Light Field Superresolution

Tom E. Bishop   Sara Zanetti   Paolo Favaro

Department of Engineering and Physical Sciences
Heriot-Watt University, Edinburgh, UK

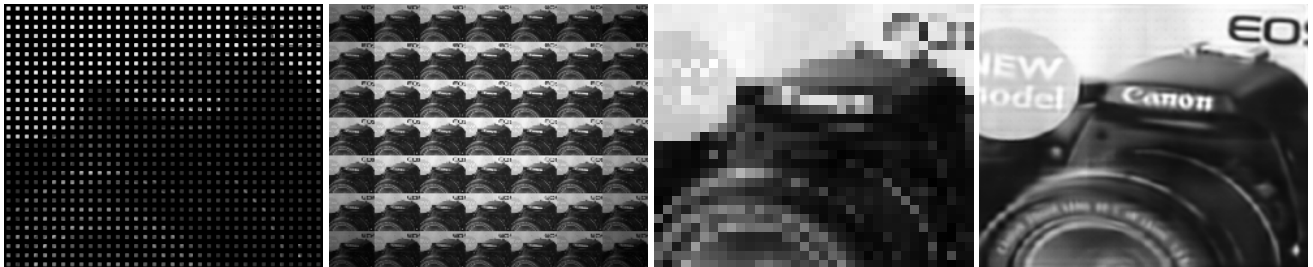{t.e.bishop,sz73,paolo.favaro}@hw.ac.uk

Figure 1. From left to right: Light field image captured with a plenoptic camera (detail); the light field image on the left is rearranged as a collection of several views; central view extracted from the light field, with one pixel per microlens, as in a traditional rendering [23]; central view superresolved with our method.

Abstract

Light field cameras have been recently shown to be very effective in applications such as digital refocusing and 3D reconstruction. In a single snapshot these cameras provide a sample of the light field of a scene by trading off spatial resolution with angular resolution. Current methods produce images at a resolution that is much lower than that of traditional imaging devices. However, by explicitly modeling the image formation process and incorporating priors such as Lambertianity and texture statistics, these types of images can be reconstructed at a higher resolution. We formulate this method in a variational Bayesian framework and perform the reconstruction of both the surface of the scene and the (superresolved) light field. The method is demonstrated on both synthetic and real images captured with our light-field camera prototype.

1. Introduction

Recently, we have seen that not only it is possible to build practical integral imaging and mask enhanced systems based on commercial cameras [1, 12, 23, 28], but also that such cameras provide an advantage over traditional imaging systems by enabling, for instance, *digital refocusing* [23] and the recovery of transparent objects in microscopy [18] from a single snapshot.

The performance of such systems, however, has been limited by the resolution of the camera sensor and of the
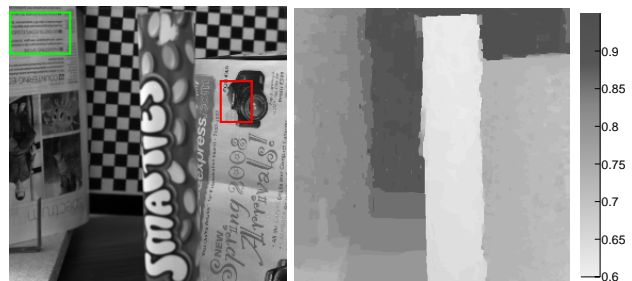

Figure 2. Left: One view captured by our plenoptic camera. In Figures 1 and 7 we restore the red and green highlighted regions. Right: The estimated depth map in meters.

microlens array. These define, according to the sampling theorem, the tradeoff between spatial and angular resolution of the recovered light field [23, 17]. Furthermore, due to diffraction, the image resolution of the system is restricted by the size of the microlenses [13].

Instead of increasing pixel density, we enhance detail by designing *superresolution (SR)* algorithms which extract additional information from the available data (see Figure 1). More specifically, we exploit the fact that light fields of natural scenes are not a collection of random signals. Rather, they generally satisfy models of limited complexity [17]. A general way to describe the properties of such light fields is via the bidirectional reflectance distribution function (BRDF), *e.g.*, Ward's model [29]. We are interested in exploring different BRDF models of increasing or-

der of complexity. In this paper we focus on the Lambertian model, which is the simplest instance.

Our main contribution is two-fold: First, we provide an image formation model by characterizing the point-spread function (PSF) of a plenoptic camera under Gaussian optics assumptions for a depth varying scene; second, we formulate the reconstruction of the light field in a Bayesian framework by explicitly introducing Lambertian reflectance priors in the image formation model. The Bayesian formulation allows us to design a SR algorithm which recovers more information than the one predicted by the basic sampling theorem. In particular, we show that, in the Lambertian case, the captured light field is equivalent to capturing several low resolution images with unknown optic flow. We formulate the problem of recovering the light field as an optimization problem where we first recover a depth map of the scene and then superresolve the light field in a variational Bayesian framework.

1.1. Prior Work and Contributions

This work relates to *computational photography*, a new emerging field encompassing several methods to enhance the capabilities and overcome limitations of standard digital photography by jointly designing an imaging device and a reconstruction algorithm.

One of the first devices based on the principles of integral photography [20] is the *plenoptic camera*, first proposed in computer vision by Adelson and Wang [1] to infer depth from a single snapshot and then more recently engineered into a single package chip [11]. In its original design, the plenoptic camera consists of a camera body with a single main lens and a lenticular array replacing the conventional camera sensor, as well as an additional relay lens to form the image on the sensor. Ng *et al.* [23] present a similar design, but produced in a portable hand-held device. They propose digital refocusing, *i.e.*, the ability to change the focus setting after the image has been taken. While their method yields impressive results, there is one caveat: The refocused images possess a spatial resolution that is lower than that of the image sensor, and equivalent to just the number of microlenses in the camera — *e.g.*, as low as 60K pixels from a 16MP camera.

