Snapshot Hyperspectral Light Field Imaging

Zhiwei Xiong[1] Lizhi Wang[2] Huiqun Li[1] Dong Liu[1] Feng Wu[1]
[1]University of Science and Technology of China
[2]Beijing Institute of Technology

Abstract

This paper presents the first snapshot hyperspectral light field imager in practice. Specifically, we design a novel hybrid camera system to obtain two complementary measurements that sample the angular and spectral dimensions respectively. To recover the full 5D hyperspectral light field from severely undersampled measurements, we then propose an efficient computational reconstruction algorithm by exploiting the large correlations across the angular and spectral dimensions through self-learned dictionaries. Simulation on an elaborate hyperspectral light field dataset validates the effectiveness of the proposed approach. Hardware experimental results demonstrate that, for the first time to our knowledge, a 5D hyperspectral light field containing 9×9 angular views and 27 spectral bands can be acquired in a single shot.

1. Introduction

Computational imaging has seen a tremendous progress in the past decades, owing to the rapid advancement of optical instruments and the explosive growth of computing power. The ultimate goal of computational imaging is to simultaneously resolve the 7 dimensions of the plenoptic function, *i.e.*, 3D in space (2D planar + 1D depth), 1D in time, 1D in spectrum, and 2D in angle [33]. The dimensions that are beyond traditional digital imaging, *i.e.*, depth, spectrum, and angle, have been extensively explored in literature under the scope of depth/3D imaging, multi/hyper-spectral imaging, and light field imaging, respectively [4, 34, 8, 12]. Moreover, commercial depth cameras (*e.g.*, Kinect) and light field cameras (*e.g.*, Lytro) are readily available for daily use, which provides new opportunities for solving difficult computer vision tasks and also enables new applications [11, 22, 17].

The trend of computational imaging is to integrate higher plenoptic dimensions together, while maintaining the respective resolution as much as possible. Following this trend, current research frontiers extend to depth from light field [13, 30], light field super-resolution [31, 35], hyperspectral video acquisition [20, 28], hyperspectral 3D imaging [14, 29], and so on. Thanks to the large correlations across different plenoptic dimensions, it is possible to recover high-dimensional light information from severely undersampled measurements. Still, elaborate hardware systems and computational reconstruction algorithms are indispensable to guarantee a decent performance.

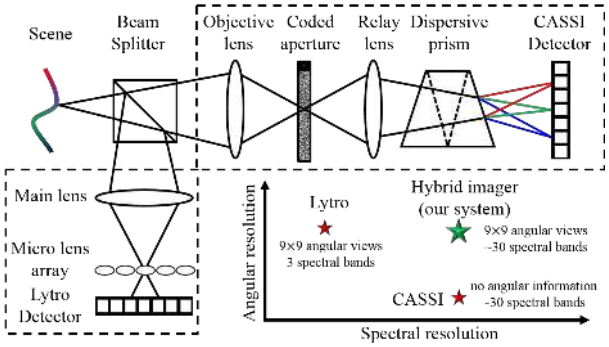


Figure 1. Schematic of snapshot hyperspectral light field imager.

In this paper, we explore a new direction in computational imaging, *i.e.*, snapshot hyperspectral light field imaging, which is an essential step towards encompassing all dimensions of the plenoptic function. Previously, the acquisition of a hyperspectral light field needs to be conducted in a scanning manner, either using a spectrometer mounted on a gantry to scan in the angular and spectral dimensions sequentially [32], or using a micro-lens-array together with tunable filters to scan in the spectral dimension alone [16]. However, the above scanning approaches are not applicable in time-critical scenarios, such as scenes with dynamic objects or varying illumination. How to acquire a hyperspectral light field without sacrificing the temporal resolution remains a key challenge.

To this end, we design a novel hybrid camera system as shown in Figure 1, which consists of an off-the-shelf light field camera (Lytro) and a coded aperture snapshot spectral imager (CASSI). The two branches are co-located via a beam splitter and calibrated in spatial, angular, and spectral dimensions. The incident light from the scene is equally divided by the beam splitter and then captured by Lytro and

CASSI, respectively. The RGB light field obtained by Lytro contains angular information of the scene but lacks spectral resolution, while the compressive measurement obtained by CASSI encodes hyperspectral information of the scene but lacks angular resolution. The hybrid imager thus provides complementary measurements for recovering the 5D hyperspectral light field with high angular and spectral resolutions, as illustrated in Figure 1.

Still, recovering the full 5D hyperspectral light field from such undersampled measurements is a severely underdetermined problem, due to the large dimensionality gap. The correlations across the angular and spectral dimensions should be exploited to assist the reconstruction. A key observation here is that, the 5D hyperspectral light field can be treated as a concatenation of 4D band-wise light fields, each of which shares similar structures to one of the blue, green, and red light fields obtained from Lytro according to the spectral proximity. Therefore, the RGB light field provides a strong prior for the band-wise light fields to be recovered. Specifically, three over-complete 4D dictionaries can be learned from the RGB light field to sparsely represent each band-wise light field. We then propose to formulate the hyperspectral light field reconstruction as a sparse-constraint optimization problem, given the undersampled measurements and the self-learned dictionaries.

For the evaluation purpose, we prepare a hyperspectral light field dataset by scanning a set of static scenes using a spectrometer mounted on a gantry. This dataset is used to validate the effectiveness of the computational reconstruction algorithm by simulation, as well as to optimize the parameters during the reconstruction. We then conduct hardware experiments with the developed hybrid camera system and the proposed reconstruction algorithm. Both simulation and hardware experimental results demonstrate that, for the first time to our knowledge, a 5D hyperspectral light field with high angular and spectral resolutions can be acquired in a single shot.

The main contributions of this work can be summarized into three aspects:

(1) The first hardware system for snapshot hyperspectral light field acquisition.

(2) An effective computational reconstruction algorithm for recovering the full 5D hyperspectral light field from severely undersampled measurements.

(3) An elaborate hyperspectral light field dataset that will be made publicly available for developing new computational imaging systems and algorithms.

