the aperture of a camera, which includes both camera systems that permit the capture of richer data streams and image processing algorithms that take advantage of this captured information.

Plenoptic cameras instantaneously capture the full light field entering the optical system. Various designs have been investigated and implemented [Adelson and Wang 1992; Okano et al. 1999; Naemura et al. 2001; Ng 2005; Georgeiv et al. 2006]. These designs vary in size and optical components, but, in principle, plenoptic cameras trade spatial resolution to capture directional information about the rays entering the optical system. This also can be seen to split the main aperture into a number of rectangular areas and form a separate image from each of these sub-apertures(Fig. 2(a)). A typical drawback of these approaches is a severely reduced spatial resolution, where the grid subdivision of the aperture results in a reduction that is quadratic in the number of samples along one axis. An advantage of these approaches is that the final image can be a simple linear combination of the recorded data [Ng 2005]. Non-linear reconstruction can afford better resolution trade-offs, but is more prone to artifacts.

Another interesting way of splitting the light entering an optical system is to use a pyramid mirror placed behind the main lens [Aggarwal and Ahuja 2004]. This effectively subdivides the aperture into "pie slices" and each of these sub-apertures is captured using a separate sensor (Fig. 2(b)).

Perhaps the most common way of splitting the light entering an optical system is to use beam splitters to replicate the optical path (Fig. 2(c)). Prisms and half-silvered mirrors are typical elements used to perform this task. In this context, 3-CCD cameras use a dichroic prism to split the light and create three copies of the image, each with a different spectral band. Many other designs have been investigated. In particular, McGuire et al. use different aperture and focus settings to perform matting [2005]. Watanabe et al. have demonstrated a real-time depth from defocus system that uses beam splitters and active illumination [1996]. We have considered designs with beam splitters to decompose the aperture, but they usually require multiple sensors and lose light because they need to rely on occlusion by a mask to select a sub-region of the aperture.

Hasinoff and Kutulakos use a brute force approach by capturing all possible combinations of aperture and focus settings for use in a depth from focus method [2006]. This method produces very high quality depth maps but requires several hundred exposures.

Applications of splitting the aperture include: extending dynamic range [Aggarwal and Ahuja 2004; Narasimhan and Nayar 2005], computing depth [Adelson and Wang 1992; Farid and Simoncelli 1996; Hiura and Matsuyama 1998], alpha matting [McGuire et al. 2005], multi-spectral imaging [Narasimhan and Nayar 2005], high-speed imaging [Harvey 1998], changing viewpoint [Okano et al. 1999; Naemura et al. 2001; Ng 2005], digital refocusing [Isaksen et al. 2000; Ng 2005; Georgeiv et al. 2006], synthetically changing depth of field [Georgeiv et al. 2006], and extending depth of field [Ng 2005; McGuire et al. 2007].

3 Optical Design

3.1 General Principle

The optical system must accomplish two tasks simultaneously: 1) split the circular aperture of the main photographic lens into a central "pinhole image" and several concentric rings and 2) re-sort and image the light rays from the "pinhole" and rings onto the imaging sensor of a digital camera.

We use a relay system to image the physical aperture diaphragm of the photographic lens to a plane outside of the lens, called the exit pupil [Hecht 2002]. The exit pupil is then divided into a pinhole and a number of concentric rings. Refractive/reflective optical elements are used to steer the light rays passing through different rings. Finally, additional lenses are used to form images on a single imaging sensor.

3.2 Our Design

Our optical design for splitting the aperture into a central pinhole and a set of concentric rings is conceptually similar to a Cassegrain lens. A schematic is shown in Fig. 3. The self-contained optical assembly is placed between a regular photographic lens and the camera body. The entire optical package includes the relay optics tube, a 4-way multi-reflection mirror to split the lens aperture, and four sets of image forming mirrors and lenses. We chose to divide the full photographic lens aperture into *N=4* sub-aperture areas because this division achieves a good trade-off between the loss of sensor resolution and the ability to perform our proposed post-exposure edits. We use a 12.8MP *Canon EOS-5D* digital SLR camera, and achieve around 3MP spatial resolution for each of the four images. From four images we are able to acquire depth maps, interpolate and extrapolate depth-of-field, and synthetically refocus.

