The Light Field Camera: Extended Depth of Field, Aliasing, and Superresolution

Tom E. Bishop, *Member*, *IEEE*, and Paolo Favaro, *Member*, *IEEE*

Abstract—Portable light field (LF) cameras have demonstrated capabilities beyond conventional cameras. In a single snapshot, they enable digital image refocusing and 3D reconstruction. We show that they obtain a larger depth of field but maintain the ability to reconstruct detail at high resolution. In fact, all depths are approximately focused, except for a thin slab where blur size is bounded, i.e., their depth of field is essentially inverted compared to regular cameras. Crucial to their success is the way they sample the LF, trading off spatial versus angular resolution, and how aliasing affects the LF. We show that applying traditional multiview stereo methods to the extracted low-resolution views can result in reconstruction errors due to aliasing. We address these challenges using an explicit image formation model, and incorporate Lambertian and texture preserving priors to reconstruct both scene depth and its superresolved texture in a variational Bayesian framework, eliminating aliasing by fusing multiview information. We demonstrate the method on synthetic and real images captured with our LF camera, and show that it can outperform other computational camera systems.

Index Terms—Computational photography, superresolution, deconvolution, blind deconvolution, multiview stereo, shape from defocus.

1 Introduction

Recently, we have seen that not only is it possible to build practical integral imaging and mask enhanced systems based on commercial cameras [1], [2], [3], [4], but also that such cameras provide an advantage over traditional imaging systems. The insertion of a microlens array in a conventional camera results in a *plenoptic* or *light field (LF)* camera. These designs enable new imaging modalities, for instance, *digital refocusing* [3] and the recovery of transparent objects in microscopy [5] from a single snapshot.

Surprisingly, the LF camera design enables a dramatic depth of field (DoF) extension. In a regular camera, blur is small only close to the focal plane in space and grows very large elsewhere; in the LF camera, blur behaves in exactly the opposite manner: It is small everywhere except nearby the focal plane, where it is bounded by the microlens size (see Fig. 2).

Unfortunately, existing approaches to producing extended DoF images from Plenoptic cameras suffer from a number of drawbacks, such as rendering images only at the microlens array resolution [3] or dealing with blur by requiring small microlens apertures that sacrifice light [6]. We investigate a method to obtain both depth maps and high resolution extended DoF images from a single plenoptic exposure.

- *T.E. Bishop is with Anthropics Technology Ltd., 22 Baseline Studios, Whitchurch Road, London W11 4AT, United Kingdom. E-mail: T.E.Bishop@gmail.com.*
- *P. Favaro is with the Joint Research Institute on Image and Signal Processing, Heriot Watt University and University of Edinburgh, Riccarton, Edinburgh EH14 4AS, United Kingdom. E-mail: Paolo.Favaro@hw.ac.uk.*

Manuscript received 24 Mar. 2010; revised 13 Apr. 2011; accepted 21 July 2011; published online 4 Aug. 2011.
Recommended for acceptance by R. Ramamoorthi.
For information on obtaining reprints of this article, please send e-mail to: tpami@computer.org, and reference IEEECS Log Number TPAMI-2010-03-0220.
Digital Object Identifier no. 10.1109/TPAMI.2011.168.

Our strategy to enhance resolution is to exploit the fact that LFs of natural scenes are not a collection of random signals. Rather, they generally satisfy models of limited complexity [7], such as the Lambertian model we consider here. We notice that an LF can be interpreted as a collection of views with unknown shifts between them (see Fig. 1) containing complementary, but related, information. We can fuse this information with a superresolution (SR) algorithm to recover the high-resolution image. Solving the SR problem requires recovering the shifts, or in our case, the depth map of the scene, but this converts the original problem of DoF extension into a simpler one, compared to a regular camera where the corresponding blind deconvolution problem is more ill posed.

Thus, we propose a two stage algorithm where we first recover the depth by establishing correspondence between the views, and then use this to form the space-varying point spread function (PSF) model, which is employed in a Bayesian deconvolution approach to estimate the SR extended DoF image. We will study the sampling patterns involved, how aliasing affects the views, and under what conditions we can hope to obtain good SR results. In [1], it is noted that plenoptic cameras tend to avoid the aliasing of angular samples that is experienced with a camera array, but do not consider spatial aliasing (as in Fig. 3), which makes aligning the views problematic.[1] We propose an iterative depth-dependent antialiasing method to solve this challenge. Our analysis also shows that the samples in an LF camera periodically overlap, which results in the reconstruction quality of an all-in-focus image varying periodically with depth (see Fig. 16). Nevertheless, the method still outperforms competing systems at these depths.

1. On the contrary, note the well-known spatio-angular information tradeoff: In fact, it is only *because* the views are aliased that superresolved refocused images may be obtained since they contain new information.

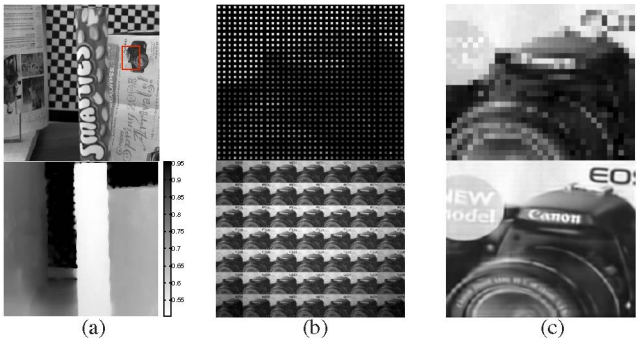

Fig. 1. Top row: (a) One view from our LF image, (b) detail of corresponding LF image, (c) detail of central view, (one pixel per microlens, as in a traditional rendering [3]). Bottom row: (a) Estimated depth map (scale in meters), (b) above LF image rearranged as views, (c) superresolved central view.

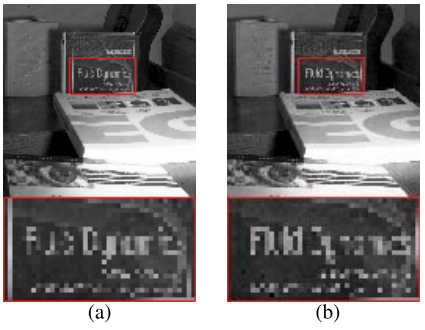

Fig. 3. Example of plenoptic view aliasing. (a) One view extracted from an LF image (courtesy of Lumsdaine and Georgiev [6]) with one region enlarged at the bottom. (b) Another view from the same data set (along the horizontal axis). Notice the aliasing affecting these views. Establishing correspondence for depth inference in these views is therefore prone to errors.

2 Prior Work and Contributions

This work relates to *computational photography*, an emerging field encompassing several methods to enhance the capabilities and overcome limitations of standard digital photography by jointly designing imaging devices and reconstruction algorithms. One of the first devices based on the principles of integral photography [8] is the *plenoptic camera*, proposed in computer vision by Adelson and Wang [1] to infer depth in a single snapshot, and more recently engineered into a single package chip [9]. In its original design, the plenoptic camera consists of a camera body with a single main lens, a lenticular array at its focal plane, and an additional relay lens to form the image on a sensor. Ng et al. [3] present a similar design, but in a portable hand-held device, and propose digital refocusing, i.e., the ability to change the focus setting after capturing the image. While their method yields impressive results, there is one caveat: The spatial resolution of the refocused images is lower than that of the image sensor, just equal to the number of microlenses in the camera—e.g., as low as 90K pixels from a 16MP camera.

An alternative to the plenoptic camera is the *programmable aperture* camera [10], which captures LF data by multiplexing views of the scene. While this approach allows recovery of images at full sensor resolution, the price to pay is a long exposure time or a low (SNR). More importantly, the scene has to be static. If motion occurs, views must be aligned, which adds further complexity and computational cost to the system. Another interesting design proposed by Veeraraghavan et al. [4] is the heterodyne camera, where the LF is modulated using an attenuating mask close to the sensor plane. The authors mention an advantage of this system is the reconstruction of high-resolution images at the plane in focus in addition to the sampled light field, but with a considerable limitation: The SNR is much reduced due to light attenuation at the mask. Georgiev and Intwala [2] suggest variants on the LF camera design. Instead of internal microlenses, they add optics external to the main lens, such as arrays of positive/negative lenses or prisms. Unfortunately, while appealing in their simplicity, these designs are bulky and tend to suffer from higher order optical aberrations. Ben-Ezra et al. [11] propose a novel resolution-enhancing design, which captures multiple frames, shifted by known subpixel amounts. To achieve this without motion blur, microactuators instantaneously shift the sensor before capturing each frame. The frames are combined to reconstruct a single high-resolution image. As in [10], this method trades off exposure time for spatial resolution. Other optical designs such as wavefront coding [12] and focal sweep [13] have been proposed that attempt to recover all-focused images via *approximately* depth-invariant blurs that are easily deconvolved. This offers a simplified approach but does not give the possibility of refocusing, as the depth map cannot be estimated.

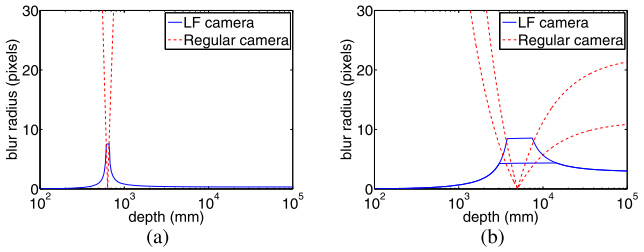

Fig. 2. Conventional versus LF camera. The size of the blur disc in a conventional camera increases quickly away from the plane in focus ((a) 635 mm, (b) 5 m). The camera has an 80 mm main lens at f/3.2 (thin lines) or f/6.3 (thick lines). By comparison, the blur under each microlens in the LF camera remains small over most depths, irrespective of main lens f/#. Its size (see (11)) behaves in the opposite manner to the regular camera blur, obtaining a maximum around the main lens plane in focus.

One could also aim to improve the sampled LF resolution by designing algorithms, rather than hardware, that exploit prior knowledge about the scene. We rely on reconstructing the depth map and pose the problem as *superresolving* the LF, starting from multiple low-resolution images with unknown translations. This approach relates to a large bulk of image processing literature [14], [15], [16], [17], [18], [19]. However, in the computational photography field, prior work is limited to: Chan et al. [20], where a compound eye system is only simulated, Levin et al. [7], who describe tradeoffs between different camera designs for LF recovery, and Lumsdaine and Georgiev's method [21] for rendering high-resolution images from a plenoptic camera.

In [21], the authors discuss when subimages under each microlens are flipped (telescopic) or not (binocular), then scale up their central part, assuming that the scene is at a

constant user-defined depth plane. This approach does not fully address LF superresolution. First, no depth map (i.e., the alignment between subimages) is estimated. Second, they only use interpolation to restore the LF, without deconvolution, meaning that overlapping subimage pixels are dropped or averaged instead of being fused. Without deconvolution, small apertures are required on the microlenses which does not efficiently use available light. Moreover, their results are not attained under a globally consistent restoration model, and no regularization is used. We will see in Fig. 17 that this is suboptimal.

In concurrent work, Levin et al. [7] describe analysis and algorithms closely related to our method. They focus on the tradeoffs in recovering the LF of a scene by comparing different camera designs and consider the use of priors in a Bayesian framework. Our approach differs in several ways: First, we derive and fully analyze an LF camera image formation model and verify its validity on real images; second, we explicitly enforce Lambertianity and use image texture priors that are unlike their mixture of Gaussians derivative priors.

Aliasing in systems similar to the plenoptic camera has been analyzed before. In particular, researchers have studied how camera geometry affects the sampling pattern. Georgiev and Lumsdaine [22] consider the choice of sensor-to-microlens spacing in a plenoptic camera. The magnification for different image planes inside the camera is then investigated, along with the depth of focus (although the scene DoF is not explicitly computed). Stewart et al. [23] examine aliasing of LFs sampled by camera arrays, suggesting that sufficiently large apertures (equal to the intercamera spacing) and pixels with full fill-factor provide the required prealiasing, so long as only one scene depth needs to be in focus. Stewart et al. [24] describe LF rendering methods to deal with such aliasing when these conditions are not met, based on combining band-limited [25] and wide-aperture [26] reconstructions. However, these methods do not consider using the depth map in the LF reconstruction—which the recovery of a high-resolution all-focused image requires (e.g., when the focal plane is moved in [24, Figs. 7c and 7e]).

In [25], Chai et al. study the sampling rates required to avoid aliased LF rendering, observing (as we do) that correct antialiasing is depth dependent. We emphasize, however, that the sampling pattern in a plenoptic camera leads to different requirements. Ng [27] discusses *postaliasing* artifacts, resulting from approximate LF refocusing. Adelson and Wang's pioneering work [1] estimated depth from the LF views, but without considering aliasing. Vaish et al. [28] perform multiview depth estimation from an array of about a hundred cameras, a system that is structurally similar to a plenoptic camera. However, we will see some fundamental differences in the aliasing of LFs captured by the two systems.

