Spatial Analysis of Discrete Plenoptic Sampling

Andrew Lumsdaine[a], Todor Georgiev[b], and Georgi Chunev[a]

[a]Indiana University, Bloomington, IN, USA;
[b]Adobe Systems, San Jose, CA, USA

ABSTRACT

Plenoptic cameras are intended to fully capture the light rays in a scene. Using this information, optical elements can be applied to a scene computationally rather than physically—allowing an infinite variety of pictures to be rendered after the fact from the same plenoptic data. Practical plenoptic cameras necessarily capture discrete samples of the plenoptic function, which together with the overall camera design, can constrain the variety and quality of rendered images. In this paper we specifically analyze the nature of the discrete data that plenoptic cameras capture, in a manner that unifies the traditional and focused plenoptic camera designs. We further present a resolution analysis for plenoptic cameras and develop design guidelines for maximizing resolution. A generalized rendering algorithm is presented that minimizes artifacts resulting from the lower resolution angular sampling that accompanies high-resolution spatial sampling. Experimental results using a real-time GPU implementation of our algorithms demonstrates the effectiveness of our approach.

Keywords: plenoptic camera, rendering, discrete sampling

1. INTRODUCTION

Integral photography has more than 100 years of history, starting with Ives[1] and Lippmann.[2] Lippmann motivated his work in this area by observing that even the "most perfect photographic print only shows one aspect of reality; it reduces to a single image fixed on a plane, similar to a drawing or a hand-drawn painting." With integral photography, he attempted instead to render infinitely more—"the full variety offered by the direct observation of objects." Unfortunately, because of inherent limitations of available technologies, the potential of integral photography was not realized.

Integral photography has recently re-emerged with the introduction of the plenoptic camera, a device that captured the distribution of light rays in space (i.e., the 4D plenoptic function) in order to capture 3D imagery and solve computer vision problems.[3] The lightfield and lumigraph, introduced to the computer graphics community respectively in[4] and,[5] established a theoretical framework for analyzing the plenoptic function. A handheld plenoptic camera, along with new methods of processing, was introduced by Ng in.[6]

Practical plenoptic cameras necessarily capture discrete samples of the plenoptic function, which together with the overall camera design, can limit the variety and quality of rendered images. The discretization comes from two sources: discrete microlenses in the microlens array and discrete pixels in the camera sensor. The effect of these sources is to impose a specific geometry on the overall sampling of the plenoptic function by plenoptic cameras. This geometry is subject to computational transforms applied when rendering images from the captured plenoptic function. Understanding the geometry of the captured lightfield is therefore fundamental to understanding the properties of the rendered image (most notably, resolution and depth of field).

In this paper we analyze plenoptic cameras specifically from the point of view of discrete sampling. We use the optical properties of plenoptic cameras to derive the geometry of discrete plenoptic function capture. Based on this geometry, we derive expressions for expected resolution from a captured plenoptic function as well as necessary conditions in the optical design for overcoming the resolution limitations of the traditional plenoptic camera.

2. RELATED WORK

The basic structure of the focused plenoptic camera can first be seen in the early work of Lippmann,[2] who considered the array of lenslets as an array of cameras focused on the photographed object. Lippmann proposed the idea of using a lenslet array to capture the radiance of a scene and produce what he called *integral photographs*. More recently, the basic principles of plenoptic cameras have been independently considered by a number of researchers including Ng,[7] Fife,[8] and Lumsdaine.[9]
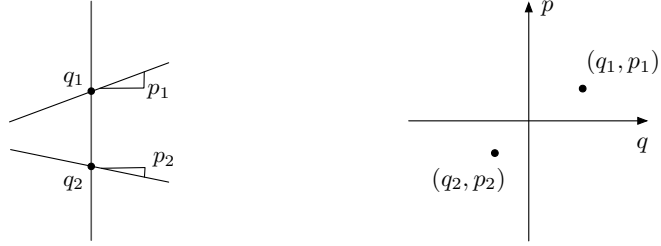
Figure 1: Rays are represented with coordinates q and p. The space of all rays comprises the q-p phase space.

