Research Article | Vol. 25, No. 13 | 26 Jun 2017 | OPTICS EXPRESS 14008
Optics EXPRESS

Heterogeneous camera array for multispectral light field imaging

YANG ZHAO,[1] TAO YUE,[1,3] LINSEN CHEN,[1] HONGYUAN WANG,[2] ZHAN MA,[1] DAVID J. BRADY,[2] AND XUN CAO[1,*]

[1]*School of Electronic Science and Engineering, Nanjing University, Nanjing 210000, China*
[2]*Department of Electrical and Computer Engineering and The Fitzpatrick Institute for Photonics, Duke University, Durham, NC 27708, USA*
[3]*Authors contributed equally to this work.*
[*]*caoxun@nju.edu.cn*

Abstract: Multispectral light field acquisition is challenging due to the increased dimensionality of the problem. In this paper, inspired by anaglyph theory (i.e. the ability of human eyes to synthesize colored stereo perception from color-complementary (such as red and cyan) views), we propose to capture the multispectral light field using multiple cameras with different wide band filters. A convolutional neural network is used to extract the joint information of different spectral channels and to pair the cross-channel images. In our experiment, results on both synthetic data and real data captured by our prototype system validate the effectiveness and accuracy of proposed method.

OCIS codes: (110.4234) Multispectral and hyperspectral imaging; (200.4260) Neural networks; (100.4145) Motion, hyperspectral image processing.

References and links

1. N. Gat, "Imaging spectroscopy using tunable filters: a review," Proc. SPIE **4056**(1), 50–64 (2000).
2. K. C. Lawrence, B. Park, W. R. Windham, and C. Mao, "Calibration of a pushbroom hyperspectral imaging system for agricultural inspection," Trans. Ame. Soc. of Agril. Engg. **46**(2), 513 (2003).
3. 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 Sig. Proc. Mag. **33**(5), 95–108 (2016).
4. M. Descour and E. Dereniak, "Computed-tomography imaging spectrometer: Experimental calibration and reconstruction results," Appl. Opt. **34**(22), 4817–4817 (1995).
5. D. J. Brady and M. E. Gehm, "Compressive imaging spectrometers using coded apertures," Proc. SPIE **6246**, 62460A (2006).
6. A. Mrozack, D. L.Marks, and D. J. Brady, "Coded aperture spectroscopy with denoising through sparsity," Opt. Express **20**(3), 2297–2309 (2012).
7. D. J. Brady, K. Choi, D. L.Marks, R. Horisaki, and S. Lim, "Compressive Holography," Opt. Express **17**(15), 13040–13049 (2009).
8. X. Lin, G. Wetzstein, Y. Liu, and Q. Dai, "Dual-coded compressive hyperspectral imaging," Opt. Lett. **39**(7), 2044–2047 (2014).
9. H. Rueda, H. Arguello, and G. R. Arce, "Dual-ARM VIS/NIR compressive spectral imager," in Proceedings of IEEE International Conference on Image Processing (IEEE, 2006), pp. 2572–2576.
10. X. Cao X, H. Du, and X. Tong, "A Prism-Mask System for Multispectral Video Acquisition," IEEE Trans. Pat. Ana. Machine Intell. **33**(12), 2423–2435 (2011).
11. J. Jia, K. J. Barnard, and K. Hirakawa, "Fourier Spectral Filter Array for Optimal Multispectral Imaging," IEEE Trans. Image Process. **25**(4), 1 (2016).
12. L. Mcmillan and G. Bishop, "Plenoptic modeling: an image-based rendering system," in Proceedings of Conference on Computer Graphics and Interactive Techniques. ACM **29**(5), 39–46 (1995).
13. M. Landy and J. Movshon, "The plenoptic function and the elements of early vision," MIT Press **1**, 3–20 (1997).
14. M. Levoy, "Light fields and computational imaging," Computer **39**(8) 46–55 (2006).
15. S. Gortler, R. Grzeszczuk R. Szeliski, and M. Cohen, "The Lumigraph," in Proceedings of Conference on Computer Graphics and Interactive Techniques. ACM **96**, 43–54 (2001).
16. D. Wood, D. Azuma, K. Aldinger, B. Curless, T. Duchamp, D. H. Salesin, and W. Stuetzle, "Surface light fields for 3D photography," in Proceedings of the Conference on Computer Graphics and Interactive Techniques, ACM 287-296 (2000).
17. B. Wilburn, N. Joshi, V. Vaish, E. V. Talvala, E. Antunez, A. Barth, A. Adams, M. Horowitz, and M. Levoy, "High performance imaging using large camera arrays," ACM Trans. on Graphics **24**(3), 765–776 (2005).

#290154 https://doi.org/10.1364/OE.25.014009
Journal © 2017 Received 9 Mar 2017; revised 29 Apr 2017; accepted 17 May 2017; published 12 Jun 2017

Optics EXPRESS

18. E. H. Adelson and J. Y. Wang, "Single lens stereo with a plenoptic camera," IEEE Trans. on Pattern Analysis and Machine Intelligence **14**(2), 99–106 (1992).
19. L. Marc, N. Ren, A. Andrew, F. Matthew, and H. Mark, "Light field microscopy," ACM Trans. on Graphics **25**(3), 924–934 (2006).
20. J. P. Lewis, "Fast normalized cross-correlation," Vision Interface **10**(1), 120–123 (1995).
21. X. Shen, L. Xu, Q. Zhang, and J. Jia, "Multi-modal and multi-spectral registration for natural images," Euro. Conf. Comput. Vision 2014, pp. 309–324.
22. E. T. Psota, J. Kowalczuk, M. Mittek, and L. C. Perez, "MAP Disparity Estimation Using Hidden Markov Trees," in Proceedings of the IEEE International Conference on Computer Vision (IEEE, 2015), pp. 2219–2227.
23. J. Zbontar and Y. LeCun, "Computing the stereo matching cost with a convolutional neural network," in Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (IEEE, 2015), pp. 1592–1599.
24. J. Zbontar and Y. LeCun, "Stereo matching by training a convolutional neural network to compare image patches," J. Mach. Learn. Res. **17**, 1–32 (2016).
25. Y. Lecun, "Learning Invariant Feature Hierarchies," Euro. Conf. Comput. Vision 2012, pp. 496–505.
26. M. Menze and A. Geiger, "Object scene flow for autonomous vehicles," in Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (IEEE, 2015), pp. 3061–3070.
27. M. Menze, C. Heipke, and A. Geiger, "Joint 3d estimation of vehicles and scene flow," ISPRS Ann. Photogramm. Remote Sens. Spatial Inf. Sci., II-3/W5, 427–434 (2015).
28. H. Hirschmüller and D. Scharstein, "Evaluation of cost functions for stereo matching," in Proceedings of the IEEE Computer Society Conference on Computer Vision and Pattern Recognition (IEEE, 2007), pp. 1–8.
29. D. Scharstein, H. Hirschmüller, Y. Kitajima, G. Krathwohl, N. Nesic, X. Wang, and P. Westling, "High-resolution stereo datasets with subpixel-accurate ground truth," German Conference on Pattern Recognition **8753**, 31–42 (2014).
30. J. I. Park, M. H. Lee, M. D. Grossberg, and S. K. Nayar, "Multispectral imaging using multiplexed illumination," in Proceedings of IEEE International Conference on Computer Vision (IEEE, 2007), pp. 1–8.
31. PointGrey, "Grasshopper3 5.0 MP Color USB3 Vision," https://www.ptgrey.com/grasshopper3-50-mp-color-usb3-vision-sony-pregius-imx250.
32. J. Heikkila and O. Silvén, "A four-step camera calibration procedure with implicit image correction," in Proceedings of the IEEE Computer Society Conference on Computer Vision and Pattern Recognition (IEEE, 1997), pp. 1106–1112.
33. Y. S. Kang, C. Lee, and Y. S. Ho, "An efficient rectification algorithm for multi-view images in parallel camera array," 3DTV Conference: The True Vision-Capture, Transmission and Display of 3D Video (IEEE, 2008), pp. 61–64.
34. 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. Vis. **110**(2), 141–155 (2014).
35. Autodesk 3ds Max, "3D computer graphics program" http://www.autodesk.com/products/3ds-max/overview.
36. Nguyen, M. H. Rang, D. K. Prasad, and M. S. Brown, "Training-Based Spectral Reconstruction from a Single RGB Image," Euro. Conf. Comput. Vision 2014, 186–201.