To the best of our knowledge, none of the prior work, with the exception of the initial versions of this work [29] and [30], covered the following contributions of our paper:

1. an explicit image formation model obtained by characterizing the spatially varying PSF of a plenoptic camera under Gaussian optics assumptions for a depth varying scene;
2. novel analysis of aliasing in views and the DoF of a plenoptic camera (that takes into account sampling and the nonnegligible blur generated by the microlenses);
3. a method to reconstruct the light field in a Bayesian framework, explicitly introducing Lambertian reflectance priors in the image formation model; notice this allows us to design an SR algorithm which recovers more information than the one predicted by the basic sampling theorem;
4. a method to reduce aliasing of views via space-varying filtering of the recorded LF, and an iterative multiview depth estimation procedure, that benefits from this reduction;
5. a comparison of the actual resolution attainable at different depths with the SR approach versus other methods. We demonstrate that the LF camera outperforms regular, coded aperture, and focal sweep cameras, in noisy conditions.

Note that our approach is general and not designed for any particular camera settings, although well-chosen parameters will likely give better results for a particular working depth range. One limitation of the depth estimation method is that the depth map is only found at the microlens array resolution, but this is often sufficient in practice. Currently, we do not model occlusions; this could be handled with a revision of our framework; however, for typical lens apertures the amount of occlusion present does not generate significant artifacts in our experiments.

We begin in Section 3 with a general overview of our approach to plenoptic depth estimation and SR. We provide definitions used in our model in Section 4, and describe how scene points map to the sensor using an equivalent internal representation. In Section 5, we describe several effects governed by the choice of camera parameters, and their relation to restoration quality. In Section 6, we analyze two interpretations of the LF data: as views or subimages, considering how sampling introduces aliasing, making depth estimation challenging. We define ideal and practical antialiased filtering solutions that we use in our regularized depth map estimation solution. From this depth map and the optical model we compute the camera's space-varying PSF matrix in Section 7.1, which is used in our Bayesian image restoration algorithm in Section 7.2. Finally, we present experimental results with both algorithms in Section 8.

3 Superresolution and Depth Estimation with the Light Field Camera

In this section, we outline how we will tackle the SR and depth estimation tasks, leaving most of the fine (but important) details to later. In order to restore the images obtained by the plenoptic imaging model at a resolution that is higher than the number of microlenses, we need to first determine an image formation model of such a system. This model allows us to simulate a light field image $\boldsymbol{l}$ given the scene radiance (the all-focused image) $\boldsymbol{r}$ and the *space-varying* PSF matrix of the camera $\boldsymbol{H}_s$. As will be shown in the next sections, when there is no noise, these quantities can be related via the simple linear relationship

$$\boldsymbol{l} = \boldsymbol{H}_s \boldsymbol{r}, \tag{1}$$

where $\boldsymbol{l}$ and $\boldsymbol{r}$ are rearranged as column vectors. $\boldsymbol{H}_s$ embeds both the camera geometry (e.g., its internal structure, the number, size, and parameters of the optics) and the scene disparity map $\boldsymbol{s}$. In general, the only quantities directly observable are the LF image $\boldsymbol{l}$ and the camera geometry, and one has to recover both $\boldsymbol{r}$ and $\boldsymbol{s}$. Due to the dimensionality of the problem, in this manuscript we consider a two-step approach where we first estimate the disparity map $\boldsymbol{s}$ and then recover the radiance $\boldsymbol{r}$ given $\boldsymbol{s}$. In both steps, we formulate inference as an energy minimization.

For now, assume that the disparity map $\boldsymbol{s}$ is known. Then, one can employ (SR) by estimating $\boldsymbol{r}$ directly from the observations. Due to the fact that the problem may be particularly ill-posed depending on the extent of the complete system, proper regularization of the solution through prior modeling of the image data is essential. We can then formulate the estimation of $\boldsymbol{r}$ in the Bayesian framework. Under the typical assumption of additive Gaussian observation noise $\boldsymbol{w}$, the model becomes $\boldsymbol{l} = \boldsymbol{H}_s \boldsymbol{r} + \boldsymbol{w}$, to which we can associate a conditional probability funtion (PDF), the likelihood $\mathrm{p}(\boldsymbol{l}|\, \boldsymbol{r}, \boldsymbol{H}_s.)$.

We then introduce priors on $\boldsymbol{r}$. We use a recently developed prior [31], [32], which can locally recover texture, in addition to smooth and edge regions in the image. By combining the prior $\mathrm{p}(\boldsymbol{r})$ with the likelihood from the noisy image formation model we can then solve the maximum a posteriori (MAP) problem:

$$\hat{\boldsymbol{r}} = \arg\max_{\boldsymbol{r}} \mathrm{p}(\boldsymbol{l}|\, \boldsymbol{r}, \boldsymbol{H}_s)\mathrm{p}(\boldsymbol{r}). \tag{2}$$

The MAP problem requires evaluating $\boldsymbol{H}_s$, which depends on the unknown disparity map $\boldsymbol{s}$. To obtain $\boldsymbol{s}$ we consider extracting views (images from different viewpoints) from the LF so that our input data are suitable for a multiview geometry algorithm (see Section 6). The multiview depth estimation problem can then be formulated as inferring a disparity map $\boldsymbol{s} \doteq \{s(\boldsymbol{c_k})\}$ by finding correspondences between the views for each 2D location $\boldsymbol{c_k}$ visible in the scene. Let $\hat{V}_{\boldsymbol{q}}$ denote the sampled view from the 2D viewing direction $\boldsymbol{q}$ and $\hat{V}_{\boldsymbol{q}}(\boldsymbol{k})$ the color measured at a pixel $\boldsymbol{k}$ within $\hat{V}_{\boldsymbol{q}}$. Then, as we will see, depth estimation can be posed as the minimization of the joint matching error (plus a suitable regularization term) between all combinations of pairs of views:

$$E_{\mathrm{data}}(\boldsymbol{s}) = \sum_{\forall \boldsymbol{q}_1, \boldsymbol{q}_2, \boldsymbol{k}} \Phi(\hat{V}_{\boldsymbol{q}_1}(\boldsymbol{k} + s(\boldsymbol{c_k})\boldsymbol{q}_1) - \hat{V}_{\boldsymbol{q}_2}(\boldsymbol{k} + s(\boldsymbol{c_k})\boldsymbol{q}_2)), \tag{3}$$

where Φ is some robust norm and $\boldsymbol{q}_1, \boldsymbol{q}_2$ are the 2D offsets between each view and the central view (the exact definition is given in Section 6.1). In practice, to save computational effort, only a subset of view pairs $\{\boldsymbol{q}_1, \boldsymbol{q}_2\}$ may be used in (3). Notice that this definition of the 2D offset implicitly fixes the central view as the reference frame for the disparity map $\boldsymbol{s}$.

As the views may be aliased, minimizing (3) is liable to cause incorrect depth estimates around areas of high-spatial frequency in the scene. Put simply, even when scene objects are Lambertian and without the presence of noise, the views might not satisfy the photoconsistency criterion sufficiently well so that E_{data} may not have a minimum at the true depth map. Moreover, subpixel accuracy is usually obtained through interpolation. This might be a reasonable approximation when the views collect samples of a band-limited (i.e., sufficiently smooth) texture. However, as shown in Fig. 3, this is not the case with LF cameras. Therefore, we have to explicitly define how samples are interpolated and study how this affects the matching of views.

We shall also see that there are certain planes where the sample locations from different views coincide. At these planes, aliasing no longer affects depth estimation, but extra information for (SR) is diminished.

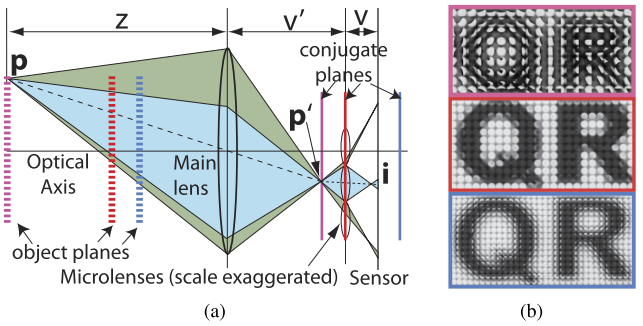

Fig. 4. (a) 2D Schematic of a LF camera. Rays from a point $\boldsymbol{p}$ are split into several beams by the microlens array. (b) Three example images corresponding to the colored planes in space (dashed) and their conjugates (solid). Top: $\boldsymbol{p}'$ before microlenses; subimages flipped. Middle: $\boldsymbol{p}'$ on microlenses; no repetitions. Bottom: $\boldsymbol{p}'$ is virtual, beyond the microlenses; no flipping.

4 Image Formation of a Light-Field Camera

In this section, we derive the image formation model of a plenoptic camera, and define the relationship between different camera parameters. To yield a practical computational model suitable for our algorithm (Section 7), we investigate the imaging process with tools from geometric optics [33], ignoring diffraction effects, and using the thin lens model. We will also analyze sampling of the LF camera by using the phase-space domain [7].

4.1 Imaging Model

In our investigation, we rebuilt a light field camera similar to that of Ng et al. [3]—essentially a regular camera with a microlens array placed near the sensor (see Fig. 4)—but, as in [21], we consider the imaging system under a general configuration of the optical elements. However, unlike in any previous work, we determine the image formation model of the camera so that it can be used for SR or more general tasks.

We use a general 3D scene representation (ignoring occlusions), consisting of the all-focused image, or *radiance*, $r(\boldsymbol{u})$ (as captured by a pinhole camera, i.e., with an infinitesimally small aperture), plus a depth map $z(\boldsymbol{u})$ associated with each point $\boldsymbol{u}$. Both r and z are defined at the microlens array plane such that r is the all-focused image that would be captured by a regular camera. In this way, we can analyze both the PSF of each point in space (corresponding to a column of $\boldsymbol{H}_s$) and sampling and aliasing effects in the captured LF with a less bulky notation. We first consider the equivalence between using points in space

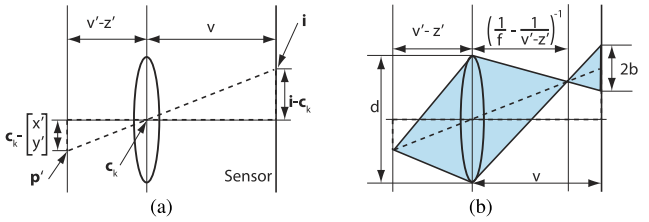

Fig. 5. (a) Imaging the conjugate object onto the sensor via one microlens. (b) Defocus blur under one microlens. The radius b is obtained by similar triangles.

or inside the camera (Section 4.1.1). Then, we specify continuous and discrete versions of the coordinates used to parameterize the captured LF. The reader already familiar with the topic may skip this section.

4.1.1 Conjugate Object Representation

As stated above, we consider an equivalent representation defined entirely inside the camera. Each point $\boldsymbol{p} \doteq [x\ y\ z]^T \in \mathbb{R}^3$ in space has a unique corresponding *conjugate* point $\boldsymbol{p}' \doteq [x'\ y'\ z']^T \in \mathbb{R}^3$ where its rays focus behind the lens, and thus the two are interchangeable (Fig. 4). Therefore, a surface in space corresponds to a conjugate surface $z'(\boldsymbol{u})$ inside the camera, parameterized by coordinates $\boldsymbol{u} \in \mathbb{R}^2$ on the microlens plane. Consider the projection of a point in space on such a surface. The thin lens law implies

$$\boldsymbol{p}' = \frac{F}{z - F}\begin{bmatrix} -1 & 0 & 0 \\ 0 & -1 & 0 \\ 0 & 0 & 1 \end{bmatrix}\boldsymbol{p}, \tag{4}$$

where F denotes the main lens focal length. The microlens array is located at a distance v' behind the main lens. It consists of $K_1 \times K_2$ microlenses, which we index with $\boldsymbol{k} = [k_1, k_2]^T, k_1 \in \{1 \cdots K_1\}, k_2 \in \{1 \cdots K_2\}$. Their centers are located at $\boldsymbol{c_k} = d\boldsymbol{k}$, with spacing d. The projection $\boldsymbol{i} = [i, j]^T$ of $\boldsymbol{p}'$ onto the sensor plane through a microlens at $\boldsymbol{c_k}$ is then computed as

$$\boldsymbol{i} = \boldsymbol{c_k} + \frac{v}{v' - z'}(\boldsymbol{c_k} - [x', y']^T), \tag{5}$$

as shown in Fig. 5. Coordinates $\boldsymbol{i}$, $\boldsymbol{c_k}$, and $[x'\ y']^T$ all have their origin in the center of the sensor, coinciding with the optical axis.