In the case of,[9] a distinction is made between the plenoptic camera presented there (the "focused plenoptic camera") and plenoptic cameras presented earlier ("traditional plenoptic cameras"). It is shown that higher spatial resolution can be achieved with the focused plenoptic camera than with traditional plenoptic cameras, although it is unclear to what extent that is due to the rendering algorithm used rather than the camera geometry itself. Moreover, the approach described there seemed to be suited only to a single depth plane. Noticeable artifacts resulted in the rendered image for portions of the scene that were not in the focal plane.

In,[7] Ng considers different configurations of the handheld plenoptic camera in which the sensor is not located precisely at the focal plane of the microlenses. The microlens arrays in[8] are focused on the image plane and the use of the overlapping microimages for high-resolution 2D image reconstruction as well as for depth estimation is discussed. While[8] briefly discusses image rendering, the significant contribution of the work is the multi-aperture image sensor itself. Given the similarity between the multi-aperture image sensor and the focused plenoptic camera, we should expect that this sensor also samples similarly to the camera reported in.[9] Accordingly, we would expect that the multi-aperture system would benefit from rendering algorithms specifically targeted for that situation.

Concerns about sampling of lightfield data, and consequential aliasing during rendering, have long been present in the lightfield literature. A number of different approaches have been proposed to deal with under-sampling, including optical methods[10] and computational methods.[11–13] In these papers, the primary concern is spatial, rather than angular, under-sampling. A discussion about balancing the tradeoffs between spatial and angular resolution appears in the paper.[14] The authors propose that lower angular resolution could be overcome via interpolation (morphing) techniques to make more sensor real-estate available for positional information.

3. BACKGROUND

In this section we briefly recapitulate terminology and notation used in the remainder of this paper.

The plenoptic function[3] (also called the *lightfield*[4] or *radiance*[15]) is a density function describing the light rays in a scene. Since the radiance is a density function over the ray space, we describe radiance transformations via transformations applied to elements of the underlying ray space. Rays in three dimensional space are represented as four dimensional vectors: Two coordinates are required to describe position and two are required to describe direction. Following the physics-based convention of,[16] we denote the radiance at a given plane perpendicular to the optical axis as $r(q, p)$, where q describes the location of a ray in the plane and p describes its direction. (These coordinates are also used in optics texts such as.[17,18]) For illustrative purposes, and without loss of generality, we adopt the convention of a two-dimensional q-p plane in this paper.

Translation and refraction by a lens are two fundamental transformations that can be applied to rays. Rays are transformed due to translation a distance t in the direction of the optical axis according to $(q', p') = (q + tp, p)$, corresponding to a linear transformation $x' = T_t x$, where

$$T_t = \begin{bmatrix} 1 & t \\ 0 & 1 \end{bmatrix}. \tag{1}$$

Similarly, rays are transformed due to optical refraction of a lens with focal length f according to $(q', p') = (q, p - \frac{1}{f}q)$, the linear transformation for which is $x' = L_f x$ where

$$L_f = \begin{bmatrix} 1 & 0 \\ -\frac{1}{f} & 1 \end{bmatrix}. \tag{2}$$
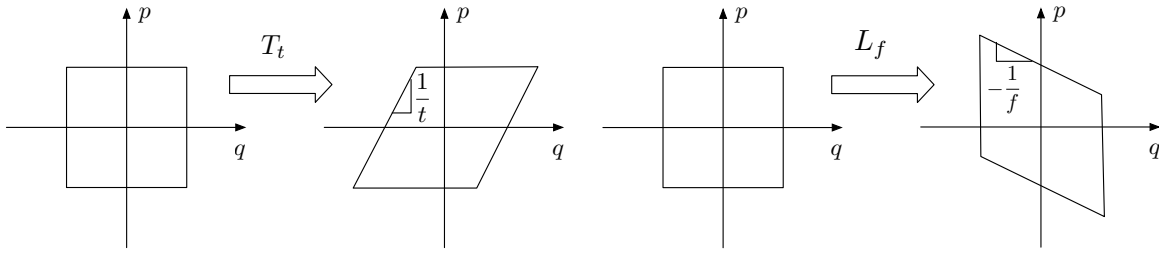
(a) Shearing due to translation.   (b) Shearing due to refraction by a lens.

