Light Field Photography with a Hand-held Plenoptic Camera

Ren Ng*   Marc Levoy*   Mathieu Brédif*   Gene Duval†   Mark Horowitz*   Pat Hanrahan*

*Stanford University   †Duval Design

Abstract

This paper presents a camera that samples the 4D light field on its sensor in a single photographic exposure. This is achieved by inserting a microlens array between the sensor and main lens, creating a plenoptic camera. Each microlens measures not just the total amount of light deposited at that location, but how much light arrives along each ray. By re-sorting the measured rays of light to where they would have terminated in slightly different, synthetic cameras, we can compute sharp photographs focused at different depths. We show that a linear increase in the resolution of images under each microlens results in a linear increase in the sharpness of the refocused photographs. This property allows us to extend the depth of field of the camera without reducing the aperture, enabling shorter exposures and lower image noise. Especially in the macrophotography regime, we demonstrate that we can also compute synthetic photographs from a range of different viewpoints. These capabilities argue for a different strategy in designing photographic imaging systems.

To the photographer, the plenoptic camera operates exactly like an ordinary hand-held camera. We have used our prototype to take hundreds of light field photographs, and we present examples of portraits, high-speed action and macro close-ups.

Keywords: Digital photography, light field, microlens array, synthetic photography, refocusing.

1   Introduction

Conventional cameras do not record most of the information about the light distribution entering from the world. The goal of the camera presented in this paper is to re-capture this lost information: to measure not just a 2D photograph of the total amount of light at each point on the photosensor, but rather the full 4D light field measuring the amount of light traveling along each ray that intersects the sensor. One can also think of this as capturing the directional lighting distribution arriving at each location on the sensor.

The purpose of capturing the additional two dimensions of data is to allow us to apply ray-tracing techniques to compute synthetic photographs flexibly from the acquired light. The overall concept is to re-sort the rays of light to where they would have termninated if the camera had been configured as desired. For example, we demonstrate that we can shoot exposures with a relatively large $f/4$ lens aperture (for a short exposure time and low image noise), and yet compute photographs where objects at any depth are as sharp as if taken with a relatively small $f/22$ aperture. This result is a way of decoupling the traditional trade-off in photography between aperture size and depth of field.

Externally, our hand-held light field camera looks and operates exactly like a conventional camera: the viewfinder, focusing mechanism, length of exposure, etc. are identical. Internally, we augment the 2D photosensor by placing a microlens array in front of it, as proposed by Adelson and Wang [1992] in their work on the "plenoptic camera" (They did not build this device, but prototyped a non-portable version containing a relay lens.) Each microlens forms a tiny sharp image of the lens aperture, measuring the directional distribution of light at that microlens.

This paper explains the optical recipe of this camera in detail, and develops its theory of operation. We describe an implementation using a medium format digital camera and microlens array. Using this prototype, we have performed resolution experiments that corroborate the limits of refocusing predicted by the theory. Finally, we demonstrate examples of refocusing and view-point manipulation involving close-up macro subjects, human portraits, and high-speed action.

2   Related Work

The optical design of our camera is very similar to that of Adelson and Wang's plenoptic camera [1992]. Compared to Adelson and Wang, our prototype contains two fewer lenses, which significantly shortens the optical path, resulting in a portable camera. These differences are explained in more detail Section 3.1 once sufficient technical background has been introduced. The other main difference between our work is in application. We demonstrate use of the camera for synthetic image formation, especially refocusing of photographs, which was not mentioned by Adelson and Wang. They proposed the camera primarily as a device for range-finding, where depth is deduced by analyzing the continuum of stereo views coming from different portions of the main lens aperture. We would like to acknowledge their foresight, however, in anticipating classical light field rendering by describing how to move the photographer's viewpoint within the disk of the lens aperture.

The plenoptic camera has its roots in the integral photography methods pioneered by Lippman [1908] and Ives [1930]. Numerous variants of integral cameras have been built over the last century, and many are described in books on 3D imaging [Javidi and Okano 2002; Okoshi 1976]. For example, systems very similar to Adelson and Wang's were built by Okano et al. [1999] and Naemura et al. [2001], using graded-index (GRIN) microlens arrays. Another integral imaging system is the Shack-Hartmann sensor used for measuring aberrations in a lens [Tyson 1991]. A different approach to capturing light fields in a single exposure is an array of cameras [Wilburn et al. 2005].

It is also worth comparing our optical design to three other existing optical systems. The first is the modern, conventional photosensor array that uses microlenses in front of every pixel to concentrate light onto the photosensitive region [Ishihara and Tanigaki 1983; Gordon et al. 1991]. One can interpret the optical design in this paper as an evolutionary step in which we use not a single detector beneath each microlens, but rather an array of detectors capable of forming an image.

The second comparison is to artifical compound eye sensors (insect eyes) composed of a microlens array and photosensor. This is

essentially our sensor without a main lens. The first 2D version of such a system appears to have been built by Ogata et al. [1994], and has been replicated and augmented more recently using updated microlens technology [Tanida et al. 2001; Tanida et al. 2003; Duparré et al. 2004]. These projects endeavor to flatten the traditional camera to a plane sensor, and have achieved thicknesses as thin as a sheet of paper. However, the imaging quality of these optical designs is fundamentally inferior to a camera system with a large main lens; the resolution past these small lens arrays is severely limited by diffraction, as first noted by Barlow [1952] in comparing human and insect eyes.

As an aside from the biological perspective, it is interesting to note that our optical design can be thought of as taking a human eye (camera) and replacing its retina with an insect eye (microlens / photosensor array). No animal has been discovered that possesses such a hybrid eye [Land and Nilsson 2001], but this paper (and the work of Adelson and Wang) shows that such a design posseses unique and compelling capabilities when coupled with sufficient processing power (a computer).

The third optical system to be compared against is the "Wavefront Coding" system of Dowski and Johnson [1999]. Their system is similar to ours in that it provides a way to decouple the trade-off between aperture size and depth of field, but their design is very different. Rather than collecting and re-sorting rays of light, they use aspheric lenses that produce images with a depth-independent blur. Deconvolution of these images retrieves image detail at all depths. While their results in producing extended depth of field images is compelling, our design provides greater flexibility in image formation, since we can re-sort the measured rays of light in different ways to produce different images.

The concept of the 4D light field as a representation of all rays of light in free-space was introduced to the graphics community by Levoy and Hanrahan [1996] and Gortler et al. [1996]. The method of computing images through a virtual aperture from light field data was proposed by Levoy and Hanrahan [1996], first demonstrated by Isaksen et al. [2000], and goes under the name of synthetic aperture photography in current work [Vaish et al. 2004; Levoy et al. 2004].

Existing demonstrations of refocusing from light fields suffers from two problems. First, it is difficult to capture the light field datasets, requiring lengthy scanning with a moving camera, or large arrays of cameras that are not suitable for conventional hand-held photography. Second, the results tend to exhibit high aliasing in blurred regions due to incomplete sampling of the virtual aperture (e.g. due to gaps between cameras). Our design addresses both these issues: our light field camera is very easy to use, in that it behaves exactly like a conventional hand-held camera. Furtheremore, our optical design reduces aliasing drastically by integrating all the rays of light passing through the aperture.

