Plenoptic Depth Estimation From Multiple Aliased Views

Tom E. Bishop Paolo Favaro
Department of Engineering and Physical Sciences
Heriot-Watt University, Edinburgh, UK
{t.e.bishop,p.favaro}@hw.ac.uk

Abstract

With a sampled light field acquired from a plenoptic camera, several low-resolution views of the scene are available from which to infer depth. Unlike traditional multi-view stereo, these views may be highly aliased due to the sparse sampling lattice in space, which can lead to reconstruction errors. We first analyse the conditions under which aliasing is a problem, and discuss the trade-offs for different parameter choices in plenoptic cameras. We then propose a method to compensate for the aliasing, whilst fusing the information from the multiple views to correctly recover depth maps. We show results on synthetic and real data, demonstrating the effectiveness of our method.

1. Introduction

A plenoptic, or light field, camera allows multiple views of a scene to be captured in a single snapshot, in a single compact device [8]. Such cameras sample the light field providing angular as well as spatial information on the distribution of light rays in space. They have recently found application in areas such as 3D reconstruction and digital refocusing. In this paper we are interested in investigating how the 3D surfaces of objects in the scene can be reconstructed with a plenoptic camera.

So far the reconstruction of 3D objects has been addressed in stereoscopic systems and, more in general, camera arrays. Although planar camera arrays and plenoptic cameras share a similar design, they have different sampling patterns, which then lead to different tradeoffs in angular and spatial aliasing. As pointed out by Adelson & Wang [1], plenopic cameras tend to avoid the aliasing of angular samples that is experienced with a camera array, by obtaining a continuum of viewpoints. However, [1] did not consider that spatial aliasing is instead very much present in plenoptic cameras (see for example Fig. 1).[1] Thus traditional multi-view stereo methods that do not take these factors into account may generate false matches and incorrect depth estimates.

Figure 1. **Example of aliasing in plenoptic views.** Left: One view extracted from a plenoptic image (courtesy of [6]) with one enlarged region shown at the bottom. Right: Another view extracted from the same data set (along the horizontal axis). Notice the apparent aliasing affecting these views. Establishing correspondence for depth inference in these views is therefore prone to errors.

At first sight, the removal of aliasing in views of the plenoptic camera does not seem to be a very challenging problem. Based on Shannon's sampling theorem, one simply needs to filter the signal so that its bandwidth is within Nyquist cutoff frequency. In practice, however, this poses two main challenges: first, antialiasing filtering needs to be applied before constructing the views, and second, aliasing in plenoptic cameras varies with the depth of the scene. In particular, the second challenge is a chicken-and-egg problem, as to obtain depth we need to apply an antialiasing filter, and to build the antialiasing filter we need to know the depth. We propose an iterative procedure to solve both of

[1] Note the well known trade off between angular and spatial information: In fact it is only *because* the views are aliased that superresolved refocused images may be obtained [6, 2]; these methods rely on the fact that each view contains different information. This however makes alignment of the views challenging.

2009 IEEE 12th International Conference on Computer Vision Workshops, ICCV Workshops
978-1-4244-4441-0/09/$25.00 ©2009 IEEE

these challenges, where one starts with the antialiasing filter corresponding to the smallest admissible bandwidth and then updates it as depth is estimated.

1.1. Related work

Aliasing in systems similar or equivalent to the plenoptic camera has been analyzed before. In particular, other researchers have studied how the sampling pattern changes depending on the camera geometry. In [4], the choice of sensor-to-microlens spacing in a plenoptic camera is considered. The magnification for different image planes inside the camera is then investigated, along with the depth of focus (although the actual depth of field in the scene is not explicitly computed). In [5], aliasing of light fields sampled from camera arrays is examined; it is suggested that using sufficiently large apertures (equal to spacing between cameras) and pixels with full fill-factor provide the required pre-aliasing, so long as only one part of the scene needs to be in focus. In [10] reconstruction methods for dealing with such aliasing when these conditions are not met are described. The work in [3] appears to be the most closely related to ours. The authors study the requirements for sampling rates to avoid aliased light field rendering. Indeed they observe, as we do, that correct antialiasing depends on depth in the scene. We note again however that the different sampling arrangement of a plenoptic camera leads to different requirements. Finally, Ng [7] discusses *postaliasing* artifacts, introduced in approximate reconstruction of a refocused light field.

Our contributions. To the best of our knowledge, none of the above prior work covered the following two contributions: 1) Novel analysis of aliasing in views captured by a plenoptic camera (that also takes into account the non-negligible blur generated by the microlenses); 2) A method to reduce aliasing in views via space-varying filtering of the captured light field, and an iterative procedure for depth estimation, that is better or equal than no filtering by construction.