In addition, it is worth mentioning that the proposed approach can be naturally extended for 6D hyperspectral light field video acquisition due to its snapshot property. In this sense, the proposed approach proceeds a large step towards the ultimate goal of resolving the 7D plenoptic function simultaneously.

2. Related work

Light field imaging. Light field imaging technology is getting mature in the sense that commercial cameras using a micro-lens-array (*e.g.*, Lytro and Raytrix) are readily available for consumer and laboratory use. Generally, these cameras output RGB light fields of the scene at a reduced spatial resolution of the detector in exchange of the angular resolution. Alternatively, the compressive light field photography [21] does not sacrifice the spatial resolution, but requires an over-complete 4D dictionary learned from an external light field database to assist the computational reconstruction. Our proposed system uses Lytro to directly capture an RGB light field, from which three over-complete 4D dictionaries are learned to exploit the correlations across the angular and spectral dimensions. These self-learned dictionaries ensure high sparsity when they are used to represent the band-wise light fields to be recovered.

Snapshot hyperspectral imaging. Various snapshot hyperspectral imaging prototypes have been developed recently, among which CASSI [2, 25] and its variants [18, 19] demonstrate impressive performance and attract increasing attention. By employing a coded aperture and a disperser to optically encode the 3D spectral information onto a 2D detector, CASSI recovers a full hyperspectral image relying on the compressive sensing theory and computational reconstruction, which supports video recording of dynamic scenes [26]. To boost the reconstruction fidelity of CASSI while maintaining its snapshot advantage, the dual-camera design of CASSI has been proposed [27, 28], where measurements from CASSI and a co-located grayscale camera are jointly used for the hyperspectral reconstruction. Inspired by that, our proposed system integrates Lytro with CASSI to acquire a hyperspectral light field without sacrificing the temporal resolution.

Hybrid imaging. Hybrid camera systems have been employed to break the capability limitation of individual cameras in many occasions. For example, a high-speed, low-resolution camera and a low-speed, high-resolution camera can be combined for high-speed, high-resolution imaging [3]; a low-resolution hyperspectral imager and a high-resolution RGB camera can be used together for high-resolution hyperspectral imaging [20]; and a low-resolution light field camera can be combined with a high-resolution RGB camera for high-resolution light field imaging [6]. Our proposed system inherits the idea of hybrid imaging for hyperspectral light field acquisition, which combines the angular and spectral dimensions for the first time.

3. System principles

Figure 2 illustrates the data flow in the proposed system. After the beam splitter, the Lytro branch captures an RGB light field that contains angular information of the scene

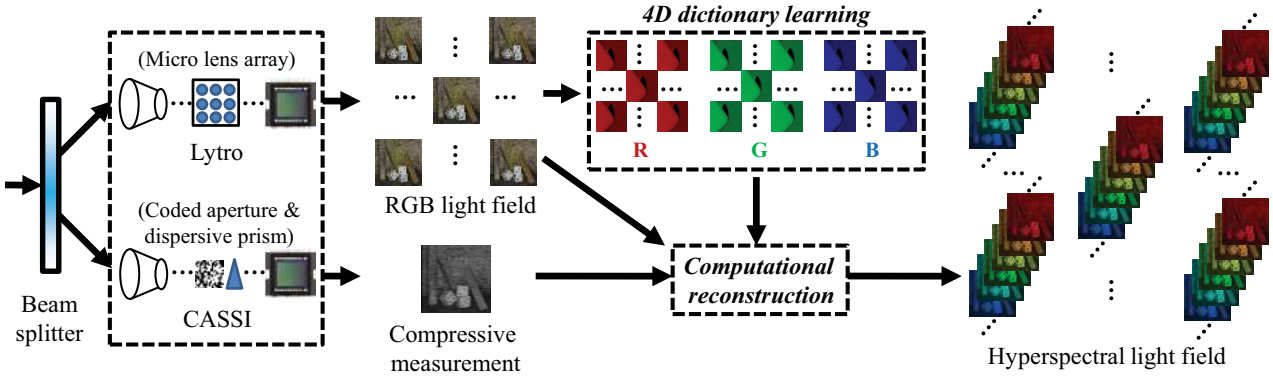


Figure 2. Data flow in the proposed system. The Lytro branch captures an RGB light field containing angular information while the CASSI branch captures a compressive measurement encoding hyperspectral information. The RGB light field is used to learn three over-complete 4D dictionaries, on which the band-wise light fields to be recovered can be sparsely represented. The underlying 5D hyperspectral light field is reconstructed through computational reconstruction relying on the undersampled measurements and the self-learned dictionaries.

relying on a built-in micro-lens-array, while the CASSI branch captures a compressive measurement that encodes hyperspectral information of the scene using a coded aperture and a dispersive prism. The Lytro branch lacks spectral resolution while the CASSI branch lacks angular resolution. Considering the large correlations across the angular and spectral dimensions, it is possible to recover the full 5D hyperspectral light field by jointly using measurements from these two branches. Note that the two branches should be calibrated in spatial, angular, and spectral dimensions. Specifically, the f-numbers of the main objective lenses in Lytro and CASSI should be matched so they collect light within the same angular range. Also, optical filters with the same passband should be used in the two branches so they collect light within the same spectral range.

Denote $f(x, y, u, v, \lambda)$ the discrete 5D hyperspectral light field to be recovered, where $1 \leq x \leq W$ and $1 \leq y \leq H$ index the spatial coordinates, $1 \leq u, v \leq S$ index the angular coordinates (here we assume a square light field, which is usually the case in practice), and $1 \leq \lambda \leq \Omega$ indexes the spectral coordinate. The output from Lytro, *i.e.*, the RGB light field, can be written as

$$g^l(x, y, u, v, k) = 0.5 \sum_{\lambda=1}^{\Omega} \omega_k(\lambda) f(x, y, u, v, \lambda) \quad (1)$$

where $k = 1, 2, 3$ corresponds to blue, green, and red channels respectively, and $\omega_k(\lambda)$ denotes the spectral response of the Lytro detector corresponding to each channel. This equation can be rewritten in a linear matrix form as

$$G^l_{u,v} = \Phi^l F_{u,v} \quad (2)$$

where $G^l_{u,v}$ and $F_{u,v}$ are the vectorized representations of g^l and f at angular coordinates (u, v), *i.e.*, view-wise RGB and hyperspectral images, and Φ^l is the angular-invariant observation matrix of Lytro which is determined by $\omega_k(\lambda)$.