1. Introduction

Both light field and multispectral imaging are hot research topics in computational photography for their potential applications on various computer vision tasks and many other scenarios such as remote sensing. Compared with traditional photography, they provides extra information from either angular or spectral dimensions of light rays.

Many commercial multispectral cameras capture different channels sequentially by the aid of tunable filters [1] or push-broom imaging frameworks [2]. Lots of snap-shot hyperspectral imaging systems [3] such as Computed Tomography Imaging Spectrometer (CTIS) [4], Coded Aperture Snapshot Spectral Imager (CASSI) [5–9] and Prism-based Multispectral Video Imaging Spectrometer (PMVIS) [10] have been proposed to capture videos [11]. Besides, 4D light fields are proposed to simplify the 7D plenoptic function [12–15]and several methods have been proposed for capturing the light fields for both static and dynamic scenes [16–19]. However, capturing the multispectral light field is still difficult due to the increased dimensionality of the problem.

Inspired by Anaglyph 3D theory, i.e., the ability of humans to synthesize full-color stereoscopic perceptual by encoding binocular views with chromatically complemented color filters (typically red and cyan), we try to extend the binocular stereo sensing to multi-camera cases for multispectral light field imaging. The main challenge for this idea is how to pair the heterogeneous images captured at different views with different color filters. Existing stereo matching algorithms such as Normalized Cross Correlation (NCC) [20, 21] and hidden Markov tree model [22] cannot

Optics EXPRESS

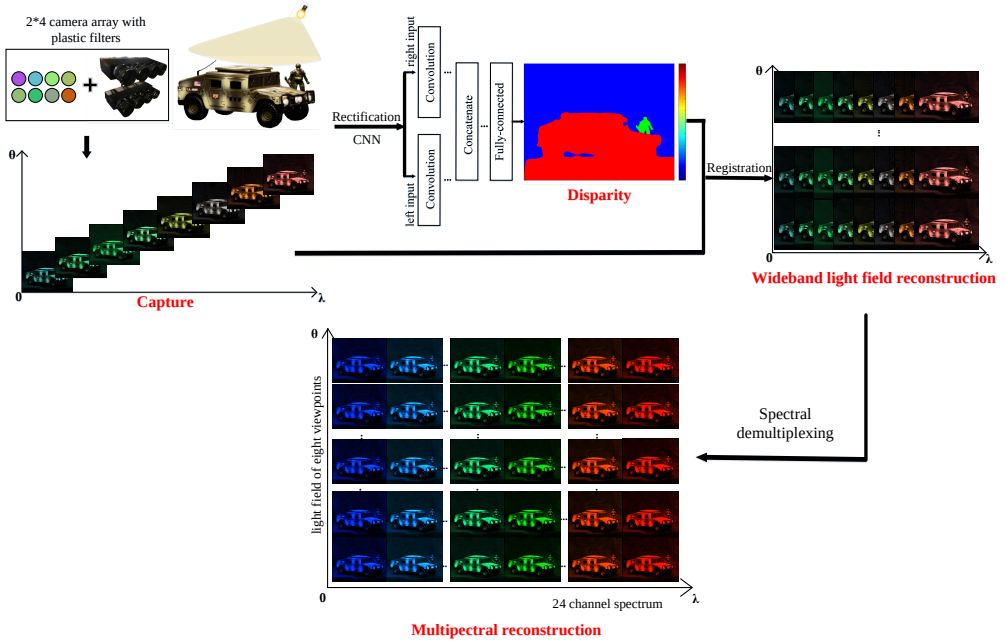

Fig. 1. This image shows an overview of our system. Our system introduces the stereo matching method with convolutional neural network and exploits the different spectral sensitivities of the filter array to reconstruct multispectral light field through the CNN-based heterogeneous stereo matching and spectral demultiplexing.

handle this problem because they assume the corresponding points in different views share the same intensities, which leads to the most important fidelity term in the objective function of their matching algorithm. In the cases that this assumption does not hold, most existing stereo matching algorithms fail to offer reasonable output.

However, in natural physiological phenomena, we, human beings can handle two heterogeneous views easily, even without any pre-training. Hence, it implies the solvability of aforementioned problem, and further more indicates that the intensity fidelity constraint does not play an important role in human visual system.

Recently, with the development of deep neural networks, tasks that have puzzled computer scientists a long time but that can be well interpreted by human brains, are perfectly resolved. For example, Zbontar and Yan [23, 24] propose a stereo matching method by training a convolutional neural network (CNN) to simulate human eye's behavior for image patch comparison and depth information extraction from a rectified image pair. Motivated by this work, we attempt to train a deep network which can handle the heterogeneous stereo matching like human brains do. The Siamese network (Bromley et.al., 1993), which is composed by two identical or similar sub-networks, is applied to handle the stereo image pairs [25]. Different channels of images in standard stereo matching datasets,such as KITTI [26, 27] and Middleburry vision benchmark [28, 29] are used to generate the heterogeneous training images.

Meanwhile, we present a prototype array system composed of eight cameras using heterogeneous wide-band color filters to capture the multispectral light field at the same time. A spectral de-multiplexing algorithm is proposed to extract 24 spectral channels from eight heterogeneously filtered trichromatic cameras. In particular, we conduct the stereo matching among different wideband-filtered images captured at different views by training a convolutional neural network,

Optics EXPRESS

and construct a 24-channel light field with different spectral response curves by warping the images according to the estimated stereo matching. By delicately designing the broad band spectral filters, the 24-spectral-channel light field can be computed using the demultiplexing algorithm [30].

