Unified Frequency Domain Analysis of Lightfield Cameras

Todor Georgiev, Chintan Intwala, Sevkit Babakan, and Andrew Lumsdaine

[1] Adobe Systems, 345 Park Ave, San Jose, CA 95110
[2] Computer Science Department, Indiana University, Bloomington, IN 47405

Abstract. This paper presents a theory that encompasses both "plenoptic" (microlens based) and "heterodyning" (mask based) cameras in a single frequency-domain mathematical formalism. Light-field capture has traditionally been analyzed using spatio-angular representation, with the exception of the frequency-domain "heterodyning" work. In this paper we interpret "heterodyning" as a general theory of multiplexing the radiance in the frequency domain. Using this interpretation, we derive a mathematical theory of recovering the 4D spatial and angular information from the multiplexed 2D frequency representation. The resulting method is applicable to all lightfield cameras, lens-based and mask-based. The generality of our approach suggests new designs for lightfield cameras. We present one such novel lightfield camera, based on a mask outside a conventional camera. Experimental results are presented for all cameras described.

1 Introduction

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

The *light itself*, or *light-field*, is mathematically described by the *radiance density function*, which is a complete record of light energy flowing along "all rays" in $3D$ space. This density is a field defined in the $4D$ domain of optical phase space, that is, the space of all lines in $3D$ with symplectic structure [1].

Conventional cameras, based on $2D$ image sensors, do not record the $4D$ radiance. Instead, they simply act as "integration devices". In a typical setting, they integrate over the $2D$ aperture to produce a $2D$ projection of the full $4D$ radiance. Integral Photography [2, 3] was proposed more than a century ago to "undo" this integration and measure the complete $4D$ radiance arriving at all points on a film plane or sensor.

As demonstrated by Levoy and Hanrahan [4] and Gortler et al. [5], 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. [6] 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.

A number of works have developed techniques for analyzing radiance in the frequency domain [7–9]. Several important results have been derived, among which are: Applications of Poisson summation formula to lightfields, lightfield displays, and depth representation of scenes; light transport and optical transforms; Fourier slice theorem applied to refocusing, and others. However, with the notable exception of "heterodyning" [10], frequency analysis of radiance has not been applied to the general understanding and design of light-field cameras. The main goal of this paper is to expand the results of "heterodyning."

The contributions of our paper are:

1. We provide a general mathematical framework for analysing lightfield cameras in the frequency domain. Our theory is applicable to both lens-based and mask-based light-field cameras.
2. We show that the concept of "heterodyning" is not just a new mask-based camera design. Rather, it is a *new method* for processing data from any lightfield camera in the freqyency domain.
3. We develop a generalized F/number matching condition and show that it is required for all lightfield camera designs, lens-based and mask-based.
4. We propose and demonstrate a new lightfield camera design that has a transparency mask placed in front of the main lens.

2   Prior Work and Basic Equations

2.1   Frequency Domain Representation

Recent works [7–9] have analyzed radiance, or light field, in frequency representation. Let $r(\boldsymbol{x})$ be the radiance in conventional $\boldsymbol{x}-$space, describing position and angle. In more detail, the spatio-angular coordinates of a ray at a given plane orthogonal to the optical axis are represented as a vector

$$\boldsymbol{x} = \begin{pmatrix} \boldsymbol{q} \\ \boldsymbol{p} \end{pmatrix},$$ (1)

where $\boldsymbol{q}$ is the location of ray-plane intersection, and $\boldsymbol{p}$ is a vector defining the two angles of that ray at location $\boldsymbol{q}$. Following texts on optics [1, 11], we use paraxial approximation assuming the angle is small. A simplified 2-dimensional vector representation of a ray is shown in Figure 1. The physical ray-space is $4D$.