3 Optical Design

The basic optical configuration, as outlined in the introduction, comprises a photographic main lens (such as a 50 mm lens on a 35 mm format camera), a microlens array, and a photosensor array of finer pitch. Figure 1 illustrates the layout of these components. The main lens may be translated along its optical axis, exactly as in a conventional camera, to focus on a subject of interest at a desired depth. As shown in Figure 1, rays of light from a single point on the subject are brought to a single convergence point on the focal plane of the microlens array. The microlens at that location separates these rays of light based on direction, creating a focused image of the aperture of the main lens on the array of pixels underneath the microlens.

Figure 13 is a dataset collected by the camera depicted in Figure 1. Macroscopically, the raw data is essentially the same as a
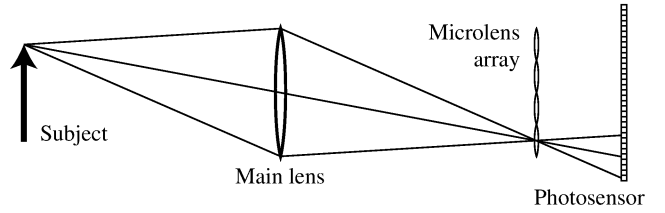

Figure 1: Conceptual schematic (not drawn to scale) of our camera, which is composed of a main lens, microlens array and a photosensor. The main lens focuses the subject onto the microlens array. The microlens array separates the converging rays into an image on the photosensor behind it.

conventional photograph. Microscopically, however, one can see the subimages of the main lens aperture captured by each microlens. These microlens images capture the structure of light in the world, and reveal, for example, the depth of objects. An introduction to this structure is described in the caption of the figure, and it is analyzed in detail in Adelson and Wang's paper.

3.1 Focusing Microlenses at Optical Infinity

The image under a microlens dictates the directional resolution of the system for that location on the film. To maximize the directional resolution, we want the sharpest microlens images possible. This means that we should focus the microlenses on the principal plane of the main lens. Since the microlenses are vanishingly small compared to the main lens, the main lens is effectively fixed at the microlenses' optical infinity. Thus, to focus the microlenses we cement the photosensor plane at the microlenses' focal depth.

Deviations from this separation result in misfocus blurring in the microlens subimages. For a sharp image within the depth of field of the microlenses, we require that the separation between the microlenses and photosensor be accurate to within $\Delta x_p \cdot (f_m / \Delta x_m)$, where Δx_p is the width of a sensor pixel, f_m is the focal depth of the microlenses, and Δx_m is the width of a microlens. For example, in our prototype, Δx_p = 9 microns, f_m = 500 microns, and Δx_m = 125 microns. Thus, we require that the separation between microlenses and photosensor be accurate to $\sim$ 36 microns.

This level of accuracy is probably one of the reasons that Adelson and Wang [1992] did not build the design in Figure 1. Instead, they introduced a relay lens between the microlens array and the photosensor, using it to focus the focal plane of the array onto the sensor. This compromise makes the camera easier to assemble and calibrate (we built an early prototype with a relay lens). However, it makes the overall device much longer and not portable because the relay lens focus is very sensitive. In the prototype described in this paper, we eliminate the relay lens by solving the problem of positioning the photosensor accurately at the focal plane of the microlenses.

Another simplification that we make over Adelson and Wang's prototype is elimination of the field lens that they position in front of the microlens array, which they use to ensure that images focus directly beneath every microlens. We do not require the images to be at exactly the same pitch as the microlenses.

3.2 Matching Main Lens and Microlens f-Numbers

The directional resolution relies not just on the clarity of the images under each microlens, but also on their size. We want them to cover as many photosensor pixels as possible.

The idea here is to choose the relative sizes of the main lens and microlens apertures so that the images are as large as possible without overlapping. A simple ray diagram (see Figure 2) shows that this occurs when the two f-numbers are equal. If the main
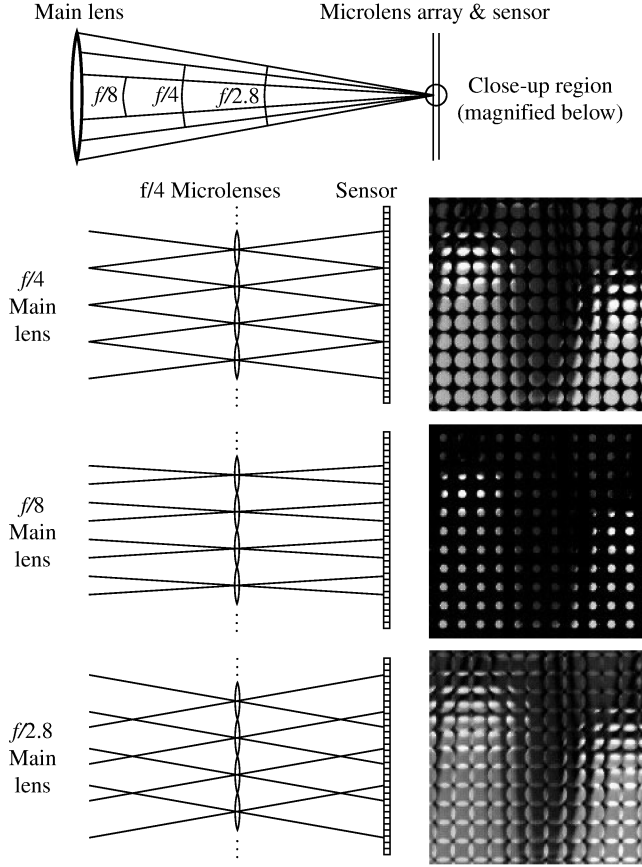
Figure 2: Illustration of matching main lens and microlens f-numbers. *Top*: Extreme convergence rays for a main lens stopped down to $f/2.8$, $f/4$ and $f/8$. The circled region is shown magnified for each of these f-stops, with the extreme convergence rays arriving at microlenses in the manigifed region. The images show close-ups of raw light field data collected under conditions shown in the ray diagrams. When the main lens and microlens f-numbers are matched at $f/4$, the images under the microlenses are maximal in size without overlapping. When the main lens is stopped down to $f/8$, the images are too small, and resolution is wasted. When the main lens is opened up to $f/2.8$, the images are too large and overlap.

lens' f-number is higher (i.e. the aperture is smaller relative to its focal length), then the images under each microlens are cropped, many pixels are black, and resolution is wasted. Conversely, if the main lens' f-number is lower (i.e. the aperture is larger), then the images under each microlens overlap, contaminating each other's signal through "cross-talk". Figure 2 illustrates these effects.

It is worth noting that in this context, the f-number of the main lens is not simply its aperture diameter divided by its intrinsic focal length, f. Rather, we are interested in the *image-side f-number*, which is the diameter divided by the separation between the principal plane of the main lens and the microlens plane. This separation is larger than f in general, when focusing on subjects that are relatively close to the camera.