2. Plenoptic Depth Estimation

As mentioned in the introduction, we intend to develop an algorithm to reconstruct a 3D map from a single light field snapshot obtained from a plenoptic camera. In our approach we consider extracting views (images) from the light field, so that our input data is suitable for a multiview geometry algorithm. The multiview depth estimation problem can then be formulated as the inference of a disparity map $s(\boldsymbol{u})$ by finding correspondences between the views for each pixel $\boldsymbol{u} \in \mathbb{R}^2$ visible in the scene. Let $V_{\bar{\boldsymbol{u}}}$ denote the view from location $\bar{\boldsymbol{u}} \in \mathbb{R}^2$ on the main lens and $V_{\bar{\boldsymbol{u}}}(\boldsymbol{u})$ the color measured at a pixel $\boldsymbol{u}$ of that view. Then, as we will see in the next sections, depth estimation can be posed as the minimization of the following joint matching error between all combinations of pairs of views[2]

$$E_{\text{data}}(s) = \sum_{\forall \bar{\boldsymbol{u}}, \hat{\boldsymbol{u}}, \boldsymbol{u}} \Phi\big(V_{\bar{\boldsymbol{u}}}(\boldsymbol{u} - s(\boldsymbol{u})\boldsymbol{\Delta}_{\bar{\boldsymbol{u}}}) - V_{\hat{\boldsymbol{u}}}(\boldsymbol{u} - s(\boldsymbol{u})\boldsymbol{\Delta}_{\hat{\boldsymbol{u}}})\big) \tag{1}$$

where Φ is some robust norm, and $\boldsymbol{\Delta}_{\bar{\boldsymbol{u}}} \in \mathbb{R}^2$ is the 2D offset between each view and the central view (the exact definition is given in the next sections). In practice, to save computational effort, we may choose to use only a subset of views $\bar{\boldsymbol{u}}$ in eq. (1) rather than all combinations. Notice that the above definition of the 2D offset implicitly fixes the central view as the reference frame for the disparity map s.

As the views may be aliased, minimising this error is liable to cause incorrect depth estimates around areas of high spatial frequency in the scene. Put in simple terms, even when objects in the scene are Lambertian, the views might not satisfy sufficiently well the photoconsistency criterion so that, even in the presence of no noise, E_{data} may not have a minimum at the true depth map. Moreover, a common procedure to achieve subpixel accuracy is to use interpolation to extend the functions to a continuous representation. This might be a reasonable approximation when the views collect samples of a band-limited (*i.e.*, sufficiently smooth) texture. As we have shown in Fig. 1, however, this is not the case with plenoptic cameras. Therefore, we have to explicitly define how samples are interpolated and study how this affects the matching term between views.

Remark 1 *Notice that, for the purpose of view matching, aliasing does not always constitute a problem. To illustrate this consider the plane in space that is brought in focus onto the microlens array by the main lens. At these spatial locations all views capture light rays exactly from the same sampling grid in space, just along different directions. Hence, if objects are Lambertian, the photoconsistency criterion is always satisfied and matching can be correctly attained (in this trivial case there is no shift). Microlenses that satisfy this property yield a sub-image in the light field with uniform intensity.*

In section 3 we carry out the analysis of the plenoptic camera by investigating aliasing from a purely geometric optics point of view. We start by recalling basic notions and definitions of quantities related to optics of the camera and then in section 3.2 we introduce two dual quantities that are used throughout the paper: the view-based light field and the subimage-based light field. These quantities are instrumental to arrive at a first definition of view aliasing in

[2]Notice that due to the ill-posedness of the reconstruction problem, a regularization term needs to be added. While we do so in our implementation, for the sake of clarity we focus only on the data term in this analysis.

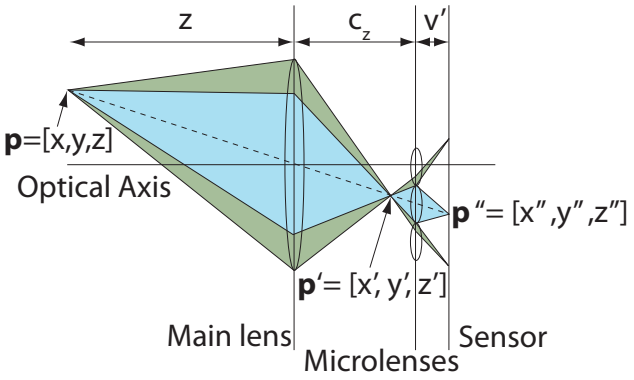

Figure 2. **Schematic of a 2D section of a light field camera.** The plenoptic camera consists of a main lens, a microlens array, and a sensor. The light emitted by a point in space $\mathbf{p}$ is deflected by the main lens and then split into several beams by the microlens array. The size of some microlenses has been exaggerated only for visualization purposes.

Table 1. Light field camera symbols and their description.

Camera parameters	
D	Main lens diameter
d	Microlens diameter & spacing
F	Main lens focal length
f	Microlens focal length
c_z	Microlens to main lens length
v'	Microlenses to CCD sensor distance
μ	Size of a CCD sensor element
Scene parameters	
$\mathbf{p}$	3D point in space
$\mathbf{p}'$	Focused (conjugate) image of $\mathbf{p}$ inside the camera
$\mathbf{p}''$	Projection of $\mathbf{p}$ onto the CCD sensor

section 3.4. In section 4 we define the ideal filtering procedure that also includes a high-resolution image reconstruction step. In section 5 we introduce a more practical solution based on filtering the subimage-based light field and an iterative algorithm. Finally, in section 6 we take into account microlens blur and present the final filtering algorithm.

3. Light Field Aliasing

The plenoptic camera consists of a main lens, as in a standard camera, and an array of microlenses positioned close to the sensor (see Fig. 2, where, for clarity, we show a 2D cross section with a 1D sensor, as the geometry can be readily extended to 3D with 2D images). We use the thin lens model and the geometric optics approximation to describe the system. For ease of reference, we collect the parameters used to describe the model in Table 1. Notice that we assume a general version of the camera in which the spacing of the microlenses from the sensor v' may be different from their focal length f. While Bishop et al. [2] derived a characterization of the depth- and spatially-varying point-spread function (PSF) model for superresolution, here we describe how the model applies to depth estimation.

