EUROGRAPHICS 2016 / J. Jorge and M. Lin
(Guest Editors)

Volume 35 (2016), Number 2

Multisampling Compressive Video Spectroscopy

Daniel S. Jeon Inchang Choi Min H. Kim †

Korea Advanced Institute of Science and Technology (KAIST)

Abstract

The coded aperture snapshot spectral imaging (CASSI) architecture has been employed widely for capturing hyperspectral video. Despite allowing concurrent capture of hyperspectral video, spatial modulation in CASSI sacrifices image resolution significantly while reconstructing spectral projection via sparse sampling. Several multiview alternatives have been proposed to handle this low spatial resolution problem and improve measurement accuracy, for instance, by adding a translation stage for the coded aperture or changing the static coded aperture with a digital micromirror device for dynamic modulation. State-of-the-art solutions enhance spatial resolution significantly but are incapable of capturing video using CASSI. In this paper, we present a novel compressive coded aperture imaging design that increases spatial resolution while capturing 4D hyperspectral video of dynamic scenes. We revise the traditional CASSI design to allow for multiple sampling of the randomness of spatial modulation in a single frame. We demonstrate that our compressive video spectroscopy approach yields enhanced spatial resolution and consistent measurements, compared with the traditional CASSI design.

Categories and Subject Descriptors (according to ACM CCS): I.4.1 [Computer Graphics]: Image Processing and Computer Vision—Digitization and Image Capture

1. Introduction

Hyperspectral imaging spectroscopy has been practiced broadly to acquire three-dimensional spectral information of scenes. Many different designs of imaging spectroscopy have been proposed and widely used in many fields, e.g., scientific study, product inspection, aerial/satellite imaging, military applications, etc. Spectral imagers are built with two-dimensional imaging sensors. Consequently, there has been a long-lasting tradeoff between the spatial and the spectral resolutions toward *video capability*. For instance, while bandpass filter-based systems [MRK*13, LK14] can provide a high spatial resolution, their spectral resolution is limited according to the number of filters. Pushbroom spectral imagers such as [HF13] can provide a high spectral resolution as well as spatial resolution. However, both scanning approaches are limited to static scenes due to temporal scanning and are incapable of *video spectroscopy* of dynamic scenes.

Snapshot-based imaging approaches, such as coded aperture snapshot spectral imaging (CASSI) [WJWB08, DTCL09], have been proposed to capture *spectral video* of dynamic scenes. However, the fundamental tradeoff problem remains for CASSI to support only a *very low spatial resolution* of spectral data, compared to conventional trichromatic imagers. Several alternatives for CASSI have therefore been proposed to better handle this tradeoff. Kittle et al. [KCWB10] revised the original design [WJWB08] with a micro-translation stage of a coded aperture mask, allowing for multiple sampling, and Tsai et al. [TB13] utilized the same design for temporal modulation for low-resolution video spectroscopy. Several methods meanwhile employ a digital micromirror device (DMD) for using a dynamic mask, rather than a static pattern, to enhance spatial resolution in CASSI [WGSN13, RAA15]. However, these solutions require multiple snapshots and thus are incapable of video spectroscopy. Alternatively, there are several approaches that combine two different types of cameras, a low-resolution hyperspectral camera and a high-resolution RGB camera, to increase spatial and spectral resolution [KWT*11, CDT*11, WXG*15b]. However, these systems approximate spectral information indirectly in hyperspectral videos, rather than directly measuring the actual spectrum per pixel. Furthermore, these multi-camera approaches are more costly than conventional single camera solutions.

In this paper, we propose combining compressive coded imaging and kaleidoscopic imaging, allowing for multiple sampling of the compressive codes. A kaleidoscopic imaging configuration has been used to capture multiple views of dynamic scenes or a large number of views efficiently. We build an alternative CASSI system for *video spectroscopy with enhanced spatial resolution*. We developed an optical design that allows for multisampling of compressive coded snapshots. The proposed design increases the sampling ratio significantly while capturing hyperspectral video on a single image sensor architecture. We reconstruct input into a four-dimensional hyperspectral video (x, y, λ, t) by solving a sparsity-

† Author emails: {sjjeon; inchangchoi; minhkim (corresponding author)}@vclab.kaist.ac.kr

constrained optimization problem with total variation. We demonstrate that the proposed method improves the spatial resolution of compressive coded imaging without sacrificing video capability relative to the traditional CASSI design.

Our **contributions** are:

- a novel optical design that enables multisampling for compressive video spectroscopy and
- a system implementation as a prototype to validate the design.

2. Related Work

Imaging spectroscopy has been researched extensively in recent decades. For brevity's sake, we refer readers to [Bra09] for background and an overview of this subject. This section exclusively surveys dispersion-based spectroscopy in detail.

Fundamental Resolution Tradeoff. Hyperspectral imaging is a spectroscopy technology that captures image information with an additional dimension of spectra. Traditional hyperspectral imaging systems require temporal scanning to reconstruct a three-dimensional spatio-spectral data cube. A hyperspectral imager employs a two-dimensional imaging sensor to capture input. Since the dimensions of the sensor are lower than those of the hyperspectral data cube, a tradeoff between the spatial and the spectral resolution in these systems is inevitable. Diverse optical designs and algorithms have been proposed to better handle this tradeoff of spatio-spectral information.

Filter-Based Spectroscopy. Bandpass-filter-based spectroscopy captures a sequence of images with narrow bandpass filters [RB05] or a liquid crystal tunable filter [LK14, NK14], coupled with a monochromatic camera, and reconstructs a hyperspectral image by packing spectral channels. The spectral resolution depends on the number of filters, and the spatial resolution is determined by the resolution of the sensor. Filter-based imaging provides high spatial resolution but with limited spectral resolution. Since multiple channels must be captured via temporal scanning, subjects being captured are limited to static objects. Recently, Manakov et al. [MRK*13] introduced a filter-based snapshot multispectral imaging system based on a kaleidoscope. The kaleidoscope produces $N \times N$ identical copies of the original image. Each image is filtered by a bandpass-filter with a different wavelength band. We are inspired by this optical design of image duplication, and apply it to *compressive coded aperture snapshot imaging*. Different from Manakov et al.'s approach, we duplicate images with diverse random apertures, allowing for multisampling of compressive coding.

Pushbroom Spectroscopy. A pushbroom-based system isolates an image into a narrow column through a single slit, disperses each column by a prism or a diffraction grating to mechanically scan optical dispersion, and then packs the column-wise dispersion into a spectral image. The drawbacks of this design are that the spatial resolution along the mechanically moving axis is lower than that of the other axis direction, and that these systems, like filter-based systems, can capture only *static scenes*. The spatial resolution of pushbroom-based systems is, however, higher than that of filter-based systems. The spectral resolution is determined by the number of pixels within the range of spectral dispersion in the sensor. Recently, Hoye et al. [HF13] presented a system by physically attaching a set of light mixing chambers on the slit to reduce typical artifacts in the pushbroom architecture.

Snapshot Spectroscopy. Compared with the pushbroom- or filter-based systems, snapshot-based systems capture a full 3D spatial-spectral data with a snapshot and are capable of capturing dynamic scenes. Although *snapshot* spectroscopy is capable of hyperspectral video, the technical tradeoff between spatial and spectral resolution still remains as a severe problem. Snapshot spectroscopy also utilizes dispersion by a prism or a diffraction grating coupled with a coded mask to reconstruct spectral information by solving a projection problem. Gehm et al. [GJB*07] introduced a pioneering snapshot spectroscopy system, so-called dual-disperser coded aperture snapshot spectral imaging (DD-CASSI), which uses a coded aperture to modulate spectral data. Du et al. [DTCL09] and Wagadarikar et al. [WPSB08, WPSB09] proposed single-disperser CASSI systems (SD-CASSI) that allow for video spectroscopy. Rajwade et al. [RKT*13] developed CASSI system by using a Bayesian implementation of blind compressive sensing. Habel et al. [HKW12] proposed a hyperspectral imager by revising a conventional camera to reconstruct spectral information via computed tomography imaging. However, the spatial resolution of the system is limited to 120×120 pixels. These snapshot-based systems suffer from a lack of spatial resolution due to the aforementioned tradeoff.

Multi-Snapshot Spectroscopy. In order to improve spatial resolution, several alternative designs have been proposed to increase the sampling rate in CASSI. Kittle et al. [KCWB10, KHK*12] proposed a multiple snapshot CASSI that captures many snapshots posing a coded mask on a micro translation stage to randomly translate the position of the mask. Wu et al. [WMAP11] utilized a DMD as a programmable coded aperture to diversify the random pattern of the coded mask. Even though these systems improve the spatial resolution significantly compared to single snapshot approaches, they are incapable of capturing dynamic scenes for hyperspectral video.

