Spatio-Angular Resolution Tradeoff in Integral Photography

Todor Georgeiv[1], Ke Colin Zheng[2], Brian Curless[2], David Salesin[1,2], Shree Nayar[3], and Chintan Intwala[1]

[1]Adobe Systems
[2]University of Washington
[3]Columbia University

Abstract

An integral camera samples the 4D light field of a scene within a single photograph. This paper explores the fundamental tradeoff between spatial resolution and angular resolution that is inherent to integral photography. Based on our analysis we divide previous integral camera designs into two classes depending on how the 4D light field is distributed (multiplexed) over the 2D sensor. Our optical treatment is mathematically rigorous and extensible to the broader area of light field research.

We argue that for many real-world scenes it is beneficial to sacrifice angular resolution for higher spatial resolution. The missing angular resolution is then interpolated using techniques from computer vision. We have developed a prototype integral camera that uses a system of lenses and prisms as an external attachment to a conventional camera. We have used this prototype to capture the light fields of a variety of scenes. We show examples of novel view synthesis and refocusing where the spatial resolution is significantly higher than is possible with previous designs.

Categories and Subject Descriptors (according to ACM CCS): I.3.3 [Computer Graphics]: Digital Photography

1. Introduction

The *light field* or *radiance density function* is a complete representation of light energy flowing along "all rays" in 3D space. This density is a field defined in the 4D domain of the optical phase space, the space of all lines in 3D with symplectic structure [GS85].

Conventional cameras, based on 2D image sensors, are simply integration devices. In a typical setting, they integrate over a 2D aperture to produce a 2D projection of the full 4D light field density. Integral Photography [Lip08] was proposed almost a century ago to "undo" the integration and measure the complete 4D light field arriving at all points on a film plane or sensor.

As demonstrated by Levoy and Hanrahan [LH96] and Gortler et. al. [GGSC96], capturing the additional two dimensions of radiance data allows us to re-sort the rays of light to synthesize new photographs, sometimes referred to as novel views. In the last decade, significant progress has been made in light field rendering to simulate a realistic camera with a finite aperture, producing depth of field effects. In this way, *synthetic-aperture photography* [LH96, IMG00]

can compute photographs focused at different depths from a single light field by simple numerical integration over the desired aperture.

Recently, Ng et al. [NLB*05] have shown that a full 4D light field can be captured even with a hand-held plenoptic camera. This approach makes light field photography practical, giving the photographer the freedom and the power to make adjustments of focus and aperture *after* the picture has been taken. In a way, it transfers the optics of the lens of the camera into the digital domain, greatly extending the types of post-processing possible with software like Photoshop.

However, one drawback of the design of Ng et al. is that they require a large number of samples of the radiance: With their design, even with a 16-megapixel image sensor, the spatial resolution of the sampled light field is limited to 300×300 pixels.

This paper surveys some of the previously proposed light field camera designs. Integral or light field photography is approached from the perspective of radiance analysis in geometrical optics. This provides a new way of looking at integral photography and the associated light field rendering. We

then propose new camera designs that produce higher spatial resolution than the camera of Ng et al., while trading-off the light field's angular sampling density. However, this lower angular resolution in the input is compensated for by inserting data synthesized by view interpolation of the measured light field.

We use three-view morphing to interpolate the missing angular samples of radiance. We demonstrate that such interpolated light fields generated from sparsely sampled radiance are generally good enough to produce synthetic aperture effects, new view synthesis, and refocusing with minimal loss in quality.

We have built an integral camera that uses a system of lenses and prisms as an external optical attachment to a conventional camera. Using a computer-vision based view interpolation algorithm, we demonstrate how our camera can be used to adjust the depth of field and synthesize novel views for scenes with high-speed action, which are impossible to do with conventional cameras. Moreover, with the same 16-megapixel sensor used by Ng et al. we are able to achieve a much higher spatial resolution of 700×700 pixels in the computed images.