3.3 Characterization of Acquired Data

We can characterize our camera's data by considering the two-plane light field, L, inside the camera, where $L(u, v, s, t)$ denotes the light traveling along the ray that intersects the main lens at (u, v) and the microlens plane at (s, t).
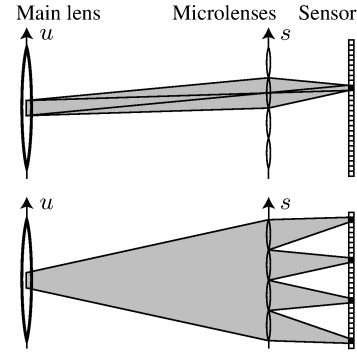

Figure 3: *Top:* All the light that passes through a pixel passes through its parent microlens and through its conjugate square (sub-aperture) on the main lens. *Bottom:* All rays passing through the sub-aperture are focused through corresponding pixels under different microlenses. These pixels form the photograph seen through this sub-aperture (see Figure 4).

Assuming ideal microlenses and pixels on aligned grids, all the light that passes through a pixel must (see top image of Figure 3)

- pass through its square parent microlens, and

- pass through the pixel's conjugate square on the main lens.

These two square regions specify a small 4D box in the light field. The pixel measures the integral of this box. Since this argument applies to all pixels, and the pixels and microlenses are arranged in regular lattices, we see that the dataset measured by all pixels is a box-filtered, rectilinear sampling of $L(u, v, s, t)$.

Sub-Aperture Images

It is instructive to examine the images formed by extracting the same pixel under each microlens, as described by Adelson and Wang [1992]. Such extraction corresponds to holding (u, v) fixed and considering all (s, t). The bottom image of Figure 3 shows that all rays passing through these pixels comes through the same sub-aperture on the main lens, and, conversely, every ray passing through this aperture is deposited in one of these pixels. Thus, the extracted image is the conventional photograph that would have resulted if taken with that sub-aperture as the lens opening (see Figure 4). Choosing a different pixel under the microlenses corresponds to choosing a different sub-aperture, and the sum of all these sub-apertures is of course the lens' original aperture. It is important to note that if the image of the main lens under the microlenses is N pixels across, then the width of the sub-aperture is N times smaller than the width of the lens' orginal aperture.

4 Image Synthesis

4.1 Synthetic Photography Equation

In this paper we concentrate on using the acquired light field to compute photographs as if taken with a synthetic conventional camera that were positioned and focused differently than the acquisition camera. For simplicity and clarity we model the everyday camera in which we vary just four parameters: the aperture size and location, and the depths of the (parallel) lens and sensor planes (see Figure 5). However, it should be noted that even though we do not demonstrate it in this paper, the availability of a light field permits ray-tracing simulations of fully general imaging configurations, such as view
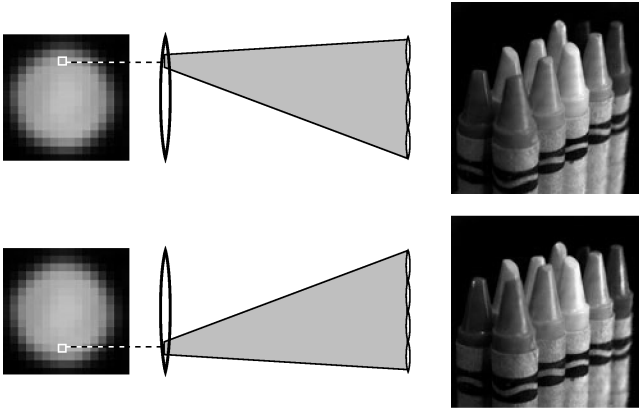
Figure 4: Two sub-aperture photographs obtained from a light field by extracting the shown pixel under each microlens (depicted on left). Note that the images are not the same, but exhibit vertical parallax.

cameras where the lens and sensor are not parallel, or even non-physical models such as general linear cameras [Yu and McMillan 2004] or imaging where each pixel is focused at a different depth.

The main point here is that our image formation technique is a physically-based simulation of a synthetic conventional camera. The remainder of this section simply develops the relevant imaging equation.

Let us introduce the concept of the synthetic light field L' parameterized by the synthetic $u' v'$ and $s' t'$ planes shown in Figure 5, such that $L'(u' v', s', t')$ is the light travelling between (u', v') on the synthetic aperture plane and (s', t') on the synthetic film plane. With this definition, it is well known from the physics literature (see for example Stroebel et al. [1986]) that the irradiance image value that would have appeared on the synthetic film plane is given by:

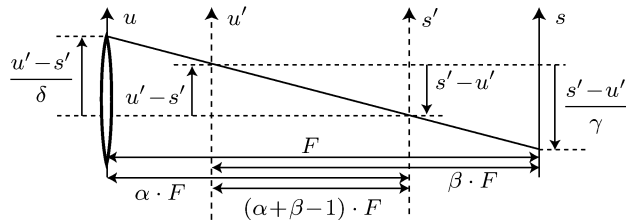$$E(s', t') = \frac{1}{D^2} \iint L'(u', v', s', t') A(u', v') \cos^4\theta \, du \, dv, \quad (1)$$

where D is the separation between the film and aperture, A is an aperture function (e.g. one within the opening and zero outside it), and θ is the angle of incidence that ray (u', v', s', t') makes with the film plane.

We invoke a paraxial approximation to eliminate the $\cos^4\theta$ term, and further simplify the equations by ignoring the constant $1/D^2$, to define

$$\overline{E}(s', t') = \iint L'(u', v', s', t') A(u', v') \, du \, dv \quad (2)$$

as the imaging equation that we will consider.

We want to express this equation in terms of the acquired light field, $L(u, v, s, t)$. The following diagram illustrates the relationship between L' and L.

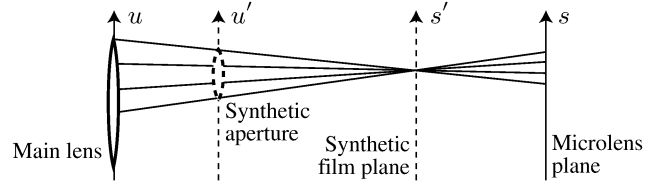Note the implicit definitions of α and β in the diagram. In addition,

Figure 5: Conceptual model for synthetic photography, shown in 2D. The u and s planes are the physical surfaces in the light field camera. u' is a virtual plane containing the synthetic aperture shown in dotted line, and s' is the synthetic film plane, together forming a synthetic camera. Note that these planes need not be between the acquisition planes. The image value that forms on the convergence point on the synthetic film is given by the sum of the illustrated cone of rays (see Equation 5). We find these rays in the acquired light field by their intersection points with the u and s planes.

we define

$$\gamma = \frac{\alpha + \beta - 1}{\alpha} \quad \text{and} \quad \delta = \frac{\alpha + \beta - 1}{\beta} \quad (3)$$

for notational convenience. The diagram shows that the ray intersecting u' and s' also intersects the u plane at $s' + (u' - s')/\delta$ and the s plane at $u' + (s' - u')/\gamma$. Thus,

$$L'(u' v', s', t') \quad (4)$$
$$= L\left(s' + \frac{u' - s'}{\delta}, \ t' + \frac{v' - t'}{\delta}, \ u' + \frac{s' - u'}{\gamma}, \ v' + \frac{t' - v'}{\gamma}\right).$$

Applying Equation 4 to Equation 2 produces the Synthetic Photography Equation that we use as the basis of image formation:

$$\overline{E}(s', t') = \iint L\left(s' + \frac{u' - s'}{\delta}, \ t' + \frac{v' - t'}{\delta}, \quad (5)\right.$$
$$\left. u' + \frac{s' - u'}{\gamma}, \ v' + \frac{t' - v'}{\gamma}\right) A(u', v') \, du \, dv.$$

Our rendering implementations are simply different methods of numerically approximating this integral.