In all, there are three main contributions of this work,e.g., (1) we propose to capture the multispectral light field using a camera array where each camera is coupled with a heterogeneous wide-band color filter; (2) we demonstrate a heterogeneously matching algorithm by using Convolutional Neural Network to simulate human eyes; (3) we present a prototype system with eight cameras for high quality 24-channel spectral light field imaging to capture both indoor and outdoor, static and dynamic scenes.

2. Camera array system for multispectral light field imaging

In this section, we present our camera array system using heterogeneous wide-band color filters for multispectral light field imaging.

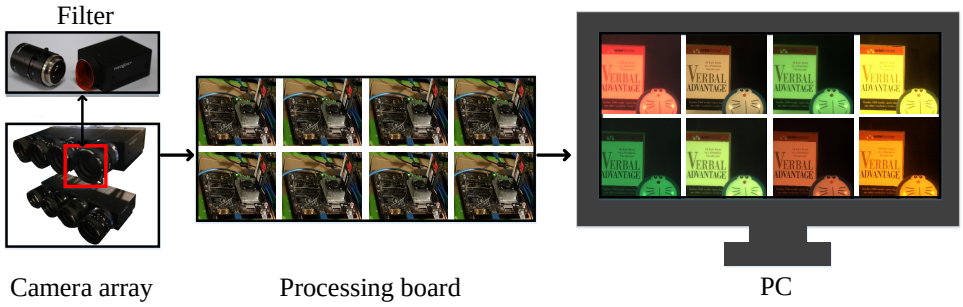

Fig. 2. Camera array configuration.

2.1. System overview

As shown in Fig. 1, we reconstruct the multispectral light field by capturing the heterogeneous multiview images by using our eight camera array system. Fig. 2 shows the system architecture of our proposed system, which consists of three main subsystems: heterogeneous cameras, processing boards (Nvidia Jetson TX1), and a PC. Considering the trade off between the system complexity and the number of captured channels, without loss of generality, a prototype system composed of eight cameras are used to take 24 channels per snapshot in this paper. However, the proposed method can be easily extended to capture light fields with more spectral channels and viewpoints. We use eight off-the-shelf RGB cameras, e.g., Point Grey GS3-U3-51S5C-C [31] with $25mm(F/16)$ lenses, each of which offers spatial resolution at 2448×2048 and temporal resolution up-to 75 frames per seconds (FPS). These eight cameras are mounted parallelly on a printed metal stand as a 2×4 camera array to enable us to move them flexibly. Each camera is connected to a Nvidia Jetson TX1 board for image processing (eg. data compression, ISP and data storage) before sending it to the PC in either raw/JPEG form or as an MPEG2 video stream. The PC host controls the system configuration (such as initialization, synchronous triggering) and data postprocessing (such as image stitching). Eight plastic filters with different color bands are mounted with cameras (as shown in Fig. 2) to capture spectrally heterogeneous images. Besides, it is noted that the default white balance must be turned off to avoid the unnecessary color manipulation during acquisition.

Optics EXPRESS

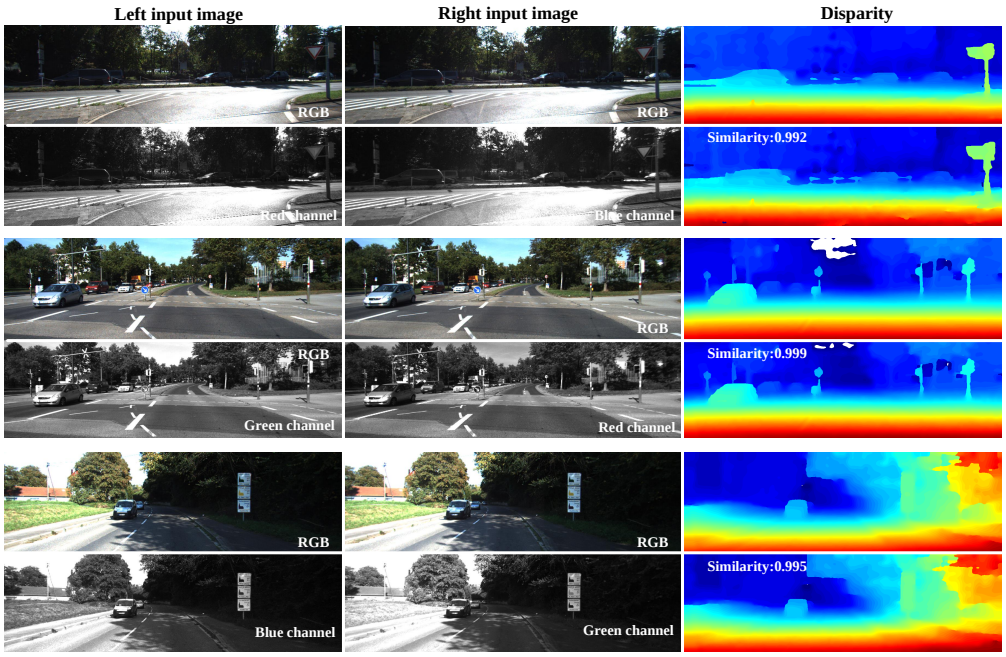

Fig. 3. Examples of predicted disparity maps on the KITTI 2015 dataset [26, 27] using our proposed method with different channel inputs (the even rows), as well as the results of Zbontar et.al. [24] with full-channel inputs (the odd rows). note that objects closer to the camera have larger disparities than objects farther away, with warmer colors representing larger values of disparity and smaller values of depth. When taking the single channel as the input, we try different pairs of RGB channels, and they all get very high similarities.

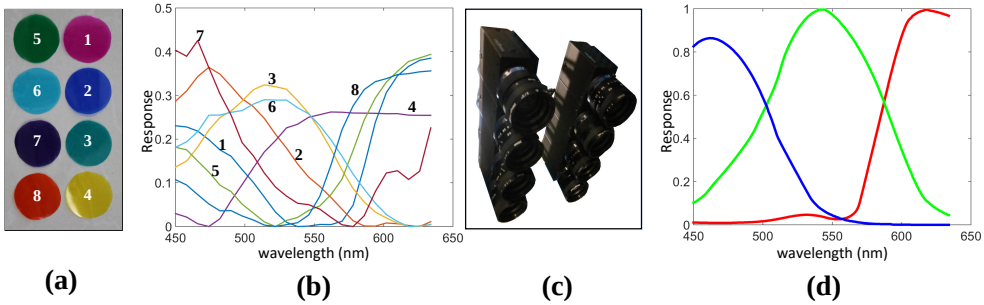

Fig. 4. Camera array with heterogeneous wide-band color filters. (a) shows eight qualified plastic filters used in our prototype camera-array system; (b) respectively illustrates their spectral sensitivities, which provide enough independent measurements of incoming light spectrum. The standard spectral response of the Point Grey GS3-U3-51S5C-C camera sensor array (c) [31] used in our prototype system is shown in (d).

After capturing the spectrally heterogeneous images/videos, the disparity maps between views can be computed using the proposed heterogeneous stereo matching network and spectral de-multiplexing algorithm respectively. The rectification is applied before matching to correct

Optics EXPRESS

the system errors. The heterogeneous measurements from different views can be warped to their correspondences and form 24 wide-band channels of all the views. By applying the spectral de-multiplexing algorithm on 24-channel images of each view, the final multispectral light field can be reconstructed.

