Depth-of-Field and Coded Aperture Imaging on XSlit Lens

Jinwei Ye, Yu Ji, Wei Yang, and Jingyi Yu

University of Delaware, Newark, DE 19716, USA

Abstract. Recent coded aperture imaging systems have shown great success in scene reconstruction, extended depth-of-field and light field imaging. By far nearly all solutions are built on top of commodity cameras equipped with a single spherical lens. In this paper, we explore coded aperture solutions on a special non-centric lens called the crossed-slit (XSlit) lens. An XSlit lens uses a relay of two orthogonal cylindrical lenses, each coupled with a slit-shaped aperture. Through ray geometry analysis, we first show that the XSlit lens produces a different and potentially advantageous depth-of-field than the regular spherical lens. We then present a coded aperture strategy that individually encodes each slit aperture, one with broadband code and the other with high depth discrepancy code, for scene recovery. Synthetic and real experiments validate our theory and demonstrate the advantages of XSlit coded aperture solutions over the spherical lens ones.

1 Introduction

Recent advances in computational imaging and photography have enabled many new solutions to tackle traditionally challenging computer vision problems. A notable class of solutions is coded computational photography: by strategically blocking light over time [11], space [5,14], and wavelength [1], etc. These solutions can facilitate scene reconstruction as well as preserve image quality. For example, coded aperture, which is initially developed in astronomy or X-ray imaging, has been extended onto commodity cameras. The coded pattern correlates the frequency characteristics of defocus blurs with scene depth to enable reliable deconvolution and depth estimation.

By far nearly all coded aperture systems are built on top of commodity cameras equipped with a spherical thin lens and circular aperture. Spherical lenses can effectively emulate pinhole projection when the aperture is small. This model also facilitates easy analysis of the depth-of-field in terms of aperture size and object distance. To implement coded aperture, it is common practice to replace the circular aperture with the desired coded patterns. Earlier approaches insert printed masks [5,14,21] between the lens and the sensor whereas more recent solutions replace the mask with programmable Liquid Crystal on Silicon (LCoS) to enable dynamic aperture encoding [7]. Tremendous efforts have been focused on developing reliable coding schemes and deconvolution algorithms. In contrast, very little research has been conducted from the perspective of the lens.

D. Fleet et al. (Eds.): ECCV 2014, Part III, LNCS 8691, pp. 753–766, 2014.

In this paper, we explore coded aperture imaging on a special non-centric lens called the crossed-slit (XSlit) lens. Geometrically, an XSlit lens aims to collect rays that simultaneously pass through two oblique slits in 3D space. Its imaging model is non-centric and the acquired images are multi-perspective [18]. Comprehensive studies have been focused on studying XSlit imaging property [23], geometry [9,10], and its applications in image-based rendering [12,19]. The theoretical foundations have recently led to the constructing real anamorphic or XSlit lenses. Yuan et al.[20] developed an XSlit anamorphot and derived the first-order monochromatic aberration. Ye et al.[15] constructed a portable version of the anamorphot by using a relay of two orthogonal cylindrical lenses, each coupled with a (rectangular) slit-shaped aperture. They have further demonstrated using the XSlit lens in Manhattan scene recovery [15] and stereo matching [16].

This paper makes the first attempt to implement coded aperture on an XSlit lens. Clearly, the first question is whether XSlit coded aperture would bring any benefits to the regular spherical lens based solutions. We hence first conduct a ray geometry analysis to study the depth-of-field (DoF) in the XSlit lens. We show that, with the same light throughput, the XSlit lens can produce a different and potentially advantageous DoF than the spherical lens. Specifically, for each XSlit lens, we introduce its throughput equivalent spherical lens (TESL) with a square aperture. We show that the XSlit blur kernels are 2D rectangularly-shaped where the shape is depth-dependent. However, compared with its TESL's square kernels, they are generally smaller in one dimension and remain the same in the other.

Based on our analysis, we further present a strategy to separately encode each slit aperture for scene reconstruction. Developing coded patterns is a dilemma: an ideal pattern will have to have two conflicting properties, i.e., reliable deconvolution vs. high depth discrepancy. The former requires the aperture pattern to be broadband to ensure robust deconvolution whereas the latter requires the code to contain many zero crossings in the frequency domain to distinguish different depth layers. Our strategy is to encode the smaller dimension to the TESL's using broadband codes and the other using high depth discrepancy codes. Synthetic and real experiments demonstrate that our XSlit coded aperture outperforms its TESL's in reliability and accuracy.

2 XSlit vs. Spherical Lenses in Depth-of-Field

We first compare depth-of-field produced by an XSlit lens vs. the traditional spherical lens. We adopt the XSlit lens model in [15] that uses a relay of two cylindrical lenses. For simplicity, we focus our analysis on Parallel Orthogonal XSlit (POXSlit) lenses where the cylindrical axis of the second lens is orthogonal to the first. The general case (when the two slits are not orthogonal) follows a similar derivation and can be found in the supplementary material.

