Rotational Crossed-Slit Light Fields

Nianyi Li[1] Haiting Lin[1] Bilin Sun[1] Mingyuan Zhou[1] Jingyi Yu[2,1]
[1]University of Delaware, Newark, DE, USA. {nianyi,haiting,sunbilin,mzhou}@eecis.udel.edu
[2]ShanghaiTech University, Shanghai, China. yujy1@shanghaitech.edu.cn

Abstract

Light fields (LFs) are image-based representation that records the radiance along all rays along every direction through every point in space. Traditionally LFs are acquired by using a 2D grid of evenly spaced pinhole cameras or by translating a pinhole camera along the 2D grid using a robot arm. In this paper, we present a novel LF sampling scheme by exploiting a special non-centric camera called the crossed-slit or XSlit camera. An XSlit camera acquires rays that simultaneously pass through two oblique slits. We show that, instead of translating the camera as in the pinhole case, we can effectively sample the LF by rotating individual or both slits while keeping the camera fixed. This leads a "fixed-location" LF acquisition scheme. We further show through theoretical analysis and experiments that the resulting XSlit LFs provide several advantages: they provide more dense spatial-angular sampling, are amenable multi-view stereo matching and volumetric reconstruction, and can synthesize unique refocusing effects.

1. Introduction

Light fields are image-based representation that records the amount of light (radiance) falling in every direction through every point in space. The original light field representation can be described using the 5D plenoptic function (3D for location and 2D for directions). If we assume that the radiance remains constant from point to point along the ray, the plenoptic field is redundant in one dimension and it is possible to describe the light field using a 4D representation. Most notable example for representation such a 4D function is to use the two-plane parametrization or 2PP where a pair of parallel planes Π_{st} and Π_{uv} are given in prior 3D space and each ray is represented by its intersection with the planes as (s, t, u, v) [13].

One of the most important tasks in image-based modeling and later computational photography and imaging is to conduct efficient sampling of the 4D light field. Early examples include capturing the scene using a camera array. The MIT LF camera array uses a grid of 64 1.3 megapixel usb webcams whereas the Stanford array is a two-dimensional grid composed of 128 1.3 megapixel Firewire cameras. At each camera location (s, t), it samples a uv slice that corresponds the image captured camera. Yu and McMillan [27] have shown that each sampled image actually corresponds to a 2D planar slice in the 4D field. The camera array, in essence, samples the 4D space using a sequence of 2D slices. More recent designs such as the light field camera [16] follows the same sampling strategy using a microlenslet array. Compared with the camera array, they can sample more densely on the st dimension due to small microlenslet baselines but sacrifices the uv resolution.

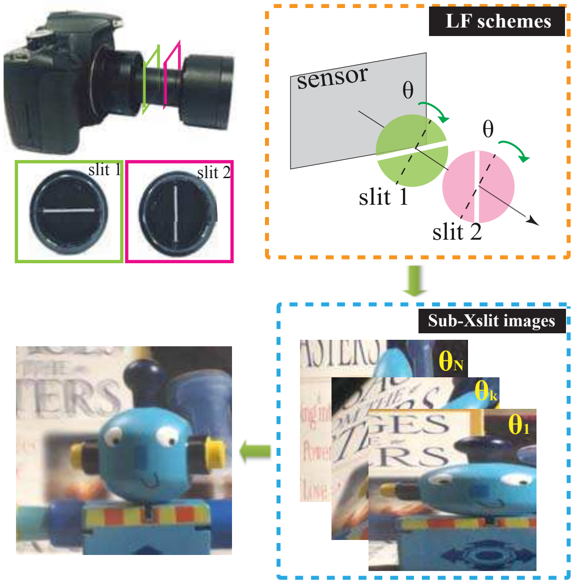
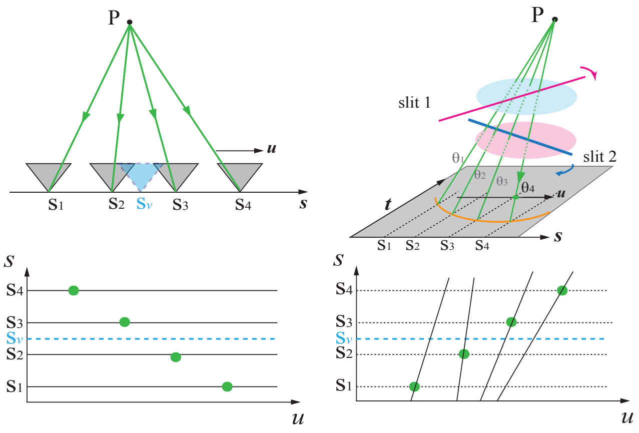


Figure 1. Acquiring a Rotational XSlit Light Field (LF). Top left: The XSlit camera. Top right: The slit rotation scheme. Bottom right: Sample acquired views. Bottom left: Dynamic refocusing effects.

In this paper, we demonstrate an alternative LF sampling scheme. Specifically, we exploit the non-centric crossed-slit or XSlit camera for acquiring the LF. An XSlit camera captures rays that simultaneously pass through two oblique (neither parallel nor intersecting) slits in 3D space [30]. If the two slits are parallel to the 2PP, the captured rays lie on a 2D planar surface in the 4D ray space [24]. In fact, the pin-



(a) Perspective camera array (b) Rotational XSlit LF

Figure 2. The LF sampling pattern using a pinhole camera array (a) and using rotational XSlit camera (b). A new perspective view (blue line) may not contain any samples in the pinhole case but is guaranteed to contain samples in the XSlit case.

hole camera can be viewed as a special XSlit camera where the two slits intersect. Although XSlit geometry has been thoroughly studied [30, 27], only recently practical designs [24] have put them into uses for computer vision tasks such as scene understanding and reconstruction [24, 25].