2. Trading angular for spatial resolution

Work in integral / light field photography falls into two major classes:

1. The earliest works of Lipmann [Lip08] and Ives [Ive28] among others, known as *integral photography*, used arrays of lenslets or pinholes placed directly in front of film, creating multiple images on it like an array of cameras. Optically similar to that is a physical array of digital cameras, which is the main approach used in current light field research (e.g., [WJV*05]). A related type of integral photography design places an array of positive lenses in front of a conventional camera to create an array of real images between the lenses and the camera. Then the camera takes a picture focused on those images (e.g., [OHAY97]). This approach is closest to ours. Also consider the following related works [OAHY99, NYH01, SH02].
2. The more recent approaches of Adelson et al. [AW92] and Ng et al. [NLB*05], known as *plenoptic cameras*, effectively place a big lens in front of the array of lenslets (or cameras) considered in the first approach, forming an image on the array of lenslets. Each lenslet itself creates an image sampling the angular distribution of radiance at that point, which corresponds to one single direction observed from multiple points of view on the main lens aperture. This approach swaps the placement of spatial and angular samples on the image plane: instead of producing an array of ordinary images, as in integral photography, it creates what appears as a single, recognizable "image" consisting of small 2D arrays of angular samples of a single point in the scene. A related technique

is that of the Hartman-Shack sensor [Tys91], which was also proposed a century ago to study wavefront shape in optics.

Both types of light field cameras share a single goal — that of increasing angular resolution of the measured light field, which often comes at the cost of spatial resolution of the final $2D$ images that are generated. In the rest of this section, we explore this trade-off between angular and spatial resolution and show that for typical scenes it can be advantageous to use higher spatial resolution at the cost of angular resolution.

### 2.1. Drawbacks of the plenoptic camera design

In the plenoptic camera recently built and studied in detail by Ng et al. [NLB*05], the light field is captured by an array of 296^2 lenslets inside a conventional camera. Each lenslet in this setting corresponds to a little camera producing an approximately 14×14 pixel image of the main lens aperture. Each pixel within that small image corresponds to one viewpoint on the aperture, while different lenslets correspond to different pixels in the final image. The result is an approximately 100-view light field with 90,000 pixels per view. (The number of effective views is 100 instead of 14^2 due to losses, which will be discussed later.)

Unfortunately, from the standpoint of professional photographers, this system produces images with very low spatial resolution. An obvious way to remedy this problem would be to use more lenslets (for example, 1,000,000), with fewer views/pixels under each lenslet (for example, 16). The difficulty with such a remedy is that each small image of the main lens aperture created by a lenslet includes pixels at the aperture boundary that are either lost entirely, or noisy. Such boundary pixels are only partially covered by the image of the aperture. In order to reconstruct the true irradiance corresponding to the illuminated part of each pixel we would need to know exactly what percentage of it has been illuminated, and correct for that in software. In other words, we would need very precise calibration of all pixels in the camera. However, captured pixel values are affected by tiny misalignments: A misalignment of a micrometer can change a boundary pixel value by more than 10%. This problem gets very visible when the lenslets get smaller. In the limiting case of a 2×2 or 4×4 pixel image under each lenslet (depending on Bayer array), all the pixels become boundary pixels, providing no reliable $3D$ information at all.

2.2. How do we capture $4D$ radiance with a $2D$ sensor?

It turns out that the original integral camera designs have certain advantages when it comes to acquiring images with sufficient spatial resolution on moderate resolution sensors.

For visualization purposes, suppose that optical phase space (a.k.a. "light field space") were 2-dimensional (instead
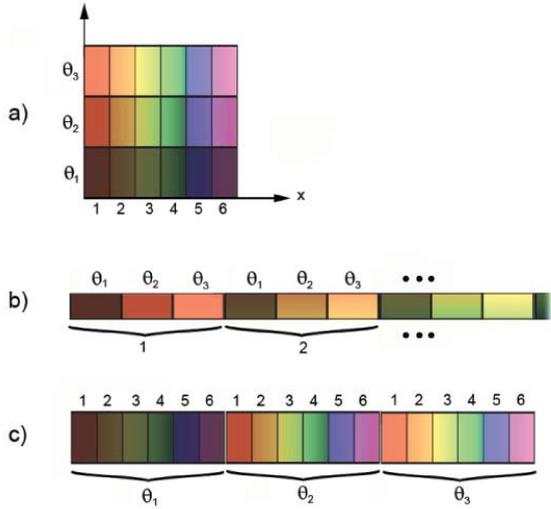
Figure 1: *(a) Partitioning optical phase space* (x, θ) *into sampling regions for the light field camera. (b) The "Plenoptic" way of sampling. (c) The "Integral Photography" way of sampling.*

of $4D$), with one spatial dimension x, and one angular dimension θ. We will consider the possible designs for a $1D$ image detector, as shown in Figure 1.