We start with explaining our notation and lens/aperture settings. We assume the two orthogonal cylindrical lenses lie at $z = l_1$ and $z = l_2$ ($l_1 > l_2$) with focal lengths f_1 and f_2, respectively. Further, we assume the cylindrical axis of first

lens is horizontal and the one of the second is vertical, as shown in Fig. 1. The simplest apertures are rectangular-shaped apertures of width w_1 and w_2 on the two lenses respectively. Same as the spherical lens, we can define the f-number of the lenses as $N_i = f_i/w_i$, $i = 1, 2$.

2.1 Cylindrical Lens Transform

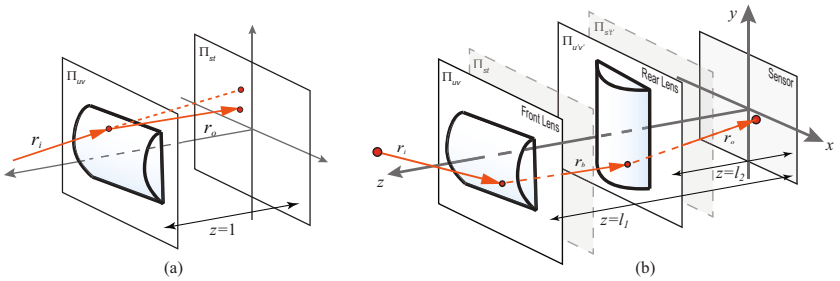


Fig. 1. (a)Ray transformation through a single cylindrical lens; (b) Ray transformation through an XSlit lens

To conduct a DoF analysis, we first study how rays are transformed through the XSlit lens. We start with a single cylindrical lens transform and then concatenate the transforms of two lenses. In the language of ray geometry[17], we set out to derive the cylindrical lens operator (CLO), analogous to the spherical lens operator [8,3].

To parameterize rays, we adopt the two-plane parametrization [6]. Each ray is parameterized by its intersections with two parallel planes Π_{uv} and Π_{st}. We place the cylindrical lens on Π_{uv}. Given an incident ray $\boldsymbol{r}_i = [s_i, t_i, u_i, v_i]^\top$ emitted from the scene towards the lens, the CLO maps $\boldsymbol{r}_i$ to an exit ray $\boldsymbol{r}_o = [s_o, t_o, u_o, v_o]^\top$, leaving the lens towards the sensor. When the cylindrical lens is horizontal, it leaves the u component of the ray unaltered and focuses only the v component; when the lens is vertical, it only focuses the u component. By applying the thin-lens law and similitude relationship, we have the horizontal and vertical CLO as

$$[u_o, v_o, s_o, t_o]^\top = \mathrm{C}(f)[u_i, v_i, s_i, t_i]^\top$$

$$\text{horizontal: } \mathrm{C}_h(f) = \begin{pmatrix} 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 1 & 0 \\ 0 & -1/f & 0 & 1 \end{pmatrix}; \text{ vertical: } \mathrm{C}_v(f) = \begin{pmatrix} 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ -1/f & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 \end{pmatrix} \tag{1}$$

where f is the focal length of the cylindrical lens.

Next we study the concatenation of two cylindrical lenses and derive the XSlit lens operator (XSLO). Specifically, we can trace along an incident ray $\boldsymbol{r}_i$ through the horizontal lens, map it to $\boldsymbol{r}_b$ incident the vertical lens, and compute the final

exit ray $\boldsymbol{r}_o$ reaching (incident to) the sensor. By simply applying the horizontal CLO (Eqn. (1)), we have

$$[u_b, v_b, s_b, t_b]^\top = \mathrm{C}_h(f_1)[u_i, v_i, s_i, t_i]^\top \tag{2}$$

To reuse CLO on the second lens, we represent $\boldsymbol{r}_b$ and $\boldsymbol{r}_o$ under a new 2PP $u'v's't'$, where $\Pi_{u'v'}$ is the lens plane of the second lens. By applying the vertical CLO, we have

$$[u'_o, v'_o, s'_o, t'_o]^\top = \mathrm{C}_v(f_2)[u'_b, v'_b, s'_b, t'_b]^\top \tag{3}$$

To concatenate the two ray transforms, we reparameterize $[u', v', s', t']^\top$ to $[u, v, s, t]^\top$ using similitude transform as

$$\begin{aligned} [u, v, s, t]^\top &= \mathrm{L}(l)[u', v', s', t']^\top \\ \mathrm{L}(l) &= \begin{pmatrix} 1{+}l & 0 & -l & 0 \\ 0 & 1{+}l & 0 & -l \\ l & 0 & 1{-}l & 0 \\ 0 & l & 0 & 1{-}l \end{pmatrix} \end{aligned} \tag{4}$$

where $l = l_1 - l_2$ is the separation between two cylindrical lenses. A similar 2PP reparameterization has been used in previous work for correlating in-lens light fields to out-of-lens ones [8].

Finally, by applying $L^{-1}(l)$ on Eqn. (2) and substituting $\boldsymbol{r}_b$ into Eqn. (3), we have the XSlit Lens Operator (XSLO): $\mathrm{S}(f_1, f_2, l)$ as

$$\begin{aligned} [u_o, v_o, s_o, t_o]^\top &= \mathrm{L}(l)\mathrm{C}_v(f_2)\mathrm{L}^{-1}(l)\mathrm{C}_h(f_1)[u_i, v_i, s_i, t_i]^\top \\ &= \mathrm{S}(f_1, f_2, l)[u_i, v_i, s_i, t_i]^\top \end{aligned} \tag{5}$$

For the more general case where the two cylindrical lenses are not orthogonal, we need to consider their angle and the derivation can be found in the supplementary material.