We adopt the design by Ye *et al*. [24] that relays two cylinderical lenses with slit apertures as the XSlit camera. To sample the LF, our approach is to rotate the XSlit camera along its optical axis, and we show the resulting Rotational XSlit (or RXSlit) sampling scheme provides substantial benefits. On the acquisition front, our new sampling scheme can achieve “fixed-location” light field acquisition. By rotating rather than translating the camera, we eliminate the need of building the camera array or moving the camera along the grid. On the reconstruction front, we show that the new sampling pattern enables more effective view synthesis and dynamic refocusing. Recall that the previous camera array samples uv slices at discrete st locations. Therefore, a new uv slice at an undersampled $s't'$ location does not contain any samples and brute-force interpolation leads to severe ghosting or aliasing [28], as shown in Fig. 2(a). In contrast, we show under the rotational XSlit sampling scheme every perspective view will contain some minimum number of samples, as presented in Fig. 2(b).

We further validate our analysis on using the R-XSlit light field for dynamic refocusing and volumetric reconstruction. Analogous to refocusing with a camera array, we specify a proxy geometry plane and then project all XSlit views onto the plane. The refocused results exhibit some unique effects: defocus blurs become more severe on pixels farther away from the image center. This leads to a novel refocusing effects that we call “Conic Blur”. For 3D reconstruction, we discretize the scene into voxels and apply the XSlit back-projection to map the voxels onto each XSlit view. Finally, we apply the graph-cut algorithm to optimize the 3D embedded voxel graph. Experiments on synthetic and real scenes show that our schemes are robust and reliable.

2. Related Work

Most related to our work are the emerging approaches on light field acquisition and multi-perspective imaging and reconstruction.

The concept of light fields can be back dated to 1936 by Gershun to describe radiometric properties of lights in 3D space [9]. Adelson [1] first introduced notation to the field of computer vision and graphics via the 5D plenoptic function, which later became the foundation to image-based modeling and rendering. The plenoptic function expresses the image of a scene from all possible viewing positions and directions but its high dimensionality has prohibited it from practical uses. Levoy and Hanranhan [13] introduced a practical LF representation using two-plane-parametrization or 2PP where each plane describes a 2D subset and the overall LF is 4D.

By far most commonly used devices of acquiring light fields include a moving held camera or robotically controlled camera [16, 21], a 1D array of cameras [29](as used in capturing the bullet time effect used in the film The Matrix), a dense array of cameras [21], and most recently handheld light field cameras [16] based on the lenslet array or coded apertures [22]. It is also possible, with the help of registration, to capture an unstructured light field by waving a camera 3D space. Nearly all existing solutions use (e.g., in camera array) or emulate (as in lenslet array) perspective cameras as the main acquisition apparatus. The sampling theory under perspective camera sampling has also been well studied, in both spatial and frequency domains [7]. In this paper, we explore a completely different LF sampling scheme based on non-perspective cameras.

While the pinhole camera has long served as workhorses for imaging (including acquiring the light fields), there is an emerging on adopting a non-centric cameras. Classic examples include the pushbroom camera [27] which collects rays along parallel planes from points swept along a linear trajectory and the crossed-slit camera which collects all rays passing through two oblique lines. The General Linear Camera framework [26] discovers that rays collected by both pushbroom and XSlit, along with the classical perspective and orthographic cameras, correspond to 2D planar slices in the 4D light field space. The GLC framework, however, does not discuss the sampling difference when using multi-perspective for acquiring the light field, which is the focus of this paper.

Finally, our work is related to 3D reconstruction. The two most widely adopted reconstruction frameworks are stereo matching and volumetric reconstruction, both can be

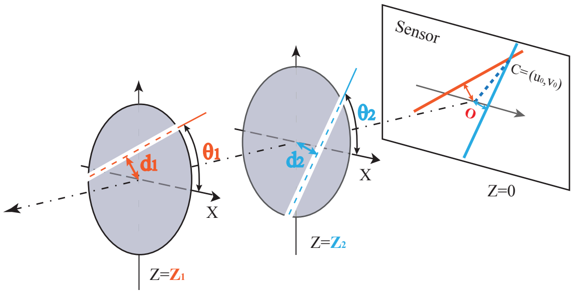


Figure 3. Illustration of our XSlit camera models. The center ray drifts off the image center.

conducted using multi-perspective cameras. For the former, Seitz [20] and Pajdla [18] independently classified all possible stereo pairs in terms of their epipolar geometry. Their results show that beyond perspective camera pairs whose epipolar geometry is a plane, two more varieties of epipolar geometry exist: hyperboloids, and hyperbolic-paraboloids, both corresponding to double ruled surfaces. Ye *et al.* [24] developed a rotational XSlit stereo matching based on hyperboloids and validated the Seitz's theory. For the latter, most adopted solution falls into the category of space carving framework where an initial bounding volume is divided into regular grids and voxels inconsistent with the observation are then pruned. In this paper, we demonstrate that this scheme can be effectively extended to non-centric cameras such as our R-XSlit light fields.

The rest of the paper is structured as follows. Section 3 presents rotational XSlit LF acquisition scheme, discussing its LF sampling pattern, the blur kernel and exploring the epipolar geometry problem. Section 4 discusses the rendering technique and 3D reconstruction method used in rotational XSlit light field. In Section 5, we present the new refocusing and volumetric reconstruction based stereo-matching results on real scene data. Section 6 concludes our work and presents future directions.

3. Rotational XSlit LF

In this section, we discuss how to acquire an LF through rotations using an XSlit camera. Before proceeding, we explain our notation. An XSlit camera collects rays that simultaneously pass through two oblique (neither parallel nor coplanar) slits in 3D space [17, 30, 27]. In this paper, we adopt the light field two-plane parametrization [13] for its simplicity. Specifically, we choose two planes Π_{uv} and Π_{st} parallel to both slits but containing neither slits.