4.1.2 Image Flipping

Consider for now that the microlenses have tiny apertures behaving as pinholes (we consider microlens blur later). Then each microlens subimage is a projection of a portion of the conjugate image onto the sensor, with the relation between points as in (5). This image may also be flipped along both axes, with respect to the conjugate image[2] (see third row of Fig. 4). Flipping occurs when the direction of a vector on the sensor $\Delta\boldsymbol{i}$ disagrees with that of its projection in the conjugate image, i.e., using $\Delta\boldsymbol{i} = -\frac{v}{v'-z'}\Delta[x'\ y']^T$, when $z' < v'$ (assuming that $z' > 0$, i.e., objects in space are further than F from the camera).

2. We use as a reference the conjugate image, as this has the same orientation as the image captured by a conventional camera, which is automatically flipped in software to return it to the same orientation as the object in space.

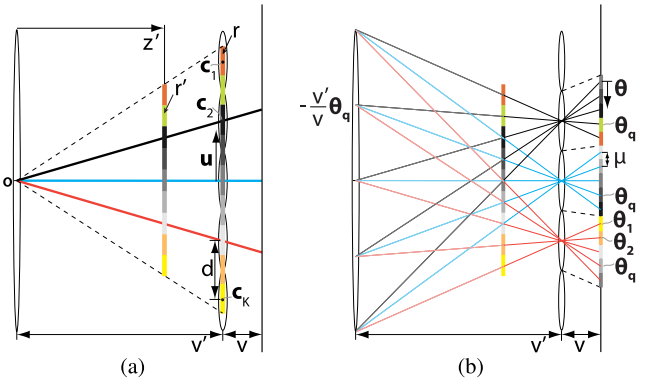

Fig. 6. Coordinates and sampling in the plenoptic camera. (a) Radiance r, at v', projected through O to the *conjugate image* r' at z'. Bold rays indicate a view. (b) Imaging r' onto *subimages* behind each microlens. View aliasing occurs: The frequency in r is higher than the microlens pitch d, so views contain samples unrelated to their neighbors via interpolation.

4.1.3 Angular Coordinates

We parameterize the 4D LF as in Fig. 6. We have defined the 2D spatial coordinates $\boldsymbol{u}$; the continuous 2D angular coordinates $\boldsymbol{\theta} \in \Theta \subset \mathbb{R}^2$ are locally defined on the sensor plane relative to each microlens. Θ defines the set of $\boldsymbol{\theta}$ contained within the main lens aperture's projection onto the sensor, which we assume for now is square. $\boldsymbol{\theta}$ under each microlens $\boldsymbol{k}$ is defined so that it projects to the same main lens position (the local origin $\boldsymbol{\theta}_0$ is the projection of the main lens center through $\boldsymbol{c_k}$).[3] Its discretized version is $\boldsymbol{\theta_q} \doteq \mu\boldsymbol{q}$, where μ is the pixel width, $\boldsymbol{q} = [q_1, q_2]^T$ and $q_1, q_2 \in \{-\lfloor (Q-1)/2 \rfloor, \ldots, 0, \ldots, +\lfloor (Q-1)/2 \rfloor\}$, with $\lfloor a \rfloor$ the floor operator. $\boldsymbol{\theta_q}$ indexes the sensor pixels relative to each microlens, so that we have a total of Q^2 angular samples (when Q is odd).

4.1.4 Views and Subimages

The set of pixels, one per microlens, that map to the same point on the main lens (i.e., with the same $\boldsymbol{\theta}$) form the image $V_{\boldsymbol{\theta}}(\boldsymbol{c})$, which we term a *view*. The pixels under a microlens at $\boldsymbol{c}$ form a *subimage* $S_{\boldsymbol{c}}(\boldsymbol{\theta})$. In Fig. 6, we represent the plenoptic image formation process as a mapping between $r(\boldsymbol{u})$ and each view or subimage. First, it is scaled down by $\frac{z'}{v'}$ to give the conjugate images $r'(\boldsymbol{u}')$ at z'. Each pixel is scaled to a different conjugate image depending on the depth map at that position. These are then sampled by the set of rays passing through a point on the main lens or the microlenses; we discuss this process further in Section 6.

If the main lens aperture and microlens spacing d are chosen so that the subimages fully tile the sensor without overlap, the subimage size is (see Fig. 7, left)

$$D\frac{v}{v'} = d\frac{v' + v}{v'}, \qquad \text{thus} \qquad Q = \frac{d}{\mu}\frac{v' + v}{v'}. \tag{6}$$

Finally, we denote the discretized and lexicographically ordered pixel coordinates by $\boldsymbol{i}_m = [i_m, j_m]^T$ with $m \in \{1 \cdots M\}$, where $M = Q^2K_1K_2$ and satisfy $\boldsymbol{i} = \mu[i_m, j_m]^T$.

3. The $(\boldsymbol{u}, \boldsymbol{\theta})$ pair is similar to the parameterization used in [7], [23], where LF coordinates are defined on two parallel planes. Here, the corresponding planes are the microlens plane ($\boldsymbol{u}$) and the main lens (implicitly via $\boldsymbol{\theta}$).

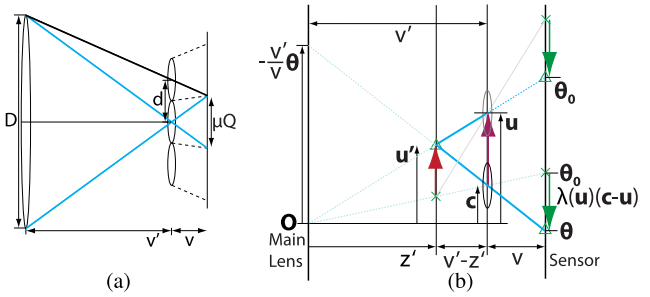

Fig. 7. (a) Choice of aperture sizes for fully tiling subimages. (b) Imaging the purple vector under different microlenses. A point $\boldsymbol{u}$ on the radiance has a conjugate point indicated by the triangle at $\boldsymbol{u}'$, which is imaged by a microlens centered at an arbitrary center $\boldsymbol{c}$, giving a projected point at angle $\boldsymbol{\theta}$. A second microlens positioned at $\boldsymbol{u}$ images the same point at the central view $\boldsymbol{\theta} = 0$ on the line through O. The scaling of the purple vector to the green vectors is λ.

The coordinates $(\boldsymbol{q}, \boldsymbol{k})$ will be used often in the next sections to parameterize the light field, which is originally captured with respect to the coordinates $[i_m, j_m]^T = \boldsymbol{q} + Q\boldsymbol{k}$. The inverse mapping is

$$(\boldsymbol{q}, \boldsymbol{k}) = \left(\operatorname{mod}\left(\begin{bmatrix} i_m \\ j_m \end{bmatrix} + \frac{Q}{2}, Q \right) - \frac{Q}{2}, \left\lfloor \frac{[i_m, j_m]^T}{Q} + \frac{1}{2} \right\rfloor \right). \quad (7)$$

4.1.5 Ray Space Representation

In Fig. 8, we show an alternate view of sampling, aliasing, and blurring effects in the plenoptic camera, via the (2D) ray space representation of Levin et al. [7]. In our version, we use the internal camera coordinates: $\boldsymbol{u}$ (spatial) and the projection of $\boldsymbol{\theta}$ onto the main lens (angular). A point in this space represents a ray through the corresponding positions on the main lens and microlens array, while a ray with constant color corresponds to a particular conjugate point (e.g., the vertical red rays are points on the microlens array, conjugate to the main lens plane in focus in space; different slopes represent other depths, with the same colors as the planes in Fig. 4). Each gray parallelogram indicates the rays in the LF that a sensor pixel integrates; their shear angle corresponds to the conjugate plane the microlenses are focused on. A column of parallelograms is one subimage, and a row represents a view. Gaps between parallelograms are due to pixel fill-factor vertically, and microlens apertures horizontally. To achieve the highest sampling accuracy achievable, the gray parallelogram should be aligned along a ray of constant intensity (so that pixel integration does not lead to a loss of information). $r(\boldsymbol{u})$ is the slice of the rays along the x-axis.

5 Superresolution Limits and Analysis of Model

5.1 Superresolution

So far we have found relations between points in space and their projections on the sensor, assuming infinitesimal microlens apertures. However, in general the image of a point $\boldsymbol{p}'$ on the sensor is a pattern called the system PSF, whose shape depends on the projection of the finite apertures onto the sensor. We will see that these blur sizes, as well as the number of times a point is imaged under different microlenses, can affect the SR quality.

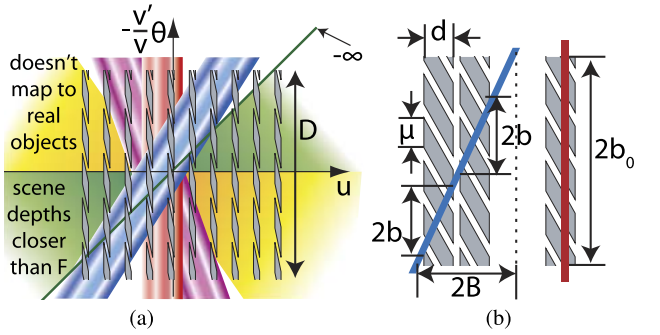

Fig. 8. Ray space diagram [7], with internal camera coordinates. (a) Apertures on the microlenses correspond to parallelograms. The yellow region corresponds to objects inside the camera, while the green line marks the conjugate image at $-\infty$, i.e., the object at distance F from the camera. (b) Microlens blur with finite apertures (vertical scale here is in θ rather than $-\frac{v'}{v}\theta$).

We would like to superresolve at resolutions close to that of the original sensor. While we may render the estimate of r at any resolution, the actual detail that will be recovered will depend on a combination of factors related to how ill posed the inversion is, and how good our system calibration and priors are. Previous SR studies [34], [35], [36] showed in general that the performance of SR algorithms decreases as blur size increases; it also decreases when the ratio of upsampling factor to number of observations remains constant, but the upsampling factor increases. While our imaging model is rather different from those mentioned for regular SR, the same general principles apply. We also point out some important design considerations based upon these limitations.

5.1.1 Main Lens Defocus

The cone of rays passing through $\boldsymbol{p}'$ causes a blur on the microlens array determined by the main lens aperture. This blur disc determines *how many* microlenses capture light from $\boldsymbol{p}'$. With the Lambertian assumption, $\boldsymbol{p}'$ casts the same light on each microlens, and this results in multiple copies of $\boldsymbol{p}'$ in the LF (see first and third image in Fig. 4). Approximating this blur by a Pillbox with radius $B \geq 0$ (i.e., the unit volume cylinder, $h(\boldsymbol{u}) = \frac{1}{\pi B^2}$ for $\|\boldsymbol{u}\|_2^2 \leq B^2$ and 0 otherwise), the main lens blur radius is

$$B = \frac{Dv'}{2}\left|\frac{1}{z'} - \frac{1}{v'}\right|. \quad (8)$$

To characterize the number of repetitions of the same pattern in the scene, we must count how many microlenses fall inside the main lens blur disc. Therefore, in each direction we have

$$\#\text{repetitions} = \frac{2B}{d} = \frac{Dv'}{d}\left|\frac{1}{z'} - \frac{1}{v'}\right|. \quad (9)$$

This ratio is also evident in the ray space (Fig. 8) by considering how many columns (subimages) the blue ray on the right covers. In our SR framework, this number determines how many subimages can be used to superresolve the LF.

5.1.2 Microlens Blur

A necessary condition to superresolve the LF is that the input views are aliased so that they sample different

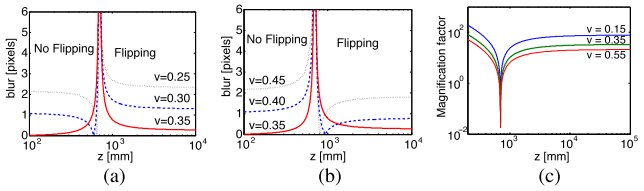

Fig. 9. (a) and (b) Microlens blur radius b versus scene depth z (in log scale), for several settings of the microlens-to-CCD spacing v. The microlens focal length f is 0.35 mm (note that Ng et al. [3] sets $v = f$). The main lens plane in focus is at 700 mm. Our model can work with any suitable settings. (c) Magnification factor λ between the radiance at the microlens plane and each subimage, under similar settings.

information, i.e., they are not just shifted and interpolated versions of the same image. We discuss aliasing further in Section 6, but here it suffices to note that increasing subimage blur reduces complementary information in the LF available to perform SR. The blur of each microlens that is fully covered by the main lens PSF (some are not, see Section 7.1.2) is also a pillbox due to the microlens aperture, with radius (see Fig. 5):

$$b = \frac{dv}{2}\left|\frac{1}{f} - \frac{1}{v' - z'} - \frac{1}{v}\right|, \tag{10}$$

where f is the microlens focal length. In Fig. 8, the microlens blur is the horizontal projection of a ray onto the subimage pixels it intersects. Restoration performance is optimal at depths where $b \to 0$ (rays like the purple one in Fig. 8, with the same slope as the pixel shear), obtained for points $\boldsymbol{p}'$ at a distance $z' = v' - \frac{vf}{v-f}$, and it will degrade away from these depths. When a microlens is not fully inside the main lens PSF, then its blur radius is smaller (see Section 7). For simplicity, consider the blur radius b_0 when both microlens and main lens share the same optical axis (see the general case in Section 7):

$$b_0 = \min\left\{\frac{2Bb}{d}, b\right\}. \tag{11}$$

Notice that as $z' \to v'$, although $b \to \infty$, the radius b_0 converges to $\frac{Dv}{2v'}$ (see Fig. 2). In Fig. 8b, this case occurs with the red ray where microlens blur radius b_0 is bounded by the subimage size. In Fig. 9, we show how b varies according to scene depth, for a few different settings of the spacing v. It can be seen that this blur behaves roughly in a complementary way to that of a conventional camera, attaining a maximum at the depth where the main lens blur onto the microlens array attains a minimum. In terms of the ray space, changing v (while keeping the x-axis attached to the microlens array) corresponds to a vertical shear of each integration region. With Ng's et al. [3] setting (microlenses focused on the main lens), the regions become rectangular.