2.2 Aperture Operator

Same as the spherical lens, a wide aperture can introduce defocus blurs and reduce the DoF. We therefore analyze defocus blurs under the regular shaped slit apertures with width w_1 and w_2. We introduce the two aperture operators: A_1 for the horizontal lens and A_2 for the vertical lens. Notice they are parameterized in v and u' on their lens planes respectively.

$$A_1(v) = \begin{cases} 1 & |v| \le w_1/2 \\ 0 & else \end{cases} \quad \text{and} \quad A_2(u') = \begin{cases} 1 & |u'| \le w_2/2 \\ 0 & else \end{cases} \tag{6}$$

Since only rays that passing through both apertures can reach the sensor, we can derive the closed-form point spread function (PSF) in the XSlit lens as follows. Consider a scene point $\dot{P}$ at depth z, we map all rays originating from $\dot{P}$

to pixels. Without loss of generosity, we assume $\dot{P}$ is on the z-axis. The incident rays $\boldsymbol{r}_i$ originated from $\dot{P}$ can be written as:

$$\boldsymbol{r}_i = [u_i, v_i, (\frac{z - l_1 + 1}{z - l_1})u_i, (\frac{z - l_1 + 1}{z - l_1})v_i]^\top \tag{7}$$

By applying XSLO on $\boldsymbol{r}_i$, we compute the exit ray $\boldsymbol{r}_o$ as

$$\boldsymbol{r}_o = [\frac{\xi}{f_2 z - l_1 f_2}u_i, v_i, \frac{\xi - z + l_2 + f_2}{f_2 z - l_1 f_2}u_i, (\frac{z - l_1 + 1}{z - l_1})v_i]^\top \tag{8}$$

where $\xi = l_1 z - l_2 z - l_1 l_2 + l_2^2 - l_1 f_2 + f_2 z$.

The incident rays are constrained by the two apertures (Eqn. (6)):

$$|u_i| \leq \frac{w_2(z - l_1)}{2(z - l_2)}\ ,\ |v_i| \leq w_1/2 \tag{9}$$

Substituting Eqn. (8) into Eqn. (9), we can map the aperture constraint onto $\boldsymbol{r}_o$:

$$|u_o| \leq \frac{w_2 \xi}{2(f_2 z - f_2 l_2)},\ |v_o| \leq w_1/2 \tag{10}$$

Since the sensor is the xy-plane at $z = 0$, we can directly obtain the PSF from $\boldsymbol{r}_o$ by mapping Eqn. (10) onto the xy-plane:

$$\mathrm{PSF}(x, y) = \begin{cases} 1 & |x| \leq \frac{w_2}{2}(\frac{z}{z - l_2} - \frac{l_2}{f_2}) \text{ and } |y| \leq \frac{w_1}{2}(\frac{z}{z - l_1} - \frac{l_1}{f_1}) \\ 0 & else \end{cases} \tag{11}$$

The XSlit PSF is non-isotropic and its shape is depth dependent as shown in Fig. 2.

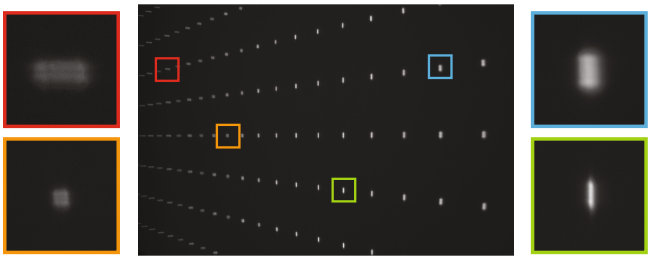


Fig. 2. Captured XSlit PSFs at different depths. We use an XSlit camera with focal length 50mm & 50mm to capture a 32×24 dot-array projected on a slanted screen (ranging from 15cm to 36cm w.r.t.the camera). The XSlit PSFs vary both in shape and scale w.r.t.depth.

2.3 Throughput Equivalent Spherical Lens

One of the most interesting questions regarding the XSlit lens is: is it better than the regular spherical thin lens? We therefore compare the XSlit PSF with the spherical thin lens PSF. For fairness, we need to ensure that the two types

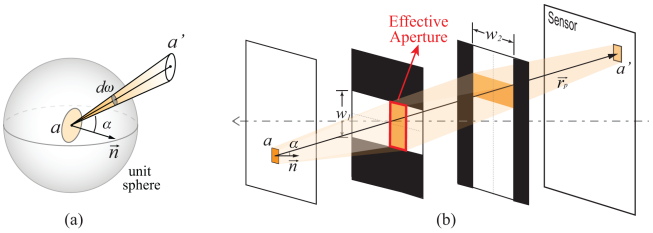


Fig. 3. Light throughput analysis on the XSlit lens. (a) Measure the flux from a to a' using solid angle $d\omega$. (b) Measuring the light enters the camera through the XSlit lens using the effective aperture.

of lenses have the same light throughput. We thus further conduct a radiometry analysis. Specifically, for every XSlit lens, we find its dual *throughput equivalent spherical lens (TESL)*.

