



# Multispectral video acquisition using spectral sweep camera

XUEMEI HU,<sup>1,3</sup> XING LIN,<sup>2,3</sup> TAO YUE,<sup>1</sup> AND QIONGHAI DAI<sup>2,\*</sup>

<sup>1</sup>*School of Electronic Science and Engineering, Nanjing University, Jiangsu, 210023, China*

<sup>2</sup>*Department of Automation, Tsinghua University, Beijing, 100084, China*

<sup>3</sup>*Authors contributed equally*

\**ghdai@tsinghua.edu.cn*

**Abstract:** This paper proposes a novel multispectral video acquisition method for dynamic scenes by using the Spectral-Sweep camera. To fully utilize the redundancies of multispectral videos in the spatial, temporal and spectral dimensions, we propose a Complex Optical Flow (COF) method that could extract the spatial and spectral signal variations between adjacent spectral-sweep frames. A complex  $L_1$ -norm constrained optimization algorithm is proposed to compute the COF maps, with which we recover the entire multispectral video by temporally propagating the captured spectral-sweep frames under the guidance of reconstructed COF maps. We demonstrate the promising accuracy of reconstructing full spatial and temporal sensor resolution multispectral videos with our method both quantitatively and qualitatively. Compared with state-of-the-art multispectral imagers, our computational multispectral imaging system can significantly reduce the hardware complexities, while achieves comparable or even better performance.

© 2019 Optical Society of America under the terms of the [OSA Open Access Publishing Agreement](#)

## 1. Introduction

Multispectral images/videos are of great significance for many applications, such as surveillance [1], medicine [2], material analysis [3], etc. Although multispectral imaging has been actively investigated for decades, capturing multispectral videos of high spatial resolution remains a challenging task. In terms of implementation, there are mainly 4 categories of methods, i.e., dispersion based methods [4–18], diffractive optical element (DOE) based methods [19–21], multi-reflection based methods [22,23] and customized-filter based methods [24–36].

3D-volume dispersion [4], pixel-wise dispersion [5,6] and single dispersion [7,8,37] are successively proposed to realize multispectral imaging in a snapshot way. Among these dispersion based methods, the spatial resolution is low due to the trade-off for the spectral resolution. To retrieve higher spatial information, coded aperture snapshot spectral imaging (CASSI) methods [9,10] are proposed. Augmented with statistically learned spatial-spectral priors [11–14], color-coding strategies [13,15], dual-coding design [16] or hybrid-system implementations [17,18], CASSI are promising for higher spatial and spectral resolution. However, CASSI-based multispectral imaging systems is not compact and requires sophisticated optical/geometry calibrations.

Recently, DOE based methods [19–21] are proposed to encode the multispectral information with spectrally-varying PSFs and realize snapshot multispectral imaging [19–21]. Among these methods, scene images of different wavelength are encoded with varying PSFs and decoded with optimization based on the calibrated PSFs of different wavelength. These kind of novel multispectral methods could enable compact, snapshot multispectral imaging and video-taking, while the reconstruction quality of these methods in both spatial and spectral domain are limited by the diffraction efficiency of the diffractive optical elements.

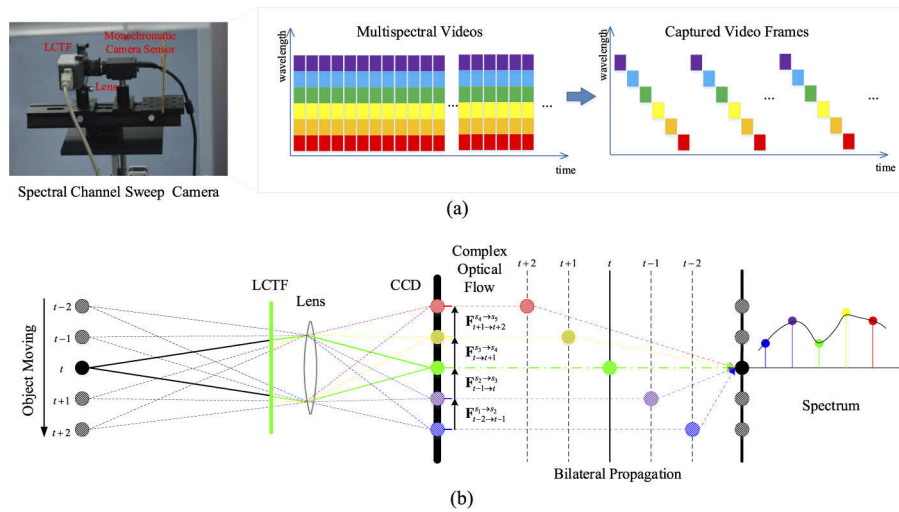
Multi-reflectance based methods [22,23] emerged to realize snapshot multispectral imaging. Through multi-reflectance optical system, the images of object is duplicated and located in

different spatial positions. Introducing multispectral encoding during the multi-reflectance process [23] or in the duplicated images [22] enables multispectral imaging. However, the field-of-view is directly sacrificed to obtain multiple reflected images in a single image and the spatial resolution is also limited.

Among all the above three main methods, the spatial information is either mixed with the spectral information or directly sacrificed for the spectral information. Multispectral imaging could also be realized through directly coding in the spectral domain with different color filters. RGB images/videos are exploited to recover spectral information [24–28]. While these method could turns the commercial RGB camera into a multispectral camera, high spectral frequency component is hard to be recovered with quite limited number (3) of spectral sampling. To reduce the loss of high frequency information, customized multispectral filtering methods are proposed. Spectral filters optimized either in the primal domain [29–31] or in the frequency domain [32,33] could help to realize efficient encoding of spectral information. However, these methods require customized spectral filter fabrication which are not easy to access. Liquid crystal tunable filter (LCTF) are widely utilized in hyperspectral imaging [34–36] due to their portability, fast tunability, convenient controllability, etc. Conventionally, the scene is spectrally scanned by tuning the center transmission wavelength of LCTF and the time resolution is limited for LCTF based multispectral imaging method. Mian [38] et al. take the first step to explore multispectral video restoration with undersampled spectral video information taken by LCTF (5 spectral wavelength images at each time, 10nm differences among different time). Sparse priors in the space, time and spectral domain are well exploited through optical flow and sparse coding, and promising restoration results are demonstrated. However, taking 5 spectral information in each instant is not physically implementable. Beyond that, combining CASSI with LCTF based spectral imaging method are introduced to enhance the spatial and spectral resolution [36], while the time resolution is still limited.

In this paper, we propose a compact LCTF-based method for taking multispectral videos through fully utilizing the temporal redundancy of natural dynamic scenes. As shown in Fig. 1(a), the proposed acquisition system is composed of a monochromatic camera and a synchronized Liquid Crystal Tunable Filter (LCTF). The transmission wavelength of the LCTF changes frame by frame through synchronization. As shown in Fig. 1(b), the acquisition system captures the moving object, each time with a different spectral channel. Combined with the proposed COF method and bilateral propagation, multispectral videos at the full spatial resolution and same frame rate of the monochromatic camera can be reconstructed. Our method is based on such an assumption that in a natural video, the object or background appears for at least a period of time and moves continuously during that time. In other words, the scene points in a natural video neither suddenly appear and then disappear in a very short time nor discontinuously move from one location to another non-neighboring area. Under this basic assumption, we propose to use the spectral-sweep camera system (Fig. 1(a)) to capture multispectral videos and then computationally reconstruct the full spatial sensor resolution multispectral channels of all frames by the aid of optical flow maps.

