Light-Field Capture by Multiplexing in the Frequency Domain

Todor Georgiev, Chintan Intwala, Derin Babacan
Adobe Systems Incorporated
345 Park Ave, San Jose, CA 95110
tgeorgie@adobe.com, cintwala@adobe.com, babacan@adobe.com

Abstract

In previous works light-field capture has been analyzed in spatio-angular representation. A light-field camera samples the optical signal within a single photograph by multiplexing the 4D radiance onto the physical 2D surface of the sensor.

In this paper we develop a theory of multiplexing the radiance in the frequency domain by optically mixing different spatial and angular frequency components and then capturing the signal by the sensor. We derive a mathematical method for recovering the multiplexed spatial and angular information from the frequency representation.

We have developed several prototypes of those frequency domain light-field cameras. They fall into two classes: (1) based on arrays of "active" optical elements, like lenses and prisms, and (2) based on "passive" optical elements, like masks, meshes or pinholes. Both types of cameras are understood and described in a single mathematical formalism in frequency domain. We show examples of novel view synthesis and refocusing based on this frequency domain multiplexing and discuss a number of advantages of the new method.

Table of Contents

1 Introduction

A central area of research in computational photography is capturing "light itself" as opposed to capturing a flat 2D picture. Advantages of this *light-field* or *integral* photography are gaining information about the 3D structure of the scene, and the new ability of optical manipulation or editing of the images, like refocusing and novel view synthesis.

The *light itself*, or *light-field* is mathematically described by the *radiance density function*, which is a complete representation of light energy flowing along "all rays" in 3*D* space. This density is a field defined in the 4*D* domain of the optical phase space, the space of all lines in 3*D* with symplectic structure [7].

Conventional cameras, based on 2D image sensors, are simply integration devices. In a typical setting, they integrate over a 2*D* aperture to produce a 2*D* projection of the full 4*D* radiance. Integral Photography [8, 11] was proposed more than a century ago to "undo" the integration and measure the complete 4*D* light field arriving at all points on a film plane or sensor.

As demonstrated by Levoy and Hanrahan [10] and Gortler et. al. [6], 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.

Recently, Ng et al. [14] have shown that a full 4*D* 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.

A number of works have developed techniques for analyzing radiance in the frequency domain [2, 3, 13]. Several important results have been derived, among which are application of Poisson summation formula to depth representation of scenes, light fields and displays, light transport and optical transforms, Fourier slice theorem applied to refocusing, and others. However, frequency analysis has not been applied to the understanding and design of light-field cameras.

2 Previous work

2.1 Light-field photography

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

1. The earliest works of Ives [8], Lippmann [11] and Ives [9] among others, known as *integral photography*, used arrays of pinholes or lenses 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., [21]). 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., [16]). Other related works [15, 12, 18] partition the 4D radiance function into a 2D array of 2D images, and project this array onto the surface of the film or sensor.

2. The more recent approaches of Adelson et al. [1] and Ng et al. [14], 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 [19], which was also proposed a century ago to study wavefront shape in optics.

These two approaches are based on understanding radiance as a distribution in 4D position-angle space, where the goal is to map it to the 2D surface of the sensor.

2.2 Frequency domain representation

Recent works [2, 3, 13] have analyzed radiance in frequency representation. Let $r(x)$ be the radiance in conventional $x-$space. It can be represented in frequency domain as follows:

$$R(\omega) = \int r(x)e^{i\omega \cdot x}dx \qquad (1)$$
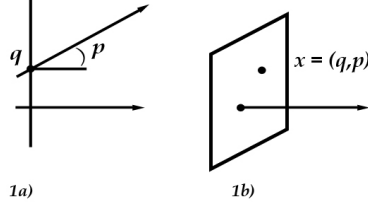
_1a)_   _1b)_

Figure 1: (a) Geometric representation of a ray as position and angle in an optical system. (b) Same ray described as a point, or a vector $x = (q, p)$, in a 2D space.

We are using the following notations. The spatio-angular coordinates of a ray at a given plane orthogonal to the optical axis are represented as a vector

$$x = \begin{pmatrix} q \\ p \end{pmatrix}, \tag{2}$$

where q is the location of ray-plane intersection, and p is a vector defining the two angles of that ray at location q. Following texts on optics [7, 5], we use paraxial approximation assuming the angle is small. A 2-dimensional vector representation of a ray is shown in Figure 1.

The spatial frequency ω_q and the angular frequency ω_p are represented in a similar way as a 4D vector:

$$\omega = \begin{pmatrix} \omega_q \\ \omega_p \end{pmatrix}. \tag{3}$$

To simplify our text and figures we will often use 2D radiance with 1-dimensional position q and angle p for each ray. The dot product is defined as usual, $\omega \cdot x = \omega_q q + \omega_p p$.

2.3   Transformations of the radiance

We will summarize and extend some results of previous works like [2, 3, 4] about transformations of radiance in optical systems. A ray x is transformed as explained in the following.

Both lens L and translation T can be described by linear transforms $x' = Ax$ of the ray as a position-angle vector (2) by the following matrices:

$$L = \begin{pmatrix} 1 & 0 \\ -\frac{1}{f} & 1 \end{pmatrix}, \tag{4}$$

$$T = \begin{pmatrix} 1 & t \\ 0 & 1 \end{pmatrix}, \tag{5}$$

where f is the focal length of the lens, and t is the translation (distance of flight). A prism deviates the ray by a fixed angle p_{prism}, so that $p' = p + p_{prism}$.