5.1.3 Magnification

The microlens blur size alone does not tell the whole story. We must also take into account the *magnification factor* $|\lambda|$ which represents the scaling between the regular image that would form at the microlens plane and the actual image that forms under each microlens. It is defined as (see Fig. 7)

$$\lambda \doteq \frac{z'}{v'}\frac{v}{v' - z'}. \tag{12}$$

Notice the relation to Lumsdaine and Georgiev's magnification factor [21]. Here, however, we refer everything to a common reference frame in r in order to compare multiple depths. Note that the number of repetitions may also be rewritten in terms of λ as

$$\#\text{reps} = \frac{D}{d}\left|\frac{v' - z'}{z'}\right| = \frac{v'}{z'}\left|\frac{v' - z'}{v}\right|\frac{v' + v}{v'} = \frac{1}{|\lambda|}\frac{\mu Q}{d}, \tag{13}$$

where we used (6). Thus, since $\frac{\mu Q}{d}$ is constant, as the number of repetitions increases, the size of subimage features decreases.

The above equation formalizes the constraint that, for each depth, the amount of information from r remains constant, but is split across a different number of subimages. What does change, however, is the *effective* blur, or the integration region in r of a sensor pixel, as it scales with the magnification factor λ.

5.1.4 Coincidence of Samples and Undersampling

In this section, we show that samples from different microlenses coincide in space on some fronto-parallel planes. On these planes the aliasing requirements are not satisfied and the SR restoration performance will decrease. We shall see this experimentally in Section 8.2.2. One such plane is when the conjugate image lies on the microlens plane. Another is shown in Fig. 6, where r' is positioned such that the blue, red and black rays intersect inside the camera at this depth. In Fig. 8, this degeneracy corresponds to a ray passing through exactly the same point in a parallelogram in different subimages.

To have an exact replica of a sample of r' under two microlenses, it must simultaneously project on two discrete pixel coordinates. Considering two microlenses separated by Nd, with $N \in \mathbb{Z}$, then the coordinate $\boldsymbol{\theta_0}$ under the first microlens must correspond to a coordinate $Nd\lambda(\boldsymbol{u}) = T$, where $T \in \mathbb{Z}$. Also, for the corresponding pixel to be fully inside the subimage, $T < \frac{\mu Q}{2}$. By substituting the expression for T, we see that such microlenses must satisfy $N < \frac{\mu Q}{2d\lambda(\boldsymbol{u})} = \frac{1}{2}\#\text{repetitions}$. Moreover, the corresponding depths are those such that $\lambda(\boldsymbol{u}) = \frac{T}{Nd}$ with N and T integers. Finally, we can also determine the total number of genuinely new samples by analyzing the overlap between different microlenses. When $\exists\, N, T \geq 1$ coprime (if not, we could always find additional matching samples), then the total number is

$$\#\text{new pixels} = [NQ + T(K_1 - N)][NQ + T(K_2 - N)]. \tag{14}$$

6 Light Field Antialiased Depth Estimation

We now consider how the extracted LF views or subimages can be related to infer depth, and why classical multiview stereo methods must be adapted to cope with aliasing of the views.

6.1 View and Subimage Correspondences

We begin by using a simple pinhole approximation of the system and by generalizing the microlenses to *virtual* ones centered in the continuous coordinates $\boldsymbol{c}$, rather than discrete positions $\boldsymbol{c_k}$.

Let us use Fig. 7b to consider: 1) how a vector or a point is mapped from the radiance onto the sensor under an arbitrary microlens at $\boldsymbol{c}$, and 2) where the correspondences of the point $\boldsymbol{c}$ lie in other views if this point is at angle $\boldsymbol{\theta}$ in one view.

To find 1, begin with the purple vector $\boldsymbol{u}-\boldsymbol{c}$. By similar triangles and by projecting first through O to the red vector and then through $\boldsymbol{c}$ to the green one, we see that the purple vector image under lens $\boldsymbol{c}$ is scaled by λ (see (12)). Noting that the local origin is $\boldsymbol{\theta}_0$, we can equivalently express the mapping of the point $\boldsymbol{u}$ in r (the tip of the vector) through a lens at $\boldsymbol{c}$ to a subimage correspondence at

$$\boldsymbol{\theta} = \frac{v}{v'-z'(\boldsymbol{u})}\frac{z'(\boldsymbol{u})}{v'}(\boldsymbol{c}-\boldsymbol{u}) \tag{15}$$

$$= \lambda(\boldsymbol{u})(\boldsymbol{c}-\boldsymbol{u}). \tag{16}$$

By inverting this relation, the original point $\boldsymbol{u}$ in r corresponding to any $\boldsymbol{\theta}$ and $\boldsymbol{c}$ is $\boldsymbol{u}(\boldsymbol{\theta},\boldsymbol{c}) = \boldsymbol{c} - \frac{\boldsymbol{\theta}}{\lambda(\boldsymbol{u})}$, and the views and subimages are related to the radiance as: $V_{\boldsymbol{\theta}}(\boldsymbol{c}) = S_{\boldsymbol{c}}(\boldsymbol{\theta}) = r(\boldsymbol{c}-\frac{\boldsymbol{\theta}}{\lambda(\boldsymbol{u})})$. $V_{\boldsymbol{\theta}}(\boldsymbol{c})$ and $S_{\boldsymbol{c}}(\boldsymbol{\theta})$ differ only by which of $\boldsymbol{\theta}$ or $\boldsymbol{c}$ we hold fixed.

Considering (2), we can reformulate the above ideas. For a point $\boldsymbol{c}_1$ in a particular view at angle $\boldsymbol{\theta}_1$, we can find its correspondence $\boldsymbol{u}(\boldsymbol{\theta}_1,\boldsymbol{c}_1)$ in the radiance, and then solve for $\boldsymbol{c}_2$ so that $V_{\boldsymbol{\theta}_1}(\boldsymbol{c}_1) = r(\boldsymbol{u}) = V_{\boldsymbol{\theta}_2}(\boldsymbol{c}_2)$, for arbitrary $\boldsymbol{\theta}_2$. The trick is to refer everything to a common reference frame where $\lambda(\boldsymbol{u})$ is defined (the points share the same depth/magnification). We choose this reference frame to be the central view $\boldsymbol{\theta}_0 = \boldsymbol{0}$, where we have $\boldsymbol{c} = \boldsymbol{u}$ and $V_0(\boldsymbol{c}) = V_0(\boldsymbol{u}) = r(\boldsymbol{u})$, i.e., this view samples the radiance directly. This can be seen in Fig. 7b as the microlens placed at $\boldsymbol{u}$. The result is that $\boldsymbol{c}_1 = \boldsymbol{u} + \frac{\boldsymbol{\theta}_1}{\lambda(\boldsymbol{u})}$ and $\boldsymbol{c}_2 = \boldsymbol{u} + \frac{\boldsymbol{\theta}_2}{\lambda(\boldsymbol{u})}$. The discrete version of these equations, which we describe below, leads us to the view matching in (3). We may also interpret these matches as positions $(\boldsymbol{c}_1,\boldsymbol{\theta}_1)$ and $(\boldsymbol{c}_2,\boldsymbol{\theta}_2)$ on the same ray in Fig. 8, where $\frac{1}{\lambda}$ is the slope of the ray and $\boldsymbol{u}$ is where the ray intersects the x-axis.

6.2 Discretization of Views and Subimages

$V_{\boldsymbol{\theta}}(\boldsymbol{c})$ and $S_{\boldsymbol{c}}(\boldsymbol{\theta})$ are defined for all possible $\boldsymbol{c}$ and $\boldsymbol{\theta}$. In practice, if we approximate the microlens array with an array of pinholes,[4] only a discrete set of samples in each view is available, corresponding to the pinholes at positions $\boldsymbol{c} = \boldsymbol{c}_{\boldsymbol{k}}$. Furthermore, the pixels in each subimage sample the possible views at $\boldsymbol{\theta}_{\boldsymbol{q}}$. Therefore, we define the discrete observed view $\hat{V}_{\boldsymbol{q}}$ at angle $\boldsymbol{\theta}_{\boldsymbol{q}}$ as the image

$$\hat{V}_{\boldsymbol{q}}(\boldsymbol{k}) \doteq V_{\boldsymbol{\theta}_{\boldsymbol{q}}}(\boldsymbol{c}_{\boldsymbol{k}}) = r\left(\boldsymbol{c}_{\boldsymbol{k}} - \frac{\boldsymbol{\theta}_{\boldsymbol{q}}}{\lambda(\boldsymbol{c}_{\boldsymbol{k}})}\right) = r(d\boldsymbol{k} - s(\boldsymbol{c}_{\boldsymbol{k}})d\boldsymbol{q}), \tag{17}$$

where we defined the *view disparity*, in pixels per view, as

$$s(\boldsymbol{u}) \doteq \frac{\mu}{d}\frac{1}{\lambda(\boldsymbol{u})}. \tag{18}$$

The discrete disparity is $s(\boldsymbol{c}_{\boldsymbol{k}})$ and depends on the depth z. The discretized subimages are just a rearrangement of the LF samples; in fact they are also defined by (17), i.e., $\hat{S}_{\boldsymbol{k}}(\boldsymbol{q}) = \hat{V}_{\boldsymbol{q}}(\boldsymbol{k})$.

In a similar manner to the continuous case, two discrete views at $\boldsymbol{q}_1$ and $\boldsymbol{q}_2$ can be related via the reference view as

$$\hat{V}_{\boldsymbol{q}_1}(\boldsymbol{k} + s(\boldsymbol{c}_{\boldsymbol{k}})\boldsymbol{q}_1) = \hat{V}_0(\boldsymbol{k}) = \hat{V}_{\boldsymbol{q}_2}(\boldsymbol{k} + s(\boldsymbol{c}_{\boldsymbol{k}})\boldsymbol{q}_2), \tag{19}$$

thus obtaining the matching terms in (3). By defining the *subimage disparity*, $t(\boldsymbol{c}_{\boldsymbol{k}}) \doteq \frac{1}{s(\boldsymbol{c}_{\boldsymbol{k}})}$, subimages may also be related via

$$\hat{S}_{\boldsymbol{k}_0+\boldsymbol{k}_1}(\boldsymbol{q} + t(\boldsymbol{c}_{\boldsymbol{k}_0})\boldsymbol{k}_1) = \hat{S}_{\boldsymbol{k}_0}(\boldsymbol{q}) = \hat{S}_{\boldsymbol{k}_0+\boldsymbol{k}_2}(\boldsymbol{q} + t(\boldsymbol{c}_{\boldsymbol{k}_0})\boldsymbol{k}_2). \tag{20}$$

The discrete views in (17) are just samples of r with spacing d, but different shifts $s(\boldsymbol{u})d\boldsymbol{q}$, depending on the view angle and depth. The multiview disparity estimation task is to estimate $s(\boldsymbol{u})$ by shifting the views so that they are best aligned. However, this requires subpixel accuracy, i.e., an implicit or explicit reconstruction of r in the continuum. According to the sampling theorem, r may be reconstructed exactly from the samples taken at spacing d so long as the original radiance image contains no frequencies higher than the Nyquist rate $f_0 = \frac{1}{2d}$. In practice, this condition is often not satisfied due to the low resolution of the views, and aliasing occurs. Observe that a larger microlens pitch leads to greater aliasing of the views.