2.2. CNN-based heterogeneous stereo matching

A convolutional neural network (CNN) based heterogeneous stereo matching algorithm is developed to compute the correspondences between the heterogeneous multiview images.

2.2.1. Network architecture

We use the same network architecture as presented in [24]. To make the paper self-complete, we briefly introduce the networks here. The Siamese network [25], i.e., two shared-weight sub-networks with joint top layers are applied. Two sub-networks are composed of four spatial convolutional layers, followed by a rectified linear unit for each layer. For each convolutional layer, 112 3×3 filters are used to extract features. Four fully connected layers with 384 units are followed to estimate the disparity from the features. Each pixel is computed using a 9×9 patch where it locates at the center of the template and other positions are filled with neighbors. The raw outputs of the network still have some errors, especially in low-texture regions and occluded areas. Thus, a series of post-processing steps, i.e., cross-based cost aggregation, semi-global matching, a left-right consistency check, subpixel enhancement, a median, and a bilateral filter, are applied to refine the quality of raw disparity maps.

2.2.2. Model training

With the network architecture aforementioned, we train the model using heterogeneous images generated from the training datasets of KITTI 2015 [26, 27] and test the model with the testing dataset of KITTI 2015.

We train the network parameters using image pairs with different color channels. Specifically, we extract a single channel from the left image of a image pair, and as for the right image, one of the rest two channels are selected to make sure the input image pair has different channels. By traversing all the possbile combinations, 200×6 image pairs are generated to train our network.

The results of proposed method with different channel inputs, as well as the results of Zbontar et.al. [24] with full-channel inputs are shown in Fig. 3. It contains three pairs of examples from the KITTI2015 dataset [26,27], together with the disparity predictions under two single-color-channel inputs. Similarity is measured as the percentage of pixels where the two disparity maps differ less than two pixels. Although the inputs are reduced to single channel, our results are quite similar to those with full-channel inputs, except for some occluded areas. The results are in accordance with our expectations since the Convolutional Neural Network mimics the human eye neurons well and can extract feature vectors more accurately than traditional stereo methods without the need of intensity fidelity constraint. Therefore, the proposed CNN-based stereo matching method using anaglyph glasses can potentially work as well as human eyes, and thus we can get the light field information through camera-arrays comprising multiple single-channel spectral cameras (realized by placing filters in front of commercial RGB cameras).

2.3. Spectral demultiplexing

Given the camera spectral response, by assuming Lambertian scenes, the imaging model of proposed heterogeneous camera array system can be expressed as:

$$p_{m,k}(x) = \int_{\Omega} s(\lambda, x) c_k^{camera}(\lambda) c_m^{filter}(\lambda) d\lambda, \tag{1}$$

Optics EXPRESS

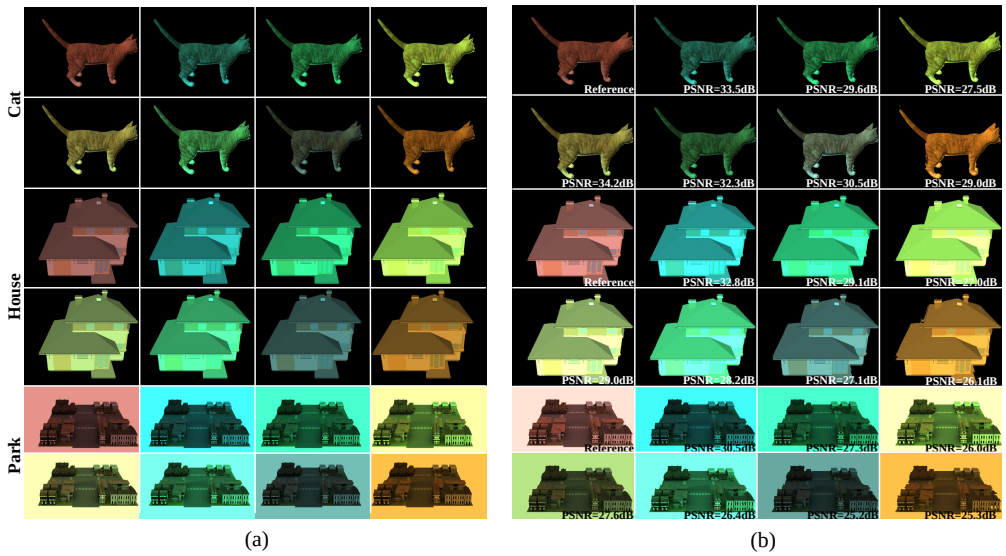

Fig. 5. (a) Simulated color images captured by 2×4 camera arrays with filters in front of them. (b) Simulated image registration with upper left image as the reference image.We also measure the Peak-Signal-to-Noise-Ratio(PSNR) for images in different viewpoints, and an increasing distance between target camera and reference camera decrease the accuracy of image registration, hence the multispectral reconstruction quality.

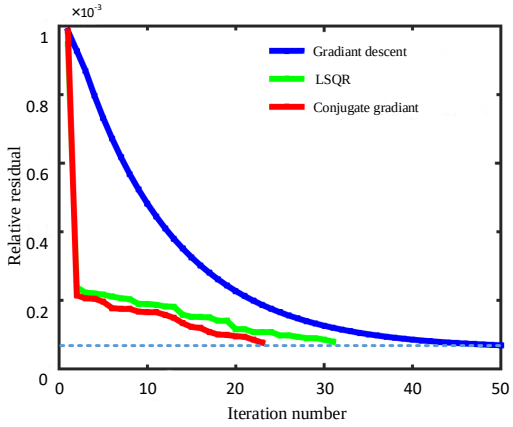

Fig. 6. The comparison of three commonly used optimization methods. These three optimization methods converge to the same solution, hence share the same accuracy. Furthermore, the running time is fastest for conjugate gradient method and slowest for gradient descent method with iteration steps in the same order of magnitude.

where $p_{m,k}(x)$ is the intensity of pixel x, $k \in \{r, g, b\}$ is the channel index of the image, and m is the camera/view index, $\Omega = [400nm, 700nm]$ is the range of the visible spectrum, $s(\lambda, x)$ is the spectral reflectance of scene point x, and $c_k^{camera}(\lambda)$ is the camera response curve of the k-th channel, $c_m^{filter}(\lambda)$ is the transmission curve of the filter at camera m.

