

Catadioptric HyperSpectral Light Field Imaging

Yujia Xue¹ Kang Zhu^{2,3} Qiang Fu² Xilin Chen³ Jingyi Yu^{1,2}

¹Plex-VR Inc.

yujia.xue@plex-vr.com

²ShanghaiTech University

{zhukang, fuqiang, yujingyi}@shanghaitech.edu.cn

³Institute of Computing Technology, Chinese Academy of Sciences

xlchen@ict.ac.cn

Abstract

The complete plenoptic function records radiance of rays from every location, at every angle, for every wavelength and at every time. The signal is multi-dimensional and has long relied on multi-modal sensing such as hybrid light field camera arrays. In this paper, we present a single camera hyperspectral light field imaging solution that we call Snapshot Plenoptic Imager (SPI). SPI uses spectral coded catadioptric mirror arrays for simultaneously acquiring the spatial, angular and spectral dimensions. We further apply a learning-based approach to improve the spectral resolution from very few measurements. Specifically, we demonstrate and then employ a new spectral sparsity prior that allows the hyperspectral profiles to be sparsely represented under a pre-trained dictionary. Comprehensive experiments on synthetic and real data show that our technique is effective, reliable, and accurate. In particular, we are able to produce the first wide FoV multi-spectral light field database.

1. Introduction

The complete plenoptic function records radiance of rays from every location, at every angle, for every wavelength and at every time. The function hence is 6D: 2D in position (spatial), 2D in direction (angular), 1D in spectrum (spectral), and 1D in time (temporal). Existing techniques have been focused on reconstructing a subset of these dimensions: a static camera for spatial, a video camera for spatial and temporal, a light field camera for spatial and angular, and imaging spectrometer for spatial and spectral. Simultaneously capturing multiple dimensions of the plenoptic function is challenging and successful solutions have generally adopted a hybrid sensing approach: cameras of different modalities are integrated and tailored algorithms are then employed for fusing different dimensions.

A notable example is the multispectral light field cam-



Figure 1. Our Snapshot Plenoptic Imager (SPI) combines a single DSLR camera with an array of spectral coded catadioptric mirrors.

era array[13] where each camera is equipped with a specific band-pass filter and image warping is used to combine the spatial-angular and the spectral dimensions. Such a system can efficiently capture high quality plenoptic function but the solution is cumbersome and expensive. Most recently, [26] constructed a hybrid system composed of a commercial light field camera (Lytro) and a coded aperture spectral imager (CASSI). This imaging system is much smaller than camera array solution in size and potentially portable but it requires complex optical configuration and extremely accurate calibration. In this paper, we present a novel snapshot plenoptic imaging solution by using only a single camera. Instead of constructing a hybrid sensor, we exploit using catadioptric mirror arrays for simultaneously acquiring the spatial, angular and spectral dimensions. We call Snapshot Plenoptic Imager or SPI.

Existing catadioptric systems combine cameras with curved mirrors and have shown promising results in broad applications, ranging from panoramic imaging [2, 17], wide FOV light field imaging and rendering[24] to robust volumetric reconstruction[9]. Our SPI modifies existing cata-

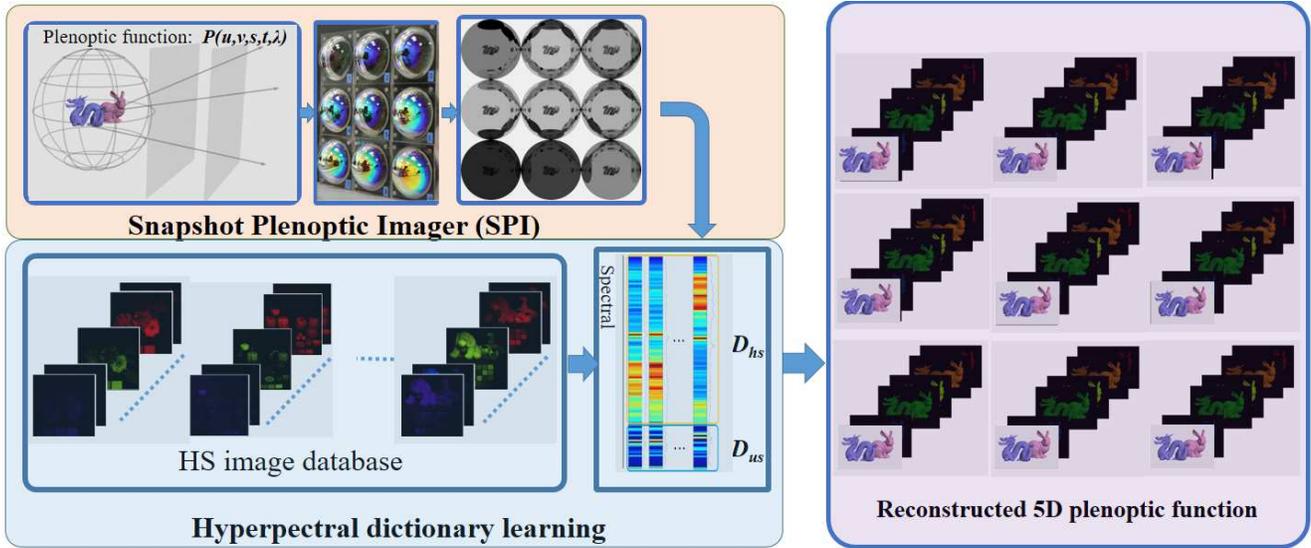


Figure 2. The SPI system samples the plenoptic function in different dimensions and reconstructs the signal using a pre-trained hyperspectral dictionary.

dioptric mirror arrays by coating each mirror with a different spectral filter. The spectral-coded catadioptric mirror array in essence acquires compressively sensed plenoptic function. We apply multi-perspective image registration for aligning individual spectral image on the catadioptric mirror. Figure 1 shows the configuration of our SPI prototype.