The temporary consistency has already been successfully applied for image-sequences/videos denoising [39], super-resolution [40], deblurring [41], etc. However, the main challenge of this strategy for our application is how to compute optical flow between different spectral channels. Considering the conventional optical flow [42,43], which follows the brightness constancy assumption, the input frames are captured with the same spectral response characteristics, and thus the corresponding points in adjacent frames are of the same intensity. This property leads to the fidelity constraint, which is the most important term for the conventional optical flow algorithms. To deal with the illumination variation among frames, the illumination-changing optical flow methods are further proposed based on the assumption of the spatial or temporal smoothness prior and managed to recover high accuracy optical flow [44–47]. These methods



**Fig. 1. The overview of the proposed multispectral imaging method.** (a) By sweeping the passing band of the Liquid Crystal Tunable Filter (LCTF) frame by frame, the light emitted from a certain moving point pass through the LCTF with different wavelengths and project onto the sensor at different locations. (b) By introducing the COF map and bilateral propagation algorithm, different spectral projections of the same point are aligned and the entire spectrum is reconstructed.

take the illumination variations into consideration when estimating the optical flow, while in the proposed spectral-sweep imaging method, the captured multispectral videos are made up of frames with different wavelengths, not only the illumination but also the spectral reflectance of the captured video frames are changing over time. Since the spectral reflectance depends on the distribution of the surface materials of the scene, which is diverse and locally inconsistent in space, the observed intensity changes are no longer smooth in space and existing optical flow methods [42–47] can not be applied in our problem.

In our method, we model the transformations between adjacent spectral-sweep frames with the complex optical flow (COF), which could model the intensity variation and spatial displacement together. The amplitude of COF models the intensity variation, including both illumination and reflectance changes, and the imagery part models the spatial displacement. Since the reflectance change depends mainly on the surface material composition, leading to the piece-wise smooth patterns [48,49]. In this paper, we enforce the  $L_1$ -sparsity of the derivative of the COF to handle the piece-wise reflectance change. Through optimizing the intensity variation and spatial displacement simultaneously in the complex form, we demonstrate that our method could handle the spectral-sweeping videos accurately. With the estimated COF, the multispectral video is reconstructed from the spectral-sweep video by bilateral propagating.

In conclusion, we propose a novel spatial-temporal sampling framework for taking multispectral video. Through co-designing of spectral-sweep acquisition system and COF-based propagation algorithm, our method achieves comparable reconstruction performance compared with state-of-the-art methods.

## 2. Complex optical flow model

Optical flow is widely used to describe the apparent motion between two adjacent video frames. Based on optical flow maps, the corresponding points in adjacent frames are connected, and the temporal correlation between video frames can be accurately explored. However, almost all

the existing optical flow algorithms are highly dependent on the intensity fidelity term, which constrains that the corresponding points in adjacent frames maintain the same intensity value,

$$\mathbf{I}_{t+1}(x, y) = \mathbf{I}_t(x + \delta x, y + \delta y), \quad (1)$$

where  $x, y$  are the horizontal and vertical coordinates of current frame,  $\delta x$  and  $\delta y$  are the spatial displacements in  $x$  and  $y$  directions between the adjacent frames,  $\mathbf{I}_t$  and  $\mathbf{I}_{t+1}$  denote the image in frame  $t$  and frame  $t + 1$  respectively.

In our application, the adjacent frames of the video captured by our spectral-sweep camera system is of different spectral channel, thus the corresponding points have different intensities. In order to model the intensity variances between adjacent spectral-sweep frames, an additional parameter besides spatial displacement should be introduced. Considering the most common case where the scene is illuminated by a single light source, according to the intrinsic decomposition model [48], the observed images  $s_o$  can be decomposed into the light source component, intrinsic reflectance component and shading component,

$$s_o(x, y, \lambda) = s_l(x, y, \lambda) \cdot s_r(x, y, \lambda) \cdot r_{\text{shading}}(x, y), \quad (2)$$

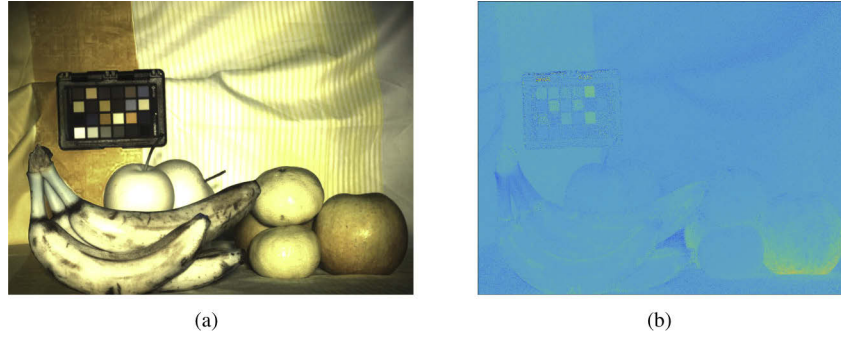
where  $x, y$  are the spatial coordinates,  $\lambda$  denotes the spectral wavelength,  $s_o$  denotes the observed spectral images,  $s_l$  is the spectrum of light source,  $s_r$  is the spectral reflectance of the surface, and  $r_{\text{shading}}$  represents the shading effects caused by 3D structure of the surface. The illumination-changing optical flow [44–47] takes the variation of light source component  $s_l$  into consideration, while our paper deals with both the global brightness changing of  $s_l$  and the locally reflectance changing of  $s_r$  together. Specifically, as for the illumination varying cases, the light source influence  $s_l$  are usually consistent in local areas and varying smoothly. In contrast, as for the reflectance variation caused by spectral sweeping, the spectral reflectance  $s_r$  relies on the surface materials and thus can be very inconsistent in a local area, i.e., the reflectance of a region with various materials can change very differently. Besides, the light sources in practice usually vary very slowly, but the spectral-sweeping caused brightness changing can be very intense.

Accordingly to the characteristics of the illumination-changed cases above, i.e., the spatially local consistence and temporal smooth variation, the smoothness regularization in both spatial and temporal domain can be applied for solving the problem [44–47]. Differently, in the proposed spectral-sweeping based method, different spectral channels are captured along time, not only the illumination  $s_l$  but also the spectral reflectance  $s_r$  of the captured video frames are changing over time. The cross-channel intensity transfer can be model by

$$\begin{aligned} r_{\lambda_1 \rightarrow \lambda_2}(x, y) &= \frac{s_o(x, y, \lambda_1)}{s_o(x, y, \lambda_2)} \\ &= \frac{s_l(x, y, \lambda_1) \cdot s_r(x, y, \lambda_1) \cdot r_{\text{shading}}(x, y)}{s_l(x, y, \lambda_2) \cdot s_r(x, y, \lambda_2) \cdot r_{\text{shading}}(x, y)} \\ &= \frac{s_l(x, y, \lambda_1) \cdot s_r(x, y, \lambda_1)}{s_l(x, y, \lambda_2) \cdot s_r(x, y, \lambda_2)}. \end{aligned} \quad (3)$$

It is obvious that the ratio between different spectral channels can counteract the shading effect and thus the intensity transfer between different spectral channels can be modeled with the multiplicative model. Figure 2 shows an example of the multiplicative transfer maps between spectral wavelength 630nm and 650nm.