In the frequency domain radiance is represented by the Fourier transform of $r(\boldsymbol{x})$:

$$R(\boldsymbol{\omega}) = \int r(\boldsymbol{x}) e^{i\boldsymbol{\omega}\cdot\boldsymbol{x}} d\boldsymbol{x}$$ (2)
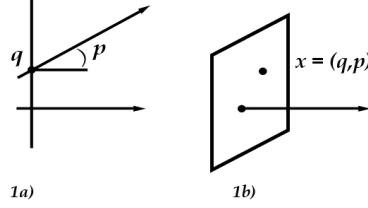
Fig. 1. (a) Geometric representation of a ray as position and angle in an optical system. (b) Same ray described as a vector $\boldsymbol{x} = (q, p)$, in 2D ray space.

where the spatial frequency $\boldsymbol{\omega}_q$ and the angular frequency $\boldsymbol{\omega}_p$ are grouped together in a similar way as a $4D$ vector:

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

To simplify our text and figures we will often use $2D$ radiance with 1-dimensional position q and angle p for each ray.

2.2 Optical Transforms

We summarize and extend some results of previous work (e.g., [7, 8, 12]) about transformations of radiance in optical systems.

The operation of a lens L or a translation T can be described by a linear transformation $\boldsymbol{x}' = A\boldsymbol{x}$ operating on a ray $\boldsymbol{x}$ [11]. If a ray is described as a position-angle vector (3), then L and T are given by the following matrices:

$$L = \begin{pmatrix} 1 & 0 \\ -\frac{1}{f} & 1 \end{pmatrix} \qquad \text{and} \qquad T = \begin{pmatrix} 1 & t \\ 0 & 1 \end{pmatrix}. \tag{4}$$

Here, f is the focal length of the lens, and t is the translation (distance of flight).

The combined action of several such elements is described by the composition of their transforms. Now, since all elements in an optical system are described only by the above transforms, we can mathematically model any optical system (like a multielement camera lens or a light-field camera) as a single combined transform.

Another important observation is that radiance is conserved in a non-absorbing system. 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 [1]. Therefore, the determinant of any optical matrix is unity, i.e.,

$$det\,A = 1, \tag{5}$$

which can also be seen directly from equation (4).

Based on this conservation property, the radiance r' after a transform is related to the radiance r before the transform by the following equation:

$$r'(\boldsymbol{x}) = r(\boldsymbol{x}_0) = r(A^{-1}\boldsymbol{x}), \qquad (6)$$

where $\boldsymbol{x}_0$ is the ray which has been mapped into $\boldsymbol{x}$ by the optical transformation A, i.e., $\boldsymbol{x} = A\boldsymbol{x}_0$.

The above transformation from r to r' can be expressed in frequency representation as:

$$
\begin{aligned}
R'(\boldsymbol{\omega}) &= \int r'(\boldsymbol{x})e^{i\boldsymbol{\omega}\cdot\boldsymbol{x}}d\boldsymbol{x} \\
&= \int r(A^{-1}\boldsymbol{x})e^{i\boldsymbol{\omega}\cdot\boldsymbol{x}}d\boldsymbol{x} \\
&= \int r(A^{-1}\boldsymbol{x})e^{i\boldsymbol{\omega}AA^{-1}\boldsymbol{x}}d\boldsymbol{x} \\
&= \int r(\boldsymbol{x}_0)e^{i\boldsymbol{\omega}A\cdot\boldsymbol{x}_0}d\boldsymbol{x}_0 \\
&= R(A^T\boldsymbol{\omega}), \qquad (7)
\end{aligned}
$$

where A^T is the transposed matrix, $R(\boldsymbol{\omega})$ is the Fourier transform of $r(\boldsymbol{x})$, and we have used (6) for the change of variables from $\boldsymbol{x}$ to $\boldsymbol{x}_0$. Note that this expression is derived for *any* optical transform A, while previous works have considered special cases [8].

Summarizing the basic equations that will be used throughout the paper, we have:

$$\boldsymbol{x}' = A\boldsymbol{x} \qquad (8)$$
$$r'(\boldsymbol{x}) = r(A^{-1}\boldsymbol{x}) \qquad (9)$$
$$R'(\boldsymbol{\omega}) = R(A^T\boldsymbol{\omega}) \qquad (10)$$

3   Light-field Cameras in the Frequency Domain

3.1   Ives' Camera

The first light-field camera (called "Process of Making Parallax Stereograms") was invented in 1903 by Frederick Ives [2]. This camera 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.