The combined action of several such elements is described by the composition of all those elements. This gives us the power to build the model of essentially any optical system, like a multielement camera lens or light-field camera as a linear or affine transform.

In a non-absorbing optical system the radiance is conserved. In other words, the radiance does not change along a ray during travel or transformation by optical elements. The mathematical representation of this fact is that any optical matrix is symplectic [7]. In this paper we will only use the property of the transforms that determinant of any optical matrix is 1,

$$det\, A = 1, \tag{6}$$

which can also be seen directly from equations (4) and (5).

Based on the above-mentioned conservation law, the radiance r' after a transform is related to the radiance r before the transform by the following equation:

$$r'(x) = r(x_0) = r(A^{-1}x), \tag{7}$$

where x_0 is the ray, which has been mapped into x by the optical transformation A, i.e., $x = Ax_0$.

The above equation can be expressed in frequency representation as follows:

$$\begin{aligned} R'(\omega) &= \int r'(x)e^{i\omega \cdot x}dx \\ &= \int r(A^{-1}x)e^{i\omega \cdot x}dx \\ &= \int r(A^{-1}x)e^{i\omega AA^{-1}x}dx \\ &= \int r(x_0)e^{i\omega A \cdot x_0}dx_0 \\ &= R(A^T\omega), \end{aligned} \tag{8}$$

where A^T is the transposed matrix, and we have used (6) for the change of variables from x to x_0. Note that this expression is derived for _any_ optical transform A, while previous works have only considered special cases.

The above results can be summarized as follows:

$$x' = Ax \tag{9}$$
$$r'(x) = r(A^{-1}x) \tag{10}$$
$$R'(\omega) = R(A^T\omega) \tag{11}$$

3 Light-field cameras in the frequency domain

3.1 The Ives' camera

The first light-field camera (called "Process of Making Parallax Stereograms") was invented back in 1903 by Frederick Ives [8]. It can be described as an array of pinhole cameras with the same focal distance f, as shown in Figure 2. This array of cameras is placed at the focal plane of a conventional large format camera.
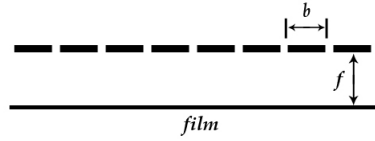

Figure 2: The Ives' light-field camera. Only the focal plane with the pinholes is represented in this figure.

We will develop the mathematical representation for the radiance transformations inside Ives' camera in frequency domain. This representation will be used throughout the paper.

Consider a 1-dimensional Ives' camera and the corresponding 2D radiance. Just before the array of pinholes the radiance is $r(x) = r(q, p)$. Just after the array of pinholes the radiance is

$$r'(q, p) = r(q, p) \sum_{m=-\infty}^{\infty} \delta(q - mb), \quad (12)$$

where b is the pitch (distance between pinholes). In frequency representation this radiance can be written based on the Poisson summation formula [17] as

$$
\begin{aligned}
R'(\omega) &= \int r(q, p) \sum_m \delta(q - mb) e^{i\omega \cdot x} dx \\
&= \frac{1}{b} \int r(q, p) \sum_n e^{in\frac{2\pi q}{b}} e^{i(\omega_q q + \omega_p p)} dq dp \\
&= \frac{1}{b} \sum_n R(\omega_q + n\frac{2\pi}{b}, \omega_p). \quad (13)
\end{aligned}
$$

Assuming a bandlimited signal, this result shows that the radiance after the pinholes consists of multiple copies of the original radiance, shifted in their frequencies by $n\frac{2\pi}{b}$ for all integers n, as shown in Figure 3a.
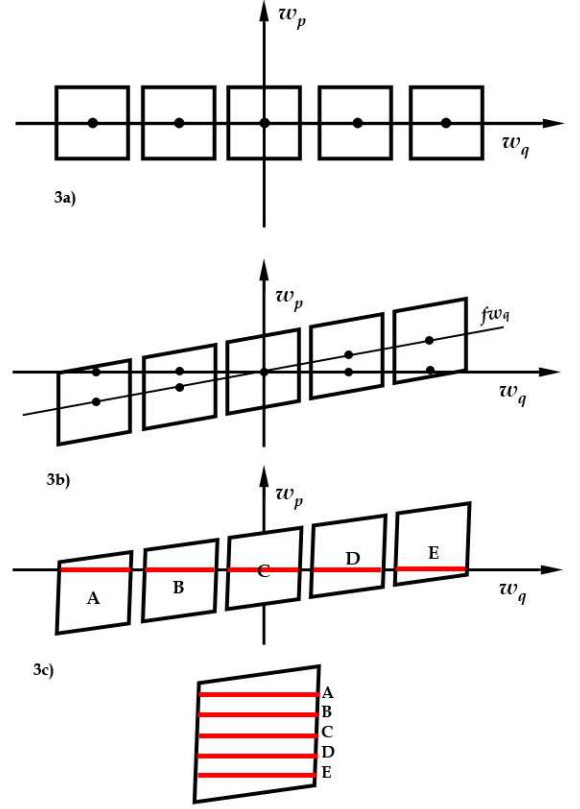

Figure 3: (a) Baldlimited signal after the array of pinholes. (b) Shear of the signal after travelling a distance f. (c) Reconstructing the original signal before the pinholes by combining samples at different intersections with the ω_q axis.