Snapshot hyperspectral imaging has long suffered from the trade-off between spatial and spectral resolution. Typical snapshot spectral imager, such as tomographic imaging spectrometer, use diffraction grating to disperse the transmitted light onto the detector. Higher order dispersion patterns provide dense spectral samples but significantly decrease spatial resolution. We hence further apply a learning-based approach to circumvent this limitation. Specifically, we demonstrate and then employ a new spectral sparsity prior that allows the hyperspectral profiles to be sparsely represented under a pre-trained dictionary. The dictionary is obtained from publicly available hyperspectral image database and we reconstruct the signal under the sparsity prior. Figure 2 shows the pipeline of proposed SPI system. We evaluate our solution on both synthetic data and real indoor scenes. Specifically, we generate simulated virtual SPI images as well as use the hyperspectral image benchmark [28]. Quantitative evaluations show that our technique achieves very high accuracy in reconstruction. In particular, since our catadioptric mirror array provides equal-baselined viewpoints, we are able to produce the first wide hyperspectral light field database.

2. Related Work

Hyperspectral Imaging. Hyperspectral imaging provides densely sampled spectral signals and benefits high

quality remote sensing [12], with a wide range of applications in agriculture, military, astronomy, surveillance and *etc.* [3, 6, 8, 7]. Traditional hyperspectral cameras and imaging spectrometers have very limited temporal resolution because they record scene reflectance in a scanning manner over either spatial or spectral domain. Non-scanning methods like computed tomographic imaging spectrometer measure a 2D projection of the 3D datacube but suffer from low spatial resolution. To circumvent these limitations, computational methods and spectral priors are proposed. Recent studies in coded aperture imaging have boosted snapshot spectral imaging by encoding 3D datacube with a coded mask on aperture plane and reconstruct spectral information computationally. To recover scene spectra from a single color image, Brown *et al.* [21] introduced a new strategy of learning a non-linear radial function network between multispectral images and single RGB images. They further managed to recover both scene reflectance and illumination spectra. Similar work of Arad *et al.* [1] recovered scene spectra from RGB images by exploring spectral sparsity. Certain methods trying to retrieve spectra from RGB inputs can not handle metamerism well, meaning different spectral power distributions can have the same RGB response.

Light Field Imaging. Recently there has been an emerging trend of light field photography in both industry and consumer market. By acquiring spatial and angular dimensions of plenoptic function, light field cameras provides novel functionalities including depth estimation, scene recovery, digital refocusing, dynamic depth-of-field and view synthesis. To acquire angular variation, micro-lens array

has been widely used in light field imaging [22]. Off-the-shelf light field cameras (Lytro and Raytrix) both place a lens array above the sensor to record directional rays but lead to low spatial resolution. To improve spatial resolution, Veeraraghavan [25] applied coded mask instead of lens-array on the optical path to sample the 4D plenoptic function in a different pattern. Most recently, Li and Ye [19, 29] acquired light field by non-perspective cameras using rotational X-slits lenses without sacrificing spatial resolution. This novel sampling strategy provides potential advantages on stereo matching, volumetric reconstruction and refocusing effects. Our SPI system exploits spectral-coded catadioptric mirror array to record different views and different bands in a single shot, thus achieving 5D plenoptic imaging.

Sparse Representation. Research in sparse representation has produced impressive improvement in image processing, including image de-noising, restoration and single image super-resolution. In these tasks, the aim is to find a compact over-complete dictionary from a given bunch of training data so that any natural image can be a sparse combination of atoms with respect to the dictionary. To super-resolve a single image, Yang *et al.* [27] jointly learned coupled dictionaries for high resolution image patches and low resolution ones. In this learning strategy, high resolution image patch and its corresponding low resolution one are forced to have the same sparse decomposition in two dictionaries. Inspired by their work, we devise a similar learning method to recover hyperspectral signals from undersampled measurements.

Catadioptric Imaging. Catadioptric imaging systems combine lenses with curved reflective surfaces to capture a wide view of world with a single sensor and have been broadly used in panoramic imaging tasks including surveillance, autonomous vehicles and immersive virtual reality content generation [17]. Baker and Nayar [2] presented a complete and detailed class of single viewpoint catadioptric imaging configurations. To handle non-single viewpoint catadioptric systems and complex reflection surfaces, Yu [30] unified different camera models into a general and linear representation. Catadioptric imaging system is also capable of light field acquisition. Levoy [18] aimed a camera at a planar mirror array to capture light fields and simulated wide aperture photography. Taguchi [24] proposed geometric modeling for non-single viewpoint catadioptric cameras and used a spherical mirror array for wide field-of-view light field acquisition. Our work combines catadioptric system with hyperspectral sensing for the first time to our knowledge. Our SPI system can provide high spatial-spectral resolution hyperspectral images.

3. Spectral Sparsity Prior

In this paper, we describe our exploitation of the spectral sparsity of natural hyperspectral images. We first experimentally prove that hyperspectral signals can be sparsely represented in an over-complete dictionary by exploiting a large number of publicly available spectral image datasets. Then we designed a spectral sampling strategy to compressively measure scene spectra. Finally, the hyperspectral datacube was recovered via sparse decomposition.

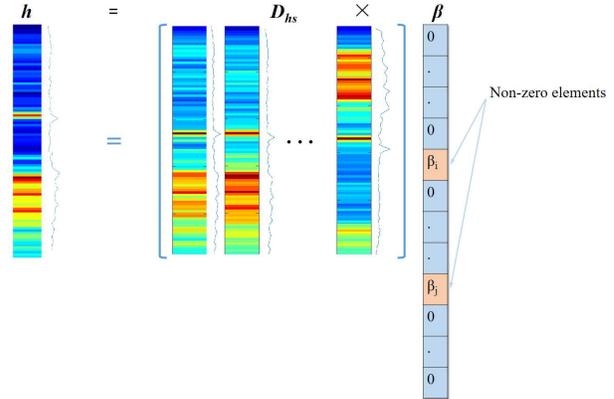


Figure 3. We sparsely represent hyperspectral signals using an over-complete dictionary.