Consider a thin pencil of light emanating from a small surface element a of an object at depth z. Assume the radiance of a is B, the light energy (flux) radiated per second from a to the another surface element a' is given by the Lambert's law:

$$d\Phi = B \cdot a \cos\alpha d\omega \tag{12}$$

where α is the angle between a's normal and the line connecting a and a'; $d\omega$ is the solid angle subtended by a' at a, as shown in Fig. 3(a).

By Eqn. (6), the area of effective aperture is $\frac{z-l_1}{z-l_2} w_1 w_2$. The project area of the effective aperture as seen by a can be computed as $\frac{z-l_1}{z-l_2} w_1 w_2 \cos\alpha$. By mapping the project area onto the unit sphere, we obtain the solid angle

$$\omega = \frac{z-l_1}{z-l_2} w_1 w_2 \cos^3\alpha \tag{13}$$

Substituting Eqn. (13) into Eqn. (12), we have the flux received at the effective aperture

$$\Phi = B \cdot a \cos\alpha\omega = B \cdot a \frac{w_1 w_2}{(z-l_1)(z-l_2)} \cos^4\alpha \tag{14}$$

Eqn. (14) indicates the total amount of light entering the camera. We can further map Φ to irradiance (flux per area) received at sensor. We have the imaged area of a as $a' = \frac{l_1 l_2 a}{(z-l_1)(z-l_2)}$ by using similar triangles. Dividing Φ by a', we have the irradiance received at sensor:

$$E_{\text{XSlit}} = \frac{\Phi}{a'} = B \frac{w_1 w_2}{l_1 l_2} \cos^4\alpha \tag{15}$$

Since the lens-to-scene distance is much larger than the lens-to-sensor distance, we can approximate l_i/w_i using f_i/w_i. Therefore, we rewrite Eqn. (15) in terms of f-numbers

$$E_{\text{XSlit}} = B N_1 N_2 \cos^4\alpha \tag{16}$$

Recall that, for a spherical lens with focal length f_p and f-number N_p, the irradiance can be computed as:

$$E_{\text{spherical}} = BN_p^2 \cos^4 \alpha \tag{17}$$

By Eqn. (16) and Eqn. (17), given a XSlit lens with f-numbers N_1 and N_2, its TESL will have f-number $N_p = \sqrt{N_1 N_2}$.

2.4 Depth-of-Field Comparisons

Now that we have derived the TESL of an XSlit lens, we can compare their DoFs. It is very important to note that, different from the spherical lens, if the two cylindrical lenses in an XSlit lens have identical focus length, it is not practical to focus the two lenses at the same depth (the only in-focus depth is $z = l_1 + l_2$, which is too close to the front horizontal lens at $z = l_1$). We therefore focus the two lenses at two different depths. At the first glance, this focus setting may appear highly undesirable as no 3D point can be clearly focused in the image. A deeper analysis as follows, however, reveals that it has several advantages.

Recall that the front lens is horizontal thus producing vertical blurs and the rear lens (closer to the sensor) is vertical resulting in horizontal blurs. By Eqn. (11), the horizontal and vertical PSF scales are:

$$b_v = (\frac{z}{z - l_1} - \frac{l_1}{f_1})w_1 \quad \text{and} \quad b_h = (\frac{z}{z - l_2} - \frac{l_2}{f_2})w_2 \tag{18}$$

We assume that the TESL coincides with the the front lens. Recall that the spherical lens' PSF is uniform. Therefore, the spherical PSF scale b is always identical to the vertical scale of XSlit PSF ($b = b_v$). Given the lens parameters and focused depth, we can plot the curves of PSF scales vs. scene depth, as shown in Fig. 4(b). The red curve corresponds to b_v and the green one corresponds to b_h. Assume z_{front} refers to the depth that the front lens focuses at, z_{rear} refers to the one that the rear lens focuses at, and z_{inter} between z_{front} and z_{rear} refers to the depth the two curves intersect.

At z_{front}, the TESL has zero PSF whereas the XSlit lens has only zero vertical scale but appear defocused horizontally. This is a major disadvantage of the XSlit, i.e., no point in the scene can be clearly focused either horizontally or vertically when using the same focal length for each cylindrical lens.

Let us consider an important depth z_{inter} where the two curves intersects. Before z_{inter}, the XSlit performs worse, i.e., it incurs more horizontal blur than the TESL ($b_h > b_v = b$). Although this is undesirable, this range is much smaller than the complete scene range in a typical scene. Therefore, under the same setup, only a small range of depth exhibits worse PSF in XSlit than in the pinhole lens while majority depths in XSlit will appear much clearly focused.