Relay optics are necessary for two reasons. First, to relay the intermediate image formed by the photographic lens to the camera's sensor. More importantly, relay optics are necessary to image the physical aperture diaphragm of the photographic lens out of the lens barrel, i.e., forming a new exit pupil at the *4*-way multi-reflection mirror. From the conjugation relation between the object and image [Hecht 2002], we know that splitting the exit pupil is equivalent to splitting the physical aperture itself. By using the *4*-way multi-reflection mirror at the new exit pupil, we reflect the incident light rays to four different directions according to where it passes through the aperture. For example, the size of the central pinhole mirror is equivalent to the lens aperture size of $f/8$. Therefore, all the rays which pass through the virtual $f/8$ aperture are steered along the optical path denoted in red, as shown in Fig. 3. Please note: the red, yellow, green and blue colors in Fig. 3 are used only to distinguish the four different light propagation paths, and are not related to any real color filtering/modification. The outer radius of the other three rings are chosen to correspond to virtual aperture sizes of $f/5$, $f/3.7$ and $f/2.8$, respectively, from the smallest ring to the largest ring. The corresponding folding mirror at this optical path reflects the light back in the direction of the camera. An

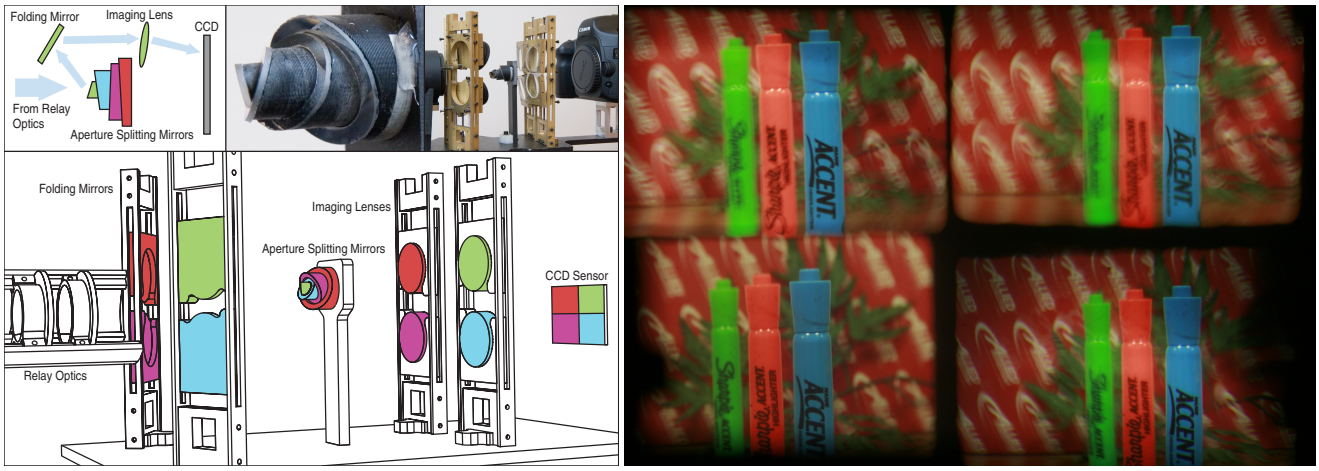


Figure 3: *A schematic diagram of our optical system and a sample image taken with the camera. Our design consists of a main photographic lens imaged through relay optics, and split using a set of tilted mirrors. The relay optics produce an image of the main lens' aperture onto the aperture splitting mirrors. The diagram is color coded to display the four separate optical paths. The right image shows data acquired from our camera. Each quadrant of the sensor captures an image from a different aperture ring.*

imaging lens is used between the folding mirror and the camera, to reduce the imaging distance and ensure that the final image size is reduced to 1/4 of the size of the camera sensor. Another function of the imaging forming lens is to correct the image plane tilt according to the Scheimpflug principle [Ray 1988]. As one can see from Fig. 3, the optical axes of all four optical paths are deviated from the original photographic lens optical axis, which is also the normal direction of the camera sensor. The four sub-images are perpendicular to their corresponding optical axes, thus do not align with the camera sensor plane. This problem can be corrected by tilting the image forming lenses to their proper positions according to the Scheimpflug Principle.