After travelling a distance f from the pinholes to the image plane, the radiance is transformed by the translation matrix (5) transposed, according to (11). The resultant radiance R_f reaching the film plane is:

$$R_f(\omega) = \sum_{n=-\infty}^{\infty} R(\omega_q + n\frac{2\pi}{b}, f\omega_q + \omega_p). \quad (14)$$

We see that the signal is sheared in the direction of angular frequency. This is represented in Figure 3b. The key observation here is that a different angular part of each copy intersects with the ω_q axis. Since the film (or sensor) responds only to the zero angular frequency, it records only the thin slice where the spectrum intersects with the ω_q axis.

By picking up slices in the image at different angular frequencies and stacking them up along the ω_q axis, we can reconstruct the original signal

$R(\omega_q, \omega_p)$, as shown in Figure 3c. Finally, an inverse Fourier transform is applied to convert the radiance into the familiar spatio-angular representation $r(x)$.

The idea that we can pick up slices at zero angular frequency, and use them to recover angular information has been proposed in the context of a different camera in [20]. We will cover this in more detail in subsection 3.3.

3.2 Replacing the pinhole array with a lens array

The idea of replacing the pinholes in the Ives' design with lenses was proposed by Lippmann back in 1908 [11]. Just as with a single pinhole camera, lenses gather much more light and produce better image quality than small holes. Lipmann called his device *Integral camera*. Different versions of it have been proposed throughout the years, the most recent one being the *plenoptic camera* [1, 14].

Our analysis of the integral camera in frequency space will be done in two steps. (1) We consider an array of pinholes as in the Ives' camera, only shifted by a constant (for all pinholes) vector a. Each pinhole is covered by a prism with angle of deviation depending on the shift, defined as $p_{prism} = \frac{a}{f}$. (2) We consider the superposition of multiple arrays of such pinhole-prisms, and show that they all contribute to the final image in the same way. Lippmann's integral camera is based on this coherent action of different arrays. It can be viewed as the limiting case where the plane is made completely of pinhole-prisms and all the light goes through. Each microlens is formed by the corresponding prisms, as a Fresnel lens.

Following the above derivation for the Ives camera in equation (13), the radiance after the pinhole-prism array can be expressed as

$$R'(\omega) = \int r(q, p + \frac{a}{f}) \sum_m \delta(q - mb - a) e^{i\omega \cdot x} dx$$
$$= \frac{1}{b} \int r(q, p + \frac{a}{f}) \sum_n e^{in\frac{2\pi(q-a)}{b}} e^{i(\omega_q q + \omega_p p)} dq dp$$
$$= \frac{1}{b} \sum_n e^{-i(\omega_p \frac{a}{f} + n\frac{2\pi a}{b})} R(\omega_q + n\frac{2\pi}{b}, \omega_p).$$
$$(15)$$

Note that now there exist additional phase multipliers in each term of the sum. After the pinhole-prism array, the light travels a distance f to the film

plane. Using equations (5) and (9) we obtain the following expression for the radiance at the film (sensor):

$$R_f(\omega) = \frac{1}{b} \sum_n e^{-i((f\omega_q + \omega_p)\frac{a}{f} + n\frac{2\pi a}{b})} R(\omega_q + n\frac{2\pi}{b}, f\omega_q + \omega_p)$$

As explained above, the film (or sensor) only records zero angular frequencies. Therefore, by restricting ω to the ω_q axis, we obtain the following final expression:

$$R_f(\omega_q, 0) = \frac{1}{b} \sum_n e^{-i(\omega_q a + n\frac{2\pi a}{b})} R(\omega_q + n\frac{2\pi}{b}, f\omega_q) \quad (16)$$

The effect of coherence would be easily observed for small a. It takes place due to the term $\omega_q a + n\frac{2\pi a}{b}$, where ω_q is within $\frac{\pi}{b}$ from the corresponding center (peak), which is at frequency $n\frac{2\pi}{b}$ in each block. For every exponential term with frequency ω_q there is another term with frequency $-n\frac{2\pi}{b} - \omega_q$ inside the same block, but on the other side of the center. Those two frequencies produce opposite phases, which results in a real positive term, $cos((\omega_q + n\frac{2\pi}{b})a)$. This term is close to 1 for all rays.

Based on this analysis, the integral camera will also work with lenses for which a can be as big as $\frac{b}{2}$, and the area of the plane is completely covered. All the terms are still positive, but the efficiency of rays far from the center is lower, and high frequencies will be attenuated.

The above analysis proves that the frequency method for multiplexing radiance, described in the case of Ives' pinhole camera, is also valid for Lippmann's microlens based integral camera. Similarly, the plenoptic camera, and other light-field cameras that can be shown equivalent to it, can be analyzed using this formulation.

3.3 Replacing the pinhole array with a mask

A light-field camera that uses a mask instead of pinholes or microlenses in front of the sensor was first proposed in [20].