Figure 2: Translation and refraction by a lens act as shearing transformations in q-p phase space.
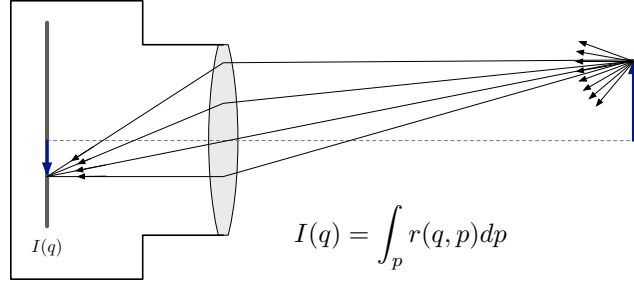

$$I(q) = \int_p r(q,p)dp$$

Figure 3: The sensor pixels in a traditional camera capture an image by physically integrating the intensities of all of the rays impinging on them.

There is an important geometric interpretation to these transformations. In phase space, translation and refraction by a lens are both *shearing transformations.* As shown in Figure 2, translation is a horizontal shear, while refraction by a lens is a vertical shear. The shearing property of translation in particular will be an important part of the analysis in the body of this paper.

A two-dimensional image is created from the four-dimensional radiance by integration over all rays incident at a given q. Given a radiance $r(q,p)$ at the image plane of a camera, an image $I(q)$ is rendered for a given range of the available p values according to

$$I(q) = \int_p (q,p)dp. \tag{3}$$

4. DISCRETE PLENOPTIC CAMERAS

As shown in Figure 3, a traditional camera captures an image by physically integrating the intensities of all of the rays impinging on each sensor pixel.

A plenoptic camera, on the other hand, captures each ray separately. One approach to separating, and then individually capturing, the rays in a scene is to put a pinhole where the sensor pixel would be while placing the sensor itself some distance b behind the pinhole. In this case, the rays that converge at the pinhole will diverge as they propagate behind the pinhole. The separate pixels in the sensor now capture separate rays. I.e., the intensity as a function of position at the sensor represents the radiance as a function of direction at the position q of the pinhole. A single image captured behind the pinhole thus samples a vertical stripe in the q-p plane, while an array of pinholes samples a grid.