Another alternative approach, a dual-camera system with a low-resolution hyperspectral camera and a high-resolution DSLR camera, was proposed [CTDL11, WXG*15b, MCWD14]. This method propagates low-resolution spectral information to the high-resolution pixel domain from the RGB camera by estimating pixel similarity or a basis that represents reflectance spectra via matrix factorization [KWT*11]. However, these high-resolution spectra are approximations rather than direct measurements of actual spectra.

Kaleidoscopic Imaging. Reshetouski et al. [RMSI11] proposed calibration and imaging theory for kaleidoscopic imaging configurations. They then investigated optical geometry in a room of planar mirrors [RMB*13]. Planar mirrors have been broadly used in various imaging applications. For instance, a planar mirror system was proposed for multiple views in capturing reflectance functions [HP03]. A reconfigurable kaleidoscopic imaging system was also introduced for high-dynamic-range imaging, light-field imaging and multispectral imaging, which is based on bandpass filters [MRK*13]. To the best of our knowledge, our work is the first to apply kaleidoscopic imaging to compressive coded imaging to achieve high resolution in hyperspectral videos.

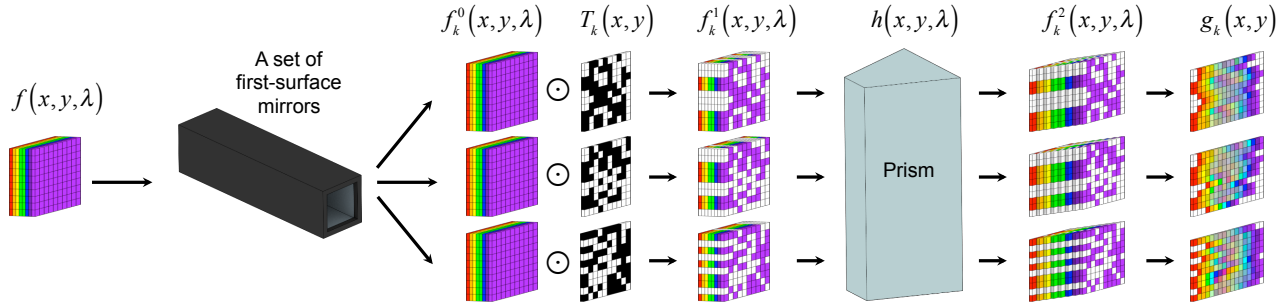

Figure 1: *Schematic diagram of our multisampling compressive coded imaging. An initial hyperspectral image f passing an objective lens is duplicated by first-surface mirrors and modulated by* different *coded aperture masks $\mathbf{T} = [T_1, T_2, \cdots, T_k]$ to coded images $\mathbf{f}^1 = [f_1^1, f_2^1, \cdots, f_k^1]$. $\mathbf{f}^1$ is then dispersed as a shear spectral function $\mathbf{f}^2 = [f_1^2, f_2^2, \cdots, f_k^2]$, and projected to a monochromatic sensor $\mathbf{g} = [g_1, g_2, \cdots, g_k]$.*

3. Multisampling Compressive Spectroscopy

We were motivated to capture hyperspectral video with enhanced spatial resolution. We built our CASSI system design from scratch but extended the compressive design toward multiple sampling to enhance the spatial resolution of traditional CASSI without sacrificing video capture capability. We begin by describing the foundations of compressive imaging, followed by our design.

3.1. Compressive Imaging with Multisampling

Let us define the spectral intensity of light in a frame as a function f of wavelengths λ and spatial locations (x,y). The input image with continuous spectra f, shown on the left in Figure 1, is duplicated into k instances of f_k^0 images

$$\mathbf{f} = [f_1^0, f_2^0, \cdots, f_k^0],$$

where the initial image f on a diffuse screen is multiplied by four first-surface mirrors. See Section 4.1 for details of optical implementation.

Each duplicated image f_k^0 is filtered by a *different* coded aperture transmission function $T_k(x,y)$, respectively. Note that the random mask functions $\mathbf{T} = [T_1, T_2, \cdots, T_k]$ are different while the duplicated f_k^0 images are identical. The modulated spectral density function $\mathbf{f}^1 = [f_1^1, f_2^1, \cdots, f_k^1]$ via the coded aperture $\mathbf{T}$ is computed by a product:

$$f_k^1(x,y,\lambda) = f_k^0(x,y,\lambda) T_k(x,y)\ .$$

A prism disperses the duplicated coded images $\mathbf{f}^1$ along a horizontal axis. The length of dispersion of a wavelength λ can be obtained from the calibration of dispersion as a function of $\phi(\lambda)$, which describes the amount of pixel shift with respect to wavelength λ. We can describe the spectral density f_k^2 after the diffraction grating as

$$\begin{aligned} f_k^2(x,y,\lambda) &= \iint \delta(x' - [x + \phi(\lambda)])\delta(y' - y) f_k^1(x',y',\lambda)dx'dy' \\ &= \iint h(x' - \phi(\lambda), x, y', y, \lambda) f_k^1(x',y',\lambda)dx'dy', \end{aligned}$$

where h describes the two-dimensional dispersion of $\phi(\lambda)$ via the prism as a combined function of two Dirac delta functions $\delta(x' - [x + \phi(\lambda)])\delta(y' - y)$.

The detector array only measures the intensity of the light rather than the spectral density. The intensity of a position (x,y) is integration over a set of wavelengths Λ in the k-th duplication:

$$\begin{aligned} g_k(x,y) &= \int_\Lambda \iint h(x' - \phi(\lambda), x, y', y, \lambda) f_k^1(x',y',\lambda)dx'dy'd\lambda \\ &= \int_\Lambda \iint h(x' - \phi(\lambda), x, y', y, \lambda) T_k(x,y) \\ &\quad \times f_k^0(x',y',\lambda)dx'dy'd\lambda. \end{aligned} \tag{1}$$

Suppose we have an image sensor that pixelates the light intensity as a two-dimensional array of a pixel size Δ. We can rewrite the discrete pixel intensity $\mathbf{g}$ at a position (i,j) in the k-th duplication as:

$$\begin{aligned} \mathbf{g}_{ijk} &= \iint g_k(x,y) \mathrm{rect}\left(\frac{x}{\Delta} - i, \frac{y}{\Delta} - j\right) dxdy, \\ &= \iint \int_\Lambda \iint h(x' - \phi(\lambda), x, y', y, \lambda) T_k(x,y) \\ &\quad \times f_k^0(x',y',\lambda) \mathrm{rect}\left(\frac{x}{\Delta} - i, \frac{y}{\Delta} - j\right) dx'dy'd\lambda dxdy. \end{aligned} \tag{2}$$

In the same manner, the coded aperture mask $T_k(x,y)$ can be formulated as a set of discrete pinholes $\mathbf{T}_{i'j'k}$ with a pixel size of Δ':

$$T_k(x,y) = \sum_{i',j'} \mathbf{T}_{i'j'k} \mathrm{rect}\left(\frac{x}{\Delta'} - i', \frac{y}{\Delta'} - j'\right). \tag{3}$$

We can substitute $T_k(x,y)$ in Equation (2) with Equation (3):

$$\begin{aligned} \mathbf{g}_{ijk} = \sum_{i',j'} \mathbf{T}_{i'j'k} \iint \int_\Lambda \iint & h(x' - \phi(\lambda), x, y', y, \lambda) \\ & \times \mathrm{rect}(\frac{x}{\Delta'} - i', \frac{y}{\Delta'} - j') \\ & \times f_k^0(x',y',\lambda) \mathrm{rect}(\frac{x}{\Delta} - i, \frac{y}{\Delta} - j) dx'dy'd\lambda dxdy. \end{aligned} \tag{4}$$