The key advantage of the XSlit is that after z_{inter} the horizontal scale of the XSlit PSF is always smaller than the spherical lens one ($b_h < b_v = b$). One can explain this phenomena by either algebraic reasoning or geometry/optics reasoning. Algebraically, since the two lenses focuses at different scene depth,

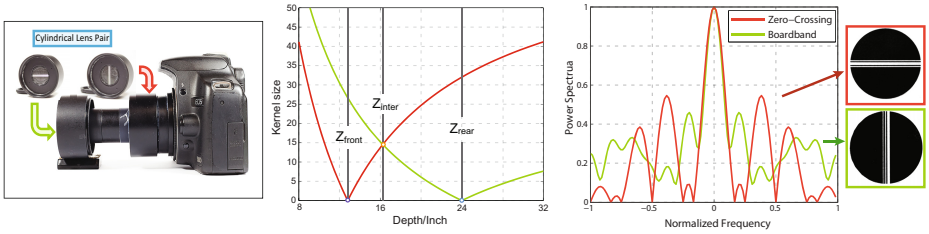


Fig. 4. (a) XSlit camera prototype; (b) XSlit PSF scales vs. depth; (c) The power Spectra of our selected coded apertures

b_h and b_v map to two different curves. The two curves will intersect somewhere between z_{front} and z_{rear}, after which b_h will always be smaller than b_v. From the geometric optics perspective, we can also reason this phenomena. The rear lens focuses farther away than the front lens. Since the two lenses have an identical focal length, after depth z_{rear}, the rear lens will incur smaller defocus blur because the relative depth of the corresponding point to its focal plane is larger than to front one.

3 XSlit Coded Aperture Imaging

The analysis above indicates that the XSlit lens have a different and potentially advantageous DoF than its TESL. In this section, we exploit the special property of XSlit DoF and implement a coded aperture system for scene recovery.

Depth recovering using coded aperture has been thoroughly explored on spherical lenses in the past decade [5,14,7,22,13]. The basis idea is to analyze the coded pattern and defocused images to recover scene depth and produce an all-focus image. However, designing the code is a dilemma. To discriminate depth, the aperture pattern should have zero-crossings in the frequency domain to purposely introduce variations among blurry images in terms of depths [5]. However, to ensure robust deconvolution, the aperture pattern should be broadband, i.e., its frequency profile should have few zero-crossings [14].

When only a single coded aperture is used on the spherical lens, one can potentially combine the two types of codes through multiplexing, e.g., horizontally using the broadband and vertically using the high depth discrepancy ones. Recall that the DoF analysis shows that compared with its TESL, the XSlit lens (under the focus configuration as discussed in Section 2.4) exhibits less horizontal blurs and approximately the same vertical blurs under the same light throughput. Our strategy hence is to encode the first cylindrical lens (the horizontal one) using the high discrepancy kernel and the vertical lens with a broadband one. We choose this strategy to remain the same depth discrepancy as its TESL (since they have identical vertical blur scale) whereas the other dimension is less blurred to provide more robust deconvolution. If we switch the coding scheme (i.e., the first with broadband pattern and the second with high depth discrepancy one), although

the all-focus image can be more easily restored, the depth discriminative ability is compromised.

3.1 Code Selection

Next we discuss how to choose the appropriate coded pattern for each cylindrical lens. Assume the vertical pattern is $K_v(x)$ and the horizontal pattern is $K_h(y)$, the overall blur kernel is therefore $K(x, y) = K_v(x) \cdot K_v(y)$ where x and y are further constrained by the close-form PSF (Eqn. (11)). For each 1D pattern, we choose from a series of randomly sampled 13-bit codes. Since the vertical code K_v is required to be broadband, we select the one with the maximum min-value in the frequency domain.

Then we fix K_v and find the optimal K_h. Assume the blur kernel at depth i is $K^i = K_v^i \cdot K_h^i$. To have better depth discrepancy, we want to maximize the distance between blurry image distributions caused by kernels at different depths, i.e., K^i and K^j $(i \neq j)$. We use the commonly used Kullback-Leibler (KL) divergence to measure the distance between two blurry image distributions

$$D(P^i(y), P^j(y)) = \int_y P^i(y)(\log P^i(y) - \log P^j(y))dy \tag{19}$$

where P^i and P^j are the blurry image distribution for K^i and K^j respectively and we use the hyper-Laplacian distribution of natural images for computing them [5,4].

Finally, we use "1010010011111" as the vertical code and "110011110011" as the horizontal code in our implementation. The power spectra of the selected codes are plotted in Fig. 4(c).

3.2 Depth Estimation

To estimate depth, we first precompute the corresponding point spread function (PSF) for each layer using the coded pattern and the closed-form PSF derived in Section 2. Once we acquire the image, we check which PSF yields to the optimal results. Specifically, we conduct deconvolution using PSFs of different depth layers: when the scale is larger than the actual one, the result will exhibit strong ringing artifacts; when the scale is smaller than the actual one, the image would appear less sharp but still without ringing. We use Gabor filter to detect ringing in the deconvolved image, i.e., greater responses correspond to more severe ringing. A Gabor filter is a Gaussian kernel function modulated by a sinusoidal plane wave and can be written as

$$G_{\lambda,\theta}(x, y) = \exp(-\frac{x'^2 + \gamma y'^2}{2\sigma^2}) \cos(2\pi \frac{x'}{\lambda}) \tag{20}$$

$$x' = x\cos\theta - y\sin\theta,\ y' = x\sin\theta - y\cos\theta$$

where λ is the wavelength (reciprocal of the spatial frequency), θ is the orientation of the filter, γ is aspect ratio, and σ is the standard deviation of Gaussian.