4.1 Discrete Lippmann Sensors

Although the ideal pinhole makes an ideal "ray separator," in practice microlenses are used instead to gather sufficient light and to avoid diffraction effects. Figure 4 shows a diagram of such a sensor, which we will refer to as the "Lippmann sensor" (after Gabriel Lippmann who first proposed it[2]). In the diagram, b is the distance from the sensor the microlens plane and a is the distance from the microlens plane to the main lens image plane. The microlens focal length is f; a, b,
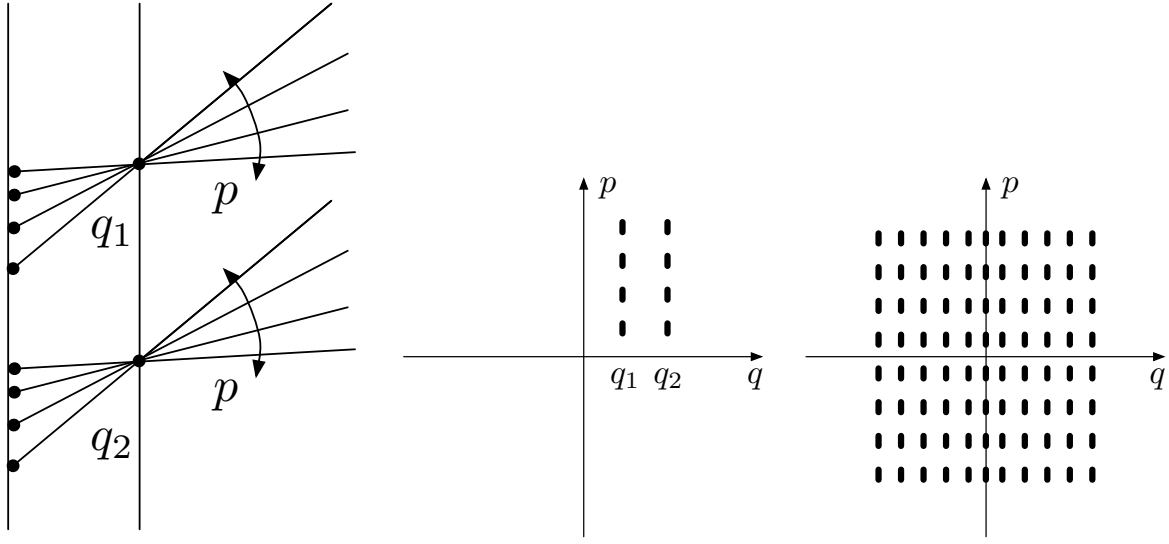
(a) Rays converging at a pinhole will separate from each other as they travel behind it and can be captured individually by a sensor behind the pinhole.

(b) Individual pinholes sample one position in the q-p plane, while individual pixels sample different directions.

(c) A single image captured behind the pinhole thus samples a vertical stripe in the q-p plane, while an array of pinholes samples a grid.
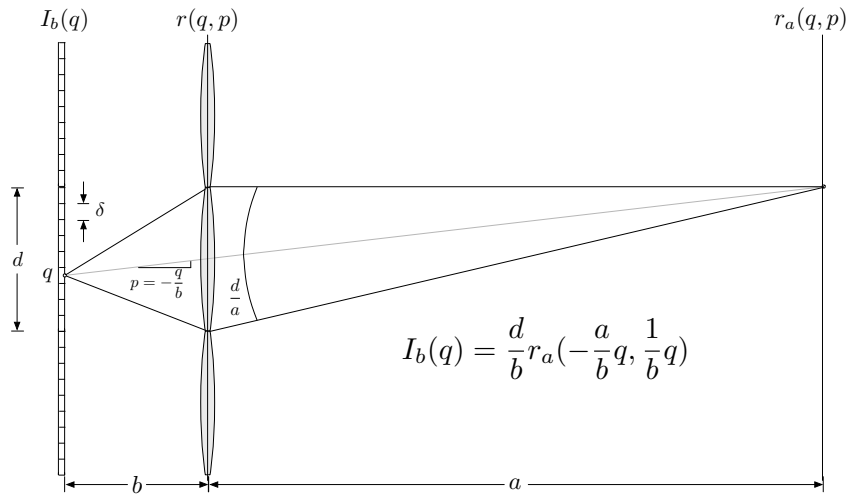