Let h be an M -dimensional hyperspectral signal (spectral power distribution). The spectral sparsity prior says: given a pre-trained over-complete dictionary D_{hs} , where $D_{hs} \in \mathbb{R}^{M \times N}$, arbitrary hyperspectral signal h can be a product of the dictionary D_{hs} and a coefficient vector β , where $\beta \in \mathbb{R}^N$, has very few non-zero elements as shown in Figure 3.

$$h = D_{hs} \cdot \beta, \|\beta\|_0 \ll N \quad (1)$$

Given a training set: h_1, h_2, \dots , an over-complete hyperspectral dictionary D_{hs} is learned by minimizing l_0 of β and residual error. Since l_0 norm minimization is well-known to be NP-hard, Donoho [10] mathematically proved that the ill-posed problem can be efficiently solved by l_1 norm minimization if the coefficient vector β is sufficiently sparse. That gives:

$$D_{hs} = \arg \min_D \frac{1}{n} \sum_{i=1}^n \frac{1}{2} \|h_i - D\beta_i\|_2^2 + \lambda \|\beta_i\|_1 \quad (2)$$

where λ is a balance term to adjust sparsity and residual error. The optimization in 2 is solved by applying the online dictionary learning (ODL) algorithm in [20].

To verify the spectral sparsity prior, we tested on two publicly available hyperspectral datasets with different scales. First on a multispectral image dataset from Columbia University [28] (31 bands), and second on a

hyperspectral dataset from ImageVal (396 bands). These databases contain a large number of indoor scenes and objects of various material under CIE standard illumination. In both datasets, we learn a compact over-complete hyperspectral dictionary D_{hs} containing 150 and 1000 spectral atoms respectively. To verify the validity of the dictionary and the sparse prior, we found the sparsest solution $\hat{\beta}$ using the algorithm in [11].

$$\hat{\beta} = \arg \min_{\beta} \frac{1}{2} \|h - D_{hs} \cdot \beta\|_2^2 \quad s.t. \quad \|\beta\|_1 < k \quad (3)$$

Figure 4 illustrates Peak Signal-to-Noise Ratio (PSNR) of recovered hyperspectral images under different parameters k and at different wavelengths. Since hyperspectral signals

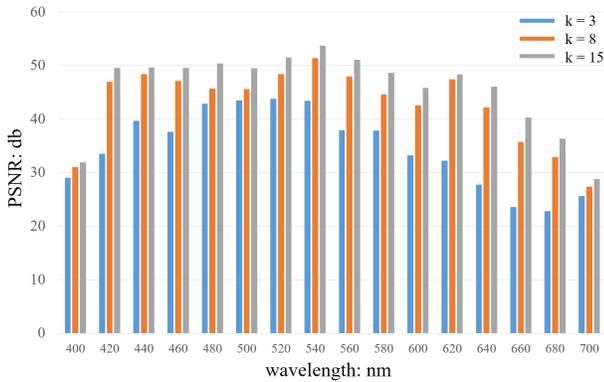


Figure 4. Band-wise Peak Signal-to-Noise Ratio (PSNR) under different parameters k

are sparse under a well-trained dictionary, it is compressible and can be compressively measured, then computationally recovered. The theory of compressive sensing indicates that sparse signals can be faithfully recovered with much less measurements required by the Shannon-Nyquist sampling theorem. Let S denote the sampling matrix, $S \in \mathbb{R}^{K \times M}$, the measurement vector m can be formulated as:

$$m = S \cdot h = S \cdot D_{hs} \cdot \beta = M_s \cdot \beta \quad (4)$$

M_s is the measuring matrix $M_s \in \mathbb{R}^{K \times N}$, $K \ll N$ of sparse signal β . In order to faithfully retrieve the sparse signal, measuring matrix M_s should satisfy the Restricted Isometry Property: for k -sparse signals, arbitrary $2k$ column vectors in M_s should be linearly independent. We solved the underdetermined linear equation 4 using the algorithm in [11]:

$$\hat{\beta} = \arg \min_{\beta} \frac{1}{2} \|m - M_s \cdot \beta\|_2^2 + \lambda \|\beta\|_1 \quad (5)$$

In our work, we adopted a strategy of co-training two dictionaries: hyperspectral dictionary D_{hs} and undersampled dictionary D_{us} . To efficiently train two dictionaries,

we coupled D_{hs} and D_{us} together to form a compound dictionary D^* :

$$D^* = \begin{pmatrix} D_{hs} \\ D_{us} \end{pmatrix} \quad h'_i = \begin{pmatrix} h \\ \delta \cdot S \cdot h \end{pmatrix} \quad (6)$$

where δ is a balance term and simply set to be 1 in our experiments. D^* is learned from a training set of n samples by solving a similar optimization task as 2:

$$D^* = \arg \min_{D'} \frac{1}{n} \sum_{i=1}^n \frac{1}{2} \|h'_i - D' \beta_i\|_2^2 + \lambda \|\beta_i\|_1 \quad (7)$$

We combined two dictionaries together in the training stage which forced the hyperspectral signal and its corresponding undersampled vector to have the same sparse coefficients with respect to the dictionaries. The undersampled dictionary D_{us} was used to find sparse coefficients $\hat{\beta}$. Then the hyperspectral signal h was recovered by $h = D_{hs} \cdot \hat{\beta}$.

4. Snapshot Plenoptic Imager

We constructed a SPI system for 5D light field acquisition. The SPI consists of a catadioptric mirror array coated with different spectral filters. The spectral coating implements the idea of compressively spectral sensing discussed in Section 3.

