

RESEARCH

Open Access



Snapshot compressive spectral - depth imaging based on light field

Xiaoming Ding^{1,2*} , QiangQiang Yan², Liang Hu³, Shubo Zhou⁴, Ruyi Wei^{2,5}, Xiaocheng Wang¹ and Yupeng Li¹

*Correspondence:

xmding@tjnu.edu.cn

¹ Tianjin Key Laboratory

of Wireless Mobile

Communications and Power

Transmission, Tianjin Normal

University, Tianjin 300387,

China

Full list of author information

is available at the end of the
article

Abstract

This paper proposes a snapshot Compressed Light Field Imaging Spectrometer based on compressed sensing and light field concept, which can acquire the two-dimensional spatial distribution, depth estimation and spectral intensity of input scenes simultaneously. The primary structure of the system contains fore optics, coded aperture, dispersion element and light field sensor. The detected data can record the coded mixture spatial-spectral information of the input scene with direction information of light rays. The datacube containing depth estimation can be recovered with the compressed sensing and digital refocus framework. We establish the mathematical model of the system and conduct simulations for verification. The reconstruction strategy is demonstrated for the simulation data.

Keywords: Multidimensional Imaging, Spectral imaging, Compressed sensing, Light field, Depth estimation

1 Introduction

As the development of 5G wireless technology, artificial intelligence (AI) and the increasing number of vehicles in big cities, Internet of Vehicles (IoV) emerges and develops rapidly. IoV has become a foundation of smart transportation and automatic drive [1, 2]. The object identification based on visual information of the road condition has an important role in automatic drive. The visual information is collected by imaging system. Usually, traditional imaging systems just detect the two-dimensional (2D) spatial data of targets, which can only provide limited information for object detection. Snapshot multidimensional imaging techniques develop rapidly in recent year, which can acquire plenty of optical information of targets as much as nine dimensions ($x, y, z, \theta, \phi, \psi, \chi, \lambda, t$), including three-dimensional (3D) spatial intensity distribution (x, y, z), propagation polar angles (θ, ϕ), propagations (ψ, χ), wavelength (λ) for spectral intensity and time (t) [3]. Multidimensional imager has a variety of applications in remote sensing, astrophysics, security, biochemistry and autonomous driving, etc. [4–7]. Especially, 3D spatial distribution and one-dimensional (1D) spectral intensity are significant in target detection, recognition, tracking and scene classification et al. in computer vision field. The depth information and spectral information of the objects can help to identify the overlap

boundary, estimate the distances among multiply objects and distinguish between the real living beings and figures. These advantages are significant for automatic drive [8, 9].

The 3D imaging integrated with spectral information produces a mass of data; however, the detection sensor is usually a two-dimensional (2D) format. So that, in order to collect and record the whole 3D spatial data, the input light field distribution should be modulated onto the sensor. There are two categories of methods to calculate the depth information of targets. One includes structured light and Time-of-Flight (ToF) techniques based on active imaging strategy, and the other includes binocular vision and light field techniques based on passive imaging strategy [10]. To measure the spectral characteristics of each spatial location in real time, the snapshot spectral imaging techniques include direct measurement strategy and computational strategy. The former includes the approaches of image-division, aperture-division and optical-path-division, and the latter includes the approaches based on computed tomography, compressed sensing (CS) and Fourier transform [11, 12].

To acquire 3D spatial and spectral information with high resolution in real time, much data need to be detected and recorded by the large format sensor. However, the increasing huge amounts of data would be a burden to the system of detection, storage and transmission in hardware. The 3D spatial and spectral information is sparse in some transform domains [13]. Hence, snapshot multidimensional imaging with CS and light field theory utilizes the correlation and redundancy of input scenes to realize data compression before sampling. This strategy can obtain 3D spatial and spectral information simultaneously in a single explosion time.

This paper proposes a snapshot Compressed Light Field Imaging Spectrometer (CLFIS) for multidimensional imaging, which can record the 2D spatial distribution and the input light direction of the targets directly. In addition, this method can collect the spectral information for each spatial location with coded aperture and dispersion element based on CS theory. The datacube (contains 2D spatial distribution and spectral intensity) with depth estimation of the targets can be recovered using digital refocus approach and CS algorithm. CLFIS has potential advantages of high light throughput, high spatial resolution and high spectral resolution.

2 Related works

Vast efforts have been reported in recent years focusing on the 3D spatial imaging with spectral measurement. According to the different approaches of 3D spatial measurement, the previous works mainly contain three types: ToF technique, binocular vision and light field.

The representation based on ToF is the Snapshot Compressive ToF + Spectral Imaging (SCTSI) proposed by Hoover Rueda-Chacon [14, 15]. This imager can send optical pulse, and the optical pulse will be reflected from targets. ToF sensor can detect the time difference between pulse sending and receipting, based on which can calculate the depth of objects. In addition, SCTSI can obtain the mixture 2D spatial-spectral information and depth information at the same time together with ToF and Coded Aperture Snapshot Spectral Imaging technique (CASSI). However, because of the need of active light, the system structure is complicated, and the range of depth and spectral measurement is limited.

The technique based on binocular vision mainly utilizes the data measured by two imaging paths to recover the 3D spatial information of targets by reconstruction algorithm. Kim et al. developed such a system firstly, which comprises a laser scanner and a customized CASSI. The system can obtain high-resolution 3D images with spectral intensity, called 3D imaging spectroscopy [16]. Wang et al. developed a cross-model stereo system that integrates the CASSI and an ordinary grayscale camera. The CASSI measures the spatial-spectral information of input scene, and the depth is reconstructed under the stereo configuration with grayscale data of targets [17]. Heist et al. utilized double snapshot spectral imagers in two paths to realize joint spectral and depth imaging in real time, together with a structured light imaging system to improve depth calculation [18]. Zhang et al. made use of two general spectral imaging paths to capture datacube and reconstruct the 3D spatial information of targets under each single spectral band through binocular vision reconstruction algorithm [19]. Yao et al. also utilize the binocular vision method based on two imaging paths [20]. One path is traditional RGB camera, the other is spectral imaging camera by optical filters. The RGB images and spectral images of targets contain different view angles, which can be used to reconstruct 3D spectral images. The binocular vision strategy needs two imaging paths, so that the optical structure is complicated. The shutter time and exposure time of these two paths should be controlled precisely, especially for the dynamic targets changing rapidly.

The technique based on light field utilizes the microlenses array to record the spatial and direction information of targets and reconstructs the 3D spatial images through digital refocus algorithm. In order to measure the spectral information, this strategy usually obtains the mixture spatial-spectral data through coded aperture approach based on CS theory. The 3D spatial datacube is recovered by digital refocus and CS reconstruction algorithm. The typical representations include Compressive Spectral Light Field Imager (CSLFI) proposed by Marquez et al. [21] and 3D compressive spectral integral imager (3D-CSII) proposed by Feng et al. [22]. CSLFI records the light directions through microlenses array and reconstructs the spectral information of targets using the mixture spatial-spectral data obtained by coded color filters. The light throughput is low for the using of coded filters, and the spectral bands is limited by the number of filters. Since the 3D-CSII utilizes a lens array as the objective lens, the quality of images would be limited for the aberration introduced by the lens array.