We will also use position-direction parametrization $[u, v, \sigma, \tau]$ where $\sigma = s - u$ and $\tau = t - v$ to simplify certain analysis. We choose Π_{uv} as the default image (sensor) plane so that (u, v) can be directly used as the pixel coordinate and $(\sigma, \tau, 1)$ can be viewed as the direction of the ray. Ye *et al.* [25] assumed that the origin of the coordinate system is the intersection point of two slits' projected lines on Π_{uv}. In this paper, we explore a more general case, *i.e.*, the origin biases that intersection point and two slits rotate along z-axis. We assume that the two slits, l_1 and l_2, lie at $z = Z_1$ and $z = Z_2$ and have angle θ_1 and θ_2 w.r.t. the x-axis, and distance of their projected lines on Π_{uv} to origin point are d_1 and d_2, where $Z_1 > Z_2 > 0$ and $\theta_1 \neq \theta_2$, as shown in Fig. 3. Each XSlit camera can be represented as $\mathcal{C}(Z_1, Z_2, \theta_1, \theta_2, d_1, d_2)$. We applied this notation for sampling the LF by changing θ_1 and/or θ_2. Under this representation, each pixel (u, v) in $\mathcal{C}$ maps to a ray with direction $(\sigma, \tau, 1)$ (see Appendix I) as:

$$\begin{cases} \sigma &= (Au + Bv + F)/E \\ \tau &= (Cu + Dv + G)/E \end{cases} \tag{1}$$

where

$$\begin{aligned}
A &= Z_2 \cos\theta_2 \sin\theta_1 - Z_1 \cos\theta_1 \sin\theta_2, \\
B &= (Z_1 - Z_2)\cos\theta_1 \cos\theta_2, \\
C &= (Z_2 - Z_1)\sin\theta_1 \sin\theta_2, \\
D &= Z_1 \cos\theta_2 \sin\theta_1 - Z_2 \cos\theta_1 \sin\theta_2, \\
E &= Z_1 Z_2 \sin(\theta_2 - \theta_1), \\
F &= (d_1 \cdot Z_2)\cos\theta_2 - (d_2 \cdot Z_1)\cos\theta_1, \\
G &= (d_1 \cdot Z_2)\sin\theta_2 - (d_2 \cdot Z_1)\sin\theta_1.
\end{aligned}$$

To capture R-XSlit LF, we simultaneously rotate both slits while maintaining their relative angle. To simplify our model, we assume POX-Slit camera where the angle between the two slits remains as 90°. [25] captured two such images by rotating the camera by 90 degrees to conduct stereo matching. We characterize ray sampling pattern when exhausting all possible rotation angles and denote the LF sampling scheme as $\mathcal{C}(Z_1, Z_2, \theta + 90°, \theta, d_1, d_2)$ (abbreviated as $\mathcal{C}_\theta$ for simplicity), for all θ. A major advantage of such sampling scheme is that we can rotate the XSlit camera or the XSlits lens set as a unit instead of rotating individual slits.

3.1. Sampling Pattern

To analyze the LF sampling pattern, we fix pixel $p = (u_0, v_0)$ on the sensor plane Π_{uv} and then analyze the sampled rays that pass through p. Specifically, we characterize the sampling function with respect to Π_{st}, *i.e.*, the plane recording the angular information of all rays when rotating the camera. Our analysis assumes l_1 and l_2 have an infinite length. By Eqn. 1, we compute (σ, τ) for (u_0, v_0) in camera $\mathcal{C}_\theta$. Since $s = \sigma + u$, $t = \tau + v$, we have: We prove (see supplementary materials) that the collect rays form a ring on the st plane as:

$$\begin{cases} s &= c_s + r_{\alpha_s}\cos(\theta + \alpha_s) + r_{\beta_s}\cos(2\theta - \beta_s) \\ t &= c_t + r_{\alpha_s}\sin(\theta + \alpha_s) + r_{\beta_s}\sin(2\theta - \beta_s) \end{cases} \tag{2}$$

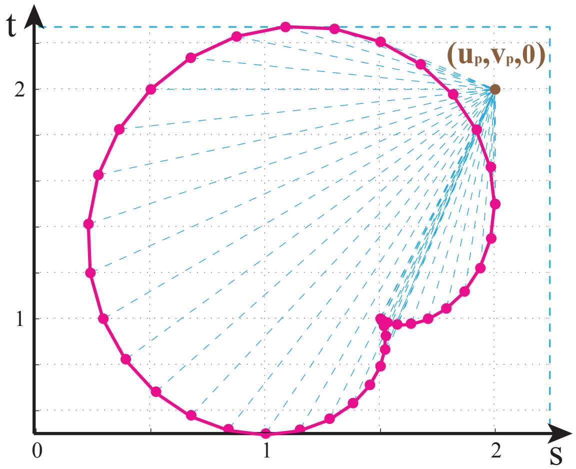
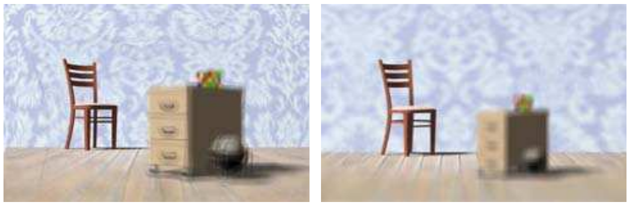


Figure 4. The (s,t) locus of (u_p, v_p) when varying θ from 0 to 2π.

where $c_s = u_0(1 - \frac{1}{2Z_1} - \frac{1}{2Z_2})$, $c_t = v_0(1 - \frac{1}{2Z_1} - \frac{1}{2Z_2})$, $r_{\alpha_s} = \sqrt{(\frac{d_1}{Z_1})^2 + (\frac{d_2}{Z_2})^2}$, $r_{\beta_s} = \sqrt{u_0^2 + v_0^2}(\frac{1}{2Z_2} - \frac{1}{2Z_1})$, $\alpha_s = \arctan\frac{d_2 Z_1}{d_1 Z_2}$ and $\beta_s = \arctan(v_0/u_0)$. This reveals that all (s, t) lie on a Limacon of Pascal curve, as shown in Fig. 4. It is important to note that when $d_1 = d_2 = 0$ the Limacon of Pascal will degrade to a circle.