We will develop the mathematical representation for the radiance transforms inside Ives' camera in the frequency domain. This representation is a new result, and it will be used extensively throughout the paper.

Consider a 1-dimensional Ives' camera and the corresponding 2D radiance. Modeling the array of pinholes as a delta-train, the radiance, just before this
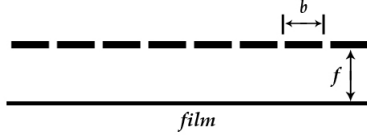
Fig. 2. The Ives' light-field camera. Only the focal plane with the pinholes is represented in this figure.

array is $r(\boldsymbol{x}) = r(q, p)$. Just after the array of pinholes the radiance is

$$r'(q, p) = r(q, p) \sum_{m=-\infty}^{\infty} \delta(q - mb), \tag{11}$$

where b is the pitch (distance between pinholes). In frequency domain the modulated train of delta-functions (11) can be represented using the Poisson summation formula [13] as

$$
\begin{aligned}
R'(\boldsymbol{\omega}) &= \int r(q, p) \sum_m \delta(q - mb) e^{i\boldsymbol{\omega}\cdot\boldsymbol{x}} d\boldsymbol{x} \\
&= \frac{1}{b} \int r(q, p) \sum_n e^{in\frac{2\pi q}{b}} e^{i(\omega_q q + \omega_p p)} dq dp \\
&= \frac{1}{b} \sum_n R(\omega_q + n\frac{2\pi}{b},\, \omega_p).
\end{aligned}
\tag{12}
$$

Assuming a band-limited 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. Note that the representation in Figure 3 follows the heterodyning method proposed in [10].

Due to traveling a distance f from the pinholes to the image plane, the radiance is transformed by the transpose of the translation matrix T, according to (10). The resultant radiance R_f reaching the film plane is:

$$R_f(\boldsymbol{\omega}) = \sum_{n=-\infty}^{\infty} R(\omega_q + n\frac{2\pi}{b},\, f\omega_q + \omega_p). \tag{13}$$

This equation shows that the signal is sheared in the direction of angular frequency. This is represented in Figure 3b. The key observation here, first made in [10], is that a different angular frequency part of each copy of the spectrum intersects with the ω_q axis. Since the film responds only to the DC component
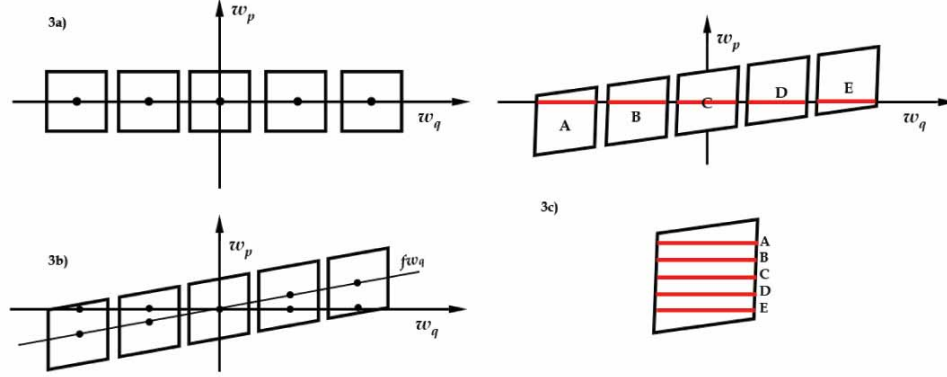
Fig. 3. (a) Bandlimited signal after the array of pinholes. (b) Shear of the signal after travelling a distance f. (c) Reconstructing the original signal before the pinholes by combining samples at different intersections with the ω_q axis.

(zero angular frequency), it records only the thin "slice" where the spectrum intersects with the ω_q axis.