In conclusion, CLFIS proposed in this paper has advantages of no need of active illumination compared with ToF system and also benefits from simpler optical structure than binocular vision system. In addition, compared with the techniques based on filter coded aperture for spatial-spectral encoding, CLFIS has higher light throughput. So that, CLFIS has compact optical structure which can ensure high robustness, and has high light efficiency which can reach 50% in theory, and has a wider range of applications without the need of active illumination.

3 Methods

3.1 General principle

The schematic layout of CLFIS is shown in Fig. 1. The entire system includes two main parts which are fore optics and imaging spectrometer. The fore optics is telecentric in

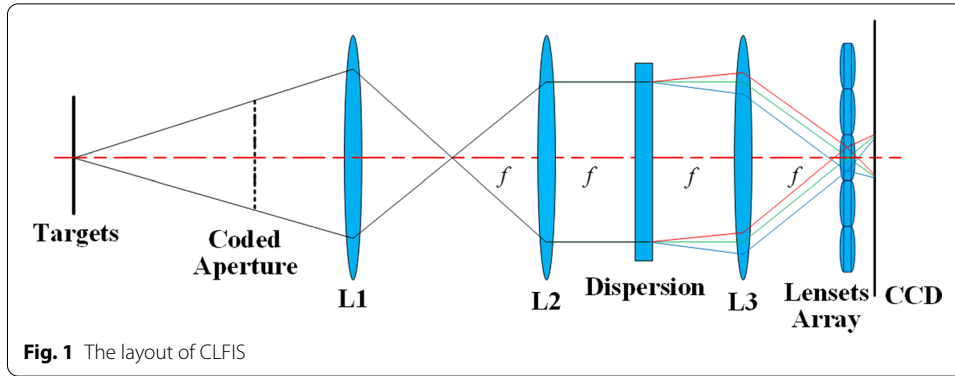
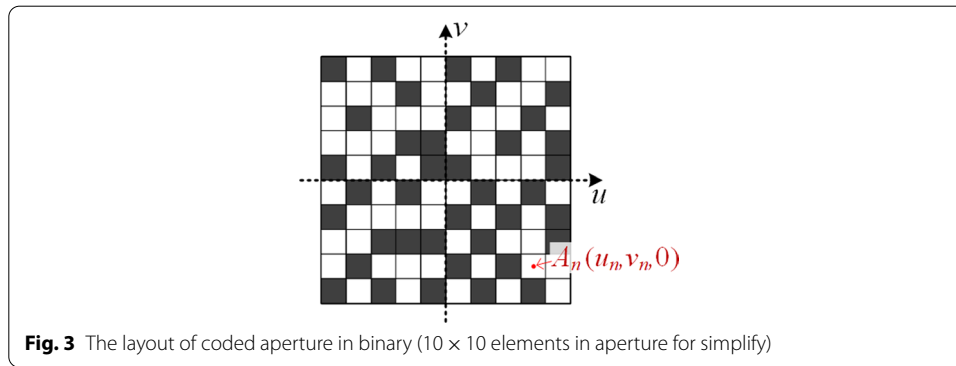
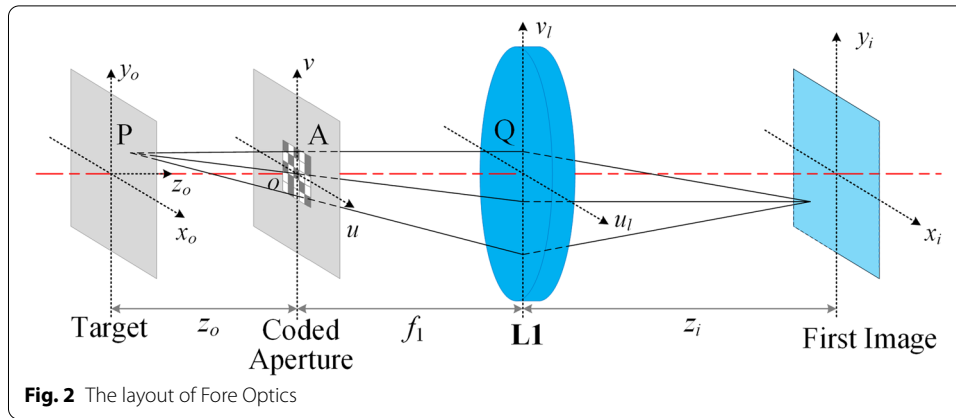


image space in order to make the chief ray from L1 is parallel to the optical axis. The telecentric structure ensures the chief ray from lenses array converge on the focus, so that the microlens center and the sub image center are at the same location in x and y directions. The aperture is placed at the focal plane of L1 and is encoded in binary randomly. L2 and L3 have the same focal length and form a $4f$ system. The light from first image formed by L1 is dispersed by an Amici prism on the focal plane of L2 and L3. Then, the light is collected by L3 and imaged on “light field sensor,” which is composed by lenses array and a format detector. The lenses array consists of hundreds of microlenses which have the same focal length and diameter. The distance between lenses array and the format detector is the focal length of microlens. The “light field sensor” can record the intensity and directions of the light simultaneously. As shown in Fig. 1, the object plane and lenses array are conjugate; the coded aperture and format detector are conjugate. From the above, CLFIS obtains the encoded light field data with spectral information. Based on the light field theory and digital refocus algorithm, we can recover the depth information of targets using the light field data. Since the coded aperture is imaged on the detector, so that the sub-image under each microlens is encoded in binary. The sub-image is sparsity in some transform domain, such as Fourier domain and wavelet domain. As results, the spatial and spectral information of targets can be recovered based on CS theory and CASSI reconstruction technique [23].

3.2 Optical model

We assume that the light intensity from arbitrarily target point is noted by $L(x_o, y_o, z_o, \lambda)$. The coordinate of object plane is represented by (x_o, y_o, z_o) . λ is the wavelength that is considered. The coordinates of coded aperture are noted by (u, v) , which is set as the origin of optical axis (z_o axis, and the positive direction is the light path direction). The coordinate of the first image plane, lenses array and detector is noted by (x_i, y_i, z_i) , (s, t) and (x_d, y_d) , respectively. The optical layout of the fore optics is shown in Fig. 2. The width of the aperture is D_1 , which is considered as a square. The first image point of the arbitrarily object point $P(x_o, y_o, z_o)$ is represented by $P'(x_i, y_i, z_i)$. According to the Gauss image formula, (x_i, y_i, z_i) is given as



$$\begin{cases} x_i = x_o f_1 / z_o \\ y_i = y_o f_1 / z_o \\ z_i = (z_o - f_1) f_1 / z_o \end{cases} \quad (1)$$

The light field data record the light direction information. Since the coded aperture conjugates with detector, the pixels under each sub-image are encoded by the aperture. The number of coded elements in the aperture is equal to that of pixels under each microlens. The position of each sub-pupil is noted by (u_m, v_n) ($m = 1, 2, \dots, M, n = 1, 2, \dots, N$; M, N are the numbers of encoded elements of the aperture in u direction and v direction, respectively, and m, n are the indexes). The aperture stop is encoded randomly in binary, which is illustrated in Fig. 3. Besides, we assume that the transmission function printed on the aperture is represented as $T(u, v)$,

$$T(u, v) = \sum_m^M \sum_n^N t_{m,n} \text{rect}\left(\frac{u}{d_a} - u_m, \frac{v}{d_a} - v_n\right) \quad (2)$$

where d_a is the side length of each element, and $t_{m,n}$ represents the coded status (1 for open and 0 for closed) at location (u_m, v_n) .