For a multi-camera system with M trichromatic cameras ($M = 8$ in our system), we can capture

Optics EXPRESS

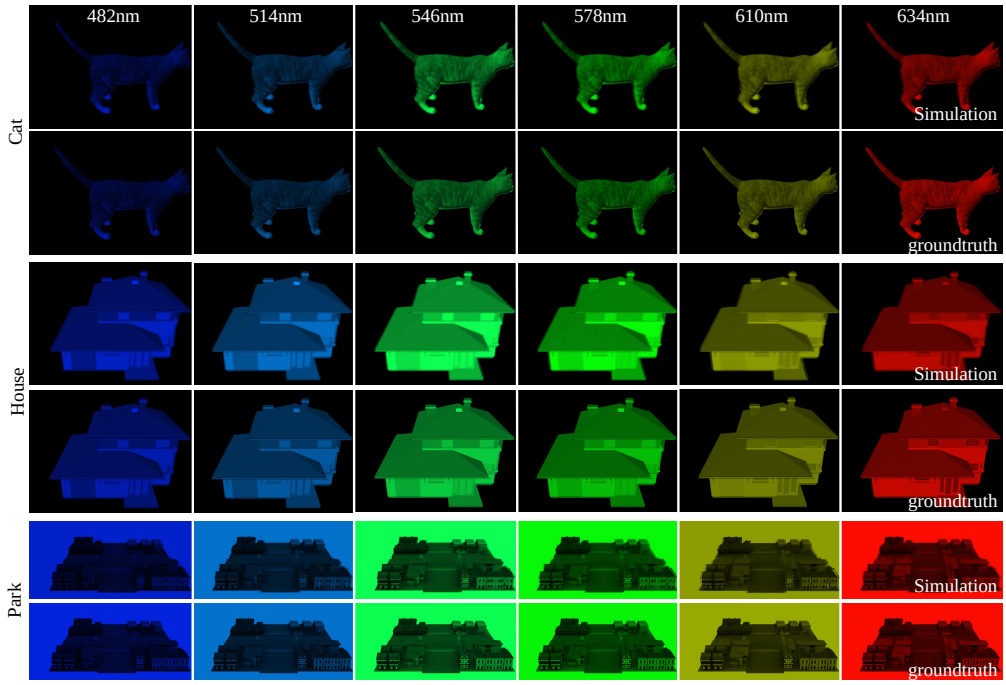

Fig. 7. Reconstructed multispectral channels of the first (top left) view of our eight camera array system. We select six single-spectral reflectance from all 24 reconstructed channels and compare the results with simulated ground truth reflectance.

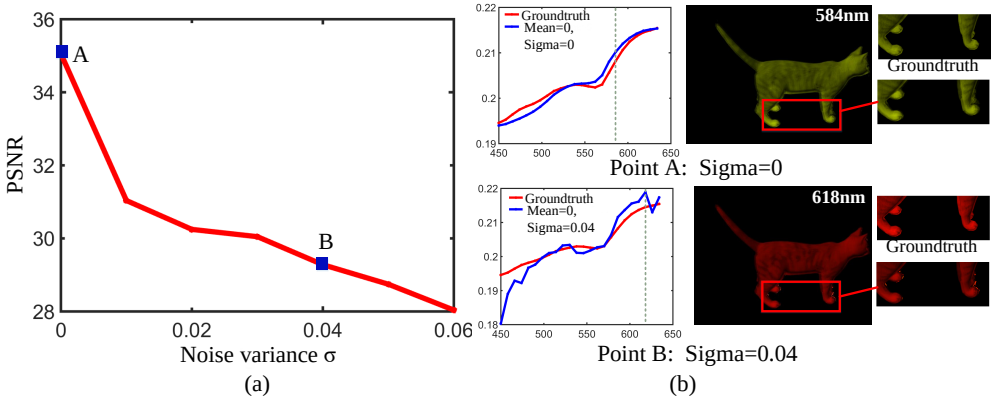

Fig. 8. (a) PSNR of simulated CAT image with different parameter σ of additional Gaussian noise. (b)illustrates both reconstructed and GroundTruth spectral reflectance curves of two selected points of (a), and pseudo-color images in chosen spectrum marked by dotted line (586nm for point A, 618nm for point B).

the wideband spectrally multiplexed images with $3 \times M$ channels. By selecting $N = 3 \times M$ spectral channels from the visible spectrum range, we obtain the spectral sensing matrix $\mathbf{C}$

Optics EXPRESS

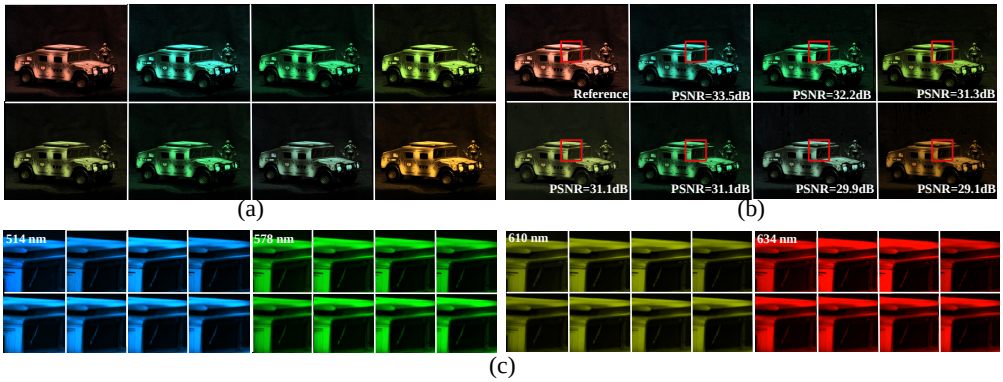

Fig. 9. Testing results of our proposed method on the light field dataset *ToyHumveeandSoldier* captured by Computer Graphics Laboratory in Stanford University. We choose a part of 2×4 image arrays from the whole 256 views on a 16×16 grid which have been calibrated already.(a) shows simulated light field color images with filters,(b) simulates image registration of the first(top left)view of the eight images. (c) illustrates four single-spectral reflectance images from all 24 reconstructed channels warped to all the views for 100×100 red patches in (b).

combining both camera responses and filter transmissions together:

$$\mathbf{C} = \begin{bmatrix} c_{1,1} & c_{1,2} & \cdots & c_{3\times M,N} \\ c_{2,1} & c_{2,2} & \cdots & c_{3\times M,N} \\ \vdots & \vdots & \ddots & \vdots \\ c_{3\times M,1} & c_{3\times M,2} & \cdots & c_{3\times M,N} \end{bmatrix}, \tag{2}$$

where $C_{3\times(m-1)+k,i}$ denotes the spectral sensitivities of i-th narrowband channel in the k-th channel of camera m. Specifically, each row of $\mathbf{C}$ is the combination of spectral response curves of both camera sensors and our wideband filters. For a scene point with spectrum $\mathbf{s} = [\mathbf{s_1}, \mathbf{s_2}, ..., \mathbf{s_N}]^\mathbf{T}$, we get the discrete version of Eq. (1),

$$p_{m,k} = \sum_{i=1}^{N} C_{3\times(m-1)+k,i} \cdot s. \tag{3}$$

Considering we have m cameras and each of them has 3 channels, the above equation system can be expressed in a matrix format as:

$$\mathbf{P} = \mathbf{Cs}, \tag{4}$$

where $\mathbf{P} = [\ p_1 \quad p_2 \quad \cdots \quad p_N\ ]$ is the heterogeneous wideband measurements of a single pixel (which can be derived by matching the correpondences of the captured heterogeneous images), and the narrowband spectrum $\mathbf{s} = [\ s_1 \quad s_2 \quad \cdots \quad s_N\ ]$ can be computed by solving the matrix with given $\mathbf{C}$.