The *4*-way multi-reflection mirror used to divide the lens aperture is made by cutting a commercial first surface mirror to the shape of an elliptical ring and gluing it onto a plastic mirror base. An image of the mirror base is shown in Fig. 3. The angles of the directions to which light is reflected must be large enough to ensure the folding mirrors and their mounts do not interfere with the relay optics tube or block the incident light from the relay optics. However, this angle cannot be too large, as the larger the angle, the more the light deviates from the original optical axis which can cause several field related optical aberrations such as coma and astigmatism. Additionally, large angles increase the possibility for vignetting from the camera mount opening to occur. Finally, larger reflecting angles at the aperture-splitting mirror increase the amount of occlusion due to splitting the aperture. Further details are discussed in Section 3.4.

We have designed the relay optics to extend the exit pupil $60mm$ behind the relay optics tube. The *4*-way multi-reflection mirror is placed at this location. The innermost pinhole mirror and the small ring mirror are tilted 25^o to left (around x-axis), and 18^o up and down (around y-axis) respectively. The two largest rings are tilted 18^o to the right (around x-axis), and 16^o up and down (around y-axis) respectively. The tilt angle for this arm is slightly smaller because these two rings are farther behind the relay optics. To generate the same amount of lateral shift at the position of the folding mirrors, the desired deviation angle is smaller.

The position of each folding mirror is determined by the tilting angle of the corresponding pinhole or ring mirror. The folding mirrors and imaging lenses are mounted on four, six-degree of freedom kinetic mounts, which ensure that the mirrors and lenses can be configured to the correct position and angle to form four sub-images at the four quadratures of the camera sensor (See Fig. 1 and 3).

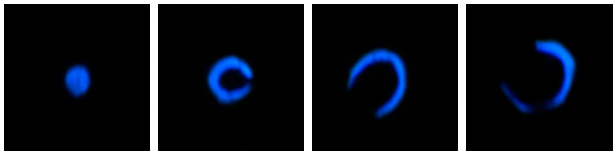
3.3 Calibration

Our system needs to be both geometrically and radiometrically calibrated. Because we have used stock optical elements, and built all the mounts and enclosures, there are significant distortions and aberrations in each image. We have observed spherical field curvature, radial distortion, tilt in the image plane, and variations in the focal length of each ring image (due to slight differences in the optical path lengths resulting from imprecise alignment and positioning of the mirrors and lenses). To geometrically calibrate for distortions between ring images, we photograph a calibration checkerboard and perform alignment between images. Through calibration we can alleviate some of the radial distortion, as well as find the mapping between images. In addition, imaging an LED (see Fig. 4) was very useful to perform fine scale adjustments of the mirror and lens angles.

We radiometrically calibrate the rings by imaging a diffuse white card. This allows us to perform vignetting correction as well as calculate the relative exposures between the different rings. Finally, we apply a weighting to each ring, proportional to its aperture size.

3.4 Occlusion Analysis

The four reflecting surfaces on the *4*-way multi-reflection mirror are tilted to different directions. They are placed in a spiral-step configuration as shown in Fig. 3. Thus each of the outer rings is partially occluded by its neighboring inner ring's extruded supporting base. Consequently, the aperture of the central pinhole area is unaffected, but a small portion of each of the other three ring apertures is occluded. The occlusion can be reduced by arranging the sequence of the four reflection surfaces, such that the normal direction transition between each of the adjacent surface pairs is minimized. For example, as shown in Fig. 3, the angle between the normal direction of the center pinhole and that of the first ring is 36^o, but the angle between that of center pinhole and the second largest ring is 49.1^o.



(a) (b) (c) (d)

Figure 4: *Point Spread Functions of our system captured by imaging a defocused point source imaged through the apertures of (a) central pinhole, (b) the first ring, (c) the second largest ring, and (d) the largest ring. The horseshoe shape of the outer rings is caused by occlusion.*

This arrangement produces less occlusion than if the reflection direction of the first and second rings is swapped.

We captured the images of the occluded apertures by probing the camera system with an LED point source at a position off the plane of focus. Figure 4 demonstrates that occlusion is low in our design.

4 Applications

In the previous section we described an optical system to capture images taken from $N = 4$ annular apertures simultaneously. Using our representation, we can synthesize a sequence of N images of different aperture sizes by accumulating the rings. In a single exposure, our technique can generate multiple images of the same scene, each as if taken with a different aperture setting. This set of multiple images then can be used to recover a *defocus gradient map*, which measures at each pixel the change in defocus blur as a function of aperture size. Our defocus gradient map is very similar in concept to a traditional depth map, and in fact we could compute depth from the sequence of aperture images using standard depth from defocus algorithms [Chaudhuri and Rajagopalan 1999]. The defocus gradient map is integral to accomplishing sophisticated operations, such as extrapolating shallow depth of field beyond the limits of the largest aperture, changing the apparent plane of focus, and increasing image sharpness using a depth guided deconvolution scheme.

