High Resolution Multispectral Video Capture with a Hybrid Camera System

Xun Cao[§]   Xin Tong[‡]   Qionghai Dai[§]   Stephen Lin[‡]

[§]Tsinghua University   [‡]Microsoft Research Asia

Abstract

We present a new approach to capture video at high spatial and spectral resolutions using a hybrid camera system. Composed of an RGB video camera, a grayscale video camera and several optical elements, the hybrid camera system simultaneously records two video streams: an RGB video with high spatial resolution, and a multispectral video with low spatial resolution. After registration of the two video streams, our system propagates the multispectral information into the RGB video to produce a video with both high spectral and spatial resolution. This propagation between videos is guided by color similarity of pixels in the spectral domain, proximity in the spatial domain, and the consistent color of each scene point in the temporal domain. The propagation algorithm is designed for rapid computation to allow real-time video generation at the original frame rate, and can thus facilitate real-time video analysis tasks such as tracking and surveillance. Hardware implementation details and design tradeoffs are discussed.

We evaluate the proposed system using both simulations with ground truth data and on real-world scenes. The utility of this high resolution multispectral video data is demonstrated in dynamic white balance adjustment and tracking.

1. Introduction

Within the human eye, light from the real world is sensed in color. The human eye perceives color from three types of cone cells sensitive to different parts of the light spectrum, namely those corresponding to red, green, and blue (RGB). Conventional cameras record RGB measurements that adequately replicate colors for human viewing, but in fact the spectrum of visible light may contain a profusion of detail that is lost in the coarse three-channel sensing of RGB. These spectral details can reveal much about the objects and lighting in the scene, and computer vision algorithms may thus have much to gain by imaging tens or hundreds of color samples over the light spectrum of each scene point.

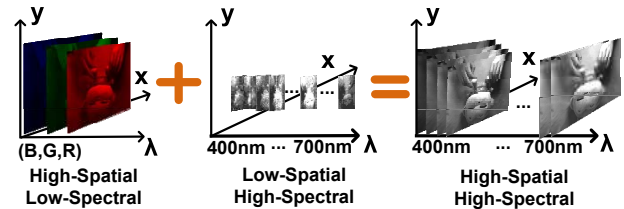Multispectral image capture has drawn much attention in the past several years. For imaging of static scenes,

Figure 1. Basic idea of multispectral video capture using the hybrid camera system. We integrate the high-spatial, low-spectral resolution frames from the RGB camera and the high-spectral, low-spatial resolution frames from the grayscale camera to generate videos with both high-spectral and high-spatial resolution. Temporal consistency in scene point color is also used, but not illustrated in this figure.

systems have been designed to record high spectral resolution at the expense of acquisition time. Spectrometers [11, 21] employ dispersive optical elements that require scanning over the scene to capture a full image. Other methods utilize a sequence of bandpass color filters in front of a camera to measure different spectral components [17, 24]. While both of these approaches provide direct measurements of scene spectra, they are unsuitable for multispectral imaging at video rates. To record dynamic scenes, reconstruction-based imaging techniques using computed tomographic imaging spectrometry (CTIS) [6, 13] and coded apertures [2, 23] have been presented. These methods allow for multispectral imaging at video rates, but the considerable post-processing needed to reconstruct multispectral videos makes them unsuitable for real-time vision applications such as tracking and surveillance. Recently, a system based on a prism and occlusion mask was proposed for direct, real-time capture of multispectral video [7]. However, due to the sacrifice of spatial resolution for additional spectral resolution, the image frames contain little spatial detail, which may limit the ability to perform image analysis. Video capture with *real-time multispectral output and both high spatial and spectral resolution* remains an open problem.

In this paper, we propose a solution to this problem based on a hybrid camera system and a method to integrate the data from the two cameras as illustrated in Figure 1. The imaging hardware, shown in Figure 2, consists of a regu-

lar RGB video camera, a grayscale video camera and several off-the-shelf optical elements including a prism, a beam splitter, an occlusion mask and a mirror. Incoming light from the scene is first equally divided towards two directions by the beam splitter. In one direction, the light passes through the occlusion mask and prism to form dispersed spectra on the grayscale sensor, similar to the system in [7]. While high spectral resolution is obtained in this manner, it is obtained with a significant loss in spatial resolution. Light in the other direction is reflected by the mirror and captured by the RGB camera, which has low spectral resolution but high spatial resolution. Frames from the two cameras are aligned so that each scene point measured by the multispectral camera has a known corresponding pixel in the RGB camera. Captured multispectral data is then transferred to corresponding pixels in the RGB frames, and is propagated to other pixels according to spectral similarity, spatial proximity, and temporal color consistency of scene points tracked by optical flow [3]. This propagation is implemented using a bilateral filter [14], whose fast implementation enables multispectral video generation in real-time.