We assume that the light from $P(x_o, y_o, z_o)$ passing through sub-pupil $A(u_m, v_n, 0)$ has the intersection point with L1 at $Q(u_l, v_l, f_1)$. The vector \mathbf{PA} is parallel to \mathbf{AQ} according to the geometrical optical theory [24], then we can get

$$\frac{u_n - x_o}{u_l - u_n} = \frac{v_n - y_o}{v_l - v_n} = -\frac{z_o}{f_1} \quad (3)$$

The intersection point with L1 is given by

$$\begin{cases} u_l = \frac{f_1}{z_o} \cdot x_o + \left(1 - \frac{f_1}{z_o}\right) \cdot u_n \\ v_l = \frac{f_1}{z_o} \cdot y_o + \left(1 + \frac{f_1}{z_o}\right) \cdot v_n \end{cases} \quad (4)$$

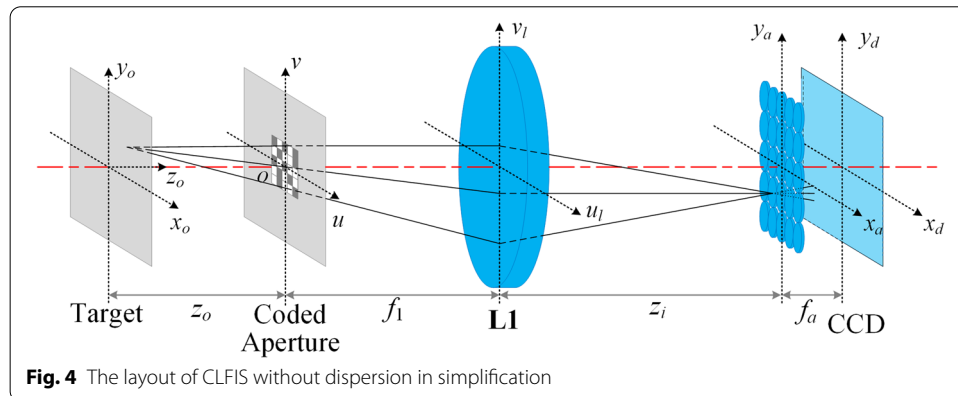
Since the L2 and L3 form a $4f$ system, the lenses array conjugates with the first image plane, CLFIS without dispersion element can be considered as the system shown in Fig. 4 for simplification. In this system, the light from $P(x_o, y_o, z_o)$ passing through $A(u_m, v_n, 0)$ intersects with lenses array at $P'(x'_a, y'_a, f_1 + l_i)$. Here, l_i is the distance between first image plane and L1. Then,

$$\begin{cases} x'_a = \frac{f_1 + l_i - z_i}{z_i - f_1} \cdot u_l + \frac{l_i}{z_i - f_1} \cdot x_i \\ y'_a = \frac{f_1 + l_i - z_i}{z_i - f_1} \cdot v_l - \frac{l_i}{z_i - f_1} \cdot y_i \end{cases} \quad (5)$$

As seen in Fig. 1, the Amici prism is direct vision, through which the light at center wavelength has no deflection. The dispersion coefficient is assumed to be $\alpha(\lambda)$. Then, the position of P' in CLFIS with dispersion is given by

$$\begin{cases} x_a = \frac{f_1 + l_i - z_i}{z_i - f_1} \cdot u_l - \frac{l_i}{z_i - f_1} \cdot x_i + \alpha(\lambda - \lambda_o) \\ y_a = \frac{f_1 + l_i - z_i}{z_i - f_1} \cdot v_l - \frac{l_i}{z_i - f_1} \cdot y_i \end{cases} \quad (6)$$

where λ_o is the center wavelength, and the “smile,” “keystone” and nonlinear dispersion are not considered. Finally, considering the light encoded by aperture stop, the intensity of the pixel on detector related to the sub-pupil (u_m, v_n) under the microlens (p_x, p_y) is expressed as



$$\begin{aligned}
 I(p_x, p_y, u_m, v_n) \\
 = \int_0^{X_o} \int_0^{Y_o} \int_0^{Z_o} L(x_o, y_o, z_o, u_m, v_n) \cdot T(u_m, v_n) \cdot dx_o dy_o dz_o
 \end{aligned} \quad (7)$$

where $L(*)$ is the light intensity from $P(x_o, y_o, z_o)$ passing through sub-pupil (u_m, v_n) . Considering the discrete pixels of sensor for sampling, the index of pixels of images (p_x, p_y) can be given by

$$\begin{cases} p_x = \frac{x_a + 0.5w}{d} + 1 \\ p_y = \frac{y_a + 0.5w}{d} + 1 \end{cases} \quad (8)$$

where w is the width of sensor format and d is the width of each pixel of the sensor.

3.3 Datacube reconstruction

The detector records the compressed spectral light field data in CLFIS. In order to obtain the spectral images with depth estimation, we should firstly reconstruct the light field data under different wavelengths based on CS algorithm and then recover each spectral images for depth estimation from light field data based on digital refocus technique. The sub-images of spectral light field images have smooth spatial structure, for these pixels, measure the light intensity from different directions of the same object point. As results, the sub-images are sparsity in some transform domain, like Fourier domain, wavelet domain, or some orthogonal dictionary domain et al. We assume that the original mixture data noted by g , and the transform expression in 2D wavelet domain is given by

$$g = W\theta \quad (9)$$

where g is the vector of 3D spatial information and spectral information, W is the transform matrix for 2D wavelets. θ is the transform coefficients of g in 2D wavelets domain. The imaging processing can be represented by

$$I = HW\theta \quad (10)$$

where H is the imaging matrix which can be derived by Eq. (7). If I , H and W are the known values, we can estimate θ under the sparsity assumption based on CS reconstruction algorithm. The estimation processing of θ is to solve the problem as followed, the estimation of θ is represented by θ' ,

$$\theta' = \arg \min \left[\|I - HW\theta'\|_2^2 + \tau \|\theta'\|_1 \right] \quad (11)$$

where the $\| \cdot \|_2$ term is l_2 norm of $(I - HW\theta')$ and the $\| \cdot \|_1$ term is l_1 norm of estimation value θ' . The first term minimizes the l_2 error, and the second term minimizes the number of elements in θ to ensure the sparsity of θ . Plenty of approaches are proposed to solve this optimization problem, such as Two-step Iterative Shrinkage/Thresholding (TwIST) algorithm [25], Gradient Projection for Sparse Reconstruction (GPSR) algorithm [26], Orthogonal Matching Pursuit (OMP) algorithm [27] and some learning methods based on deep network [28–30]. In this paper, we chose the traditional TwIST

for the optimal sparse solution. This algorithm is to use the regularization function to penalize the estimations of θ' that are undesirable to appear in the estimated θ' . τ is a tuning parameter to control the sparsity of θ' . The code of TwIST is available online (<http://www.lx.it.pt/~bioucas/code.htm>). After estimating θ , we can calculate the estimated g through Eq. (9). Since the estimated g is spectral light field data, we need digital refocus for g to obtain the depth information of targets. The digital refocused method used in this paper is proposed in [31]. The refocused processing is to remap the light field onto the selected imaging plane and integrate within the pupil plane. According to Fourier projection-slice theorem, the projection integration in spatial domain is equivalent to the slicing in Fourier domain. The derived Fourier slice imaging theorem shows that the imaging projection of the light field in the spatial domain is equivalent to the slicing of the Fourier transform of light field in frequency domain, and the slicing angle is perpendicular to the projection angle. If the image distance after digital refocus is assumed to be z_i' , which is represented as $z_i' = \gamma z_p$, the image after refocusing is given by

$$f'(x_d', y_d')_\lambda = \iint f_\lambda(u_m, v_n, x_d, y_d) du_m dv_n \quad (12)$$

where

$$\begin{cases} x_d = \frac{x_d'}{\gamma} + u_m \left(1 - \frac{1}{\gamma}\right) \\ y_d = \frac{y_d'}{\gamma} + v_n \left(1 - \frac{1}{\gamma}\right) \end{cases} \quad (13)$$

According to Eq. (13), we could choose γ value in order for acquiring the distinct spectral images under different depths. The depth estimation approaches using light field camera data usually include three main categories: sub-aperture image matching-based methods, Epipolar-Plane Image (EPI)-based methods and learning-based methods. In this paper, we use the first one integrating with defocus cue mentioned in Ng's work [31]. The spectral light field images are refocused at a series continues depth candidates. Among these candidates, the most distinct one for the selected area judged by an average gradient function is thought to be the actual depth.