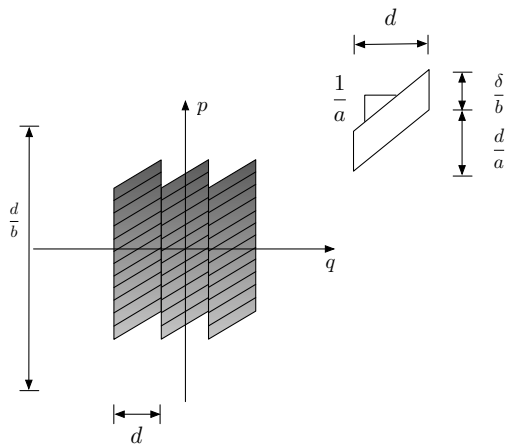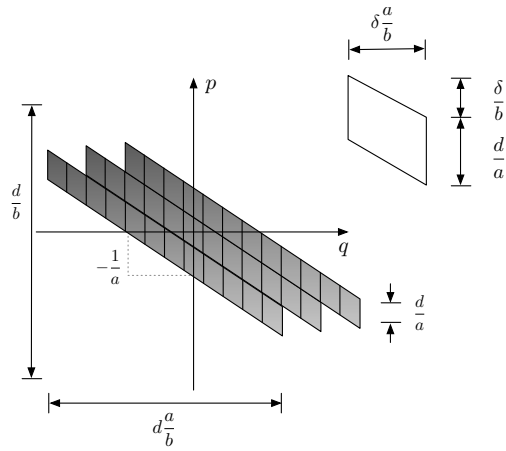$$I_b(q) = \frac{d}{b} r_a(-\frac{a}{b}q, \frac{1}{b}q)$$

Figure 4: Geometry of Lippmann sensor for a plenoptic camera. The CCD (or CMOS) sensor is placed at a distance b behind an array of microlenses. In one notable limit, $b \to f$ and $a \to \infty$.
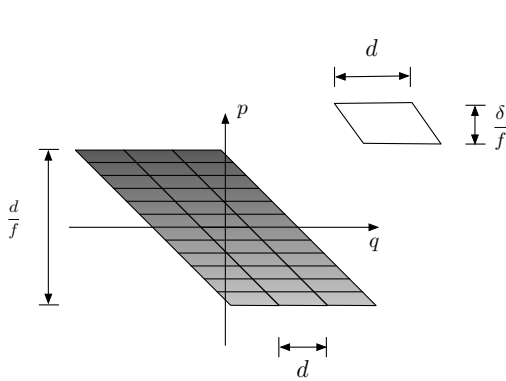
(a) Sampling by the focused Lippmann sensor at the microlens array. This radiance is related to the one shown in Figure 5b by a shearing transform of distance a.

(b) Sampling by the focused Lippmann sensor at a finite distance in front of the microlenses, where a, b, and f satisfy the lens equation. Note that shearing decreases the width of the spatial region sampled by a single pixel.
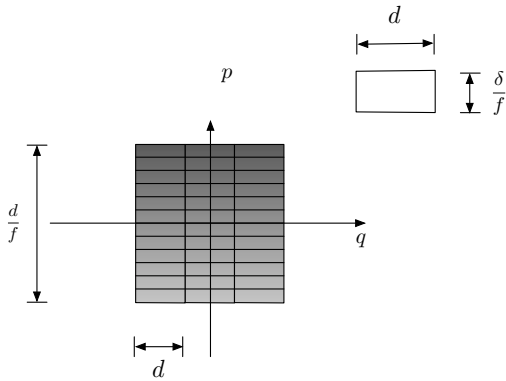
(c) Sampling by the traditional Lippmann sensor at the microlens array when the distance between the microlens array and the sensor is equal to the microlens focal length (i.e., when $b = f$).

(d) Sampling by the traditional Lippmann sensor at distance a in front of the microlenses. In this case, shearing does *not* decrease the width of the spatial region sampled by a single pixel.

Figure 5: Radiance sampling by the Lippmann sensor. The geometry of a single pixel for each case is shown in the upper right.

and f are assumed to satisfy the lens equation $1/a + 1/b = 1/f$. Sensor pixels have size δ and without loss of generality we take d to be the microlens aperture and the spacing between microlenses.