On the other hand, the compressive measurement from CASSI can be written as

$$g^c(x, y) = 0.5 \sum_{\lambda=1}^{\Omega} \sum_{u,v=1}^{S} \omega(\lambda) T(x, y - \psi(\lambda)) f(x, y - \psi(\lambda), u, v, \lambda) \quad (3)$$

where $\omega(\lambda)$ denotes the spectral response of the CASSI detector, $T(x, y)$ the transmission function of the coded aperture, and $\psi(\lambda)$ the wavelength-dependent dispersion function of the prism. (Please refer to [26] for a detailed formulation of the CASSI measurement.) Similar to the Lytro branch, the output from CASSI can be rewritten as

$$G^c = \Phi^c F \quad (4)$$

where G^c is the vectorized representation of g^c, $F = \{F_{u,v}\}_{u,v=1}^{S}$ can be regarded as the concatenation of the view-wise hyperspectral images, and Φ^c is the observation matrix of CASSI which is jointly determined by $\omega(\lambda)$, $T(x, y)$, and $\psi(\lambda)$.

By combining Eq. 2 and Eq. 4, the hybrid camera system model can then be expressed as

$$\begin{pmatrix} G^c \\ G^l_{1,1} \\ G^l_{1,2} \\ \vdots \\ G^l_{S,S} \end{pmatrix} = \begin{pmatrix} \Phi^c & & & \\ \Phi^l & \mathbf{0} & \dots & \mathbf{0} \\ \mathbf{0} & \Phi^l & \dots & \mathbf{0} \\ \vdots & \vdots & \ddots & \vdots \\ \mathbf{0} & \mathbf{0} & \dots & \Phi^l \end{pmatrix} \begin{pmatrix} F_{1,1} \\ F_{1,2} \\ \vdots \\ F_{S,S} \end{pmatrix} \quad (5)$$

A more simplified expression will be

$$G = \Phi F \quad (6)$$

where G comprises measurements from both two branches and Φ is a sparse matrix representing the overall system forward operation.

4. Computational reconstruction

Due to the large dimensionality gap, recovering the full 5D hyperspectral light field F from its undersampled measurements G is a severely underdetermined problem and has rarely been investigated before. To tackle this problem, we exploit the large correlations across the angular and spectral dimensions of the 5D signal through the sparsity prior. Specifically, the 5D hyperspectral light field can be regarded as a concatenation of 4D band-wise light fields. If we treat the RGB light field from Lytro as three separate light fields, then each 4D band-wise light field shares similar structures to one of them according to the spectral proximity. For example, a band-wise light field falling in the green spectrum is observed to have similar structures to the green light field. Therefore, the RGB light field can be used to learn three over-complete 4D dictionaries to sparsely represent each 4D band-wise light field.

To learn the three over-complete dictionaries, we randomly sample a number of 4D patches sized $m = w \times h \times s \times s$ from the blue, green, and red light fields, separately. The corresponding dictionary $D_k \in \mathbb{R}^{m \times n} (k = 1, 2, 3)$ is then derived by KSVD [1], where $n\,(n > m)$ is the number of atoms (*i.e.*, vectorized 4D patches) remaining in the dictionary. These self-learned dictionaries ensure high sparsity when they are used to represent the band-wise light fields to be recovered. Once the dictionaries are ready, the 5D hyperspectral light field can be sparsely represented as

$$F = [F_1, F_2, \ldots, F_\Omega]^T = [D_1, D_2, D_3] \circ [\alpha_1, \alpha_2, \ldots, \alpha_\Omega]^T \tag{7}$$

where $F_\lambda (1 \leq \lambda \leq \Omega)$ denotes a band-wise light field, $\alpha_\lambda (1 \leq \lambda \leq \Omega)$ denotes the sparse coefficient vector that represents F_λ on D_k, and the operation $\circ$ is defined as

$$F = [D_1(\alpha_1, \ldots, \alpha_i), D_2(\alpha_{i+1}, \ldots, \alpha_j), D_3(\alpha_{j+1}, \ldots, \alpha_\Omega)]^T \tag{8}$$

where $1 \leq i < j \leq \Omega$ specify which dictionary should be used for each band-wise light field and are determined by the spectral response of the Lytro detector. A more simplified expression will be

$$F = D \circ \alpha \tag{9}$$

where D is composed of $\{D_k\}_{k=1}^3$ and α is the concatenation of $\{\alpha_\lambda\}_{\lambda=1}^\Omega$.

According to the compressive sensing theory [7, 9], F can be recovered by solving the following optimization problem instead

$$\hat{\alpha} = \arg\min_\alpha ||G - \Phi D \circ \alpha||_2^2 + \tau ||\alpha||_0 \tag{10}$$

where τ is a regularization parameter. This optimization problem can be efficiently solved by employing the orthogonal matching pursuit algorithm [24].

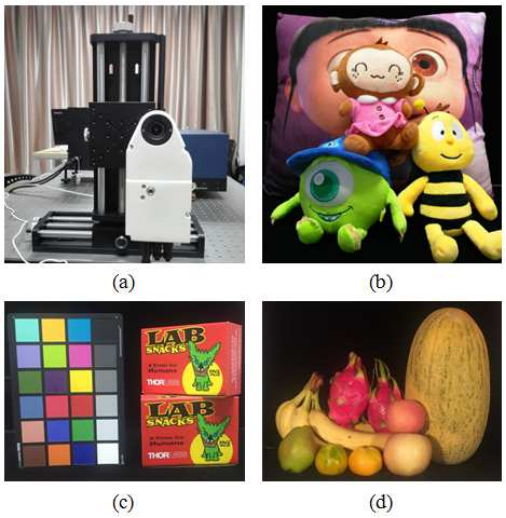


Figure 3. (a) The platform for preparing the hyperspectral light field dataset. (b)-(d) three static scenes for generating the dataset: *Toys*, *Boards*, and *Fruits*.