As earlier researchers have pointed out [LH96], we want to sample most densely the dimension that changes the most — i.e., the spatial rather than angular dimension. Adopting this space-saving approach we end up with the partitioning of the light space into rectangles or *radiance pixels*, "long" in the direction of θ, and "short" in the direction of x. (See Figure 1a) Radiance pixels are relatively sparse (e.g., 3 samples) along the angular direction, but relatively dense (e.g., 6 samples) along the spatial direction. This is the type of partitioning of optical phase space practically chosen in most light field cameras.

In order to fit the $2D$ phase space of Figure 1a into a single dimension, we need to rearrange, or multiplex, light field data to fit into a single row, as a $1D$ array. The "plenoptic camera arrangement" Figure 1b puts all angular samples for pixel 1 (the first column) in a row, then all angular samples for pixel 2 (the second column) next to them, and so on. Now we encounter the problem at the boundaries discussed above. Out of 3 angular samples only one is left intact. The left and right pixels in each sub-image, θ_1 and θ_3, are lost.

To avoid this problem we rearrange the optical data as in Figure 1c. All spatial samples at a given angle θ are grouped together. In this way we get a coherent image of lots of pixels representing θ_1-samples, then next to them we place all θ_2 samples, and so on. Again boundary pixels are lost, but now

they are much fewer as a percentage of all pixels in a sub-image.

Obviously, the method is more efficient. Of course, any array of conventional cameras already samples the light field in exactly this way, so the technique is not new. What we gain here is an intuitive understanding of why this approach is better and also some motivation for developing new camera designs, as we discuss next.

### 2.3. Derivation of our camera design

In a traditional approach to light field photography we would use an array of cameras to capture an array of $2D$ images as in Figure 1c. For example, the classic arrangement of lenses from integral photography, shown in Figure 3a, produces just such a result. In this section, we develop a series of equivalent camera designs based on a formula from *affine optics*, which will be derived next. The proposed affine optics treatment of optical phase space can be used in other light field constructions.

Conventional Gaussian optics [GB94] is linear in the following sense. All the equations are written relative to the *optical axis*, which plays the role of origin (or zero point) in optical phase space ("light field space"), treated as a vector space. (See Figure 2.) For the purposes of this discussion, we will continue working in a $2D$ optical phase space, with spatial dimension x and angular dimension θ. More precisely, we will use θ to denote the tangent of the angle relative to the optical axis, at which a ray intersects a plane (line in our case) perpendicular to the optical axis. Finally, let f donate focal length.

With this notation, a lens is defined by the linear transform:

$$\begin{pmatrix} x' \\ \theta' \end{pmatrix} = \begin{pmatrix} 1 & 0 \\ -\frac{1}{f} & 1 \end{pmatrix} \begin{pmatrix} x \\ \theta \end{pmatrix}, \qquad (1)$$

and a space translation of the light field from one plane to another separated by distance T is represented by the linear transform
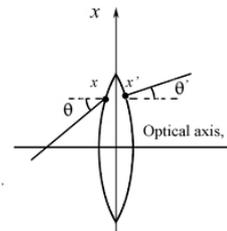
Figure 2: *Light field transformation at the plane of a lens.*

$$\begin{pmatrix} x' \\ \theta' \end{pmatrix} = \begin{pmatrix} 1 & T \\ 0 & 1 \end{pmatrix} \begin{pmatrix} x \\ \theta \end{pmatrix}. \qquad (2)$$

These and all other transforms used in Gaussian optics are *linear transforms* relative to the optical axis. Unfortunately, in linear optics there is no representation for a lens shifted from the optical axis, as we would need in light field photography. For example, in Figure 3a one would pick an arbitrary optical axis through one of the lenses, and then all the other lenses would be considered shifted relative to the optical axis, and not representable as linear transforms in this coordinate system. In linear optics we have no way of writing an expression for the radiance valid at the same time everywhere in a light field camera.