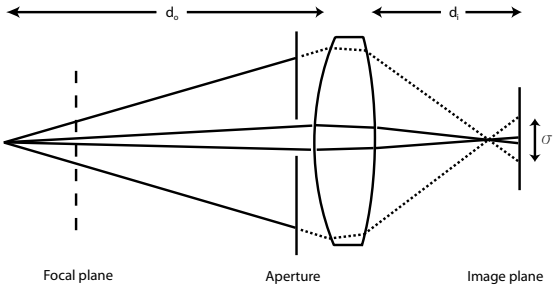


Figure 5: *The relationship between aperture size and depth of field. The size σ of the defocus blur is dependent on the aperture size. Consequently, the depth of field is also dependent on the aperture size.*

4.1 Defocus Gradient Map

Assuming that our scene is composed of planar patches parallel to the image plane, we can approximate defocus blur over each patch as a convolution, where the filter size is determined by the patches' distance from the plane in focus. In his original work on depth from defocus, Pentland [1987] derives an equation relating the object distance d_o to internal camera parameters and the defocus blur kernel size σ (see Fig. 5):

$$d_o = \frac{f d_i}{d_i - f - \sigma N}, \tag{1}$$

where f is the focal length, d_i is the distance between the lens and the imager plane, and N is the f-number (the ratio of the focal length to the diameter of the lens). Solving for σ we have:

$$\sigma = \frac{(d_i - f)d_o - f d_i}{N d_o}. \tag{2}$$

Substituting $M = [(d_i - f)d_o - f d_i]/d_o$ and $g = 1/N$, we can rewrite Eq. 2 in the simple linear form:

$$\sigma = Mg, \tag{3}$$

where it is trivial to see that M is the derivative of σ with respect to the inverse f-number $1/N$. The utility of Eq. 3 is that if M is known, the blur kernel size can be calculated for an arbitrary f-number. We call our estimate of M at each pixel the "defocus gradient map".

The defocus gradient map measures the change in size of blurring kernels as a function of aperture size. The defocus gradient map is related to the distance of an object from the plane of focus. An object on the focus plane will always be sharp (hence its blurring kernel will be zero for all aperture sizes). An object away from the focus plane will become blurrier as the aperture size is increased, and in particular, the rate at which it becomes blurry is directly proportional to its depth from the focal plane.

It is possible to calculate the defocus gradient map by running standard depth from defocus algorithms to recover a depth map and then directly converting the depth map to a defocus gradient map. However, we do not require exact depth per se, and in fact we are more interested in the apparent change of defocus blur with respect to aperture size. The defocus gradient map is a simpler, more direct representation for our applications.

We can use Eq. 3 to compute the defocus gradient map. At a pixel p, the change in blur should lie on the line $\sigma_p = M_p g$. Therefore, if we can estimate σ_p in each of our aperture images, we can directly calculate M_p. Unfortunately, it is difficult to directly estimate σ_p, and instead we adopt a hypothesis-and-test framework. For a set $\{M_i\}$ of discrete values of M, we hypothesize that pixel p has defocus gradient M_i, and test this hypothesis against our observed data.

In practice we use a Markov Random Field [Boykov et al. 2001; Szeliski et al. 2006] framework to solve for the defocus gradient map. We chose MRFs to solve for the defocus gradient map because it globally optimizes our data objective while simultaneously applying spatial regularization. We set up a MRF where the labels for each pixel are assigned from a set $\{M_i\}$ of discrete values of M. The optimization objective function is a standard combination of an error term E_i^p (Eq. 4) and a smoothness term, S. The penalty E_i^p for assigning a node p the label M_i is calculated as:

$$E_i^p = \sum_{j=1}^{N} (I_0 \otimes \sigma_j)(p) - I_j(p). \tag{4}$$

Equation 4 measures the error at pixel p between the smallest aperture (pinhole) image I_0, convolved with the expected blur $\sigma_j = M_i(1/N_j)$ and the measured blur size (as observed in image I_j).