6.3 Ideal and Approximate Antialiasing Filtering

Ideally the LF should be antialiased *before* views are extracted, i.e., we should combine information *across* views. We make use of an extension of the sampling theorem by Papoulis [37], showing that if r is bandlimited with a bandwidth $f_r = Qf_0/\kappa$, then it can be accurately reconstructed on a grid with spacing $\kappa\mu$ if we have Q/κ sets of samples available, with any shifts or linear filtering of the original signal. This implies that we obtain the correctly antialiased views $\tilde{V}_{\boldsymbol{q}}(\boldsymbol{k})$ from the sampled light field as follows:

1. Use a reconstruction method $\mathcal{F}(\cdot)$ jointly on all samples to obtain
$$r(\boldsymbol{u}) = \mathcal{F}(\{\hat{V}_{\boldsymbol{q}'}(\boldsymbol{k}')\},\boldsymbol{u}) = \sum_{\boldsymbol{k}',\boldsymbol{q}'} \Psi_{\boldsymbol{k}',\boldsymbol{q}'}(\boldsymbol{u})\hat{V}_{\boldsymbol{q}'}(\boldsymbol{k}')$$
for some set of interpolating kernels $\Psi_{\boldsymbol{k}',\boldsymbol{q}'}$ (we could use the theorem from [37] to define these kernels, but essentially this operation corresponds to applying any (SR) method).
2. Filter these samples with an antialiasing filter h_{f_0} at the correct Nyquist rate f_0 to obtain
$$\tilde{r}(\boldsymbol{u}) = (h_{f_0} \star r)(\boldsymbol{u}).$$
3. Resample to obtain $\tilde{V}_{\boldsymbol{q}}(\boldsymbol{k}) = \tilde{r}(\boldsymbol{c}_{\boldsymbol{k}} + s(\boldsymbol{u})\frac{d}{\mu}\boldsymbol{\theta}_{\boldsymbol{q}})$.

A drawback of this approach is that a computationally demanding (SR), as well as filtering at a high resolution before extracting low resolution views, is required. Moreover, a chicken-and-egg type problem is apparent: The depth-dependent filters depend on the unknown depth map. Thus, we look at an approximate but efficient method.

Rather than filtering the whole LF simultaneously, we filter each subimage directly, bypassing the reconstruction

4. We will see that the addition of microlens blur due to finite apertures will integrate around these sample locations.

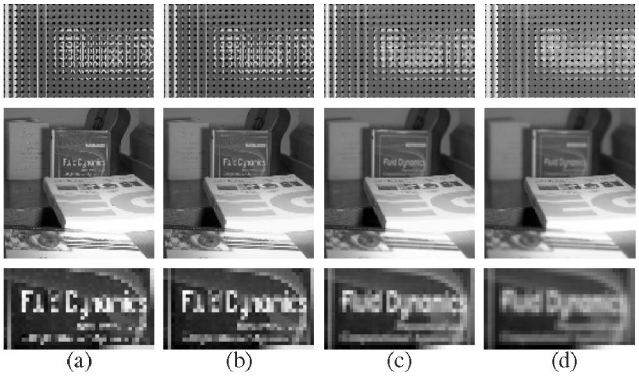

Fig. 10. Antialiasing filtering, increasing from (a) to (d). Top row: Detail of subimages. Middle: Corresponding filtered full view. Bottom: Magnified detail of the view.

step. Since each subimage is a windowed projection of r onto the sensor (ignoring blur for now), we may equivalently project the filters in the same way. This is approximate at subimage boundaries, where we must use filters with a support limited to the domain of Θ. Hence, we upper bound the filter size using a Lanczos windowed version of the ideal Sinc kernel. The antialiasing filter h_{f_0}, defined in r, is projected onto the sensor via the conjugate image at z', i.e., scaling by $|\lambda|$, as in (16). Hence, the scaled filter has physical cutoff frequency $f_0|\lambda|$. We propose an iterative method, beginning with a strong antialiasing filter, and refining the estimate based upon the current depth map. Too much filtering might remove detail for valid matches, while too little may leave aliasing behind (see Fig. 10). We summarize the algorithm as follows:

1. Initialize all filters with cutoff $f_0|\lambda|_{\max}$, i.e., assuming the depth which yields the most aliasing in the working volume.
2. Estimate the disparity map $s(\boldsymbol{c_k})$ (see Section 6.5).
3. Rearrange the views as subimages $\hat{S}_{\boldsymbol{k}}(\boldsymbol{q})$.
4. For each $\boldsymbol{k}$, filter $\hat{S}_{\boldsymbol{k}}(\boldsymbol{q})$ by $h_{f_0|\lambda|}$, using $\lambda = \frac{\mu}{ds(\boldsymbol{c_k})}$.
5. Repeat from 2 until the disparity map update is negligible.

6.4 Microlens Blur

With finite microlens apertures each pixel integrates over a larger area and aliasing is reduced due to additional blur (see Fig. 8). By taking this into account we can use milder antialiasing.

As the antialiasing filter for an array of pinhole lenses is a Sinc filter, we define the *antialiasing kernel size* as this filter's first zero crossing, i.e., $\frac{1}{2f_0|\lambda|}$. The correct amount of antialiasing is readily obtained by comparing this size with the blur radius b. Then, the final antialiasing filter has a radius approximated as $|\frac{1}{2f_0|\lambda|} - b|$, clipped from below at 0 and from above by $\frac{d}{2}$. Fig. 11 shows the resulting filter sizes for the settings used in Section 8.1.2.

6.5 Regularized Depth Estimation

We now have all the necessary ingredients to work on the energy introduced in (3). The depth map s is discretized at $\boldsymbol{c_k}$ as a vector $\boldsymbol{s} = \{s(\boldsymbol{u})\}_{\boldsymbol{u}\in\{\boldsymbol{c_k},\forall \boldsymbol{k}\}}$. Due to the ill-posedness of the problem, we introduce regularization, favoring piecewise constant solutions by using the total variation

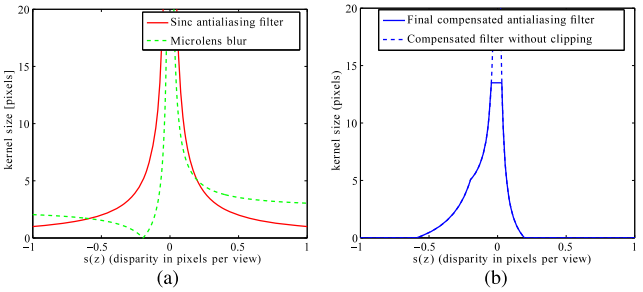

Fig. 11. Microlens blur and antialiasing filter sizes versus depth. (a) Overlap of filter kernel size and microlens blur radius for different disparity (depth) values. (b) Resulting antialiasing kernel size for different depth values.

term $\|\nabla s(\boldsymbol{u})\|_1$, where ∇ is the 2D gradient with respect to $\boldsymbol{u}$. Hence, we wish to solve

$$\tilde{\boldsymbol{s}} = \arg\min_{\boldsymbol{s}} E_{\text{data}}(\boldsymbol{s}) + \gamma\|\nabla s(\boldsymbol{u})\|_1, \tag{21}$$

where $\gamma > 0$ determines the tradeoff between regularization and data fidelity (in our experiments we chose $\gamma = 10^{-3}$). We minimize this energy by using an iterative solution. By noticing that E_{data} can be written as a sum of terms depending on a single entry of $\boldsymbol{s}$ at once, we find an initialization $\boldsymbol{s}_0$ by performing a fast brute force search in E_{data} for each $\boldsymbol{c_k}$ independently. Then, we approximate E_{data} via a second order Taylor expansion, i.e.,

$$\begin{aligned} E_{\text{data}}(\boldsymbol{s}_{t+1}) \simeq\ & E_{\text{data}}(\boldsymbol{s}_t) + \nabla E_{\text{data}}(\boldsymbol{s}_t)(\boldsymbol{s}_{t+1} - \boldsymbol{s}_t) \\ & + \frac{1}{2}(\boldsymbol{s}_{t+1} - \boldsymbol{s}_t)^T H_{E\text{data}}(\boldsymbol{s}_t)(\boldsymbol{s}_{t+1} - \boldsymbol{s}_t), \end{aligned} \tag{22}$$

where ∇E_{data} and $H_{E\text{data}}$ are the gradient and Hessian of E_{data}, and subscripts t and $t+1$ denote iteration number. To ensure our local approximation is convex we take the absolute value (component wise) of $H_{E\text{data}}(\boldsymbol{s}_t)$. In the case of the term $\|\nabla s(\boldsymbol{u})\|_1$, we use a first order Taylor expansion of its gradient. Computing the Euler-Lagrange equations of the approximate energy E with respect to $\boldsymbol{s}_{t+1}$ this linearization results in

$$\nabla E_{\text{data}}(\boldsymbol{s}_t) + |H_{E\text{data}}(\boldsymbol{s}_t)|(\boldsymbol{s}_{t+1} - \boldsymbol{s}_t) - \gamma\nabla \cdot \frac{\nabla(\boldsymbol{s}_{t+1} - \boldsymbol{s}_t)}{|\nabla \boldsymbol{s}_t|} = 0, \tag{23}$$

which is a linear system in the unknown $\boldsymbol{s}_{t+1}$, and can be efficiently solved using conjugate gradients (CG).

7 Light Field Superresolution

So far we devised an algorithm to reduce aliasing in views and estimate the depth map. We now define a computational PSF model, and formulate the MAP problem presented in Section 3.

7.1 Light Field Camera Point Spread Function

7.1.1 PSF Definition

Combining the analysis from Sections 4 and 5, we can determine the system PSF of the plenoptic camera h_s^{LI}—which is unique for each point in 3D space, and will be a combination of main lens and microlens array blurs. We define this PSF such that the intensity at a pixel $\boldsymbol{i}$ caused by a unit radiance point at $\boldsymbol{u}$ with a disparity $s(\boldsymbol{u})$ is

$$h_s^{LI}(\boldsymbol{i},\boldsymbol{u}) = h_{\boldsymbol{k}(\boldsymbol{i})}^{ML}(\boldsymbol{\theta}_{\boldsymbol{q}(\boldsymbol{i})},\boldsymbol{u}) h_{\boldsymbol{k}(\boldsymbol{i})}^{\mu L}(\boldsymbol{\theta}_{\boldsymbol{q}(\boldsymbol{i})},\boldsymbol{u}), \tag{24}$$

where $\boldsymbol{k}(\boldsymbol{i})$ and $\boldsymbol{q}(\boldsymbol{i})$ are given by (7). In a Lambertian scene, the image l captured by the light field camera is then

$$l(\boldsymbol{i}) = \int h_s^{LI}(\boldsymbol{i},\boldsymbol{u}) r(\boldsymbol{u}) d\boldsymbol{u}. \tag{25}$$

We define the microlens point spread function $h_{\boldsymbol{k}(\boldsymbol{i})}^{\mu L}$ considering the main lens diameter to be infinite. This gives

$$h_{\boldsymbol{k}(\boldsymbol{i})}^{\mu L}(\boldsymbol{\theta}_{\boldsymbol{q}(\boldsymbol{i})},\boldsymbol{u}) = \begin{cases} \dfrac{1}{\pi b^2(\boldsymbol{u})} & \left\|\boldsymbol{\theta}_{\boldsymbol{q}(\boldsymbol{i})} - \lambda(\boldsymbol{u})(\boldsymbol{c}_{\boldsymbol{k}(\boldsymbol{i})} - \boldsymbol{u})\right\|_2 < b(\boldsymbol{u}) \\ 0 & \text{otherwise.} \end{cases} \tag{26}$$

We define the main lens point spread function $h_{\boldsymbol{k}(\boldsymbol{i})}^{ML}$ assuming, instead, the microlens diameter is infinite. We obtain

$$h_{\boldsymbol{k}(\boldsymbol{i})}^{ML}(\boldsymbol{\theta}_{\boldsymbol{q}(\boldsymbol{i})},\boldsymbol{u}) = \begin{cases} \dfrac{d^2}{4\pi\beta^2}, & \left\|\boldsymbol{\theta}_{\boldsymbol{q}(\boldsymbol{i})} \pm \frac{2b(\boldsymbol{u})}{d}(\boldsymbol{c}_{\boldsymbol{k}(\boldsymbol{i})} - \boldsymbol{u})\right\|_2 < \frac{2\beta}{d} \\ 0, & \text{otherwise,} \end{cases} \tag{27}$$

where $\beta(\boldsymbol{u}) \doteq B(\boldsymbol{u}) b(\boldsymbol{u})$ and the sign of the $\boldsymbol{c}_{\boldsymbol{k}(\boldsymbol{i})} - \boldsymbol{u}$ term is positive when $v > \frac{z'f}{z'-f}$ and negative otherwise.

7.1.2 Main Lens Vignetting

As seen in Fig. 4 a microlens may only be partially hit by the main lens blur disc, which results in clipped microlens PSF; this effect is modeled by the product of $h_{\boldsymbol{k}(\boldsymbol{i})}^{\mu L}$ and $h_{\boldsymbol{k}(\boldsymbol{i})}^{ML}$. Also notice that depending on the camera settings and the distance of the object from the camera, the PSF under each microlens may be flipped.