Now we can rewrite the above equation in a matrix-vector form. Suppose the captured input image of k duplication, is $\mathbf{g} \in \mathbb{R}^{ijk}$. A hyperspectral image function $f_k^0(x',y',\lambda)\mathrm{rect}(\frac{x}{\Delta} - i, \frac{y}{\Delta} - j)$ of k duplication with l number of wavelengths is $\mathbf{f} \in \mathbb{R}^{ijl}$. The projection of the spectral dispersion function $h(x' - \phi(\lambda), x, y', y, \lambda)\mathrm{rect}(\frac{x}{\Delta'} - i', \frac{y}{\Delta'} - j')$ combined with the compressive coded mask $\mathbf{T}$ can be presented as a non-negative and binary matrix $\mathbf{H} \in \mathbb{R}^{ijkl \times ijl}$. Now we can rewrite the entire process of spectral dispersion and projection as the product of $\mathbf{H}$ and $\mathbf{f}$:

$$\mathbf{g} = \mathbf{H}\mathbf{f}. \tag{5}$$

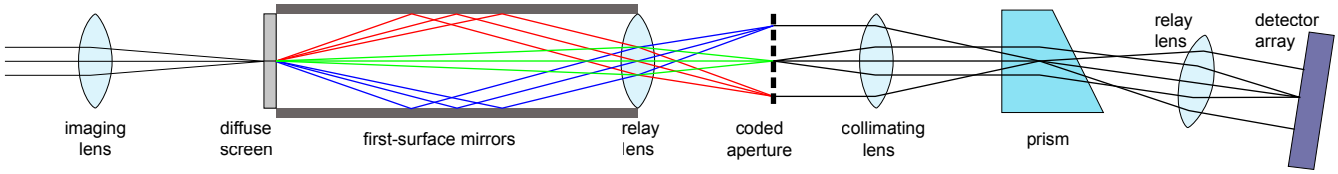

Figure 2: *Optical path in our hyperspectral imager. An objective lens (left) forms an image on a diffuse screen. This image is duplicated as an array of 3-by-3 images by four first-surface mirrors and projected to a coded aperture (middle). The duplicated images with different coded masks are collimated and transformed to a monochromatic sensor (right) via a prism.*

3.2. Spectral Reconstruction

State-of-the-art CASSI systems find an optimal solution to the ill-posed inverse problem of reconstructing input spectra by using an expectation maximization with total variation [GJB*07, WJWB08, KCWB10, WMAP11] or learning an over-complete dictionary via sparse representation [LWLD14, LLWD14, PMX*14, WXG*15b, SHG*16]. The foundation of our reconstruction workflow follows the traditional reconstruction of CASSI using a total variation (TV) regularizer. Our reconstruction problem of multisampling hyperspectral channels is to seek $\mathbf{f}$ that can minimize $\|\mathbf{g}-\mathbf{H}\mathbf{f}\|_2^2$. We formulate the reconstruction as a Lagrangian relaxation problem of a constrained optimization:

$$\min_{\mathbf{f}} \frac{1}{2}\|\mathbf{g}-\mathbf{H}\mathbf{f}\|_2^2 + \tau\Gamma(\mathbf{f}), \tag{6}$$

where τ is set to around 0.1 in our experiments. We solve this data term with the TV-L1 regularizer. We compute an isotropic L1 norm $\Gamma(\mathbf{f})$ of variation in the horizontal and the vertical axis of the spectral data cube [BDF07]:

$$\begin{aligned}\Gamma(\mathbf{f}) = \textstyle\sum_l \sum_{i,j} \{ &|\mathbf{f}(i+1,j,l)-\mathbf{f}(i,j,l)| \\ &+ |\mathbf{f}(i,j+1,l)-\mathbf{f}(i,j,l)|\}.\end{aligned} \tag{7}$$

Since the L1 norm is known to be robust, and it enforces the sparsity of the gradients of $\mathbf{f}$, we select the L1 norm over the L2 norm to find a robust and smooth solution. We sum the spatial variations of the horizontal and vertical axes within the spectral data cube. Note that while calculating the total variation, spatial smoothness is considered explicitly without spectral smoothness, thereby allowing steep gradient changes along the spectral axis.

4. System Implementation

This section describes technical details for implementing our system prototype of 4D hyperspectral video spectroscopy.

4.1. Kaleidoscopic Imaging in CASSI

Geometric Optics. The design of our system originates from a single disperser architecture with a coded aperture of CASSI proposed by Wagadarikar [WJWB08]. Figure 2 shows a schematic diagram of light transport in our system. A conventional objective lens forms an image on a diffuse screen, which is surrounded by four first-surface mirrors to duplicate the image. The duplicated images are then relayed to a coded aperture mask for spatial modulation. The compressed rays are collimated to be dispersed by a prism. Note that the collimating step should be placed prior to the dispersion to avoid inconsistent focusing among wavelengths. The last relay lens focuses the dispersed light on an image sensor. The captured light is the projection of the sheared spectrum of a scene (refer to Section 3.1). In summary, our novel optical design with image duplication allows for *multiple sampling* of randomness of the coded aperture to enhance spatial resolution with benefits of compressive coded aperture snapshot imaging for video spectroscopy.

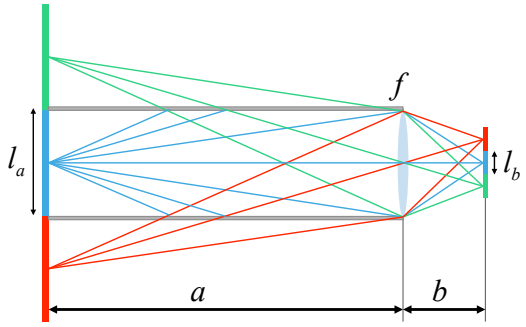

Figure 3: *The ratio of image minification is determined by the length of the mirrors a and focal length b.*

View Multiplication. Once an objective lens forms an image on a diffuse screen, four surrounding first-surface mirrors duplicate the image on a virtual plane at the same focusing distance parallel to the diffuser. The duplicated 3×3 views of the diffuser are then compressed to the size of the image sensor. The minification ratio of copied images $m = l_b/l_a$ is determined by the horizontal length of the mirrors a and the focal length f of the relay lens (see Figure 3):

$$m = \frac{b}{a} = \frac{1}{a/f-1} \quad \text{s.t.} \quad \frac{1}{a}+\frac{1}{b}=\frac{1}{f}.$$

For image multiplication, we built a four-sided kaleidoscope in dimensions of $24\times24\times240$ mm with four first-surface mirrors from Edmund Optics and a mirror holder created by a 3D printer.

Geometric Calibration. Misalignment of the mirrors introduces perspective distortion of copied images. To solve the misalignment problem, we performed geometric calibration on captured views, following Manakov et al. [MRK*13]. Capturing a checkerboard allows us to estimate homographies between duplicated views and the original view using corresponding points of the checkerboard. By warping the copied views with the estimated homographies, all views can be aligned with respect to center view coordinates. This geometric calibration is independent of the scene because optical distortion occurs after objective image formation. Note that the halves of the duplicated views, on the main diagonal and anti-diagonal excluding the center one, are duplicated from their horizontal neighbors, and the other halves come from the vertical

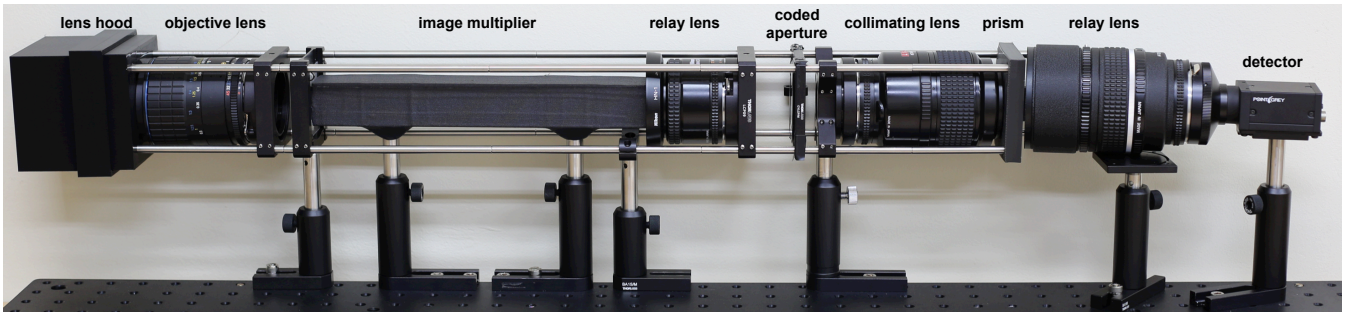

Figure 4: *Hardware prototype of our hyperspectral imager. For actual operation we fully encapsulate the entire system by a 3D-printed cover hood to block interfering light.*

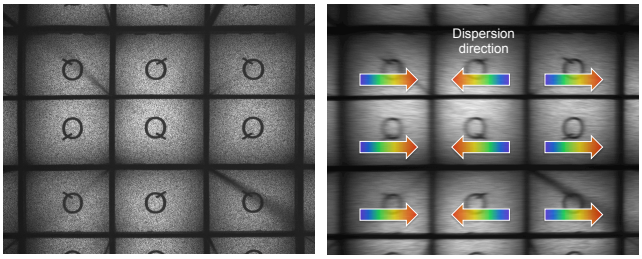

Figure 5: *(a) A multiview image of a printed letter at a wavelength of 550 nm. Neighboring views are flipped vertically and horizontally. (b) The dispersion direction of the first and the third columns is from the blue (left) to red light (right), which is opposite to the direction of the second column.*