One way to analyze this camera would be to start again with the pinhole formula we derived for the Ives' camera, and instead of prisms assume appropriate attenuation at each pinhole. On the other hand, it is also possible to directly derive the result for periodic attenuation functions, like

$\frac{1}{2}(1 + cos(\omega_o q))$. Then the radiance after the attenuating mask would be represented as

$$R'(\boldsymbol{\omega}) = \frac{1}{2}R(\boldsymbol{\omega}) + \frac{1}{2} \int r(\boldsymbol{x})cos(\omega_o q)\, e^{i\boldsymbol{\omega}\cdot\boldsymbol{x}}d\boldsymbol{x}$$
$$= \frac{1}{2}R(\boldsymbol{\omega}) + \frac{1}{4} \int r(\boldsymbol{x})(e^{i\omega_o q} + e^{-i\omega_o q})\, e^{i\boldsymbol{\omega}\cdot\boldsymbol{x}}d\boldsymbol{x}$$
$$= \frac{1}{2}R(\boldsymbol{\omega}) + \frac{1}{4}(R(\omega_q + \omega_o, \omega_p) + R(\omega_q - \omega_o, \omega_p)). \tag{17}$$

After the mask the signal travels a distance f to the sensor. Again using equations (5) and (11) we obtain the final expression for the radiance:

$$R_f(\omega_q, \omega_p) = \frac{1}{2}R(\omega_q, f\omega_q + \omega_p)$$
$$+ \frac{1}{4}(R(\omega_q + \omega_o, f\omega_q + \omega_p) + R(\omega_q - \omega_o, f\omega_q + \omega_p)). \tag{18}$$

Again we observe duplication of our bandlimited signal into multiple blocks, and shearing proportional to the travel distance. Any periodic mask can be analyzed this way based on Fourier series expansion and considering individual component frequencies. As first observed in [20], samples of the signal on the ω_q axis can be used to reconstruct the complete radiance $R(\boldsymbol{\omega})$.

3.4 Placing the array in front of the camera

Another type of light-field cameras can be described as putting any of the arrays used on previous camera designs in front of a regular camera, and focusing it slightly behind the array.

The idea for this design is based on the fact that the image inside any camera is 3-dimensional, and is a distorted copy of the outside world. It is clear that the structures we place inside the camera have their corresponding structures in the outside world. This is based on the mapping defined by the main camera lens.

The sensor plane corresponds to the plane of focus, and any optical elements in front of it could be replaced by their enlarged copies in the real world, in front of the plane of focus.

Because of this correspondence, and based on the lens formula, we can build structures in front of the camera and use them as if they were microstructures inside. In the Results section we will demonstrate how a fine mask in front of the sensor, in an area not accessible due to the cover glass, can be replaced by a mosquito mesh in front of the camera.

3.5 Matching the F/numbers

The rule of matching the F/numbers of the main camera lens and the microlenses was first explicitly stated in [14]. If densely packed microlenses in an integral camera had smaller F/number than the main lens, then parts of the images of the main lens would extend beyond the area covered by the corresponding microlens, and would interfere with the image in the neighboring microlens. If the F/number of the microlenses was bigger, then part of the area under each microlens would remain unused. Clearly, we want F/numbers to match.

A light-field photographer is not free to change the aperture without considering the current aperture of the microlenses. This raises several important questions: Is this restriction relaxed in any way for other light-field cameras that are not based on microlenses? Is there a quantity equivalent to F/number in cases other than microlenses, and can we gain additional knowledge in this important area based on frequency analysis?

The final expression for the radiance in all cameras has second (angular frequency) argument in R equal to $f\omega_q$, where f is the distance from the pinholes, microlenses or mask - to the sensor. This is a measure for the amount of shear, which can be seen as the tilt of the line $f\omega_q$ in Figure 3b.

Assume a camera is sampling the angular frequency N times, i.e., copies of the signal that intersect with the ω_q axis N times. For example, this could be a mask containing N frequencies at interval ω_o, or N peaks, including the zero frequency peak. The frequency spectrum of this signal covers an interval of $N\omega_o$ in the horizontal axis.

Because of the tilt, those peaks are spread in the vertical ω_p axis in an interval of $f N\omega_o$. Therefore, the following expression holds:

$$2\omega_{po} = f N\omega_o, \tag{19}$$

where ω_{po} is the maximal angular frequency of the original signal.

The width of the cone of rays (maximal angle of rays) coming to a point on the film plane in a camera is $\frac{1}{F}$, where F is the F/number of the main lens. If the maximal resolution (number of lines) in a light-field camera in angular direction is N, then the maximal angular frequency would be $\omega_{po} = 2\pi NF$. By substituting in equation (19), we obtain

a) FFT of a typical lens-based image, see figure 7.