The above observation suggest a method of reconstructing the complete spatio-angular representation of the radiance $r(\boldsymbol{x})$. This can be done by stacking the above slices along the ω_p axis and performing inverse Fourier transform of the volume image.

The original idea that we can pick up the above type of "slices" of a similar spatio-angular representation of a periodic mask in the frequency domain, and use them to recover angular information, has been first proposed in the context of a different camera in [10]. We will cover this in more detail in subsection 3.3.

3.2   Replacing the Pinhole Array with a (Micro) Lens Array

The idea of replacing the array of pinholes in front of the film with lenses was first proposed by Lippmann back in 1908 [3]. Just as with a single pinhole camera, lenses gather much more light and produce better image quality than small holes. Lipmann called his approach *Integral photography*. Different versions of it have been proposed throughout the years, the most recent one being the *plenoptic camera* [14, 6].

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 $\boldsymbol{a}$. Each pinhole is covered by a prism with angle of deviation depending on the shift, defined as $\boldsymbol{p}_{prism} = \frac{\boldsymbol{a}}{f}$. (2) We consider the superposition of multiple shifted arrays of such pinhole-prisms, and show that they all contribute to the final image in a similar way. Lippmann's integral photography 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 (12), the radiance after the pinhole-prism array can be expressed as

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

$$(14)$$

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 (4) and (10) we obtain the following expression for the radiance at the film (sensor):

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

As explained above, the film (or sensor) only records zero angular frequencies. Therefore, by restricting $\boldsymbol{\omega}$ to the $\boldsymbol{\omega}_q$ axis, we obtain the following final expression:

$$
R_f(\omega_q, 0) = \frac{1}{b} \sum_n e^{-i(\omega_q a + n\frac{2\pi a}{b})} R(\omega_q + n\frac{2\pi}{b},\, f\omega_q) \tag{15}
$$

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 leads to a surprising new result: The frequency method of demultiplexing radiance, described in the case of Ives' pinhole camera, is also applicable to Lippmann's microlens based integral photography. Similarly, the plenoptic camera, as well as other light-field cameras that can be shown to be equivalent to it, can be analyzed using this new 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 [10].

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 $\frac{1}{2}(1 + cos\,(\omega_0 q))$. This attenuating mask modulates the light field to produce two spectral copies seen mathematically as follows

$$R'(\boldsymbol{\omega}) = \frac{1}{2}R(\boldsymbol{\omega}) + \frac{1}{2}\int r(\boldsymbol{x})cos\,(\omega_0 q)\, e^{i\boldsymbol{\omega}\cdot\boldsymbol{x}}d\boldsymbol{x}$$

$$= \frac{1}{2}R(\boldsymbol{\omega}) + \frac{1}{4}\int r(\boldsymbol{x})(e^{i\omega_0 q} + e^{-i\omega_0 q})\, e^{i\boldsymbol{\omega}\cdot\boldsymbol{x}}d\boldsymbol{x}$$

$$= \frac{1}{2}R(\boldsymbol{\omega}) + \frac{1}{4}(R(\omega_q + \omega_0,\, \omega_p) + R(\omega_q - \omega_0,\, \omega_p)). \quad (16)$$

After the mask the signal travels a distance f to the sensor. Again using equations (4) and (10) we obtain the final expression for the radiance:

$$R_f(\omega_q,\omega_p) = \frac{1}{2}R(\omega_q,\, f\omega_q + \omega_p)$$

$$+ \frac{1}{4}(R(\omega_q + \omega_0,\, f\omega_q + \omega_p) + R(\omega_q - \omega_0,\, f\omega_q + \omega_p)). \quad (17)$$

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 [10], samples of the "mask signal" on the ω_q axis can be used to reconstruct the complete spatio-angular attenuation function of the mask. In our case we use the method to reconstruct the radiance $R(\boldsymbol{\omega})$.

