Scene-adaptive Coded Apertures Imaging

Xuehui Wang\textsuperscript{1}, Jinli Suo\textsuperscript{2}, Jingyi Yu\textsuperscript{3}, Yongdong Zhang\textsuperscript{1}, and Qionghai Dai\textsuperscript{2}

\textsuperscript{1}Institute of Computing Technology, Chinese Academy Of Sciences
\textsuperscript{2}Department of Automation, Tsinghua University
\textsuperscript{3}University of Delaware

Abstract

Coded aperture imaging systems have recently shown great success in recovering scene depth and extending the depth-of-field. The ideal pattern, however, would have to serve two conflicting purposes: 1) be broadband to ensure robust deconvolution and 2) has sufficient zero-crossings for a high depth discrepancy. This paper presents a simple but effective scene-adaptive coded aperture solution to bridge this gap. We observe that the geometric structures in a natural scene often exhibit only a few edge directions, and the successive frames are closely correlated. Therefore we adopt a spatial partitioning and temporal propagation scheme. In each frame, we address one principal direction by applying depth-discriminative codes along it and broadband codes along its orthogonal direction. Since within a frame only the regions with edge direction corresponding to its aperture code behaves well, we utilize the close among-frame correlation to propagate the high quality single frame results temporally to obtain high performance over the whole image lattice. To physically implement this scheme, we use a Liquid Crystal on Silicon (LCoS) microdisplay that permits fast changing pattern codes. Firstly, we capture the scene with a pinhole and analyze the scene content to determine primary edge orientations. Secondly, we sequentially apply the proposed coding scheme with these orientations in the following frames. Experiments on both synthetic and real scenes show that our technique is able to combine advantages of the state-of-the-art patterns for recovering better quality depth map and all-focus images.

1. Introduction

In the past decade, computational photography has emerged as a vibrant field of research. Rapid progress in this area has brought forth new imaging systems to facilitate easy and robust extraction of scene information. A notable example is the coded aperture imaging system. The idea can be backdated to 1965 when it was used to identify the first cosmic X-ray source. The technique has since been widely employed in astronomy to improve the signal-to-noise ratios of lensless imaging detectors for x-ray and gamma-ray sources. In optics, it can be used to encode spatial information into Point Spread Function (PSF) for depth estimation. Aslantas \cite{1} treats the aperture as a two-dimensional Gaussian function whose spread parameter is related to the object depth. Dowski and Cathey\cite{4} introduced the wavefront coding by modulating the aperture using a 3D phase plate. Their generated PSF is depth invariant, and the sharp all-focus image can be recovered using spatially invariant deconvolution.

More recent coded aperture techniques exploit depth-variant PSFs whose imaging process can be essentially viewed as depth-dependent convolution: the result is generated by convolving a sharp, all-focus image with different scales of a fixed aperture PSF. The scaled PSFs correlate the frequency characteristics of the blur kernel with scene depth so that one can apply special deconvolution algorithms to simultaneously reduce blurs and recover the scene. The technique closely resembles depth-from-defocus (DfD) \cite{5} except that it only needs one input image.

Coded Aperture Design: A Dilemma. Tremendous efforts have been focused on designing good coded patterns. However, an ideal pattern will have to have two conflicting properties: reliable deconvolution and high depth discrepancy. For the first, the aperture pattern should be broadband and have few zero-crossings in the frequency domain, to allow robust deconvolution for better preserving high frequency components such as edge features. For example, Veeraraghavan \textit{et al}. \textsuperscript{15} used such coded patterns to maximize the minimum spectrum amplitude. However, a broadband kernel has poor depth discrimination capability and using it for depth estimation is suboptimal. For the latter, Levin \textit{et al}. \textsuperscript{9} have shown that a good depth
discrepancy pattern should have multiple zero-crossings in the frequency domain. These zero-crossings map different depth layers to kernels with significantly different frequency characteristics. Coded aperture patterns as such have shown great success in recovering scene depth. However, deconvolving these kernels is challenging and previous solutions have resorted to special and computationally expensive algorithms such as the iterative re-weighted least squares (IRLS).