7.1.3 Discretization

To arrive at a computational model, we discretize the spatial coordinates as $\boldsymbol{u}_n$ with $n \in \{1 \dots N\}$ and use the pixel coordinates $[i_m, j_m]^T$ with $m \in \{1 \dots M\}$. Then, (25) can be rewritten in matrix-vector form $\boldsymbol{l} = \boldsymbol{H}_s \boldsymbol{r}$ as in (1). $\boldsymbol{r}(n) \doteq r(\boldsymbol{u}_n)$ and $\boldsymbol{l}(m) \doteq l(i_m, j_m)$ are the discrete vectorized versions of l and r. $\boldsymbol{H}_s \in \mathbb{R}^{M \times N}$ is the sparse matrix

$$\boldsymbol{H}_s(m,n) \doteq h_{s(\boldsymbol{u}_n)}^{LI}(\boldsymbol{i}_m, \boldsymbol{u}_n), \tag{28}$$

where each column is the PSF corresponding to the upsampled depth map at that point (i.e., it is a nonstationary operator). Also, we are free to choose the spacing of the samples $\boldsymbol{u}_n$ as $\kappa\mu$; setting $\kappa = 1$ recovers the same resolution as the original sensor; however, depending on the camera settings, too high a resolution will not reveal additional detail. Hence, we choose κ based on our analysis of the limits of the LF camera as described in the previous sections. Note that, in particular, if the microlens diameter is reduced, more image detail will be visible (although less light efficient). Typically, we use $\kappa = 2$ to reduce computational load. The upsampling factor compared to the original views is Q/κ.

7.2 Bayesian Superresolution

We use the Bayesian framework to estimate $\boldsymbol{r}$, where all unknowns are treated as stochastic quantities. With additive Gaussian observation noise $\boldsymbol{w} \sim \mathcal{N}(\boldsymbol{0}, \sigma_w^2 \boldsymbol{I})$, the model becomes $\boldsymbol{l} = \boldsymbol{H}_s \boldsymbol{r} + \boldsymbol{w}$, and the probability of observing a given light field $\boldsymbol{l}$ in (1) may be written as $\mathrm{p}(\boldsymbol{l} \mid \boldsymbol{r}, \boldsymbol{H}_s, \sigma_w^2, s) = \mathcal{N}(\boldsymbol{l} \mid \boldsymbol{H}_s, \boldsymbol{r}, \sigma_w^2 \boldsymbol{I})$.

We then introduce priors on the unknowns (assuming $\boldsymbol{s}$ is already estimated). Many recent image restoration works make use of *nonstationary* edge preserving priors. For example, total variation or modeling heavy-tailed distributions of image gradients or wavelet subbands are popular [38]. We apply a recently developed Markov random field (MRF) prior [31], [32] which extends such ideas to modeling higher order neighborhoods. It uses a local autoregressive (AR) model, whose parameters are also inferred, and leads to a conditionally Gaussian prior $\mathrm{p}(\boldsymbol{r} \mid \boldsymbol{a}, \boldsymbol{\sigma}_v.) = \mathcal{N}(\boldsymbol{r} \mid \boldsymbol{0}, \mathbf{C}^{-1} \boldsymbol{Q}_v \mathbf{C}^{-T}.)$, where matrix $\boldsymbol{C}$ applies locally adaptive regularization using AR parameters $\boldsymbol{a}$; $\boldsymbol{Q}_v$ is a diagonal matrix of local variances $\boldsymbol{\sigma}_v$. We estimate these parameters using conjugate priors, which lets us set confidences on their likely values. The resulting Gaussian—inverse-gamma combination also represents inference with a heavy-tailed Student-t, considering the marginal distribution, corresponding to sparsity in the texture model.

The SR inference procedure therefore involves finding an estimate of the parameters $\boldsymbol{r}, \boldsymbol{a}, \boldsymbol{\sigma}_v, \sigma_w^2$ given the observations $\boldsymbol{l}$ and an estimate of $\boldsymbol{H}_s$. Direct maximization of the posterior $\mathrm{p}(\boldsymbol{r}, \boldsymbol{a}, \boldsymbol{\sigma}_v, \sigma_w \mid \boldsymbol{l}, \boldsymbol{H}_s) \propto \mathrm{p}(\boldsymbol{l} \mid \boldsymbol{H}_s) \mathrm{p}(\boldsymbol{r} \mid \boldsymbol{a}, \boldsymbol{\sigma}_v) \mathrm{p}(\boldsymbol{a}, \boldsymbol{\sigma}_v, \sigma_w^2)$ is intractable; hence, we use variational Bayes estimation with the mean field approximation to obtain an estimate of the parameters.

7.3 Numerical Implementation

The variational Bayesian procedure requires alternate updating of approximate distributions of each of the unknown variables. The approximate distribution for $\boldsymbol{r}$ is a Gaussian with

$$\mathbb{E}[\boldsymbol{r}] = \mathrm{cov}[\boldsymbol{r}] \sigma_w^{-2} \boldsymbol{H}_s^T \boldsymbol{l}, \tag{29}$$

$$\mathrm{cov}[\boldsymbol{r}]^{-1} = \mathbf{C}^T \boldsymbol{Q}_v^{-1} \mathbf{C} + \sigma_w^{-2} \boldsymbol{H}_s^T \boldsymbol{H}_s. \tag{30}$$

This update of $\mathbb{E}[\boldsymbol{r}]$ can be found solving $\boldsymbol{M}^T \boldsymbol{M} \boldsymbol{r} = \boldsymbol{M}^T \boldsymbol{y}$, with $\boldsymbol{Q}_v^{-1} = \boldsymbol{L}^T \boldsymbol{L}$, $\boldsymbol{M} = [\sigma_w^{-1} \boldsymbol{H}_s^T | \mathbf{C}^T \boldsymbol{L}]^T$, and $\boldsymbol{y} = [\sigma_w^{-1} \boldsymbol{l}^T | \boldsymbol{0}]^T$. However, due to the large size of this linear system, we solve efficiently using CG for each step. Each CG iteration requires multiplying by $\boldsymbol{H}_s$ and its transpose once, which we implement as a sparse matrix formed using a look-up table of precomputed PSFs from each point in 3D space. The image restoration procedure is run in parallel tasks across restored tiles of size up to around 400×400 pixels, which are trimmed to the valid region and then seamlessly mosaiced (size is limited such that the nonstationary $\boldsymbol{H}_s$ can be preloaded into memory). We run experiments in MATLAB on an 8-core Intel Xeon processor with 2 GB of memory per task. Precalculating the look-up table can take up to 10 minutes per depth plane (depending on PSF size); however, restoration is faster: Each CG iteration takes around 0.1-0.5 seconds (again depending on depth). Good convergence is achieved after typically 100 to 150 iterations. The AR model parameters are recomputed every few CG iterations, taking around 1-2 seconds. Notice that since the model is linear in the unknown $\boldsymbol{r}$, convergence is guaranteed

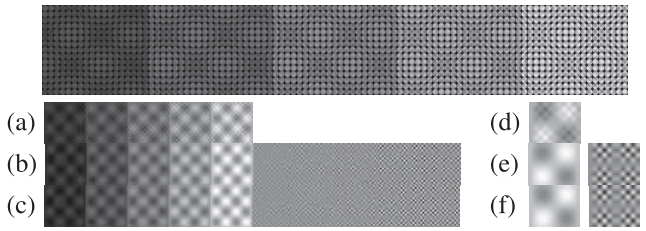

Fig. 12. Synthetic data: antialiasing filtering. Top: LF image of a sum-of-two-sinusoids texture (at five depth planes). Bottom: (a) One extracted view, containing aliasing. (b) The view filtered with the estimated depth (left) and the corresponding high-frequency image (right). (c) As in (b), but using the true depth; there is little difference between the two. (d)-(f) Enlarged portions of the last depth step in (a)-(c).

by the convexity of the cost functional. The total runtime per tile is typically 30-60 seconds, depending on depth.

8 Experiments

The antialiased 3D depth estimation method is tested on both synthetic and real data in Section 8.1. Then, the proposed SR method is tested similarly in Section 8.2. We also evaluate restoration performance in Section 8.2.2, comparing to other computational and regular camera systems.

8.1 Antialiased Depth Estimation

8.1.1 Synthetic Data

The scene in Fig. 12 consists of five steps at different depths, with a texture that is the sum of sinusoids at 0.2 and 1.2 times the views' Nyquist rate f_0, therefore the higher frequency is aliased in the views (but not in the subimages). The scene has disparities in the range $s = 0.24$ to 0.44 pixels per view. We simulate a camera with $Q = 15$, $\mu = 9.05 \times 10^{-6}$ m, $d = 0.135$ mm, $v = 0.5$ mm, $f = 0.5$ mm, $v' = 91.5$ mm, and $F = 80$ mm. Note that for these settings, the microlens blur is small but nonnegligible, varying between 1.2 and 2.1 pixels radius. We use the 9×9 pixel central region of each subimage for depth estimation. Fig. 12 shows one of the 81 extracted views, and the result of filtering with the estimated and true disparity maps, with the aliased component separated out. In Fig. 13, we compare the resulting disparity maps recovered with no antialiasing filtering, the iterative method, with the correct antialiasing filter (the errors at depth transitions occur due to lack of occlusion modeling in the synthesized data), and the ground truth.

We also test the algorithm's performance, repeating the experiment at 16 depths (for $s = 0.1$ to 0.6). Results in Table 1 show the average L_2 norm per pixel of the error between the ground truth and the disparity maps obtained with different filtering. We use the sinusoidal pattern and three textures taken from the Brodatz data set (http://sipi.usc.edu/database/database.cgi?volume=textures). Each texture is resized to 380×380 pixels, and tested again with the central 50 percent enlarged to this size to give a coarser scale. We test both the noise-free case and with additive Gaussian observation noise at two different levels. The textures contain different proportions of high and low frequencies; when more high frequencies are removed due to aliasing, matching performance decreases as less texture content remains to match.

TABLE 1
Average L_2 Norm Disparity Error (Pixels per View $\times 10^{-3}$), against Noise Levels, Filtering Method: No (A), Ideal (B), and Iterative Filtering (C), 4 textures and at 2 scales: Fine (1), Coarse (2)

Method (Noise)	Sin	Bark/1	Bark/2	Straw/1	Straw/2	Bubbles/1	Bubbles/2
A (0%)	2.656	2.340	1.822	1.927	2.256	2.455	1.025
B (0%)	0.568	1.346	0.954	1.383	1.152	1.129	0.666
C (0%)	0.662	1.456	1.032	1.512	1.273	1.230	0.689
A (1.25%)	2.798	2.543	2.008	2.555	2.530	2.678	1.209
B (1.25%)	0.807	1.468	1.033	1.669	1.317	1.295	0.751
C (1.25%)	0.811	1.556	1.123	1.838	1.421	1.385	0.775
A (2.5%)	3.1794	2.944	2.346	3.660	3.039	3.120	1.537
B (2.5%)	1.115	1.625	1.192	2.330	1.637	1.694	0.877
C (2.5%)	1.138	1.813	1.314	2.481	1.732	1.723	0.918

Row header is in the format (Texture/Scale).

8.1.2 Real Data

The use of small $f/10$ microlens apertures and a large microlens spacing (27 pixels per subimage) results in significant aliasing in the data from Georgiev and Lumsdaine [22] (kindly made available by Georgiev at http://www.tgeorgiev.net), hence it is a good test for the antialiasing algorithm. Other camera settings are as described in [6], [22]. Subimages and view details of this data are shown in Fig. 10, for different filters, to appreciate the effect of an incorrect size. In Fig. 14, we show the disparity maps obtained at different steps of the iterative algorithm (the first iteration is essentially a standard multiview stereo result), along with the final regularised result. Notice that there is a progressive reduction in the number of artifacts and that the disparity map becomes more and more accurate. The estimated

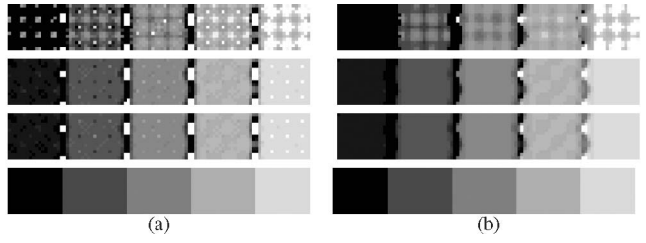

Fig. 13. Depth estimates. From top to bottom: Results obtained without filtering, with the iterative method, with the correct filtering, and the ground truth. Each row shows the disparity map (a) without and (b) with regularization. Notice how the results obtained with the estimated depth are extremely similar to those obtained with the correct filter.