Compared with LF acquisition using a projective camera array, such rotation-based sampling scheme has a few advantages. Our scheme can acquire many more angular samples. The angular resolution in the projective camera array corresponds to the number of cameras whereas it corresponds to the number of different rotation angles in our case. Mechanically, it would be much easier to rotate the slits than to build a camera array or translation stage for controlling the camera. What is more important is that rotational XSlit light field provides a much denser angular sampling. In the camera array case, its density depends on the spacing (baseline) between cameras and generally it is difficult to make the baseline small enough to avoid undersampling (aliasing). In contrast, In the rotational XSlit, we can make the rotation step very small to acquire a highly dense LF. Although the emerging light field camera can potentially do the same by using tailored optical unit (e.g., a microlenslet array), our sampling scheme will not require using any special optical device.

Fig. 2 shows the sampling differences between the traditional perspective camera array and our rotational XSlit camera. We show a 2D slice su from a 4D light field captured by conventional camera/lenticular array. Under this sampling, each image captured by a camera maps to a 2D parallel slice. Since the space between adjacent slices are "empty", any new perspective view (which corresponds to a slice in between) will not contain any sampled rays and traditional approaches rely on geometry-guided ray interpolation [11]. In contrast, the LF captured under our rotational XSlit camera setup samples the space in a different way: each XSlit camera also maps to a 2D slice [27] but under the rotational setup the recorded slices are not axis-aligned in the 4D ray space. As a result, if we render a new perspective view (2D slice), it is guaranteed to intersect with the sampled XSlit slices and therefore contain some minimal number of ray samples. A detailed analysis can be found in the Appendix III.



R-XSlit LF Refocusing Camera array LF Refocusing

Figure 5. Refocusing rendering comparison between the R-XSlit LF and regular camera array LF.

3.2. Blur Kernel

Given a Rotational LF that captured by $\mathcal{C}_{\theta_i}, i = 1, ..., N$, and a 3D point $P = (x_0, y_0, z_0)$ in the world, we set out to analyze the shape and size of blur kernel by finding the pattern of all the projections of P on a plane Π_f at $z = f$ parallel to the sensor plane. We compute the projection (u_f, v_f) as:

$$\begin{cases} u_f &= (1 - f/z_0)u + x_0 f/z_0 \\ v_f &= (1 - f/z_0)v + y_0 f/z_0 \end{cases} \tag{3}$$

with (u, v) computed as:

$$\begin{cases} u &= c_u + r_{\alpha_b}\cos(\theta + \alpha_b) + r_{\beta_b}\cos(2\theta - \beta_b) \\ v &= c_v + r_{\alpha_b}\sin(\theta + \alpha_b) + r_{\beta_b}\sin(2\theta - \beta_b) \end{cases} \tag{4}$$

where $c_u = -\frac{x_0}{2}(\frac{Z_1}{z_0 - Z_1} + \frac{Z_2}{z_0 - Z_2})$, $c_v = -\frac{y_0}{2}(\frac{Z_1}{z_0 - Z_1} + \frac{Z_2}{z_0 - Z_2})$, $r_{\alpha_b} = z_0\sqrt{(\frac{d_1}{z_0 - Z_1})^2 + (\frac{d_2}{z_0 - Z_2})^2}$, $r_{\beta_b} = \frac{\sqrt{x_0^2 + y_0^2}}{2}(\frac{Z_2}{z_0 - Z_2} - \frac{Z_1}{z_0 - Z_1})$, $\alpha_b = \arctan\frac{d_2(z_0 - Z_1)}{d_1(z_0 - Z_2)}$ and $\beta_b = \arctan(y_0/x_0)$ Details of this derivation shows in the appendix.

According to Eqn. 3 and 4, the projection trajectory of P on plane Π_f is a Limacon of Pascal. The kernel size, depends on the spatial location of P. Getting Closer to the center optical axis or further away from the slits will result in smaller blur kernel size. This dependency of blur size on depth and spatial center is consistent with our vision habit: we focus at an important object and make it centered in the view. Previous studies in biology [8, 19] have shown that human eyes capture a much higher resolution near the center of the retina than near the boundary. This resembles our XSlit light field acquisition system where rays are much more densely sampled (angularly) near the center. Consequently, when we conduct LF refocusing via ray blending,

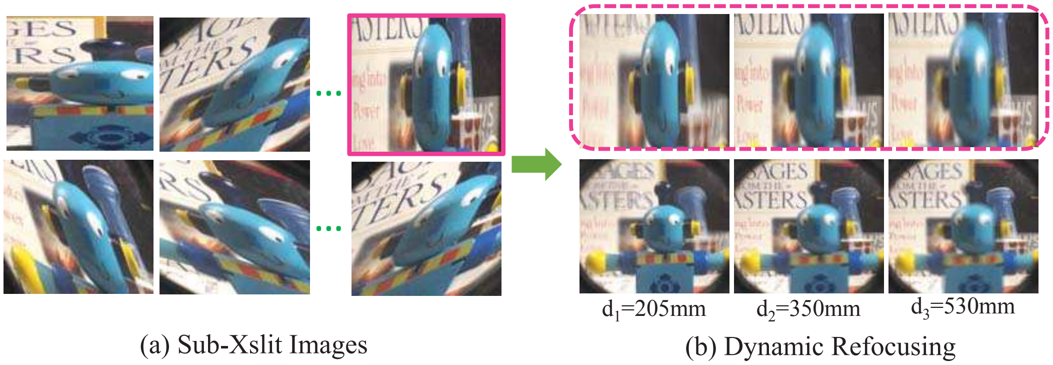


Figure 6. Dynamic refocusing images rendered form R-XSlit Light Field. (a) The Sub-XSlit images are captured by our prototype R-XSlit LF camera. (b) Two different rendering effects. The first row shows the a focus stack using a sub-XSlit image as a reference image; The second row shows refocusing rendering from a perspective view.

our uneven ray sampling leads to non-uniform refocusing. Such a phenomena is very common to the human perception system [15, 6] and [3, 2] have already explored this ”Conic Blur” property in video extrapolation. From above reasons, we believe that the refocusing rendering from the R-XSlit will naturally pleases our vision system.