b) FFT of a mask-based image with vertical gratings.
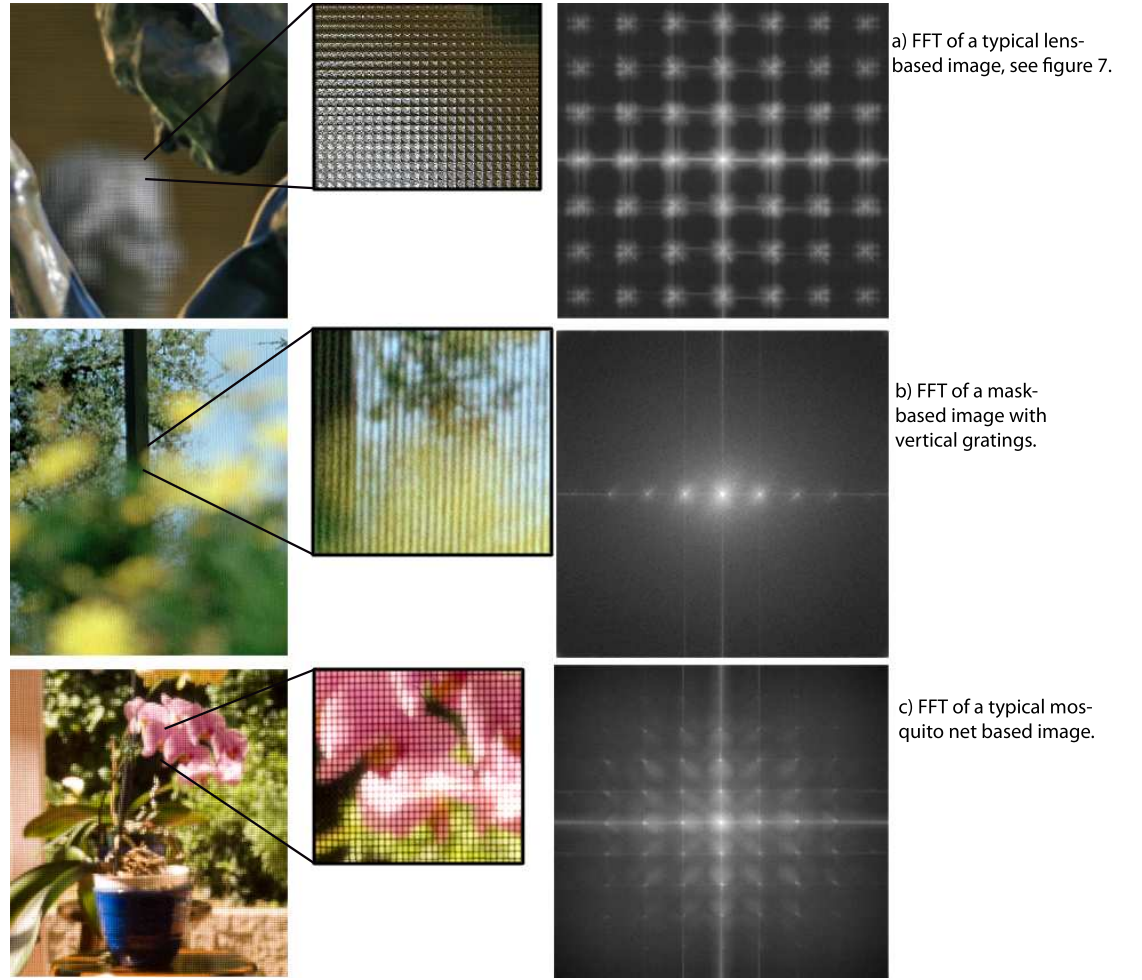
c) FFT of a typical mosquito net based image.

Figure 4: *Left:* Typical pictures obtained from lens-based (top), mask-based (middle), mosquito-net-based (bottom) cameras. *Right:* Magnitude of the Fourier Transforms of the images (adjusted for display).

$$f\omega_o = 4\pi F. \qquad (20)$$

Since the wavelength is b, so that $\omega_o = \frac{2\pi}{b}$, we obtain

$$f\frac{2\pi}{b} = 4\pi F. \qquad (21)$$

The maximal spatial frequency in the initial bandlimited spectrum is $\frac{\omega_o}{2}$, and the signal has wavelength $2b$. In this way we obtain the final result:

$$\frac{f}{b} = F. \qquad (22)$$

All cameras multiplexing in the frequency domain have to satisfy the F/number matching condition (22), where b is the pitch of the pinholes or microlenses, or the period of the lowest frequency in the mask, and f is the distance from the mask or array of pinholes/microlenses to the sensor.

One of the points in the Results section is that we refer the reader to a series of movies (available in the electronic version of the paper), showing how mask-based cameras work or fail to work for different F/numbers.

4 Results

We have applied the method of frequency domain analysis to the images captured by the camera designs mentioned in Section 3. In this section we first explain the process of demultiplexing in the frequency domain to render the images from different camera designs, and then we proceed to show examples from each of the cameras.
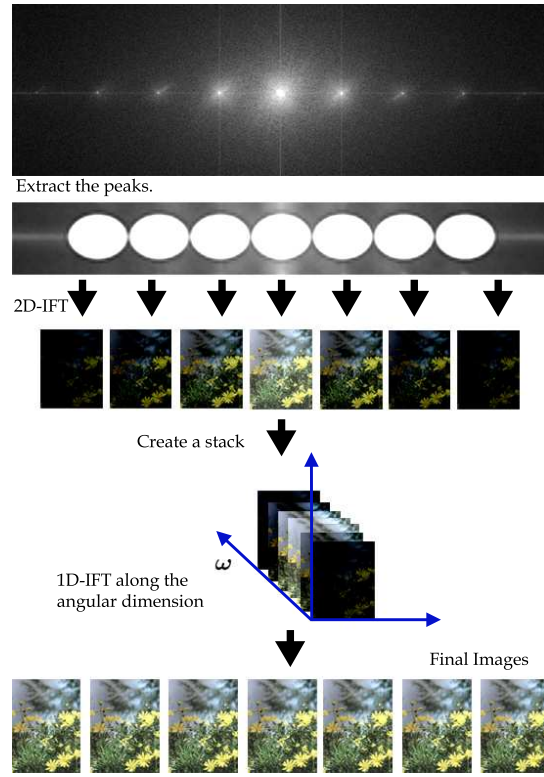
Extract the peaks.