4 Simulations

In order to evaluate the feasibility and effectiveness of CLFIS proposed in this paper, we conduct simulations for imaging and reconstruction. The target used for simulations is chosen from the iCVL portal for the database of hyperspectral images that are available online [32]. The size of target is 320×320 , and the bands rang is from 450 to 660 nm with the number of 8. The input datacube is shown in Fig. 5. The lenses array contains 64×64 microlenses, of which each lens covers 32×32 pixels on detector. If each pixel is with size of $10 \mu\text{m} \times 10 \mu\text{m}$, then the size of each microlens is $320 \mu\text{m} \times 320 \mu\text{m}$. The object distance is set 92 mm, 100 mm, 108 mm, respectively. The distance 100 mm is assumed to be the ideal object distance with which the image of targets forms on the lenses array. The aperture stop is the entrance pupil of the system, which is encoded in binary randomly with 32×32 elements. As described above, the sub-image under each microlens is encoded the same as the entrance pupil.

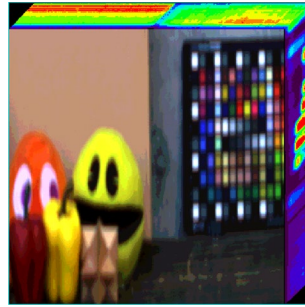


Fig. 5 The input datacube

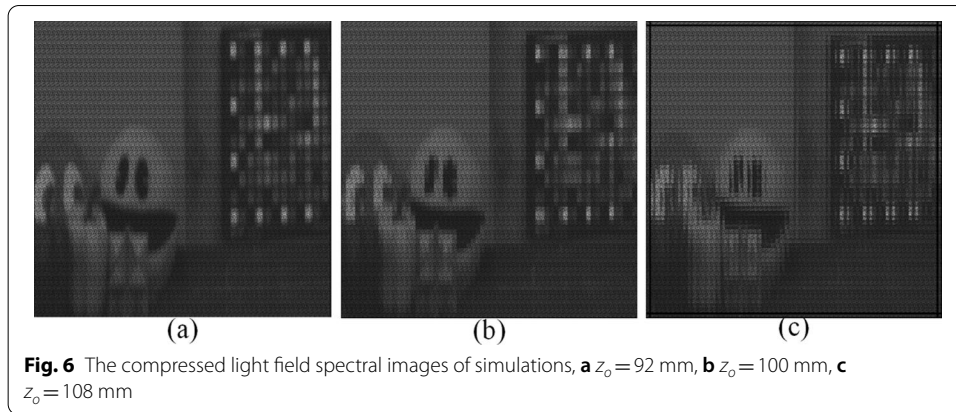
Table 1 The system parameters for simulations

System parameters	Values
Target size	20.48 mm × 20.48 mm
Band range and number	450 nm—660 nm, 8
Object length	92 mm, 100 mm, 108 mm
Size of entrance pupil	10 mm × 10 mm
Number of encoded elements	32 × 32
Number of microlenses	64 × 64
Pixel size on the sensor	10 μm × 10 μm
Focal length of L1, L2, L3	50 mm

We conduct two simulations. The first one is conducted using the entire datacube with the object distances of 92 mm, 100 mm and 108 mm, respectively, to test the feasibility of the imaging principle and reconstruction method. The second one is conducted with a datacube in which different parts are related to different object distances. This simulation is performed to evaluate the accuracy of depth estimation in CLFIS.

For the first simulation, the object distance is set 92 mm, 100 mm, and 108 mm, respectively, and the size of target is assumed to be 20.48 mm × 20.48 mm. To make the image cover the entire lenses array, the magnification of fore optics is 0.5. The entrance pupil is a square with size of 10 mm × 10 mm. The focal lengths of L1, L2, and L3 have the same value of 50 mm. Since the image of entrance pupil is of size 320 μm × 320 μm, so that the focal length of microlens is 0.8 mm according to the conjugate relationship. The main parameters of this system are summarized in Table 1. The dispersion coefficient $\alpha(\lambda)$ is considered to be the value that makes the input images from adjacent wavelengths shift a distance of the size of a single sensor pixel from each other.

The simulations are conducted by MATLAB based on Eq. (7) which should be discretized to adjust to the computer calculations. The input datacube is with the size of 320 × 320 × 8, and the number of encoded elements in aperture is 32 × 32, so that there should be about 800 million light rays to be calculated for tracing. The simulation results are shown in Fig. 6. The mixture data are with the size of 2048 × 2055.



5 Results and discussion

The reconstructed spectral images performed by CS algorithm and digital refocus method are given in Fig. 7. As seen in this figure, we reconstruct the spectral light field images from the compressed light field data by TwIST algorithm. The parameter τ is an empirical value, which is chosen as 0.05 by plenty trials. With the processing of reconstruction, the datacube with light field information is sparsely represented by a 2D wavelet coefficient. Though the TwIST and 2D wavelet transform may not be the best approach for reconstruction; however, this should be a widely used method. And the work in this paper we performed is to verify the feasibility of CLFIS. So that a universal and robust algorithm should be the best choice. In addition, some future work will be conducted to explore more accurate and efficient algorithm for better reconstruction results. After reconstructing the datacube with light field information, we obtain the light field data for each single spectrum from 450 to 660 nm. The digital refocus method is performed on each spectral light field data to get the quality spectral images at the object distance, which means that the refocused image distances should be about 77.174 mm, 75 mm and 73.148 mm, respectively, calculated based on the conjugate relation between object and image. According to the ideal image distances, the values of γ are set to be 1.029, 1 and 0.975, respectively. The digital refocus method used in this paper is conducted in spatial domain. The clearly focused images at any position can be recovered by this method within the range of depth of field.

To validate the ability of CLFIS to reconstruct the spatial and spectral feature of the objects in the scenes, we calculate PSNR of each spectral images and get the average value for each object distance comparing with the ideal input datacube. The PSNR values are 39.4 dB, 40.1 dB, 38.8 dB for object distance of 92 mm, 100 mm, and 102 mm, respectively. We choose five points in the targets at random and plot the spectral curves. The curves are normalized and illustrated in Fig. 8.