We begin by analyzing aliasing in a plenoptic camera where the microlenses have been replaced by pinholes and the pixels have negligible size (but spacing μ). We do so because scaling and projection are not affected by this assumption (to be removed later on), and the notation is easier to follow.

3.1. Definitions and Notation

In this section we formalize the notation in the image generation process. The reader mostly interested in understanding the general gist of the analysis can skip this section. In the following we heavily refer to the scheme in Fig. 3. We perform the analysis using projections entirely inside the camera. The whole scene is mapped to the volume behind the main lens, such that each scene plane at a depth z is focused to a sharp *conjugate image* plane at depth $z' = \frac{zF}{z-F}$ (given by the thin lens law) from the lens. We can then treat these conjugate images as the objects to be imaged by the pinhole or microlens array. Notice that in general the scene may not be a plane, so that the depth z varies spatially. We use the coordinates $\boldsymbol{u} = [u, v]^T$ on the microlens plane to define the depth z, so that, in general we have $z(\boldsymbol{u})$ and $z'(\boldsymbol{u})$. The microlens array is located at c_z and consists of $K_1 \times K_2$ microlenses. We then index the microlenses with $\boldsymbol{k} = [k_1, k_2]^T, k_1 = 1 \cdots K_1, k_2 = 1 \cdots K_2$; their centers are located at $\boldsymbol{u} = \boldsymbol{c_k}$, with spacing d, *i.e.*, $\boldsymbol{c_k} = d\boldsymbol{k}$. The angular coordinates $\bar{\boldsymbol{u}} = [\bar{u}, \bar{v}]^T$ are located on the main lens aperture. Their discretised version $\boldsymbol{q} = [q_1, q_2]^T$, with $q_1, q_2 = 1 \cdots Q$, is instead located on the sensor[3] and indexes the pixel centers under each microlens, such that we extract a total of Q^2 views from the captured light field (note that only those lying inside the projection of the main lens aperture onto the sensor contain a useful image). The reprojection of each coordinate $\boldsymbol{q}$ on the sensor onto the main lens is at $\bar{\boldsymbol{u}} = \boldsymbol{\theta_q}$, with spacing $\frac{c_z}{v'}\mu$, *i.e.*, $\boldsymbol{\theta_q} = \frac{c_z}{v'}\mu\boldsymbol{q}$.

3.2. View-Based Vs. Subimage-based Sampling

In order to understand how views are affected by aliasing, we need to study how the plenoptic camera samples the light field. We do so by using the view-based and the subimage-based light field representations.

In the plenoptic camera, each microlens forms its own *subimage* $S_{\boldsymbol{u}}(\bar{\boldsymbol{u}})$ on the sensor, and pixels in each subimage

[3] In the case that the apertures are chosen such that the subimages fully tile the sensor without overlap, each subimage is $d'' = \frac{v'}{c_z + v'}d$ in size and hence $Q = \frac{d''}{\mu}$.

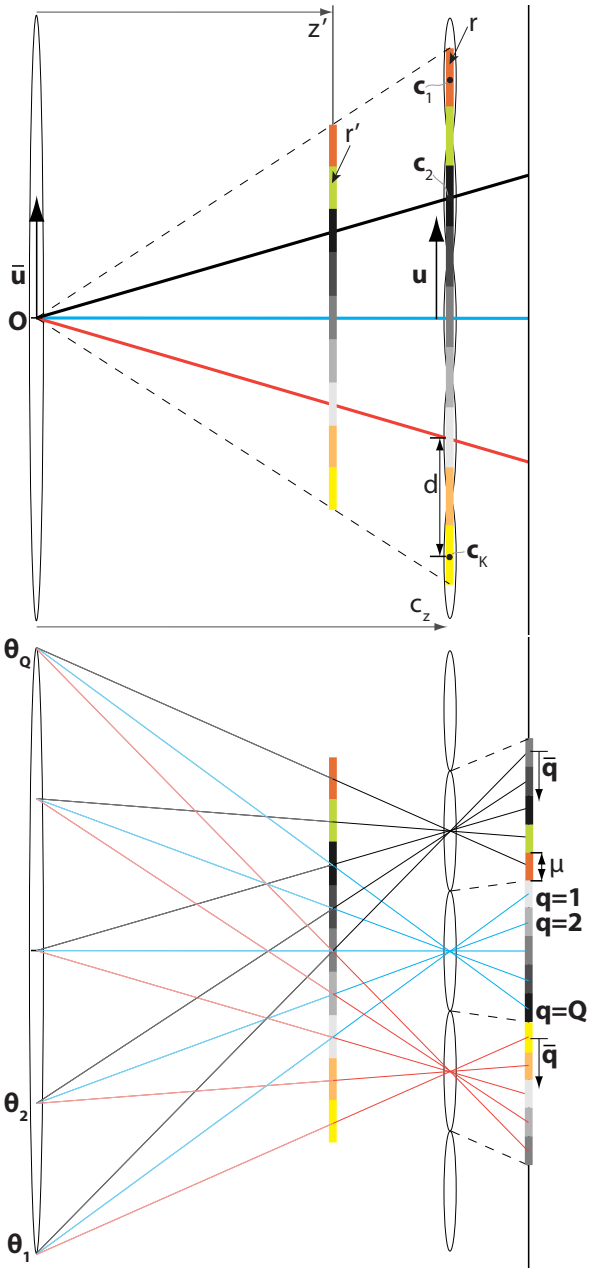

Figure 3. **Coordinates and sampling in the light field camera**. Top: In the main lens pinhole array model, we begin with an image r at c_z, and scale it down to r' at z', with $\boldsymbol{O}$ the optical center of the main lens. Bottom: This conjugate image is then imaged behind each microlens; for instance the blue rays show the samples obtained in the subimage under the central microlens. The central view is formed from the sensor pixels hit by the solid bold rays. Aliasing of the views is present: the spatial frequency in r is higher than the microlens pitch d, and clearly the adjacent view contains samples unrelated to the central view via interpolation.