More recently, a specific form of the Lippmann sensor was proposed by Ng,[6] in which the distance b was taken to be equal to the microlens focal length f. In this case, as has been derived elsewhere,[6] we have the following expression for how the image captured on the sensor (I_f) samples the radiance at the microlens plane (r):

$$I_f(q) = \frac{d}{f} r(0, \frac{1}{f} q).$$ (4)

This sampling is shown graphically in Figure 5c.

An alternative form of the Lippmann sensor was proposed by Lumsdaine and Georgiev,[9] in which the distance b was chosen not to be equal to f in order to form a relay system with the camera main lens. To distinguish this case from the specific case where $b = f$, we will refer to this sensor as the focused Lippmann sensor (in keeping with the terminology of the "focused plenoptic camera"). In this case, as derived in,[9] we have the following expression for how the image captured on the sensor (I_b) samples the radiance at the microlens focal plane (r_a):

$$I_b(q) = \frac{d}{b} r_a(-\frac{a}{b} q, \frac{1}{b} q)$$ (5)

This sampling is shown graphically in Figure 5b. As we discuss below, it is important to note that r_a and r represent the radiance in the scene at different planes.

As reported in,[9] each microimage captured by the focused Lippmann sensor captures a slanted stripe of the radiance at distance a in front of the microlens array. This difference was described as a significant one in[9] because it allowed significantly higher spatial resolution when rendering.

We note, however, that the difference is not as significant as may have been described. A comparison of Figure 5b with Figure 5c is not necessarily the right one to make to compare the traditional plenoptic camera and the focused plenoptic camera (equivalently, Lippmann sensors with spacing f and spacing b). The former case describes the radiance at a distance a in front of the sensor, whereas the latter describes it at the sensor. Indeed, if we instead consider the radiance captured by the focused Lippmann sensor as it is captured at the sensor itself, we obtain the sampling as shown in Figure 5a and see that the focused Lippmann sensor captures the radiance in a manner quite similar to the traditional sensor (compare to Figure 5c).

The primary difference between the radiance as captured by Figure 5c) and Figure 5a) is the tilt (of $\frac{1}{a}$) in the latter case of the region sampled by each pixel. (Compare the pixel geometry in Figure 5c with the pixel geometry in Figure 5a. This small difference turns out to be the significant one between the two types of Lippmann sensors, as described below in Section 5.

Finally, we remark that the sampling geometry is determined both by the microlens dimension (d) as well as by the pixel size δ.

4.2 Camera Design

Another aspect of plenoptic camera design is the main lens optics. The main lens of the camera will map the light in the exterior world into the light in the interior world of the camera. This mapping will present a certain plenoptic function inside the camera. However, it is important to distinguish between how the different Lippmann sensors sample the radiance with which they are presented and what radiance they are presented with. In the case of Ng's hand-held plenoptic camera, the main lens focal plane (the primary plane of photographic interest) was mapped to the microlens array. In the case of the focused plenoptic camera, the main lens focal plane was mapped to a distance a in front of the microlens array.

5. RESOLUTION

5.1 Spatial Resolution Analysis

To allow basic comparisons between different samplings of the $q - p$ plane, we use the following definition of spatial resolution.
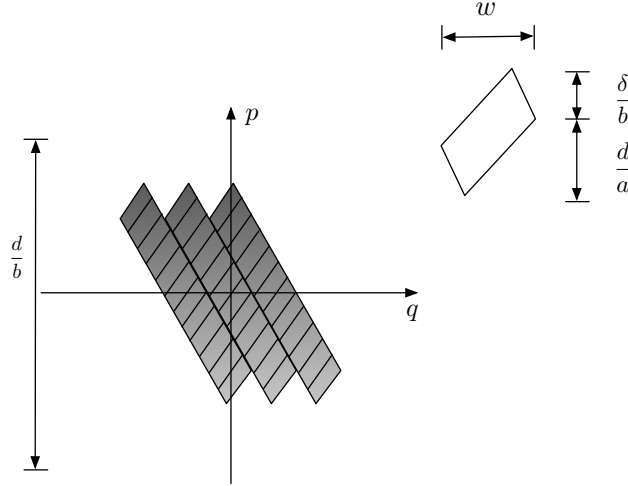
Figure 6: A pixel sampling the radiance at distance c in front of the sensor.