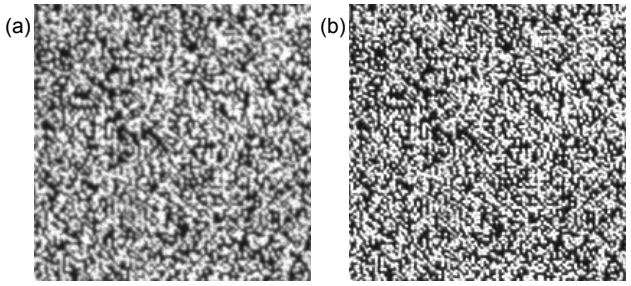

Figure 6: *Overcoming the diffraction limit. (a) presents the raw capture of the aperture, and (b) shows the result of the deconvolution method.*

neighbors. For instance, the top-left view is a mixture of the top-mid and mid-left views. Since these views suffer from black diagonal line artifacts, which originate from the gap between two orthogonal first-surface mirrors (see Figure 5a), we use five views for our experiments, the original center and its four directly neighboring views.

We adjust the minification ratio of mirrors to configure 3×3 views of duplication. Figure 5 shows flipped duplication by mirrors. These changes of view orientations are initially calibrated by calculating a homography transform per each view $\mathbf{g}_k$. This initial transformation registers duplicated views to the center view $\mathbf{g}_0$ in Equation (5). We then refine per-pixel registration in $\mathbf{g}_k$ by applying optical flow [Liu09]. Figure 5(a) presents a captured image of multiple views. Figure 5(b) shows direction changes of spectral dispersion in multiple views, following the kaleidoscopic imaging theory [RMSI11]. Our homography transformations include these direction changes of spectral dispersion.

Dispersion Directions. One of the insights of our optical design is that our mirror configuration diversifies the directions of dispersion of the coded aperture. The $\mathbf{H}_k$ matrices in Equation (5) are flipped along the vertical axis, as shown in Figure 5. This allows more robust reconstruction of the horizontal dispersion while solving the inverse problem.

4.2. Radiometric Calibration

The camera response function of our hyperspectral imager can be described as the linear product of the sensor's quantum efficiency $\mathbf{q}$ and the overall light efficiency $\mathbf{e}$ of the optical system. We then define the reconstructed signals $\mathbf{f}$ of each wavelength in Equation (6) as $\mathbf{f} = \mathbf{qel}$, where $\mathbf{l}$ is the radiance measured by the system. By determining a linear transformation $\mathbf{c} = (\mathbf{qe})^{-1}$, we can convert raw signals $\mathbf{f}$ to incident radiance $\mathbf{l}$. To derive this calibration model, we captured and measured 24 colors in a standard color target (X-rite ColorChecker) to regress model coefficients using least squares. The multiplication of $\mathbf{c}$ and $\mathbf{f}$ yields physically-meaningful radiance $\mathbf{l}$ from the reconstructed signals $\mathbf{f}$.

Once we obtain a hyperspectral radiance map, we convert it to sRGB color vectors to visualize visible spectral information as an sRGB color image. We first project spectra $\mathbf{l}$ to tristimulus values of CIEXYZ using the CIE color matching functions $\mathbf{M}_{XYZ}$ of 2-degree observation [CIE86] following [KRf*14]. We then transform the tristimulus values to sRGB color vectors $\mathbf{s}$ using the standard sRGB transform $\mathbf{M}_{sRGB}$ [NS98]: $\mathbf{s} = \mathbf{M}_{sRGB}\mathbf{M}_{XYZ}\mathbf{l}$. Finally, we apply a white balancing algorithm [Buc80] to $\mathbf{s}$ via gamma correction (γ=2.2) to obtain sRGB color images.

4.3. Technical Specifications

We implemented our design as the prototype shown in Figure 4. Our system consists of four lenses, four first-surface mirrors, a diffuse screen, a coded mask, a prism, and a monochromatic camera. This section provides technical specifications for hardware implementation.

Coded Aperture. A pixel in the coded aperture with random binary patterns corresponds to two-by-two pixels of the CCD camera, PointGrey Grasshopper 3 (9 megapixels with the resolution of

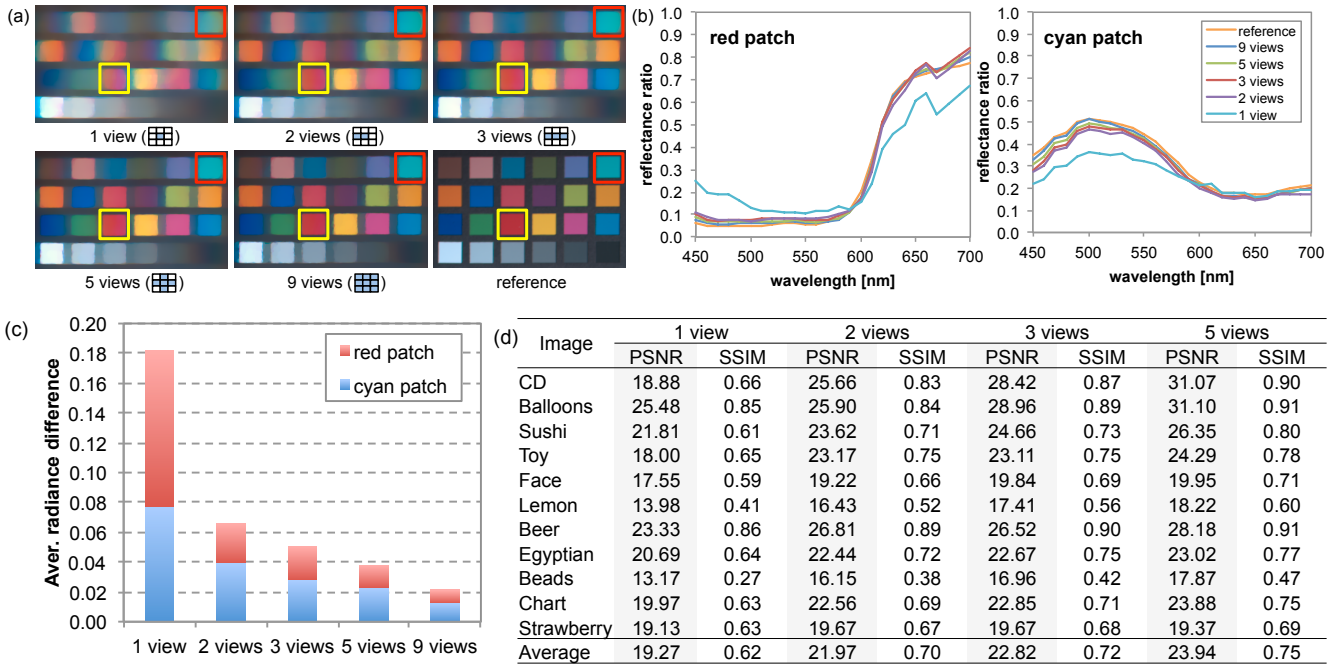

Image	1 view		2 views		3 views		5 views	
	PSNR	SSIM	PSNR	SSIM	PSNR	SSIM	PSNR	SSIM
CD	18.88	0.66	25.66	0.83	28.42	0.87	31.07	0.90
Balloons	25.48	0.85	25.90	0.84	28.96	0.89	31.10	0.91
Sushi	21.81	0.61	23.62	0.71	24.66	0.73	26.35	0.80
Toy	18.00	0.65	23.17	0.75	23.11	0.75	24.29	0.78
Face	17.55	0.59	19.22	0.66	19.84	0.69	19.95	0.71
Lemon	13.98	0.41	16.43	0.52	17.41	0.56	18.22	0.60
Beer	23.33	0.86	26.81	0.89	26.52	0.90	28.18	0.91
Egyptian	20.69	0.64	22.44	0.72	22.67	0.75	23.02	0.77
Beads	13.17	0.27	16.15	0.38	16.96	0.42	17.87	0.47
Chart	19.97	0.63	22.56	0.69	22.85	0.71	23.88	0.75
Strawberry	19.13	0.63	19.67	0.67	19.67	0.68	19.37	0.69
Average	19.27	0.62	21.97	0.70	22.82	0.72	23.94	0.75