An alternative to the plenoptic camera is the *programmable aperture* camera [19]. This device captures light field data by exploiting multiplexing of views of the scene. While this approach allows to exploit the full resolution of the camera sensor, the price to pay is a long exposure time or a low signal to noise ratio. Another interesting design proposed by Veeraraghavan *et al.* [28] is that of the heterodyne camera, where the light field is modulated using an attenuating mask close to the sensor plane. While the authors mention that the advantage of this system is the reconstruction of high resolution images at the plane in fo-

cus in addition to the sampled light field, there is a considerable limitation: The mask blocks much of the light that could reach the sensor and thus reduces the signal-to-noise ratio (SNR). Finally, Georgiev and Intwala [12] suggest a variety of different camera designs to capture the light field. Instead of internal microlens arrays, they use additional external optical elements, such as multiple prisms or an array of positive/negative lenses placed in front of the main lens. Unfortunately whilst appealing in their simplicity, these designs tend to suffer from higher order optical aberrations. Ben-Ezra *et al.* [2] propose a novel camera design to enhance the resolution of images where a randomly moving sensor collects multiple frames from slightly different positions and is synchronized to be motion-less during image capture so as to avoid motion-blur. The multiple frames are then combined to reconstruct a single high resolution frame. As in [19], this method needs to trade off exposure time for spatial resolution.

In contrast to the above approaches, one could aim at improving the resolution of the measured light field by designing algorithms, rather than hardware, that exploit prior knowledge about the scene. Stewart *et al.* [26] propose a method for recovering light fields for the purpose of rendering, based on combining the band-limited reconstruction in [7] with the wide-aperture reconstruction of [15]. These methods, however, do not consider using the depth map in the reconstruction of the light field. However, the recovery of a high-resolution image that is focused everywhere requires knowing the depth map (for instance, see the reconstructions obtained by moving the focal plane in Figure 7c and 7e in [26]).

We rely on the reconstruction of the depth map and pose the problem as that of *superresolving* the light field by starting from multiple low resolution images with unknown translational misalignment. This approach relates to a large bulk of literature in image processing [24, 16, 5, 22, 9, 10, 14]. While this problem has been extensively investigated in image processing, prior work in the context of computational photography is limited to work by Chan *et al.* [8], where a compound eye system is only simulated, to work by Lumsdaine and Georgiev [21], that propose a method to superresolve images captured with a plenoptic camera, and to work by Levin *et al.* [17], that describe trade-offs between different camera designs in recovering the light field of a scene.

Lumsdaine and Georgiev detect whether subimages under each microlens are flipped (telescopic) or not (binocular), and then scale up their central part by assuming that the scene is an equifocal plane at a user-defined depth. Their approach does not fully address SR of a light field. First, they do not reconstruct the depth map of the scene, which corresponds to finding the alignment between the subimages. Second, they do not use a deconvolution method to re-

store the light field, but only interpolation. This means that overlapping pixels in the subimages are dropped instead of being fused. Moreover, their results are not performed under a globally consistent restoration model, and there is no regularization in their algorithm.

In concurrent work, Levin *et al.* [17] describe analysis and algorithms that are closely related to our method. They focus on the trade-offs in recovering the light field 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 image formation model of a plenoptic camera and verify its validity on real images; second, we explicitly enforce Lambertianity and make use of image texture priors that are unlike their mixture of Gaussians derivative priors.

In the case of light field images obtained from the plenoptic camera, the unknown translational misalignment between the image views is due to the unknown depth map of the scene. The estimation of such depth map is therefore a fundamental step in our SR algorithm. Vaish *et al.* [27] perform multiview depth estimation from an array of about a hundred cameras, a system that is structurally similar to a plenoptic camera. Their method addresses the rejection of outliers by employing robust multiview matching, a strategy that we also employ in our depth estimation method.

2. Image Formation of a Light-Field Camera

In this section we derive the image formation model of a plenoptic camera and then analyze under what conditions SR can be best addressed. To arrive at a practical computational model suitable for our algorithm (section 3), we investigate the imaging process with tools from geometrical optics [6]. Our basic approximation is to ignore effects due to diffraction and use the thin lens model for both the main lens and each lens in the microlens array. We start by defining the basic parameters in our camera and establish their relationship in the image formation model (section 2.1). We summarize the model by characterizing the light field camera PSF (section 2.2). Then, we analyze such model and study its behavior under different modalities of operation (section 2.3).

2.1. Imaging Model

In our investigation we rebuilt the same light field camera that was used by Ng *et al.* [23], but did not restrict our analysis to the same camera parameters. Functionally, the light field camera is approximately equivalent to a camera with two types of optical elements: a main lens and a microlens array. As in [21], we consider the imaging system under a general configuration of these optical elements; however, unlike in any previous work, we determine the image formation model of the camera so that it can be used for
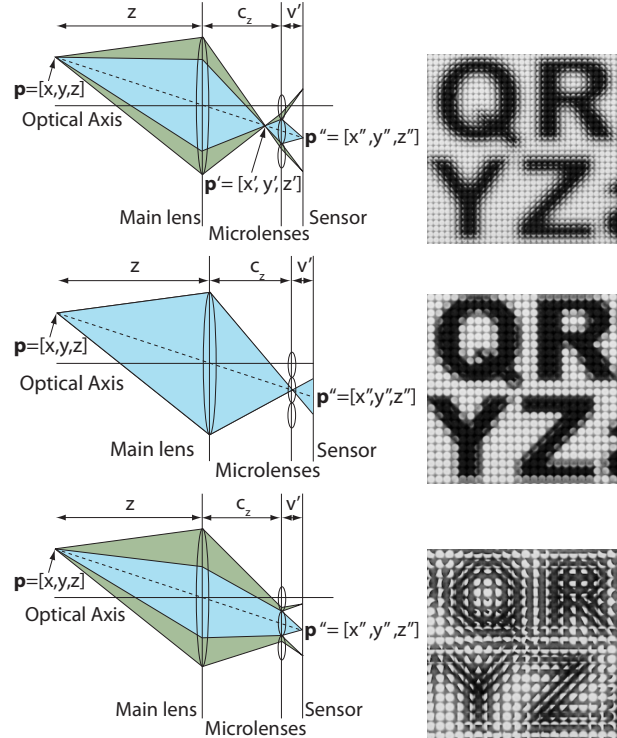