3.3. Epipolar Geometry Existency

The image sequence captured by rotating both slits generally does not form valid epipolar geometry. In fact, Ye *et al*. [25] have shown that the necessary and sufficient condition for two XSlit cameras to form valid epipolar geometry is when the directions of the two slits get switched, *i.e.* between $\mathcal{C}(Z_1, Z_2, 0, 90^\circ, 0, 0)$ and $\mathcal{C}(Z_1, Z_2, 90^\circ, 0, 0, 0)$. However, in the special POX-Slit case, where the two slits are perpendicular, every image in the captured sequence can form epipolar geometry with the other in the sequence (i.e., the one whose slit directions are flipped) if we rotate the camera to cover 360 degrees. Finally, it is worth noting that even for cases when valid epipolar geometry does not exist, we can conduct efficient volumetric reconstruction.

4. Applications

In this section, we demonstrate applications of our rotational XSlit light field acquisition scheme.

4.1. Image-based Rendering

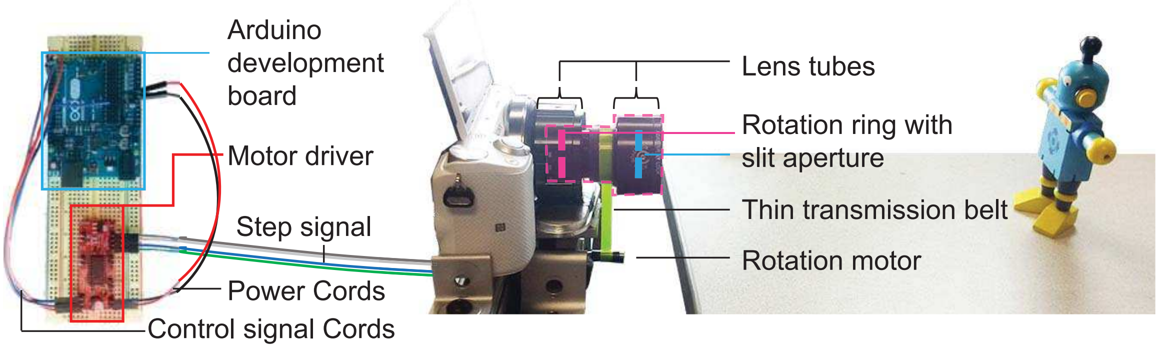
The original goal of acquiring a LF is to conduct image-based rendering, *e.g*., to synthesizing new refocused (perspective) images. For LFs acquired by a pinhole camera array, the refocusing results are synthesized by interpolating between the sampled images. This can be done by first imposing a geometry proxy, e.g., a 3D plane (as shown in the lumigraph [10]), then projecting rays from a reference view to intersect with the proxy, and finally tracing the intersections back to the sampled images to fetch the recorded radiances. Alternatively, one can use a disparity value, if epipolar geometry exists, to directly represent the proxy geometry and to query corresponding pixels from the LF views. As discussed in Section 3.3, there’s no homogenous epipolar geometry in R-XSlit LF, we adopt the first scheme to render focus stacks.

XSlit Refocusing. For the rotational XSlit LF $\{\mathcal{C}_\theta | \theta \in \Omega_\theta = \{\beta_1, \beta_2 ... \beta_N\}\}$, we render refocusing result $\mathcal{J}_\beta^f$ corresponding to XSlit view $\mathcal{C}_\beta$, where superscript f indicates that the focal depth is $z_f = f$. Specifically, we first specify a geometry proxy plane and conduct backward tracing for view blending. Alternatively, we can forward project each XSlit image onto the proxy plane and then combine all images via multi-texturing using the graphics pipeline. In fact, the forward projection of an XSlit image to an arbitrary 2D plane corresponds to a collineation that can be efficiently computed. We can further control the aperture size by varying the number of views involved in the blending . Using a small number of views will result in an image of deep depth-of-field. While a large number will result in shallow depth-of-field effects.

Perspective Refocusing. Using this rotational XSlit LF, we can also render a new perspective image focusing at some focal depth $z_f = f$. We sample a grid of voxels on the plane z_f to render a perspective image. For each voxel $P = (x, y, z_f)$, we trace the rays back to all the XSlit views to fetch the recorded radiances. According to the projection Eqn. 3, we can compute the pixel location q_θ at $\mathcal{I}_\theta$ corresponding to P. Thus the refocusing image $\mathcal{J}_P^f$ can be rendered as:

$$\mathcal{J}_P^f(p) = \frac{1}{N} \sum_{\theta \in \Omega_\theta} \mathcal{I}_\theta(q_\theta). \tag{5}$$

The most notable difference between perspective over R-XSlit LF is the defocus blur kernel.Fig. 6(b) shows examples of perspective view refocusing. In particularly, refo-



(a) Motor Control System

(b) System setup

Figure 7. Our rotational XSlit LF acquisition system prototype. (a) The control circuit for the rotation motor. (b) System setup overview.

cused images exhibit a "conic blur" effect ,*i.e.*, the blurriness is much more severe near the boundary and is nearly invisible near the center as shown in Fig. 6(b). In Fig. 6, we conduct real refocusing on a double-slit rotation LF. From which we can see nice blurring due to dense angular sampling.

4.2. Volumetric Reconstruction

Recall that R-XSlit camera does not have epipolar geometry across all views. The only case that there're epipolar pairs existing is when $d_1 = d_2 = 0$. Such a sampling scheme can be viewed as multiple stereo pairs although no uniform epipolar geometry exists across all pairs. In this case, we can adopt the volumetric reconstruction scheme for both 3D recovery and rendering.

The problem of reconstruction can be formulated as a variation to the foundational space carving framework by Kutulakos and Seitz [12], in which a set of N perspective input camera views are used to recover a 3D volumetric representation of the scene. In classical volumetric reconstruction, the scene is first discretized into voxels of size coherent with the input image resolution. In our case, we first position a virtual perspective camera whose Center-of-Projection lies at $(0, 0, Z)$, where $Z=(Z_1 + Z_2)/2$ with the size of its view frustum matching the extent of both horizontal and vertical slits. To measure the color consistency, we need to first determine the projection of the voxel in each XSlit view. We use the XSlit projection Eqn. 3 to map every voxel to all individual XSlit cameras.