4.1. Spectral Coded Catadioptric Mirror Array

Catadioptric cameras are widely used for wide field-of-view imaging. We couple spectral and angular dimension together to simultaneously capture 5D plenoptic functions in a single snapshot. The proposed configuration consists of L^2 curved mirrors in a $L \times L$ array and a single DSLR camera. We physically implemented the sampling matrix S by combining the Bayer filter array and spectral coating. Bayer filter arrays are broad bandpass filters that provide the simplest spectral samples. To faithfully measure hyperspectral signals, the spectral coded catadioptric mirrors first encode scene spectra, giving L^2 spectral measurements of the scene. Then Bayer filters further modulate spectral measurements into $3L^2$ values. Denote spectral reflectance function of catadioptric mirrors as column vectors $R_1(\lambda), R_2(\lambda), \dots, R_{L^2}(\lambda)$, and Bayer filters' quantum efficiency as $B_R(\lambda), B_G(\lambda), B_B(\lambda)$, the sampling matrix S in

section 3 is given:

$$S = \begin{bmatrix} B_R(\lambda_{min}) \cdot R_1(\lambda_{min}) \cdots B_R(\lambda_{max}) \cdot R_1(\lambda_{max}) \\ \vdots \\ B_R(\lambda_{min}) \cdot R_{L^2}(\lambda_{min}) \cdots B_R(\lambda_{max}) \cdot R_{L^2}(\lambda_{max}) \\ B_G(\lambda_{min}) \cdot R_1(\lambda_{min}) \cdots B_G(\lambda_{max}) \cdot R_1(\lambda_{max}) \\ \vdots \\ B_G(\lambda_{min}) \cdot R_{L^2}(\lambda_{min}) \cdots B_G(\lambda_{max}) \cdot R_{L^2}(\lambda_{max}) \\ B_B(\lambda_{min}) \cdot R_1(\lambda_{min}) \cdots B_B(\lambda_{max}) \cdot R_1(\lambda_{max}) \\ \vdots \\ B_B(\lambda_{min}) \cdot R_{L^2}(\lambda_{min}) \cdots B_B(\lambda_{max}) \cdot R_{L^2}(\lambda_{max}) \end{bmatrix} \quad (8)$$

We extracted raw data from the DSLR to keep spectral fidelity from on-board post-capture processing, like white balance and non-linear normalization. Figure 5 plots spectral reflectance functions of catadioptric mirrors and Bayer filter array's quantum efficiency of the sensor used in our experiment (Canon 5D MarkIII).

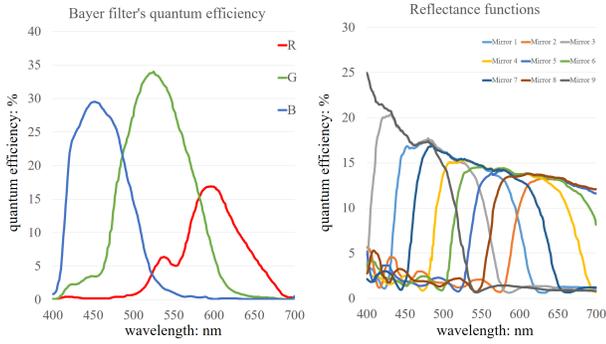


Figure 5. Quantum efficiency of Bayer filters (left) and spectral reflectance functions of catadioptric mirrors (right).

Captured subviews present non-singleview projections of the scene from equal-baselined viewpoints. We re-projected distorted views using geometric modeling in [24]. Consider the setup in Figure 6: a spherical mirror of radius r is located at the origin and a perspective camera aims at the sphere at a distance l . A ray of incident angle θ_i is reflected by the spherical surface with the angle of reflectance θ_r . In 3D free space, all incident rays with the same θ_i can be reprojected to a bundle of rays with θ'_i by a virtual camera placed at l' on the axis. Due to the fact that the camera is very far away from the mirrors: $l \gg r$, so the virtual camera position is approximately constant within a desired field-of-view.

$$\tan \theta_i (l - r \sin \gamma) = r \cdot \cos \gamma \quad (9)$$

where γ is the direction of surface normal. For any given

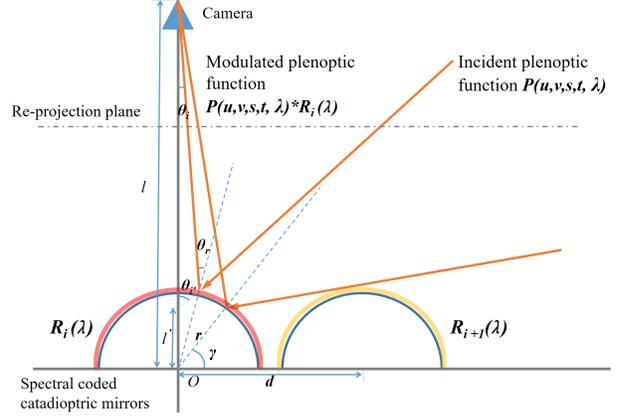


Figure 6. Re-projection geometry of catadioptric mirror array.

θ_i , we have:

$$\gamma = \arcsin\left(\frac{l \tan^2 \theta_i + \sqrt{r^2 \tan^4 \theta_i - l^2 \tan^2 \theta_i + r^2}}{r(1 + \tan^2 \theta_i)}\right) \quad (10)$$

Further using the law of reflection, the equivalent incident angle θ'_i and l' can be obtained as follows:

$$\theta'_i = \left(\frac{\pi}{2} - \gamma\right) + \theta_r = \left(\frac{\pi}{2} - \gamma\right) + \left(\frac{\pi}{2} - \gamma + \theta_i\right) = \theta_i - 2\gamma + \pi \quad (11)$$

$$l' = r \sin \gamma - \frac{r \cos \gamma}{\tan \theta'_i} \quad (12)$$

To correct distortion in catadioptric imaging, we performed a ray-wise re-projection to a virtual focal plane using Eqn. 11 and 12. In Figure 7, we compare our re-projected image with ground truth. To keep the FoV constant, both images were rendered in POV-RAY. There is noticeable resolution loss in the corrected image due to decreasing angular sampling in the catadioptric imaging system.