The smoothness (regularization) term, S, defines how similar we would like spatial neighbors to be. S is specified as horizontal S_x

and vertical S_y pairwise weights between adjacent pixels. S_x is calculated as $S_x = \exp(-(I_{0x})^2 \times \sigma)$, where I_{0x} is the horizontal spatial derivative of I_0, and σ is a bandwidth parameter. S_y is calculated analogously. Our assumption is that depth discontinuities often occur across intensity edges. In flat intensity regions, our smoothness term encourages nearby pixels to have the same labels. However, neighborhoods with large gradients (e.g., edges) incur a small smoothness weight, and thus are less penalized for having different depth labels. Similar discontinuity-preserving smoothness terms have been used previously [Boykov et al. 2001; Levin et al. 2004].

4.2 Interpolating and Extrapolating Aperture Size

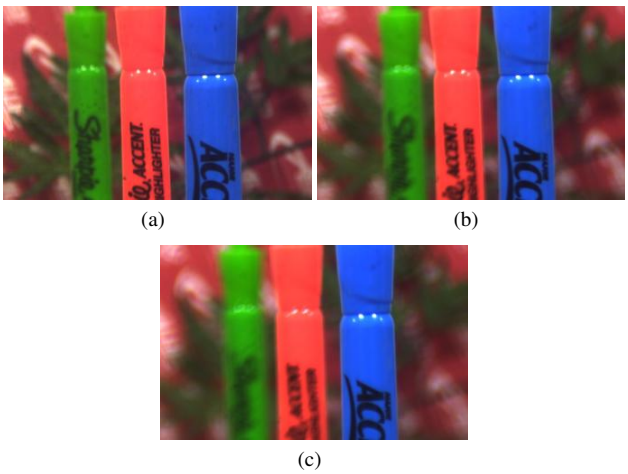


Figure 6: *Images created by summing rings of the aperture. (a) the pinhole aperture. (b) the sum of the pinhole and the first ring. (c) the sum of the pinhole and the first two rings. Notice that the depth of field is decreased as aperture rings are added.*

Our optical system captures four images of the scene simultaneously, each from an annular section of the aperture (see Fig. 3). It is possible to reconstruct the four aperture images by successively accumulating rings, e.g., the third aperture image is constructed by summing the inner disc and the next two aperture rings. A continuum of images with aperture sizes between the smallest and largest captured apertures can be synthesized by interpolating between the reconstructed images. Interpolating between apertures provides a way to continuously adjust the depth of field in an image. Figure 6 shows several images with varying aperture sizes constructed by summing the individual rings.

Using our defocus gradient map we can extrapolate shallow depth of field beyond the physical constraints of the maximum aperture. This is accomplished by extrapolating the size of the blurring kernel (using the defocus gradient) and blurring the pinhole image. Figure 8(a) shows an image taken at $f/1.8$ and Fig. 8(b) shows a synthesized version computed using our defocus gradient map technique. The defocus gradient map was computed from four separate exposures ($f/\# = 22$, 13, 8, and 4, respectively). The difference image is shown in Fig. 8(c). Figure 7 shows some extrapolated images taken with our camera.

Noise Characteristics Interpolated and extrapolated images have different noise characteristics. Images created using the interpolation technique show noise characteristics similar to a standard image of the same aperture size. Interpolated images have decreased shot noise by summing multiple aperture rings. Extrapolated images use only the pinhole image, and thus points on the image plane exhibit the noise characteristics of the pinhole image. Additionally, some light efficiency is lost due to the added elements in the relay and mirror system.