The following sections deal with two important special cases, refocusing and moving the observer, and present the theoretical performance limits in these two regimes.

4.2 Digital Refocusing

Refocusing is the major focus of the experiments and results in this paper, because of the favorable theoretical limits described here. In refocusing, only the synthetic film plane moves (i.e. $\beta = 1$), and we use a full aperture (i.e. $A(u', v') = 1$. In this case $\delta = \alpha$ and $\gamma = 1$, and the synthetic photography equation simplifies to:

$$\overline{E}(s', t') = \iint L\left(u', v', \ u' + \frac{s' - u'}{\alpha}, \ v' + \frac{t' - v'}{\alpha}\right) \, du' \, dv'.$$
$$(6)$$

Examining this equation reveals the important observation that refocusing is conceptually just a summation of shifted versions of the images that form through pinholes (fix u' and v' and let s' and t' vary) over the entire uv aperture. In quantized form, this corresponds to shifting and adding the sub-aperture images, which is the technique used (but not physically derived) in previous papers [Vaish et al. 2004; Levoy et al. 2004].

Theoretical Sharpness of Refocused Photographs

This method of estimating Equation 6 implies that we can render any focal plane as sharp as it appears in the sub-aperture images. The circles of confusion (blur) in these sub-aperture images is N times narrower than in an un-refocused full-aperture photograph. Furthermore, the depth of field is as for a lens N times narrower.

This reasoning leads to the following important characterization of the limits of digital refocusing:

Linear Increase in Sharpness with Directional Resolution

If the image under the microlenses is N pixels across, then we can digitally refocus such that any desired region is geometrically sharper by a factor of N compared to a conventional camera. For equal clarity everwhere, the conventional photograph would have to be exposed with an aperture N times narrower.

Given this limit, it is easy to deduce the range of synthetic focal depths for which an optically sharp image can be produced (i.e. one that is essentially indistinguishable from a conventional photograph focused at that depth). This occurs simply when the synthetic focal plane falls within the depth of field of the lens aperture N times narrower.

It is worth noting that the preceding deductions can be proven in closed analytic form [Ng 2005], under the assumption that the light field camera provides a band-limited (rather than 4D-box filtered) light field. This Fourier-space analysis provides firm mathematical foundations for the theory described here.

Section 6 presents an experiment that we did to test the extent that we could refocus our system, showing that it comes within a factor of 2 of the theory described here.

Reducing Noise by Refocusing

A consequence of the sharpening capabilities described in the previous paragraph is that a light field camera can provide superior image signal-to-noise ratio (SNR) compared to a conventional camera with equivalent depth of field. This is achieved by shooting the light field camera with an aperture N times larger, resulting in an N^2 times increase in the acquired light signal. In a conventional camera the larger aperture would reduce depth of field and blur the image, but in the light field camera refocusing is used to match sharpness even with the larger aperture.

This N^2 increase in light level will result in an $O(N^2)$ or $O(N)$ increase in the overall image SNR, depending on the characteristics of the sensor and shooting conditions. If the camera is limited (e.g. in low light) by sensor noise that is independent of the signal, the SNR will increase as $O(N^2)$. On the other hand, if the limiting source of noise is photon shot noise (e.g. in high light), then the SNR increases as $O(N)$. Section 6 presents results of an experiment measuring the noise scaling in our system, showing that it is $O(N)$ in the experimental conditions that we used.

4.3 Moving the Observer

In classical light field rendering, we consider renderings from pinholes (i.e. $A(u, v)$, is a Dirac delta function centered at the pinhole at (u_o, v_o) on the lens plane). In this case it does not matter where the focal plane is, as it simply scales the image, and we set $\alpha = 1$. In this case, $\gamma = 1$ and $\delta = \beta$, and the synthetic photography equation (5) simplifies to

$$\overline{E}(s', t') = L\left(s' + \frac{u_o - s'}{\beta}, t' + \frac{v_o - t'}{\beta}, s', t'\right) \quad (7)$$

This equation shows that pinhole rendering is substantially faster than rendering lens-formed images, because we do not need to perform the double integral over the lens plane.
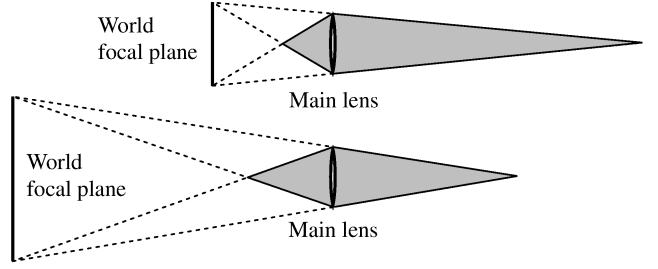

Figure 6: The space of locations (shaded regions) for a synthetic pinhole camera that result in a synthetic photograph without vignetting. The same optical system is depicted at two focus settings (resulting in different image magnifications, higher on top). Note that the space shrinks relative to the distance fo the scene as the magnification decreases. Thus, close-up photography (macrophotography) results in the greatest ability to move the observer.

Vignetting Analysis

The light fields that we acquire provide a subset of the space of rays limited in the uv plane by the bounds of the aperture, and in the st plane by the bounds of the microlens array. If we attempt to render synthetic photographs that require rays outside these bounds, then vignetting occurs in the synthetic image.

Figure 6 illustrates the subspace of pinholes that can be rendered without vignetting, derived by tracing extremal rays in the system. The top image of Figure 7 illustrates a vignetted synthetic photograph, where the pinhole has been moved towards the subject beyond the bounds of the subspace.

Using Closest Available Rays to Alleviate Vignetting

The one modification that we have made to our phsyically-based synthetic camera model is a non-physical technique to extend the vignetted images to a full field of view. The idea (see Figure 7 is to simply clamp the rays that extend beyond the bounds of the physical aperture to the periphery of the aperture (i.e. we use the closest rays that are available). It is interesting to note that this results in a multi-perspective image where the center of projection varies slightly for different pixels on the image plane.

5 Implementation

Our goal for the prototype was to create a hand-held light field camera that could be used in regular photographic scenarios and would highlight the capabilities of light field photography.

5.1 Hardware

The two main issues driving our component choices were resolution and physical working volume. Resolution-wise, we would ideally like an image sensor with a very large number of small pixels, since refocused image pixels consist of the sum of a large number of pixel values. In terms of working volume, we wanted to be able to access the sensor very easily to attach the microlens array.

These considerations led us to choose a medium format digital camera for our prototype. Medium format cameras provide the maximum sensor resolution available on the market. They also provide easiest access to the sensor, because the digital "back," which contains the sensor, detaches completely from the body.