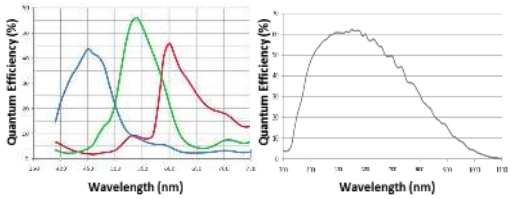


Figure 4. Spectral sensitivity curves of the Lytro (left) and CASSI (right) detectors used in our hardware system.

5. Simulation

5.1. Dataset

Since the hyperspectral light field data is rarely available in public, we prepare a hyperspectral light field dataset ourselves for the evaluation purpose. The dataset is obtained by scanning three static scenes using a spectrometer (Senop Rikola) mounted on a gantry. The platform used to collect the data is shown in Figure 3(a). The spectrometer uses a liquid crystal tunable filter and captures a narrow-band spectral image with up to 1nm bandwidth in a single shot. The gantry supports 2D translation of the spectrometer with 0.01mm precision on each direction. The three scenes, as shown in Figure 3(b)-(d), are placed at a distance of around 1m to the platform and contain a variety of materials with diverse geometry and reflectance characteristics. The raw data for each scene contains 9×9 angular views and 25 spectral bands (ranging from 450nm to 690nm in 10nm increments) at a spatial resolution of 512×512. In total, we capture 2025 images for each scene when the spectrometer is placed at different locations (the distance between two neighboring views is 10mm). Note a one-time calibration is needed in advance to address the extrinsic and intrinsic

camera parameters for rectifying the captured images.

5.2. Algorithm evaluation

The spectral sensitivity curves of the Lytro and CASSI detectors used in our hardware system are shown in Figure 4, which are discretized to generate the required spectral response functions for simulation. In addition, for the CASSI branch, the transmission function of the coded aperture is generated as a random Bernoulli distribution with $p = 0.5$, and the dispersion function of the prism is assumed to be a linear distribution for simplicity. For the Lytro branch, we further divide the spectral coordinate into three intervals as 450-530nm, 540-600nm, and 610-690nm, which correspond to the blue, green, and red channels and specify the dictionary that should be used for recovering a certain band-wise light field (*i.e.*, determining i and j in Eq. 8).

For simulation, we test three different angular resolutions of $S = 5, 7, 9$. The spectral resolution remains as $\Omega = 25$, and the spatial resolution is slightly cropped after rectification. The parameters used in our proposed dictionary-based reconstruction (DBR) algorithm are selected optimally through a cross-validation process. For dictionary learning, the 4D patch size is set as $m = 6 \times 6 \times S \times S$, and 30000 patches are randomly sampled from each of the blue, green, and red light fields, respectively. After KSVD, there are $n = 2m$ atoms remaining in the dictionary. The maximum iteration number of DBR is set to 80 and τ is set to 0.004, 0.002, and 0.0005 when S equals to 5,7, and 9, respectively. For comparison, we also generate the reconstruction results using the two-step iterative shrinkage/thresholding (TwIST) algorithm [5] along with the total variation regularizer.

Quantitative evaluation. Two quantitative image quality metrics, peak signal-to-noise ratio (PSNR) and spectral angle mapping (SAM) [15], are adopted to evaluate the reconstruction fidelity. PSNR measures the spatial fidelity of reconstruction, which is calculated based on each 2D spatial image and then averaged over the spectral and angular dimensions. SAM measures the spectral fidelity of reconstruction, which is calculated based on each 1D spectral vector and then averaged over the spatial and angular dimensions. The PSNR and SAM results of TwIST and DBR are reported in Table 1. It can be seen that both TwIST and DBR decently recover the 5D hyperspectral light field at different angular resolutions, which demonstrates the feasibility of our proposed hybrid imaging model. Moreover, DBR outperforms TwIST with an average of 2.45dB gain in PSNR and 15% decrease in SAM (a smaller SAM indicates a higher fidelity reconstruction), which validates the effectiveness of the sparse-constraint reconstruction using the self-learned dictionaries.

Qualitative evaluation. Figure 5 shows one selected band from the central view of the reconstructed hyperspectral light field for each scene when S equals to 5. We can see that, on the one hand, the original image is decently recovered under the hybrid imaging model through either TwIST or DBR. On the other hand, as can be easily observed from the zoom-in results, TwIST tends to smear out the object details due to its local smoothness prior, while DBR better preserves the object details by further exploiting the correlations across the angular and spectral dimensions through the self-learned dictionaries.

Table 1. Quantitative evaluation of two reconstruction methods at different angular resolutions.

Views	Scene	PSNR		SAM	
		TwIST	DBR	TwIST	DBR
5×5	*Boards*	34.44	37.29	0.097	0.079
	Toys	36.43	38.60	0.077	0.065
	Fruits	36.35	38.86	0.074	0.065
7×7	*Boards*	33.98	36.80	0.098	0.082
	Toys	36.18	38.12	0.078	0.066
	Fruits	35.97	37.9	0.077	0.067
9×9	*Boards*	32.99	36.30	0.108	0.092
	Toys	35.71	38.00	0.087	0.074
	Fruits	35.72	38.01	0.082	0.072
Average		**35.31**	**37.76**	**0.086**	**0.073**

View-wise and band-wise evaluation. For an inspection across the angular and spectral dimensions, Figure 6 shows the view-wise and band-wise PSNR results of the DBR reconstruction of the *Boards* scene when S equals to 5. As demonstrated, in terms of the spectral dimension, the PSNR of each band in a certain view varies along with wavelength, due to the non-uniform spectral sensitivity of the detectors. In terms of the angular dimensions, all views share a similar PSNR distribution, as they are treated equally in simulation. In practice, due to the vignetting effect of the micro-lens-array in the light field camera, the central view generally has higher reconstruction fidelity than the corner views, as can be seen from the hardware experimental results.