To derive a rigorous description of this new situation we need a more general mathematical framework that extends linear optics into what should be called *affine optics* (it adds translations to linear optics). A typical element representing an affine transform would be the prism. It tilts all rays by the same *fixed* angle α that depends only on the prism itself. Expressed in terms of the ray coordinates the prism transform is:

$$\begin{pmatrix} x' \\ \theta' \end{pmatrix} = \begin{pmatrix} x \\ \theta \end{pmatrix} + \begin{pmatrix} 0 \\ \alpha \end{pmatrix}. \qquad (3)$$

Now, a lens shifted a distance s from the optical axis would be treated as follows:

1. Convert to new lens-centered coordinates by subtracting s.

$$\begin{pmatrix} x' \\ \theta' \end{pmatrix} = \begin{pmatrix} x \\ \theta \end{pmatrix} - \begin{pmatrix} s \\ 0 \end{pmatrix} \qquad (4)$$

2. Apply the usual linear lens transform.

$$\begin{pmatrix} x'' \\ \theta'' \end{pmatrix} = \begin{pmatrix} 1 & 0 \\ -\frac{1}{f} & 1 \end{pmatrix} \begin{pmatrix} x - s \\ \theta \end{pmatrix} \qquad (5)$$

3. Convert to the original optical axis coordinates by adding back s.

$$\begin{pmatrix} q''' \\ \theta''' \end{pmatrix} = \begin{pmatrix} 1 & 0 \\ -\frac{1}{f} & 1 \end{pmatrix} \begin{pmatrix} x - s \\ \theta \end{pmatrix} + \begin{pmatrix} s \\ 0 \end{pmatrix} \qquad (6)$$

We can re-write this equation as:

$$\begin{pmatrix} q''' \\ \theta''' \end{pmatrix} = \begin{pmatrix} 1 & 0 \\ -\frac{1}{f} & 1 \end{pmatrix} \begin{pmatrix} x \\ \theta \end{pmatrix} + \begin{pmatrix} 0 \\ \frac{s}{f} \end{pmatrix}. \qquad (7)$$

Thus, we see that a shifted lens is equivalent to a lens with a prism. This result will be used to show that our proposed new designs are optically equivalent to arrays of cameras. This equivalence is exact.
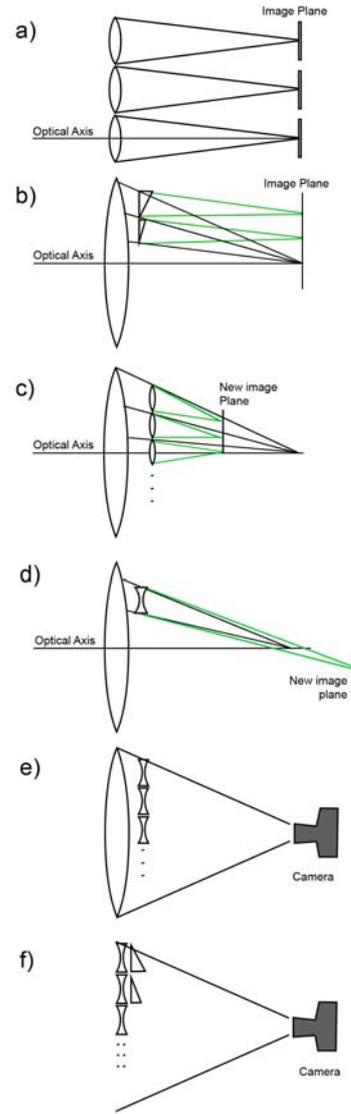
Figure 3: *Six designs of light field cameras. (a) Classical integral photography. (b) One lens and multiple prisms. (c) Main lens and a lens array. (d) Main lens and an array of negative lenses. (e) Same as (d), only implemented as external for the camera. (f) Example of external design of negative lenses and prisms that has no analog as internal.*

Based on equation 7, Figure 3a is optically equivalent to Figure 3b. The array of shifted lenses has been replaced with one central lens and a set of prisms. Equation 7 represents the relations between focal lengths, shifts, and prism angles that make the two systems equivalent.

Intuitively, different prisms tilt rays that would otherwise converge to the same point in different directions, separating

them onto different locations in the image plane and forming different sub-images. Those different sub-images are of the type Figure 1c, which is the more efficient design. (Note that intuition is not sufficient to convince us that this approach is exact. Intuition only tells us that "this should work at least approximately.")