Figure 3. Schematic of a 2D section of a light field camera. By starting from the left in each diagram, the plenoptic camera consists of a main lens, a microlens array, and a sensor. The light emitted by a point in space **p** is deflected by the main lens and then split into several beams by the microlens array. The size of some microlenses has been exaggerated only for visualization purposes. Top row: Case of multiple repetitions with no flipping (main lens out of focus). Middle row: Case with no repetitions (main lens in focus). Bottom row: Case of multiple repetitions with flipping (main lens out of focus).

SR or more general tasks. Due to symmetry and for simplicity, we visualize our analysis of the model in 2D sections as shown in Figure 3.

We summarize all the symbols and their meaning in Table 1. As shown in Figure 3, the image of a 3D point in space $\mathbf{p} \doteq [x\ y\ z]^T \in \mathbb{R}^3$ results in a collection of blur discs whose shape depends on three factors: the blur introduced by the main lens, the masking due to the microlens array, and the blur introduced by the microlenses. We will see all of these effects in the following analysis and, in particular, in section 2.2. By applying the thin lens law[1] to the main lens, we can find that the point **p** is brought into focus inside the camera at the position $\mathbf{p}' \doteq [x'\ y'\ z']^T \in \mathbb{R}^3$

---

[1]The thin lens law establishes that a point in space at a distance z from the lens is imaged in focus at a distance $\dfrac{Fz}{z-F}$ from the lens (inside the camera) where F is the lens focal length [6].

Table 1. Light field camera symbols and their description.

Camera parameters	
D	Main lens diameter
d	Microlens diameter
F	Main lens focal length
f	Microlens focal length
$\mathbf{c}$	Microlens center in 3D space
v'	Microlenses to CCD sensor distance
μ	Size of a CCD sensor element
Scene parameters	
$\mathbf{p}$	3D point in space
$\mathbf{p}'$	Focused image of $\mathbf{p}$ inside the camera
$\mathbf{p}''$	Projection of $\mathbf{p}$ onto the CCD sensor
Point spread function parameters	
$\mathbf{p}_b$	Main lens blur center in 3D space
B	Main lens blur radius
b	Microlens blur radius

where

$$\mathbf{p}' = \frac{F}{z-F} \begin{bmatrix} -1 & 0 & 0 \\ 0 & -1 & 0 \\ 0 & 0 & 1 \end{bmatrix} \mathbf{p}. \tag{1}$$

By using similar triangles, we can easily find that the projection $\mathbf{p}_b$ of $\mathbf{p}'$ onto the microlens array plane is

$$\mathbf{p}_b = \mathbf{p}' \frac{z-F}{F} \frac{c_z}{z}. \tag{2}$$

The projection $\mathbf{p}_b$ is the center of the blur generated by the main lens blur. If we approximate the blur generated by the main lens with a Pillbox function[2] the main lens blur radius B is

$$B = \frac{Dc_z}{2} \left| \frac{1}{F} - \frac{1}{z} - \frac{1}{c_z} \right|. \tag{3}$$

Finally, the projection of $\mathbf{p}$ onto the sensor plane through a microlens centered in $\mathbf{c} = [c_x \ c_y \ c_z]^T \in \mathbb{R}^3$ is instead computed as

$$\mathbf{p}'' = \mathbf{c} + \frac{v'}{c_z - z'} \left(\mathbf{p}\frac{z'}{z} + \mathbf{c} \right) \tag{4}$$

and this microlens generates a small blur disc with radius

$$b = \frac{dv'}{2} \left| \frac{1}{f} - \frac{1}{c_z - z'} - \frac{1}{v'} \right|. \tag{5}$$

We are now ready to analyze two important effects introduced by the use of a microlens array. First, the main lens may cause a vignetting effect on the microlenses. Second, each microlens might flip the image of an object in the scene, depending on the camera parameters and the position of the object in space.

---

[2]The Pillbox function is defined as the unit area cylinder with base the disc generated by the aperture.

2.1.1 Main Lens Vignetting

As one can observe in Figure 3 a microlens may only be partially hit by the blur disc cast by the main lens. This will then affect the shape of the blur disc generated by the microlens on the sensor plane/Furthermore, because each microlens has a finite aperture, the main lens blur disc will be masked by discs of the size of each microlens. Due to the Pillbox model for the main lens, microlenses that are completely or partially hit by light emitted from a point $\mathbf{p}$ satisfy

$$\|\mathbf{p}_b - \mathbf{c}\| < B + \frac{d}{2}. \tag{6}$$

2.1.2 Image Flipping

The pattern generated by each microlens may not only vary in position and blur, but it might also appear flipped along both the abscissa and ordinate axes (third row of Figure 3). Flipping can be easily characterized as follows. Let us consider a point moving in space of Δ along either the abscissa or the ordinate axis. If this movement generates a shift Δ'' on the sensor plane in the same direction (*i.e.*, with the same sign) then there is no flipping. In formulas, we have that

$$\Delta'' = \Delta \frac{v'z'}{z(c_z - z')} \tag{7}$$

and, therefore, there is no flipping when

$$c_z - z' > 0 \tag{8}$$

if we assume that $z' > 0$ (*i.e.*, when objects in space are at a distance from the camera of at least the main lens focal length F). This scenario is shown on both the first and second row of Figure 3. If instead we have $c_z < z'$ then there is flipping (third row of Figure 3).