Spectral signature evaluation. For a more comprehensive comparison, Figure 7 shows the recovered spectral signatures at two selected spatial points from the central view of the *Boards* scene when S equals to 5. Besides TwIST and DBR, we also generate the spectral signature through interpolation from the RGB values [23]. As can be seen, the interpolation results have a large deviation from the groundtruth, while the DBR results are the closest to the groundtruth. Table 2 gives the corresponding root-mean-square-error (RMSE) of the recovered signatures by different methods, which confirms the superiority of DBR.

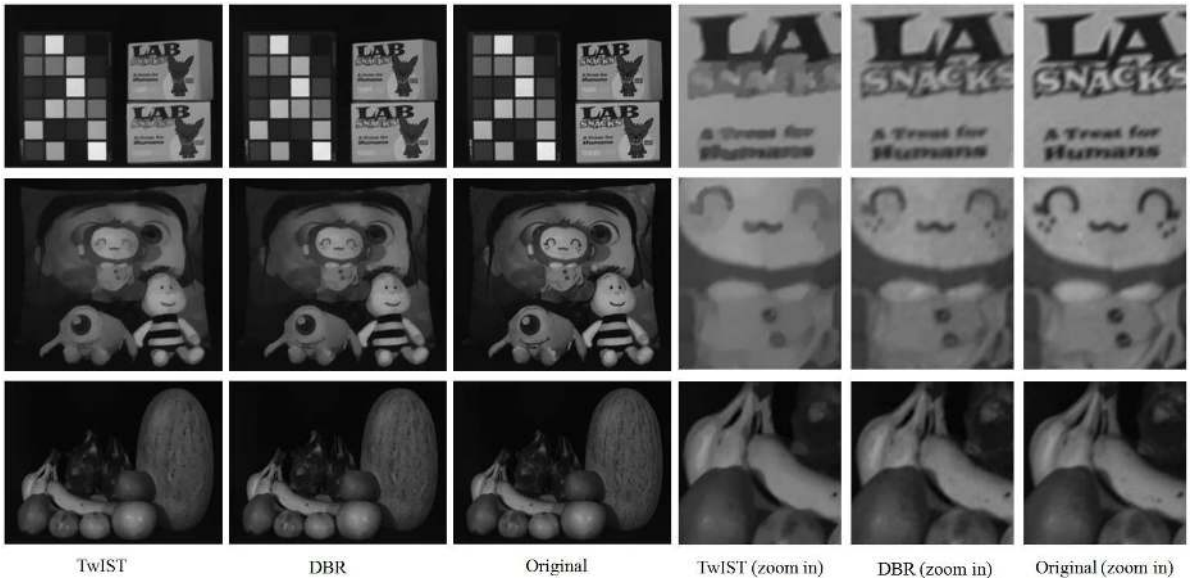


Figure 5. Reconstruction results of one selected band from the central view of three scenes (5×5 views). From top to bottom: *Boards* (620nm), *Toys* (570nm), and *Fruits* (520nm). (Please see the electronic version for better visualization.)

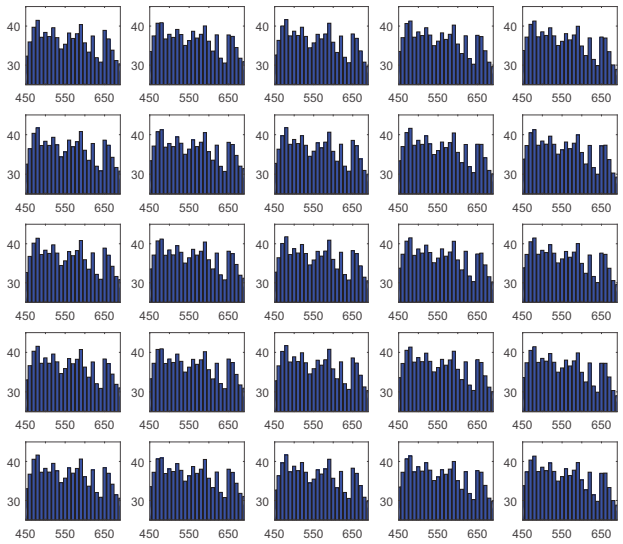
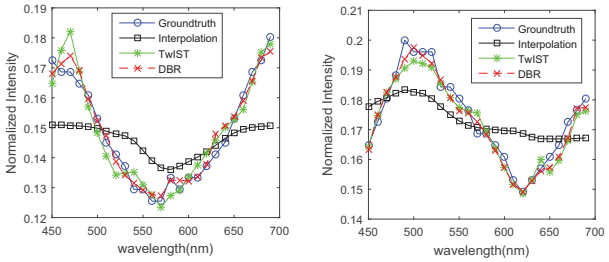


Figure 6. View-wise and band-wise PSNR results of the DBR reconstruction of the *Boards* scene (5×5 views). Horizontal axis for wavelength and vertical axis for PSNR in each view.



(a) (b)

Figure 7. Spectral signatures at two selected spatial points from the central view of the *Boards* scene (5×5 views).

Table 2. RMSE of spectral signatures at two points in Figure 7.

Point	DBR	TwIST	Interpolation
(a)	0.0029	0.0038	0.0106
(b)	0.0031	0.0050	0.0132

6. Experiments

6.1. Hardware system

Figure 8 demonstrates the prototype system we have developed for snapshot hyperspectral light field imaging. The incident light from the scene is equally divided by a beam splitter and captured by Lytro and CASSI, respectively. The Lytro branch captures an RGB light field with 9×9 views at a spatial resolution of 380×380. The main objective lens of Lytro has a fixed f-number of f/2 [10]. In the CASSI branch, an 8mm objective lens is used to project the scene onto a coded aperture, for which the f-number is also set to f/2 to match with Lytro. The manufactured coded aperture is a random binary pattern with 300×300 elements and each element has a size of $10\mu\text{m} \times 10\mu\text{m}$. A double Amici prism vertically disperses the spectrum with the center wavelength at 550nm. Each element on the coded aperture is mapped to 2×2 pixels on a panchromatic detector (PointGrey FL3-U3-13Y3M-C) by a relay lens (Edmund 45-762), so the spatial resolution of the CASSI measurement is 600×600. An optical filter with a passband from 500nm to 700nm is used in each branch to restrict the spectrum to the same range.