Figure 3c is also self-explanatory from the point of view of intuition. The additional small lenses focus light rays closer than the original focal plane of the main lens. Thus they form individual images instead of being integrated into one image as in traditional one-optical-axis cameras. Again, this is "at least approximately correct" as a design, and we need formula 7 to prove that it is exactly correct and to find the exact values of the parameters (in terms of equivalence with Figure 3b.)

In more detail, each of the shifted lenses in Figure 3c is equivalent to a big lens on the optical axis and a prism. The big lens can be combined in one with the main lens, and we get equivalence with Figure 3b.

Figure 3d is similar, only with negative lenses. Designs Figure 3c and Figure 3d can be used practically if we integrate an array of 10-20 lenslets into the barrel of a conventional camera lens and use it with a high resolution camera as a compact light field camera.

Figure 3e describes a design external to the camera. It is used in this paper for the examples with 20 negative lenses. The whole optical device looks like a telephoto lens, which can be added as an attachment to the main camera lens. See Figure 6.

Figure 3f is our best design. We have implemented a version made up of 19 lenses and 18 prisms. See Figure 4. It is lightweight compared to similar design with a big lens. Also, an array of prisms is cheaper than a big lens.

Figure 4: *Our optical device consisting of lens-prism pairs.*

As in the design of Figure 3e, the camera sees an array of virtual images created by the negative lenses, in front of the optical device and focuses upon them. The prisms shift these images appropriately, so the result is as if the scene is viewed by an array of parallel cameras. Again the idea is that a camera with a lens shifted from the optical axis is equivalent to a camera on the axis, a lens and a prism. We should also note that practically, the role of the negative lenses is to expand the field of view in each image, and that the prisms can be viewed as making up a Fresnel lens focused at the camera's center of projection. Other external designs are possible with an array of positive lenses creating real images between the array of lenses and the main camera lens.

We have built prototypes for two of the designs: Figure 3e with 20 lenses, cut into squares, and Figure 3f with 19 lenses and 18 prisms. Because of chromatic problems with our prisms currently we produce better images with the design in Figure 3e, which is used to obtain the results in this paper. Also, our lenses and prisms for the design Figure 3f are not cut into squares, which leads to loss of pixels even with hexagonal packing, Figure 4. We are planning to build a version based on quality optical elements.

3. Synthetic aperture photography

Light fields can be used to simulate the defocus blur of a conventional lens, by re-projecting some or all of the images onto a (real or virtual) focal plane in the scene, and computing their average. Objects on this plane will appear sharp (in focus), while those not on this plane will appear blurred (out of focus) in the resulting image. This synthetic focus can be thought of as resulting from a large-aperture lens, the viewpoints of light field images being point samples on the lens surface. This method was proposed by Levoy and Hanrahan [LH96], first demonstrated by Isaksen et al. [IMG00], and goes under the name of *synthetic aperture photography* in current work [VWJL04, WJV*05]. It creates a strong sense of 3D; further, summing and averaging all the rays serves as a noise reduction filter, hence the resulting image has superior signal-to-noise ratio (SNR) compared to the original inputs.

The projection and averaging approach to synthetic aperture requires a dense light field. However, we are working with relatively sparse samplings comprised of 20 images. Simply projecting and averaging such an image set results in pronounced ghosting artifacts, essentially the result of aliasing in the sampled light field. Stewart et al. [SYGM03] explore reconstruction filters to reduce the aliasing in undersampled light fields; however, even with 256 images some artifacts remain.

Instead, we address the aliasing problem by generating more camera views than those provided directly by the camera array through view morphing [SD96]. This is equivalent to generating a *synthetic light field* by carefully interpolating between the samples in our sparse camera data. Fundamentally, this is possible because of the well known "redundancy" of the light field [LH96], which in the Lambertian

case is constant along angular dimensions at each point on the surface that is being observed. In the following subsections, we describe our method for filling out the light field and for using it to generate synthetic aperture images.

3.1. Synthetic light field by tri-view morphing

Our sampling consists of viewpoints that lie on a grid. We tessellate this grid into a triangular mesh, as illustrated in Figure 5. Our goal is to be able to fill in arbitrary viewpoints within the grid. As described below, we do this by computing warps that allow view morphing between each pair of views connected by an edge. These warps are then combined to allow barycentric interpolation of views within each triangle of viewpoints.