In general, it is difficult to simultaneously achieve high-accuracy depth estimation and high-quality deblurring using a single coded aperture. To address that, Hiura and Matsuyama [7] used a pair of pinhole apertures on a multi-focus camera to recover scene depth. However, the pinhole pattern is notorious for its low light efficiency. Zhou et al. [17] used a pair of asymmetric half-ring aperture patterns whose frequency responses are complementary while each has sufficient zero-crossings. Although highly effective, their approach requires capturing two images at the same viewpoint and thus cannot handle dynamic scenes. In the same vein, Takeda et al. [14] used Zhou’s patterns on a pair of cameras focusing at different depths to simultaneously conduct stereo matching and DfD. However, the use of a stereo system is less practical and desirable.

Another issue with existing coded aperture imaging systems is that the code is chosen independent of scene. Consider a scene that contains mostly horizontal edge features. Horizontal blurs will provide less useful information in depth discrepancy than vertical blurs. This indicates that a scene-adaptive coded aperture is most desirable. A notable exception is the noise-dependent coded aperture by Zhou et al. [17], although the goal is to improve the signal-to-noise ratio of the reconstruction rather than accuracy and speed in deconvolution. Further, their techniques rely on sophisticated and computationally optimization techniques (i.e., genetic algorithms) to find the suitable codes.

Our Approach. This paper presents a new scene-adaptive coded aperture solution for dynamic imaging. We observe that the edges in a natural image are of only a few principal orientations, which can be caused by structure, texture, shading, etc. In Fig. 1(a), we randomly select 12 images from a large collected database including 100 Internet images; the cumulative distribution of whose edge direction number is plotted in Fig. 1(b), in which we discretize the edge orientations into nine 20° sectors. Considering that adjacent video frames are closely correlated, we propose to design codes addressing one principal direction and then make use of the temporal redundancy to obtain high performance across a whole nature image.

For the regions with one dominant direction, we propose to apply depth-discriminative codes along it and broadband codes along its orthogonal counterpart, with the aperture pattern and its frequency distribution shown in Fig. 1(c).
Although depth-discriminative code contains multiple zero-crossings and therefore can introduce strong deconvolution artifacts, these artifacts will occur along the dominant direction where image gradients are small. Therefore, they are less perceivable to the human eyes. Orthogonal to the dominant edge direction, the gradients are much larger and the use of broadband codes ensures minimal deconvolution artifacts. Fig. 1(d) and (e) shows the depth estimation and deconvolution results using our code vs. the state-of-the-art codes respectively. Results show that our code combines the advantages of both the broadband and high discrepancy codes and it produces high depth accuracy comparable to [9] and high image quality comparable to [17].

To physically implement our spatial partitioning and temporal propagation scheme, we snap \( N + 1 \) shots for a dynamic scene with \( N \) dominant directions, using a Spatial Light Modulator (SLM) consists of a high fill factor Liquid Crystal on Silicon (LCOS) [11][12] that permits fast code pattern changes. In the first shot we capture the scene with a pinhole aperture and conduct ultra-fast analysis of the scene content to determine the edge directions. In the following, we acquire a sequence of defocused images by applying our proposed coding scheme with respect to one of the corresponding principal orientations in each shot. Our multiple-shot solution differentiates from other two-shot/multiple-shot solutions [16] in two main aspects: We use the first shot to guide the afterwards shots and utilize the temporal redundancy among successive video frames. Finally, we develop robust deconvolution algorithms to recover the all-focus images and estimate scene depths, and then combine them together to generate high performance result across the whole frame sequence. We validate our technique on a broad range of synthetic and real scenes.

2. Aperture Designs

We start with discussing how to design the coded pattern for regions with one dominant direction such as Fig. 2(a), and the design can be applied to close correlated successive frames to handle general nature scenes with multiple principal directions. Along the dominant direction the intensity varies slightly and artifacts in this direction is imperceivable. Oppositely, there exist large and rich intensity changes along the orthogonal direction and frequency lost in this direction will cause large degeneration in visual quality. Therefore, we attempt to design an aperture pattern with broadband frequency along the orthogonal direction, while of high depth discrimination power at expense of refocusing performance along the dominant direction. For convenience, we denote the dominant direction as \( y \)− and its perpendicular direction as \( x \)− in this paper. Correspondingly, the \( y \)− and \( x \)− slices of image \( I \) are respectively represented as \( I_y \) and \( I_x \).