Figure 7: *Comparison of hyperspectral images (a) and spectral measurements (b), synthetically reconstructed with different numbers of input views. The PNSRs of the reconstructed results from 1/2/3/5/9 views are 15.93/18.62/18.99/20.96/24.96. (b) compares the measurement accuracy of our method on the red and cyan patches by varying the number of input views. (c) compares the averaged radiance differences between the reference and the reconstruction of the red and cyan patches as the number of views increases. (d) compares the PSNRs and SSIMs between the reference [YMIN08] and the synthetically reconstructed hyperspectral images with different numbers of views. We used* $\tau = 0.15$*, and took 50 iterations for TV-L1 optimization.*

3376×2704). The spatial resolution of the mask patterns is critical to that of the output spectra. Although smaller features would yield a higher resolution, we also must consider the diffraction effect with smaller features. Owing to the diffraction limit of the system, the coded aperture is blurred, as shown in Figure 6(a). In order to overcome the diffraction effect, we apply the Richard-Lucy deconvolution method [Luc74] to the captured coded aperture mask image. This deconvolved image is used to build a matrix $\mathbf{H}$ in Equation (6). The radius of the Airy disk kernel R is calculated by

$$R = k\frac{L\lambda}{d}, \tag{8}$$

where L is the distance from the camera aperture to the sensor plane, λ is the wavelength of light and d is the diameter of the camera aperture, and the constant $k = 1.22$ accounts for the shape of our circular aperture. The estimated radius of the diffraction pattern in visible spectra is $\sim$6.10 μm, which is larger than the sensor pitch $\sim$3.69 μm. The PSF radius in our system is $\sim$1.65 pixel. The overall transmittance of this coded mask is $\sim$50% as 50% of the mask area is occluded by chrome mask on quartz, where the pixel pitch of the coded mask is 7.40 μm.

Spectral Dispersion. We mount a UV-IR cut filter to select incident spectral energy, wavelengths ranging between 450nm and 700nm. A solid-state plasma light source (Thorlabs HPLS-30-04) was used as a light source. Suppose we target reconstruction of about 30 wavelength channels of 10 nm intervals from a 2:1 ratio of the mask and the sensor pixel sizes in our system. The pixel range of spectral dispersion must span at least 60 pixels for visible spectra from 450 nm to 700 nm. We choose a prism made of BK-7. The refractive index of the material is 1.5168. To fit our measurement goal, we fabricate the apex angle of the prism as 17 °.

Diffuse Screen. We installed a diffuse screen at the focal length of the objective imaging lens to duplicate an image with four neighboring mirrors. We tested three types of diffusers: a ground glass and an opal and a holographic diffuser. The ground glass diffuser degrades image quality by grain and the opal diffuser shows very low transmittance. We therefore decided to use a holographic diffuser, the transmittance of which is specified as $\sim$85% (Edmund Optics holographic diffuser #55-440).

Objective and Relay Lenses. We install a Coastal Optics 60 mm f/4 UV-VIS-IR lens as an objective lens that is apochromatic from approximately 315 nm to 1.1 μm. We employ three Nikon lenses for optics relaying and imaging. In particular, the second lens from the left serves as an imaging lens, and the third one functions as a collimating lens by configuring the focal distance to infinity in these lenses. The ratio of the focal lengths of these lenses determines the zoom factor of our imaging system. We use the lenses of the same focal length of 50 mm to preserve 1:1 imaging.

5. Results

In this section, we validate our multisampling compressive video spectroscopy by presenting a series of quantitative and qualitative

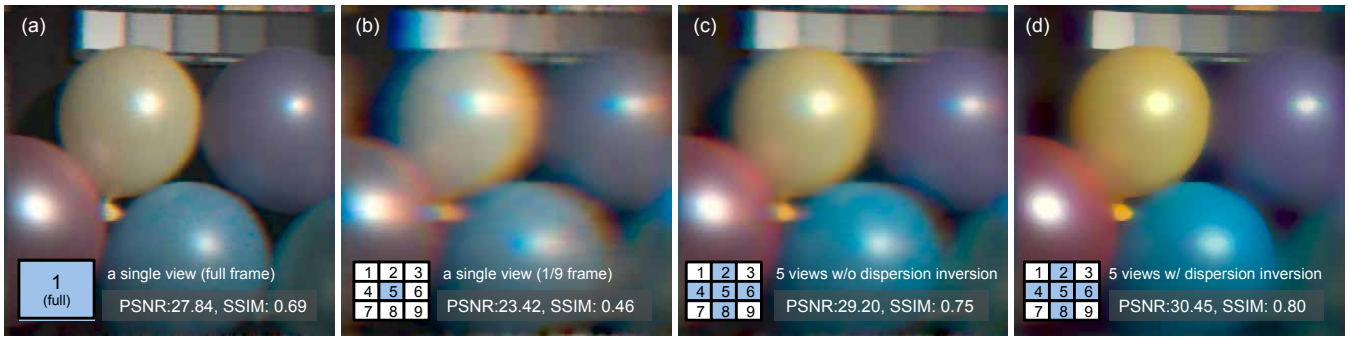

Figure 8: *(a) The traditional CASSI system [WJWB08] synthesized with a full-frame input (1935×1503), which has a nine times larger resolution than that of a view (645×501) in our multiview configuration. (b) a reconstruction result using a single view of the multiview configuration. (c) and (d) compare spectral reconstruction results using five different views that we use with our prototype. (c) shows a reconstruction result without dispersion inversion (see Figure 3b), while (d) presents a reconstruction results with dispersion inversion. The five-view reconstruction with dispersion inversion improves not only PSNR but also SSIM of spectral channels.*

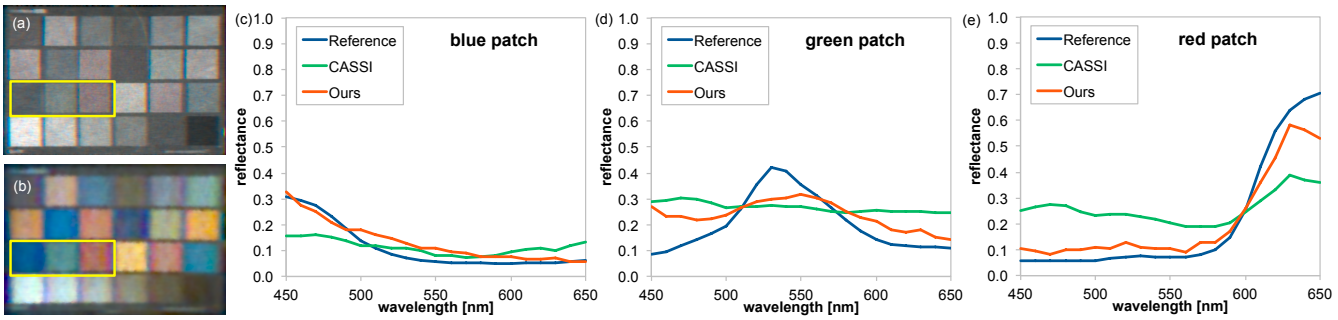

Figure 9: *We compare the reflectances of color patches in a ColorChecker target measured by a spectroradiometer (Jeti Specbos 1200), a traditional CASSI system [WJWB08] and our system. (a) shows a hyperspectral image of the target captured by the traditional CASSI system. (b) shows a hyperspectral image captured by our system. (c), (d) and (e) compare the reconstructed spectral reflectance of the blue, the green and the red patch in the target using three instruments. Even though our system's spatial resolution is lower than the traditional CASSI system, the accuracy of spectral reconstruction is significantly improved in our system.*

analysis of experiment results. The experiments include evaluations via synthetic hyperspectral images and via real images captured by our imager. We reconstructed hyperspectral video footage using non-optimized Matlab codes. The reconstruction process for one frame (645×501) took approximately 200 seconds by a machine with an Intel i7-3770 CPU 3.4 GHz with 32 GB memory. Refer to the supplemental video for video results.

Reconstruction Accuracy. Here we validate the effectiveness of our multisampling design by simulating the reconstruction process, described in Section 3, with hyperspectral image datasets of the real world [YMIN08, CZ11]. The reference images are scaled to the same size as the employed sensor resolution. The active sensing area is segmented to nine views. We compare reconstruction results from five different multisampling configurations of 1/2/3/5/9 views with the reference hyperspectral image in terms of the peak signal-to-noise ratio (PSNR) and structural similarity (SSIM) [WBSS04].