We use Gabor filters with $\theta = 0°$ and $90°$ for ringing detection. We define the response of a Gabor filter G_θ as

$$R_\theta(x, y) = \iint I(x, y) G_\theta(x - u, y - v) du dv \tag{21}$$

We sum up the horizontal and vertical Gabor responses on each deconvolved image and thus the one with smallest value corresponds to the optimal depth. We discretize the scene to N depth layers and reuse the graph-cut algorithm [2] for assigned depth labels. We use the Gabor response as the penalty term for building the graph. Therefore, the energy function E of assigning a depth label d_i to a pixel p is formulated as

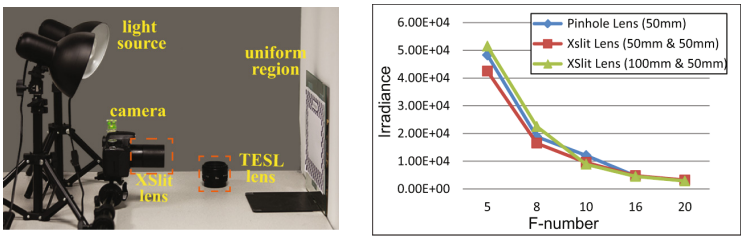
$$E(d_i) = \alpha \cdot \sum_{p \in P} E_d(p, d_i(p)) + \sum_{p_1, p_2 \in N} E_s(p_1(d_i), p_2(d_j)) \tag{22}$$

where P represents all pixels in the image; N represents the pixel neighborhood; $E_d(p, d_i(p))$ is the Gabor response as the data term; E_s is the smooth term; and the non-negative coefficient α balances the data term and the smooth term.

To recover all-focus image, since our blur kernel only compromises 1D scene information, we simply reuse the modified Wienner deconvolution using natural image priors [21] which is much faster than the Iteratively Reweighted Least Squares (IRLS) deconvolution [5] that can handle kernels with many zero-crossings.

4 Experiments

We validate our DoF analysis and coded aperture imaging model on real XSlit lenses as shown in Fig. 4(a). We follow the same design as [15] to construct the XSlit lens using a relay of two cylindrical lenses with focal length 50mm. We use a 50mm spherical lens as its TESL. We use the camera body of Canon 60D. For all experiments, we use ISO 100 and shutter speed 1/100s.



(a) (b)

Fig. 5. Light throughput comparison between the XSlit lens and its TESL. (a) Experimental setup; (b) The irradiance vs. f-number curve for the three lenses.

Throughput. Given a specific f-number, e.g., $f/10$ (i.e., with both slit aperture widths 5mm), we can compute the aperture size of its TESL by dividing the f-number with the focal length (in this case, a square aperture of 5mm×5mm). We further construct the second XSlit lens where the two cylindrical lenses have different focal lengths (the first of 100mm and the second of 50mm). By the definition of f-number in Section 2, for the second XSlit lens to have the same throughput as the first and its TESL, we need to double the aperture width of the 100mm lens (e.g., in this example 10mm). In our experiments, we compare the throughput of all three lenses with $f/5$, $f/8$, $f/10$, $f/16$ and $f/20$.

We use all three lenses to capture a uniform white diffuse surface under the same lighting condition, as shown in Fig. 5(a). We use the averaged intensity value of the linearly tone-mapped raw image from the camera as irradiance measure. Fig. 5(b) plots irradiance vs. f-number curves of all three lenses. Our results are consistent with the throughput analysis: as far as the XSlit lenses and the TESL have the same f-number, they have equivalent light throughput.

Notice though that when the aperture size gets too big (e.g., $f/5$), the XSlit lenses exhibit strong vignetting.

Depth-of-Field. Next we valid our DoF analysis of the XSlit lens vs. its TESL. We construct a scene that consists of three objects lying at depth 38cm, 50cm and 63cm w.r.t.the camera. For the XSlit lens, the front lens focuses at 40cm and the rear lens at 45cm whereas its TESL focuses at 40cm. The image captured by the XSlit and its TESL are shown in Fig. 6(a) and (b) respectively. Our analysis predicts that the PSF of the XSlit should be narrow in the horizontal direction than the TESL for objects lying far away. Notice that layer 2 and layer 3 have depth greater than z_{inter} and in the acquired images, they appear less horizontally blurred in the XSlit image than the TESL one, which is consistent with our prediction. Furthermore, the vertical blur scales are approximately the same in both the XSlit and its TESL. The XSlit hence improves the depth-of-field for objects at depth greater than z_{inter}, e.g., the texts in layer 3 appear clearer in the XSlit image than in its TESL's.

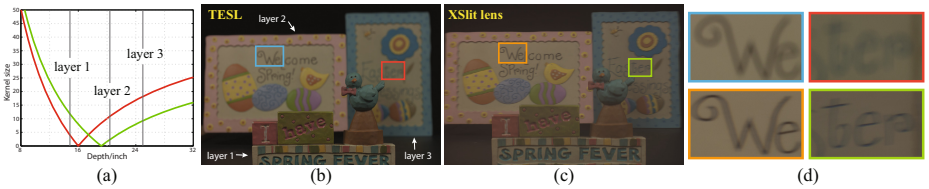


Fig. 6. DoF comparison between the XSlit lens and its TESL. (a) PSF scale vs. depth curve; (b) and (c) are the images captured by the TESL and XSlit lens respectively; (d) shows close-up views in (b) and (c).