look through different portions of the main lens (see Fig. 9 for examples of subimages). Thus, the image formed by collecting the pixels under each microlens that look through the same portion of the main lens (at an angle given by $\bar{\boldsymbol{u}}$) form a *view*, which we denote $V_{\bar{\boldsymbol{u}}}(\boldsymbol{u})$. The views and the subimages are two equivalent representations of the same quantity, the sampled light field. Establishing the relationship between different views or different subimages requires using an explicit light field representation. Because we assume that the scene is made of Lambertian objects, we can represent the continuous light field with a function $r(\boldsymbol{u})$ that does not depend on the viewing angle. In our model, $r(\boldsymbol{u})$ is called radiance image or scene texture, and it relates to the light field via a reprojection in space of 2D coordinates on the microlens plane through the centre of the main lens. That is, $r(\boldsymbol{u})$ is the high resolution all-focused image that would be captured by a standard pinhole camera with the sensor placed at the microlens plane. This provides a common reference frame, independent of depth. We also define the depth map on this plane. With reference to Fig. 3, we then represent the plenoptic image formation process as a mapping between $r(\boldsymbol{u})$ and each view or microlens subimage, via the conjugate images at z', which we denote r'. Notice that each pixel is scaled to a different conjugate image depending on the depth map at that position.

A conjugate image $r'(\boldsymbol{u})$ located at z' can be written as a function of $r(\boldsymbol{u})$ by projecting $\boldsymbol{u}$ through $\boldsymbol{O}$, and then by scaling down by $\frac{z'}{c_z}$, *i.e.*, $r'(\boldsymbol{u}) = r\left(\frac{c_z}{z'}\boldsymbol{u}\right)$ (see Fig. 3). We can now define a continuum of views as a sampling of the conjugate image, at points that intersect the rays exiting from $\bar{\boldsymbol{u}}$ and passing through each position $\boldsymbol{u}$ on the array (see Fig. 3). These rays intersect the plane at z' at positions $\bar{\boldsymbol{u}} + \frac{z'}{c_z}(\boldsymbol{u} - \bar{\boldsymbol{u}})$, such that a view is defined as

$$\begin{aligned} V_{\bar{\boldsymbol{u}}}(\boldsymbol{u}) &= r'\left(\bar{\boldsymbol{u}} + \frac{z'}{c_z}(\boldsymbol{u} - \bar{\boldsymbol{u}})\right) \\ &= r\left(\frac{c_z}{z'}\left(\bar{\boldsymbol{u}} + \frac{z'}{c_z}(\boldsymbol{u} - \bar{\boldsymbol{u}})\right)\right) \\ &= r\left(\boldsymbol{u} + s(\boldsymbol{u})\frac{v'}{dc_z}\bar{\boldsymbol{u}}\right) \end{aligned} \tag{2}$$

where we defined $s(\boldsymbol{u}) = \frac{dc_z}{v'}\frac{c_z - z'(\boldsymbol{u})}{z'(\boldsymbol{u})}$. Using the same coordinates, the continuous subimages are defined in exactly the same way, except we now fix $\boldsymbol{u}$ for a particular subimage, and consider the rays from all the different angles $\bar{\boldsymbol{u}}$

$$S_{\boldsymbol{u}}(\bar{\boldsymbol{u}}) = r\left(\boldsymbol{u} + s(\boldsymbol{u})\frac{v'}{dc_z}\bar{\boldsymbol{u}}\right). \tag{3}$$

3.3. Discretisation of Views and Subimages

$V_{\bar{\boldsymbol{u}}}(\boldsymbol{u})$ and $S_{\boldsymbol{u}}(\bar{\boldsymbol{u}})$ are defined for all possible $\boldsymbol{u}$ and $\bar{\boldsymbol{u}}$. In practice, if we approximate the microlens array with an array of pinholes, only a discrete set of samples in each view is available, corresponding to the pinholes at positions $\boldsymbol{u} = \boldsymbol{c_k}$. Furthermore, the pixels in each subimage sample the possible views at $\bar{\boldsymbol{u}} = \boldsymbol{\theta_q}$. Therefore, we define the discrete

observed view $\hat{V}_{\boldsymbol{q}}$ at angle $\boldsymbol{\theta}_{\boldsymbol{q}}$ as the image given by the samples for each $\boldsymbol{k}$

$$\begin{aligned}\hat{V}_{\boldsymbol{q}}(\boldsymbol{k}) &\doteq r\left(\boldsymbol{c}_{\boldsymbol{k}} + \frac{c_z - z'}{z'}\frac{c_z}{v'}\mu\boldsymbol{q}\right) \\ &= r\left(\boldsymbol{c}_{\boldsymbol{k}} + s(\boldsymbol{c}_{\boldsymbol{k}})d\mu\boldsymbol{q}\right). \end{aligned} \tag{4}$$

Here $s(\boldsymbol{c}_{\boldsymbol{k}})$ is the disparity at the discrete locations $\boldsymbol{c}_{\boldsymbol{k}}$ and it depends on the depth z; we also use $\bar{s}(\boldsymbol{c}_{\boldsymbol{k}}) = s(\boldsymbol{c}_{\boldsymbol{k}})\mu$, the view disparity in pixels. The corresponding discretized subimages are just a rearrangement of the pixels in the views; in fact they are also defined by eq. (4), *i.e.*, we have that $\hat{S}_{\boldsymbol{k}}(\boldsymbol{q}) = \hat{V}_{\boldsymbol{q}}(\boldsymbol{k})$.