Remark 1 *Notice that the subimage flipping that we have analyzed in this section does **not** correspond to the flipping of the blur generated by a single point in space through a single microlens (that instead occurs when $v' > \frac{(c_z - z')f}{c_z - z' - f}$). The microlens blur inversion is usually insignificant because the PSF is usually symmetric.*

2.2. Light Field Camera Point Spread Function

If we combine the analysis carried out in the previous sections we can determine the PSF of the light field camera, which will be a combination of the blur generated by the main lens and the blur generated by the microlens array. In our notation, we define the PSF of the light field camera as a function h_{LI} such that the intensity at a pixel[3] (i, j) caused

---

[3]For simplicity, we assume that the pixel coordinates have their zero in the center of the sensor plane, which we assume to coincide with the optical axis. Also, i follows the abscissa axis and j follows the ordinate axis.

by a unit radiance point $\mathbf{p}$ in front of the camera

$$h_{LI}(\mathbf{p}, i, j) = h_{ML}(\mathbf{p}, i, j) h_{\mu L}(\mathbf{p}, i, j) \quad (9)$$

so that in a Lambertian scene the image l captured by the light field camera is

$$l(i, j) = \int h_{LI}(\mathbf{p}, i, j) r(\mathbf{p}) d\mathbf{p} \quad (10)$$

where r is the light field defined at each point in space.[4] h_{ML} is the main lens PSF and it is defined as

$$h_{ML}(\mathbf{p}, i, j) = \begin{cases} \frac{1}{\pi B^2} & \|\mathbf{p}_s - \mathbf{p}_b\| < B \\ 0 & \text{otherwise} \end{cases} \quad (11)$$

where

$$\mathbf{p}_s \doteq \left(\begin{bmatrix} i\mu \\ j\mu \\ v'z' \end{bmatrix} - \mathbf{p} \frac{v'z'}{z} \right) \frac{c_z - z'}{c_z - z' + v'} \quad (12)$$

if there is no microlens blur inversion[5] (*i.e.*, $v' < \frac{(c_z - z')f}{C_z - z' - f}$); otherwise if the microlens blur is inverted, the inverted coordinates $\hat{\mathbf{p}}_s$ can be obtained directly from the previous ones via

$$\hat{\mathbf{p}}_s \doteq d \left\lfloor \frac{\mathbf{p}_s}{d} + \frac{1}{2} \right\rfloor - \left(\mathbf{p}_s + \frac{d}{2} \right) \% \, d + \frac{d}{2} \quad (13)$$

where $\lfloor . \rfloor$ denotes the closest lower integer and $a \% b$ denotes a modulo b. Finally, the microlens array PSF $h_{\mu L}$ is defined as

$$h_{\mu L}(\mathbf{p}, u, v) = \begin{cases} \frac{4}{\pi b^2} & \left\| (\mathbf{p}_s + \frac{d}{2}) \% \, d - \frac{d}{2} \right\| < b \\ 0 & \text{otherwise.} \end{cases} \quad (14)$$

To arrive at a computational model, we discretize the spatial coordinates as $p_n = [x_n \; y_n \; z_n]^T$ with $n = [1, N]$ and order the pixel coordinates as $[i_m, j_m]$ with $m = [1, M]$ so that eq. (10) can be rewritten in matrix-vector notation as

$$\mathbf{l} = \mathbf{H}\mathbf{r} \quad (15)$$

where $\mathbf{l} \in \mathbb{R}^M$ is the captured image rearranged as a column vector, $\mathbf{r} \in \mathbb{R}^N$ is the unknown reflectance of the discretized volume also rearranged as a column vector, and $\mathbf{H} \in \mathbb{R}^{M \times N}$ is a (sparse) matrix representing the PSF of the light field camera. As we assume that the scene can be described by a single discretized depth map $s : \mathbb{R}^2 \mapsto [0, \infty)$, we define

$$\begin{aligned} \mathbf{r}(n) &\doteq r([x_n \; y_n \; s(x_n, y_n)]^T) \\ \mathbf{l}(m) &\doteq l(i_m, j_m) \\ \mathbf{H}(m, n) &\doteq h_{LI}([x_n \; y_n \; s(x_n, y_n)]^T, i_m, j_m). \end{aligned} \quad (16)$$

---

[4]Because illumination is constant and we assume that the scene is Lambertian (within the field of view of the camera) the light field does not depend on the emitting direction and can be represented by a function of a 3D point in space

[5]Notice that, as mentioned in Remark 1, the microlens blur inversion is *not* the flipping of a subimage.

2.3. Analysis of the Imaging Model

Although the PSF of the light field camera obtained in the previous sections fully characterizes how a light field is imaged on the sensor, it also does not provide an intuitive tool to analyze the imaging process. In this section we will see that more insight can be gained by isolating defocusing due to the main lens from defocusing due to the microlenses.

2.3.1 Main Lens Defocus

The blur disc generated by a point in space $\mathbf{p}$ onto the microlens array determines the *number* of microlenses that capture light from $\mathbf{p}$. Under the Lambertian assumption, $\mathbf{p}$ casts the same light on each microlens, and this results in multiple copies of $\mathbf{p}$ in the light field (see first and third row of Figure 3).

To characterize the number of repetitions of the same pattern in the scene, or, equivalently, the number of microlenses that simultaneously image the same points in space, we need to count how many microlenses fall inside the main lens blur disc, *i.e.*, we need to pick the ratio between the main lens blur diameter $2B$ and the microlens diameter d

$$\#\text{repetitions} = \frac{2B}{d} = \frac{Dc_z}{d} \left| \frac{1}{F} - \frac{1}{z} - \frac{1}{c_z} \right|. \quad (17)$$

In our SR framework, the number of repeating patterns is extremely important as it determines the number of subsampled images that we can use to superresolve the light field. It is also immediate to conclude that objects that are brought into focus by the main lens (*i.e.*, with a single repetition such as on the second row in Figure 3) will have the least accuracy in the reconstruction process, and vice versa in the case of objects that are brought out of focus by the main lens.