Eq. (4) is the final formulation that forms the core of our spectral de-multiplexing reconstruction system, and can be solved by minimizing the following objective function in a least squares sense with respect to $\mathbf{s}$:

$$\hat{\mathbf{s}} = \arg\min_{\mathbf{s}} ||\mathbf{P} - \mathbf{Cs}||^2, \tag{5}$$

where $\hat{\mathbf{s}}$ denotes estimated spectrum from given measurements.

Optics EXPRESS

Since the illumination and surface spectra are generally continuous with seldom sharp edges in real world, and the surface spectra should be positive all the time, we introduce the smoothness and non-negative constraints into the objective function to make it an optimization problem:

$$\begin{aligned} \hat{\mathbf{s}} &= \arg\min_{\mathbf{s}} ||\mathbf{P} - \mathbf{C}\mathbf{s}||^2 + \lambda||\nabla\mathbf{s}||^2 \\ s.t. \quad & s(i) \geq 0 \qquad for\ all\ i, \end{aligned} \tag{6}$$

where ∇ is the differential operator, λ is the weight for smoothness constraint, and in our experiment λ is set to be 0.01 experimentally.

The projected Gradient based method is applied for minimizing this problem. Specifically, for each iterative step of normal gradient descent, the projecting manipulation is added to keep the searching inside the feasible region. Considering the speed and convergence properties of different optimization methods which as discussed in detail in Sec.4.1, we finally apply conjugate gradient method to solving this optimization problem.

3. Implementation details

3.1. System configuration

Our proposed prototype system consists of eight cameras in two rows. Each row has four cameras that are placed without any gap, so that the horizontal baseline of proposed system is exactly the width of the camera (30 mm). As for the vertical space, the eight camera are placed on two rungs with about 50mm interval, leading to 50mm vertical baseline.

The stereo matching problem assumes all the cameras are placed parallelly. However, in practice, it is impossible to meet this requirement. Thus, the system calibration and rectification are required to correct the system errors and make the corresponding points in different cameras in the same epiploar line. In this paper, we use the camera calibration toolbox [32] to calibrate the intrinsic and extrinsic parameters. After knowing the stereo camera projection matrices, the rectifying transformation can be calculated by solving relationships between original projection matrices and rectified projection matrices with the line through two camera centers as the baseline. Then,we can rectify the real captured images by applying the rectifying transformation to the camera array [33].

3.2. Responses calibration and filter selection

Response calibration To make the problem solvable, we need to calibrate the spectral sensitivities of the camera array with filters employed in our system, in other words, calibrate sensing matrix $\mathbf{C}$ in Eq. (6) since the accuracy of C greatly depends on the detection accuracy of the eight wide band filters' spectra. To estimate the spectral curves of filter arrays, we measured the spectral signals $s0$ of the Macbeth color chart using a hyperspectral camera (Prism-Mask Imaging Spectrometer [34]) and $s1$ of the same color chart with wideband plastic filters covered in front of the hyperspectral camera. The spectral sensitivities of each filter c_{filter} can be obtained through $s1/s0$.

The total error of sensing matrix $\mathbf{C}$ can be calculated as:

$$E_C = \frac{\frac{s1+\Delta s1}{s0+\Delta s0} - \frac{s1}{s0}}{\frac{s1}{s0}} = \frac{\frac{\Delta s1 s0 - \Delta s0 s1}{s0^2}}{\frac{s1}{s0}} = |\frac{\Delta s1}{s1}| + |\frac{\Delta s0}{s0}| \tag{7}$$

where $|\frac{\Delta s1}{s1}|$ and $|\frac{\Delta s0}{s0}|$ are both hyperspectral camera's relative measurement error, which can be presented by E as upper limit, and thus will be no larger than $2E$. That is to say, the total error E_C of sensing matrix C is in a controllable range. Thus, the accuracy of the matrix C only

Optics EXPRESS

depends on the detection accuracy of the hyperspectral camera, which can be further calibrated by using high-sensitivity hyperspectral camera sensor.

After test, we have chosen a group of eight qualified plastic filters with lowest condition number, whose spectral curves are plotted in Fig. 4. These eight spectral responses are well-conditioned and robust to small changes of the inputs in the function, and thus provide enough variance to solve our problem in Eq. (6). The standard spectral response of the Point Grey GS3-U3-51S5C-C camera sensor array (Fig. 4(c)) is provided in Point Grey website and can be downloaded directly, as shown in Fig. 4(d).

Filter selection The premise of our work is that correspondences in different cameras provide uncorrelated measurements of the spectra of a single point to enable the full reconstruction of the spectral curve. Thus, the accuracy of reconstruction depends on the correlation between the spectral sensitivities of different cameras. The best scenario would arise when the spectral responses of different filters are completely uncorrelated. The worst case would be the spectral sensitivities of different filters are almost identical. We analyze the spectral sensitivities of different filters to validate that they provide enough independent measurements of the incoming light spectrum. About 18 types of plastic filters are tested, and eight of them are selected by minimizing the condition number of resulted sensing matrix $\mathbf{C}$ as shown in Fig. 4(a). From Fig. 4(b), we can see the selected eight filters has different transmission curves and thus can sense the spectrum accurately.

3.3. Image registration

As for the two images at different rows and different columns, instead of computing the stereo matching directly, we introduce the intermediary image to facilitate system calibration and computation. Given an arbitrary image pair at different rows and columns, there exist two intermediary images which are at the same row of one input image and at the same column with the other. By introducing the intermediary images, the corresponding points in arbitrary image pairs can be matched by the aid of their common correspondences in intermediary images. By using the intermediary images, any image pairs can be easily aligned and warped without the need of rectification between images in different rows and columns, which is difficult in these cases since epipolar lines are neither horizontal or vertical.

By applying the stereo matching between all image pairs, we derive disparity maps between all image pairs, and thus can warp all the images to any view of eight cameras. In fact, to derive the whole multispectral light field, all the images are warped to all the views, so that eight 24-channel images can be derived. Note that if the users are only interested in a certain view, the rest images do not need to be warped, since the spectral multiplexing can be achieved independently on a single 24-channel image.

4. Experiments

4.1. Experiments on synthetic data

We first perform experiments on synthetic data to validate our algorithm as well as to analyze the effect of the number of the cameras. We have synthesized three groups of data using *Autodesk 3ds Max* software [35] with off-the-shelf 3D models and arranged the simulated cameras 2 inches apart and paralleled so that their fields of views overlap completely about 10 feet from the array.

(a) Data synthesis To derive the spectral images, the training based algorithm [36] is applied to generate the spectral images from RGB images directly. Then, the filter transmission curves and the camera response curves are used to estimate real capture images by using Eq. (3). As is shown in Fig. 5(a), three examples of multiview images simulated as wideband-filtered RGB images are presented.