Figure 7 validates that our multisampling design improves not only the spatial resolution (a) but also the accuracy of spectral measurements (b), as we increase input randomness. Figure 7(b) compares the differences between the reference measured by a calibrated spectroradiometer (Jeti Specbos 1200) and the reconstructed spectral reflectance over the red and the cyan patch, varying the number of input views. Figure 7(c) shows the averaged differences between the reference and the reconstructed radiance over the red and the blue patch. This experiment validates that as the number of input frames increases, the accuracy of the reconstructed hyperspectral images increases consistently. We also test the performance of our multisampling approach with 11 reference hyperspectral images [YMIN08] by comparing the reference and the reconstructed images. See Figure 7(d). The averaged PSNRs and SSIMs of reconstructed hyperspectral channels are improved consistently as the number of input views increases.

Multiview Tradeoff. The proposed method produces multiviews on a camera sensor by segmenting the sensing area into nine windows of 3×3 views. This means that we utilize nine times more samples than a single-frame CASSI in theory. On the other hand, this design reduces the spatial resolution of the reconstructed image (645×501) by $1/9^{th}$, compared to the traditional CASSI [WJWB08] (1935×1503) with the same hardware resources. Now we compare the two different design configurations under the same resources.

Figure 8(a) shows a hyperspectral image, synthetically re-

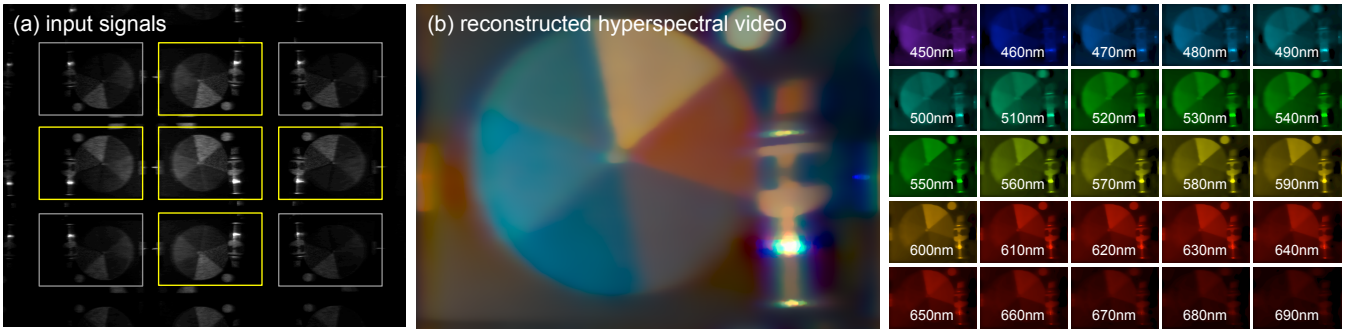

Figure 10: *We captured hyperspectral video footage using our prototype. (a) shows the raw input of a single frame, where the yellow rectangles indicate the subviews that we used for reconstruction. (b) presents the sRGB color visualization of the reconstructed spectral channels from 450 nm to 700 nm in 10 nm intervals. The right-most image array presents the spectral power distributions of each wavelength in the video footage. Refer to the supplemental video for more results.*

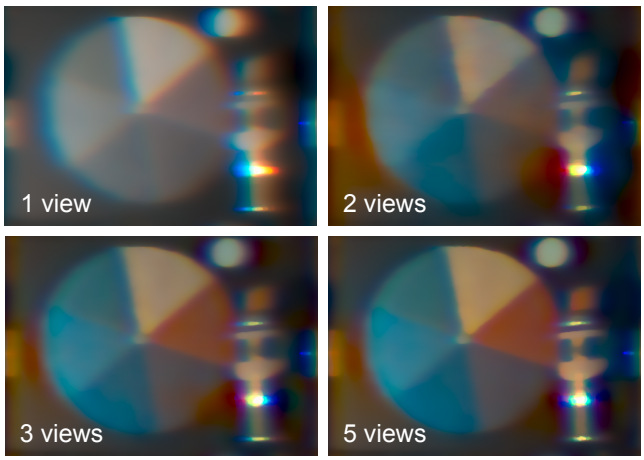

Figure 11: *Reconstructed hyperspectral video footage with 1/2/3/5 multiviews, captured by our prototype.*

constructed from a single image input of a traditional CASSI method [WJWB08]. The image resolution in the tradition CASSI is nine times higher than that of our nine-view configuration, while the remaining optical configurations such as the dispersion power are the same. Figure 8(b) presents a reconstruction result using a single subview among nine duplications in our method. Since the resolution of the input image is reduced by nine times compared to the original, it is not surprising that our reconstruction using a single view is worse than the full-frame reconstruction.

As shown in Figure 3(b), the spectral dispersion of these columns is inverted due to the design of kaleidoscopic imaging after image registration. We conduct a simulation to compare this effect of *dispersion inversion* in reconstructing the spectrum. Figures 8(c) and (d) compare results reconstructed from the five subviews without and with *dispersion inversion*. It is worth noting that although the number of input subviews is the same, the spatial resolution with dispersion inversion increases relatively in terms of the PSNR and SSIM of reconstructed hyperspectral channels. The configuration presented in Figure 8(d) is the one that we finally chose for our system implementation.

We implemented the traditional CASSI method [WJWB08] and the optical design described in Section 3.1 as our system prototype (Section 4). Note that we implemented both systems using the same hardware configuration just except the kaleidoscope unit. The traditional CASSI system captures images with nine times higher resolution than our proposed imaging system of nine subviews. Figure 3 shows a photograph of the system prototype. Figures 9(a) and (b) compare the hyperspectral images captured by the two imaging systems. It is not surprising that the traditional CASSI can provide a higher spatial resolution than the proposed method. However, comparing the reconstructed spectral resolutions from both systems, the spectral reflectance captured by our multisampling imager is significantly more accurate than the traditional CASSI system with the same hardware resource. Figures 9(c), (d) and (e) compare spectral reflectances of the blue, the green and the red patch, measured by both systems, with respect to reference measurements (Jeti Specbos 1200 spectroradiometer).

Figure 10 presents a hyperspectral video footage of a moving object of a color wheel. Refer to the supplemental video for more results. Figure 10(a) shows an example of a raw input frame for hyperspectral video captured by the system prototype (10 FPS). Among nine multiviews we choose five clean views as input for spectral reconstruction due to diagonal artifacts. The yellow rectangles indicate the used frames. We then reconstruct them as a frame of the hyperspectral video footage, which contains per-pixel spectra in the 4D hyperspectral video footage (x, y, λ, t). Figure 10(b) presents visible color visualization converted from reconstructed hyperspectral data. The right-hand-side array of photographs shows individual spectral channels. Figure 11 compares the quality of the reconstructed videos with different numbers of multiviews. We quantitatively evaluated the measurement accuracy by comparing our image-based measurements with spectroradiometer measurements. For instance, the measurement differences of the red sector in Figure 11 between our system and the spectroradiometer decreases by 21, 19, 18, and 10 in ΔE^*_{00} gradually as we increase the number of input views. Figure 12 compares more hyperspectral video footage captured by our prototype. Refer to the supplemental video for more results.