3.4   Placing the Array in Front of the Camera

Another family 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 1-to-1 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.
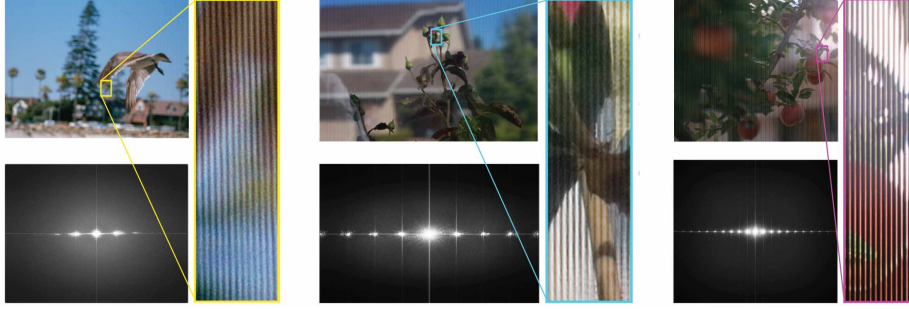
Fig. 4. Typical pictures as obtained from our camera designs and their corresponding Fourier transforms (magnitude shown). From left to right are images from,(a) mask-based, (b) mosquito-net based, and, (c) lens-based cameras.

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. We will return to this idea in section 4.3.

3.5 Generalized F/Number Matching

The angle of the cone of rays coming to a point on the sensor in a camera is $\Omega = \frac{1}{F}$, where F is the F/number of the main lens. As shown in [6], for a plenoptic camera the F/numbers of the main lens and the microlenses must match. A generalization of the F/Number matching rule to other radiance camera systems can be derived using frequency domain 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 of the amount of shear, which can be seen as the tilt of the line $f\omega_q$ in Figure 3b.

Assume a camera samples the angular frequency N times, i.e. sheared copies of the signal intersect the ω_q axis N times. For example, this could be a mask containing N frequencies at interval ω_0. The frequency spectrum of this signal covers an interval of $N\omega_0$ in the horizontal axis.

Because of the tilt, those peaks are spread in the vertical ω_p direction to cover an interval of $fN\omega_0$. Therefore, the following expression holds:

$$2\omega_{p0} = fN\omega_0, \tag{18}$$

where ω_{p0} is the maximal angular frequency of the original signal.

If the maximal number of samples in a light-field camera in angular direction is N, then the maximal angular frequency would be $\omega_{p0} = 2\pi\frac{N}{\Omega} = 2\pi NF$. By

substituting in (18), we obtain

$$4\pi F = f\omega_0$$

Denote by $\omega_b = \frac{2\pi}{b}$ the base frequency of the modulating structure with period b. We have seen that our bandlimited signal captured with this structure needs to have maximum spatial frequency $\frac{\omega_0}{2} = \omega_b$. Then $\omega_0 = \frac{4\pi}{b}$. If we substiture in the above equation,

$$4\pi F = f\frac{4\pi}{b}.$$

In this way we obtain the final result:

$$F = \frac{f}{b}. \tag{19}$$

Here, b is the pitch of the pinholes/microlenses or the period of the lowest frequency in the mask; f is the distance from the array of pinholes/microlenses or the mask to the sensor. *All* cameras multiplexing in the frequency domain must satisfy this condition.

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 demultiplexing to images captured with the different types of lightfield cameras described in Section 3. In this section we explain our process of demultiplexing in the frequency domain, similar to the algorithm in [10], to render images produced with different camera designs. Then we proceed to discuss examples rendered from each of our cameras.

4.1   Demultiplexing Method

In all the aforementioned camera designs, the 4-dimensional radiance is multiplexed onto the 2-dimensional camera sensor (or film). This process of radiance multiplexing is given by (13), (15) and (17) for the respective camera designs. Figure 4 shows typical images obtained from the three camera designs, and the magnitudes of the 2D Fourier transforms of these images. Notice that the spectral tiles are placed at equal distances horizontally. This is due to our choice of masks consisting of only vertical lines, or lenticular lenses. 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. The entire 4D radiance is encoded into these type of 2D images in spatial-domain or equally spaced angular slices in frequency-domain. So in the case of lens based cameras,

pixels belonging to each angular view can be extracted spatially from the images under each microlens. Pixels belonging to the same angular view can be rearranged and put together into individual images, representing novel views or refocusing [6]. The frequency counterpart of this process is accomplished by creating a stack of spectral slices and performing an appropriate inverse Fourier-transform. In our case of extracting horizontal parallax, we perform an inverse 3D-Fourier-transform. This process is essentially equivalent to the "heterodyning" method in [10].