Unlike systems that tradeoff either temporal or spatial resolution for additional spectral information, the proposed approach does not require such sacrifices while maintaining high spectral accuracy. Different from reconstruction based systems, our capture device can be constructed from widely available components and is much simpler to calibrate in practice. The effectiveness of this system is demonstrated with experiments on different computer vision applications including dynamic white balance adjustment and tracking.

2. Related Work

Depending on the target application, most existing methods for multispectral imaging forgo high spatial and/or temporal resolution to increase spectral sampling. For example, the spectrometer in [11] obtains a very high spectral resolution of $0.1nm$ but measures only a single pixel at a time. Methods that employ multiple color filters, such as a rotating filter wheel in front of the camera [24] or different filters distributed over the sensor [17], require multiple exposures to record different parts of the spectrum at each pixel. While such methods are effective for static scenes, they lack the efficiency needed to capture video of dynamic scenes.

Instead of directly measuring the spectral data of each scene point, CTIS systems [5, 6, 12] and coded aperture methods [2, 22, 23] treat multispectral imaging as a reconstruction problem. They regard the two-dimensional spatial information plus one spectral dimension as a 3D datacube, and reconstruct the 3D datacube from a set of 2D projections. CTIS systems utilize 2D projections that integrate spectral signals from different scene positions on the detector. The multiple projections are recorded on the sensor in a single snapshot, which gives this method the potential

to be used for multispectral video acquisition. However, CTIS systems need custom-made optical elements and are sensitive in practice to calibration noise. Only simulations and snapshots of very simple scenes have been reported in the literature [10]. Coded aperture methods take multiplexed 2D projections of the 3D datacube using specially designed apertures, namely the coded aperture snapshot imager (CASSI). The key idea of CASSI is to use apertures with certain patterns to code and decode the optical field. Recently, Wagadarikar *et al*. presented a video rate CASSI system [23]. A video of a lit candle was demonstrated, though with some reported reconstruction error. Although CTIS and CASSI systems have high spectral and temporal resolution, the spatial resolution is relatively low. Also, a time-consuming reconstruction step is necessary for both CTIS and CASSI, which makes them unsuitable for use in real-time video processing applications.

Other types of systems have also been presented for multispectral imaging with high temporal resolution. A device based on optical fiber bundles and a diffraction grating was proposed in [8] for real-time video capture, but provides only a 14×14 image resolution due to practical limitations on minimum fiber thickness. Park *et al*. developed a system that captures six-channel video at $30fps$ by taking advantage of multiplexed coded illumination [15], but is limited by the need for controlled lighting. In [7], a prism-based system for multispectral video acquisition was demonstrated with relatively low spatial resolution. Our acquisition system differs from previous methods in that it achieves high resolution in the spectral, spatial and temporal domains, with real-time video generation.

3. System Overview

In this section, we first introduce the basic principles and configuration of the proposed hybrid camera system. The entire multispectral capture pipeline is then described including the propagation algorithm and its temporal enhancement. Finally, system distortion and calibration issues are discussed together with several design tradeoffs.