2.1. Depth Discriminative 1D Code

Assume \( I_y \)’s derivative follows a zero mean Gaussian distribution

\[
p(I_y) \propto \prod_i e^{-\frac{1}{2} (I_y(i) - I_y(i+1))^2} = N(0, \Psi),
\]

where \( i \) indexes the entries in \( I_y \), \( \Psi^{-1} = \lambda C_f^T C_f \) with \( C_f \) being the convolution matrix corresponding to the differential kernel \( f = [1 \ - \ 1] \), i.e., \( C_f I_y \equiv I_y \otimes f \). We can interpret the prior above in the frequency domain:

\[
p(I_y) \propto e^{-\frac{1}{2} \tilde{\Psi} \tilde{\tau_y}^T \tilde{\tau_y}}
\]

where \( \tilde{\Psi}^{-1} = \lambda \text{diag}(|G_f|^2) \) is diagonal.

Let \( p(I_{y,k}) \) be the distribution of an observed blurred version of signal \( I_y \) under blur \( k \), i.e., \( I_{y,k} = I_y \otimes k + n \). Here \( n \sim N(0, \eta^2) \) is the sensor noise. Since the blur kernel \( k \) is a linear operator, \( p(I_{y,k}) \) remains as Gaussian:

\[
p(I_{y,k}) \sim N(0, \Phi_k).
\]

The covariance matrix \( \Phi_k = C_k \Psi C_f^T + \eta^2 I \) transforms the original one \( \Psi \), and \( I \) denotes the identity matrix. We can map \( \Phi_k \) into the frequency domain and obtain a diagonal covariance matrix \( \tilde{\Phi}_k = \text{diag}(G_k) \tilde{\Psi} \text{diag}(G_k)^T + \eta^2 I \). The Gaussian distribution of the blurred signal in Fourier domain becomes:

\[
p(I_{y,k}) \propto e^{-\frac{1}{2} \sum u \tilde{\Psi}_{y,k}(u)/\sigma_k(u)},
\]

where \( \sigma_k(u) \) is \( u \)-th diagonal entry of \( \tilde{\Phi}_k \):

\[
\sigma_k(u) = |G_k(u)|^2 (\lambda |G_f(u)|^2)^{-1} + \eta^2.
\]

Levin et al.[9] have shown that the depth discriminative aperture codes should ensure that the blurred images at different blur levels are statistically different. We quantitatively measure this difference using the Kullback-Leibler (KL) divergence:

\[
D_{KL}(p(I_{y,k_1})||p(I_{y,k_2})) = \sum_1 p(I_{y,k_1}) \left( \log \frac{p(I_{y,k_1})}{p(I_{y,k_2})} \right).
\]

Since \( p(I_{y,k_1}), P(I_{y,k_2}) \) are both Gaussian, using the frequency domain representation, we can simply Eqn. 6 as:

\[
D_{KL}(p(I_{y,k_1})||p(I_{y,k_2})) = \sum \left( \frac{\sigma_{k_1}(u)}{\sigma_{k_2}(u)} - \log \frac{\sigma_{k_1}(u)}{\sigma_{k_2}(u)} \right).
\]

Next, we traverse all of the depth pairs to search for an optimal depth discriminative 1D binary code. Fig. 2(c) shows our resulting 1D code and its frequency distribution.
2.2. Weighted Broadband Frequency 1D Code

Along the $x$-direction, we search for a 1D broadband kernel. Recall that such kernels resemble the one used in fluttered shutter for motion blur removal [13]. Specifically, the goal in fluttered shutter is to encode the shutter sequence to reduce the loss of high frequency information. A good code in the frequency domain should have a large minimal magnitude and be of small fluctuations. Although good at high quality deblurring, such codes are not directly applicable to our task. This is because fluttered shutter assumes constant velocity motions, i.e., uniform frequency distribution whereas there exist apparent non-uniform frequency distributions in natural images, as shown in Fig. 2(b). In fact, the optimal kernel for constant velocity motion is suboptimal for acceleration or harmonic motions [3].

Statistically the frequency of natural images, either with or without a dominant direction, concentrates around zero and decays towards the boundary. Therefore, a code preserving the lower frequency with a higher priority is more preferable. We therefore adopt a Gaussian weighting scheme to combine the optimal codes at different velocities in [13], and the parameters of the Gaussian weight is determined by $I_x$’s frequency. The frequency of the designed broadband code is plotted as a blue solid line in Fig. 2(c).