2D-IFT

Create a stack

1D-IFT along the angular dimension

Final Images

Figure 5: Flow-chart of the method of demultiplexing of the angular information.

4.1 Method of demultiplexing

In all the aforementioned camera designs, the 4-dimensional radiance is multiplexed onto the 2-dimensional camera sensor (or the film). This process of radiance multiplexing is given by (14), (16) and (18) for the respective camera designs.

Figure 4 shows typical images obtained from the three camera designs. The magnitudes of the 2D Fourier transforms of these images are shown on the right side of Figure 4, where the shifted tiles of the transform are visible in each of the images. Notice that these tiles are placed in equal distances both horizontally and vertically in the case of our lens-based and mosquito-net based camera images, and only horizontally in the case of our mask-based camera. This is due to our choice of the mask consisting of only vertical lines. In all three cases we will only show examples of extracting horizontal parallax, but extending the method to obtain parallax in both directions is straightforward.

Figure 5 shows an example application of the demultiplexing method in the frequency domain. We first note that the entire 4D light field is encoded

Figure 6: Correcting the effect of waves due to small shifts or misalignments in the FFT.

in these images. There are several ways of extracting individual parallax views. For instance, in the case of lens based cameras, pixels belonging to each camera can be extracted from the captured image, rearranged and put into individual images, so that a 2D array of 2D images is obtained.

A frequency domain method would be based on the separability of the Fourier transform. Following Figure 3, the tiles of the 2D Fourier transform are first extracted, as shown in the second row of Figure 5. First, a 2D inverse Fourier transform individually applied to each the tiles to obtain 7 images. These images are stacked together to form a 3D image. Final horizontal parallax images can be obtained by applying a 1D inverse Fourier transform (IFFT) along the third dimension of this 3D image, and unstacking the result. This process is effectively performing a 3D IFFT. In the general case of horizontal and vertical parallax we need to extend it to 4D IFFT.

Good artifact-free results are very sensitive to determining the location of the centers of the tiles in the Fourier transforms. The Fourier transforms of the images are obtained by Fast Fourier Transform, which makes the location of the centers ambiguous, due to the discretization. There is a misplacement error within one pixel around each center, which causes low-frequency waves in the final parallax images. This problem can be fixed by multiplying the images before the last 1D IFFT by a linear phase that corresponds to the subpixel shift in the FFT. Figure 6 shows an image from our mask-based camera before and after the phase correction is applied to eliminate the low-frequency waves.

4.2 Lens based cameras

Our lens based camera is a simulation of the plenoptic/integral camera. We took 49 pictures of
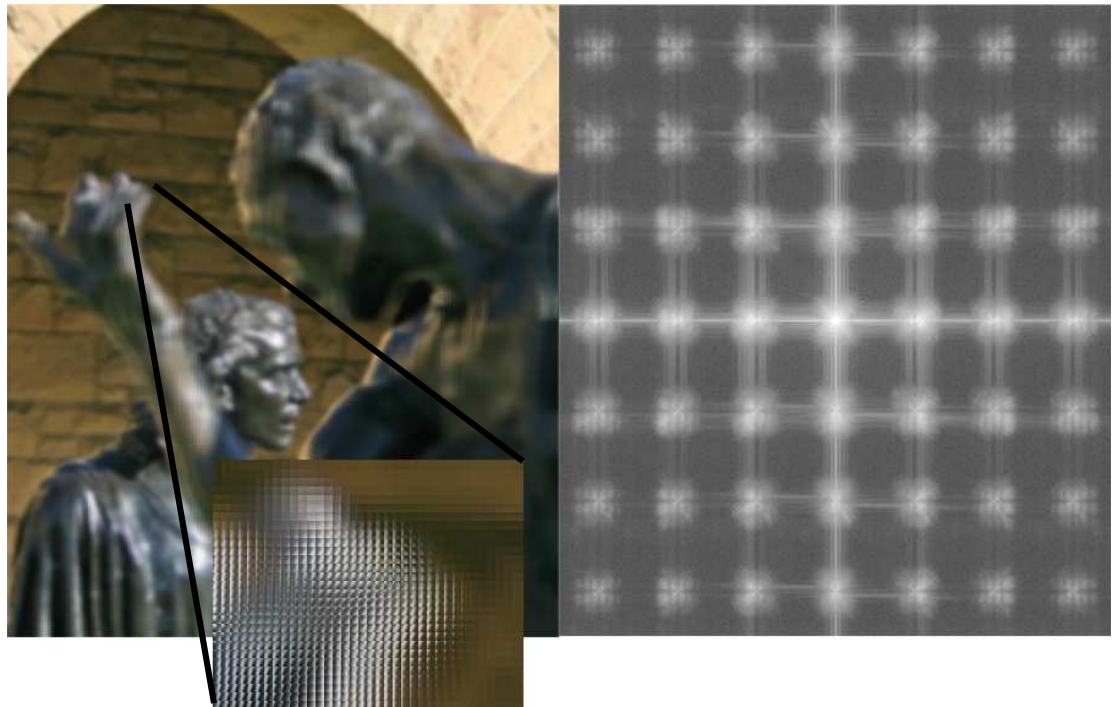
Figure 7: *Left:* The light-field image of Rodin sculptures. *Right:* Absolute value of the Fourier transform of the image.

the scene from equally spaced locations. The centers of projection are arranged on a plane as a 7×7 grid, with the cameras pointed perpendicular to the plane. The final image is made up of 7×7 blocks each of which consists of 49 pixels taken from the same location in all 49 images. An example image is shown in Figure 7. Notice the zoomed area which clearly displays the formed blocks.