The voxel depth assignment problem is solved via the graph-cut algorithm [5, 4]. Specifically, we traverse the spatial voxels through plane sweeping. For each voxel, we fetch corresponding pixels from respective XSlit images and compute their color variance as the data cost. We also adopt color weighted smooth prior for depth estimation. Fig. 11 shows the reconstruction results.

5. Experiments

We validate our proposed Rotational XSlit light field scheme on both synthetic and real scenes. In this section, we first talk about our acquisition devices and our camera structure. Next, we address the calibration problem of R-XSlit LF and also evaluate the practicability of our scheme. We show both the rendering and stereo matching results using different sampling densities.

5.1. Camera Construction

Fig. 7 illustrates our prototype R-XSlit camera. We mount the XSlit lens on a commodity interchangeable lens camera (*e.g*. Sony NEX-5T). We align the two cylindrical lenses orthogonally using two lens tube. Each tube contains a rotation ring, with which we can control the rotation degree of each slit precisely.

In [25], R-XSlit pairs are acquired though rotating XSlit camera. However, this methods only works when capturing small amount of data. To form valid light field, we need capture large numbers of images as accurate as possible. Nevertheless, it is hard to eliminate or even evaluate the slight bias of rotation axis when rotating the camera, and those small errors are accumulated and can lead to huge inaccuracy. To overcome this, we mount each slit to lens tube with a rotation ring which can rotate 360° freely without affecting the tube. In stead of rotating the camera, we rotate the lens tube. Moreover, to minimize the inaccuracy, we adopt a stepper motor to control the rotation procedure. The lens tube and the motor lever are connected by a flat ribbon to make sure that stepper motor and the lens tube are rotating equally in the same speed. To control the rev rate, we employ a Arduino Uno R3 board, *i.e*., a board that can control the rotation mode of stepper motor by an uploaded program from computer. By applying the stepper motor to the XSlit camera, we are capable of capturing R-XSlit LF through video mode. In this way, we can therefore minimize the manual errors and capture R-XSlit light field without moving the camera. Another advantage of adopting

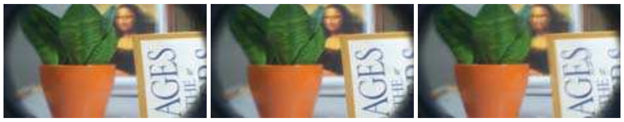

(a) d$_1$=-0.07mm d$_2$=0.258mm (b) d$_1$=-0.17mm d$_2$=0.258mm (c) d$_1$=-0.07mm d$_2$=0.358mm

Figure 8. The refocusing images under different d_1, d_2. (a)(b) have 0.1 difference in d_1, (a)(c) have 0.1 difference in d_2.

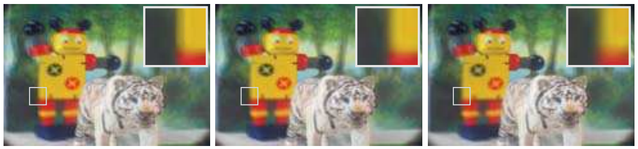

(a) 50 sub-XSlit images (b) 200 sub-XSlit images (c) 500 sub-XSlit images

Figure 9. Refocusing rendering results using different sampling density along the rotation angle. In this example, we focus at the head of the tiger. The out-of-focus region is smooth even using a small number of sub-XSlit images.

stepper motor is that it is easy for us to control the density of the light field. In our implement, we set the rotation rate at 12° per second, the frame rate at 30. Typically, we can capture about 900 images for each light field, *i.e*., when rotating the lens tube 360°.

To ensure the stability of the rotation, we also adopted an additional calibration step beforehand: we captured 3 LFs of a checkerboard calibration target, each with a different rotating speed of the motor and extracted their corners for verification. We found that the LF views align almost perfectly with the theoretical computation. In fact, if they were missed aligned due to uneven rotation speed, the results could also be used to adjust the sequence in the following experiments. The wiggle of axis also seems to marginally affect the results. We suspect this is due to the rigidity of between the camera and stepper motor which make the jiggles nearly negligible. Finally, the lens tube sets were sealed to the camera body and we did not observe obvious changing stray light patterns during the acquisition.

On improving light accumulation, we adopted the dual cylindrical lens design and focus adjustment schemes [23] which has significantly improved the light throughput.

5.2. Calibration

Rather than trying to align the optical axis (*i.e*., the central ray), we set out to calibrate the camera by finding out the bias d_1, d_2 of l_1 and l_2. The two slits' position w.r.t. the image sensor are $Z_1 = 62mm$ and $Z_2 = 26mm$ and have width of $2mm$. For a 3D point $P = (x, y, z)$ in a scene, we capture it three times by rotating the lens tube by 90° on a rotation ring to generate 3 XSlit images. According to Eqn. 4, the projection locations of P on image sensor should be:

$$\begin{cases} u_0 &= \dfrac{Z_1 x - d_1 z}{Z_1 - z} \quad v_0 = \dfrac{Z_2 y - d_2 z}{Z_2 - z} \\ u_{90} &= \dfrac{Z_2 x + d_2 z}{Z_2 - z} \quad v_{90} = \dfrac{Z_1 y - d_1 z}{Z_1 - z} \\ u_{180} &= \dfrac{Z_1 x + d_1 z}{Z_1 - z} \quad v_{180} = \dfrac{Z_2 y + d_2 z}{Z_2 - z} \end{cases} \tag{6}$$

By solving Eqn. 6 we can get that:

$$\begin{cases} d_1 &= -\dfrac{(u_{90} + v_0)(u_0 - u_{180})(Z_1 - Z_2)}{2Z_2(u_{90} - u_{180} + v_0 - v_{90})} \\ d_2 &= \dfrac{(u_0 + v_{90})(v_0 - v_{180})(Z_1 - Z_2)}{2Z_1(u_0 - u_{90} - v_{90} + v_{180})} \end{cases} \tag{7}$$

We therefore choose 30 calibration points on $\mathcal{I}_0$ and find their corresponding points on $\mathcal{I}_{90}$ and $\mathcal{I}_{180}$ respectively. From Eqn. 6 and Eqn. 7, we derive 30 sets of (d_1, d_2). $d_1 = 0.05mm$, $d_2 = 0.28mm$ are the average value of those 30 results.