Finally, we can combine the two 1D kernels together. Recall that our target aperture’s marginal distributions along the two orthogonal directions should be similar to the two 1D codes discussed above. We therefore simply generate a 2D aperture pattern as the outer product of two 1D codes, as shown in Fig. 2(d) and Fig. 1(c).

In Fig. 2(d), we observe that the designed aperture pattern is of rich zero crossings along the dominant direction while broadband frequency along the orthogonal direction. Fig. 2(e) compares the frequency of our aperture pattern with those of previously proposed ones from two vertical slices, i.e., in terms of depth estimation and refocusing. The richness of zero-crossing is comparable to Levin’s [9], while the frequency profile along the dominant direction is similar to the one of Veeraraghavan et al. [15].

3. Scene-adaptive Coded Aperture Imaging

Recall that both depth inference and refocusing from a single defocused image are ill-posed where strong priors are necessary for reliably solve the problem. For example, one cannot distinguish between a defocused image of sharp texture and a focused image of smooth texture. Therefore, we present a two-shot solution, the first with a pinhole aperture for determining scene structure and hence the code, the second with the tailored coded aperture for depth estimation and refocusing. The first shot also provides important priors
for the second.

We firstly capture a noisy but sharp pinhole image $I_0$ for scene structure analysis. To reduce noise, we adopt widely used Wiener filtering for denoising, which is of sufficiently high efficiency, to generate $\hat{I}_0$. $\hat{I}_0$ will lose some fine details due to denoising but should still preserve a sufficient number of large scale edges/structures for scene structure analysis. We can obtain the principal directions of the scene and group the pixels into several corresponding regions with a dominant direction. Next, for each region(s) with a specific dominant direction, the frequency distribution along the orthogonal direction is calculated, from which we can determine the weights of broadband code, and compute its outer product with the depth discriminative code as a 2D aperture pattern. Finally the generated 2D patterns are rotated to ensure that the depth discriminative direction coincides the dominant direction of its corresponding region.

Although a two shot strategy is beneficial for both depth estimation and all-focus image recovery, there exist apparent misalignment between the pinhole shot $I_0$ and encoded shot, we denote which as $I_2$. To address this problem, we adopt the motion estimation model proposed by Chambolle and Pock [2] to compensate the object motion. Experiments show that the algorithm gives fairly good motion estimation, and Pock [2] to compensate the object motion. Experiments show that the algorithm gives fairly good motion estimation, even one frame is blurred.

Let the motion compensated pinhole image as $I_1$, the depth estimation and refocusing can be performed by combining the pinhole image $I_1$, aperture pattern $K$ and the second acquired image $I_2$. Recall that $I_1$ and $I_2$ will have different intensity levels due to aperture sizes. We therefore scale $I_1$ with the aperture ration to make the two images have approximately the same intensity level.

### 3.1. Depth Calculation

Denoting the latent sharp image as $I$, we have

$$I_1 = I + \eta_1,$$

where $\eta_1$ is the sensor noise. Here we adopt the widely used noise model in time constrained photography proposed in [6]. The noise is zero mean Gaussian with a constant variance over the image, and it is a combination of a constant additive read noise and a multiplicative photon noise.

Denoting the aperture pattern as $K$, the blur kernels at different depths should be $K$’s scaled versions. For a planar object that incurs $d$-pixel blur kernel, denoted as $K_d$, the coded defocused image $I_2$ can be written as:

$$I_2 = I \otimes K_d + \eta_2,$$

where $\eta_2$ is the sensor noise which is zero-mean Gaussian. Replacing $I$ in Eqn. 9 with $I_1$ and introducing regularization term that penalizes abrupt depth changes, we can get an energy function as

$$E_d = \|I_1 \otimes K_d - I_2\|_2 + \gamma \|\nabla d\|_1$$

with $\gamma$ being the weighting factor.

To improve the applicability in low-texture or textureless regions, we assign a spatially varying confidence map to the image lattice to account for the richness of textures (gradient magnitude). We have the new energy function as:

$$E_d = (1 - W) \cdot \|I_1 \otimes K_d - I_2\|_2 + W \cdot \gamma \|\nabla d\|_1,$$  \hspace{1cm} (11)

where $W = \exp(-\nabla I_1)$ forces low confidence in the depth estimation in uniform color regions. Further, the dot product can be rewritten in the form of sparse matrix multiplication. The objective function above is convex and can be solved via standard convex optimization.