The calibration of our system contains two steps: cal-

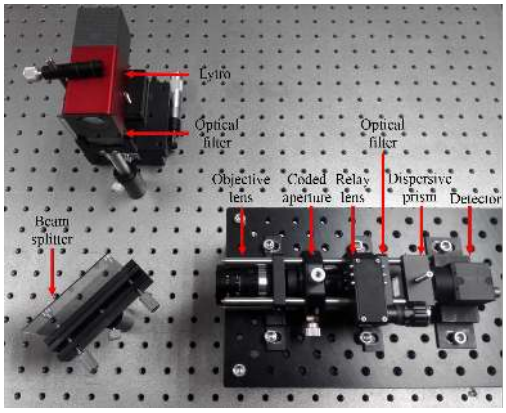


Figure 8. Prototype of snapshot hyperspectral light field imager.

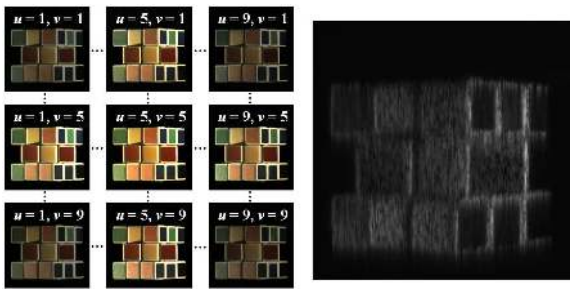


Figure 9. Measurements of the test scene. Left: RGB light field from Lytro (containing 9×9 angular views). Right: compressive measurement from CASSI (encoding 27 spectral bands).

ibration of CASSI and calibration between Lytro and CASSI. The CASSI calibration is conducted following the procedures in the seminal work [26], from which the observation matrix of CASSI is obtained and the entire spectrum spanning over the passband of the optical filter is discretized into 27 bands with different intervals. The calibration between Lytro and CASSI is conducted using a checkerboard scene. Owing to the CASSI calibration, we only need to align the central view of Lytro with the projection of one wavelength on the CASSI detector, and the alignment with other wavelengths can then be easily deduced. To this end, the checkerboard is illuminated by monochromatic light and captured by Lytro and CASSI simultaneously. Once the optical components in the CASSI branch are fixed, we adjust the position of Lytro so that the checkerboard occupies an area with the same resolution both in the central view of Lytro and on the CASSI detector. Therefore, the 5D hyperspectral light field to be recovered has a resolution of $380(W) \times 380(H) \times 9(S) \times 9(S) \times 27(\Omega)$.

6.2. Results

To evaluate the performance of our system, we test a scene of a twisted Rubik's Cube with distinct colors. The exposure time is 5ms for Lytro and 60ms for CASSI respectively, under the illumination of a tungsten halogen lamp. The measurements from Lytro and CASSI are shown in Figure 9. Note the vignetting effect in the RGB light field from Lytro, where the central view has higher intensity and higher signal-to-noise ratio than the corner views. The parameters used for the DBR reconstruction are the same as in simulation. The reconstruction results of five selected bands from the central view are shown in Figure 10(a). It can be seen that the proposed approach decently recovers the scene content, although some noise is visible due to the imperfect measurements and system observation matrices obtained in practice.

For a quantitative evaluation on the reconstruction fidelity, we inspect the spectral signatures at two selected spatial points from the central view. The reference signatures are measured by a probe spectrometer (Stellar-Net BLK-CXR-SR-50 with 1.3nm spectral precision). Figure 11 shows the comparison results with interpolation from the RGB values [23]. We can see the DBR signatures well match the reference while the interpolation ones have a large deviation, which validates the reconstruction fidelity of the proposed approach. Table 3 gives the RMSE of these signatures with respect to the reference, which confirms the superior performance of the proposed approach.

For an inspection across the angular dimensions, Figure 10(b)-(c) show the reconstruction results of two band-wise light fields (9 out of 81 views are shown due to space limitation). As can be seen, the reconstruction results are also influenced by the vignetting effect in the Lytro measurement. Consequently, the corner views have lower reconstruction fidelity than the central view. (Ideally, the band-wise light field should have a uniform intensity distribution regardless of the views as in simulation.) This issue could be alleviated by advanced vignetting correction.

7. Conclusion and discussion

In this paper, we have presented the first snapshot hyperspectral light field imager in practice. Specifically, we designed a novel hybrid camera system to obtain two complementary measurements that sample the angular and spectral dimensions respectively. To recover the full 5D hyperspectral light field from severely undersampled measurements, we then proposed an efficient computational reconstruction algorithm by exploiting the large correlations across the spectral and angular dimensions through self-learned dictionaries. Extensive simulation on an elaborate hyperspectral light field dataset and preliminary hardware experimental results validate the performance of the proposed approach.