From the reconstructed results, especially for the object distances of 92 mm and 108 mm which are defocus when imaging, the reconstruction strategy can recover the quality of datacube. For each object distances, 30 points are chosen randomly to extract spectral curves and calculate the Relative Spectral Quadratic Error (RQE) [33] and Spectral Angle (SA) [34] to evaluate the curves, the average values are given in Table 2. The RQE is defined as

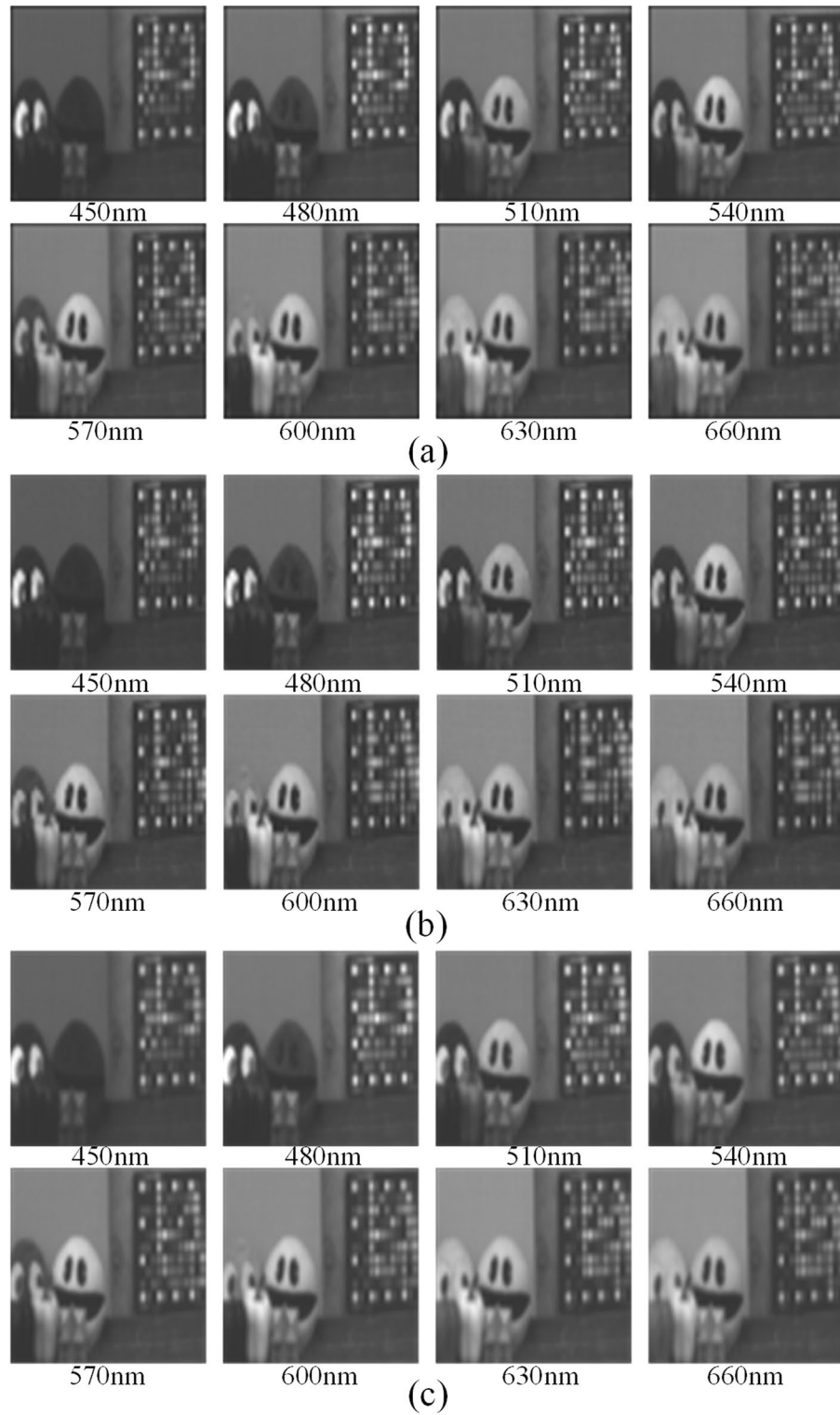
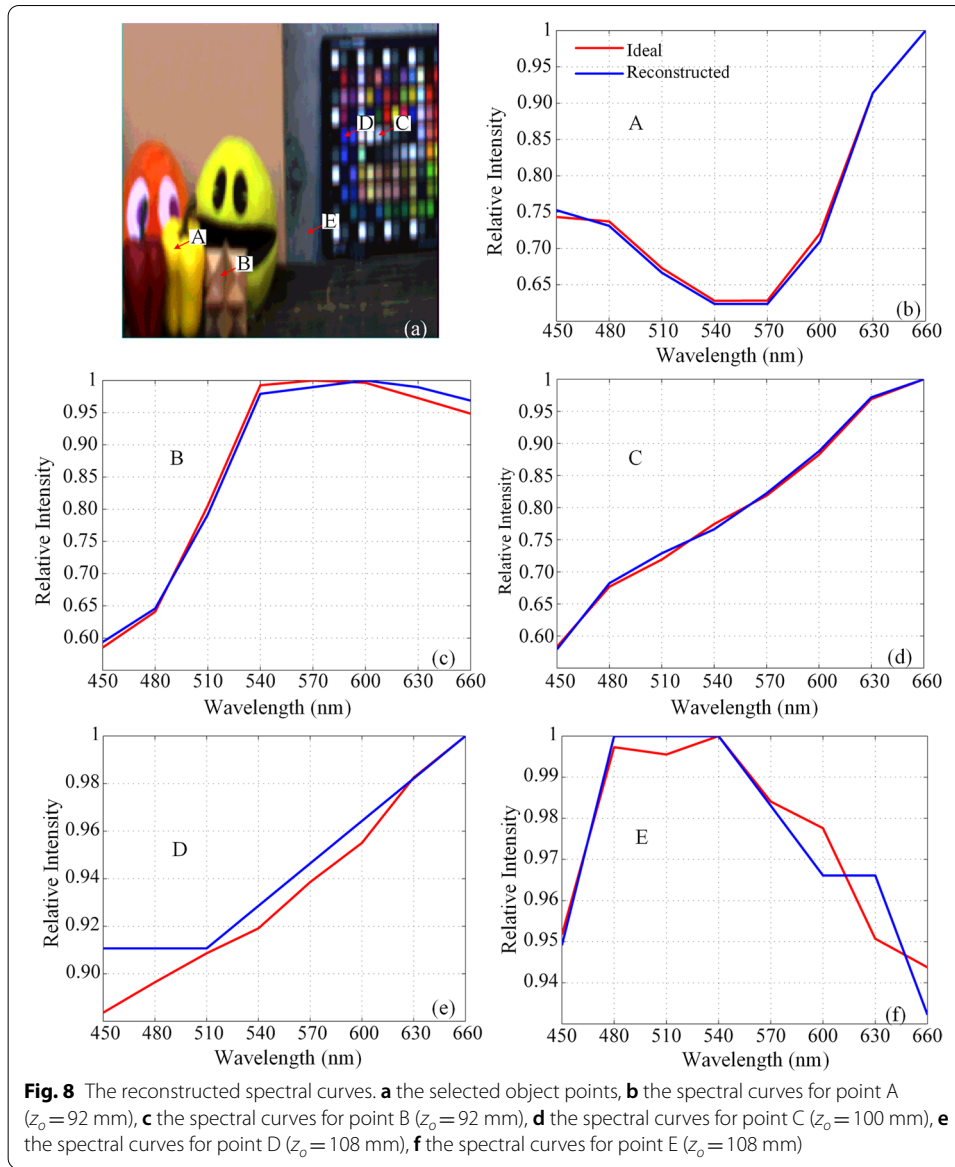