For a non-planar scene, we search for the depth map $d$ that minimizes above energy

$$d^* = \arg \min_{d \in D} ((1 - W) \cdot \|I_1 \otimes K_d - I_2\|_2 + W \cdot \gamma \|\nabla d\|_1),$$  \hspace{1cm} (12)

where $D = \{d_1, ..., d_l\}$ is the set of $l$ discretized depths.

### 3.2. Deconvolution

We can further compute an all-focus image from the coded aperture image. Specifically, we use the maximum a posteriori (MAP) to find the all-focus image $I$ under the estimated depth $d$. The process resembles the noisy/blurry pair solution in [16]:

$$p(I|I_1, I_2, d, K) \propto \max p(I_1|I)p(I_2|I, d, K)p(I).$$  \hspace{1cm} (13)

From Eq. 8 and 9, we can compute the likelihood (under the Gaussian model) of $I$ with respect to the noisy image $I_1$ and the defocused image $I_2$ as:

$$p(I_1|I) \propto \exp\left\{-\frac{1}{2\sigma^2}\|I_1 - I\|^2\right\}$$  \hspace{1cm} (14)

$$p(I_2|I, d, K) \propto \exp\left\{-\frac{1}{2\sigma^2}\|I_2 - I \otimes K_d\|^2\right\}.$$

We also incorporate the prior of $I$ whose gradients $\nabla I$ should follow a zero mean Hyper-laplace distribution[8]:

$$p(I) \propto \exp\left\{-\frac{1}{2}\|\nabla I\|^{0.5}\right\},$$  \hspace{1cm} (15)

in which $\nabla$ is the gradient operator.

Finally, we take log on above three terms and form the combined energy function as:

$$E_I = \|I - I_1\|_2 + \alpha \|I \otimes K_d - I_2\|_2 + \beta \|\nabla I\|^{0.5}. $$

The first term aims to preserve details, the second enforces consistency with the observations and the third penalizes the noise and ringing artifacts. We use parameter $\alpha, \beta$ to balance the three terms and apply the alternating minimization algorithm [8] to compute the optimal solution $I$. 


It is important to note that the optimization scheme above is convex only for spatially-invariant blur kernels (i.e., constant depth). To handle multiple depth layers, we recover the regions at different depths separately and combine them together, i.e.,

$$I^* = \sum_{d \in D} \arg \min \pi_d(E_I),$$  

where $\pi_d$ is a linear mapping for selecting the regions at depth $d$.

3.3. Orientation Guided Temporal Propagation

Since each shot is taken with a coded aperture rotated towards one principal direction, each recovered all-focus frame and corresponding estimated depth map only exhibit better performance in the regions corresponding to its aperture pattern than the other shots. Fortunately, natural videos are temporal redundant and results with good performance across the image lattice can be obtained by fusing results from neighboring frames.

Denoting a temporal window centered at frame $t$ as $\mathcal{N}(t)$, we have

$$d^t = \sum_{t' \in \mathcal{N}(t)} \mathcal{M}_{t' \rightarrow t}(C^t_{d'} \odot d^t), \quad I^t = \sum_{t' \in \mathcal{N}(t)} \mathcal{M}_{t' \rightarrow t}(C^t_{I} \odot I^t).$$

(18)

Here $\mathcal{M}_{t' \rightarrow t}$ is the transformation matrix between frame $I^t$ and its neighboring frame $I^{t'}$, and can be calculated by optical flow [10]; $C^t_{d'}$ and $C^t_{I}$ are the weight matrices for all-focus image and depth map of frame $t'$, we calculate them using the cosine and sine distance between the image edge direction and pattern orientation, respectively; $\odot$ denotes a dot product between two vectors/matrices of the same dimension. For a specific position, the weight coefficients within $\mathcal{N}(t)$ are normalized to ensure consistent pixel intensity, i.e., $\sum_{t \in \mathcal{N}(t)} C^t = 1$. The weight matrices of the red-framed exemplary scene in Fig. 1(a) are visualized in Fig. 3, from which we can see that the confidence map matches above theoretic analysis quite well. Note that for clearer display, the weights for the flat regions are set to be zero, instead of its true value 0.25.

4. Experiment

4.1. Synthetic Experiments

We first conduct experiments on synthetic toy data to evaluate our aperture coding scheme. Since the effectiveness of propagation utilizing temporal redundancy of dynamic scenes is well known, so in this section we just use static scenes directly, i.e., pre-aligned successive frames.