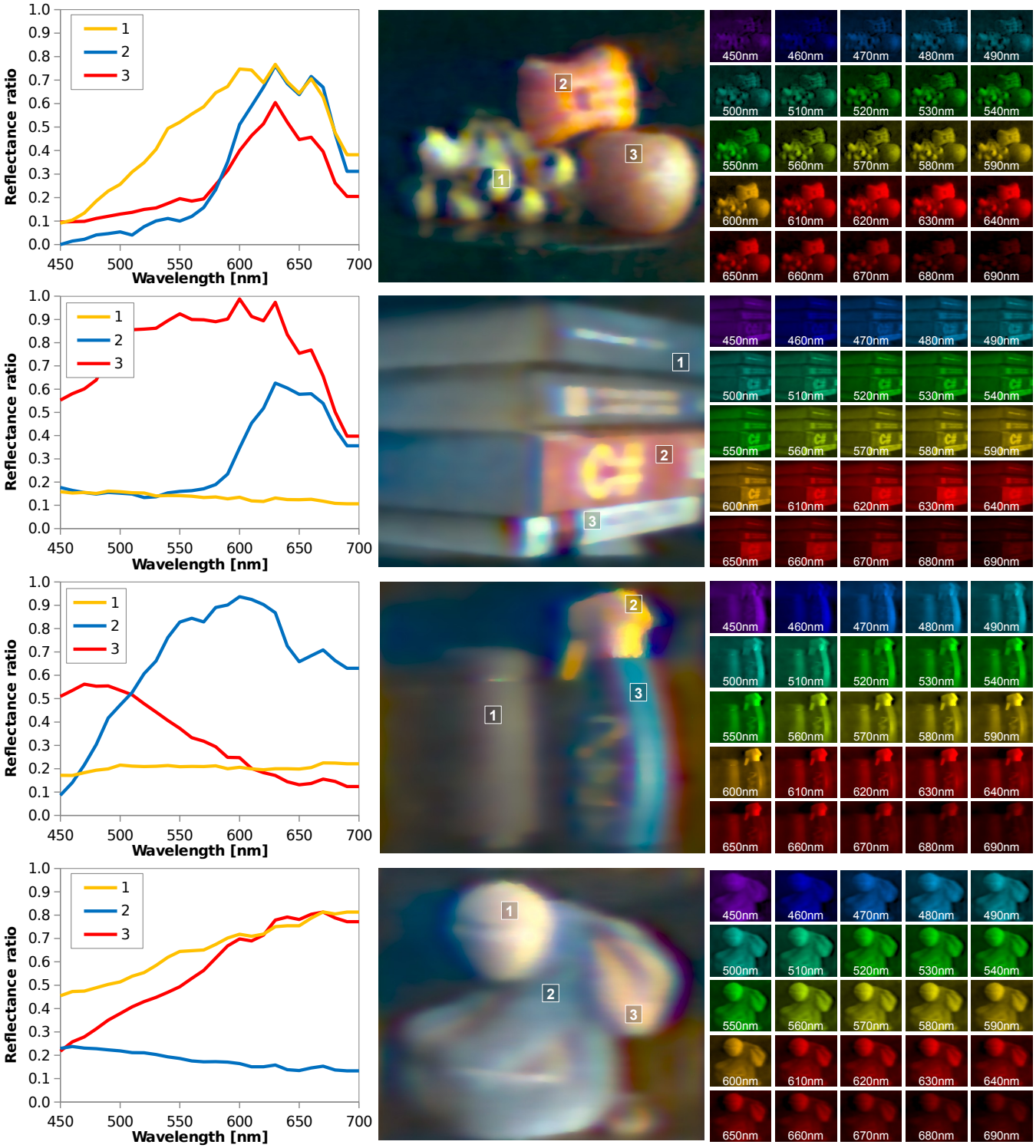

Figure 12: *We captured hyperspectral video footage using our system prototype. This figure presents measured reflectances, sRGB color images and spectral channels of moving objects. Refer to the supplemental video for more results.*

6. Discussion

We discuss a range of observations made throughout the development of our system and the acquisition of a variety of test scenes.

Tradeoff between Spatial and Spectral Resolution. Filter-based approaches [MRK*13] provide high-spatial resolution, but their spectral resolution is limited. Compressive imaging approaches [LWLD14,LLWD14] including the proposed method provide high-spectral resolution, but their spatial resolution is limited. This is a long-lasting tradeoff between spatial and spectral resolution in hyperspectral imaging. Many approaches to solving this tradeoff [KCWB10, WMAP11] by combining multiple sampling and compressive imaging have been proposed. Compressive imaging principles stand on sparsity, randomness, and convex optimization. The solution of this problem that we propose is multiple sampling of the randomness of sparse modulation of the spectrum using kaleidoscopic imaging. The proposed method accommodates multiple sampling and video imaging. We demonstrate increased sampling of randomness is beneficial to enhance spectral resolution significantly, despite that we utilize the entire sensor resolution partially as segments via a kaleidoscope.

Randomness of the Coded Mask. Regarding compressive sensing, it is well documented that randomness leads to an effective sensing mechanism in reconstructing sparsely sampled signals. We create random patterns by lithographically etching a quartz substrate with an active area of 25 mm square, where a pixel on the mask corresponds to 2-by-2 pixels on a sensor. The key insight is that the randomness provides incoherence in sampling sparse signals, which is crucial for sparsity-based optimization such as total-variation or sparse coding. In the compressive sensing field, random sensing is practiced as a near-optimal strategy.

Hardware Assembly. As shown in our experimental results, the spatial resolution captured by our prototype is noticeably lower than the synthetic results. We found that building the prototype requires sophisticated alignment of optical components including the coded aperture, the four mirrors, the prism, and the camera along the optical path. The diffraction effect through the coded mask and lens distortion, which is not present in the synthetic experiments, hindered sharp reconstruction of captured spectra. We notice that the misalignment of copied views gives rise to a critical reconstruction problem. Although the misalignment can be solved by image registration, the dispersion directions of each view are different due to misaligned orientation of views. Distortion of dispersion directions breaks the transformation matrix and corrupts the hyperspectral image reconstruction. Also, focusing on the coded mask contributes to the quality of reconstructed images. Since fine focusing requires perfect lenses and high precision positioning, we obtain a slightly blurred image of the coded mask.

Implementation of Kaleidoscope. Designing a system to use nine views originally, we found that even small corner seams near four edges of the mirrors hindered reconstruction of sparse information, as shown in the diagonal views in Figure 5(a). Owing to this issue in fabricating the mirror holder, we omitted these corner views from multiple sampling in the system. However, a more elaborately fabricated mirror holder would solve this problem.

Light Efficiency. In traditional CASSI systems, the coded mask occludes 50% of incident light to yield spatial modulation. In state-of-the-art imaging systems with multiple cameras [WXG*15a, WXG*15b], a beam splitter is necessary to divide incident light into two directions with a further 50% loss, i.e., this configuration leads to a combined 75% loss of incident energy on the camera equipped with the coded aperture. In contrast, our mirror-based multisampling approach enhances spectral resolution with a cost of only 15% loss of incident energy by the diffuser.

TV-L1 Optimization vs. Sparse Coding. Traditional CASSI reconstruction methods [GJB*07, WJWB08, KCWB10, WMAP11] formulate the reconstruction problem as an inverse problem that minimizes $\|\mathbf{g}-\mathbf{Hf}\|_2^2$ with TV-L1 including our method (refer to Equation (6)). Recent reconstruction methods using sparse representation such as [LWLD14,LLWD14,PMX*14,WXG*15b] formulate this problem as minimizing $\|\mathbf{g}-\Phi\mathbf{D}\alpha\|_2^2$, where Φ is a projection matrix, $\mathbf{D}$ is a 3D spatio-spectral dictionary, and α is a corresponding sparse code vector. This over-complete dictionary $\mathbf{D}$ is learned from a large set of 3D spatio-spectral patches. While the former methods solve the optimization problem with a *global* spatio-spectral matrix $\mathbf{H}$, the latter approaches solve the *local* optimization problems with a large set of independent spatio-spectral dictionaries. These sparse coding-based approaches require many hours for training and even reconstruction; e.g., it takes 25 hours to reconstruct an image of 374×502 pixels by a state-of-the-art method [LLWD14]. Since we are targeting video applications of spectroscopy, we were motivated to choose the traditional reconstruction approach with consideration of computational cost. We tested five methods: TWIST [BDF07], GPSR [FNW07], NeARest [SP08], SpaRSA [WNF09], and a sparse coding approach [LWLD14]. We chose TWIST because it is the most efficient and accurate. In future work, we would like to apply a sparse coding-based approach to solving the optimization problem.

7. Conclusions

We have presented a novel camera system for compressive imaging to measure hyperspectral video. We make a tradeoff between multisampling (beneficial for spatial and spectral resolution, which hinders video acquisition) and snapshot-based design (beneficial for hyperspectral video acquisition, which suffers from low spatial resolution) by combining a coded aperture and a kaleidoscope to achieve high spatial and spectral resolution in hyperspectral videos. Specifically, we provide insights for coupling multisampling and compressive imaging, offering physically-meaningful acquisition of hyperspectral video. We validated the effectiveness and consistency of our system qualitatively and quantitatively. Finally, we have provided a range of building experience and potential directions for compressive video spectroscopy.

Acknowledgments

Min H. Kim, the corresponding author, gratefully acknowledges Korea NRF grants (2013R1A1A1010165 and 2013-M3A6A6073718) and additional support by an ICT R&D program of MSIP/IITP (10041313).