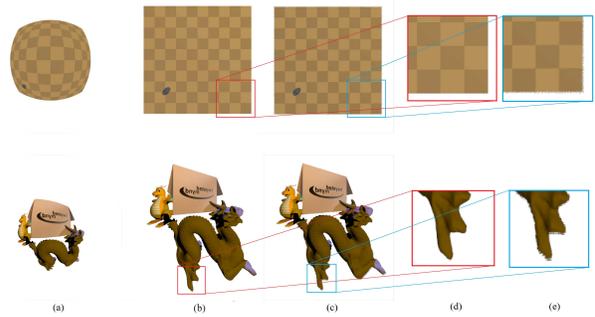


Figure 7. Perspectively corrected SPI images: a simulated SPI image (a), the groundtruth perspective views (b) and (d), and our perspectively corrected results (c) and (e).

The model of a single catadioptric mirror can be extended to a mirror array due to its rotational symmetry. Note that l is not constant for all mirrors in a planar array configuration.

4.2. Reconstruction

To reconstruct 5D data from corrected sub-views, we first conduct an image registration process. Because each catadioptric mirror is coated with different spectral filters, the sub-views share no intensity consistency. Figure 8 shows epipolar images extracted from regular light fields and SPI. It is apparent that a single scene point shows different spectral responses in different viewpoints.

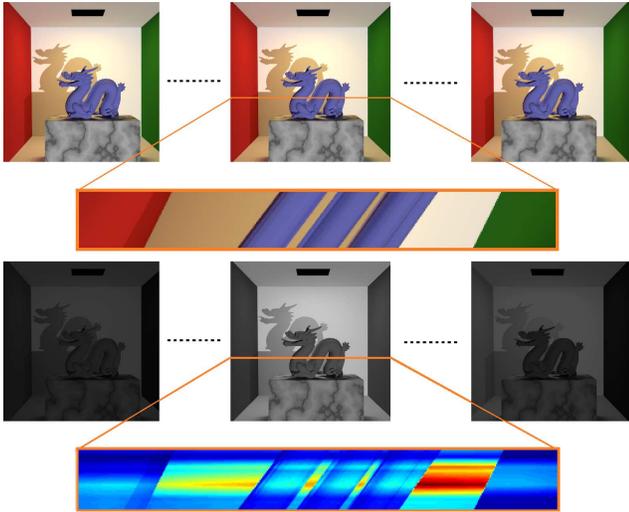


Figure 8. Epipolar images (EPI) from regular light fields (top) and SPI (bottom). Intensity inconsistencies can lead to unreliable tensor (depth) estimation.

Traditional stereo matching methods cannot handle cross-spectrum image registration due to inconsistent intensity. To register cross-spectrum images, We adopt the Robust Selective Normalized Cross Correlation (RSNCC) for reliable feature matching [23]. We then recovered a depth map by minimizing a modified energy function via α -expansion graph cuts [5, 15, 4, 14]. Finally, we restored the hyperspectral dimensions via sparse reconstruction as discussed in section 3

Apart from using the cross-spectrum correspondence, we further employ depth-from-defocus methods by fusing a hyperspectral focal stack. We generate a focal stack $I_1, I_2 \dots$ by refocusing the hyperspectral light field at a serial of depths. We then compute a patch-wise refocusing metric ξ_{rf} as:

$$\xi_{rf}(d, p) = -\frac{M[I_d(d, p)]}{\sum_{i \in \mathbf{U}} M[I_d(d, i)]} \quad (13)$$

where d is target depth, \mathbf{U} is a patch centered at p , and $M(I)$ represents gradient magnitude.

We modify traditional energy function $\xi(d, p)$ in stereo

matching by :

$$\xi(d, p) = \underbrace{\alpha \cdot \xi_{RSNCC}(d, p)}_{\text{dataterm}} + \underbrace{\beta \cdot \xi_{rf}(d, p)}_{\text{smooth penalty}} + \underbrace{P(d, p)}_{\text{smooth penalty}} \quad (14)$$

where ξ_{RSNCC} indicates the RSNCC correspondence metric, and α, β are weighting terms. The depth map can be solved by minimizing the given energy function:

$$D_{map}(p) = \arg \min_d \sum \xi(d, p) \quad (15)$$

We solved the optimization with α -expansion graph cuts. For more information of the RSNCC correspondence and graph cuts, please refer [23, 5, 15, 4, 14]

5. Experimental Results

Hardware Implementation. The SPI prototype consists of 9 spectral coded catadioptric mirrors in a 3×3 array. The reflectance functions of individual mirrors are provided in Figure 5. We used off-the-shelf Canon DSLR (5D MarkIII with EF 75-300mm lens) to capture raw SPI images 2 meters away from the mirror array. Raw SPI images have a spatial resolution of 4000×6000 . All catadioptric mirrors have the identical radius of 3.25cm and the baseline between two neighbouring mirrors is 7cm . Perspective corrected sub-views have a lower spatial resolution of 1000×1000 and a wide FoV of 90° . The learning-based algorithm further recovers 31 bands, giving the final reconstructed 5D plenoptic function a resolution of $1000 \times 1000 \times 3 \times 3 \times 31$.

Spectra Recovery. To quantitatively evaluate spectral efficiency and accuracy of our SPI system, we first simulated SPI images from a benchmark hyperspectral image database[28]. Then we reconstructed hyperspectral datacubes (31 bands) from SPI images. We compared our reconstruction results with ground truth by calculating band-wise Root Mean Square Error (RMSE) over the visible spectra (from 400nm to 700nm). Table 1 shows the performance of the SPI system in different scenes. Figure 9 illustrates a recovered datacube (view in RGB) from simulated SPI image and spectral profiles of randomly extracted scene points. The recovered hyperspectral signals provide high spectral fidelity.