Coded Aperture Imaging. Next, we demonstrate coded aperture imaging on the XSlit lens for scene reconstruction. Same as existing solutions, we first conduct experiments on synthetic data to compare the performance of XSlit coded aperture vs. its TESL's. We construct a simple scene with three depth layers at 20cm, 40cm and 80cm. The XSlit lens focuses at 16cm and 24cm. The TESL's focal plane coincides with the front focal plane of the XSlit. For the TESL, we use the optimal coded pattern presented [5] which is expected to have high depth discrepancy. For our XSlit lens, we use the codes described in Section 3.1.

Fig. 7 shows the recovered depth map and all-focus image using the two lenses. For image deconvolution, we apply IRLS (first column) and Wiener filters (the other two columns). Coded aperture result on the TESL using the optimal code [5] produces a high quality depth map although the recovered all-focus image exhibits ring artifacts near occlusion boundaries due to zero-crossings in the frequency domain, as discussed in Section 3.1. Our XSlit coded aperture solution is able to reduce the ringing artifacts thanks to smaller horizontal PSFs and special coding schemes and its recovered depth map is comparable to the TESL's.

Finally, we use our XSlit lens to acquire real scenes. We printed the coded pattern on clear masks and insert them on the aperture plane of each cylindrical lens. The scene consists of three cards at depth 50cm, 80cm, and 100cm respectively. The front lens is focused at 30cm and the rear lens at 60cm. We use the same code as in the synthetic case and apply graph-cut based depth estimation. In particular, we discretize depth to 10 labels ranging from 30cm to 120cm. We segment the captured defocus blurred image into 10×10 patches, compute the Gabor response for each depth layer, and find the optimal depth labeling. Our XSlit coded aperture imaging is able to recover satisfactory depth maps and all-focus images as shown in Fig. 8. The blocky effect in our depth output is partially due to the large window sizes.

5 Discussions and Future Work

We have presented an XSlit depth-of-field (DoF) analysis and an XSlit-based coded aperture imaging solution. Despite being highly theoretical, we have showcased that an XSlit lens is potentially advantageous than its throughput equivalent spherical lens: when the two types of lenses have a similar throughput, the XSlit lens will exhibit better DoF and hence can benefit coded aperture imaging. A major limitation of our approach though is the challenges in constructing high quality XSlit lenses. While the techniques for fabricating lenses that effectively emulate spherical thin lenses have matured in the past century, relatively little work has focused on designing high quality cylindrical lenses and let alone XSlit lenses. In fact, our self-constructed XSlit exhibits high distortions, poor anti-reflection, and strong vignetting. It is our important future work to work with optical engineers to fabricate a larger scale XSlit lens for experimenting our coded aperture scheme.

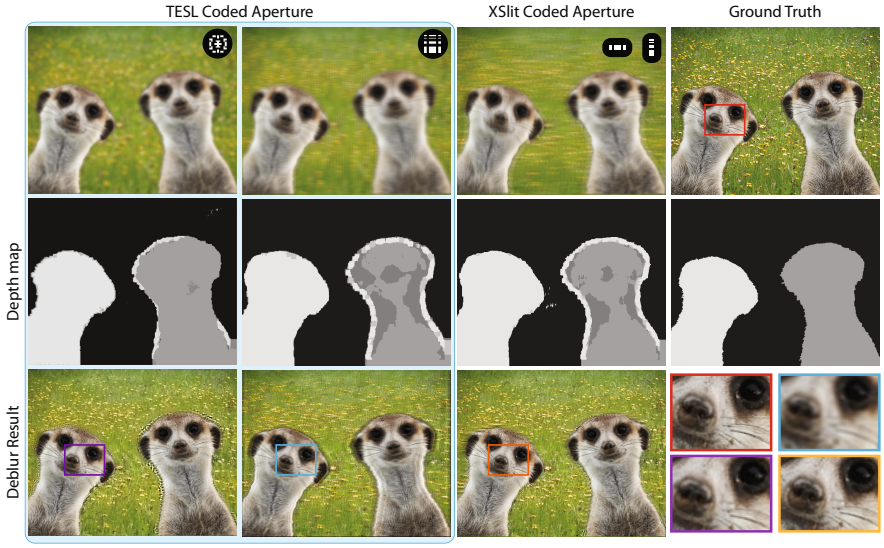


Fig. 7. Coded aperture result on a synthetic scene. We compare our XSlit coded aperture and its TESL with two different coded patterns: on the left we use the coding scheme in [5] and on the right we use the coding scheme that is a combination of the XSlit codes.