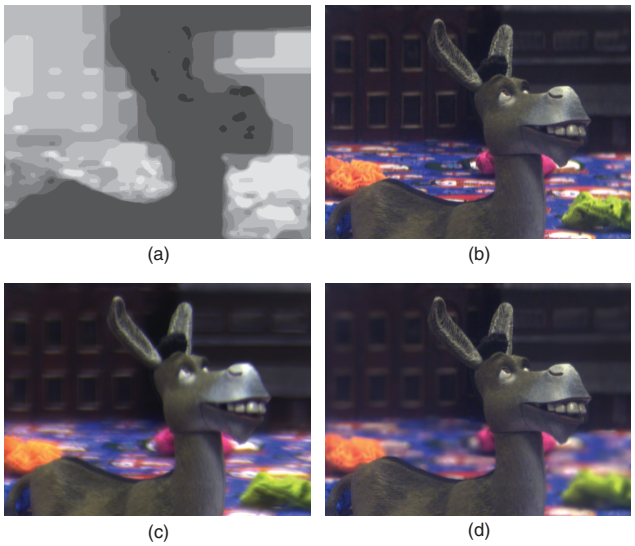


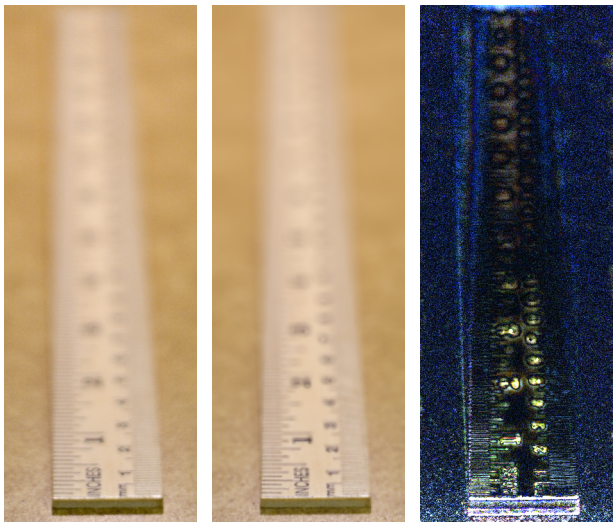
Figure 7: *Extrapolating the aperture to twice the area of the largest captured aperture. (a) the defocus gradient map; darker colors indicate smaller gradients. (b) f/8 image (smallest aperture). (c) accumulated f/2.8 image. (d) extrapolated f/2 image.*

4.3 Synthetic Refocusing and Guided Deconvolution

The defocus map is an encoding of the relative distance from the focus plane at each image point. In particular, image points near the focal plane will have a small defocus gradient, and the defocus gradient will increase the further the point is from the focal plane. Since we store discrete labels in the defocus map, we can relabel, or shift, the values in the map by an offset to achieve a synthetic refocusing effect. After offseting the labels we can perform the depth of field extrapolation technique described in the previous section. Figure 9 shows an example of our synthetic refocusing method. In Fig. 9(a) the focus is on the doll in front. In Fig. 9(b) the focus has been "moved" to the doll in the back.

It is important to note that we cannot perform actual refocusing of the image, we can only synthesize a new shallow depth of field image where the perceived image plane has been moved. In particular, we must rely on the large depth of field present in the smallest aperture image to provide all the detail at the shifted "focal plane." Additionally, to resolve the standard depth from defocus ambiguity, we assume the original photograph was focused on the closest scene point.

We now describe a form of guided deconvolution to enhance details in the pinhole image. The defocus gradient map provides an estimate of the PSF at each pixel. This PSF estimate can be used to adjust the deconvolution kernel used at each pixel. If we use K different labels when calculating the defocus gradient map (i.e., K depth values), then we run K separate deconvolutions of the pinhole image, each with a different PSF to produce a set of deconvolved images $\{D_i\}$. The size and shape of each PSF used is determined by Eq. 3 (a different value of M for each of the K labels, g is deter-



(a) (b) (c)

Figure 8: *A comparison of our extrapolation method to a reference image. (a) A reference image taken at f/1.8. (b) Our extrapolated synthetic result. (c) difference image. The images used to compute (b) were taken in multiple exposures, without our prototype optical system. The mean error is under 5% of the average image intensity.*

mined by the aperture size of the pinhole image, e.g., $f/8$). We use Lucy-Richardson deconvolution ("deconvlucy" in Matlab).

The final output of our deconvolution method is assembled by compositing the K separate deconvolutions based on the defocus gradient map labels. For example, if a pixel p has label k ($1 \leq k \leq K$) in the defocus gradient map (i.e., pixel p is at depth k), then we copy the corresponding pixel location in the k^{th} deconvolved image D_k (which has been deconvolved with a PSF corresponding to objects at depth k) into the output image. This provides a spatially adapted deconvolution method: The PSF used to calculate the deconvolved output at a pixel is determined by the estimated depth/defocus at the pixel. In contrast, traditional deconvolution methods use a single PSF for the entire image. The main benefit we have found is that our method alleviates most of the over-sharpening artifacts that are common with deconvolution methods by spatially tailoring the PSF to the local blurriness present in the image. Figures 9(d) and (e) compare refocusing with and without our deconvolution, respectively. Note the improved detail in our deconvolved version.