As shown in table 1 and 9, the recovered hyperspectral datacube has less confidence in near infra-red and ultra-violet spectra. The bottom right spectral profile in Figure 9 deviates from ground truth when it reaches near infra-red spectrum. These failure cases tend to happen in insensitive spectra of the detector. Commercial CCD and CMOS sensors for color imaging have low quantum efficiency when wavelength comes to $> 680\text{nm}$ or $< 420\text{nm}$. To prevent these failures from happening, it is recommended to use sensors with a wide and plane spectral response.

Table 1. Band-wise RMSE (8 bit spectral images)

	400nm	450nm	500nm	550nm	600nm	650nm	700nm
Toys	28.1	2.2	2.0	1.4	3.1	5.1	25.4
Food	10.5	3.0	2.9	1.7	2.6	7.2	20.0
Sliced lemon	8.4	1.1	1.2	0.8	2.4	2.8	9.6
Feather	16.0	2.5	4.0	1.7	4.6	7.3	18.2
Flower	6.4	0.7	1.2	0.6	1.3	2.9	20.1
Painting	9.8	2.0	2.7	1.1	3.0	3.5	14.1

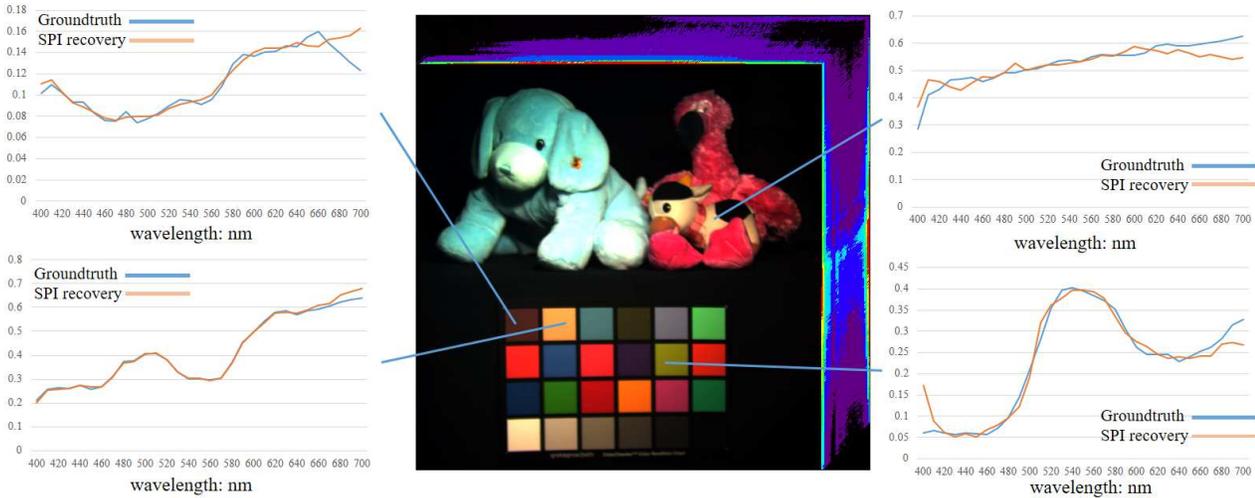


Figure 9. Recovered datacube from simulated SPI image and spectral profiles of randomly chosen scene points.

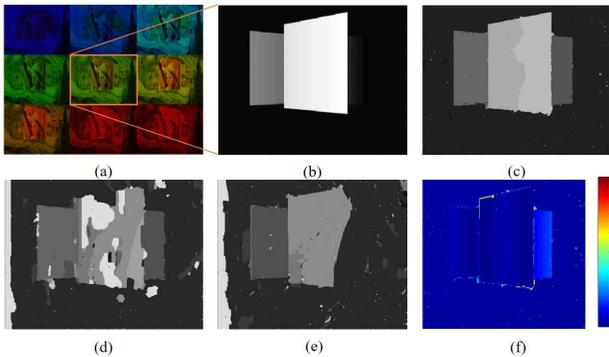


Figure 10. (a) simulated SPI image (viewed in RGB), (b) the ground truth depth map, (c) our recovered depth map, (d) SSD result, (e) NCC result, (f) the error map.

Hyperspectral Light Fields. Typical applications of light field techniques are depth estimation and dynamic refocusing. Simulations and experiments on real scenes have demonstrated that our SPI camera is capable of providing high quality depth maps and unique spectral refocusing effects.

Figure 10 shows our recovered depth map of rendered hyperspectral light fields. Simulated hyperspectral light fields and ground truth depth map are rendered in POV-RAY

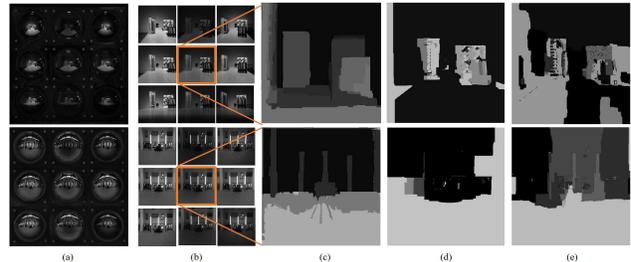


Figure 11. Real scenes captured by our SPI system of indoor (top row) and outdoor (bottom row). (a) raw SPI images, (b) perspective-corrected views, (c) our depth estimation results, (d) SSD stereo matching results, (e) NCC stereo matching results.

with spatial resolution 600×800 and 11 depth layers. We compared our depth estimation with depth maps generated from popular stereo matching methods (SSD, NCC). Traditional stereo algorithms failed to recover depth information from cross-spectrum images due to intensity inconsistency while our reconstruction method handles hyperspectral light fields well and provides high quality depth estimation with little artifacts. Figure 11 shows two real scenes captured with our SPI system. Comparison with stereo matching results demonstrate that the proposed algorithm faithfully retrieves depth information which benefits geometry recon-