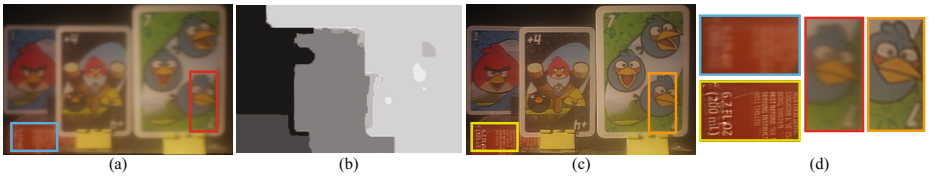


Fig. 8. Coded aperture result on a real scene. (a) The captured defocused image using the proposed XSlit coded aperture scheme; (b) Our recovered depth map; (c) Our recovered all-focus image; (d) Close-up views in (a) and (c).

Another future direction we plan to explore is scene-dependent coded aperture designs. A natural scene often contains patterns that exhibit strong directional features, e.g., a scene can contain mostly horizontal features and few vertical ones. We therefore can potentially encode the vertical and horizontal slit apertures differently to simultaneously account for depth estimation and defocus compensation. Finally, our analysis shows that XSlit defocus blurs different significantly from the ones of spherical lenses: the PSFs vary both in size and shape at different depths. In particular, the shape of the kernel can help validate depth hypothesis in scene recovery. In the future, we plan to explore reliable solutions to integrate this useful depth cue into scene recovery.

Acknowledgements. This project was supported by the National Science Foundation under grants IIS-CAREER-0845268.

References

1. Arguello, H., Arce, G.R.: Code aperture optimization for spectrally agile compressive imaging. J. Opt. Soc. Am. A 28(11), 2400–2413 (2011)
2. Boykov, Y., Veksler, O., Zabih, R.: Efficient approximate energy minimization via graph cuts. IEEE Transactions on Pattern Analysis and Machine Intelligence 20(12), 1222–1239 (2001)
3. Ding, Y., Xiao, J., Yu, J.: A theory of multi-perspective defocusing. In: The IEEE Conference on Computer Vision and Pattern Recognition (2011)
4. Krishnan, D., Fergus, R.: Fast image deconvolution using hyper-laplacian priors. In: Advances in Neural Information Processing Systems (2009)
5. Levin, A., Fergus, R., Durand, F., Freeman, W.T.: Image and depth from a conventional camera with a coded aperture. In: ACM SIGGRAPH (2007)
6. Levoy, M., Hanrahan, P.: Light field rendering. In: ACM SIGGRAPH (1996)
7. Nagahara, H., Zhou, C., Watanabe, T., Ishiguro, H., Nayar, S.K.: Programmable aperture camera using LCoS. In: Daniilidis, K., Maragos, P., Paragios, N. (eds.) ECCV 2010, Part VI. LNCS, vol. 6316, pp. 337–350. Springer, Heidelberg (2010)
8. Ng, R.: Fourier slice photography. In: ACM SIGGRAPH (2005)
9. Pajdla, T.: Geometry of two-slit camera. Tech. Rep. CTU-CMP-2002-02, Czech Technical University (2002)
10. Ponce, J.: What is a camera? In: The IEEE Conference on Computer Vision and Pattern Recognition (2009)
11. Raskar, R., Agrawal, A., Tumblin, J.: Coded exposure photography: Motion deblurring using fluttered shutter. In: ACM SIGGRAPH (2006)
12. Seitz, S.M., Kim, J.: The space of all stereo images. IJCV 48(1), 21–38 (2002)
13. Takeda, Y., Hiura, S., Sato, K.: Fusing depth from defocus and stereo with coded apertures. In: Computer Vision and Pattern Recognition (2013)
14. Veeraraghavan, A., Raskar, R., Agrawal, A., Mohan, A., Tumblin, J.: Dappled photography: Mask enhanced cameras for heterodyned light fields and coded aperture refocusing. In: ACM SIGGRAPH (2007)
15. Ye, J., Ji, Y., Yu, J.: Manhattan scene understanding via XSlit imaging. In: The IEEE Conference on Computer Vision and Pattern Recognition (2013)
16. Ye, J., Ji, Y., Yu, J.: A rotational stereo model based on XSlit imaging. In: The IEEE International Conference on Computer Vision (2013)
17. Ye, J., Yu, J.: Ray geometry in non-pinhole cameras: A survey. The Visual Computer 30(1), 93–112 (2014)
18. Yu, J., McMillan, L., Sturm, P.: Multi-perspective modelling, rendering and imaging. Eurographics 29(1), 227–246 (2010)
19. Yu, J., Ding, Y., Mcmillan, L.: Multiperspective modeling and rendering using general linear cameras. Communications in Information & Systems 7(4), 359–384 (2007)
20. Yuan, S., Sasian, J.: Aberrations of anamorphic optical systems. ii. primary aberration theory for cylindrical anamorphic systems. Appl. Opt. 48(15), 2836–2841 (2009)
21. Zhou, C., Nayar, S.K.: What are Good Apertures for Defocus Deblurring? In: IEEE International Conference on Computational Photography (2009)
22. Zhou, C., Lin, S., Nayar, S.: Coded Aperture Pairs for Depth from Defocus and Defocus Deblurring. IJCV 93(1), 53 (2011)
23. Zomet, A., Feldman, D., Peleg, S., Weinshall, D.: Mosaicing new views: The crossed-slits projection. IEEE Transactions on Pattern Analysis and Machine Intelligence 25(6), 741–754 (2003)