3.1.1. View morphing with segmentation-based stereo

View morphing [SD96] is a method for interpolating two reference images to generate geometrically correct in-between views from any point on the line connecting the two initial centers of projection. To achieve this effect, a correspondence is needed between the pair of images.

Recently, color segmentation approaches have gained in popularity for dense correspondence computation. They use color discontinuities to delineate object boundaries and thus depth discontinuities. Also, they model mixed color pixels at boundaries with fractional contributions (a.k.a. matting) to reduce artifacts at depth discontinuities.

We build on the segment-based optical flow work of Zitnick et al. [ZJK05]. The idea behind their method is to model each pixel's color as the blend of two irregularly-shaped segments with fractional contributions α and then solve for a mutual segmentation between a pair of images that gives rise to segments with similar shapes and colors. We modify their flow algorithm in two ways. First, between each pair of images, we require the matched segments to lie along epipolar lines. Second, we simultaneously compute epipolar flow between an image and two neighbors defining a triangle, so that the segments in each image are consistent between neighbors needed for tri-view morphing, described in the next subsection.

3.1.2. Tri-view blending

Seitz et al. [SD96] demonstrated that any linear combination of two parallel views gives a valid interpolated projection of the scene. Multiple image morphing [GW97, LWS98] has been used to extend two-view morphing to morphing among three or more views and into a complete image-based 3D system [Geo98]. Tri-view morphing [XS04] is a more recent system for creating the appearance of 3D via multi-image morphing, making use of the trifocal tensor to generate the warping transforms among three views.

Here we summarize our method for tri-view morphing within triangles on our camera grid. Given three images I_1, I_2 and I_3, we morph to the target image I_s using barycentric coefficients λ_1, λ_2 and λ_3. Let W_{ij} be the warping vector field (or "flow") from image I_i to image I_j, according to the disparity map from I_i to I_j obtained using the segmentation-based stereo algorithm from Section 3.1.1. Ideally, this warping function will convert image I_i into an image identical to I_j. In general, warping any image I by a vector field W will produce a new image denoted as $I(W)$. We warp each of the input images to I_s using affine (barycentric) combination of the three vector fields, and then we blend them together based on the same barycentric coefficients:

$$I_{out} = \sum_{i=1}^{3} \lambda_i I_i \left(\sum_{j=1}^{3} \lambda_j W_{ij} \right)$$

Note that we generally sample within the camera grid, so that the desired image is inside of a triangle defined by the three input images I_i, and then $\lambda_i \geq 0$ and $\Sigma_{i=1}^{3} \lambda_i = 1$. Extrapolation outside the grid is also feasible to some extent, in which case one or more barycentric coordinates will be negative.

3.2. Synthetic aperture rendering

To simulate the defocus of an ordinary camera lens, we first define an aperture location and size on the camera grid (see Figure 5). Then, we densely sample within this aperture using tri-view morphing. Finally, we determine an in-focus plane, project all images within the aperture onto this plane, and average.

4. Results

We have built working prototypes of two camera designs, shown in Figures 3e and 3f. All of our results are based on images taken with the former of these two designs, so that is the design that we describe in detail here.

4.1. Camera

Our implementation of the camera design from Figure 3e was built with an array of 4×5 negative lenses cut into squares and attached to each-other with minimal loss of space. Before being glued together the lenses were placed with their flat sides facing downward on a piece of glass, so we believe they are very well aligned on a plane and parallel to each other. Since all lenses have the same focal length, -105 mm, their focal points are on one plane. This plane is perpendicular to the direction of view to the precision of lens manufacturing.

We calibrate the camera centers using an off-the-shelf structure-from-motion (SFM) system [BL05] which recovers both the intrinsic and the extrinsic parameters of the

Figure 5: *The set of 20 images (middle) is a sparse light field captured with our camera. A close-up of one of the images is shown on the left. The hazy edges are defocused images of the boundaries of the lenses; for the results in this paper, we discard these contaminated pixels. Each vertex on the right represents one camera view. We decompose the camera plane into triangles illustrated on the right. Any novel camera view inside these triangles can be synthesized using tri-view morphing. The circular region represents a possible virtual aperture we want to simulate.*

camera. For the purposes of synthetic aperture, one could also pursue the calibration method discussed by Vaish et al. [VWJL04], in which relative camera positions are recovered.