2.3.2 Focused Subimages

A necessary condition to superresolve the light field is that the input views are subject to aliasing so that they carry different information about the scene, i.e., they are not merely shifted and interpolated versions of the same image. To satisfy this condition, we need to work away from the plane in the scene that the main lens brings in focus on the sensor plane. Also, we need the microlens blur to be as small as possible, otherwise pixels from different views blend together thus reducing the high frequency content of the light field. The condition corresponds to the microlens blur radius satisfying $b = 0$, which is verified by points $\mathbf{p}$ in space at a distance (see left plot in Figure 4)

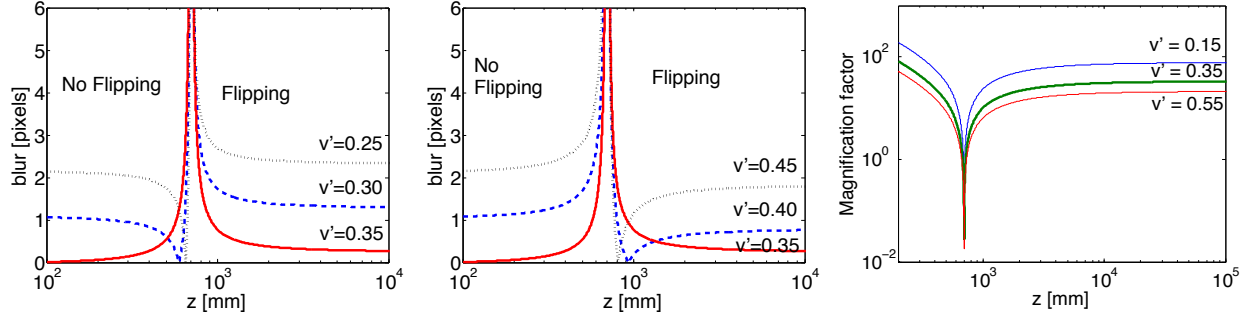$$z = \frac{F(c_z - \frac{v'f}{v'-f})}{c_z - \frac{v'f}{v'-f} - F}. \quad (18)$$

Figure 4. Left: Plot of the microlens blur for several camera settings as we vary the depth of a point in space. Each curve shows a different value of the microlens-to-CCD spacing v' with focal length of the microlenses equal to $0.35mm$ (note that Ng [23] uses v' equal to the focal length of the microlenses). The x axis shows the object depth z with the plane in focus at $700mm$ (in log scale). Middle: The same plot as on the left with v' values greater than the focal length of the microlenses. In our setting we use $v' = 0.4mm$ (dashed plot) and work in the range $800mm - 1000mm$, which results in a microlens blur below 1 pixel. Right: Magnification factor between image that would form at microlens plane, and the actual image that forms under each microlens, under the same settings.

In practice, due to the finite pixel size and diffraction, the views will be in focus for blur radii that are sufficiently small. In addition to the microlens blur, one needs to take into account the *magnification factor* $\frac{z'}{c_z} \frac{v'}{c_z - z'}$ which tells us the scaling between the image that would form at the microlens plane, and the actual image that forms under each microlens. Notice that this factor is directly related to the magnification factor in Lumsdaine and Georgiev [21].

The simplest scenario is when the magnification factor does not change much depending on the depth range. This immediately suggests to work with depths about or larger than $z = 1000mm$ (see right plot in Figure 4). In addition, as one can see in the plots to the left in Figure 4, this depth range corresponds to the flipped region (dotted line) and it leads to a maximum of 1 pixel blur. Notice that Ng's settings (solid line) yield the smallest blur radius over a general depth range, but that is not needed if we limit the working volume as we do.

3. Light Field Superresolution

In order to restore the images obtained by the plenoptic imaging model at a resolution that is higher than the number of microlenses, we employ superresolution by estimating 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 PSF, proper regularisation of the solution through prior modeling of the image data is essential.

We pose the estimation of r in the Bayesian framework, where we treat all the unknowns as stochastic quantities. We begin by noting that under the typical assumption of additive Gaussian observation noise w, the model becomes $l = Hr + w$, and the probability of observing a given light field l in (15) may be written as $\mathrm{p}\left(l \,\middle|\, r, H, \sigma_w^2, s\right) = \mathcal{N}\left(l \,\middle|\, H_s, r, \sigma_w^2 I\right)$, where $w \sim \mathcal{N}(0, \sigma_w^2 I)$. We note here the depth dependence of the matrix H_s (*i.e.*, it is a non-stationary operator).

We then introduce priors on the unknown variables. Here we have focused on the SR restoration and thus we define the imaging model on r, whilst we assume that H_s is known (*i.e.*, we already have an estimate of the depth map). Many recent approaches in image restoration have made use of *nonstationary* priors, which have edge preserving properties. For example, total variation restoration or modeling the heavy-tailed distributions of the image gradients or wavelet subbands are popular methods [25]. Here we apply a recently developed Markov random field prior [4, 3] which extends these ideas by incorporating higher-order information rather than just differences between neighboring pixels. As such, in addition to smooth and edge regions in the image, it is able to locally model texture.

