High-speed Hyperspectral Video Acquisition with a Dual-camera Architecture

Lizhi Wang[1], Zhiwei Xiong[2], Dahua Gao[3], Guangming Shi[1], Wenjun Zeng[2], Feng Wu[4]
[1]Xidian University. [2]Microsoft Research. [3]Air Force Engineering University. [4]University of Science and Technology of China.

Snapshot spectral imagers have seen a rapid development owing to the flourish of computational reconstruction. Techniques in this category, including computed tomographic imaging spectrometry (CTIS) [2], coded aperture snapshot spectral imager (CASSI) [4], and hybrid spectral video imaging system (HVIS) [1], have the ability to recover a full hyperspectral image with a single shot and so to acquire a 4D hyperspectral video. Nevertheless, the video frame rate that can be achieved with these snapshot spectral imagers is usually limited compared to the RGB/panchromatic cameras equipped with the same detector. So far, the highest frame rate reported in literature is 30fps by CASSI, for a bright scene with burning candles [4].

We propose a novel dual-camera design to acquire 4D high-speed hyperspectral (HSHS) videos with high spatial and spectral resolution. Our work has two key technical contributions. First, we build a dual-camera system that simultaneously captures a panchromatic video at a high frame rate and a hyperspectral video at a low frame rate, which jointly provide reliable projections for the underlying HSHS video. Second, we exploit the panchromatic video to learn an over-complete 3D dictionary to represent each band-wise video sparsely, and a robust computational reconstruction is then employed to recover the HSHS video based on the joint videos and the self-learned dictionary. Experimental results demonstrate that, for the first time to our knowledge, the hyperspectral video frame rate reaches up to 100fps with decent quality, even when the incident light is not strong.

The data flow in the proposed system is detailed in Fig. 1. As can be seen, there are two branches after the beam splitter. In the panchromatic camera (PanCam) branch, there is simply an objective lens in front of the detector and thus the light path is short and unobstructed. In the CASSI branch, light is first encoded by a coded aperture and then dispersed by a dispersive prism before reaching the detector, which results in considerable light intensity attenuation. Suppose the detectors in the two branches are identical, the PanCam branch can work at a much higher frame rate than the CASSI branch in practice, due to higher efficiency of light utilization. That is to say, the PanCam branch lacks in spectral resolution while the CASSI branch lacks in temporal resolution. Therefore, it is possible to recover 4D HSHS videos by jointly using the measurements from the two branches, under elaborate calibration and synchronization.

Let $f(x,y,\lambda,t)$ denote the scene information of a 4D HSHS video clip in its discrete form, where $1 \le x \le W$ and $1 \le y \le H$ index the spatial coordinates, $1 \le \lambda \le \Omega$ indexes the spectral coordinate, and $1 \le t \le K$ indexes the temporal coordinate. Since the beam splitter equally divides the incident light, the high frame-rate PanCam video captured at time t can be written as

$$g^p(x,y,t) = 0.5 \sum_{\lambda=1}^{\Omega} w(\lambda) f(x,y,\lambda,t), \tag{1}$$

where $w(\lambda)$ is the spectral response function of the detector.

The low frame-rate CASSI image captured during the whole clip can be written as

$$g^c(x,y) = 0.5 \sum_{t=1}^{K} \sum_{\lambda=1}^{\Omega} w(\lambda) S(x, y-\phi(\lambda)) f(x, y-\phi(\lambda), \lambda, t), \tag{2}$$

where $S(x,y)$ denotes the transmission function of the coded aperture and $\phi(\lambda)$ denotes the wavelength-dependent dispersion function of the prism.

The dual-camera system model can then be expressed as

$$G = \Phi F, \tag{3}$$

where G includes all the measurements, Φ is a sparse matrix representing the overall system forward operation, and F is the underlying HSHS video.

This is an extended abstract. The full paper is available at the Computer Vision Foundation webpage.

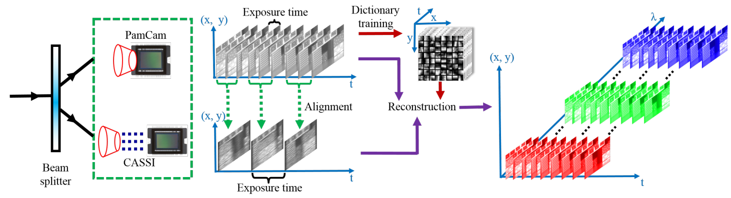

Figure 1: Data flow in the proposed dual-camera system.

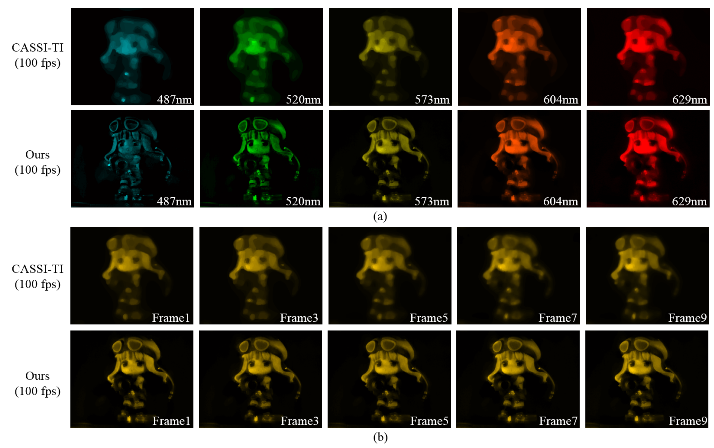

Figure 2: HSHS video reconstruction results of a fast moving doll under ordinary indoor illumination. Exposure time for CASSI and PanCam is 100ms and 10ms, respectively. (a) Results of different selected bands at one temporal location. (b) Results of one selected band at different temporal locations. CASSI-TI denotes the temporal interpolation of CASSI reconstruction.

The panchromatic video is further exploited to learn an over-complete 3D dictionary to represent each band-wise video sparsely. This is motivated by the observation that a 4D HSHS video can be treated as a concatenation of multiple band-wise videos which often have similar structural content as the panchromatic video. Therefore, the dictionary learned from the panchromatic video yields high sparsity when representing the band-wise videos. We can then solve the following minimization problem instead

$$\hat{\alpha} = \arg\min_{\alpha} ||G - \Phi D \circ \alpha||_2^2 + \tau ||\alpha||_0, \tag{4}$$

where D is the self-learned dictionary, α is the concatenation of the sparse coefficients of all band-wise patches in F when represented on D, the operation $\circ$ derives F from D and α, and τ is a regularization parameter balancing the data fidelity and the prior sparsity. Eq. 4 can be easily solved by employing the orthogonal matching pursuit algorithm [3].

Figure 2 shows a part of experimental results based on the proposed approach. For more results, please refer to our full paper.

[1] Xun Cao, Xin Tong, Qionghai Dai, and S. Lin. High resolution multispectral video capture with a hybrid camera system. In *CVPR*, 2011.

[2] Takayuki Okamoto and Ichirou Yamaguchi. Simultaneous acquisition of spectral image information. *Opt. Lett.*, 16(16):1277–1279, 1991.

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

[4] 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 – 88, 2009.