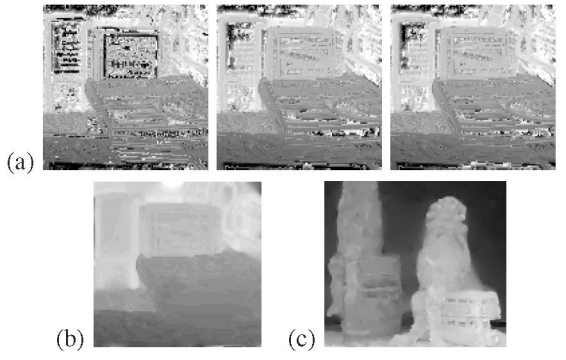

Fig. 14. Depth estimates on real data. (a) Disparity map estimate obtained without regularization (left) and for increasing filtering iterations (middle and right). (b) L_1 regularized disparity map, from the energy of the third iteration. (c) Regularized depth map for the puppets data set (Fig. 18).

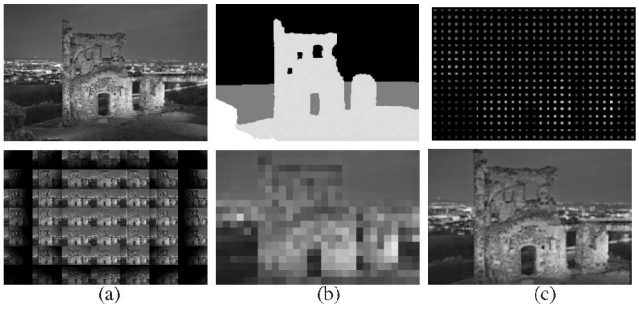

Fig. 15. Synthetic results. (a) Top: True radiance. (b) Top: True depth map. (c) Top: Light field image simulated with our model. (a) Bottom: LF image rearranged as views. (b) Bottom: Central view (as in a traditional rendering [3]). (c) Bottom: (SR) radiance restored with our method.

disparities lie in the range $s = -0.34$ to 0.31, with the middle book being around the main lens plane-in-focus (zero disparity). Regularized depth maps from other real data sets are also shown in Figs. 14 and 1.

8.2 Superresolution Results

8.2.1 Synthetic Data

In Fig. 15, we simulate LF camera data using (1) and a synthetic depth map, then apply the SR algorithm (using the known depth) to recover a high resolution focussed image. The simulated scene lies in the range 800-1,000 mm, each of the 49 views (i.e., we use only the 7×7 pixel central portion of each subimage in this experiment) is a 19×29 pixel image. The magnification gain is about seven times along each axis.

8.2.2 LF Camera/SR Performance Testing

First, we test how the proposed SR method compares to low-resolution integral refocusing [3] and the method of Lumsdaine and Georgiev [6]. We generate synthetic LF data using our model and the same settings as Section 8.1.1, with the "Bark" texture from the Brodatz database positioned on a sequence of 110 planes with depths in the range 486-1,074 mm. We omit planes near the main lens plane-in-focus where $\lambda < 1$, i.e., where space is undersampled and reconstruction better than the resolution in [3] is not possible at all points.[5] For each depth, and for a range of different additive noise levels σ_w, we restore the texture using our SR method using $\kappa = 1$ and an L_2 smoothness prior (i.e., fixing the matrix $\boldsymbol{C}$ as the discrete Laplacian), and also the method of Lumsdaine and Georgiev [6]. In the first plot of Fig. 16, we show the improvement in signal-to-noise ratio (ISNR) of these results, defined by $10\log(\frac{\|\boldsymbol{r}-\boldsymbol{r}_0\|}{\|\boldsymbol{r}-\hat{\boldsymbol{r}}\|})$, where $\boldsymbol{r}$ is the true texture, $\boldsymbol{r}_0$ an initial estimate given by the method in [3]. We see that our method performs well across depths, with a maximum when the magnification is close to 1, as predicted by our analysis, and ISNR decreasing for strong noise (20 dB) to a similar level as the method of Lumsdaine and Georgiev [6], which is resistant to noise due to averaging. Also observe there are certain depths where the performance of SR in the LF camera drops as predicted due to coinciding samples at rational disparity values (see Section 5.1.4).

5. A general limitation of LF cameras, and not our method. It can be solved by picking camera parameters to keep this region outside the working volume.

In the second experiment, we compare the simulated performance of DoF extension using the LF camera with both a traditional camera and a coded aperture camera system (for these systems we deconvolve images away from the plane-in-focus). Each camera has the same number of sensor pixels, $N_{\text{sensor}} = K_1 K_2 Q^2$, as the LF camera (the original image $\boldsymbol{r}$ has more pixels, even for $\kappa = 1$, to properly simulate boundary conditions in real devices). We used our SR algorithm with the L_2 smoothness prior to restore the images from the simulated LF camera, and deconvolution with the same prior and the known aperture PSF to restore the images in the other two cases. Noise variances are estimated automatically from the images. We normalize the main lens' open aperture area in each case so that the same amount of light reaches the sensor. For the coded aperture system, we used the $\sigma = 0.005$ optimized mask from Zhou and Nayar [39], although the mask from Veeraraghavan et al. [4] gave very similar results.

The second plot in Fig. 16 shows the L_2 norm of the error per pixel, $\frac{1}{N_{\text{sensor}}}\|\boldsymbol{r} - \hat{\boldsymbol{r}}\|$, for each system. Coded aperture (dashed lines) has a benefit over a regular camera (thin lines), mostly for lower noise levels, around the main lens plane in focus, and an LF camera (thick lines) outperforms both for depths away from this plane (again for depths where $\lambda < 1$ we cannot restore the image at all points). The horizontal dotted lines are the error from down- and re-upsampling the original $\boldsymbol{r}$ via bicubic interpolation, by the indicated factors (2×-32×). In the third experiment in Fig. 17, we quantify the image resolution with regards to coded aperture and focal sweep [13], via restoration of a simulated resolution chart at various depths. In coded aperture the aim is to obtain a high resolution depth map and image. Here, we only compare the image reconstruction when depth is known. Also, notice that in the focal sweep method the depth estimation step is not needed. Not only do we see an improvement over the method of Lumsdaine and Georgiev [6], but also that deblurring this result is not equivalent to restoration with the full model and that the LF camera reproduces detail such as the digits much more faithfully than the other methods. The top row in Fig. 17 shows that the reconstruction with the LF camera is less degraded than in other extended DoF systems. In coded aperture, texture content is lost away from the focal plane due to blur, and in focal sweep depth invariance of the PSF does not hold well for planes at the end of the sweep range. Also note that in Fig. 17d, the reconstruction improves for λ approaching 1 (i.e., at depth level 88, last row). Finally note that these results are for a circular main lens aperture; with a square aperture there are $\frac{4}{\pi}$ more pixels available, which will improve results further for the LF camera.

8.2.3 Real Data

We perform SR on data from both our camera prototype (Figs. 1 and 18) and from Georgiev and Lumsdaine [22] (Fig. 18). To capture our data, we built a portable LF camera similar to that in [3], using a Hasselblad H2 medium format camera with an 80 mm f/2.8 lens, and a Megavision E4 digital back. The 16 MP color CCD has $4{,}096 \times 4{,}096$ pixels, size $\mu = 9\ \mu\text{m}$. A custom-made adapter positions the

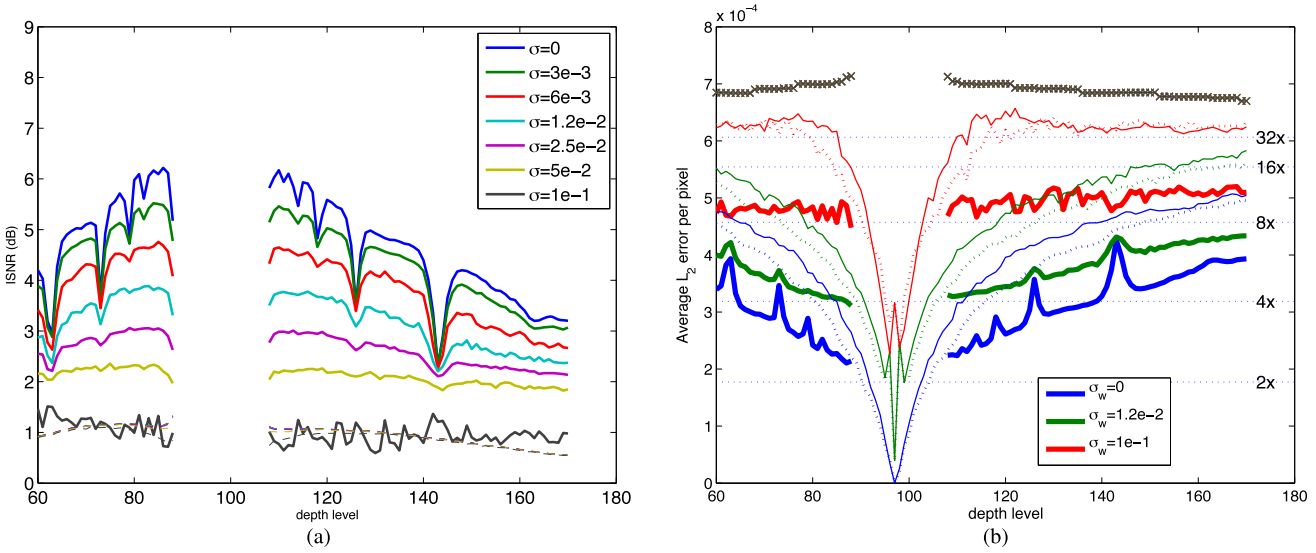

Fig. 16. L2 error results comparison. (a) Restoration performance versus depth of our method and the method of [6] on the simulated LF camera using our camera settings and the Brodatz "Bark" texture, with input intensity range 0-1. The (ISNR) is compared for several different levels of observation noise (standard deviation σ_w). Solid lines show our restoration method, dashed using the method of Lumsdaine and Georgiev [6] on the same data. We have not restored depths where $\lambda < 1$ (there are gaps in the restoration at these parts since some parts of these planes are not sampled at all). (b) Performance comparison of DoF extension between the LF camera (thick lines), a regular camera (thin lines), and a coded aperture camera (dotted lines). The crosses indicate the error from the upsampled integral refocusing result on the same LF data. See main text in Section 8.2.2 for further description.

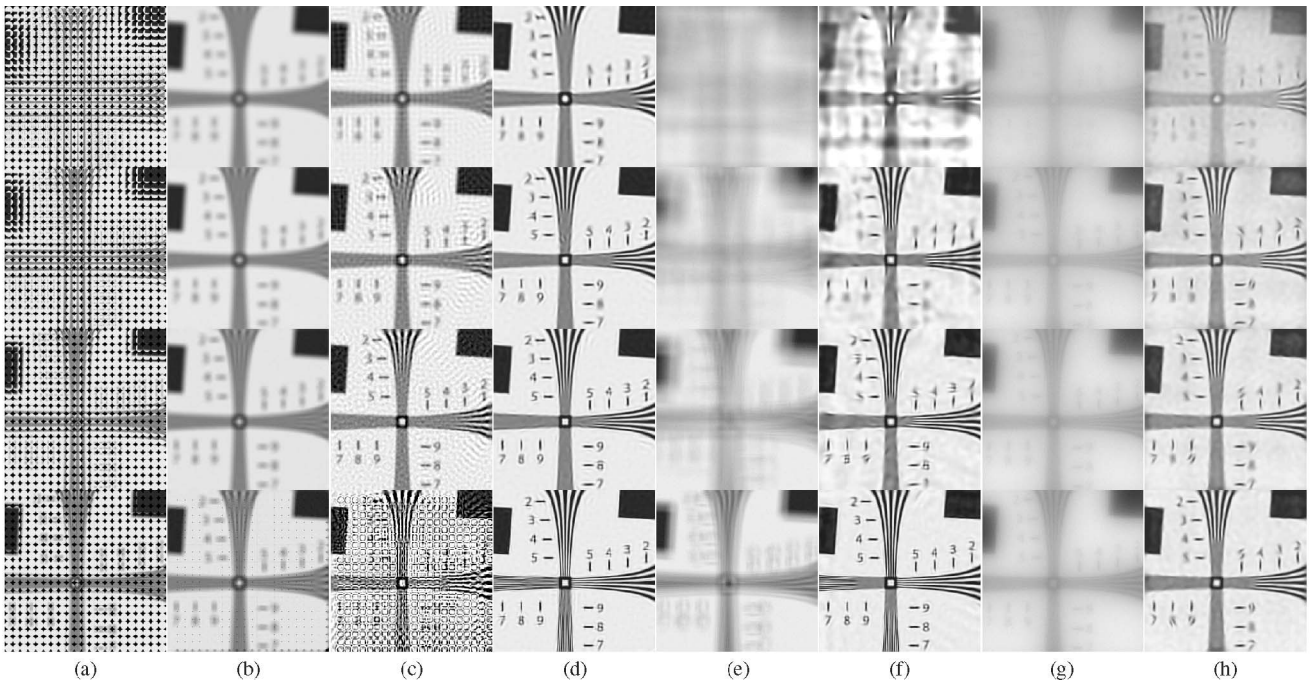

Fig. 17. Resolution tests. The experiment in Fig. 16 is repeated with the same camera settings using a resolution test chart and noise at $\sigma = 1.2 \times 10^{-2}$. Column (a) Simulated Light Field image; (b) method of [6] (used as initialization); (c) method of [6] deblurred (only for comparison); (d) input image restored with our method; (e) Simulated Coded Aperture image; (f) deconvolved CA image; (g) Simulated focal sweep image (across the whole depth range); (h) deconvolved focal sweep image, using middepth PSF. Rows, top to bottom: $\mathrm{depth} = 60, 72, 80, 88$. The plenoptic camera is seen to outperform the CA and focal sweep systems in terms of regularity and clarity of the solution away from the main-lens plane in focus when depth is known. Also, more detail is recovered via the full observation model than by deblurring the results in the second column.

microlenses near the sensor. The array has about 250×250 circular lensets, with $f \approx 0.35$ mm, $v \approx 0.4$ mm, and diameter $d = 135\ \mu$m, giving about 15×15 pixels per subimage.[6] Our microlenses have an $f/4$ aperture, though in Fig. 1 we used just the central 7×7 pixel subimages, as our initial array lacked a chromium mask in the gaps, the absence of which meant light leakage in the outer views made them unusable for SR. In the left of Fig. 18, we used a new set of microlenses with this mask, and we use the central 69 views within a 9 pixel diameter circle.