It is now simple to see that two views are related as

$$\begin{aligned}\hat{V}_{\boldsymbol{q}+\Delta_{\boldsymbol{q}}}(\boldsymbol{k}) &= r\big(\boldsymbol{c}_{\boldsymbol{k}} + \bar{s}(\boldsymbol{c}_{\boldsymbol{k}})d(\boldsymbol{q} + \Delta_{\boldsymbol{q}})\big) \\ &= \hat{V}_{\boldsymbol{q}}(\boldsymbol{k} + \bar{s}(\boldsymbol{c}_{\boldsymbol{k}})\Delta_{\boldsymbol{q}}), \end{aligned} \tag{5}$$

and, conversely, two subimages are related as

$$\begin{aligned}\hat{S}_{\boldsymbol{k}+\Delta_{\boldsymbol{k}}}(\boldsymbol{q}) &= r\left(\boldsymbol{c}_{\boldsymbol{k}} + d\Delta_{\boldsymbol{k}} + \bar{s}(\boldsymbol{c}_{\boldsymbol{k}})d\boldsymbol{q}\right) \\ &= r\left(\boldsymbol{c}_{\boldsymbol{k}} + \bar{s}(\boldsymbol{c}_{\boldsymbol{k}})d\left(\boldsymbol{q} + \bar{t}(\boldsymbol{u})\Delta_{\boldsymbol{k}}\right)\right) \\ &= \hat{S}_{\boldsymbol{k}}(\boldsymbol{q} + \bar{t}(\boldsymbol{c}_{\boldsymbol{k}})\Delta_{\boldsymbol{k}}) \end{aligned} \tag{6}$$

where $\bar{t}(\boldsymbol{c}_{\boldsymbol{k}}) \doteq \frac{1}{\bar{s}(\boldsymbol{c}_{\boldsymbol{k}})}$. Then, it is immediate to conclude that $\Delta_{\bar{\boldsymbol{u}}} = \bar{\boldsymbol{u}}\frac{v'}{dc_z}$ and arrive at the matching terms in eq. (1).

3.4. Matching Views and Aliasing

From eq. (4), it is apparent that the discrete views are just the samples of r with spacing d, but with a different shift $\bar{s}(\boldsymbol{u})d\boldsymbol{q}$, that depends on the view angle and the depth. The multi-view disparity estimation task is to estimate $\bar{s}(\boldsymbol{u})$ by shifting the available views so as to achieve the best alignment. Such alignment, however, requires subpixel accuracy, so that an implicit or explicit reconstruction of r in the continuum is needed. However, 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 a higher aliasing of the views.

4. Light Field Antialiasing

In order to eliminate aliasing from the views, it is necessary to filter the light field *before* views are extracted. Therefore, we consider how we can make use of information *across* the views to help impose the correct antialiasing operation.

Our method is based on an extension of the sampling theorem by Papoulis [9], that shows that if r is bandlimitted with no frequency content above a cutoff frequency $f_r = Qf_0$, then we can accurately reconstruct it if we have Q sets of samples available, with any shifts or linear filtering of the original signal. This implies that we can perform the following steps to obtain the correctly antialiased views $\hat{V}_{\boldsymbol{q}}(\boldsymbol{k})$ from the sampled light field:

1. Use a reconstruction method $\mathcal{F}(\cdot)$ on all of the samples together to obtain r, as $r(\boldsymbol{u}) = \mathcal{F}(\{\hat{V}_{\boldsymbol{q}'}(\boldsymbol{k}')\}, \boldsymbol{u}) = \sum_{\boldsymbol{k}',\boldsymbol{q}'} \Psi_{\boldsymbol{k}',\boldsymbol{q}'}(\boldsymbol{u})\hat{V}_{\boldsymbol{q}'}(\boldsymbol{k}')$ for some set of interpolating kernels $\Psi_{\boldsymbol{k}',\boldsymbol{q}'}$;[4]

2. Filter these samples with an antialiasing filter h_{f_0} at the correct Nyquist rate f_0 to obtain $\tilde{r}(\boldsymbol{u}) = \left(h_{f_0} \star r\right)(\boldsymbol{u})$;

3. Resample to obtain the antialiased views as $\tilde{V}_{\boldsymbol{q}}(\boldsymbol{k}) = \tilde{r}(\boldsymbol{c}_{\boldsymbol{k}} + s(\boldsymbol{c}_{\boldsymbol{k}})\mu\boldsymbol{q})$.

These steps may be combined into a single antialiasing filter acting on the original samples, as

$$\tilde{V}_{\boldsymbol{q}}(\boldsymbol{k}) = \sum_{\boldsymbol{k}',\boldsymbol{q}'} h^*_{f_0}(\boldsymbol{k}', \boldsymbol{q}', \boldsymbol{k} + \bar{s}(\boldsymbol{c}_{\boldsymbol{k}})\boldsymbol{q}))\hat{V}_{\boldsymbol{q}'}(\boldsymbol{k}') \tag{7}$$

where $h^*_{f_0}(\boldsymbol{k}', \boldsymbol{q}', \boldsymbol{k})$ is the combined reconstruction, antialiasing, and resampling kernel.