After introducing the multiplicative ratio, our cross-channel transfer model for spectral-sweep videos contains two additive offsets for location transfer and one multiplicative ratio for intensity transfer. To deal with these offsets and ratio in a unified model, we propose the Complex Optical Flow model, which is defined in 3D spherical coordinate system (i.e., includes both 1D intensity and 2D location transfer information). Specifically, the 3D spherical coordinate system



**Fig. 2. Multiplicative intensity transfer map.** (a) The synthesized RGB image from the captured multispectral images and (b) the multiplicative transfer map between spectral wavelengths 630nm and 650nm.

(composed of a radius and two angle elements) is used to denote each pixel element in the COF map

$$\mathbf{F}_{t \rightarrow t+1}^{s_1 \rightarrow s_2}(x, y) = \mathbf{F}(x, y)e^{\delta x \mathbf{i} + \delta y \mathbf{j}}, \quad (4)$$

where  $\mathbf{F}_{t \rightarrow t+1}^{s_1 \rightarrow s_2}$  denotes the COF map from time  $t$  to  $t + 1$ , the corresponding swept spectral channel are  $s_1$  and  $s_2$  respectively. The radius of the spherical vector, i.e.,  $\mathbf{F}(x, y)$ , denotes the intensity transfer ratio from the spectral channel  $s_1$  to  $s_2$ , the angles  $\delta x$  and  $\delta y$  are used to describe the locational displacements in  $x$  and  $y$  directions of the image plane,  $\mathbf{i}$  and  $\mathbf{j}$  are the unit angle vectors of the spherical coordinate system respectively. Each pixel in frame  $t, t + 1$  can be represented by  $\mathbf{I}_t^{s_1}(x, y)e^{x\mathbf{i}+y\mathbf{j}}$ ,  $\mathbf{I}_{t+1}^{s_2}(x, y)e^{x\mathbf{i}+y\mathbf{j}}$  and the transformation between the adjacent frames can be modeled by complex multiplication

$$\begin{aligned} \mathbf{I}_{t+1}^{s_2}(x, y)e^{x\mathbf{i}+y\mathbf{j}} &= \mathbf{I}_t^{s_1}(x, y)e^{x\mathbf{i}+y\mathbf{j}} \cdot \mathbf{F}_{t \rightarrow t+1}^{s_1 \rightarrow s_2}(x, y) \\ &= \mathbf{I}_t^{s_1}(x, y)\mathbf{F}(x, y)e^{(x+\delta x)\mathbf{i}+(y+\delta y)\mathbf{j}}, \end{aligned} \quad (5)$$

which can be represented briefly as

$$\mathbf{I}_{t+1}^{s_2} = \mathbf{I}_{t+1}^{s_2} \mathbf{F}_{t \rightarrow t+1}^{s_1 \rightarrow s_2}. \quad (6)$$

The COF maps are defined between the adjacent frame pairs with different spectral channels, which can be applied for our captured spectral-sweep videos. Since the reflectance change depends mainly on the surface material composition of the scene, leading to the piece-wise smooth patterns [48,49]. In this paper, we enforce the  $L_1$ -sparsity of the derivative of the complex optical flow (COF) to handle this piece-wise smooth reflectance changes. Through optimizing the intensity variation and spatial displacement simultaneously in the complex form, we demonstrate that our method could handle the spectral-sweeping videos accurately.

In the following, we propose a complex  $L_1$ -norm constrained optimization algorithm to estimate the COF maps among the captured spectral-sweep frames.

### 3. COF estimation algorithm

Based on the definition of COF maps, the fidelity term of the objective function for estimating a COF map is

$$E_f = \|\mathbf{I}_t^{s_1} \mathbf{F}_{t \rightarrow t+1}^{s_1 \rightarrow s_2} - \mathbf{I}_{t+1}^{s_2}\|_2^2. \quad (7)$$

The problem is still ill-posed by minimizing the fidelity term only. To deal with this ill-posedness, we introduce the complex  $L_1$ -norm regularization on the derivative of the COF map,

$$E_c = \|\nabla \mathbf{F}_{t \rightarrow t+1}^{s_1 \rightarrow s_2}\|_1, \quad (8)$$



where  $\nabla$  is the complex gradient operation on COF maps. Since the COF map here is a complex vector field, the complex gradient is defined as the derivative of adjacent elements in complex vector field along the  $x$ - and  $y$ - directions,

$$\begin{aligned}\nabla_x \mathbf{F} &= \lim_{\Delta x \rightarrow 0} \mathbf{F}(x + \Delta x, y) - \mathbf{F}(x, y) \\ \nabla_y \mathbf{F} &= \lim_{\Delta y \rightarrow 0} \mathbf{F}(x, y + \Delta y) - \mathbf{F}(x, y),\end{aligned}\quad (9)$$

where the subtractions in Eq. (9) are spherical vector operations. The complex  $L_1$ -norm in Eq. (8) denotes the summation of modules of complex gradient vectors.

The complete objective function is a weighted combination of Eq. (7) and Eq. (8),

$$E = E_f + \lambda_c E_c, \quad (10)$$

where  $\lambda_c$  is the weight of the complex  $L_1$ -norm constraint, and in this paper,  $\lambda_c$  is set to 0.02 empirically.

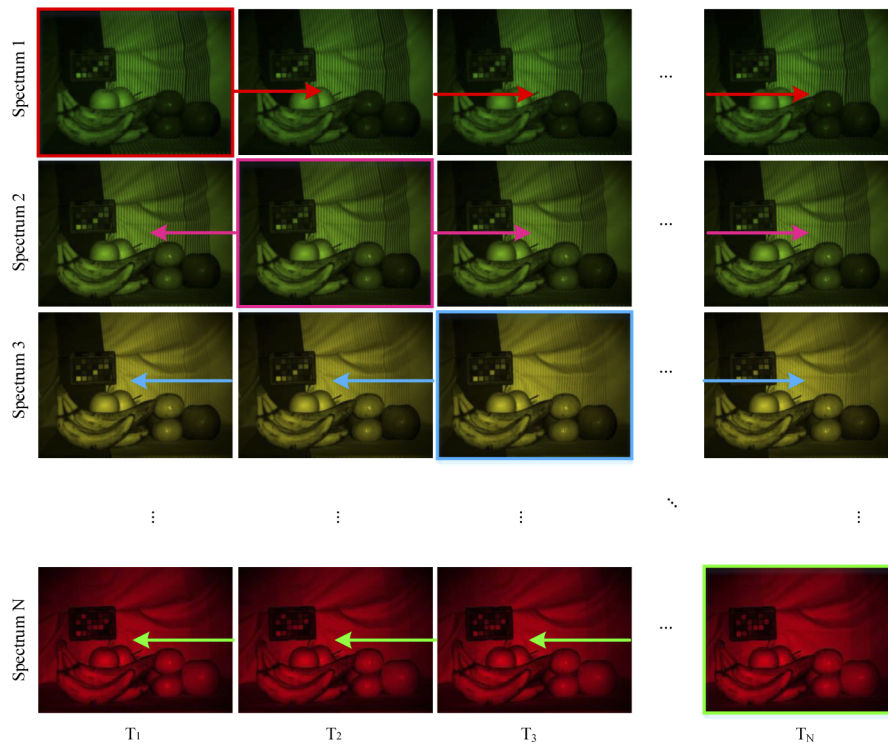
To solve Eq. (10), the Iterative Reweighted Least Squares (IRLS) method [50] is applied. The basic idea of IRLS is approximating the complex  $L_1$ -norm by using weighted  $L_2$ -norm iteratively to make the algorithm converge step by step. Specifically, during the  $k$ -th iteration, the non-quadratic term  $E_c$  in Eq. (8) is approximated by,

$$E_c^k = |\nabla \mathbf{F}_{t \rightarrow t+1}^{s_1 \rightarrow s_2}{}^{(k-1)}|^{-1} |\nabla \mathbf{F}_{t \rightarrow t+1}^{s_1 \rightarrow s_2}|^2, \quad (11)$$

where  $\nabla \mathbf{F}_{t \rightarrow t+1}^{s_1 \rightarrow s_2}{}^{(k-1)}$  is the complex gradient of the  $k - 1$  th iteration. By replacing  $E_c$  with the reweighted quadratic term  $E_c^k$  in Eq. (10), the updated objective function becomes a quadratic function and thus can be easily solved by the gradient descent-based methods. In this paper, we adopt the Conjugate Gradient (CG) algorithm.