Figure 6: *Our sparse light field camera prototype, with 2 positive lenses and an array of 20 negative lenses in front of a conventional camera.*

4.2. Renderings

With our camera prototype, twenty views are captured at a single exposure, with each view containing roughly 700 by 700 pixels. Twenty-four triangles are formed to cover the entire viewing space. The relative locations of all the cameras are recovered by running SFM on the 20 images. Once the size, location, and shape of a virtual lens is specified, we densely sample viewpoints using our tri-view morphing algorithm at one reference depth. All examples shown here were sampled with about 250 views. Sweeping through planes of different depths corresponds to shifting all views accordingly. By shifting and summing all the sampled views, we compute synthetic aperture images at different depths.

In the seagull example of Figure 7, we demonstrate refocusing at three different depths from near to distant. (Please see the supplemental materials to see a videos demonstrating a continuous change of focal depth, as well as the other results in this section.)

For the juggling example (with input images in Figure 5) we present three sets of results, shown in Figure 8. On the first row we show three synthesized novel views inside a triangle of input images. Despite the slight motion blur of the tennis balls, the interpolated views look realistic with clean and smooth boundaries. The second row shows three synthetic aperture images focusing at three different depths. The third row shows results focused on the juggler with varying depth of field. The effect is created with varying aperture size. The left image and the middle image have the exact same virtual aperture. Notice that the number of samplings makes a huge difference as the left uses only 24 views, revealing strong aliasing in blurred regions, while the middle image uses over 200 views. The right image shows an even larger aperture that spans outside the area of the input camera array, showing that view *extrapolation* also produces reasonable results for this application.

The reader is encouraged to see the electronic version of this paper for high resolution color images. The supplementary videos show sequences of synthetic aperture images as the focal plane sweeps through a family of planes that spans the depths of the scenes. The sharpness of objects on the focal plane together with the smooth blur indicates the accuracy of our technique. The size of the virtual aperture used in the seagulls example (Figure 7) and in most results of the juggling scene (Figure 8) is about one quarter of the entire viewing region.

5. Conclusion and future work

In this paper, we describe several practical light field camera designs with the specific application to synthetic aperture photography. We compare the two fundamental ways of approaching light field capture, and argue that an important point in the camera space for integral photography is in a sparse sampling of the angular dimensions of the light field in order to achieve better spatial resolution. As such, we explore how integral cameras can be used to produce results with higher spatial resolution than plenoptic cameras, using the same image sensor.

We further draw upon state-of-the-art computer vision techniques as a post- processing tool to interpolate or "fill in" the sparse light field. We demonstrate the effectiveness of this framework with realistic refocusing and depth of field results. Averaging lots of intermediate views not only reduces sampling errors, but also makes errors caused by stereo matching much more tolerable, which is one of the insights of our approach.

Most of the computing cycles are spent on generating in-between views. An analysis on the sampling bounds would be helpful for better efficiency. How densely does one have to sample the viewing space in order to create non-aliased results? Furthermore, it would be interesting to explore the possibility of skipping the entire process of view interpolation and realizing refocusing directly from the disparity map.

We use twenty views in our camera implementation. For typical scenes we get good results, but for scenes with more complex 3*D* structure we begin to observe artifacts. A detailed study of the relationship between optimal sampling rate and 3*D* scene complexity would be useful. It might be possible to dynamically adjust the number of captured views based on scene geometry so that results with optimal resolution are achieved.

In the last few years, we have experienced the fast bloom of digital photography. Sensors are gaining in resolution. Capturing a light field with a single exposure becomes achievable in a realistic hand-held camera. This adds a whole new dimension to digital photography with the possibility of capturing a sparse light field with a compact camera design, and later post-processing based on computer vision. We hope that our work will inspire others to explore the possibilities in this rich domain.