Our digital back is a Megavision FB4040. The image sensor that it contains is a Kodak KAF-16802CE color sensor, which has
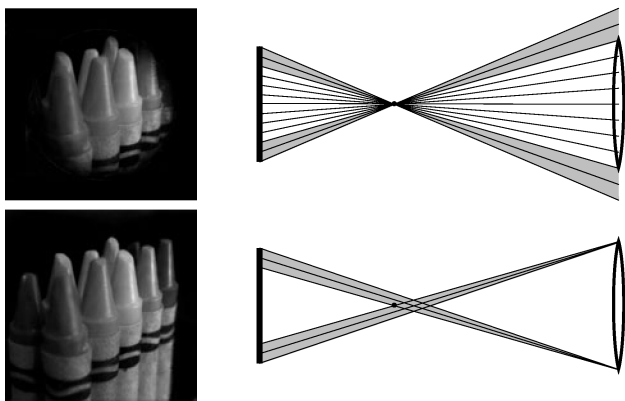
Figure 7: Technique for ameliorating vignetting. *Top:* Moving the pinhole observer beyond the bounds shown in Figure 6 results in vignetting because some required rays are unavailable (shaded gray). *Bottom:* To eliminate the vignetting, we use the closest available rays, by clamping the missing rays to the bounds of the aperture (shaded region). Note that these rays do not pass through the original pinhole, so the resulting multi-perspective image has a different center of projection for each ray in the corrected periphery.

approximately 4000×4000 pixels that are 9 microns wide. Our microlens array was made by Adaptive Optics Associates (part 0125-0.5-S). It has 296×296 lenslets that are 125 microns wide, square shaped, and square packed with very close to 100% fill-factor. The focal length of the microlenses is 500 microns, so their f-number is $f/4$. For the body of our camera we chose a Contax 645, and used two lenses: a 140 mm $f/2.8$ and 80 mm $f/2.0$. We chose lenses with wide maximum apertures so that, even with extension tubes attached for macrophotography, we could achieve an $f/4$ image-side f-number to match the f-number of the microlenses.

We glued the microlens array to a custom aluminum lens holder, screwed a custom base plate to the digital back over the photosensor, and then attached the lens holder to the base plate with three screws separated by springs (see Figure 8). Adjusting the three screws provided control over separation and tilt. The screws have 56 threads per inch, and we found that we could control separation with a mechanical resolution of 10 microns. Figure 8 shows a cross-section through the assembled parts.

We calibrated the separation—a one-time procedure—using a pinhole light source that produces an array of sharp spots on the sensor (one under each microlens) when the correct separation is achieved. The procedure took 10–20 iterations of screw adjustments. We created a high contrast pinhole source by stopping down the 140 mm main lens to its minimum aperture and attaching 78 mm of extension tubes. This creates an aperture of approximately $f/50$, which we aimed at a white sheet of paper.

The final resolution of the light fields that we capture with our prototype is 292×292 in the spatial st axes, and just under 14×14 in the uv directional axes. Figure 9 is a photograph showing our prototype in use.

5.2 Software

Our first software subsystem produces 4D light fields from the 2D sensor values. The first step is demosaicking: interpolating RGB values at every pixel from the values of the color filter array [Hamilton and Adams 1997]. We then correct for slight lateral misalignments between the microlens array and the photosensor by rotating the raw 2D image (by less than 0.1 degrees), interpolate the image upwards slightly to achieve an integral number of pixels per microlens, and then dice the array of aligned subimages to produce
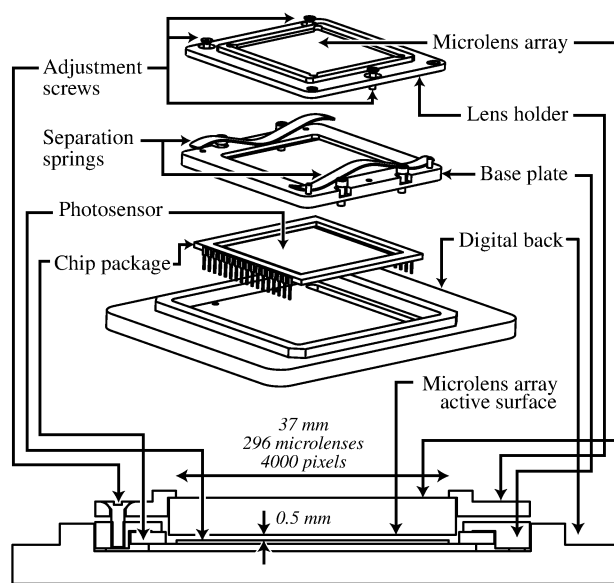

Figure 8: *Top:* Exploded view of assembly for attaching the microlens array to the digital back. *Bottom:* Cross-section through assembled parts.

Figure 9: Our light field camera in use.

the 4D light field, $L(u, v, s, t)$. (s, t) selects the subimage, and (u, v) selects the pixel within the subimage.

The second subsystem processes light fields to produce final photographs. Our various implementations of synthetic photography are simply different numerical techniques for approximating Equations 5 and Equation 6.

In the case of refocusing, we find that the traditional shifting and summing the sub-aperture images as in previous work [Vaish et al. 2004; Levoy et al. 2004] works well in most cases. For large motions in the focal plane, it can leave noticeable artifacts in blurred regions due to undersampling of the directional variation. For better image quality (with longer integration times), we use higher-order quadrature techniques, such as supersampling with a quadrilinear reconstruction filter.

To account for vignetting, we normalize the images by dividing each pixel by the fraction of integrated rays that fall within the bounds of the acquired light field. This eliminates darkening of borders in refocusing, for instance. This technique breaks down in the case of classical light field rendering, where we use the method of using closest available rays as replacements for vignetted rays, as described in Section 4.3.

Finally, we have experimented with producing extended depth of field images, by refocusing a light field at multiple depths and applying the digital photomontage technique of Agarwala et al. [2004]. Although we could produce extended depth of field simply by extracting a single sub-aperture image, this technique would be noisier because it integrates less light. Figure 15 of Section 7 illustrates this phenomenon.
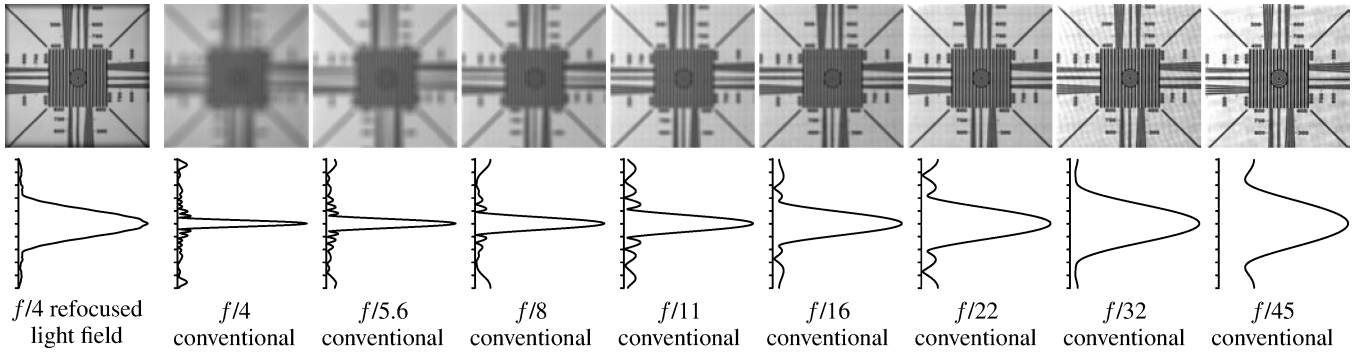
| $f/4$ refocused light field | $f/4$ conventional | $f/5.6$ conventional | $f/8$ conventional | $f/11$ conventional | $f/16$ conventional | $f/22$ conventional | $f/32$ conventional | $f/45$ conventional |

Figure 10: Top row shows photos of resolution chart. Bottom row plots (sideways, for easier comparison of widths) magnitude of center column of pixels in transfer function. Note that the $f/4$ refocused light field best matches the conventional photo taken at $f/22$.