During the implementation, a coarse-to-fine framework, which is widely used in optical flow estimation algorithms [51], is adopted to prevent the algorithm from trapping to local minima. In practice, the algorithm starts from a very coarse level (1/256 of the size of input framesets/videos), and at each level, the COF maps are computed by iteratively computing the outer loops (reweighted iteration with  $k$  increasing) and inner loops (CG iteration for minimizing the objective function with fixed weights) until the algorithm converges. After the algorithm converges at each level, the initial COF maps of next level are computed by bilinear-upsampling the current level COF maps. In the very beginning, the first level COF maps are initialized by a constant  $1e^{0i+0j}$ , and the scaling factor between adjacent levels is set to be 1.2.

Through reconstructing the forward and backward COF maps among adjacent sweep-spectral frames, multispectral videos can be recovered by bilaterally propagating the captured spectral channel from the current frame to the other frames with the optical flow part of COF maps, i.e. the 2D location transfer part, as shown in Fig. 3. In a cycle of the spectral-sweep frame set, i.e. the frames sweeping from the first channel (with the shortest wavelength) to the final one (with the longest wavelength), each frame contains one real captured spectral channel, which is propagated in both forward and backward frames according to the COF maps. By assigning the missing points by its corresponding point in the nearest real captured channel according to the COF maps, the information of the real captured spectral channels are propagated to the rest frames to fill the missing channels and the entire multispectral video can be reconstructed.



**Fig. 3. The diagram of bilateral propagation for reconstructing multispectral videos.** The input images (i.e. diagonal frames marked with colored boxes) are captured by the LCTF-based spectral-sweep acquisition system. Each  $N$  frames ( $N$  is the number of sweeping channels of the input spectral-sweep video) from the first channel (with the shortest wavelength) to the last one (with the longest wavelength) are processed as a period unit to reconstruct the corresponding multispectral video. The channels on both sides of input spectral-sweep frames (the diagonal frames) are missing and need to be reconstructed by bilateral propagation. The arrows denotes the propagation directions.

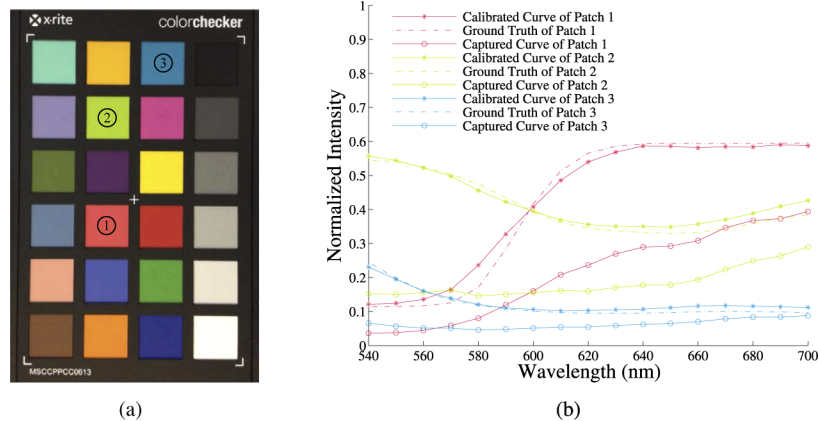
## 4. Experiment

### 4.1. System and calibration

To eliminate the differences of sensitivities on different spectral channels, which is caused by the varying LCTF transmittances and sensor response efficiencies at different wavelengths, we use the color checker (Fig. 4(a)) as a reference board, and correct the output responses by using the first-order fitting for each channel individually. We can see that after the calibration, the output spectral curves of all the color areas match well with respect to the ground truth ones (Fig. 4(b)).

### 4.2. COF demonstration

To verify the proposed COF estimation algorithm, we first tested it on a binocular image pair. The original images are provided by Liu *et al.* [52], and different channels of the images, i.e. the red channel of left view and the green channel of right view, are used as inputs, as shown in Fig. 5. The naive optical flow (NOF) map derived by the human assisted method [52], which can be regarded as the ground truth, is shown in the right of the top row of Fig. 5 as a reference. The first two images of the middle row are results of the NOF algorithm [51] on full-channel input and the synthetic spectral-sweep images (i.e., the red channel of left view and the green



**Fig. 4.** The calibration of sensor and LCTF spectral response responses. (a) The pattern of X-rite ColorChecker we used for calibration. (b) Spectral curves of the corresponding color areas before calibration, after calibration and ground truth.

channel of right view) respectively. It is obvious that although the NOF algorithm [51] works well on full-channel images (shown in the left of middle row), the spectral-sweep inputs bring great challenges and fail the algorithm (shown in the middle of middle row). In contrast, the proposed COF algorithm gives promising result on the spectral-sweep images (shown in the right of the middle row). In the last row, the ground truth of left view and the reconstructed counterpart warped from the right view by using our COF are shown. The error map between ground truth and warping image is also given. We can see that the COF algorithm achieves high accuracy reconstruction on this qualitative comparison.

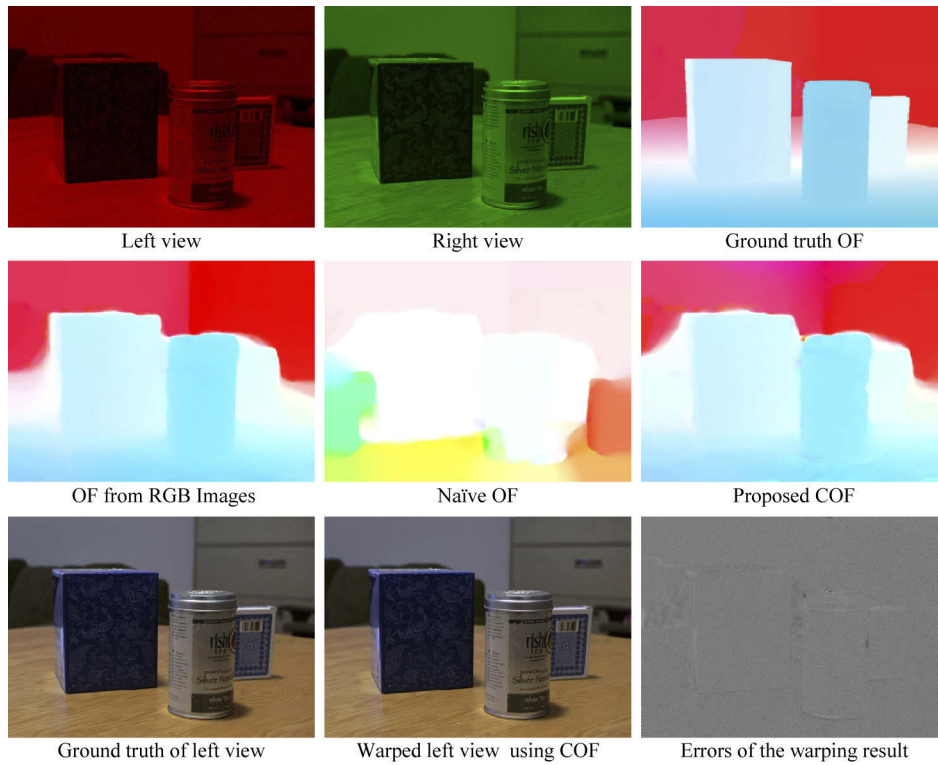
To further quantitatively verify the accuracy of proposed method, we test the proposed COF algorithm on real captured dynamic images with 17 spectral channels (from 540 nm to 700 nm with 10 nm spectral resolution). Since we cannot directly capture the multispectral images simultaneously at each time instant for dynamic scene with LCTF, we simulate the dynamic scene through keeping the scene static and moving the camera step by step. Each camera step corresponds to a time frame, and for each camera movement step, we capture all spectral channels, i.e. 17 images, as the ground truth. Only one channel of each frame is used for producing the captured spectral-sweep video. We compare the PSNR of the NOF, the proposed COF method in retrieving the high-fidelity optical flow cross spectral channels. As shown in Fig. 6, in average, the COF method could retrieve about 5dB PSNR and 0.35 SSIM improvement compared with the NOF method.