The FFT of this image is shown on the right in Figure 7. To obtain horizontal parallax, we apply the demultiplexing described in the previous section, with 7 views. Two images resulting from this process are shown in Figure 8. Small parallax is visible in this stereo pair at close examination. Left and right images are switched so that stereo fusion can be achieved with crossed eyes observation.
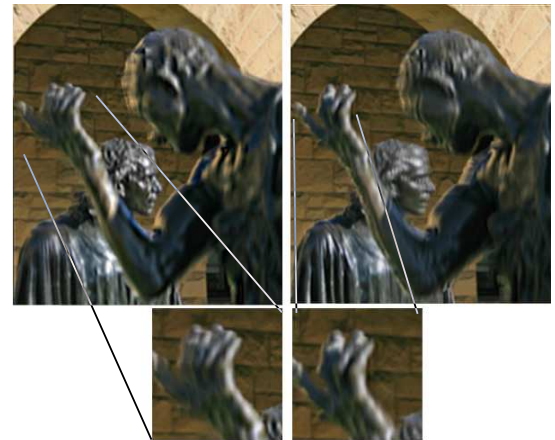
4.3 Mask-based cameras

We have built several working prototypes mask-based cameras. In order to achieve good resolution we need small value of the largest period b, on the order of 0.1 mm. With F/number of the main lens equal to 4 we need to place the mask about 0.4 mm from the surface of the sensor, which is impossible due to the cover glass. This situation forced

Figure 8: Two stereo views from the frequency domain reconstructed light field of Rodin. Left and right images are switched.

us to work with film based medium format camera. The reason for medium format is the bigger image which gives potential for higher resolution, and easier access to the film back, where we make our modifications.

We are using Contax 645 with a film back. We

Figure 9: Our camera with the 3M filter in the film back.

Figure 10: Two stereo views of flowers. Selected areas show detail in which it is easy to spot parallax. Left and right images are switched.

have experimented with two different masks. First, we take a picture of a poster displaying computer generated grid, and then use the negative as a mask in front of the film. The computer generated grid is is 2D cosine mask with 3 harmonics in both spatial dimensions. The spacing of 0.5 mm is achieved by placing the developed negative between two thin glasses. The film that is being exposed slides directly on the surface of the glass.

As the second mask we used a 3M computer screen filter. This filter contains about 14 black lines/mm, and those lines are sandwiched between transparent plastic material 0.2mm thick. As a result the F/number of the mask is about 3. This way we can capture a very high resolution light-field of 14 samples/mm, where each sample contains complete angular information. In this section we only show results obtained with this second mask, since the results from the first mask are similar.

Figure 9 shows a picture of the Contax camera and the film back with the 3M filter glued to the window just in front of the film.

A sequence of parallax movies accompanying the electronic version of this paper, which are generated from pictures at different apertures, show that the best F/number is 5.6. This value is slightly higher than the expected 3 or 4. Possible reasons are the refractive index of the plastic material, which increases optical path, and possible micro-spacing between the film and the 3M filter due to mechanical imperfection/dust.

Figure 10 shows two stereo views from the light-field generated from the image taken with the mask at F/5.6. The reader is encouraged to see the elec-
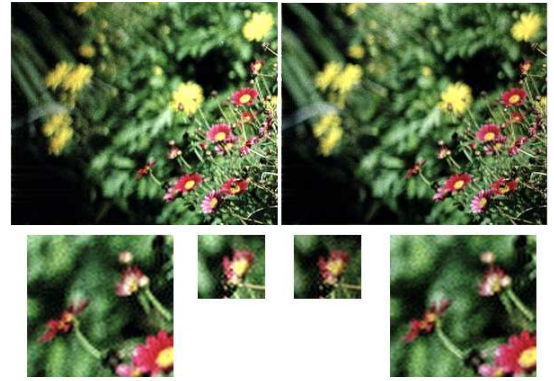
tronic version of this paper for the original high resolution images. The supplementary videos present sequences of synthetic view sweeps through a range of angles that show clear parallax and create the feeling of depth in the images.

4.4 Mosquito-net camera

In order to demonstrate our method of frequency domain multiplexing based on external mesh or mask, we take pictures through a mosquito net.

We used our Contax camera with 80mm lens and a digital back without any modifications. The net is placed approximately 2m from the camera, and the camera is focused on a plane about 10cm behind the net. With this setting we overcome the cover glass problem.

By differentiating the lens equation

$$\frac{1}{a} + \frac{1}{b} = \frac{1}{f}$$

we obtain

$$\frac{da}{a^2} = -\frac{db}{b^2},$$

so moving the focus by $da = 10cm$ away from the net produces a movement of $-da\frac{b^2}{a^2} = 0.25mm$ away from the sensor surface. At the same time the image of our 2mm grid of the net has been reduced linearly to 0.08mm, which gives us F/number of about 3, and high resolution.