5 Discussion

5.1 Alternative Optical Designs

We have investigated several alternative optical designs in addition to the design previously described in this paper. We would like to briefly discuss two of these alternative designs, with the hope that it may spur further research.

The first alternative design involves placing a complex refractive element (i.e., a lens with non-traditional surface curvature) at the exit pupil which is then used to divide and image concentric rings of the aperture. The surface of the refractive element is designed such that light striking an annular region of the lens forms an image

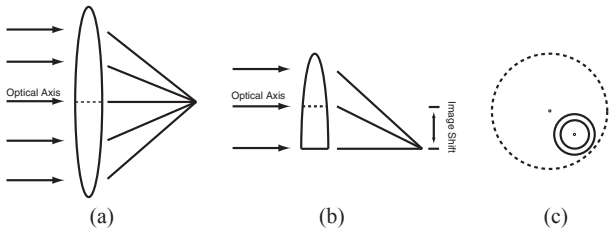


Figure 10: *Alternative designs. (a) Normal focusing through a lens. (b) Laterally shifting the focused image by decentering the optical axis. (c) Example of cutting an annular region from a larger theoretical lens. The distance from the optical axis of the annular region and the optical axis of the theoretical lens determines the amount of lateral shift.*

shifted laterally from the center of the CCD imager. Figure 10 describes how it is possible to laterally shift the image formed through a lens by decentering the lens with respect to the main optical axis. Conceptually, the proposed refractive element is composed of decentered annular sections, each cut from a theoretical larger-radius lens. See Fig. 10(c) for an example. To mimic our current design, the refractive element would be composed of four annular regions which formed images at the four quadrants of the CCD. Two advantages of this design are that, in theory, it could be easily extended to 9 or 16 regions and it splits the aperture at a single plane without occlusion problems because it uses only refractive elements. The main disadvantage we found in simulation was that because of the unusual surface shape, and the limitation to using a single lens element, the optical aberrations were unacceptably high. Additionally, it would be difficult and expensive to manufacture with glass, although it may be more practical using molded plastic optics.

The second alternative design is to place a micro-lens array over the CCD, where each lenslet is a miniaturized version of the complex refractive element described just described. This is similar to the light field camera design proposed by Ng [2005], however, instead of capturing a full light field, would integrate light from annular regions of the aperture, thus enabling higher spatial resolution. We believe that because the lenslet array is responsible for a very local resorting of light rays, the quality would be higher than any of the previously proposed designs. Unfortunately a micro-lens array cannot be removed in order to take standard photographs.

5.2 Limitations

One potential drawback of our system is that our mirror design requires very precise and difficult alignment of the optical elements in order to minimize aberrations. However, production quality manufacturing techniques could produce an optical system of a quality comparable to that of standard photographic lenses. Additionally, our system has difficulty recovering accurate depth and defocus information in regions without texture. This problem is common to many depth from defocus algorithms, and we employ the standard solution of using spatial regularization. It may be possible to use the unusual shapes of each aperture ring along with coded aperture methods to further improve the depth maps.

Another limitation is that our synthetic refocus method is unable to correctly synthesize blur effects across depth discontinuities. Unlike light field cameras, we are unable to capture the subtle parallax effects that occur across occlusion boundaries.

(a) (b) (c) (d) (e)

Figure 9: *Refocusing on near and far objects. (a) is the computed defocus gradient map. Dark values denote small defocus gradients. Using (a) we can synthesize (b) the near focus image. (c) defocus gradient map shifted to bring the far object to focus. (d) synthesized refocus image using (c). (e) synthesized refocus image using our guided deconvolution method. Notice the far object is still somewhat blurry in (d), and the detail is increased in (e).*

6 Conclusions

Our goal in this work was to explore the design space of computational cameras, and to examine the types of post-exposure edits that are possible without capturing a full light field. In this paper we describe a prototype optical system that captures images taken with multiple aperture settings in a single exposure. Our design is compact and does not require any modifications to either the camera or photographic lenses. Furthermore, the photographer is able to remove our system, and take a full resolution image if desired. We demonstrate several applications of our multi-aperture camera, including adjusting the depth of field and generating synthetically refocused images.

In future work we would like to investigate different methods of coding the aperture. In particular, we would like to extend our decomposition to a spatio-temporal splitting of the aperture. This would allow us to recover frequency content lost due either from depth defocus or motion blur. It may also be possible to design an adaptive optical system that adjusts the aperture coding based on the scene. Another avenue of future work that we would like to explore is to build a camera that simultaneously captures multiple images focused at different depths in a single exposure, using a single CCD sensor.