#### 4.3. Multispectral reconstruction

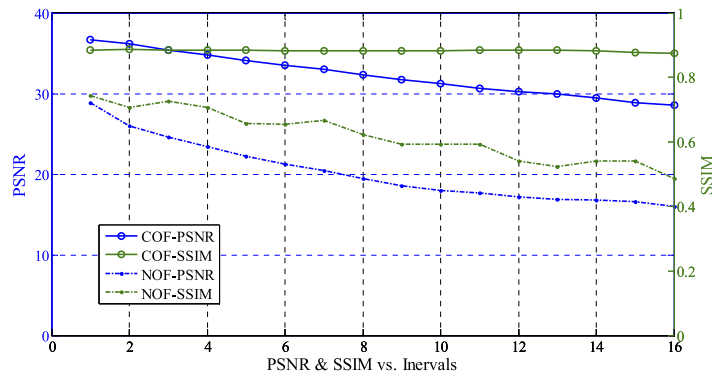
To verify the accuracy of proposed method, we reconstruct the multispectral video on the synthesized sweep-spectral frames extracted from the ground truth multispectral data (Sec. 4.2). Several input single-channel frames and reconstructed channels of the corresponding frames are shown in Fig. 7, the video ([Visualization 1](#)) is provided as a supplementary material. We can see that the reconstructed images are of excellent visual quality except for some holes in marginal regions, which is because these regions are out of the field-of-views of input frames.

To further demonstrate our method quantitatively, we compare the proposed method with existing representative single-camera multispectral imaging methods, including Coded Aperture Snapshot Spectral Imager (CASSI) [14], Prism-Mask Imaging Spectrometer (PMIS) [6], and the deep-learning based CASSI [11] (deep CASSI). To prevent the influences of system errors and noises, we test all the spectral imaging techniques by simulating the optical imaging

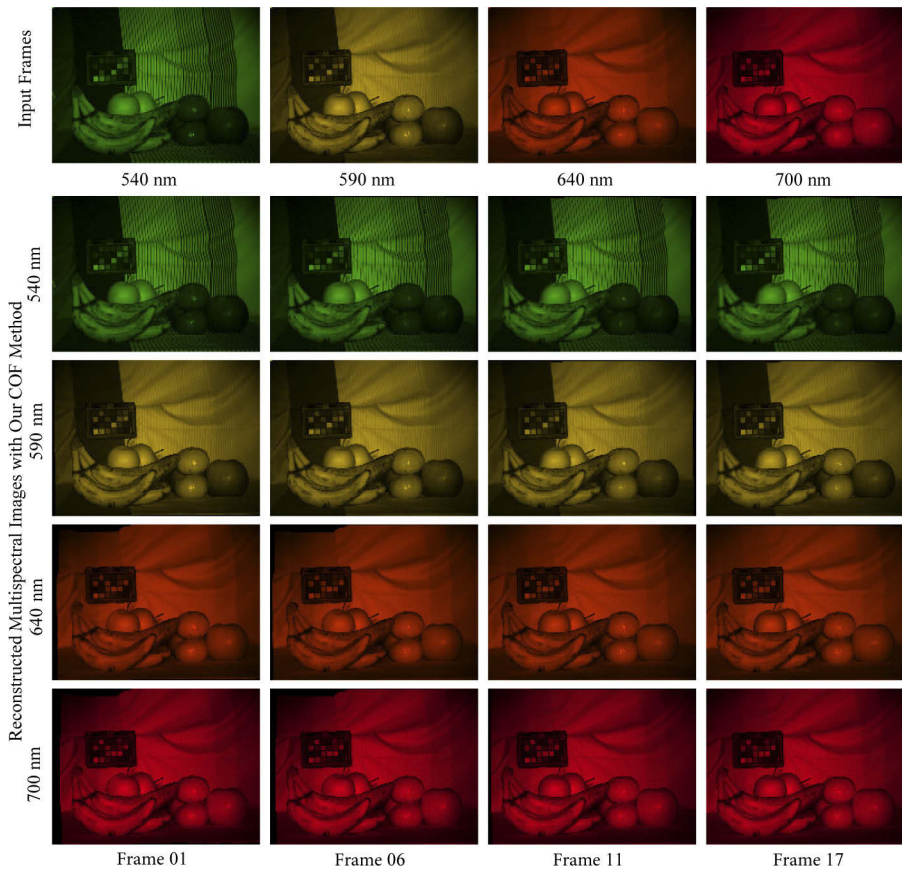




**Fig. 5. The NOF/COF between the red-green images.** We extract different color channels of the binocular images (red channel of left view and green channel of right view) from original full-channel (RGB channels) images (provided by Liu *et al.* [52]), and compare both optical flow maps computed by NOF algorithm [51] and the proposed COF method. The ground truth and estimated optical flow maps from full-channel images are also given. The ground truth of left view and its warping version from right view using our COF, as well as the error map of the warped image are shown in the bottom row.