Fig. 7 The series of reconstructed spectral images, **a** $z_o = 92$ mm, **b** $z_o = 100$ mm, **c** $z_o = 108$ mm

**Table 2** The evaluation of spectral curves

Object distance (mm)	RQE	SA (rad)
92	0.252	0.063
100	0.116	0.054
102	0.257	0.082

$$RQE(\lambda, \lambda') = \sqrt{\sum_{i=1}^{N_\lambda} (\lambda_i - \lambda'_i)^2} / \sum_{i=1}^{N_\lambda} \lambda_i \quad (14)$$

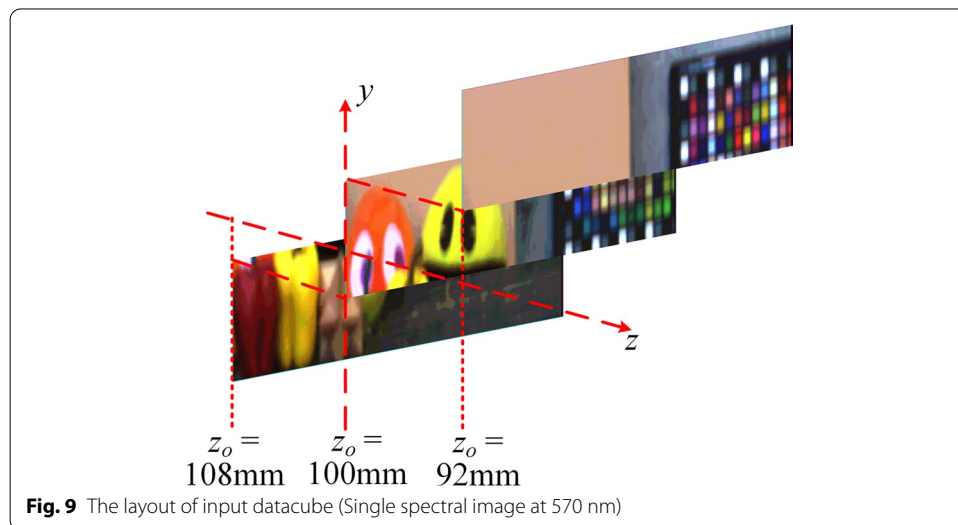
where λ is the standard spectral curve, and λ' is the measured spectral curve. N_λ is the number of sampling and i is the index. Noted by this equation, when the value of RQE is

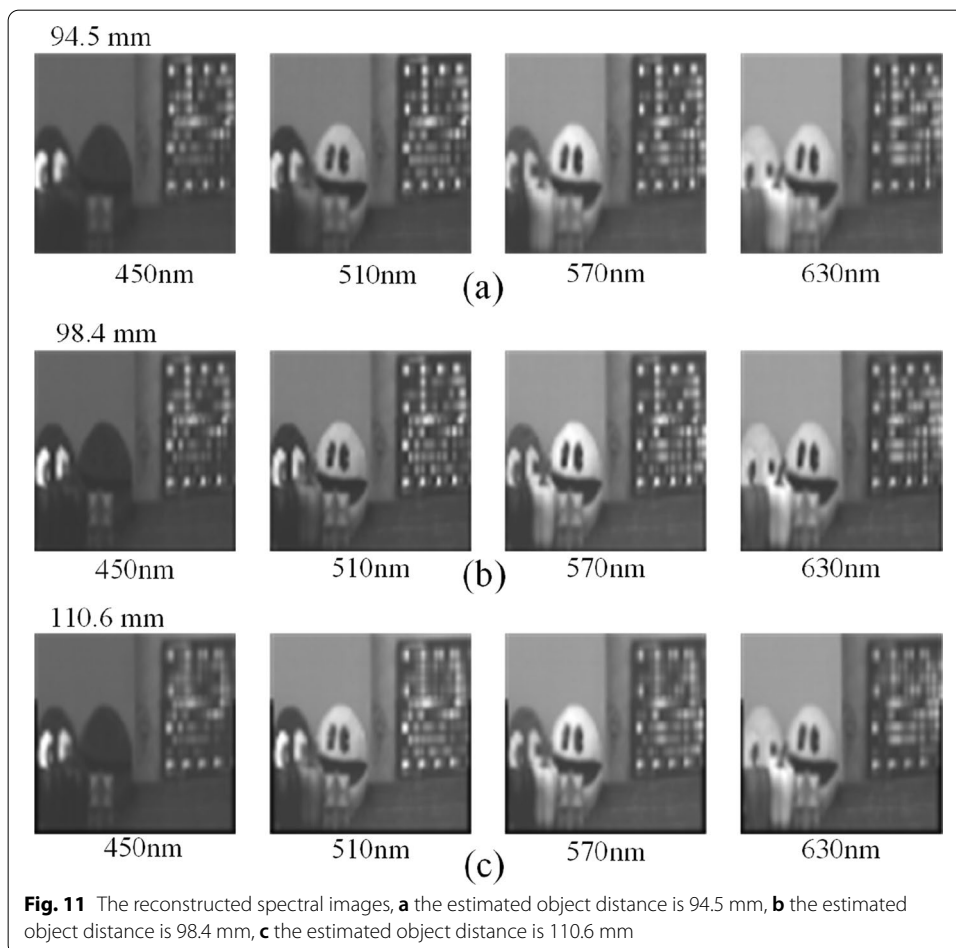
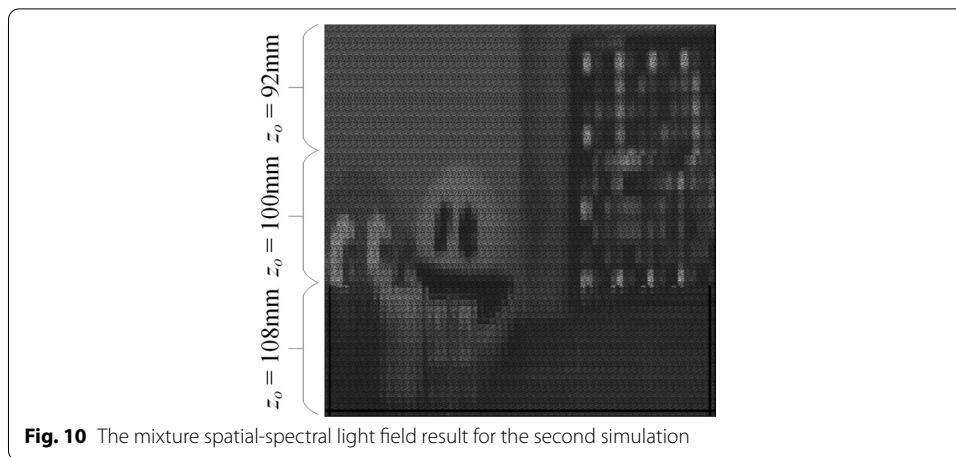
smaller, the two spectral curves are more similar. SA is another evaluation function for spectral data, in which the two spectral curves are regarded as vectors in a 2D space. The similarity of two curves is evaluated by calculation the generalized angle between these two vectors. The angle is represented by the arccosine that is defined as

$$SA(\lambda, \lambda') = \cos^{-1} \left[\frac{\sum_{i=1}^{N_\lambda} \lambda_i \lambda'_i}{\sqrt{\sum_{i=1}^{N_\lambda} \lambda_i^2} \cdot \sqrt{\sum_{i=1}^{N_\lambda} \lambda'^2_i}} \right] \quad (15)$$

The first simulation is conducted to identify the feasibility of the system proposed in this paper. The imaging simulation synthesizes the aliasing raw data, and the reconstruction method could effectively recover the 2D spatial distribution, 1D spectral intensity of input sense from different field depths based on the aliasing data. To verify the estimation of depth, we perform the second simulation. The input datacube of target is assumed to be tridimensional. Different parts of input field are related to different object distances. The target datacube inputted in this simulation is illustrated in Fig. 9. The whole object is divided to three parts whose object distances are set as 92 mm, 100 mm and 108 mm, respectively. The system parameters have the same value from above. The simulation result is given in Fig. 10.