Acknowledgments

The authors would like to thank John Barnwell and Jonathan Westhues for their immeasurable help in the machine shop. We also thank Jane Malcolm, SeBaek Oh, Daniel Vlasic, Eugene Hsu, and Tom Mertens for all their help and support. This work was supported by a NSF CAREER award 0447561 *Transient Signal Processing for Realistic Imagery* and a Ford Foundation predoctoral fellowship. Frédo Durand acknowledges a Microsoft Research New Faculty Fellowship and a Sloan fellowship.

References

ADELSON, E. H., AND WANG, J. Y. A. 1992. Single lens stereo with a plenoptic camera. *IEEE PAMI 14*, 2, 99–106.

AGGARWAL, M., AND AHUJA, N. 2004. Split aperture imaging for high dynamic range. *IJCV 58*, 1 (June), 7–17.

BOYKOV, Y., VEKSLER, O., AND ZABIH, R. 2001. Fast approximate energy minimization via graph cuts. *IEEE PAMI 23*, 11.

CHAUDHURI, S., AND RAJAGOPALAN, A. 1999. *Depth From Defocus: A Real Aperture Imaging Approach*. Springer Verlag.

FARID, H., AND SIMONCELLI, E. 1996. A differential optical range camera. In *Optical Society of America, Annual Meeting*.

GEORGEIV, T., ZHENG, K. C., CURLESS, B., SALESIN, D., NAYAR, S., AND INTWALA, C. 2006. Spatio-angular resolution tradeoffs in integral photography. In *EGSR (2006)*, 263–272.

HARVEY, R. P., 1998. Optical beam splitter and electronic high speed camera incorporating such a beam splitter. US 5734507, United States Patent.

HASINOFF, S., AND KUTULAKOS, K. 2006. Confocal stereo. In *ECCV (2006)*, 620–634.

HECHT, E. 2002. *Optics*. Addison-Wesley, Reading, MA.

HIURA, S., AND MATSUYAMA, T. 1998. Depth measurement by the multi-focus camera. In *IEEE CVPR (1998)*, 953–961.

ISAKSEN, A., MCMILLAN, L., AND GORTLER, S. J. 2000. Dynamically reparameterized light fields. In *SIGGRAPH (2000)*.

LEVIN, A., LISCHINSKI, D., AND WEISS, Y. 2004. Colorization using optimization. In *SIGGRAPH (2004)*, 689–694.

MCGUIRE, M., MATUSIK, W., PFISTER, H., HUGHES, J. F., AND DURAND, F. 2005. Defocus video matting. *ACM TOG SIGGRAPH (2005) 24*, 3 (Aug.), 567–576.

MCGUIRE, M., MATUSIK, W., CHEN, B., HUGHES, J. F., PFISTER, H., AND NAYAR, S. 2007. Optical splitting trees for high-precision monocular imaging. *IEEE CG&A (2007)* (March).

NAEMURA, T., YOSHIDA, T., AND HARASHIMA, H. 2001. 3-D computer graphics based on integral photography. *Opt. Expr. 8*.

NARASIMHAN, S., AND NAYAR, S. 2005. Enhancing resolution along multiple imaging dimensions using assorted pixels. *IEEE PAMI 27*, 4 (Apr), 518–530.

NG, R. 2005. Fourier slice photography. *ACM TOG SIGGRAPH (2005) 24*, 3 (Aug.), 735–744.

OKANO, F., ARAI, J., HOSHINO, H., AND YUYAMA, I. 1999. Three-dimensional video system based on integral photography. *Optical Engineering 38* (June), 1072–1077.

PENTLAND, A. P. 1987. A new sense for depth of field. *IEEE PAMI 9*, 4, 523–531.

RAY, S. 1988. *Applied photographic optics*. Focal Press.

SZELISKI, R., ZABIH, R., SCHARSTEIN, D., VEKSLER, O., KOLMOGOROV, V., AGARWALA, A., TAPPEN, M. F., AND ROTHER, C. 2006. A comparative study of energy minimization methods for markov random fields. In *ECCV (2006)*, 16–29.

WATANABE, M., NAYAR, S., AND NOGUCHI, M. 1996. Real-time computation of depth from defocus. In *SPIE*, vol. 2599, 14–25.