4.2   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 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 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 a 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.

Another mask we used was a 3M computer screen filter. Measurements under magnification showed that the filter contains about 14 black lines/mm, with the lines being sandwiched between transparent plastic material $0.2mm$ thick. Accordingly, the F/number of the mask is about 3. In this section we only show results obtained with this second mask, since the results from the first mask are similar.

Figure 5 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 7 shows two stereo views from the light-field generated from an image taken with the mask at F/5.6. The reader is encouraged to see the electronic version of this paper for the original high resolution images. The supplementary videos present sequences of synthetic view sweeps through a range of angles, or refocusing generated with the method of [6]. These videos create a clear sense of depth in the images.
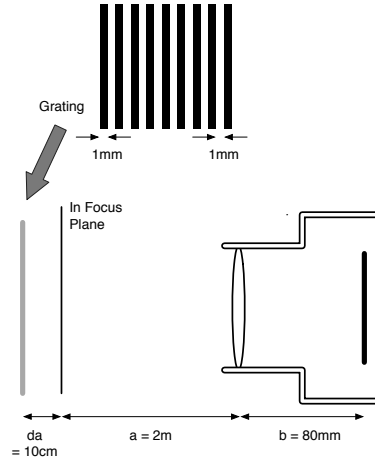
Fig. 5. Our camera with the 3M filter in the film back.

Fig. 6. Our "mosquito net" setup.

Fig. 7. Stereo images from our mask-based camera. Stereopsis can be achieved with crossed eye observation. Close examination of the marked areas will also reveal horizontal parallax between the pictures.

4.3   "Mosquito-Net" Camera

The research reported in this paper was initially motivated by an interest in explaining the unusual behavior exhibited by images taken through a mosquito net. We demonstrate the operation of frequency domain demultiplexing with an external mask (a "mesh" or a "mosquito net"). Since real mosquito nets are irregular, we constructed a regular mask by printing out a pattern of 250 vertical black lines on a transparency and mounting this transparency on glass. The printed lines were $1mm$ in width, with a spacing of $1mm$, i.e,. the period is $T = 2mm$.

Pictures were taken with a 16 megapixel digital camera having an $80mm$ lens. The transparency mask was placed approximately $2m$ from the camera, and the camera is focused on a plane about $10cm$ behind the transparency (see Figure 6)..

Fig. 8. Refocused images from our mosquito-net-based camera.

With this setting we overcome the difficult problem of implementing a mask inside the camera, which would have to be precisely positioned under the cover glass of the sensor, $0.2mm$ from the silicon.

We need to find exactly how a movement in depth in object space corresponds to a movement in image space. By differentiating the lens equation $\frac{1}{a} + \frac{1}{b} = \frac{1}{f}$ we obtain $\frac{da}{a^2} = -\frac{db}{b^2}$. Thus a displacement by $da = 10cm$ away from the plane of focus and towards the transparency produces a corresponding displacement of $-da\frac{b^2}{a^2} = 0.16mm$ away from the sensor surface. At the same time the image of our $2mm$ grid of the transparency is reduced linearly to $T\frac{b}{a} = 0.08mm$, which gives us an F/number of about 2, and high effective resolution of the final image (defined by the small effective mesh period of $0.08mm$). In this way an outside mesh is used as a much finer mesh inside the camera.

Figure 4b shows an image taken through the transparency. An example of refocusing using this image is presented in Figure 8. It is achieved by generating different views and mixing them to effectively integrate over the virtual aperture. The result is shown in Figure 8, where the two images are produced with two different registrations of the views, one on the background, and the other one - on the foreground.