5. Approximate Light Field Antialiasing

One of the main problems in the above solution is that the reconstruction of r from all the samples can be quite computationally expensive, as it essentially requires running a superresolution algorithm, and working with data at a higher resolution than the size of the views. Secondly, the correct reconstruction filter depends on the as yet unknown depth map. Therefore, we look at an approximate method.

5.1. Subimage Filtering

Rather than performing filtering across the whole light field simultaneously, we consider each subimage independently. That is, we bypass the reconstruction step, and work directly with the subimage data. Because each subimage is a windowed projection of r' onto the sensor, we can project the antialiasing filter h_{f_0} onto the sensor in the same manner, and filter the subimages to achieve a similar result. The main difference is that at the subimage boundaries there is some error due to the windowing by the main lens aperture. Hence, we only consider a limited amount of antialiasing

[4] An ideal reconstruction filter may be derived based upon the sampling theorem in [9], but note that in practice this corresponds to a superresolution operation on the light field texture and any suitable superresolution method may be used.

filtering and use views that are sufficiently close to the central view. Although different tradeoffs are possible, we consider antialiasing filters with support within half of the microlens diameter and views that are within a disc of radius $d/2$ from the central view. This will result in an upper limit of the filter size. Notice that according to the sampling theorem, an ideal low pass filter gives a Sinc kernel; in practice it is common practice to use a windowed version such as a Lanczos window to avoid the infinite support of the kernel.

5.2. An Iterative Algorithm

The whole method is now explained in detail. The antialiasing filter h_{f_0} is defined on the microlens array. If we project this onto the conjugate image at z' and then through each microlens onto the sensor, the scaled filter has physical cutoff frequency $ds(\boldsymbol{c_k})f_0$, or $\bar{\alpha}(\boldsymbol{c_k})f_0$, where $\bar{\alpha}(\boldsymbol{c_k}) = d\bar{s}(\boldsymbol{c_k})$. We begin with a strong antialiasing filter, and refine this estimate based upon the estimated depth map. Too much filtering might introduce additional solutions to the energy minimization problem (1), thus reducing sensitivity. Vice versa, too little filtering might remove the correct solution. We summarize the algorithm as follows:

1. Initialize the depth map $z(\boldsymbol{c_k})$ to the depth at which there would be the most aliasing in the working volume;
2. Estimate the disparity map by minimising eq. (1);
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 $h_{\bar{\alpha}(\boldsymbol{c_k})f_0}$, where $z(\boldsymbol{c_k})$ is the estimated depth map.
5. Repeat from the second step until the disparity map update is negligible.

6. Microlens Blur

So far we have only modelled the pinhole version of the microlenses. In fact this is a worse case scenario; in practice the same analysis holds as before, except that each pixel integrates over a larger area on the conjugate plane. Or, to put it another way, the subimages are now subject to additional blur, thus reducing the effects of aliasing. By taking the additional blur into account, we now arrive at a milder antialiasing filter, and examine the range of depths for which critical aliasing is avoided.

Given an image point on the conjugate plane z', it is known that the *microlens blur* on the sensor due to a circular microlens aperture of diameter d has radius [2]

$$b = \frac{dv'}{2}\left|\frac{1}{f} - \frac{1}{c_z - z'} - \frac{1}{v'}\right|. \qquad (8)$$

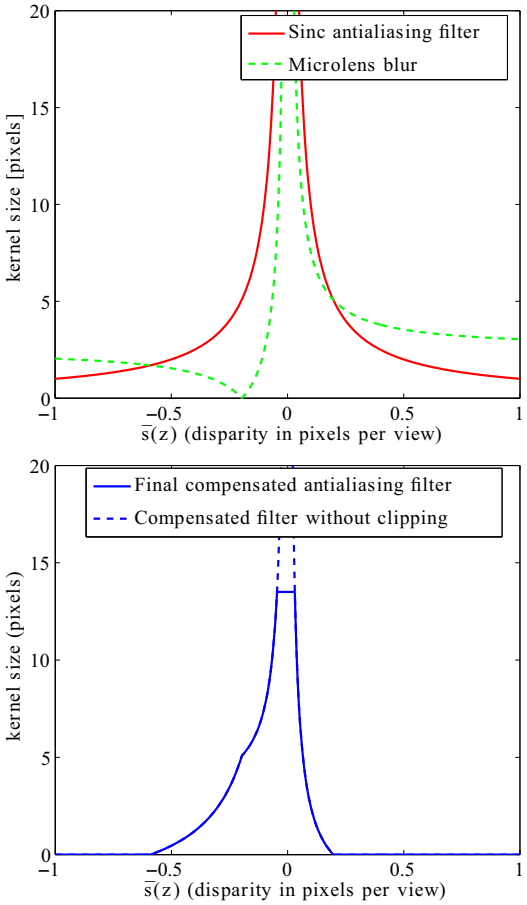

Figure 4. **Microlens blur.** Top: Overlap of filter kernel size and microlens blur radius for different depth values. Bottom: Resulting antialiasing kernel size for different depth values.