Definition. The *spatial resolution floor* $R_{sf}(z)$ of a sampled plenoptic function at plane z is the total spatial width of the function divided by the sample width of a single pixel.

Consider the plenoptic function sampled as shown in Figure 5c and 5a. In both of these cases, the total width of the plenoptic function is the number of microlenses times the width of a microlens. Similarly, in both cases, the sample width of a pixel is the width of one microlens. Hence, the spatial resolution floor of the radiance sampled at the microlens array is equal to the number of microlenses. Or,

$$R_{fb}(0) = R_{ff}(0) = N, \tag{6}$$

where N is the number of microlenses in the Lippmann sensor and $R_{ff}(z)$ and $R_{fb}(z)$ respectively represent the spatial resolution floors for the traditional and focused Lippmann sensors.

Referring to Figure 5c we can also see that for the focused Lippmann sensor,

$$R_{fb}(a) = NK/\frac{a}{b} = NM \tag{7}$$

where K is the number of pixels per microlens and $M = K/\frac{a}{b}$. Note that full sensor resolution is KN so that the focused Lippmann sensor achieves a spatial resolution floor equal to the full sensor resolution divided by the minification factor $\frac{a}{b}$. (This result agrees with that in.[9])

A more interesting question is regarding the achievable resolution at arbitrary locations in front of the sensor. Let us consider the width of a pixel from Figure 5a at a distance c in front of the sensor. Figure 6 shows this case, in which a shear corresponding to the translation by distance c has been applied. Some straightforward geometric manipulations (which we omit for space reasons) show that the resolution is:

$$R_{fb}(c) = NK\frac{ab}{Kb|a-c|+ac} = R_{fb}(a)\frac{a}{bK|1-\frac{c}{a}|+c} \tag{8}$$

When $a = c$, this reduces to equation (7).

Note also that maximum resolution for the focused Lippmann sensor is achieved when $a = c$. This result can readily be seen geometrically by referring to Figure 5b, which shows a pixel when $a = c$. In that case, a different value of $c \neq a$ will apply a shear to the pixel in one direction or the other, depending on whether $c > a$ or $c < a$. In either case, the shear will act to increase the width of the pixel.

Applying the same analysis to pixels from Figure 5c illustrates the importance of focusing. In the (ideal) traditional Lippmann sensor, the region of the plenoptic function sampled by each pixel is constant as a function of p. Hence, applying a shearing transformation will not decrease the width of the pixel. As a result:

$$R_{ff}(z) \leq N \tag{9}$$

for all z. Moreover, as can be seen from Figures 5c and 5d, the minimum pixel width (and maximum resolution) is obtained at $z = 0$, i.e., at the microlens array.

5.2 The Focused Plenoptic Principle

The difference between the traditional Lippmann sensor and the focused Lippmann sensor is only one of optical geometry, i.e., the magnitude of distance between the sensor and the microlenses. In the case of the traditional Lippmann sensor we have the maximum resolution equal to N and obtained at the $z = 0$. In the case of the focused Lippmann sensor we have the maximum resolution equal to M times N and obtained at $z = a$. The question then arises, how does the traditional Lippmann sensor become the focused Lippmann sensor as we move the sensor away from $b = f$? That is, there is a qualitative difference between the behavior of the traditional and the focused Lippmann sensors, yet there is a continuous family of sensors parameterized by the distance b between the microlenses and the sensor.

Consider the sampling of the pixel shown in Figure 5c. Clearly, as a shear is applied to this pixel, it's width will increase (as shown in Figure 5d). Consider instead the pixel shown in Figure 5a. As a shear is applied, it will assume the shape shown in Figure 6 and assume its minimum value as shown in Figure 5b. These two different responses to shearing produce the different behaviors of the traditional and focused Lippmann sensors.