Both mechanical and software calibration of the system is essential. Our microlens adapter has external screws,

6. We scale down the recorded images by $\frac{v'+v}{v'}$ to give exactly $Q = 15$, and set μ in the model to $\frac{v'+v}{v'}$ times larger than its true value to match.

Fig. 18. Superresolution on real data Top row: Data from Georgiev and Lumsdaine [22] ($\kappa = 1.4$); Bottom row: "puppets" data set ($\kappa = 3$). (a) Nearest-neighbour interpolation of one view. (b) High-resolution view using method of Lumsdaine and Georgiev [6]. (c) View superresolved with our method.

enabling full 3D repositioning and rotation without removing the back. After manual correction, any residual error in the captured images is removed by automatic homography-based rectification (consistent to $\frac{1}{20}$ pixel across the sensor), and photometric calibration.

After calibration, we estimate the depth map (see Fig. 14), and use its upsampled version to construct $\boldsymbol{H}_s$. We compare with the restoration produced by the method in [6], modified to scale the subimages locally depending on the depth map. Clearly, we obtain sharper results with our data, due to deconvolution (although there are a few errors particularly around depth transitions, that may be due to depth map errors and lack of occlusion modeling). With the data in [22], there is less improvement from deconvolution since the microlens apertures (which sacrifice light and, hence, SNR) reduce the blur size, but the use of an integrated restoration model also suppresses artifacts, for example, the lines at the bottom of the image and in the background. The results in Fig. 18 consist of 6×5 tiles of 125×125 pixels in the left column and 16×19 tiles of 135×135 pixels in the right column.

We reiterate that our methods are general enough to work with any camera settings, though certain settings may suit better certain tasks. One important tradeoff is in the choice of the microlens aperture as done in [22]. Small microlens apertures require more light, but also allow the recoverery of an all-focused image at about the detectors resolution. On the other hand, large microlens apertures are more light efficient, but result in an effective recovered image with lower resolution.

9 Conclusions

We have presented a formal methodology for the restoration of high-resolution images from light field data captured from a LF camera, which is normally limited to returning images at the lower resolution of the number of microlenses in the camera. In our methodology, the 3D depth of the scene is first recovered by matching antialiased light field views, and then deconvolution is performed. This procedure makes the LF camera more useful for traditional photography applications; we have also shown the performance benefit of the LF camera for extended depth of field over other camera designs. In the future, we hope to use simultaneous depth estimation and superresolution, as well as extending the model to non-Lambertian and occluded scenes.

Acknowledgements

The authors wish to thank Mohammad Taghizadeh and the diffractive optics group at Heriot-Watt University for providing them with the microlens arrays and for stimulating discussions, and Mark Stewart for designing and building their microlens array interface. This work has been supported by EPSRC grant EP/F023073/1(P).

REFERENCES

[1] E.H. Adelson and J.Y. Wang, "Single Lens Stereo with a Plenoptic Camera," *IEEE Trans. Pattern Analysis and Machine Intelligence,* vol. 14, no. 2, pp. 99-106, Feb. 1992.

[2] T. Georgeiv and C. Intwala, "Light Field Camera Design for Integral View Photography," technical report, Adobe Systems, 2006.

[3] R. Ng, M. Levoy, M. Brédif, G. Duval, M. Horowitz, and P. Hanrahan, "Light Field Photography with a Hand-Held Plenoptic Camera," Technical Report CSTR 2005-02, Stanford Univ., Apr. 2005.

[4] A. Veeraraghavan, R. Raskar, A.K. Agrawal, A. Mohan, and J. Tumblin, "Dappled Photography: Mask Enhanced Cameras for Heterodyned Light Fields and Coded Aperture Refocusing," *ACM Trans. Graphics,* vol. 26, no. 3, p. 69, 2007.

[5] M. Levoy, R. Ng, A. Adams, M. Footer, and M. Horowitz, "Light Field Microscopy," *ACM Trans. Graphics,* vol. 25, no. 3, pp. 924-934, 2006.

[6] A. Lumsdaine and T. Georgiev, "The Focused Plenoptic Camera," *Proc. IEEE Int'l Conf. Computational Photography,* Apr. 2009.

[7] A. Levin, W.T. Freeman, and F. Durand, "Understanding Camera Trade-Offs through a Bayesian Analysis of Light Field Projections," *Proc. European Conf. Computer Vision,* pp. 619-624, 2008.

[8] G. Lippmann, "Epreuves Reversibles Donnant la Sensation du Relief," *J. Physics,* vol. 7, no. 4, pp. 821-825, 1908.

[9] K. Fife, A. El Gamal, and H.-S. Wong, "A 3D Multi-Aperture Image Sensor Architecture," *Proc. IEEE Custom Integrated Circuits Conf.*, pp. 281-284, 2006.

[10] C.-K. Liang, G. Liu, and H.H. Chen, "Light Field Acquisition Using Programmable Aperture Camera," *Proc. IEEE Int'l Conf. Image Processing,* pp. V233-236, 2007.

[11] M. Ben-Ezra, A. Zomet, and S. Nayar, "Jitter Camera: High Resolution Video from a Low Resolution Detector," *Proc. IEEE CS Conf. Computer Vision and Pattern Recognition,* vol. 2, 2004.

[12] W. Cathey and E. Dowski, "New Paradigm for Imaging Systems," *Applied Optics,* vol. 41, no. 29, pp. 6080-6092, 2002.

[13] H. Nagahara, S. Kuthirummal, C. Zhou, and S.K. Nayar, "Flexible Depth of Field Photography," *Proc. 10th European Conf. Computer Vision,* Oct. 2008.

[14] S.C. Park, M.K. Park, and M.G. Kang, "Super-Resolution Image Reconstruction: A Technical Overview," *IEEE Signal Processing Magazine,* vol. 20, no. 3, pp. 21-36, May 2003.

[15] A. Katsaggelos, R. Molina, and J. Mateos, *Super Resolution of Images and Video.* Morgan & Claypool, 2007.

[16] S. Borman and R. Stevenson, "Super-Resolution from Image Sequences—A Review," *Proc. Midwest Symp. Circuits and Systems,* pp. 374-378, 1999.

[17] M.K. Ng and A.C. Yau, "Super-Resolution Image Restoration from Blurred Low-Resolution Images," *J. Math. Imaging and Vision,* vol. 23, pp. 367-378, 2005.

[18] S. Farsiu, D. Robinson, M. Elad, and P. Milanfar, "Advances and Challenges in Super-Resolution," *Int'l J. Imaging Systems and Technology,* vol. 14, pp. 47-57, 2004.

[19] B.R. Hunt, "Super-Resolution of Images: Algorithms, Principles, Performance," *Int'l J. Imaging Systems and Technology,* vol. 6, no. 4, pp. 297-304, 2005.

[20] W.-S. Chan, E. Lam, M. Ng, and G. Mak, "Super-Resolution Reconstruction in a Computational Compound-Eye Imaging Systems," *Multidimensional Systems and Signal Processing,* vol. 18, no. 2, pp. 83-101, Sept. 2007.

[21] A. Lumsdaine and T. Georgiev, "Full Resolution Lightfield Rendering," technical report, Indiana Univ. and Adobe Systems, 2008.

[22] T. Georgiev and A. Lumsdaine, "Depth of Field in Plenoptic Cameras," *Proc. Eurographics 2009,* 2009.

[23] M. Levoy and P. Hanrahan, "Light Field Rendering," *Proc. ACM Siggraph,* pp. 31-42, 1996.

[24] J. Stewart, J. Yu, S.J. Gortler, and L. McMillan, "A New Reconstruction Filter for Undersampled Light Fields," *Proc. 14th Eurographics Workshop Rendering,* pp. 150-156, 2003.

[25] J.-X. Chai, S.-C. Chan, H.-Y. Shum, and X. Tong, "Plenoptic Sampling," *Proc. ACM Siggraph,* pp. 307-318, 2000.

[26] A. Isaksen, L. McMillan, and S.J. Gortler, "Dynamically Reparameterized Light Fields," *Proc. ACM Siggraph,* pp. 297-306, 2000.

[27] R. Ng, "Fourier Slice Photography," *Proc. ACM Siggraph,* vol. 24, no. 3, pp. 735-744, 2005.

[28] V. Vaish, M. Levoy, R. Szeliski, C. Zitnick, and S.B. Kang, "Reconstructing Occluded Surfaces Using Synthetic Apertures: Stereo, Focus and Robust Measures," *Proc. 26th IEEE CS Conf. Computer Vision and Pattern Recognition,* vol. 2, pp. 2331-2338, 2006.

[29] T.E. Bishop, S. Zanetti, and P. Favaro, "Light Field Superresolution," *Proc. IEEE Int'l Conf. Computational Photograph,* Apr. 2009.

[30] T.E. Bishop and P. Favaro, "Plenoptic Depth Estimation from Multiple Aliased Views," *Proc. 12th IEEE Int'l Conf. Conf. Computer Vision Workshops,* 2009.

[31] T.E. Bishop, R. Molina, and J.R. Hopgood, "Blind Restoration of Blurred Photographs via AR Modelling and MCMC," *Proc. IEEE 15th Int'l Conf. Image Processing,* 2008.

[32] T.E. Bishop, "Blind Image Deconvolution: Nonstationary Bayesian Approaches to Restoring Blurred Photos," PhD dissertation, Univ. of Edinburgh, 2008.

[33] M. Born and E. Wolf, *Principles of Optics.* Pergamon, 1986.

[34] Z. Wang and F. Qi, "Analysis of Multiframe Super-Resolution Reconstruction for Image Anti-Aliasing and Deblurring," *Image and Vision Computing,* vol. 23, no. 4, pp. 393-404, Apr. 2005.

[35] D. Robinson and P. Milanfar, "Fundamental Performance Limits in Image Registration," *IEEE Trans. Image Processing,* vol. 13, no. 9, pp. 1185-1199, Sept. 2004.

[36] S. Baker and T. Kanade, "Limits on Super-Resolution and How to Break Them," *IEEE Trans. Pattern Analysis and Machine Intelligence,* vol. 24, no. 9, pp. 1167-1183, Sept, 2002.

[37] A. Papoulis, "Generalized Sampling Expansion," *IEEE Trans. Circuits and Systems,* vol. 24, no. 11, pp. 652-654, Nov. 1977.

[38] E.P. Simoncelli, "Statistical Modeling of Photographic Images," *Handbook of Image and Video Processing,* A. Bovik, ed., second ed., Academic Press, Jan. 2005.

[39] C. Zhou and S. Nayar, "What Are Good Apertures for Defocus Deblurring?" *Proc. IEEE Int'l Conf. Computational Photography,* 2009.

Tom E. Bishop received the MEng degree in electronic and information engineering from the University of Cambridge (Pembroke College) in 2004, and the PhD degree in engineering from the University of Edinburgh in 2009. From 2008 until 2011, he was a postdoctoral researcher at Heriot Watt University. Since March 2011, he has worked as a researcher at Anthropics Technology Ltd., London. His research interests include computational photography, inverse problems, image modeling, and blind deconvolution. He is a member of the IEEE.

Paolo Favaro received the DIng degree from the Università di Padova, Italy in 1999, and the MSc and PhD degrees in electrical engineering from Washington University, St. Louis, Missouri, in 2002 and 2003, respectively. He was a postdoctoral researcher in the Computer Science Department of the University of California, Los Angeles and subsequently at the University of Cambridge, United Kingdom. He is now assistant professor in the Joint Research Institute for Signal and Image Processing between the University of Edinburgh and Heriot-Watt University, United Kingdom. His research interests are in computer vision, computational photography, inverse problems, convex optimization methods, and variational techniques. He is a member of the IEEE.