The prior takes the form of a local autoregressive (AR) model, whose parameters we estimate as part of the whole inference procedure. The AR model correlates pixels within a region, and is written

$$r_\omega(i,j) = \sum_{\forall (k,l \in \mathcal{S}_a)} r_\omega(i - k, j - l) a_\omega(k, l) + u_\omega(i, j) \quad (19)$$

where u_ω is a white noise excitation signal with local variance $\sigma_{u_\omega}^2$. This model is written for all regions ω from a given segmentation of the image, in matrix form as $u = (I - A)r = Cr$, where the matrix C represents the non-stationary regularisation operator, or equivalently the synthesis model for the image, parameterized by a. The assumed independence of the excitation signal u_ω in each region allows the joint probability density function to be found as $\mathrm{p}_u\left(u \,\middle|\, \sigma_u\right) = \mathcal{N}\left(u \,\middle|\, 0, Q_u\right)$, where Q_u is a diagonal matrix. Thus via a probability transformation we obtain the image prior

$$\mathrm{p}\left(r \,\middle|\, a, \sigma_u\right) = \mathcal{N}\left(r \,\middle|\, 0, \Sigma_r\right), \quad (20)$$

where $\Sigma_r = \mathbb{E}\left[rr^T\right] = C^{-1}Q_u C^{-T}$; so we see that the model for r depends on both the local variances and au-

toregressive parameters across the image which are concatenated as $\boldsymbol{a}$ and $\boldsymbol{\sigma}_u$ respectively. Under the proposed model we also estimate these parameters by employing standard conjugate priors: Gaussian and inverse-gamma distributions, which let us set a confidence on the likely values of the parameters. Moreover, the Gaussian—inverse-gamma combination used for modeling the local variances also represents inference under a heavy-tailed Student-t if we consider the marginal distribution.

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

3.1. Numerical Implementation

In practice, the variational Bayesian procedure requires alternate updating of approximate distributions of each of the unknown variables. By far the main burden is in updating $\boldsymbol{r}$ at each iteration. The mean of the approximate distribution is found in a standard form as

$$\mathbb{E}^k\left[\boldsymbol{r}\right] = \mathrm{cov}^k\left[\boldsymbol{r}\right] \sigma_w^{-2} \mathbf{H}^T \boldsymbol{l} \qquad (21)$$

$$\mathrm{cov}^k\left[\boldsymbol{r}\right]^{-1} = \mathbf{C}^T \boldsymbol{Q}_u^{-1} \mathbf{C} + \sigma_w^{-2} \boldsymbol{H}_s^T \boldsymbol{H}_s \qquad (22)$$

This system is linear conditional on the previous estimate of the image model parameters. However it is too large in size to minimize it directly, especially given the nonstationary structures of $\boldsymbol{H}_s$ and $\boldsymbol{C}$. Therefore, we use conjugate gradients least squares (CGLS) minimization to estimate $\boldsymbol{r}$ at each step. Due to the factorization of the image prior, we can solve the least square system

$$\boldsymbol{M}^T \boldsymbol{M} \boldsymbol{r} = \boldsymbol{M}^T \boldsymbol{y} \qquad (23)$$

where $\boldsymbol{Q}_u^{-1} = \boldsymbol{L}^T \boldsymbol{L}$, $\boldsymbol{M} = \begin{bmatrix} \sigma_w^{-1} \boldsymbol{H}_s \\ \boldsymbol{L} \mathbf{C} \end{bmatrix}$, and $\boldsymbol{y} = \begin{bmatrix} \sigma_w^{-1} \boldsymbol{l} \\ \mathbf{0} \end{bmatrix}$. These iterations require multiplying by both $\boldsymbol{H}_s$ and its transpose once at each step, which we implement using a look-up table of precomputed PSFs from each position in 3D space. The image restoration procedure is run in parallel tasks across restored tiles of size 200x300 pixels, which are seamlessly joined due to full boundary conditions being used. The size is limited such that the columns of $\boldsymbol{H}_s$ for all the required depths can be preloaded into memory, avoiding disk access during restoration. We run our experiments in MATLAB on an 8-core Intel Xeon processor with 2GB of memory available per task. Pre-calculating the look-up table for $\boldsymbol{H}_s$ requires up to 10 minuites per depth

plane we use, however restoration is faster with each conjugate gradients (CG) iteration takeing around 20 seconds (the actual complexity depends on the depth), and good convergence is achieved after typically 30 to 50 iterations. The image model parameters are recomputed every few CG iterations, taking around 30 seconds. Notice that since the model is linear in the unknown $\mathbf{r}$, convergence is guaranteed by the convexity of the cost functional.

We consider that the depth estimation is performed as an initial step. In our real data experiments we have implemented a multi-view block matching procedure, which minimizes an error term across all the views extracted from the light field. We use a robust norm to help eliminate outliers; first results show that we can obtain a useful initial depth estimate from our data, however we plan to consider incorporating the depth estimation as part of the entire inference procedure, and obtaining a super-resolved depth map.

4. Experiments

4.1. Equipment Description and Calibration

For our experiments we use a Hasselblad H2 medium format camera with an $80mm$ f/2.8 lens. We use a Megavision E4 digital back. The 16MP color CCD has 4096×4096 pixels and a surface of about $3.68cm \times 3.68cm$ (the side of one pixel is $9\mu m$). A custom-fabricated adapter enables us to fit a microlens array very close to the sensor. The array has approximately 250×250 circular lensets, each with a diameter of $135\mu m$, giving about 15×15 pixels per microlens. The focal length of the microlenses is approximately $0.35mm$, and we used a distance v' of $0.4mm$.

Our microlenses have an $f/4$ aperture, although for our experiments we use a smaller main-lens f-number ($f/6.8$), and just the central 7×7 pixels under each microlens. This is because our current prototype microlenses do not have a chromium mask in the gaps, as opposed to Ng $et\ al.$'s system [23]; absence of this mask means light leaks into the outer views from the corners between each microlens, making these views currently unusable for SR.

To obtain useful results an important task is the calibration of the whole system, both mechanical and in postprocessing. It is very important to have the microlenses centered with respect to the sensor so that we can use the maximum resolution available from the light field. Our microlens adapter enables full 3D repositioning and rotation without requiring removal of the back, simply by adjusting screws on the 4 support's edges. After the initial manual correction, post-processing is done on the captured images to compensate for any residual error that is not easily appreciable by eye, including 3D rotation of the microlens images, distortion correction, and photometric calibration.