Potential applications. Snapshot hyperspectral light field imaging could unblock a number of applications where both spectral and angular information are necessary in time-critical scenarios. For example, both hyperspectral and light field imaging help computer vision tasks such as classifi-

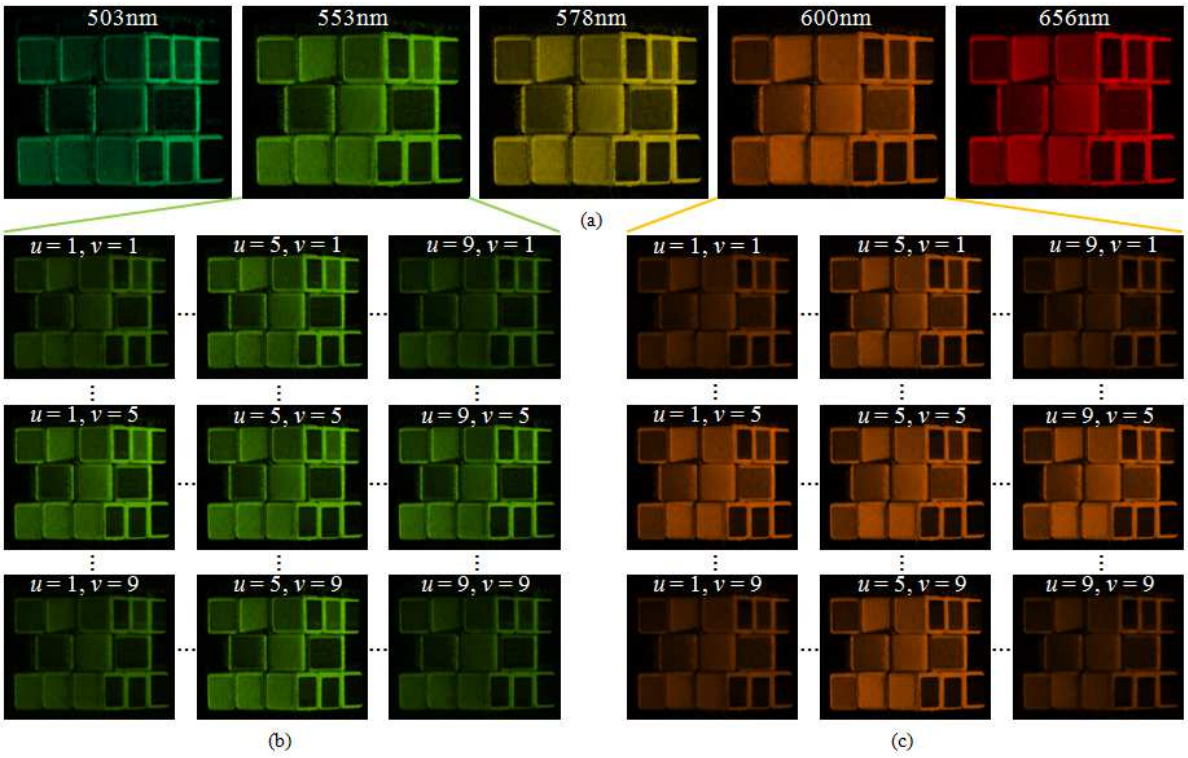


Figure 10. Reconstruction results of the test scene. (a) Five selected bands from the central view. (b) Two band-wise light fields with 9 out of 81 views. The corner views are darker than the central view due to vignetting. (Please see the electronic version for better visualization.)

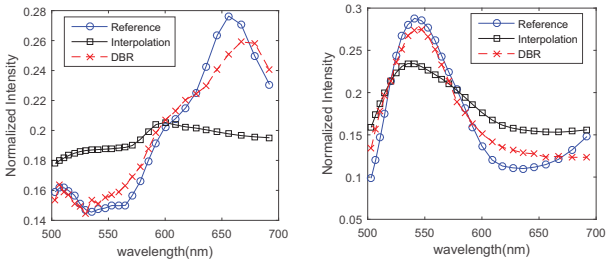


Figure 11. Spectral signatures at two selected spatial points from the central view of the test scene.

cation and recognition, and encompassing higher plenoptic dimensions in an efficient way will reasonably promote the overall performance of these tasks, especially when spatial information alone is not sufficient; from the graphics perspective, the capability of capturing a hyperspectral light field in a single shot will facilitate the rendering of dynamic scenes in a more realistic way; other occasions such as heritage digitalization and scientific observation will also benefit from snapshot hyperspectral light field imaging.

Advanced issues. While the proposed approach for hyperspectral light field imaging demonstrates encouraging results, there is still room for improvement. First, currently we assume the Lytro camera directly outputs view-wise RGB images, without taking into account the preprocessing steps, such as demosaicing and rectification, which convert the raw measurement into the RGB light field. Therefore, the observation matrix of Lytro used in this work is not accurate enough, and can be improved by incorporating the preprocessing steps. Second, due to the extremely high data dimensionality, the computational reconstruction suffers from high complexity. Our unoptimized MATLAB code takes about 16 hours for reconstructing a hyperspectral light field at a resolution of $380\times380\times9\times9\times27$ on a mainstream CPU including dictionary learning. Parallelized algorithms are needed to accelerate the reconstruction. Last but not least, the proposed approach has the potential for 6D hyperspectral light field video acquisition due to its snapshot property, which is regarded as our future work.

Table 3. RMSE of spectral signatures at two points in Figure 11.

Point	DBR	Interpolation
(a)	0.0098	0.0382
(b)	0.0177	0.0383

Acknowledgments

We acknowledge funding from the NSFC grants 61671419 and 61425026 and the CAS Pioneer Hundred Talents Program.

References

[1] M. Aharon, M. Elad, and A. Bruckstein. K-svd: An algorithm for designing overcomplete dictionaries for sparse representation. *IEEE Trans. Signal Process.*, 54(11):4311–4322, 2006.

[2] G. Arce, D. Brady, L. Carin, H. Arguello, and D. Kittle. Compressive coded aperture spectral imaging: An introduction. *IEEE Signal Process. Mag.*, 31(1):105–115, 2014.

[3] M. Ben-Ezra and S. K. Nayar. Motion deblurring using hybrid imaging. In *CVPR*, 2003.

[4] A. Bhandari and R. Raskar. Signal processing for time-of-flight imaging sensors: An introduction to inverse problems in computational 3-d imaging. *IEEE Signal Process. Mag.*, 33(5):45–58, 2016.

[5] J. Bioucas-Dias and M. Figueiredo. A new twist: Two-step iterative shrinkage/thresholding algorithms for image restoration. *IEEE Trans. Image Process.*, 16(12):2992 – 3004, 2007.

[6] V. Boominathan, K. Mitra, and A. Veeraraghavan. Improving resolution and depth-of-field of light field cameras using a hybrid imaging system. In *ICCP*, 2014.