We use Average Gradient to evaluate the images definition, when perform reconstruction and digital refocus. When the object distance of digital refocus varies from 85 to 115 mm, we obtain the most distinct images for three parts of the target at 94.5 mm, 98.4 mm and 110.6 mm, respectively. The images at wavelengths of 450 nm, 510 nm, 570 nm, 630 nm are shown in Fig. 11. Choose the spectral image at 660 nm as R band, spectral image at 540 nm as G band, and spectral image at 450 nm as B band to synthesize the true color images as shown in Fig. 12. According to the depth estimation results, we form a synthetic 3D image for the input sense in Fig. 13. From the reconstruction results, the CLFIS proposed in this paper can estimate the depth information of the targets. The error is about between 1.5% and 2.7% in simulations. The estimation error is correspondence with the object distance, angular resolution and the quantization error of the light field sensor. Larger object distance usually means





larger estimation error. The angular resolution is determined by the pixels covered by a single microlens and the pupil aperture diameter. Higher angular resolution helps to make more precise depth estimation. For the depth estimation in practical applications, we need to scan the image field to calculate the depth successively to acquire

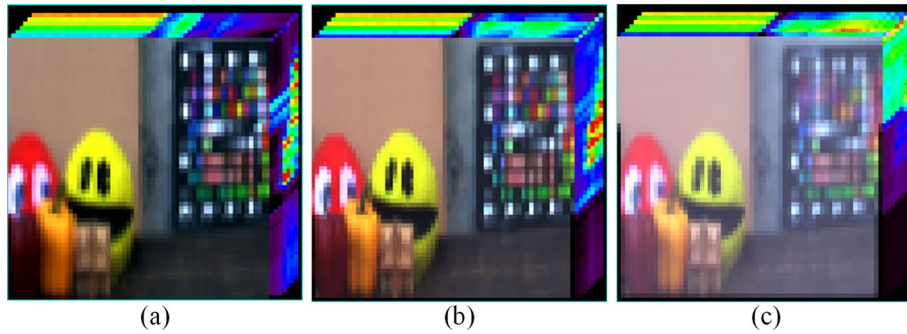


Fig. 12 The reconstructed datacube for depth estimation simulation

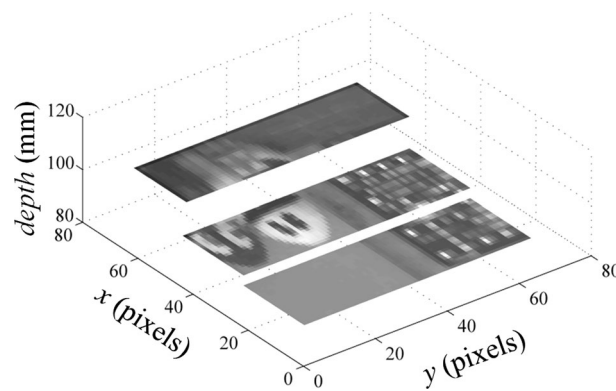


Fig. 13 Reconstructed synthetic 3D image at 570 nm

the entire 3D distribution of the targets. In the future, the digital refocus methods will be studied to calculate the 3D spatial information of targets directly for CLFIS.

6 Conclusion

In this paper, we propose the CLFIS technique, which utilizes the coded aperture, dispersion prism and microlenses array to obtain the compressed spectral light field data of targets. The 2D spatial distribution, depth estimation and spectral intensity can be recovered from the mixture data measured by sensor through CS reconstruction algorithms and digital refocus methods. This is a type of snapshot multidimensional imaging technique. This paper establishes the mathematical model according to light ray tracing theory. Imaging simulations are conducted to verify the model and reconstruction strategy. From the simulated compressed mixture data, we can reconstruct and estimate the spectral images at different depths. The images and spectral curves are evaluated by evaluation functions such as PSNR, RQE and SA, whose results reveal that the reconstructed datacube conforms the quality of ideal datacube. The refocus object distances estimated are mostly similar to the real distance in simulation, and the estimation error is from 1.5 to 2.7%. Finally, CLFIS has the potential advantages of high light throughput and high robustness of optical structure. In future works, the research will focus on the optimization of the coded aperture and the reconstruction processing.

Abbreviations

AI: Artificial intelligence; IoV: Internet of Vehicles; ToF: Time-of-Flight; CS: Compressed sensing; CLFIS: Compressed Light Field Imaging Spectrometer; SCTSI: Snapshot Compressive ToF + Spectral Imaging; CASSI: Coded Aperture Snapshot Spectral Imaging technique; CSLFI: Compressive Spectral Light Field Imager; 3D-CSII: 3D compressive spectral integral imager; TwIST: Two-step Iterative Shrinkage/Thresholding; GPSR: Gradient Projection for Sparse Reconstruction; OMP: Orthogonal Matching Pursuit; PSNR: Peak Signal-to-Noise Ratio; RQE: Relative Spectral Quadratic Error; SA: Spectral Angle.

Acknowledgements

Not applicable.

Authors' contributions

XD proposed the framework of the whole ideal, structure of the system and the model; QY, LH and RW helped to perform the simulations. XW, YL helped to conduct the analysis of the results. SZ participated in the conception and design of this research and helped to revise the manuscript. All authors read and approved the final manuscript.

Funding

This work was supported by the Scientific Research Project of Tianjin Educational Committee (2021KJ182), National Natural Science Foundation of China (NSFC) (62001328, 61803372, 62001327, 61901301), Doctoral Foundation of Tianjin Normal University (52XB2004, 52XB2005), Open Research Fund of CAS Key Laboratory of Spectral Imaging Technology (No. LSIT202004W) and Natural Science Foundation of Shaanxi Province (2021JQ-323).

Availability of data and materials

Please contact the authors for data requests.

Declarations

Ethics approval and consent to participate

Not applicable.

Consent for publication

Not applicable.

Competing interests

The authors declare that they have no competing interests.

Author details

¹Tianjin Key Laboratory of Wireless Mobile Communications and Power Transmission, Tianjin Normal University, Tianjin 300387, China. ²CAS Key Laboratory of Spectral Imaging Technology, Xi'an Institute of Optics and Precision Mechanics, Xi'an 710119, China. ³CAS Key Laboratory of Computational Optics Imaging Technology, Aerospace – Information Research Institution, Beijing 100094, China. ⁴Institute of Information Science and Technology, Donghua University, Shanghai 201620, China. ⁵Electronic Information School, Wuhan University, Wuhan 430072, China.