Figure 5. Synthetic experiment. Left: True Lambertian light field used in first experiment. Right: True depth map.

4.2. Results on Synthetic and Real Data

In Figures 5 and 6, we work with simulated plenoptic camera data. We use our image formation model (15) to compute the light field that would be obtained by a camera similar to our prototype, then apply the SR restoration algorithm to recover a high resolution focussed image from the observation. The simulated scene lies in the range $800mm - 1000mm$, each of the 49 views (*i.e.*, we use the $7{\times}7$ pixel central subimage under each microlens) may be rearranged as a 19×29 pixel image. We use the true depth parameters in the light field reconstruction. The magnification gain is about 7 times along each axis as one can appreciate by comparing the third image from the left with the rightmost image in Figure 6.

We perform a similar experiment with real data obtained from our camera (Figure 2) in Figures 1 and 7. First we estimate a depth map using the multi-view disparity estimation procedure [27], and then we restore a region of the high resolution image. Note that there are still some artifacts visible due to the imperfect calibration of the camera and hence errors in the model; as such the restoration relies heavily on the priors which have been set to perform additional smoothing. Also, the depth map is not updated based on the restored image, and improvements in accuracy are likely to be seen with a simultaneous depth estimation and SR algorithm. Finally, note that we should achieve higher resolution gains by making use of more views, once our hardware allows it.

5. Conclusions

We have presented a formal methodology for the restoration of high resolution images from light field data captured from a plenoptic camera, which is normally limited to outputting images at the lower resolution of the number of microlenses in the camera. This procedure makes the plenoptic camera more useful for traditional photography applications, as well as vision tasks such as depth estimation that we also demonstrate. In the future we hope to incorporate simultaneous depth estimation to improve the process on real scenes.

6. Acknowledgements

References

[1] E. H. Adelson and J. Y. Wang. Single lens stereo with a plenoptic camera. *IEEE J. PAMI*, 14(2):99–106, Feb 1992. 1, 2

[2] M. Ben-Ezra, A. Zomet, and S. Nayar. Jitter camera: high resolution video from a low resolution detector. In *Proc. CVPR 2004*, volume 2, pages 135–142, 2004. 2

[3] T. E. Bishop. *Blind Image Deconvolution: Nonstationary Bayesian approaches to restoring blurred photos*. PhD thesis, University of Edinburgh, 2008. 6

[4] T. E. Bishop, R. Molina, and J. R. Hopgood. Blind restoration of blurred photographs via AR modelling and MCMC. In *IEEE International Conference on Image Processing (ICIP)*, 2008. 6

[5] S. Borman and R. Stevenson. Super-resolution from image sequences - a review. In *Midwest Symposium on Circuits and Systems*, pages 374–378., 1999. 2

[6] M. Born and E. Wolf. *Principles of Optics*. Pergamon, 1986. 3

[7] J.-X. Chai, S.-C. Chan, H.-Y. Shum, and X. Tong. Plenoptic sampling. In *SIGGRAPH '00: Proceedings of the 27th annual conference on Computer graphics and interactive techniques*, pages 307–318, New York, NY, USA, 2000. ACM Press/Addison-Wesley Publishing Co. 2

[8] W.-S. Chan, E. Lam, M. Ng, and G. Mak. Super-resolution reconstruction in a computational compound-eye imaging system. *Multidimensional Systems and Signal Processing*, 18(2):83–101, Sept. 2007. 2

[9] M. Elad and A. Feuer. Restoration of a single superresolution image from several blurred, noisy, and undersampled measured images. *IEEE transactions on image processing*, 6:1646–1658, 1997. 2

[10] S. Farsiu, D. Robinson, M. Elad, and P. Milanfar. Advances and challenges in super-resolution. *International Journal of Imaging Systems and Technology*, 14:47–57, 2004. 2

[11] K. Fife, A. El Gamal, and H.-S. Wong. A 3d multi-aperture image sensor architecture. In *Custom Integrated Circuits Conference, 2006. CICC '06. IEEE*, pages 281–284, 2006. 2

[12] T. Georgeiv and C. Intwala. Light field camera design for integral view photography. Technical report, Adobe Systems Incorporated 345 Park Ave, San Jose, CA 95110 tgeorgie@adobe.com, 2006. 1, 2

[13] B. Ha and Y. Li. Scaling laws for lens systems. *Applied Optics*, 28:4996–4998, 1989. 1

[14] B. R. Hunt. Super-resolution of images: Algorithms, principles, performance. *International Journal of Imaging Systems and Technology*, 6 Issue 4:297–304, 2005. 2

[15] A. Isaksen, L. McMillan, and S. J. Gortler. Dynamically reparameterized light fields. In K. Akeley, editor, *Proc. SIGGRAPH '00*, pages 297–306. ACM Press / ACM SIGGRAPH / Addison Wesley Longman, 2000. 2

[16] A. Katsaggelos, R. Molina, and J. Mateos. *Super resolution of images and video*. Synthesis Lectures on Image, Video, and Multimedia Processing. Morgan & Claypool, 2007. 2