Fig. 8 illustrates that a slight bias of l_1 and l_2 will have significant impact on the rendering performance. It is worth noting that the average value doesn't guarantee the optimal solution. To find out the correct (d_1, d_2), we first use the average d_1 and d_2 to generate a focus stack using Eqn. 5. Next, we pick out a slice that focusing on a highly textured object at depth f. Note that the slice might still be a little blur due to the incorrect d_1, d_2 value. We then crop 10 8x8 patches from the object, and use the focusness detection methods in [14] to measure the patches' focusness degree when varying d_1 and d_2 respectively. A focuss degree for a (d_1, d_2) pair is computed by averaging all the pixels value in those 10 focusness maps, (d_1, d_2) that achieve the highest degree is regarded as the optimal solution. After the optimization procedure, we derive the best solution $d_1 = -0.07mm$, $d_2 = 0.26mm$.

5.3. Results

We conduct 3D reconstruction and refocusing rendering on both synthetic and real data.

Synthetic Data We first test our scheme on synthetic data rendered by the POV-Ray ray tracer. Fig. 10 presents the refocusing effects rendered by the R-XSlit camera $\mathcal{C}_1(-2, -6, \theta + 90°, \theta, -0.2, 0.1)$ and $\mathcal{C}_2(-2, -6, \theta + 90°, \theta, 0, 0)$. We collected 360 views by $\mathcal{C}_1$ and $\mathcal{C}_2$ with equal angular interval $\Delta\theta = 1°$. In $\mathcal{C}_2$ case, $\mathcal{I}_\theta = \mathcal{I}_{\theta+180°}$. In the refocusing results from $\mathcal{C}_2$, the center portion is always in focus. This is because that when $d_1 = d_2 = 0$, the image centers of all sub-XSlit images corresponds to a same ray. In contrast, $\mathcal{C}_1$ captured multiple rays for every pixels. The Conic Blur effect of $\mathcal{C}_2$ is more obvious than $\mathcal{C}_1$. It is worth noting that for the same reason, $\mathcal{C}_1$ achieves better reconstruction results than $\mathcal{C}_2$ for the center portion. Fig. 11 shows the depth reconstruction result of a synthetic example (first row) using $\mathcal{C}_2$.

Real Data Next, we validate our LF model on scenes acquired by our R-XSlit prototype $\mathcal{C}_\theta(62, 26, \theta + 90°, \theta, -0.07, 0.26)$ (Section 5.1). The R-XSlit LF is captured through video recording. For each captured light field, we can extract about 900 XSlit images at resolution 1920×1080 when two slits rotate 360°. Fig. 9 presents the refocusing using different numbers of XSlit images and we can see that by incorporating more sub-XSlit images, some alias such as the black lines caused by insufficient sampling can be eliminated. However, the out-of-focus region is overall smooth even using a small number of sub-XSlit images. Fig. 11 shows the depth reconstruction results of some real scenes.

6. Discussions and Future Work

We have presented a new framework on acquiring light fields of a scene by using an XSlit camera. Different from previous pinhole based approaches that require translating the cameras in 3D space, we keep the XSlit camera fixed in 3D space but rotate both slits. We have demonstrated that such acquisition scheme exhibits a significantly different sampling pattern of the light field. In particular, under this sampling pattern, any virtual perspective camera is guaranteed to contain a minimal number of acquired samples. The acquired light fields can be further used for effective 3D reconstruction (stereo matching and space carving) and for image-based rendering (new view synthesis and dynamic refocusing). We have also derived defocus blur kernels for R-XSlit LF and validated our theories through comprehensive experiments on synthetic and real data.

On the requirement of mechanically rotating the slit to acquire a light field, we admit that this is a limitation in this initial study, although compared with regular pinhole LF acquisition, our scheme has two major advantages. First, if we fix one slit but rotate the other, we will be able to acquire a 3D LF that has the same ray sampling pattern as translating a pinhole camera along a line. However, rotating the lens/camera is much easier to mechanically implement than translating the camera along a line. Second, in cases such as endoscopic imaging, it would be very difficult to translate a camera. Our rotation scheme however overcomes this limitation.

There are a number of exciting directions that we plan to explore. Our immediate future work is to conduct experiments that individually rotate each slit to acquire the complete 4D light field. There are many interesting questions regarding the resulting light field including the ray density distribution when compared with the light field camera based on microlenslet array, its effects on refocusing quality (aliasing vs. blur kernel), its usefulness in depth inference, etc.

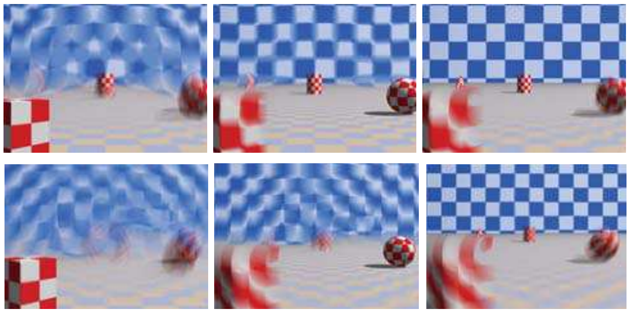


Figure 10. Refocusing effect using different R-XSlit LF camera settings. The first and second rows show the results correspond to $\mathcal{C}_1$ and $\mathcal{C}_2$ respectively. (See text for details.)

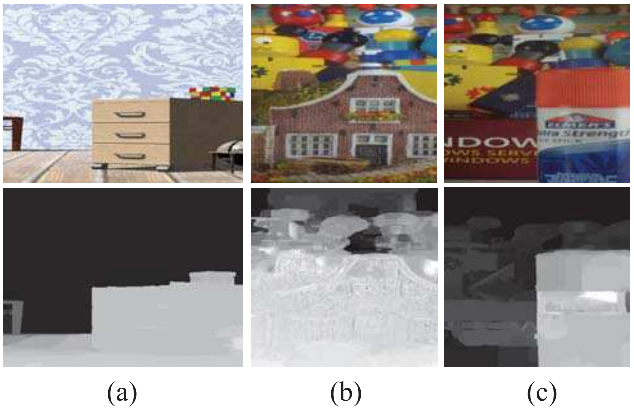


Figure 11. Depth reconstruction from R-XSlit light field on a synthetic example (a) and real examples (b)(c). The first row presents XSlit images and the second row shows their corresponding depth maps.