Because the microlenses naturally introduce blur in the sampled radiance, one does not need to introduce as much antialiasing as derived in the previous sections. The correct amount of subimage antialiasing can be readily obtained by comparing the microlens blur radius b with the antialiasing kernel size. In the previous section we obtained that the antialiasing filter for an array of pinhole lenses is a Sinc filter. We define as the *antialiasing kernel size* the location of the first zero of the Sinc filter, which corresponds to $\frac{1}{2\bar{\alpha}(\boldsymbol{c_k})f_0} = \bar{t}(\boldsymbol{c_k})$. Then, the final antialiasing filter has a radius that is approximated as the difference between $\bar{t}(\boldsymbol{c_k})$ and the microlens blur radius b, clipped from below at 0 and from above at half the diameter of the microlenses (14 pixels in the experiments with the real data). To see an illustration of the final antialiasing filter used in the experiments on real data, see Fig. 4.

7. Experiments

We have tested our method on both synthetic and real data, kindly made available by T. Georgiev at http://www.tgeorgiev.net.

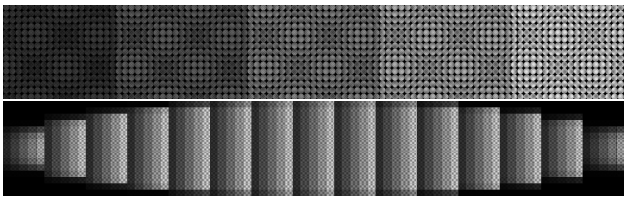

Figure 5. **Synthetic data.** Top: Light field displayed as tiled Subimages. Bottom: Light field displayed as tiled Views.

7.1. Synthetic data

In Fig. 5, we work with simulated plenoptic camera data. The scene consists of 5 steps at different depths, with a sum of sinusoids texture. The two sinusoids in $r(\boldsymbol{u})$ are at 0.2 and 1.2 times the Nyquist rate f_0, therefore the higher frequency causes aliasing of the views (but is still present without aliasing in the subimages). The simulated scene has disparities in the range $\bar{s} = 0.24$ to 0.44 pixels per view. We simulate a camera with $Q = 15$, $\mu = 9 \times 10^{-6}m$, $d = 0.135mm$, $v' = 0.5mm$, $f = 0.5mm$, $c_z = 91.5mm$, and $F = 80mm$. Each of the 81 views (*i.e.*, we use the 9×9 pixel central subimage under each microlens) may be rearranged as a 15×99 pixel image. Fig. 6 shows one of the views, before and after filtering with the estimated and true disparity maps, with the aliasing component removed. resulting subimage detail and view for antialiasing filtering with a few blur diameters. In Fig. 7, proceeding from top to bottom, we show the resulting disparity maps recovered with no antialiasing filtering, the iterative method, by using the correct antialiasing filter and the ground truth. As we are simulating data, we can also evaluate the performance of the algorithm. We repeat the experiment with 16 different depth steps, and compute the average L_2 norm per pixel of the difference between the ground truth disparity map and the solutions obtained with different filtering methods. Results are shown in Table 2, for the sinusoidal pattern and 2 other textures, taken from the Brodatz texture database.

Texture:	Sinusoids	"Bark"	"Straw"
No filtering	2.656	1.822	2.256
Ideal filtering	0.568	0.954	1.152
Iterative filtering	0.662	1.032	1.273

Table 2. L_2 norm error per pixel in disparity estimation ($\times 10^{-3}$), for different choices of filtering method and 3 different textures.

7.2. Real data

Data used in this experiments is shown in Fig. 8. The camera settings used are those described in [4] and [6]. To appreciate the effects of correct choice of antialiasing filter size, Fig. 9 shows the resulting subimage detail and view for antialiasing filtering with a few blur diameters. in Fig. 10, we show from left to right the resulting disparity maps obtained at different steps of the iterative algorithm, along with a regularised version of the final 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 disparities lie in the range -0.34 to 0.31 pixels per view, with the book in the middle being around the main lens plane-in-focus (zero disparity).

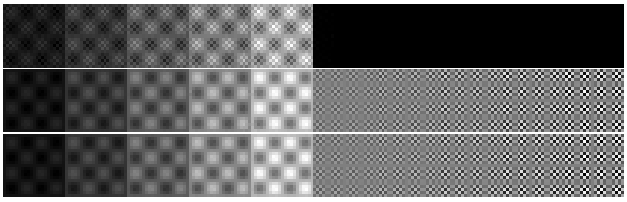

Figure 6. **Antialiasing filtering.** Top row: An unfiltered view (left) and the corresponding high frequency image (right). Middle row: View filtered with the estimated depth (left) and the corresponding high frequency image (right). Bottom row: View filtered with the true depth (left) and the corresponding high frequency image (right). The view filtered with the estimated disparity map is very close to the one obtained with the true disparity map.

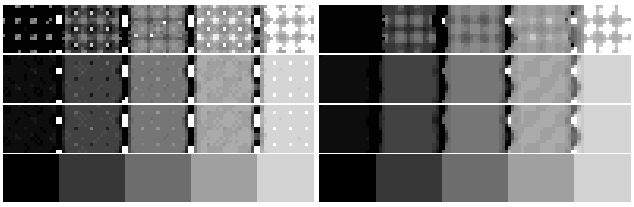

Figure 7. **Depth estimates.** From top to bottom: Results obtained without filtering, results obtained with the iterative method, results obtained with the correct filtering, and ground truth. On each row we show the disparity map obtained without regularization (left) and with regularization (right). Notice how the results obtained with the estimated depth are extremely similar to the ones obtained with the correct filter (the errors at depth transitions occur due to lack of occlusion modelling in the generative model; however this effect does not occur with real data). Depth values are shown in grayscale. Dark intensities correspond to points that are closer to the camera and bright intensities correspond to points that are far from the camera.