Optics EXPRESS

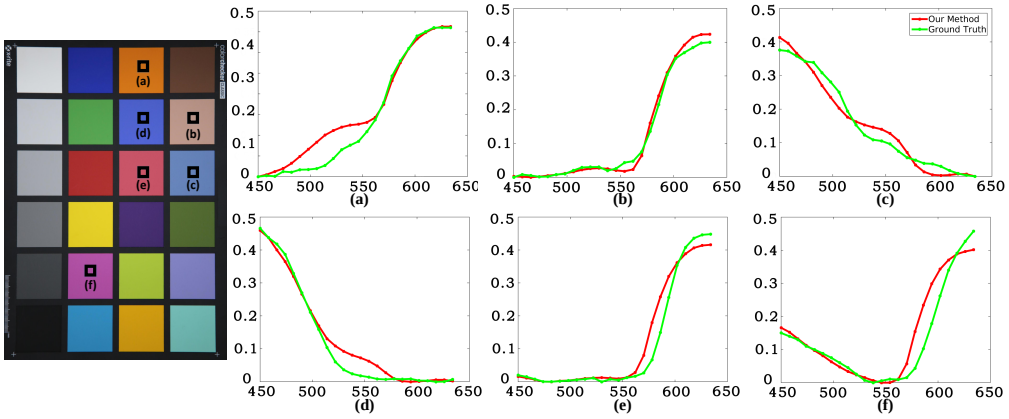

Fig. 10. Verification experiment using a Macbeth color chart. The results from our method and the Ground Truth curves of the color checker are compared. We randomly choose six patches from all 24 color patches of the colorboard and illustrate their both reconstructed and standard spectral reflectance curves of 24 channels from 450nm to 634nm, with an interval of 8nm.

Fig. 11. Real color images captured by our proposed 2×4 camera array system with heterogeneous wideband filters. These two scenes are both captured under indoor iodine-tungsten illumination and we can obtain the illumination spectra through capturing the standard white board. we also randomly select several points with different colors and illustrate their reconstructed 24-channel single-spectral reflectance curves in the rightmost column.

(b) Image registration We then do image registration for images in different viewpoints. The registration results of the first view (top left view) warped from all the other views are shown in Fig. 5(b), and the Peak-Signal-to-Noise-Ratio (PSNR) of warped images are given in the lower right corner of each images. We can see that the registration quality are much better for the nearer views than further ones. Since we show the registration results of the top left view, the warping images from the nearest views, i.e. Row 1, Column 2 and Row 2, Column 1, gives much better results than the images warped from the furthest views, i.e. bottom right. Besides, we note that the simulated scenes on the bottom are of complex periodic details and large view changes, and

Optics EXPRESS

thus have worse registration result than other examples.

(c) Optimization Multispectral reconstruction can be solved by minimizing the objective function in Eq. (6). We have compared three commonly used optimization methods, which are *gradient descent method*, *conjugate gradient method* and *least square QR factorization method* respectively on simulated CAT image, the comparison results are shown in Fig. 6, from which we can see the three optimization methods converge to the same solution, hence share the same accuracy. Furthermore, the running time is fastest for conjugate gradient method and slowest for gradient descent method, so we finally choose conjugate gradient method to solve this optimization problem considering the tradeoff between efficiency and complexity.

(d) Reflectance reconstruction We reconstruct 24-channel spectral reflectance for each view by using spectral de-multiplexing algorithm and implement Cubic Spline Interpolation function on standard trichromatic curves of camera sensor to get discrete distribution of RGB channels (the discrete value is 24). Then we respectively apply these three 24×1 distributions on each single spectral image to obtain trichromatic channels, and hence convert it into pseudo-color image. By reconstruct all the eight spectral views, the final multispectral light field is recovered. As shown in Fig. 7, six selected spectral channels for the first (top left) view of reconstructed results are presented. The quantitative evaluation of reconstruction errors for spectral reflectance both of each channel and average in Peak-Signal-to-Noise-Ratio (PSNR) and Structural-Similarity-Index (SSIM) are given in Table. 1.

To evaluate our algorithm's dependence on spectral sensing matrix C, we add Gaussian noise $N(\mu, \sigma^2)$ into filter array's spectra to simulate inaccurate matrix C. The PSNR of simulated CAT image with different parameter σ are evaluated, as is shown in Fig. 8(a), where the accuracy of reconstructed multispectral images turns down gradually as parameter σ of Gaussian noise component increases. In Fig. 8(b), we illustrate both reconstructed and Groundtruth spectral

Table 1. Evaluating multispectral reflectance reconstruction errors from three groups of 2×4 simulated light field datasets.

Cat	**PSNR (dB)**	**450nm**	**458nm**	**466nm**	**474nm**	**482nm**	**490nm**	**498nm**	**506nm**	**Avg**
		35.11	34.35	33.28	32.78	32.67	32.36	32.05	31.47	
		514nm	**522nm**	**530nm**	**538nm**	**546nm**	**554nm**	**562nm**	**570nm**	30.90
		30.67	30.06	29.67	29.91	30.10	30.56	30.91	30.88	
		578nm	**586nm**	**594nm**	**602nm**	**610nm**	**618nm**	**626nm**	**634nm**	
		29.94	29.17	28.71	28.58	28.97	29.59	29.92	30.00	
	SSIM	**450nm**	**458nm**	**466nm**	**474nm**	**482nm**	**490nm**	**498nm**	**506nm**	**Avg**
		0.9246	0.9236	0.9238	0.9231	0.9208	0.9195	0.9164	0.9153	
		514nm	**522nm**	**530nm**	**538nm**	**546nm**	**554nm**	**562nm**	**570nm**	0.9118
		0.9114	0.9064	0.9049	0.9049	0.9062	0.9090	0.9112	0.9095	
		578nm	**586nm**	**594nm**	**602nm**	**610nm**	**618nm**	**626nm**	**634nm**	
		0.9008	0.8993	0.9012	0.9038	0.9074	0.9107	0.9136	0.9163	
House	**PSNR (dB)**	**450nm**	**458nm**	**466nm**	**474nm**	**482nm**	**490nm**	**498nm**	**506nm**	**Avg**
		35.67	34.73	33.18	32.40	32.31	31.90	31.51	30.55	
		514nm	**522nm**	**530nm**	**538nm**	**546nm**	**554nm**	**562nm**	**570nm**	30.94
		29.35	28.78	28.09	28.72	29.04	29.87	30.63	31.12	
		578nm	**586nm**	**594nm**	**602nm**	**610nm**	**618nm**	**626nm**	**634nm**	
		30.81	30.37	30.04	29.93	30.36	31.02	31.29	31.10	
	SSIM	**450nm**	**458nm**	**466nm**	**474nm**	**482nm**	**490nm**	**498nm**	**506nm**	**Avg**
		0.8628	0.8679	0.8745	0.8784	0.8779	0.8735	0.8852	0.8887	
		514nm	**522nm**	**530nm**	**538nm**	**546nm**	**554nm**	**562nm**	**570nm**	0.8726
		0.8832	0.8877	0.8876	0.8874	0.8780	0.8832	0.8830	0.8713	
		578nm	**586nm**	**594nm**	**602nm**	**610nm**	**618nm**	**626nm**	**634nm**	
		0.8499	0.8459	0.8486	0.8600	0.8623	0.8646	0.8672	0.8737	
Park	**PSNR (dB)**	**450nm**	**458nm**	**466nm**	**474nm**	**482nm**	**490nm**	**498nm**	**506nm**	**Avg**
		29.55	29.22	28.21	27.79	27.73	27.50	27.17	26.52	
		514nm	**522nm**	**530nm**	**538nm**	**546nm**	**554nm**	**562nm**	**570nm**	26.72
		25.87	25.45	25.07	25.33	25.37	25.64	25.71	25.53	
		578nm	**586nm**	**594nm**	**602nm**	**610nm**	**618nm**	**626nm**	**634nm**	
		25.08	25.01	25.36	26.05	27.01	28.12	28.56	28.40	
	SSIM	**450nm**	**458nm**	**466nm**	**474nm**	**482nm**	**490nm**	**498nm**	**506nm**	**Avg**
		0.8385	0.8426	0.8429	0.8430	0.8444	0.8462	0.8442	0.8402	
		514nm	**522nm**	**530nm**	**538nm**	**546nm**	**554nm**	**562nm**	**570nm**	0.8223
		0.8314	0.8198	0.8125	0.8091	0.8050	0.8017	0.7935	0.7756	
		578nm	**586nm**	**594nm**	**602nm**	**610nm**	**618nm**	**626nm**	**634nm**	
		0.7534	0.7636	0.7921	0.8209	0.8402	0.8540	0.8601	0.8591	