6 Experimental Refocusing Performance

We experimentally measured the refocusing performance, and resulting improvement in signal-to-noise ratio, of our system by comparing the resolution of its images against those of a conventional camera system with the main lens stopped down to varying degrees.

6.1 Method

We took photographs of two test targets with our prototype. The first target is a pinhole light, to approximate a spatial impulse response, which we used to calculate the transfer function of our system. The second was a video resolution chart, which we used to provide easier visual comparison.

Physical measurement consisted of the following steps:

1. Position target 18 inches away and focus on it.
2. Move the target slightly closer so that it is blurred and stopping down the aperture reduces the blur.
3. Shoot light fields with all available aperture sizes (image-side f-numbers of $f/4$ to $f/45$).

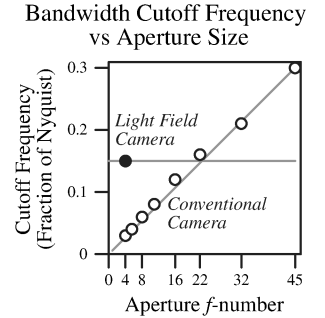Our processing procedure consisted of the following steps:

1. Compute conventional photos by adding up all light under each microlens.
2. Compute refocused photographs by varying refocus depth until maximum output image sharpness is achieved.
3. For the pinhole target, compute the Fourier transform of the resulting photographs. This is the transfer function of the system under each configuration.

6.2 Results

Figure 10 presents images of the resolution chart, and the computed transfer functions. Compare especially the reduction in blur between the light field camera at $f/4$ and the conventional camera at the same aperture size. The conventional camera produces a much blurrier image of the resolution target, and its transfer function has a much narrower bandwidth.

A visual inspection of the resolution charts shows that the light field camera at $f/4$ most closely matches the conventional camera at $f/22$, both in terms of the sharpness of the resolution chart as well as the bandwidth of its transfer function. Figure 11 presents numerical corroboration by analyzing the transfer functions.

Figure 11: Comparison of bandwidth cutoff frequencies for light field camera with refocusing versus conventional cameras with stopped down apertures. We choose the cutoff frequency as the minimum that contains 75% of the transfer function energy. Note that with this criterion the light field camera most closely matches the $f/22$ conventional camera.

6.3 Discussion

In the ideal case the light field camera would have matched the $f/45$ conventional camera. This is because the images under each microlens are 12 pixels across, and as described in Section 4.2, a directional resolution of 12×12 ideally enables refocusing to match an aperture 12 times narrower. In reality, our experiment shows that we are able to refocus within the depth of field of an $f/22$ aperture, approximately a loss of a factor of 2.

There are three main sources of directional blur that contribute to this loss. The first is diffraction. Physical optics [Goodman 1996] predicts that the blur pattern past a square aperture (our microlenses) has a central lobe of width $2 \cdot \lambda \cdot f$-number where λ is the wavelength of light and the f-number is for the smallest aperture in the system. In our system, this means that the microlenses induce a blur on the image plane that is approximately 2/3 the width of a pixel, which degrades the directional resolution. The second source of blur is the fact that there are a non-integral number of sensor pixels per microlens image, creating a vernier-like offset in the location of a pixels relative to the center of the image from microlens to microlens. This means that rays from the same portion of the main lens will project to slightly different (neighboring) pixels under different microlenses. Our software does not account for this effect, and thus our interpretation of the incident direction of rays collected by a pixel may be up to half a pixel off depending on the microlens. The third source of blur is that we use quadrilinear filtering in resampling the light field values in numerical integration of the imaging equation.

Since we can refocus to match the sharpness of a lens aperture approximately 6 times narrower, we measured the improvement in SNR that could be achieved by opening up the aperture and refocusing. We measured the SNR of a conventional $f/22$ photograph (using the average of many such photos as a noise-free standard), and the SNR of a refocused $f/4$ light field photograph (see Fig-
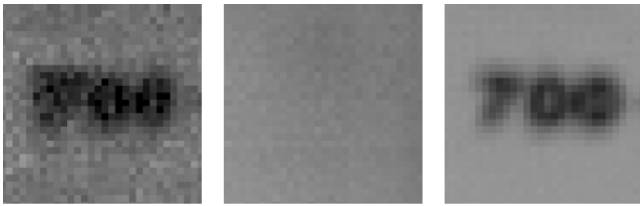
Figure 12: Noise analysis. The left image is a close-up of a conventional photo at $f/22$, the middle is a conventional photo at $f/4$ (completely blurred out due to misfocus), and the right image is a refocused light field photograph shot at $f/4$. All exposure durations were the same (1/60 sec), and the left image has been scaled to match the brightness of the other two. The light field photo (right) is much sharper than the conventional one at the same aperture size (middle). It is as sharp as the conventional $f/22$ photo, but its SNR is 5.8 times higher.

ure 12). The light field photograph integrates 25-36 times as much light due to the larger aperture, and indeed we found that the SNR of the light field camera was 5.8 times better, suggesting scaling with the square root of the light level. As an aside, this suggests that our system is limited by photon noise, since being limited by signal-independent sources such as dark current would exhibit SNR scaling linearly with light level.

7 Examples of Light Field Photography

We have acquired hundreds of light fields of a variety of photographic subjects, including portraits, live action and close-up macro subjects. This section presents a selection of these photographs.

Figure 13 presents an entire light field acquired with our camera, of a scene with a number of people arranged in depth. Because the photograph was taken indoors, a large aperture was used to keep the exposure time short. The camera was focused on the man wearing spectacles. The top-most image of Figure 14 illustrates what a photograph of this scene taken with a conventional sensor would have looked like, with severe blurring of the closest and furthest faces. The remaining images in Figure 14 show that we can digitally refocus the scene to bring any person into sharp focus. Furthermore, Figure 15 demonstrates that we can produce a photograph in which everyone is in focus at once, using the technique described in Section 5.2.

Figure 16 illustrates a different type of use, in which we took a number of candid portraits of a close-up subject. A large aperture was used to focus attention on the face and blur the background. The chosen example shows a case in which the camera did not accurately focus on the subject, due to subject movement and the extremely shallow depth of field. These examples demonstrate that even though the auto-focus is off by only a few inches, the face is severely blurred. Refocusing of the light field successfully restores critical sharpness.