To simulate a scene with multiple principal directions, the frame is generated by rotating a floor-texture with one dominant direction towards 4 directions ($0^\circ$, $45^\circ$, $90^\circ$, and $135^\circ$) and each composites one vertical stripe of the all-focus image, as shown in Fig. 4(a). To illustrate stability our designed pattern on different depths, we designed a stair-like depth map visualized in Fig. 4(b), in which the maximum blur kernel size is about 15 pixels while the minimum one is about 3 pixels. Then, we defocus this all-focus image to generate four defocused shots by convolving with our coded pattern in above 4 directions. The depth map and all-focus image are calculated using the algorithms in Sec. 3, as shown in Fig. 4(c)-(d).

Fig. 4 (e)-(h) respectively shows the depth estimated from each single shot. Comparing with (b), we can see that when our designed coded pattern is in parallel with the edge direction, the depth accuracy degenerates slightly because the broadband frequency along the orthogonal direction of the aperture code is of low depth discrimination power while lacking intensity variation along the dominant direction limits the depth discrepancy of the aperture code. Oppositely, few zero-crossing along the orthogonal direc-
tion ensures that the recovered all-focus image is of little residue at the same setting, as shown in (i)-(l). The results provide an intuitive comprehension of the performance of the proposed coding scheme: each shot can produces high performance in only some regions, either for depth estimation or all-focus recovery. By combining image fusion technology described in Sec. 3, we can achieve superb depth map and all-focus image result for nature scenes with multiple orientations.

4.2. Real Implementation

To physically implement our scene-adaptive coded aperture setup, we construct a dynamic coded aperture system similar to the one used in [12], as shown in Fig. 5. Our system consists of several off-the-shelf elements: a primary lens (Cannon 50mm/1.8#), a pair of relay lenses (Nikon 50mm/1.8#), a Holoeye LC-R 1080 Spatial Light Modulator (Reflective, 1920×1200 pixels, 0.7”), and a Point Grey Grasshopper sensor (2/3” CCD, pixel size 3.45 µm, 2448×2048 pixels at 15fps). The spatial resolution of the designed aperture pattern is 54×54, and we treat 20×20 LCoS units as a superpixel and use in all 1080×1080 LCoS pixels to maximize the light efficiency.

![Figure 5. Our scene-adaptive coded aperture imaging setup using LCoS, whose close-up is displayed in the bottom left corner, and the corresponding light path (overlaid).](image)

Since the prototype so far is not portable and the data capturing is limited in indoor scenes, we build one dynamic “street” scene that bears similar structure of real scenes. The image is 860×590 pixels, and the size of the blur kernels roughly varies from 1 to 13 pixels. Fig. 6 shows the inputs (noisy but sharp pinhole image in (a) and coded blurry frames in (e)), the distribution of the dominant directions (b) and the results, including a recovered depth map in (c) and all-focus frame in (d). The single-frame results from four sequential frames are displayed as well, in (f) and (g), respectively. The conclusions are coherent with the ones from the synthetic data: The results of each single shot displays good depth estimation accuracy in the regions with dominant direction parallel with the aperture code and excellent deconvolution results in the regions with orthogonal dominant direction; Making use of temporal redundancy of nature videos promises good performance over the whole image lattice. We refer the readers to the supplementary materials for more results and video demonstration.

5. Conclusions and Discussions

We have presented the first scene-adaptive coded aperture scheme to address the dilemma in coded pattern design. Our proposed idea is very simple but effective: assume that the scene exhibits a small number of principal directions, we take multi-shot strategy, by using the high depth discrepancy code along one of the primal direction while using the broadband code along its orthogonal direction. We have tailored algorithms for robust depth map estimation and high quality all-focus image recovery. Comprehensive experiments on both synthetic and real data show that our approach is able to come advantages of the two different coding patterns and outperform state-of-the-art solutions.

If the scene exhibits no structure (e.g., grass, random texture, etc.), our code scheme may not perform well. In the future, we plan to explore the problem of designing scene-adaptive coded patterns in the frequency domain, i.e., even if the scene does not exhibit obvious spatial structures, it may exhibit strong frequency structures. We may also resort to color coded apertures, which encode each channel differently with respect to a direction, and implement scene adaptive aperture coding applicable for static nature images. Finally, our work is inline with the emerging interest in designing motion-adaptive coded shutter[3] and it will be highly interesting to designing a combined motion- and scene-adaptive shutter sequences.