[7] E. Candes. Compressive sampling. In *Int. Congr. Math.*, 2006.

[8] X. Cao, T. Yue, X. Lin, S. Lin, X. Yuan, Q. Dai, L. Carin, and D. J. Brady. Computational snapshot multispectral cameras: Toward dynamic capture of the spectral world. *IEEE Signal Process. Mag.*, 33(5):95–108, 2016.

[9] D. Donoho. Compressed sensing. *IEEE Trans. Inf. Theory*, 52(4):1289–1306, 2006.

[10] T. Georgiev, Z. Yu, A. Lumsdaine, and S. Goma. Lytro camera technology: theory, algorithms, performance analysis. In *Proc. SPIE*, 2013.

[11] J. Han, L. Shao, D. Xu, and J. Shotton. Enhanced computer vision with microsoft kinect sensor: A review. *IEEE Trans. Cybernetics*, 43(5):1318–1334, 2013.

[12] I. Ihrke, J. Restrepo, and L. Mignard-Debise. Principles of light field imaging: Briefly revisiting 25 years of research. *IEEE Signal Process. Mag.*, 33(5):59–69, 2016.

[13] H. G. Jeon, J. Park, G. Choe, J. Park, Y. Bok, Y. W. Tai, and I. S. Kweon. Accurate depth map estimation from a lenslet light field camera. In *CVPR*, 2015.

[14] M. H. Kim, T. A. Harvey, D. S. Kittle, H. Rushmeier, J. Dorsey, R. O. Prum, and D. J. Brady. 3d imaging spectroscopy for measuring hyperspectral patterns on solid objects. *ACM Trans. Graph.*, 31(4):38, 2012.

[15] F. Kruse, A. Lefkoff, J. Boardman, K. Heidebrecht, A. Shapiro, P. Barloon, and A. Goetz. The spectral image processing system (sips)-interactive visualization and analysis of imaging spectrometer data. *Remote Sensing of Environment*, 44(2):145–163, 1993.

[16] R. Leitner, A. Kenda, and A. Tortschanoff. Hyperspectral light field imaging. In *Proc. SPIE*, 2015.

[17] M. Levoy, R. Ng, A. Adams, M. Footer, and M. Horowitz. Light field microscopy. *ACM Trans. Graph.*, 25(3):924–934, 2006.

[18] X. Lin, Y. Liu, J. Wu, and Q. Dai. Spatial-spectral encoded compressive hyperspectral imaging. *ACM Trans. Graph.*, 33(6):233:1–233:11, 2014.

[19] X. Lin, G. Wetzstein, Y. Liu, and Q. Dai. Dual-coded compressive hyperspectral imaging. *Opt. Lett.*, 39(7):2044–2047, 2014.

[20] C. Ma, X. Cao, X. Tong, Q. Dai, and S. Lin. Acquisition of high spatial and spectral resolution video with a hybrid camera system. *Int. J. Comput. Vision*, 110(2):141–155, 2014.

[21] K. Marwah, G. Wetzstein, Y. Bando, and R. Raskar. Compressive light field photography using overcomplete dictionaries and optimized projections. *ACM Trans. Graph.*, 32(4):46, 2013.

[22] R. Ng, M. Levoy, M. Brédif, G. Duval, M. Horowitz, and P. Hanrahan. Light field photography with a hand-held plenoptic camera. *Computer Science Technical Report, Stanford University*, 2(11):1–11, 2005.

[23] B. Smits. An rgb-to-spectrum conversion for reflectances. *J. Graph. Tools*, 4(4):11–22, 1999.

[24] J. A. Tropp and A. C. Gilbert. Signal recovery from random measurements via orthogonal matching pursuit. *IEEE Trans. Inf. Theory*, 53(12):4655–4666, 2007.

[25] A. Wagadarikar, R. John, R. Willett, and D. Brady. Single disperser design for coded aperture snapshot spectral imaging. *Appl. Opt.*, 47(10):B44 – B51, 2008.

[26] A. Wagadarikar, N. Pitsianis, X. Sun, and D. Brady. Video rate spectral imaging using a coded aperture snapshot spectral imager. *Opt. Express*, 17(8):6368 – 88, 2009.

[27] L. Wang, Z. Xiong, D. Gao, G. Shi, and F. Wu. Dual-camera design for coded aperture snapshot spectral imaging. *Appl. Opt.*, 54(4):848–858, 2015.

[28] L. Wang, Z. Xiong, D. Gao, G. Shi, W. Zeng, and F. Wu. High-speed hyperspectral video acquisition with a dual-camera architecture. In *CVPR*, 2015.

[29] L. Wang, Z. Xiong, G. Shi, W. Zeng, and F. Wu. Simultaneous depth and spectral imaging with a cross-modal stereo system. *IEEE Trans. Circuits Syst. Video Technol.*, to be published.

[30] T. C. Wang, A. A. Efros, and R. Ramamoorthi. Occlusion-aware depth estimation using light-field cameras. In *ICCV*, 2015.

[31] S. Wanner and B. Goldluecke. Variational light field analysis for disparity estimation and super-resolution. *IEEE Trans. Pattern Anal. Mach. Intell.*, 36(3):606–619, 2014.

[32] G. Wetzstein, I. Ihrke, A. Gukov, and W. Heidrich. Towards a database of high-dimensional plenoptic images. In *ICCP*, 2011.

[33] G. Wetzstein, I. Ihrke, D. Lanman, and W. Heidrich. Computational plenoptic imaging. *Comput. Graph. Forum*, 30(8):2397–2426, 2011.

[34] Z. Xiong, Y. Zhang, F. Wu, and W. Zeng. Computational depth sensing: Toward high-performance commodity depth cameras. *IEEE Signal Process. Mag.*, to be published.

[35] Y. Yoon, H. G. Jeon, D. Yoo, J. Y. Lee, and I. S. Kweon. Learning a deep convolutional network for light-field image super-resolution. In *ICCVW*, 2015.