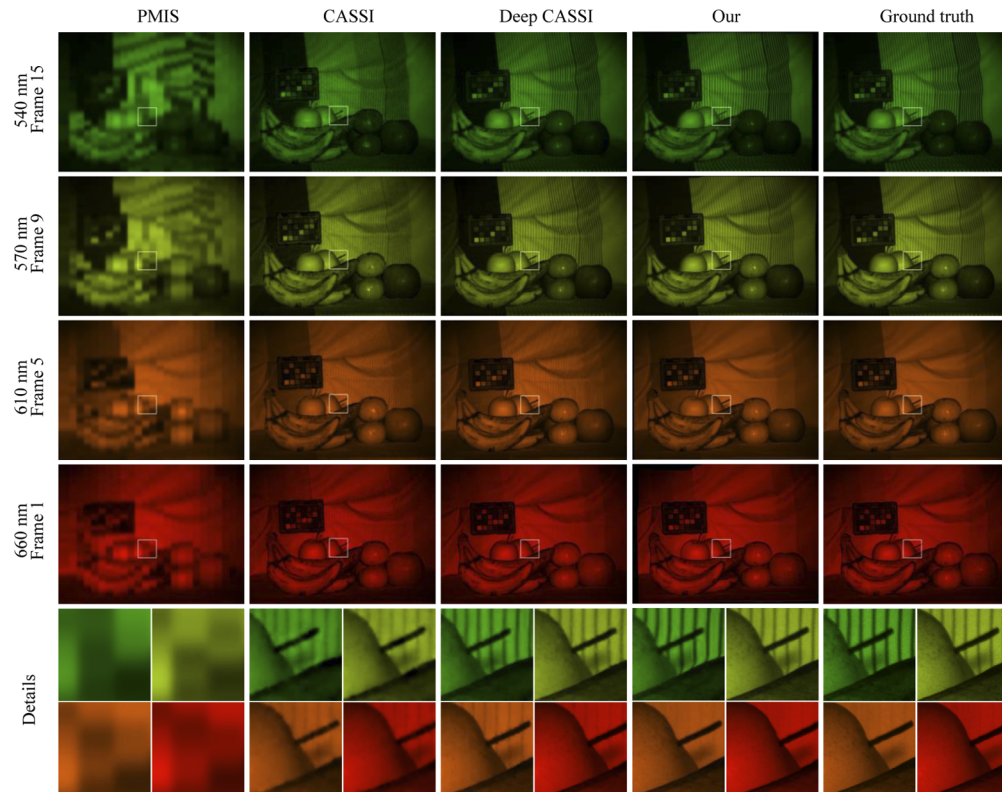
**Fig. 6. Quantitative errors with different frame intervals.** The error curves of the results of NOF algorithm and our COF method with respect to the frame intervals during propagation.



**Fig. 7.** The experimental results on real captured data with ground truth. (Visualization 1) Top row: selected input images captured at view 1, 6, 11 and 17 of spectral wavelength at 540 nm, 590 nm, 640 nm and 700 nm respectively. Middle 4 rows: selected reconstructed channels.

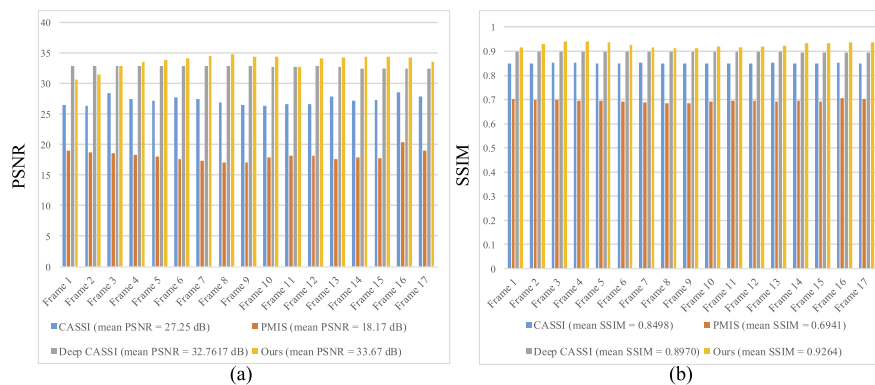
process and applied the following reconstruction algorithm on the simulated measurements from the above ground truth multispectral video. Figure 8 presents the qualitative comparisons between the results of PMIS [6], CASSI [9], Deep CASSI [11], our method and the ground truth. It is shown that the proposed methods achieves visually better or comparable results than the state-of-the-arts. As shown, the spatial resolution of single camera based PMIS is quite limited since it is directly sacrificed for spectral resolution. Conventional CASSI based spectral reconstruction method retrieves more spatial details than PMIS and Deep CASSI retrieves much more details than conventional CASSI through introducing the deep-learned spatial-spectral prior. Our method could retrieve spatial details comparable to Deep CASSI. Through further inspecting the background streaks of Deep CASSI, our method and ground truth (last two rows in Fig. 8), the results of Deep CASSI contains streaks background in 610nm and 660nm, which actually do not exist in the ground truth. Our method could handles this spectral difference and the reconstructed spectral images is consistent with the ground truth.

The quantitative comparison in Figs. 9(a) and 9(b) further verify the promising performance of our method. Note that the mean PSNR shown in Fig. 9(a) is the average PSNR of all the reconstructed spectral channels of each frame. That is because each frame has one real captured channel and 16 reconstructed channels, and without considering the effects of noises and



**Fig. 8. Qualitative comparisons with state-of-the-art spectral video acquisition methods on real captured data with ground truth.** Row 1~4: selected channels at different views. Bottom row: close-ups of the boxed regions.

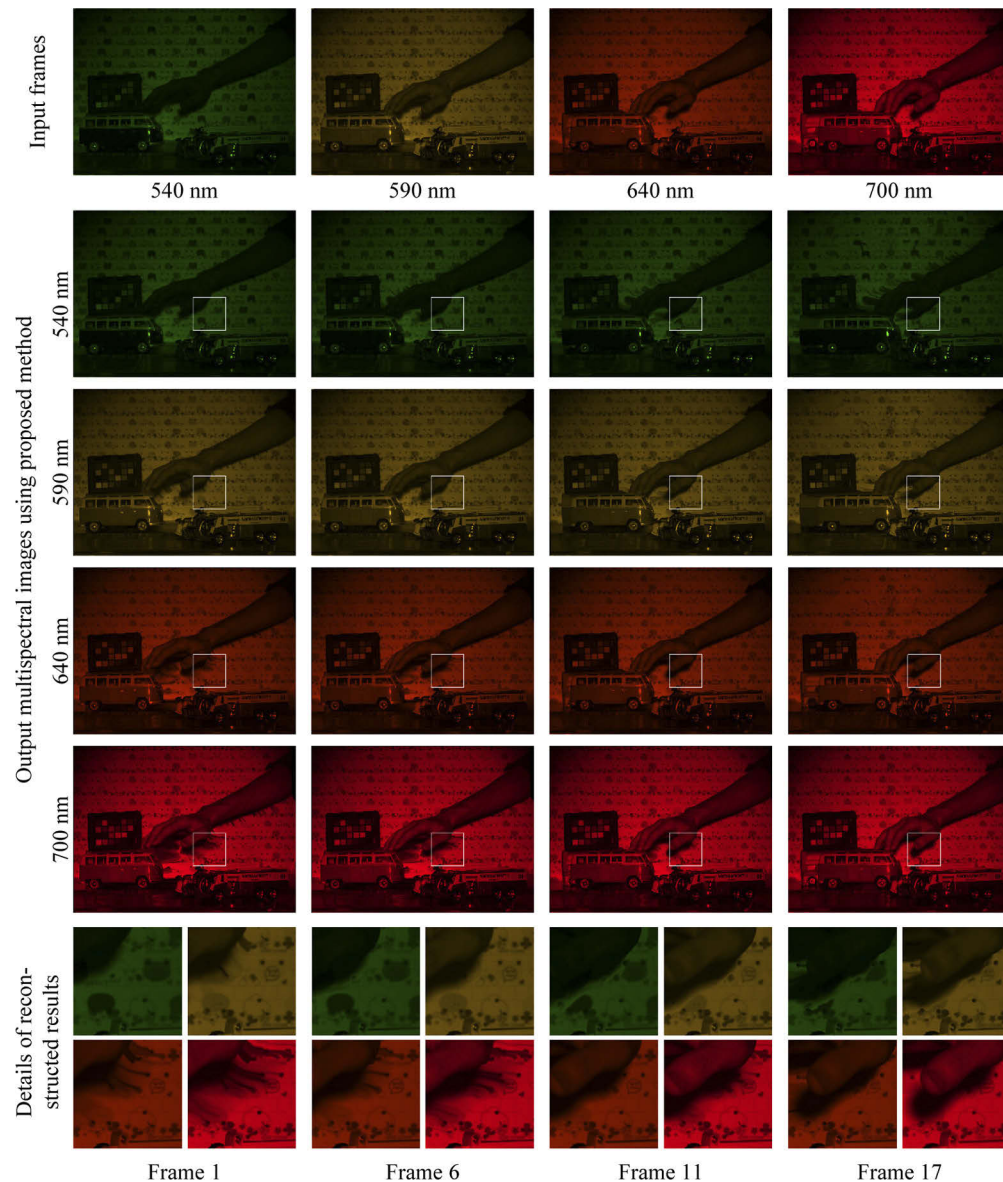
quantization errors, the PSNR of the real captured channel is infinite. To prevent the influence of these infinite values, we just calculate the mean PSNR of reconstructed channels here. Therefore, the real performance of our method should be better than the mean PSNR shown in Fig. 9(a).