References

[1] V. Aslantas. A depth estimation algorithm with a single image. Optics express, 15(8):5024–5029, 2007.
[2] A. Chambolle and T. Pock. A first-order primal-dual algorithm for convex problems with applications to imaging. Journal of Mathematical Imaging and Vision, 40(1):120–145, 2011.
[3] Y. Ding, S. McCloskey, and J. Yu. Analysis of motion blur with a flutter shutter camera for non-linear motion. In Proceedings of European Conference on Computer Vision, pages 15–30, Crete, Greeceee, 2010.
[4] E. R. Dowski Jr and W. T. Cathey. Extended depth of field through wave-front coding. Applied Optics, 34(11):1859–1866, 1995.
[5] P. Grossmann. Depth from focus. Pattern Recognition Letters, 5(1):63–69, 1987.
[6] S. W. Hasinoff, K. N. Kutulakos, F. Durand, and W. T. Freeman. Time-constrained photography. In Proceedings of IEEE International Conference on Computational Vision, pages 333–340, Kyoto, Japan, 2009.
Figure 4. Results on a toy example. (a) A sharp image with four principal directions. (b) Synthetic depth map of the image in (a). (c)-(d) The estimated depth and residue of all-focus recovery unifying four aperture codes addressing four principal directions respectively. (e)-(h) and (i)-(l) The estimated depth and residue of all-focus recovery by using four differently oriented aperture codes.

[7] S. Hiura and T. Matsuyama. Depth measurement by the multi-focus camera. In Proceedings of IEEE Conference on Computer Vision and Pattern Recognition, pages 953–959, Santa Barbara, CA, 1998.

[8] D. Krishnan and R. Fergus. Fast image deconvolution using hyper-laplacian priors. In Advances in Neural Information Processing Systems, pages 1033–1041, Vancouver, B.C., 2009.

[9] A. Levin, R. Fergus, F. Durand, and W. T. Freeman. Image and depth from a conventional camera with a coded aperture. ACM Trans. Graphics, 26(3):70, 2007.

[10] C. Liu. Beyond Pixels: Exploring New Representations and Applications for Motion Analysis. PhD thesis, Massachusetts Institute of Technology, May 2009.

[11] H. Mannami, R. Sagawa, Y. Mukaigawa, T. Echigo, and Y. Yagi. High dynamic range camera using reflective liquid crystal. In Proceedings of IEEE International Conference on Computational Vision, pages 1–8, Rio de Janeiro, Brazil, 2007.

[12] H. Nagahara, C. Zhou, T. Watanabe, H. Ishiguro, and S. Nayar. Programmable aperture camera using LCoS. In Proceedings of European Conference on Computer Vision, pages 337–350, Crete, Greece, 2010.

[13] R. Raskar, A. Agrawal, and J. Tumblin. Coded exposure photography: motion deblurring using fluttered shutter. ACM Trans. Graphics, 25(3):795–804, 2006.

[14] Y. Takeda, S. Hiura, and K. Sato. Fusing depth from defocus and stereo with coded apertures. In Proceedings of IEEE Conference on Computer Vision and Pattern Recognition, pages 209–216, Portland, OR, 2013.

[15] A. Veeraraghavan, R. Raskar, A. Agrawal, A. Mohan, and J. Tumblin. Dappled photography: Mask enhanced cameras for heterodyned light fields and coded aperture refocusing. ACM Trans. Graphics, 26(3):69, 2007.

[16] L. Yuan, J. Sun, L. Quan, and H.-Y. Shum. Image deblurring with blurred/noisy image pairs. ACM Trans. Graphics, 26(3):1, 2007.

[17] C. Zhou, S. Lin, and S. Nayar. Coded aperture pairs for depth from defocus. In Proceedings of IEEE International Conference on Computational Vision, pages 325–332, Kyoto, Japan, 2009.
Figure 6. The result on one sequence captured by our prototype system. (a) The 1st shot: pinhole image. (b) Distribution of the principal directions. (c)(d) Estimated depth and recovered all-focus image of the 2nd frame by unifying four frames. (e)–(g) The captured encoded sequence, depth estimation and all-focus recovery of four sequential frames, with regions suffering from ringing artifacts shown in closeups.