4.4   (Micro) Lens-Based Cameras

An example image obtained through our lens-based camera is shown in Figure 4c. Notice the zoomed in area which clearly displays the complex structure of the image, encoding $3D$ information. The FFT of this image is shown in the same figure.

To obtain horizontal parallax, we apply the demultiplexing described in section 4.1 to generate 11 views. Two images resulting from this process are shown in Figure 9. They show clear parallax. It can be observed at close examination, for example in the marked region. Note that left and right images are switched so that stereo parallax can be noticed with proper cross eyes observation.

Supplemental videos show the effect of smoothly changing views for this image and another "Tall Grass" image. Also, we provide an example of refocusing based on mixing those 11 views using the method of [6].

Fig. 9. Stereo images from our lens-based camera. Stereopsis can be achieved with crossed eye observation. Close examination of the marked areas will also reveal horizontal parallax between the pictures.

5   Conclusions and Future Work

In this paper, we have derived a new mathematical formalism for analyzing light-field cameras in the frequency domain. The 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.

The important conclusion we draw out of this finding is that the "heterodyning" method is not restricted to a special "heterodyning" type of camera. It is a new method of demultiplexing captured lightfield data from **any** lightfield camera, mask based or microlens based.

Analyzing light-field cameras in the frequency domain and designing new approaches based on that is a new emerging area of research, covering both mask and lens based cameras. Mask based cameras are much cheaper and easier to build. Using microlenses, prisms, and other arrays of optical elements with the frequency multiplexing approach might have unexpected new potential. There is much more to be done in this direction.

In relation to this approach, we have proposed and implemented a new "mosquito net" lightfield camera based on masks/nets in front of the main lens. We have also built prototypes of all these cameras in order to demonstrate the validity of this formalism.

In the last year, a whole new dimension has been added to to integral photography due to unexpected possibilities with frequency domain multiplexing. Compact camera designs, and later post-processing based on computer vision are opening new possibilities. We hope that our work will inspire others to explore further this rich domain.

References

1. Guillemin, V., Sternberg, S.: Symplectic techniques in physics. (1985)
2. Ives, F.: Patent US 725,567. (1903)

3. Lippmann, G.: Epreuves reversible donnant la sensation du relief. J. Phys. 7 (1908) 821–825
4. Levoy, M., Hanrahan, P.: Light field rendering. ACM Trans. Graph. (1996) 31–42
5. Gortler, S.J., Grzeszczuk, R., Szeliski, R., Cohen, M.F.: The lumigraph. ACM Trans. Graph. (1996) 43–54
6. Ng, R., Levoy, M., Brdif, M., Duval, G., Horowitz, M., Hanrahan, P.: Light field photography with a hand-held plenoptic camera. Tech. Rep. (2005)
7. Chai, J., Chan, S., Shum, H., Tong, X.: Plenoptic sampling. ACM Trans. Graph. (2000) 307–318
8. Durand, F., Holzschuch, N., Soler, C., Chan, E., Sillion, F.: A frequency analysis of light transport. ACM Trans. Graph. (2005) 1115–1126
9. Ng, R.: Fourier slice photography. ACM Trans. Graph. (2005) 735–744
10. Veeraraghavan, A., Mohan, A., Agrawal, A., Raskar, R., Tumblin, J.: Dappled photography: Mask enhanced cameras for heterodyned light fields and coded aperture refocusing. ACM Trans. Graph. **26**(3) (2007) 69
11. Gerrard, A., Burch, J.M.: Introduction to matrix methods in optics. (1994)
12. Georgiev, T., Zheng, K.C., Curless, B., Salesin, D., Nayar, S., Intwala, C.: Spatio-angular resolution tradeoffs in integral photography. In: Rendering Techniques 2006: 17th Eurographics Workshop on Rendering. (June 2006) 263–272
13. Oppenheim, A.V., Willsky, A.S.: Signals and Systems. Prentice Hall, Upper Saddle River, New Jersey (1997)
14. Adelson, T., Wang, J.: Single lens stereo with a plenoptic camera. IEEE Transactions on Pattern Analysis and Machine Intelligence (1992) 99–106