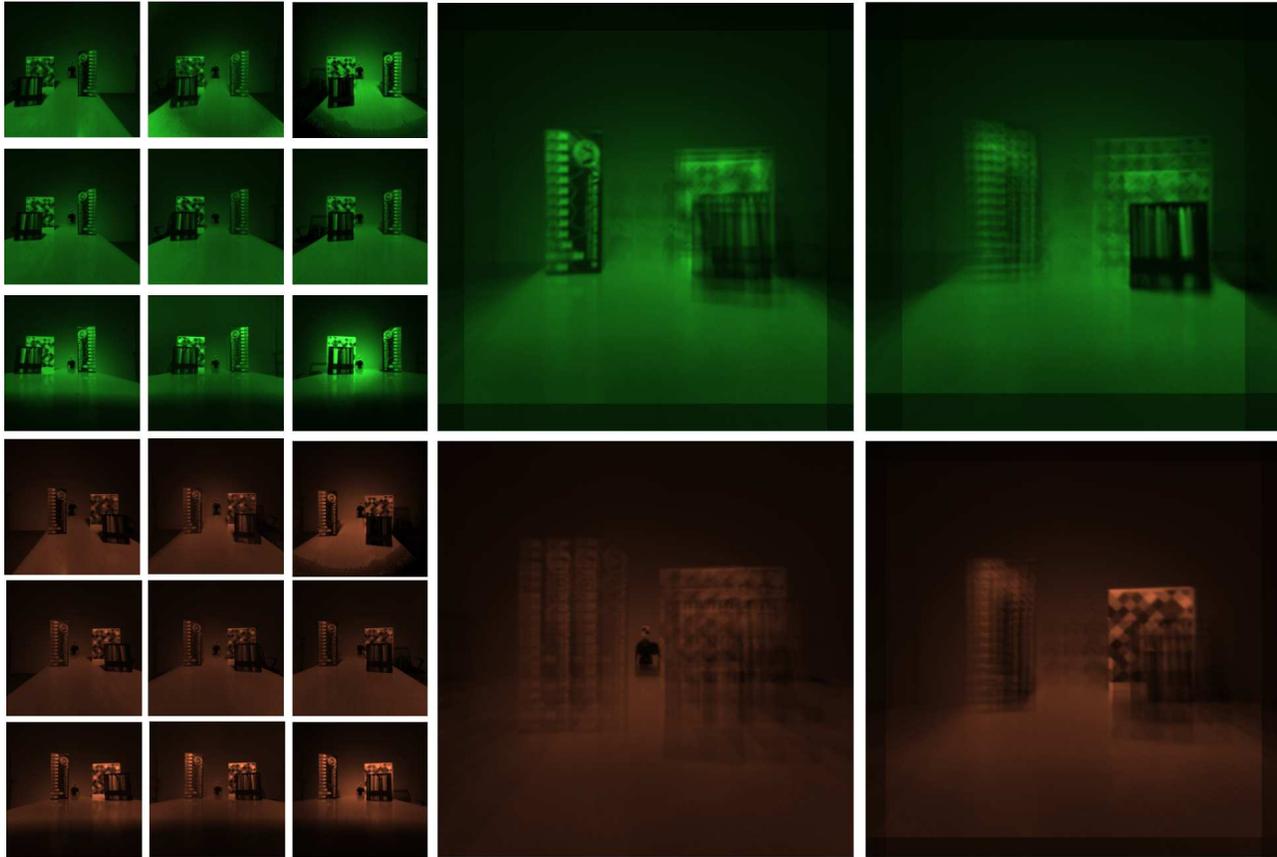


Figure 12. Spectral refocusing effects: Top row: refocusing at 550 nm. Bottom row: refocusing at 650 nm.

struction, semantic segmentation and *etc.*

The dynamic depth-of-field effect demonstrates the validity of our SPI technique. In Figure 12, we show spectral refocusing effects in a real indoor scene. We refocused the reconstructed 5D plenoptic function at 4 depth layers and at 550nm and 660nm respectively. The unique spectral refocusing effects have potential in computer vision tasks *e.g.* object detection and tracking.

6. Conclusion and future work

We have presented a Snapshot Plenoptic Imager (SPI) capable of acquiring 5D light fields in a single shot. Our system only uses a single DSLR camera and combines it with an array of spectrally coded catadioptric mirrors. In a snapshot, we are able to acquire plenoptic samples in spatial, angular and spectral dimensions. We have devised geometric correction technique to address image distortions and presented a learning based method to reconstruct the complete plenoptic function.

Our approach is the first single-shot plenoptic function imaging solution. In theory, it can be directly applied to also acquire the temporal dimension, *e.g.* by using a high

resolution video camera. In reality, the effective resolution and image quality of commodity video cameras will lead to low image quality (resolution, signal-to-noise ratio, *etc.*) and hence poor image registration. One possible solution is to interleave the capture of video and static images. A high resolution static image can compensate for low quality video frames by using learning-based image denoising and super-resolution. Our solution can also be potentially combined with additional cameras to form a hybrid sensing solution.

Our current system can only handle visible spectral range. In the future, we plan to extend the range to ultra-violet and near infra-red spectra. Krishnan *et al.* [16] has previously shown that a commodity sensor can capture under the two spectra, possibly with the help of auxiliary flash. Finally, we expect broad applications of our system in tasks such as object detection, material classification, and spectral-aware relighting.