Optics EXPRESS

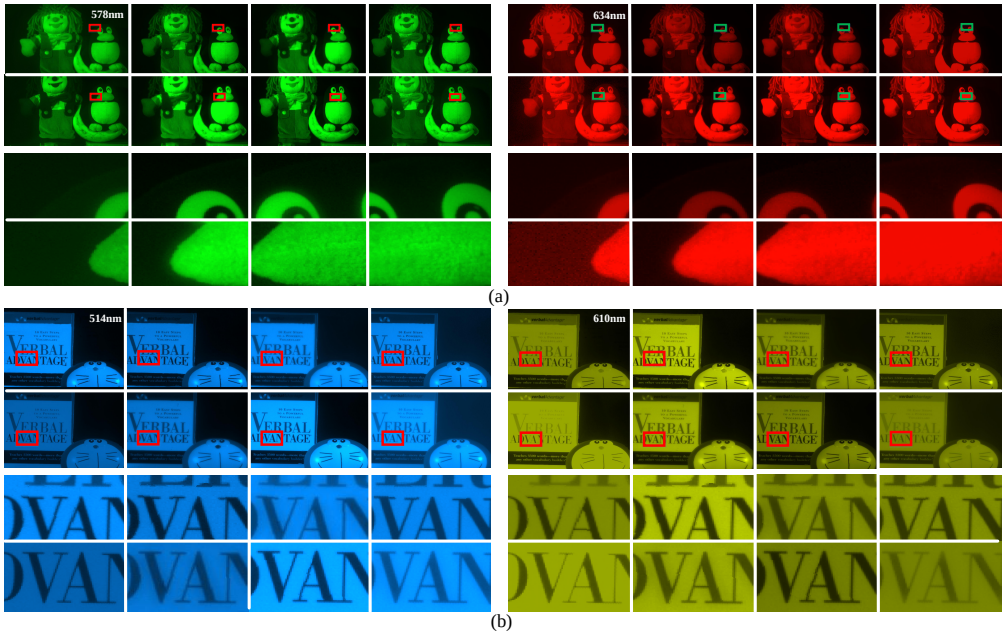

Fig. 12. Multispectral image reconstruction of various light field datasets captured by our own system under indoor iodine-tungsten illuminations and the detail results of the same patch in all eight views. We respectively select two single-spectral reflectance from all 24 reconstructed multispectral channels for each scenario, which are rendered as RGB images using the spectral sensitivities of the Point Grey GS3-U3-51S5C-C camera sensor.

reflectance curves of two selected points of Fig. 8(a), and pseudo-color images in chosen spectrum marked by dotted line (586nm for point A, 618nm for point B). We can see that although the overall reconstruction performance remains consistent when matrix C is inaccurate, some areas such as the cat feet are greatly influenced by noise, as shown in the enlarged rightmost image.

4.2. Experiments on light field datasets

We also test proposed method on the publicly available light field image datasets captured by Computer Graphics Laboratory in Stanford University using their multi-camera array [17]. As shown in Fig. 9(a), we use the training based spectral reconstruction algorithm [36] to generate multispectral light field from the existing light field data with RGB images. Fig. 9(b) shows the registration results of the first view (top left view) warped from all the other views and PSNR of warped images are given in the lower right corner of each image. Fig. 9(c) illustrates 100×100 reflectance patches in four selected spectral channels(514nm, 528nm, 610nm and 634nm) of respectively reconstructed results for all eight views. As can be seen, our method accurately reconstructs multispectral images in each viewpoint.

4.3. Experiments on real data

For the experiments on real data, we captured color images of several indoor scenes under iodine-tungsten illumination using our prototype camera array system introduced in Section 2.1, and the resolution of the examples in this paper is 1920×1080 that covers nonplanar objects.

We first compare multispectral reconstruction results of our method with ground truth curves of the standard Macbeth color checker to verify our algorithms and part of the comparison

Optics EXPRESS

results are shown in Fig. 10. As can be seen, the proposed method promisingly reconstructs 24 multispectral images of the classic color checker. It is worth noting that we should remove the illumination interference first which can be obtained through capturing a standard white board before recovering spectral reflectance of the scene. Fig. 11 illustrates various scenes under indoor iodine-tungsten illumination captured by our camera system. We select several typical points(such as red, blue and green points) from the images and illustrate their 24-channel single-spectral reflectance curves respectively in the rightmost column of Fig. 11. Meanwhile, we can also obtain the light field of the same scene simultaneously using these eight commercial digital cameras. Fig. 12 show several reconstructed single-spectral reflectance patches chosen from all 24 reconstructed channels for all the eight viewpoints, from which we can see images obtained by different cameras are registrated well except for some planer regions, where the disparity map may not be accurate enough. So far, we have successfully proved that we can obtain the light field and multispectral information simultaneously using our heterogeneous camera array system and algorithms.

5. Conclusion

We have introduced a framework for affordable and easy-to-use multispectral light field imaging using heterogeneous cameras array system. By exploiting anaglyph theory, multispectral images can be reconstructed through spectral demultiplexing. The proposed system can flexibly increase spectral channels by adding more cameras into the camera array. We have demonstrated the effectiveness and accuracy of our system using various synthesized and real examples.

The work has left out a few issues that deserve to be explored in depth. For example, accuracy of estimated multispectral images depends severely on the disparity mapping algorithms using convolutional neural networks in this paper and we hope to do further optimization in postprocessing of stereo matching in the next step to further improve the performance. Reconstruction time acceleration is also on the list of our future work.

Funding

National Natural Science Foundation of China (NSFC) (61422107, 61571215, 61627804, 61671236); National Science Foundation for Young Scholar of Jiangsu Province, China (BK20160634, BK20140610) .