**Fig. 9. Quantitative comparisons with state-of-the-art spectral video acquisition methods on real captured data with ground truth.** (a) Mean PSNR of all the reconstructed channels. (b) Mean SSIM of all the channels.



We also apply the proposed system on the moving scenes. As shown in Fig. 10, the dynamic scene with moving objects, i.e. the hand and the toy vehicle, is captured by our LCTF-based spectral-sweep camera system, and several frames of the reconstructed multispectral video (Visualization 2) with 17 channels (from 540 nm to 700 nm with 10 nm spectral resolution) are shown in Fig. 10. To demonstrate the details well, the close-ups (the corresponding regions are marked by rectangles in the full images) are given in the bottom row. All the close-ups are sampled from the same location, and thus the movement of hand is easy to tell from the zoom-in



**Fig. 10.** The experiment results on real captured dynamic scene (Visualization 2). Top row: selected input frames captured at time 1, 6, 11 and 17 with spectral wavelength 540 nm, 590 nm, 640 nm and 700 nm respectively. Middle 4 rows: selected reconstructed channels. Bottom: the details of reconstructed results.

patches. In Fig. 10, we can see that the proposed system works well on the real dynamic scenes. The moving objects are reconstructed with promising quality, and even the shadows of the moving hand are well reconstructed. However, although it is not very obvious, there indeed exist some artifacts in the occlusion regions of the background (occluded by the hand) of some channels. That is because the regions of these channels are occluded on the corresponding real captured single channel frames. For example, the 17th channel (700nm) of Frame 1 are propagated from the real captured channel (700nm) of Frame 17, and the regions with artifacts in 17th channel of Frame 1 are occluded in Frame 17, so the algorithm has to use the adjacent non-occluded pixels to fill the holes in Frame 1, as shown in the insets of Fig. 10.

## 5. Conclusion and discussion

In summary, the proposed system can capture and reconstruct the dynamic multispectral framesets/videos by using the LCTF-based spectral-sweep camera system. The proposed approach is validated by a series of experiments on both synthetic and real captured data. Yet, the current system still has some limitations: (a) The COF may fail in the occluded regions, which may lead to the failure reconstruction. (b) The transmittance of LCTF at some channels (main in the blue band) is very low, so that the system have a bad signal to noise ratio for these channels.

In the future, the adaptive exposure strategy can be used to compensate the low transmittance of the LCTF on the blue band. Besides, the lost information caused by occlusions can be completed by using the similar patches in both the intra and inter frames, i.e. the non-local similarity priors.

## Funding

Young Scholar of Jiangsu Province (BK20160634); Fundamental Research Funds for the Central Universities (021014380128); Project of BMSTC (Z181100003118014); National Natural Science Foundation of China (61627804, 61671236, 61671265, 61971465).

## References

1. S. Denman, T. Lamb, C. Fookes, V. Chandran, and S. Sridharan, "Multi-spectral fusion for surveillance systems," *Comput. Electr. Eng.* **36**(4), 643–663 (2010).
2. V. Backman, M. B. Wallace, L. Perelman, J. Arendt, R. Gurjar, M. Müller, Q. Zhang, G. Zonios, E. Kline, T. McGillican, and S. Shapshay, "Detection of preinvasive cancer cells," *Nature* **406**(6791), 35–36 (2000).
3. P. C. Gray, I. R. Shokair, S. E. Rosenthal, G. C. Tisone, J. S. Wagner, L. D. Rigdon, G. R. Siragusa, and R. J. Heinen, "Distinguishability of biological material by use of ultraviolet multispectral fluorescence," *Appl. Opt.* **37**(25), 6037–6041 (1998).
4. M. Descour and E. Dereniak, "Computed-tomography imaging spectrometer: experimental calibration and reconstruction results," *Appl. Opt.* **34**(22), 4817–4826 (1995).
5. H. Du, X. Tong, X. Cao, and S. Lin, "A prism-based system for multispectral video acquisition," in *Proceedings of IEEE International Conference on Computer Vision*, (IEEE, 2009), pp. 175–182.
6. X. Cao, H. Du, X. Tong, Q. Dai, and S. Lin, "A prism-mask system for multispectral video acquisition," *IEEE Trans. Pattern Anal. Mach. Intell.* **33**(12), 2423–2435 (2011).
7. S.-H. Baek, I. Kim, D. Gutierrez, and M. H. Kim, "Compact single-shot hyperspectral imaging using a prism," *ACM Trans. Graph.* **36**(6), 1–12 (2017).
8. Y. Zhao, X. Hu, H. Guo, Z. Ma, T. Yue, and X. Cao, "Spectral reconstruction from dispersive blur: A novel light efficient spectral imager," in *Proceedings of IEEE Conference on Computer Vision and Pattern Recognition*, (IEEE, 2019), pp. 12202–12211.
9. M. Gehm, R. John, D. Brady, R. Willett, and T. Schulz, "Single-shot compressive spectral imaging with a dual-disperser architecture," *Opt. Express* **15**(21), 14013–14027 (2007).
10. 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).
11. I. Choi, D. S. Jeon, G. Nam, D. Gutierrez, and M. H. Kim, "High-quality hyperspectral reconstruction using a spectral prior," *ACM Trans. Graph.* **36**(6), 1–13 (2017).
12. L. Wang, T. Zhang, Y. Fu, and H. Huang, "Hyperreconnet: Joint coded aperture optimization and image reconstruction for compressive hyperspectral imaging," *IEEE Trans. on Image Process.* **28**(5), 2257–2270 (2019).
13. H. Arguello and G. R. Arce, "Colored coded aperture design by concentration of measure in compressive spectral imaging," *IEEE Trans. on Image Process.* **23**(4), 1896–1908 (2014).