Figure 17 illustrates that our camera can operate with very short exposures, just as in an ordinary camera. The photographs illustrate a high-speed light field of water frozen in time as it splashes out of a broken wine glass. This example suggests the potential of light field photography in applications such as sports photography, where a large aperture is used to collect light quickly for a short exposure. The resulting shallow depth of field, coupled with moving targets, makes for challenging focus requirements, which would be reduced by digital refocusing after exposure.

Figure 18 illustrates moving the observer in the macrophotography regime. The light field was acquired with the lens imaging at 1:1 magnification (the size of the crayons projected on the photosensor is equal to their size in the world). These figures demonstrate the substantial changes in parallax and perspective that can

be achieved when the lens is large relative to the field of view.

8 Discussion

We believe that this imaging methodology expands the space for camera design. For example, the new post-exposure controls reduce the pre-exposure requirements on auto-focus systems. Faster systems might be designed by allowing the camera to focus less accurately before the shutter is released. A similar principle applies to the design of the main lens, which is often the most expensive camera subsystem. Although we have not demonstrated it in this paper, the availability of the light field should allow reduction of lens aberrations by re-sorting the acquired, distorted rays of light to where they should have terminated in a system without aberrations. This capability may allow the use of simpler and cheaper lenses.

Another part of the design space that needs to be re-examined is the sensor design. For at least two reasons, light field cameras provide stronger motivation than conventional cameras for pursuing small pixels. The first reason is that conventional cameras have already reached the limits of useful resolution for many common applications. The average commercial camera released in the second half of 2004 had 5.4 megapixels of resolution [Askey 2004], and yet studies show that more than 2 megapixels provides little perceptually noticeable improvement for the most common 4"×6" print size [Keelan 2002]. The second reason is that in conventional cameras, reducing the pixel size reduces image quality by decreasing the SNR. In light field cameras, reducing pixel size does not reduce the signal level because we sum all the pixels under a microlens to produce a final image pixel. Thus, it is the size of the *microlenses* that determines the signal level, and we want the pixels to be as small as possible for maximum directional resolution.

Of course explorations of such reductions in pixel size must take into consideration the appropriate physical limits. The most fundamental limit is diffraction, in particular the blurring of light as it passes through an aperture. As discussed in Section 6.3, classical physical optics [Goodman 1996]) predicts that the optical signal on the sensor in our camera is blurred by a diffraction kernel of 6 microns. The pixels in the sensor that we used are 9 microns across, so pursuing much smaller pixels would require using microlenses and main lenses with larger relative apertures to reduce diffraction.

Aside from further research in the design of the camera and sensor themselves, we believe that there are many interesting applications that could benefit from this technology. In photography, the capability of extending depth of field while using a large aperture may provide significant benefits in low-light or high-speed imaging, such as sports photography and security surveillance. The ability to refocus or extend the depth of field could also find many uses in movies and television. For example, effects such as the "focus pull" might be performed digitally instead of optically.

A very different kind of benefit might emerge from the popularization of light field datasets themselves. These permit interactive capabilities that are absent from conventional photographs. Playing with the focus may provide digital photo albums with new entertainment value in that different subjects of interest in each photo can be discovered as they "pop" when the correct focal depth is found (e.g. Figure 14). Changing the viewpoint interactively, even at ordinary photographic magnifications, can also create a dramatic 3D effect through parallax of foreground figures against distant backgrounds. Many of our colleagues and friends have found these two interactive qualities new and fun.

Finally, we would like to describe more serious and potentially important applications in medical and scientific microscopy, where shallow depth of field is an intimately close barrier. Employing a light field sensor instead of a conventional photosensor would enable one to extend the depth of field without being forced to reduce

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

Figure 13: A complete light field captured by our prototype. Careful examination (zoom in on electronic version, or use magnifying glass in print) reveals 292×292 microlens images, each approximately 0.5 mm wide in print. Note the corner-to-corner clarity and detail of microlens images across the light field, which illustrates the quality of our microlens focusing. (a), (b) and (c) show magnified views of regions outlined on the key in (d). These close-ups are representative of three types of edges that can be found througout the image. (a) illustrates microlenses at depths closer than the focal plane. In these *right-side up* microlens images, the woman's cheek appears on the left, as it appears in the macroscopic image. In contrast, (b) illustrates microlenses at depths further than the focal plane. In these *inverted* microlens images, the man's cheek appears on the *right*, opposite the macroscopic world. This effect is due to inversion of the microlens' rays as they pass through the world focal plane before arriving at the main lens. Finally, (c) illustrates microlenses on edges *at* the focal plane (the fingers that are clasped together). The microlenses at this depth are constant in color because all the rays arriving at the microlens originate from the same point on the fingers, which reflect light diffusely.

Figure 14: Refocusing after a single exposure of the light field camera. Top is the photo that would have resulted from a conventional camera, focused on the clasped fingers. The remaining images are photographs refocused at different depths: middle row is focused on first and second figures; last row is focused on third and last figures. Compare especially middle left and bottom right for full effective depth of field.

Figure 15: *Left*: Extended depth of field computed from a stack of photographs focused at different depths. *Right:* A single sub-aperture image, which has equal depth of field but is noisier.

Figure 16: Refocusing of a portrait. Left shows what the conventional photo would have looked like (autofocus mis-focused by only 10 cm on the girl's hair). Right shows the refocused photograph.

Figure 17: Light field photograph of water splashing out of of a broken wine glass, refocused at different depths.
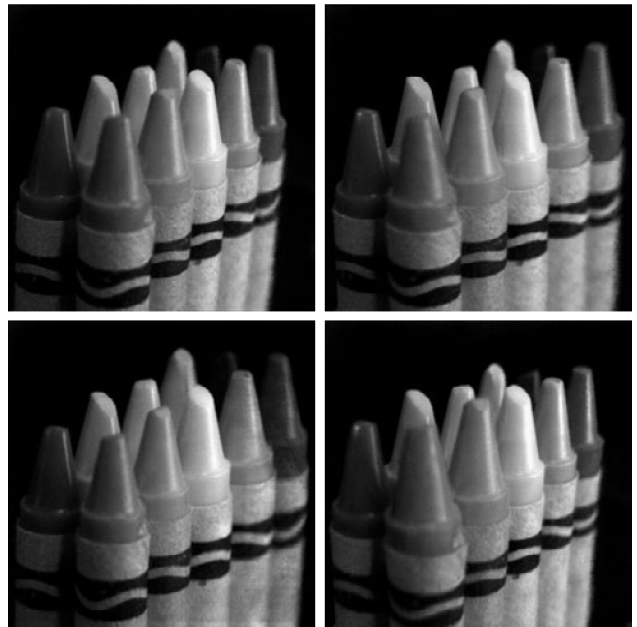

Figure 18: Moving the observer in the macrophotography regime (1:1 magnification), computed after a single light field camera exposure. Top row shows movement of the observer laterally within the lens plane, to produce changes in parallax. Bottom row illustrates changes in perspective by moving along the optical axis, away from the scene to produce a near-orthographic rendering (left) and towards the scene to produce a medium wide angle (right). In the bottom row, missing rays were filled with closest available (see Figure 7).

the numerical aperture of the system. It would also allow the ability to move the viewpoint after each exposure. This could be of particular use in video microscopy of live subjects, where there is not time to scan in depth to build up extended depth of field. In applications where one scans over different depths to build up planes in a volume, such as deconvolution microscopy [McNally et al. 1999], the ability to digitally refocus would accelerate scanning by permitting larger steps between planes.