References

[AW92] ADELSON T., WANG J.: Single lens stereo with a plenoptic camera. *IEEE Transactions on Pattern Analysis and Machine Intelligence* (1992), 99–106. 2

[BL05] BROWN M., LOWE D. G.: Unsupervised 3d object recognition and reconstruction in unordered datasets. In *Proceedings of 5th International Conference on 3D Imaging and Modelling (3DIM)* (2005), pp. 21–30. 6

[GB94] GERRARD A., BURCH J. M.: Introduction to matrix methods in optics. 3

[Geo98] GEORGIEV T.: 3d graphics based on images and morphing. *US Patent 6268846* (1998). 6

[GGSC96] GORTLER S. J., GRZESZCZUK R., SZELISKI R., COHEN M. F.: The lumigraph. *ACM Trans. Graph.* (1996), 43–54. 1

[GS85] GUILLEMIN V., STERNBERG S.: Symplectic techniques in physics. 1

[GW97] GEORGIEV T., WAINER M.: Morphing between multiple images. *Tech. Rep.* (1997). 6

[IMG00] ISAKSEN A., MCMILLAN L., GORTLER S. J.: Dynamically reparameterized light fields. *ACM Trans. Graph.* (2000), 297–306. 1, 5

[Ive28] IVES H.: Camera for making parallax panoramagrams. *J. Opt. Soc. Amer., 17* (1928), 435–439. 2

[LH96] LEVOY M., HANRAHAN P.: Light field rendering. *ACM Trans. Graph.* (1996), 31–42. 1, 3, 5

[Lip08] LIPPMANN G.: Epreuves reversible donnant la sensation du relief. *J. Phys. 7* (1908), 821–825. 1, 2

[LWS98] LEE S., WOLBERG G., SHIN S.: Polymorph: Morphing among multiple images. *IEEE Computer Graphics and Applications* (1998). 6

[NLB*05] NG R., LEVOY M., BRĹĘDIF M., DUVAL G., HOROWITZ M., HANRAHAN P.: Light field photography with a hand-held plenoptic camera. *Tech. Rep.* (2005). 1, 2

[NYH01] NAEMURA T., YOSHIDA T., HARASHIMA H.: 3d computer graphics based on integral photography. *Optics Express, Vol. 8, 2* (2001). 2

[OAHY99] OKANO F., ARAI J., HOSHINO H., YUYAMA I.: Three-dimensional video system based on integral photography. *Optical Engineering, Vol. 38, 6* (1999). 2

[OHAY97] OKANO F., HOSHINO H., ARAI J., YUYAMA I.: Real-time pickup method for a three-dimensional image based on integral photography. *Applied Optics, Vol. 36, 7* (1997), 1598–1603. 2

[SD96] SEITZ S. M., DYER C. R.: View morphing. *ACM Trans. Graph.* (1996), 21–30. 5, 6

[SH02] STEVENS R., HARVEY T.: Lens arrays for a three-dimensional imaging system. *Journal of Optics A, Vol. 4* (2002). 2

[SYGM03] STEWART J., YU J., GORTLER S. J., MCMILLAN L.: A new reconstruction filter for undersampled light field. *Eurographics Symposium on Rendering* (2003), 150–156. 5

[Tys91] TYSON R.: Principles of adaptive optics. In *Academic Press* (1991). 2

[VWJL04] VAISH V., WILBURN B., JOSHI N., LEVOY M.: Using plane + parallax to calibrate dense camera arrays. In *Proceedings of IEEE Conference on Computer Vision and Pattern Recognition (CVPR)* (2004). 5, 7

[WJV*05] WILBURN B., JOSHI N., VAISH V., TALVALA E., ANTUNEZ E., BARTH A., ADAMS A., LEVOY M., HOROWITZ M.: High performance imaging using large camera arrays. In *ACM Trans. Graph.* (2005). 2, 5

[XS04] XIAO J., SHAH M.: Tri-view morphing. *Computer Vision and Image Understanding 96*, 3 (2004), 345–366. 6

[ZJK05] ZITNICK C. L., JOJIC N., KANG S.: Consistent segmentation for optical flow estimation. In *Proceedings of IEEE International Conference on Computer Vision (ICCV)* (2005). 6

Figure 7: *Synthetic aperture photography of flying birds. Refocusing to different depths.*

(a) Three novel views generated using tri-view morphing.

(b) Synthetic aperture results with the focal plane moving from near to far.

(c) Synthetic aperture results with varying depth of field. (Left image demonstrates sparse sampling.)

Figure 8: *Synthetic aperture photography of human motion focusing at different depths.*