14. A. A. Wagadarikar, N. P. Pitsianis, X. Sun, and D. J. Brady, "Video rate spectral imaging using a coded aperture snapshot spectral imager," *Opt. Express* **17**(8), 6368–6388 (2009).
15. K. M. León-López, L. V. G. Carreño, and H. A. Fuentes, "Temporal colored coded aperture design in compressive spectral video sensing," *IEEE Trans. on Image Process.* **28**, 253–264 (2019).
16. X. Lin, Y. Liu, J. Wu, and Q. Dai, "Spatial-spectral encoded compressive hyperspectral imaging," *ACM Trans. Graph.* **33**(6), 1–11 (2014).
17. 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).
18. L. Wang, Z. Xiong, H. Huang, G. Shi, F. Wu, and W. Zeng, "High-speed hyperspectral video acquisition by combining nyquist and compressive sampling," *IEEE Trans. Pattern Anal. Mach. Intell.* **41**(4), 857–870 (2019).
19. Y. Shechtman, L. E. Weiss, A. S. Backer, M. Y. Lee, and W. Moerner, "Multicolour localization microscopy by point-spread-function engineering," *Nat. Photonics* **10**(9), 590–594 (2016).
20. J. Chen, M. Hirsch, B. Eberhardt, and H. P. Lensch, "A computational camera with programmable optics for snapshot high-resolution multispectral imaging," in *Proceedings of Asian Conference on Computer Vision*, (IEEE, 2018), pp. 685–699.
21. D. S. Jeon, S.-H. Baek, S. Yi, Q. Fu, X. Dun, W. Heidrich, and M. H. Kim, "Compact snapshot hyperspectral imaging with diffracted rotation," *ACM Trans. Graph.* **38**(4), 1–13 (2019).
22. A. Manakov, J. Restrepo, O. Klehm, R. Hegedus, E. Eisemann, H.-P. Seidel, and I. Ihrke, "A reconfigurable camera add-on for high dynamic range, multispectral, polarization, and light-field imaging," *ACM Trans. Graph.* **32**(4), 1 (2013).
23. T. Takatani, T. Aoto, and Y. Mukaigawa, "One-shot hyperspectral imaging using faced reflectors," in *Proceedings of IEEE Conference on Computer Vision and Pattern Recognition*, (IEEE, 2017), pp. 4039–4047.
24. B. Arad and O. Ben-Shahar, "Sparse recovery of hyperspectral signal from natural rgb images," in *Proceedings of European Conference on Computer Vision*, (IEEE, 2016), pp. 19–34.
25. Y. Fu, T. Zhang, Y. Zheng, D. Zhang, and H. Huang, "Joint camera spectral sensitivity selection and hyperspectral image recovery," in *Proceedings of European Conference on Computer Vision*, (IEEE, 2018), pp. 788–804.
26. Y. Fu, Y. Zheng, L. Zhang, and H. Huang, "Spectral reflectance recovery from a single rgb image," *IEEE Trans. Comput. Imaging* **4**(3), 382–394 (2018).
27. N. Akhtar and A. S. Mian, "Hyperspectral recovery from rgb images using gaussian processes," *IEEE Trans. Pattern Anal. Mach. Intell.* (2018).
28. H. Li, Z. Xiong, Z. Shi, L. Wang, D. Liu, and F. Wu, "Hsvcn: Cnn-based hyperspectral reconstruction from rgb videos," in *Proceedings of IEEE International Conference on Image Processing* (IEEE, 2018), pp. 3323–3327.
29. J. Bao and M. G. Bawendi, "A colloidal quantum dot spectrometer," *Nature* **523**(7558), 67–70 (2015).
30. P.-J. Lapray, X. Wang, J.-B. Thomas, and P. Gouton, "Multispectral filter arrays: Recent advances and practical implementation," *Sensors* **14**(11), 21626–21659 (2014).
31. S. Nie, L. Gu, Y. Zheng, A. Lam, N. Ono, and I. Sato, "Deeply learned filter response functions for hyperspectral reconstruction," in *Proceedings of IEEE Conference on Computer Vision and Pattern Recognition* (IEEE, 2018), pp. 4767–4776.
32. J. Jia, K. J. Barnard, and K. Hirakawa, "Fourier spectral filter array for optimal multispectral imaging," *IEEE Trans. on Image Process.* **25**(4), 1530–1543 (2016).
33. C. Ni, J. Jia, M. Howard, K. Hirakawa, and A. Sarangan, "Single-shot multispectral imager using spatially multiplexed fourier spectral filters," *J. Opt. Soc. Am. B* **35**(5), 1072–1079 (2018).
34. N. Gat, "Imaging spectroscopy using tunable filters: a review," in *Wavelet Applications VII*, vol. 4056 (International Society for Optics and Photonics, 2000), pp. 50–65.
35. H. Lee and M. H. Kim, "Building a two-way hyperspectral imaging system with liquid crystal tunable filters," in *Proceedings of International Conference on Image and Signal Processing* (Springer, 2014), pp. 26–34.
36. X. Wang, Y. Zhang, X. Ma, T. Xu, and G. R. Arce, "Compressive spectral imaging system based on liquid crystal tunable filter," *Opt. Express* **26**(19), 25226–25243 (2018).
37. M. Descour and E. Dereniak, "Computed-tomography imaging spectrometer: experimental calibration and reconstruction results," *Appl. Opt.* **34**(22), 4817–4826 (1995).
38. A. Mian and R. Hartley, "Hyperspectral video restoration using optical flow and sparse coding," *Opt. Express* **20**(10), 10658–10673 (2012).
39. C. Liu and W. T. Freeman, "A high-quality video denoising algorithm based on reliable motion estimation," in *Proceedings of European Conference on Computer Vision* (Springer, 2010), pp. 706–719.
40. C. Liu and D. Sun, "A bayesian approach to adaptive video super resolution," in *Proceedings of IEEE International Conference on Computer Vision and Pattern Recognition* (IEEE, 2011), pp. 209–216.
41. S. Cho, J. Wang, and S. Lee, "Video deblurring for hand-held cameras using patch-based synthesis," *ACM Trans. Graph.* **31**(4), 1–9 (2012).
42. T. Brox, A. Bruhn, N. Papenberger, and J. Weickert, "High accuracy optical flow estimation based on a theory for warping," in *Proceedings of European Conference on Computer Vision* (Springer, 2004), pp. 25–36.
43. C. Liu, J. Yuen, and A. Torralba, "Sift flow: Dense correspondence across scenes and its applications," *IEEE Trans. Pattern Anal. Mach. Intell.* **33**(5), 978–994 (2011).

44. S. Negahdaripour, "Revised definition of optical flow: Integration of radiometric and geometric cues for dynamic scene analysis," *IEEE Trans. Pattern Anal. Mach. Intell.* **20**(9), 961–979 (1998).
45. H. W. Haussecker and D. J. Fleet, "Computing optical flow with physical models of brightness variation," *IEEE Trans. Pattern Anal. Mach. Intell.* **23**(6), 661–673 (2001).
46. C.-H. Teng, S.-H. Lai, Y.-S. Chen, and W.-H. Hsu, "Accurate optical flow computation under non-uniform brightness variations," *Comput. Vision Image Understanding* **97**(3), 315–346 (2005).
47. D. Fortun, P. Bouthemy, and C. Kervrann, "Optical flow modeling and computation: a survey," *Comput. Vision Image Understanding* **134**, 1–21 (2015).
48. J. T. Barron and J. Malik, "Shape, illumination, and reflectance from shading," *IEEE Trans. Pattern Anal. Mach. Intell.* **37**(8), 1670–1687 (2015).
49. S. Bi, X. Han, and Y. Yu, "An  $l_1$  image transform for edge-preserving smoothing and scene-level intrinsic decomposition," *ACM Trans. Graph.* **34**(4), 78 (2015).
50. R. W. Wedderburn, "Quasi-likelihood functions, generalized linear models, and the gauss–newton method," *Biometrika* **61**(3), 439–447 (1974).
51. C. Liu, "Beyond pixels: Exploring new representations and applications for motion analysis," Ph.D. thesis, Massachusetts Institute of Technology (2009).
52. C. Liu, W. T. Freeman, E. H. Adelson, and Y. Weiss, "Human-assisted motion annotation," in *Proceedings of IEEE International Conference on Computer Vision and Pattern Recognition* (IEEE, 2008), pp. 1–8.