[17] A. Levin, W. T. Freeman, and F. Durand. Understanding camera trade-offs through a Bayesian analysis of light field projections. In *European Conference on Computer Vision*, number 14, pages 619–624, 2008. 1, 2, 3
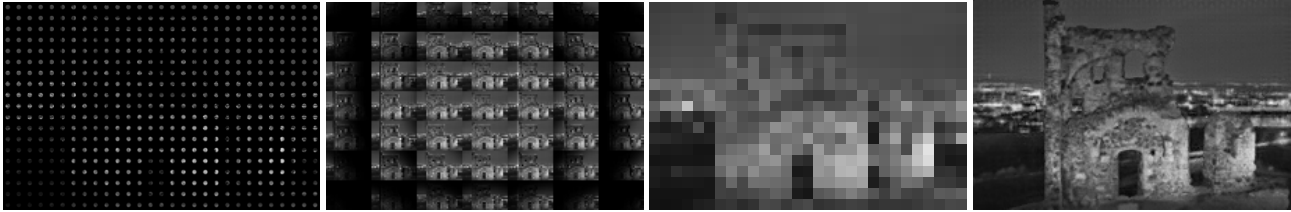
Figure 6. Synthetic results. From left to right: original light field image synthesised with our model; light field image rearranged as views; central view; and restored Lambertian light field with our superresolution method.
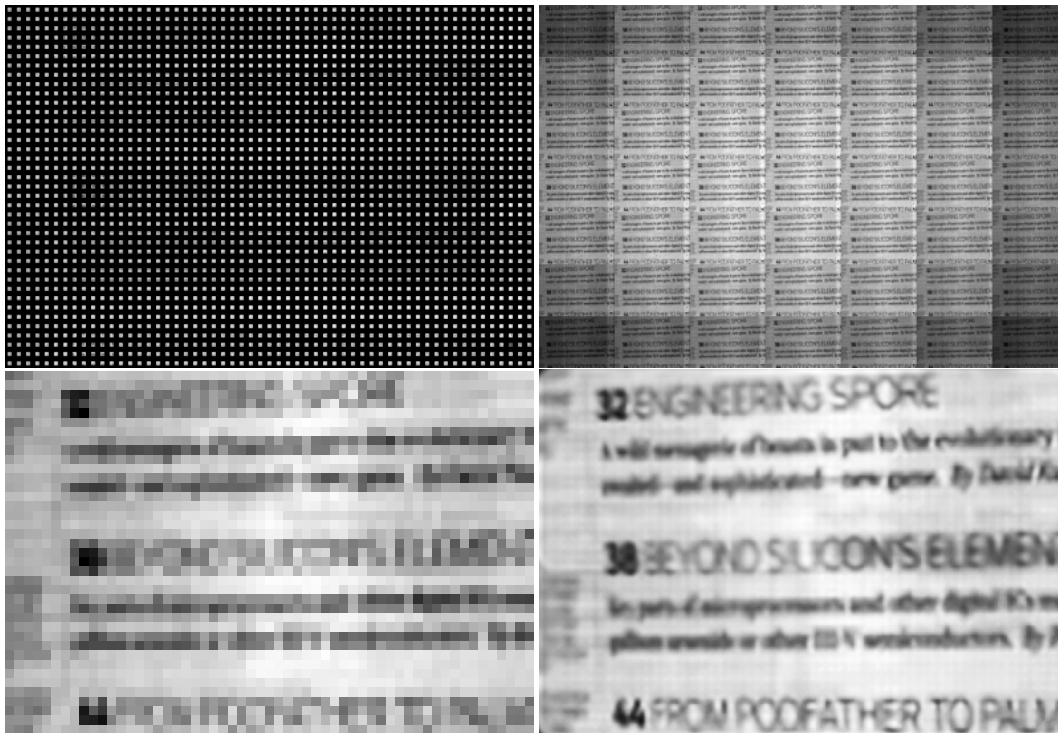

Figure 7. Real results. From left to right, top to bottom: a region of the captured light field image, masked to only show the $7{\times}7$ central views we make use of; this region rearranged as views; one of these views enlarged; superresolution restoration using our method.

[18] M. Levoy, R. Ng, A. Adams, M. Footer, and M. Horowitz. Light field microscopy. *ACM Trans. Graph.*, 25(3):924–934, 2006. 1

[19] C.-K. Liang, G. Liu, and H. H. Chen. Light field acquisition using programmable aperture camera. In *IEEE International Conference on Image Processing (ICIP)*, 2007. 2

[20] G. Lippmann. Epreuves reversibles donnant la sensation du relief. *Journal of Physics*, 7(4):821–825, 1908. 2

[21] A. Lumsdaine and T. Georgiev. Full resolution lightfield rendering. Technical report, Indiana University and Adobe Systems, 2008. 2, 3, 6

[22] M. K. Ng and A. C. Yau. Super-resolution image restoration from blurred low-resolution images. In *Journal of Mathematical Imaging and Vision*, volume 23, pages 367–378, Norwell, MA, USA, 2005. Kluwer Academic Publishers. 2

[23] 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 University, April 2005. 1, 2, 3, 6, 7

[24] S. C. Park, M. K. Park, and M. G. Kang. Super-resolution image reconstruction: A technical overview. *IEEE Signal Processing Mag-*

azine, 20(3):21–36, 2003. 2

[25] E. P. Simoncelli. Statistical modeling of photographic images. In A. Bovik, editor, *Handbook of Image and Video Processing, 2nd Edition*, chapter 4.7. Academic Press, 2 edition, Jan 2005. 6

[26] J. Stewart, J. Yu, S. J. Gortler, and L. McMillan. A new reconstruction filter for undersampled light fields. In *EGRW '03: Proceedings of the 14th Eurographics workshop on Rendering*, pages 150–156, 2003. 2

[27] V. Vaish, M. Levoy, R. Szeliski, C. Zitnick, and S. B. Kang. Reconstructing occluded surfaces using synthetic apertures: Stereo, focus and robust measures. In *Proc. CVPR '06*, volume 2, pages 2331–2338, 2006. 3, 8

[28] 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. Graph. (Proc SIGGRAPH 2007)*, 26(3):69, 2007. 1, 2

[29] G. J. Ward. Measuring and modeling anisotropic reflection. In *SIGGRAPH '92: Proceedings of the 19th annual conference on Computer graphics and interactive techniques*, pages 265–272, New York, NY, USA, 1992. ACM. 1