Our work also reveals a previously overlooked property: a light field acquired by a multi-perspective camera is potentially better for rendering perspective images. This is illustrated in the ray density analysis in image-based rendering. Conversely, the same argument can be made that a light field acquired by a perspective camera (e.g., a camera array) can better render a multi-perspective virtual view. Such phenomenon can be interpreted in terms of ray geometry in the 4D space as an image, perspective or multi-perspective, is a 2D planar cut (the General Linear Camera) in the ray space where ray samples can be viewed as intersections of the GLC plane with the sampling camera planes. In the future, we plan to study the corresponding theories and validate them through experiments using various light field acquisition solutions.

Acknowledgements

This research is supported by National Science Foundation Grant RI-1422477 and CRI-1513031.

References

[1] E. H. Adelson and J. R. Bergen. The plenoptic function and the elements of early vision. *Computational models of visual processing*, 1991. 2

[2] A. Aides, T. Avraham, and Y. Y. Schechner. Multiscale ultrawide foveated video extrapolation. In *Computational Photography (ICCP), 2011 IEEE International Conference on*, pages 1–8. IEEE, 2011. 5

[3] T. Avraham and Y. Y. Schechner. Ultrawide foveated video extrapolation. *Selected Topics in Signal Processing, IEEE Journal of*, 5(2):321–334, 2011. 5

[4] Y. Boykov and G. Funka-Lea. Graph cuts and efficient nd image segmentation. *IJCV*, 2006. 6

[5] Y. Boykov and V. Kolmogorov. An experimental comparison of min-cut/max-flow algorithms for energy minimization in vision. *TPAMI*, 2004. 6

[6] J. A. Brefczynski and E. A. DeYoe. A physiological correlate of the'spotlight'of visual attention. *Nature neuroscience*, 2(4):370–374, 1999. 5

[7] J.-X. Chai, X. Tong, S.-C. Chan, and H.-Y. Shum. Plenoptic sampling. In *Proceedings of the 27th annual conference on Computer graphics and interactive techniques*. ACM Press/Addison-Wesley Publishing Co., 2000. 2

[8] J. T. Enns and R. A. Rensink. Influence of scene-based properties on visual search. *Science*, 247(4943):721–723, 1990. 4

[9] A. Gershun. The light field. moscow. *Journal of Mathematics and Physics*, 1936. 2

[10] S. J. Gortler, R. Grzeszczuk, R. Szeliski, and M. F. Cohen. The lumigraph. In *SIGGRAPH*. ACM, 1996. 5

[11] A. Isaksen, L. McMillan, and S. J. Gortler. Dynamically reparameterized light fields. In *SIGGRAPH*. ACM, 2000. 4

[12] K. N. Kutulakos and S. M. Seitz. A theory of shape by space carving. *IJCV*, 2000. 6

[13] M. Levoy and P. Hanrahan. Light field rendering. In *SIGGRAPH*. ACM, 1996. 1, 2, 3

[14] N. Li, J. Ye, Y. Ji, H. Ling, and J. Yu. Saliency detection on light field. In *The IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, June 2014. 7

[15] M. Morgan and R. Watt. Mechanisms of interpolation in human spatial vision. *Nature*, 299(5883):553–555, 1982. 5

[16] R. Ng, M. Levoy, M. Brédif, G. Duval, M. Horowitz, and P. Hanrahan. Light field photography with a handheld plenoptic camera. *Computer Science Technical Report CSTR*, 2005. 1, 2

[17] T. Pajdla. Geometry of two-slit camera. *Rapport Technique CTU-CMP-2002-02, Center for Machine Perception, Czech Technical University, Prague*, 2002. 3

[18] T. Pajdla. Stereo with oblique cameras. *IJCV*, 2002. 3

[19] J. F. Parker Jr and V. R. West. Bioastronautics data book: Nasa sp-3006. *NASA Special Publication*, 3006, 1973. 4

[20] S. M. Seitz and J. Kim. The space of all stereo images. *IJCV*, 2002. 3

[21] S. University. The (new) stanford light field archive, 2008. http://lightfield.stanford.edu/. 2

[22] A. Veeraraghavan, R. Raskar, A. Agrawal, A. Mohan, and J. Tumblin. Dappled photography: Mask enhanced cameras for heterodyned light fields and coded aperture refocusing. *TOG*, 2007. 2

[23] J. Ye, Y. Ji, W. Yang, and J. Yu. Depth-of-field and coded aperture imaging on xslit lens. In *Computer Vision–ECCV 2014*, pages 753–766. Springer, 2014. 7

[24] J. Ye, Y. Ji, and J. Yu. Manhattan scene understanding via xslit imaging. In *CVPR*. IEEE, 2013. 1, 2, 3

[25] J. Ye, Y. Ji, and J. Yu. A rotational stereo model based on xslit imaging. In *ICCV*. IEEE, 2013. 2, 3, 5, 6

[26] J. Yu. *General linear cameras: theory and applications*. PhD thesis, Massachusetts Institute of Technology, 2005. 2

[27] J. Yu and L. McMillan. General linear cameras. In *ECCV*. Springer, 2004. 1, 2, 3, 4

[28] Z. Yu, X. Guo, H. Ling, A. Lumsdaine, and J. Yu. Line assisted light field triangulation and stereo matching. In *ICCV*. IEEE, 2013. 2

[29] C. L. Zitnick, S. B. Kang, M. Uyttendaele, S. Winder, and R. Szeliski. High-quality video view interpolation using a layered representation. In *TOG*. ACM, 2004. 2

[30] A. Zomet, D. Feldman, S. Peleg, and D. Weinshall. Mosaicing new views: The crossed-slits projection. *TPAMI*, 2003. 1, 2, 3