Received: 11 November 2021 Accepted: 4 January 2022

Published online: 31 January 2022

References

1. X. Liu, X.B. Zhai, W. Lu, C. Wu, QoS-guarantee resource allocation for multibeam satellite industrial internet of things with NOMA. *IEEE Trans. Industr. Inf.* **17**(3), 2052–2061 (2021)
2. X. Liu, X. Zhang, Rate and energy efficiency improvements for 5G-Based IoT with simultaneous transfer. *IEEE Internet Things J.* **6**(4), 5971–5980 (2019)
3. W. Feng, Research on the technology of snapshot compressive spectral integral imaging, Ph.D. dissertation, Nanjing University of Science & Technology, Nanjing, Jiangsu Province (2018)
4. A.F.H. Goetz, G. Vane, J.E. Solomon, B.N. Rock, Imaging spectrometry for earth remote sensing. *Science* **228**(4704), 1147–1153 (1985)
5. G. Cheng, P. Zhou, and J. Han, Learning rotation-invariant convolutional neural networks for object detection in VHR optical remote sensing images, *IEEE Trans. Geosci. Remote Sens.*, vol. 54, no. 12, pp. 7405–7415, Dec. (2016)
6. J.M. Ramirez, H. Arguello, Spectral image classification from multi-sensor compressive measurements. *IEEE Trans. Geosci. Remote Sens.* **58**(1), 626–636 (2020)
7. R.R. Iyer, M. Žurauskas, Q. Cui, L. Gao, S. Theodore, S.A. Boppart, Full-field spectral-domain optical interferometry for snapshot three-dimensional microscopy. *Biomed. Opt. Express* **11**(10), 5903 (2020)
8. X. Liu, X. Zhang, NOMA-based resource allocation for cluster-based cognitive industrial internet of things. *IEEE Trans. Industr. Inf.* **16**(8), 5379–5388 (2020)
9. F. Li, K. Lam, X. Liu, J. Wang, K. Zhao, L. Wang, Joint pricing and power allocation for multibeam satellite systems with dynamic game model. *IEEE Trans. Veh. Technol.* **67**(3), 2398–2408 (2018)
10. B. Javidi et al., Multidimensional optical sensing and imaging system (MOSIS): from macroscales to microscales. *Proc. IEEE* **105**(5), 850–875 (2017)
11. N. Hagen and M. W. Kudenov, Review of snapshot spectral imaging technologies, *Opt. Eng.*, vol. 52, no. 9, pp. 090901. (2013)
12. L. Gao, L.V. Wang, A review of snapshot multidimensional optical imaging: measuring photon tags in parallel. *Phys. Rep.* **616**, 1–37 (2016)

13. C. Tao et al. Hyperspectral image recovery based on fusion of coded aperture snapshot spectral imaging and RGB images by guided filtering, *Opt. Commun.*, vol. 458, pp. 124804. (2020)
14. H. Rueda et al., Single Aperture Spectral+ToF compressive camera: toward Hyperspectral+Depth Imagery. *IEEE J-STSP*. **11**(7), 992–1003 (2017)
15. H. Rueda-Chacon, J. F. Florez, D. L. Lau, and G. R. Arce, Snapshot compressive ToF+Spectral imaging via optimized color-coded apertures, *IEEE Trans. Pattern. Anal.*, vol. 1, no. 1. (2019)
16. M.H. Kim et al., 3D imaging spectroscopy for measuring hyperspectral patterns on solid objects. *ACM Trans. Graph.* **31**(4), 1–11 (2012)
17. L. Wang et al., Simultaneous depth and spectral imaging with a cross-modal stereo system. *IEEE Trans. Circuits. Syst. Video Technol.* **28**(3), 812–817 (2018)
18. S. Heist et al., 5d hyperspectral imaging: fast and accurate measurement of surface shape and spectral characteristics using structured light. *Opt. Express*. **26**(18), 23366–23379 (2018)
19. C. Zhang, M. Rosenberger, and G. Notni, 3D multispectral imaging system for contamination detection, In *Proc. SPIE 11056, Optical Measurement Systems for Industrial Inspection XI*, pp. 1105618. (2019)
20. M. Yao et al., Spectral-depth imaging with deep learning based reconstruction. *Opt. Express* **27**(26), 38312 (2019)
21. M. Marquez, H. Rueda-Chacon, H. Arguello, Compressive spectral light field image reconstruction via online tensor representation. *IEEE Trans. Image. Process.* **29**, 3558–3568 (2020)
22. W. Feng et al., 3D compressive spectral integral imaging. *Opt. Express* **24**(22), 24859 (2016)
23. X. Miao et al., λ -Net: Reconstruct Hyperspectral Images From a Snapshot Measurement, in *IEEE/CVF International Conference on Computer Vision (ICCV)*, pp. 4058–4068. (2019)
24. *Handbook of Optics*, 3rd ed. Volume I: Geometrical and Physical Optics, Polarized Light, Components and Instruments (set), McGraw-Hill, New York, USA, pp. 20.1–20.5. (2001)
25. M. Bioucas-Dias, M.A.T. Figueiredo, A new TwIST: two-step iterative shrinkage/thresholding algorithms for image restoration. *IEEE Trans. Image. Process.* **16**(12), 2992–3004 (2007)
26. M.A.T. Figueiredo, R.D. Nowak, S.J. Wright, gradient projection for sparse reconstruction: application to compressed sensing and other inverse problems. *IEEE J-STSP* **1**(4), 586–597 (2008)
27. Y. C. Pati, R. Rezaifar, P. S. Krishnaprasad, Orthogonal matching pursuit: Recursive function approximation with applications to wavelet decomposition, in *Proceedings of 27th Asilomar conference on signals, systems and computers*, IEEE, vol. 1, pp. 40–44. (1993)
28. L. Wang, T. Zhang, Y. Fu et al., HyperReconNet: joint coded aperture optimization and image reconstruction for compressive hyperspectral imaging. *IEEE Trans. Image. Process.* **28**(5), 2257–2270 (2019)
29. L. Wang, C. Sun, Y. Fu, et al., Hyperspectral Image Reconstruction Using a Deep Spatial-Spectral Prior, in *2019 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, pp. 8032–8041. (2019)
30. X. Miao, X. Yuan, Y. Pu, et al., λ -Net: Reconstruct Hyperspectral Images From a Snapshot Measurement, in *2019 IEEE/CVF International Conference on Computer Vision (ICCV)*, pp. 4058–4068. (2019)
31. R. Ng, M. Levoy, M. Brédif, et al., Light field photography with a hand-held plenoptic camera, Stanford Univ. Comput. Sci., Stanford, CA, USA, Rep. 2005-02. (2005)
32. B. Arad, O. Ben-Shahar, Sparse Recovery of Hyperspectral Signal from Natural RGB Images, in *2016 European Conference on Computer Vision (ECCV)*, vol. 9911. (2016)
33. C. Mailhes, P. Vermande, F. Castanie, Spectral image compression. *J. Optics*. **12**(3), 121–132 (1990)
34. F.A. Kruse et al., The spectral image processing system (SIPS)-interactive visualization and analysis of imaging spectrometer data. *Rem. Sens. Env.* **44**(2–3), 145–163 (1993)

Publisher's Note

Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

Submit your manuscript to a SpringerOpen[®] journal and benefit from:

- Convenient online submission
- Rigorous peer review
- Open access: articles freely available online
- High visibility within the field
- Retaining the copyright to your article

Submit your next manuscript at ► [springeropen.com](https://www.springeropen.com)