Figure 11 shows an image taken through the net at F/number 4, and Figure 12 shows two stereo views of the scene reconstructed with our method of de-multiplexing in the frequency domain. The effect of
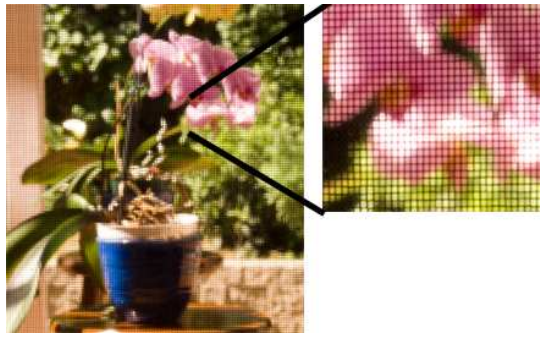
Figure 11: Picture taken through a mosquito net.

Figure 12: Two stereo views from the light-field generated from the picture taken through a mosquito net. Left and right images are switched.

changing the aperture on the appearance of the net can be seen the movie accompanying this paper.

5   Conclusions and future work

In this paper, we have derived a new mathematical formalism of analyzing light-field cameras in the frequency domain. The proposed method of multiplexing the $4D$ radiance onto the $2D$ sensor is shown to work in the frequency domain for a number of of light-field cameras, both lens-based and mask-based.

Analyzing light-field cameras in the frequency domain and designing new approaches based on that is a new emerging area of research. Mask based cameras are obviously cheaper and easier to build. Combining microlenses, prisms, and other active optical elements might have unexpected potential. There is much more to be done in that direction, including applying the methods of wave optics.

We have shown that the F/number matching condition is a requirement for all light-field cameras. This helps to construct and adjust cameras so they produce working data. We also propose new designs based on masks/nets outside the camera. Our analysis of the wave lead to an algorithm of con-

sistently producing good results based on the proposed frequency multiplexing approach.

We have built several light field cameras in order to demonstrate the main points of this paper.

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. A whole new dimension is added to to integral photography due to unexpected possibilities with frequency domain multiplexing 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

[1] T. Adelson and J. Wang. Single lens stereo with a plenoptic camera. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, pages 99–106, 1992.

[2] J. Chai, S. Chan, H. Shum, and X. Tong. Plenoptic sampling. *ACM Trans. Graph.*, pages 307–318, 2000.

[3] F. Durand, N. Holzschuch, C. Soler, E. Chan, and F. Sillion. A frequency analysis of light transport. *ACM Trans. Graph.*, pages 1115–1126, 2005.

[4] Todor Georgiev and Chintan Intwala. Light field camera design for integral view photography. *Adobe Tech. Rep.*, 2006.

[5] A. Gerrard and J. M. Burch. Introduction to matrix methods in optics. 1994.

[6] Steven J. Gortler, Radek Grzeszczuk, Richard Szeliski, and Michael F. Cohen. The lumigraph. *ACM Trans. Graph.*, pages 43–54, 1996.

[7] V. Guillemin and S. Sternberg. Symplectic techniques in physics. 1985.

[8] F. Ives. Patent US 725,567. 1903.

[9] H. Ives. Camera for making parallax panoramagrams. *J. Opt. Soc. Amer., 17*, pages 435–439, 1928.

[10] Marc Levoy and Pat Hanrahan. Light field rendering. *ACM Trans. Graph.*, pages 31–42, 1996.

[11] Gabriel Lippmann. Epreuves reversible donnant la sensation du relief. *J. Phys. 7*, pages 821–825, 1908.

[12] T. Naemura, T. Yoshida, and H. Harashima. 3d computer graphics based on integral photography. *Optics Express, Vol. 8, 2*, 2001.

[13] R. Ng. Fourier slice photography. *ACM Trans. Graph.*, pages 735–744, 2005.

[14] Ren Ng, Marc Levoy, Mathieu Brdif, Gene Duval, Mark Horowitz, and Pat Hanrahan. Light field photography with a hand-held plenoptic camera. *Stanford Tech. Rep.*, 2005.

[15] F. Okano, J. Arai, H. Hoshino, and I. Yuyama. Three-dimensional video system based on integral photography. *Optical Engineering, Vol. 38, 6*, 1999.

[16] F. Okano, H. Hoshino, J. Arai, and I Yuyama. Real-time pickup method for a three-dimensional image based on integral photography. *Applied Optics, Vol. 36, 7*, pages 1598–1603, 1997.

[17] Alan V. Oppenheim and Alan S. Willsky. *Signals and Systems*. Prentice Hall, Upper Saddle River, New Jersey, 1997.

[18] R. Stevens and T. Harvey. Lens arrays for a three-dimensional imaging system. *Journal of Optics A, Vol. 4*, 2002.

[19] R. Tyson. Principles of adaptive optics. In *Academic Press*, 1991.

[20] A. Veeraraghavan, A. Mohan, A. Agrawal, R. Raskar, and J. Tumblin. Coded aperture and optical heterodyning : A mask-based approach for digital refocusing and light field acquisition by conventional cameras. *ACM Trans. Graph.*, 2007.

[21] B. Wilburn, N. Joshi, V. Vaish, E. Talvala, E. Antunez, A. Barth, A. Adams, M. Levoy, and M. Horowitz. High performance imaging using large camera arrays. In *ACM Trans. Graph.*, 2005.