References

[BDF07] BIOUCAS-DIAS J., FIGUEIREDO M.: A new twist: two-step iterative shrinkage/thresholding for image restoration. *IEEE TIP 16*, 12 (2007), 2992–3004. 4, 10

[Bra09] BRADY D. J.: *Optical Imaging and Spectroscopy*. Wiley-OSA, 2009. 2

[Buc80] BUCHSBAUM G.: A Spatial Processor Model for Object Colour Perception. *J. the Franklin Institute 310*, 1 (1980), 1–26. 5

[CDT*11] CAO X., DU H., TONG X., DAI Q., LIN S.: A prism-mask system for multispectral video acquisition. *IEEE TPAMI 33*, 12 (2011), 2423–2435. 1

[CIE86] CIE: *Colorimetry*. CIE Pub. 15.2, Commission Internationale de l'Eclairage (CIE), Vienna, 1986. 5

[CTDL11] CAO X., TONG X., DAI Q., LIN S.: High resolution multispectral video capture with a hybrid camera system. In *CVPR* (2011). 2

[CZ11] CHAKRABARTI A., ZICKLER T.: Statistics of Real-World Hyperspectral Images. In *CVPR* (2011). 7

[DTCL09] DU H., TONG X., CAO X., LIN S.: A prism-based system for multispectral video acquisition. In *ICCV* (2009). 1, 2

[FNW07] FIGUEIREDO M., NOWAK R., WRIGHT S.: Gradient projection for sparse reconstruction: Application to compressed sensing and other inverse problems. *IEEE J-STSP 1*, 4 (2007), 586–597. 10

[GJB*07] GEHM M. E., JOHN R., BRADY D. J., WILLETT R. M., SCHULZ T. J.: Single-shot compressive spectral imaging with a dual-disperser architecture. *OSA OE 15*, 21 (2007), 14013–27. 2, 4, 10

[HF13] HOYE G., FRIDMAN A.: Mixel camera – a new push-broom camera concept for high spatial resolution keystone-free hyperspectral imaging. *OSA OE 21*, 9 (2013), 11057–11077. 1, 2

[HKW12] HABEL R., KUDENOV M., WIMMER M.: Practical spectral photography. *Wiley CGF 31* (2012), 449–458. 2

[HP03] HAN J. Y., PERLIN K.: Measuring bidirectional texture reflectance with a kaleidoscope. *ACM TOG 22*, 3 (2003), 741–748. 2

[KCWB10] KITTLE D., CHOI K., WAGADARIKAR A., BRADY D. J.: Multiframe image estimation for coded aperture snapshot spectral imagers. *OSA AO 49*, 36 (2010), 6824–33. 1, 2, 4, 10

[KHK*12] KIM M. H., HARVEY T. A., KITTLE D. S., RUSHMEIER H., DORSEY J., PRUM R. O., BRADY D. J.: 3d imaging spectroscopy for measuring hyperspectral patterns on solid objects. *ACM TOG 31*, 4 (2012), 38:1–11. 2

[KRf*14] KIM M. H., RUSHMEIER H., FFRENCH J., PASSERI I., TIDMARSH D.: Hyper3d: 3d graphics software for examining cultural artifacts. *ACM JOCCH 7*, 3 (2014), 1:1–19. 5

[KWT*11] KAWAKAMI R., WRIGHT J., TAI Y.-W., MATSUSHITA Y., BEN-EZRA M., IKEUCHI K.: High-resolution hyperspectral imaging via matrix factorization. In *CVPR* (2011), pp. 2329–2336. 1, 2

[Liu09] LIU C.: *Beyond Pixels: Exploring New Representations and Applications for Motion Analysis*. Ph.D. Thesis, Massachusetts Institute of Technology, 2009. 5

[LK14] LEE H., KIM M. H.: Building a two-way hyperspectral imaging system with liquid crystal tunable filters. In *ICISP* (2014), pp. 26–34. 1, 2

[LLWD14] LIN X., LIU Y., WU J., DAI Q.: Spatial-spectral encoded compressive hyperspectral imaging. *ACM TOG 33*, 6 (2014). 4, 10

[Luc74] LUCY L. B.: An iterative technique for the rectification of observed distributions. *The astronomical journal 79* (1974), 745. 6

[LWLD14] LIN X., WETZSTEIN G., LIU Y., DAI Q.: Dual-Coded Compressive Hyper-Spectral Imaging. *OSA OL 39* (2014), 2044–2047. 4, 10

[MCWD14] MA C., CAO X., WU R., DAI Q.: Content-adaptive high-resolution hyperspectral video acquisition with a hybrid camera system. *Optics letters 39*, 4 (2014), 937–940. 2

[MRK*13] MANAKOV A., RESTREPO J. F., KLEHM O., HEGEDÜS R., EISEMANN E., SEIDEL H.-P., IHRKE I.: A reconfigurable camera add-on for high dynamic range, multispectral, polarization, and light-field imaging. *ACM TOG 32*, 4 (2013), 47. 1, 2, 4, 10

[NK14] NAM G., KIM M. H.: Multispectral photometric stereo for acquiring high-fidelity surface normals. *IEEE CGA 34*, 6 (2014), 57–68. 2

[NS98] NIELSEN M., STOKES M.: The creation of the sRGB ICC Profile. In *CIC* (1998), IS&T. 5

[PMX*14] PENG Y., MENG D., XU Z., GAO C., YANG Y., ZHANG B.: Decomposable nonlocal tensor dictionary learning for multispectral image denoising. In *CVPR* (2014). 4, 10

[RAA15] RUEDA H., ARGUELLO H., ARCE G. R.: DMD-based implementation of patterned optical filter arrays for compressive spectral imaging. *OSA JOSAA 32*, 1 (2015), 80–89. 1

[RB05] RAPANTZIKOS K., BALAS C.: Hyperspectral imaging: potential in non-destructive analysis of palimpsests. In *ICIP* (2005), vol. 2. 2

[RKT*13] RAJWADE A., KITTLE D., TSAI T.-H., BRADY D., CARIN L.: Coded hyperspectral imaging and blind compressive sensing. *SIAM Journal on Imaging Sciences 6*, 2 (2013), 782–812. 2

[RMB*13] RESHETOUSKI I., MANAKOV A., BHANDARI A., RASKAR R., SEIDEL H.-P., IHRKE I.: Discovering the structure of a planar mirror system from multiple observations of a single point. In *CVPR* (2013). 2

[RMSI11] RESHETOUSKI I., MANAKOV A., SEIDEL H.-P., IHRKE I.: Three-dimensional kaleidoscopic imaging. In *CVPR* (2011). 2, 5

[SHG*16] SERRANO A., HEIDE F., GUTIERREZ D., WETZSTEIN G., MASIA B.: Convolutional Sparse Coding for High Dynamic Range Imaging. *Wiley CGF 35*, 2 (2016). 4

[SP08] SUN X., PITSIANIS N.: Solving non-negative linear inverse problems with the NeAREst method. *Proc. SPIE 7074* (2008), 707402. 10

[TB13] TSAI T.-H., BRADY D. J.: Coded aperture snapshot spectral polarization imaging. *OSA AO 52*, 10 (2013), 2153–2161. 1

[WBSS04] WANG Z., BOVIK A. C., SHEIKH H. R., SIMONCELLI E. P.: Image quality assessment: from error visibility to structural similarity. *IEEE TIP 13*, 4 (2004), 600–612. 7

[WGSN13] WANG L., GAO D., SHI G., NIU Y.: Double-channel compressive spectral imaging via complementary code patterns. In *ICSPCC* (2013). 1

[WJWB08] WAGADARIKAR A., JOHN R., WILLETT R., BRADY D.: Single disperser design for coded aperture snapshot spectral imaging. *OSA AO 47*, 10 (2008), B44–B51. 1, 4, 7, 8, 10

[WMAP11] WU Y., MIRZA I. O., ARCE G. R., PRATHER D. W.: Development of a digital-micromirror-device-based multishot snapshot spectral imaging system. *OSA OL 36*, 14 (2011), 2692–4. 2, 4, 10

[WNF09] WRIGHT S., NOWAK R., FIGUEIREDO M.: Sparse reconstruction by separable approximation. *IEEE TSP 57*, 7 (2009), 2479–2493. 10

[WPSB08] WAGADARIKAR A. A., PITSIANIS N. P., SUN X., BRADY D. J.: Spectral image estimation for coded aperture snapshot spectral imagers. vol. 7076, pp. 707602–707602–15. 2

[WPSB09] WAGADARIKAR A. A., PITSIANIS N. P., SUN X., BRADY D. J.: Video rate spectral imaging using a coded aperture snapshot spectral imager. *OSA OE 17*, 8 (2009), 6368–88. 2

[WXG*15a] WANG L., XIONG Z., GAO D., SHI G., WU F.: Dual-camera design for coded aperture snapshot spectral imaging. *OSA AO 54*, 4 (2015), 848–858. 10

[WXG*15b] WANG L., XIONG Z., GAO D., SHI G., ZENG W., WU F.: High-speed hyperspectral video acquisition with a dual-camera architecture. In *CVPR* (2015). 1, 2, 4, 10

[YMIN08] YASUMA F., MITSUNAGA T., ISO D., NAYAR S.: *Generalized Assorted Pixel Camera: Post-Capture Control of Resolution, Dynamic Range and Spectrum*. Tech. rep., Nov 2008. 6, 7