References

- [1] B. Arad and O. Ben-Shahar. Sparse recovery of hyperspectral signal from natural rgb images. In *European Conference*

- on *Computer Vision*, pages 19–34. Springer, 2016.
- [2] S. Baker and S. K. Nayar. A theory of single-viewpoint catadioptric image formation. *International Journal of Computer Vision*, 35(2):175–196, 1999.
 - [3] E. BenDor, R. Taylor, J. Hill, J. Demattè, M. Whiting, S. Chabrilat, and S. Sommer. Imaging spectrometry for soil applications. *Advances in Agronomy*, 97(07):321–392, 2008.
 - [4] Y. Boykov and V. Kolmogorov. An experimental comparison of min-cut/max-flow algorithms for energy minimization in vision. *IEEE Transactions on Pattern Analysis & Machine Intelligence*, 11(26):1124–37, 2004.
 - [5] Y. Boykov, O. Veksler, and R. Zabih. Fast approximate energy minimization via graph cuts. *IEEE Transactions on Pattern Analysis & Machine Intelligence*, 23(11):1222–1239, 2001.
 - [6] J. G. P. W. Clevers. The use of imaging spectrometry for agricultural applications. *Isprs Journal of Photogrammetry & Remote Sensing*, 54(5-6):299–304, 1999.
 - [7] T. W. Cooley, J. A. Gardner, T. M. Davis, and S. D. Straight. Advanced responsive tactically effective military imaging spectrometer (artemis): system overview and objectives. In *Photonic Devices + Applications*, pages 666102–666102–6, 2007.
 - [8] P. J. Curran. Imaging spectrometry for ecological applications. *International Journal of Applied Earth Observation & Geoinformation*, 3(4):305–312, 2001.
 - [9] Y. Ding, J. Yu, and P. Sturm. Multiperspective stereo matching and volumetric reconstruction. In *IEEE International Conference on Computer Vision*, pages 1827–1834, 2009.
 - [10] D. L. Donoho. For most large underdetermined systems of equations, the minimal ℓ_1 norm nearsolution approximates the sparsest nearsolution. *Communications on Pure & Applied Mathematics*, 59(6):797–829, 2006.
 - [11] B. Efron, T. Hastie, I. Johnstone, and R. Tibshirani. Least angle regression. *Mathematics*, 32(2):407–451, 2004.
 - [12] A. F. Goetz, G. Vane, J. E. Solomon, and B. N. Rock. Imaging spectrometry for earth remote sensing. *Science*, 228(4704):1147–53, 1985.
 - [13] J. Holloway, K. Mitra, S. Koppal, and A. N. Veeraraghavan. Generalized assorted camera arrays: Robust cross-channel registration and applications. *IEEE Transactions on Image Processing A Publication of the IEEE Signal Processing Society*, 24(3):823–835, 2014.
 - [14] V. Kolmogorov and R. Zabih. Computing visual correspondence with occlusions using graph cuts. In *Computer Vision, 2001. ICCV 2001. Proceedings. Eighth IEEE International Conference on*, pages 508–515, 2001.
 - [15] V. Kolmogorov and R. Zabih. What energy functions can be minimized via graph cuts? *IEEE Transactions on Pattern Analysis & Machine Intelligence*, 26(2):147–59, 2004.
 - [16] D. Krishnan and R. Fergus. Dark flash photography. *Acm Transactions on Graphics*, 28(28):96, 2009.
 - [17] G. Krishnan and S. K. Nayar. Cata-fisheye camera for panoramic imaging. pages 1–8, 2008.
 - [18] M. Levoy, B. Chen, V. Vaish, M. Horowitz, I. McDowall, and M. Bolas. Synthetic aperture confocal imaging. *Acm Transactions on Graphics*, 23(3):825–834, 2004.
 - [19] N. Li, H. Lin, B. Sun, M. Zhou, and J. Yu. Rotational crossed-slit light field. In *CVPR*, 2016.
 - [20] J. Mairal, F. Bach, J. Ponce, and G. Sapiro. Online dictionary learning for sparse coding. In *International Conference on Machine Learning, ICML 2009, Montreal, Quebec, Canada, June*, pages 689–696, 2009.
 - [21] R. M. H. Nguyen, D. K. Prasad, and M. S. Brown. *Training-Based Spectral Reconstruction from a Single RGB Image*. 2014.
 - [22] N. Ren, M. Levoy, M. Bredif, G. Duval, M. Horowitz, and P. Hanrahan. Light field photography with a hand-held plenoptic camera. 2005.
 - [23] X. Shen, L. Xu, Q. Zhang, and J. Jia. *Multi-modal and Multi-spectral Registration for Natural Images*. Springer International Publishing, 2014.
 - [24] Y. Taguchi, A. Agrawal, A. Veeraraghavan, S. Ramalingam, and R. Raskar. Axial-cones: modeling spherical catadioptric cameras for wide-angle light field rendering. *Acm Transactions on Graphics*, 29(6):7231–7238, 2010.
 - [25] 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 Transactions on Graphics*, 26(3), 2007.
 - [26] Z. Xiong, L. Wang, H. Li, D. Liu, and F. Wu. Snapshot hyperspectral light field imaging. In *The IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, July 2017.
 - [27] J. Yang, J. Wright, T. S. Huang, and Y. Ma. Image super-resolution via sparse representation. *IEEE Transactions on Image Processing A Publication of the IEEE Signal Processing Society*, 19(11):2861, 2010.
 - [28] F. Yasuma, T. Mitsunaga, D. Iso, and S. K. Nayar. Generalized assorted pixel camera: Postcapture control of resolution, dynamic range, and spectrum. *IEEE Transactions on Image Processing A Publication of the IEEE Signal Processing Society*, 19(9):2241–53, 2010.
 - [29] J. Ye, Y. Ji, and J. Yu. A rotational stereo model based on xslit imaging. In *International Conference on Computer Vision, ICCV*, pages 489–496, 2013.
 - [30] J. Yu and L. Mcmillan. General linear cameras. In *Computer Vision - ECCV 2004, European Conference on Computer Vision, Prague, Czech Republic, May 11-14, 2004. Proceedings*, pages 14–27, 2004.