8. Conclusions

We have presented analysis and a method to estimate the 3D depth map of a scene from a single plenoptic image. While in camera arrays 3D depth estimation is carried out by using standard multi-view geometry algorithms, we show that in a plenoptic camera aliasing between the views may compromise the reconstruction. Our analysis takes into account the geometry of a generic plenoptic camera and the blur of microlenses. We then propose an iterative method to compensate for aliasing, based on pre-filtering the light field image before extracting the views. We show how this yields improved results on synthetic and real data.

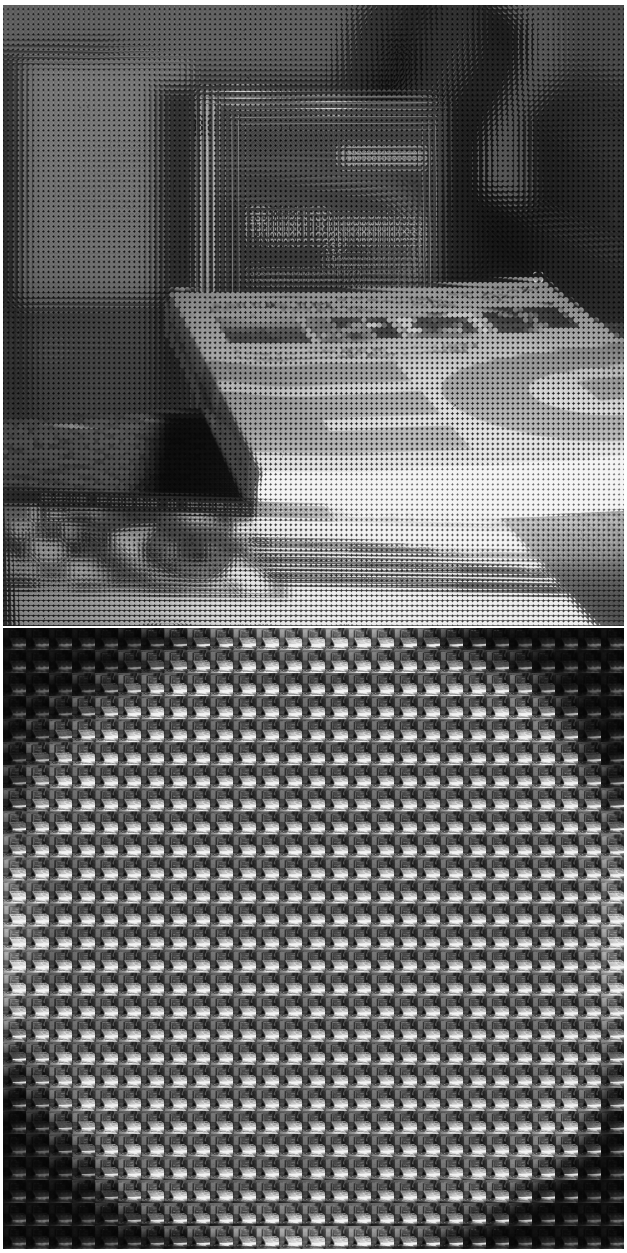

Figure 8. **Real data.** Top: Light field displayed as tiled Subimages. Bottom: Light field displayed as tiled Views.

Acknowledgment 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 cameras. 14(2):99–106, 1992.

[2] T. E. Bishop, S. Zanetti, and P. Favaro. Light field superresolution. In *ICCP 09 (IEEE International Conference on Computational Photography)*, Apr 2009.

[3] 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.

[4] T. Georgiev and A. Lumsdaine. Depth of field in plenoptic cameras. In P. Alliez and M. Magnor, editors, *Eurographics 2009*, 2009.

[5] M. Levoy and P. Hanrahan. Light field rendering. In *SIGGRAPH '96: Proceedings of the 23rd annual conference on Computer graphics and interactive techniques*, pages 31–42, New York, NY, USA, 1996. ACM Press.

[6] A. Lumsdaine and T. Georgiev. The focused plenoptic camera. In *ICCP 09 (IEEE International Conference on Computational Photography)*, Apr 2009.

[7] R. Ng. Fourier slice photography. *ACM Trans. Graph.*, 24(3):735–744, 2005.

[8] 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.

[9] A. Papoulis. Generalized sampling expansion. *IEEE Trans Circuits Syst*, CAS-24(11):652–654, 1977.

[10] 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, Aire-la-Ville, Switzerland, Switzerland, 2003. Eurographics Association.

Figure 9. **Filtered subimages (filter size increases from left to right).** Top row: Close up of a few microlens subimages. Middle row: Corresponding filtered full view. Bottom row: Magnified detail of the view.

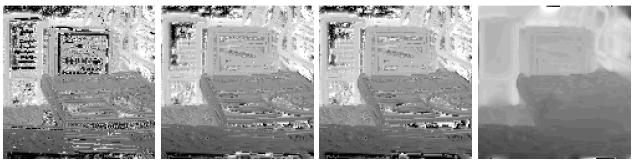

Figure 10. **Depth estimates on real data (iteration time increases from left to right).** First 3 Images: Resulting disparity map estimate obtained without regularization. Final image: Resulting disparity map estimate obtained from the energy of the third iteration, with L_1 regularization. Disparity values are shown in grayscale. Dark intensities correspond to points that are closer to the camera and bright intensities correspond to points that are far from the camera.