Note, however, that the shearing behavior of the pixel depends on its geometry. In particular, the pixel will only decrease in width if

$$\frac{d}{a} > \frac{\delta}{b}. \tag{10}$$

Geometrically, we must have the upper left corner of the pixel be at a lower value than the lower right corner. Assuming the lens equation holds and substituting for a in equation (10) we obtain the following condition that defines the boundary between traditional and focused Lippmann sensors:

$$b > f(1 + \frac{\delta}{d}). \tag{11}$$

Note that this condition is defined solely in terms of the optical geometry of the sensor.

Any sensor for which (11) holds will exhibit focused Lippmann sensor properties; any sensor for which (11) does not hold will exhibit traditional Lippmann sensor properties. Hence, we refer to equation (11) as the "focused plenoptic principle."

REFERENCES

[1] Ives, F., "US patent 725,567," (1903).

[2] Lippmann, G., "Épreuves réversibles. Photographies intégrales," *Académie des sciences* , 446–451 (March 1908).

[3] Adelson, E. and Bergen, J., "The plenoptic function and the elements of early vision," in [*Computational Models of Visual Processing*], 3–20, MIT Press, Cambridge, MA (1991).

[4] Levoy, M. and Hanrahan, P., "Light field rendering," *Proceedings of the 23rd annual conference on Computer Graphics and Interactive Techniques* (Jan 1996).

[5] Gortler, S. J., Grzeszczuk, R., Szeliski, R., and Cohen, M. F., "The lumigraph," *ACM Trans. Graph.* , 43–54 (1996).

[6] Ng, R., Levoy, M., Bredif, M., Duval, G., Horowitz, M., et al., "Light field photography with a hand-held plenoptic camera," Tech. Rep. 2005-02, Stanford University Computer Science (Jan 2005).

[7] Ng, R., *Digital light field photography*, PhD thesis, Stanford University, Stanford, CA, USA (2006). Adviser-Patrick Hanrahan.

[8] Fife, K., Gamal, A. E., and Wong, H.-S. P., "A 3Mpixel multi-aperture image sensor with $0.7\mu m$ pixels in $0.11\mu m$ CMOS," in [*IEEE ISSCC Digest of Technical Papers*], 48–49, IEEE (February 2008).

[9] Lumsdaine, A. and Georgiev, T., "The focused plenoptic camera," in [*IEEE International Conference on Computational Photography (ICCP)*], IEEE (2009).

[10] Levoy, M. and Hanrahan, P., "Light field rendering," *ACM Transactions on Graphics* , 31–42 (1996).

[11] Chai, J., Chan, S., Shum, H., and Tong, X., "Plenoptic sampling," *ACM Trans. Graph.* , 307–318 (2000).

[12] Isaksen, A., McMillan, L., and Gortler, S. J., "Dynamically reparameterized light fields," *ACM Trans. Graph.* , 297–306 (2000).

[13] Stewart, J., Yu, J., Gortler, S. J., and McMillan, L., "A new reconstruction filter for undersampled light field," *Eurographics Symposium on Rendering* , 150–156 (2003).

[14] Georgiev, T., Zheng, K., Curless, B., Salesin, D., and et al., "Spatio-angular resolution tradeoff in integral photography," *Proc. Eurographics Symposium on Rendering* (Jan 2006).

[15] Nicodemus, F. E., ed., [*Self-study manual on optical radiation measurements*], National Bureau of Standards (1978).

[16] Guillemin, V. and Sternberg, S., "Symplectic techniques in physics," (1985).

[17] Gerrard, A. and Burch, J. M., [*Introduction to Matrix Methods in Optics*], Dover Publications (1994).

[18] Wolf, K. B., [*Geometric Optics on Phase Space*], Springer (2004).