9 Conclusion

This paper demonstrates a simple optical modification to existing digital cameras that causes the photosensor to sample the in-camera light field. This modification is achieved without altering the external operation of the camera. Coupled with custom software processing, the acquired light fields provide unique photographic capabilities, such as the ability to digitally refocus the scene after exposure, extend the depth of field while maintaining high signal-to-noise ratio, and alter the viewpoint and perspective.

The fact that these capabilities scale linearly with directional resolution allows the design to be gradually adopted as the march of VLSI provides excess sensor resolution that may be allocated to directional sampling. We believe that through this mechanism the design of every lens-based digital imaging system could usefully converge to some parameterization of the light field camera paradigm.

Acknowledgments

References

ADELSON, T., AND WANG, J. Y. A. 1992. Single lens stereo with a plenoptic camera. *IEEE Transactions on Pattern Analysis and Machine Intelligence 14*, 2 (Feb), 99–106.

AGARWALA, A., DONTCHEVA, M., AGRAWALA, M., DRUCKER, S., COLBURN, A., CURLESS, B., SALESIN, D., AND COHEN, M. 2004. Interactive digital photomontage. *ACM Transactions on Graphics (Proceedings of SIGGRAPH 2004) 23*, 3, 292–300.

ASKEY, P., 2004. Digital cameras timeline: 2004. Digital Photography Review, http://www.dpreview.com/reviews/timeline.asp?start=2004.

BARLOW, H. B. 1952. The size of ommatidia in apposition eyes. *Journal of Experimental Biology 29*, 4, 667–674.

DOWSKI, E. R., AND JOHNSON, G. E. 1999. Wavefront coding: A modern method of achieving high performance and/or low cost imaging systems. In *SPIE Proceedings*, vol. 3779.

DUPARRÉ, J., DANNBERG, P., SCHREIBER, P., AND BRÄUER, A. 2004. Micro-optically fabricated artificial apposition compound eye. In *Electronic Imaging – Science and Technology, Proc. SPIE 5301*.

GOODMAN, J. W. 1996. *Introduction to Fourier Optics, 2nd edition.* McGraw-Hill.

GORDON, N. T., JONES, C. L., AND PURDY, D. R. 1991. Application of microlenses to infrared detector arrays. *Infrared Physics 31*, 6, 599–604.

GORTLER, S. J., GRZESZCZUK, R., SZELISKI, R., AND COHEN, M. F. 1996. The lumigraph. In *SIGGRAPH 96*, 43–54.

HAMILTON, J. F., AND ADAMS, J. E., 1997. Adaptive color plan interpolation in single sensor color electronic camera. U.S. Patent 5,629,734.

ISAKSEN, A., MCMILLAN, L., AND GORTLER, S. J. 2000. Dynamically reparameterized light fields. In *SIGGRAPH 2000*, 297–306.

ISHIHARA, Y., AND TANIGAKI, K. 1983. A high photosensitivity IL-CCD image sensor with monolithic resin lens array. In *Proceedings of IEEE IEDM*, 497–500.

IVES, H. E. 1930. Parallax panoramagrams made with a large diameter lens. *J. Opt. Soc. Amer. 20*, 332–342.

JAVIDI, B., AND OKANO, F., Eds. 2002. *Three-Dimensional Television, Video and Display Technologies*. Springer-Verlag.

KEELAN, B. W. 2002. *Handbook of image quality : characterization and prediction*. Marcel Dekker.

LAND, M. F., AND NILSSON, D.-E. 2001. *Animal Eyes*. Oxford University Press.

LEVOY, M., AND HANRAHAN, P. 1996. Light field rendering. In *SIGGRAPH 96*, 31–42.

LEVOY, M., CHEN, B., VAISH, V., HOROWITZ, M., MCDOWALL, I., AND BOLAS, M. 2004. Synthetic aperture confocal imaging. *ACM Transactions on Graphics (Proceedings of SIGGRAPH 2004) 23*, 3, 822–831.

LIPPMANN, G. 1908. La photographie intégrale. *Comptes-Rendus, Académie des Sciences 146*, 446–551.

MCNALLY, J. G., KARPOVA, T., COOPER, J., AND CONCHELLO, J. A. 1999. Three-dimensional imaging by deconvolution microscopy. *METHODS 19*, 373–385.

NAEMURA, T., YOSHIDA, T., AND HARASHIMA, H. 2001. 3-D computer graphics based on integral photography. *Optics Express 8*, 2, 255–262.

NG, R. 2005. Fourier slice photography. To appear at SIGGRAPH 2005.

OGATA, S., ISHIDA, J., AND SASANO, T. 1994. Optical sensor array in an artifical compound eye. *Optical Engineering 33*, 11, 3649–3655.

OKANO, F., ARAI, J., HOSHINO, H., AND YUYAMA, I. 1999. Three-dimensional video system based on integral photography. *Optical Engineering 38*, 6 (June 1999), 1072–1077.

OKOSHI, T. 1976. *Three-Dimensional Imaging Techniques*. Academic Press.

PHARR, M., AND HUMPHREYS, G. 2004. *Physically Based Rendering: From Theory to Implementation*. Morgan Kaufmann.

STROEBEL, L., COMPTON, J., CURRENT, I., AND ZAKIA, R. 1986. *Photographic Materials and Processes*. Focal Press.

TANIDA, J., KUMAGAI, T., YAMADA, K., MIYATAKE, S., ISHIDA, K., MORIMOTO, T., KONDOU, N., MIYAZAKI, D., AND ICHIOKA, Y. 2001. Thin observation module by bound optics (TOMBO): concept and experimental verification. *Applied Optics 40*, 11 (April), 1806–1813.

TANIDA, J., SHOGENJI, R., KITAMURA, Y., YAMADA, K., MIYAMOTO, M., AND MIYATAKE, S. 2003. Color imaging with an integrated compound imaging system. *Optics Express 11*, 18 (September 8), 2109–2117.

TYSON, R. K. 1991. *Principles of Adaptive Optics*. Academic Press.

VAISH, V., WILBURN, B., JOSHI, N., AND LEVOY, M. 2004. Using plane + parallax for calibrating dense camera arrays. In *Proceedings of CVPR*.

WILBURN, B., JOSHI, N., VAISH, V., TALVALA, E.-V., ANTUNEZ, E., BARTH, A., ADAMS, A., LEVOY, M., AND HOROWITZ, M. 2005. High performance imaging using large camera arrays. To appear at SIGGRAPH 2005.

YU, J., AND MCMILLAN, L. 2004. General linear cameras. In *8th European Conference on Computer Vision (ECCV)*, 61–68.