3.1. System Configuration

We illustrate the hybrid camera system in Figure 2(a). For ease of understanding, the devices and optical path are shown as a 2D slice of the actual 3D system. Incoming light from a scene point first reaches the beam splitter, which reflects half of the light along the blue path, while transmitting the remainder along the yellow path. The reflected light is again reflected at a mirror before arriving at the RGB camera. An alternative configuration for this light path is to remove the mirror and have the RGB camera directly face the beam splitter, but this results in a less compact system. The light on this path is measured in RGB at a high spatial resolution.
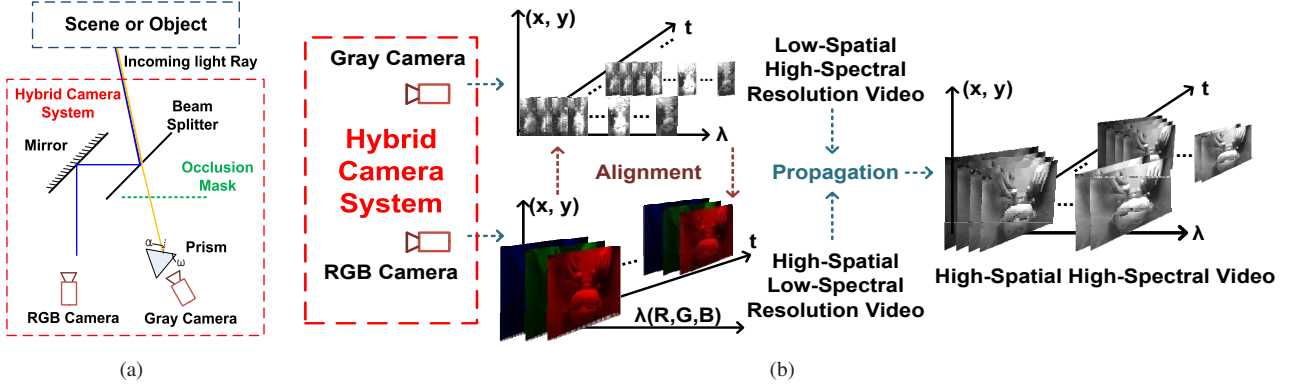
Figure 2. (a) System configuration of the proposed hybrid camera system. (b) Pipeline for high-resolution multispectral video generation.

The light transmitted through the beam splitter is dispersed by the prism onto the grayscale sensor, which measures numerous channels of its spectra. To avoid overlap of spectra from different scene points on the sensor, we employ an occlusion mask as in [7] to subsample the scene radiance. Because of this subsampling, the multispectral imaging at the grayscale camera is obtained at a low spatial resolution.

The triggering of the two cameras is synchronized, such that high resolution RGB video frames and low resolution multispectral video frames are captured simultaneously. The two cameras are aligned to capture the same view. Each sample point of the multispectral imager has a counterpart pixel in the RGB camera that shares the same incoming light ray, and these correspondences are determined as described in Section 4. The correspondences are used by the propagation algorithm to produce high resolution multispectral video.

3.2. Propagation Algorithm

The correspondence between multispectral samples and RGB pixels gives high resolution spectral values for certain pixels in each RGB frame. To obtain high spectral resolution over the entire frame, we propagate the multispectral data to the other pixels according to color similarity and spatial proximity. Although related low-resolution to high-resolution propagation methods have been presented [9, 18], they are not suitable for real-time video processing (e.g. 30 seconds per frame in [9]). This motivates us to use a bilateral filtering approach, which can be computed in real-time [25].

The bilateral filter extends the traditional Gaussian spatial filter to account for differences in intensity. In our scenario illustrated in Figure 3(a), there are RGB pixels that do not have a correspondence to a multispectral sample (as shown in blue), and RGB pixels with known multispectral data from correspondences (as shown in orange). The multispectral data of blue pixels is solved based on color simi-
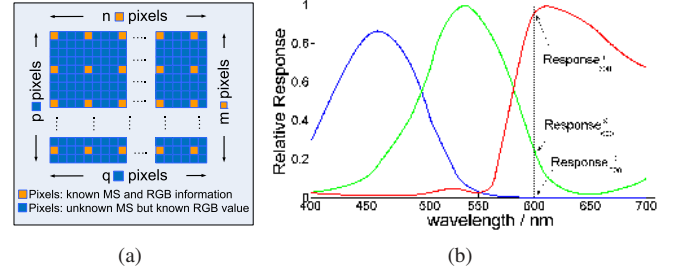

Figure 3. Propagation algorithm. (a) Distribution of multispectral pixels over RGB frame. (b) Response curves of RGB filters.

larity and spatial distance to the surrounding orange pixels.

As exhibited in Figure 3(a), suppose we have an RGB frame with a resolution of $p \times q$ pixels and a corresponding multispectral frame of resolution $m \times n$ $(m < p, n < q)$. For each *blue* pixel (i, j), its multispectral (MS) information is calculated from neighboring *orange* pixels as

$$ms_{ij} = \sum_{c \in R,G,B} \frac{\sum_{k \in \Omega} \mathcal{G}_{\sigma_r}(d_k^{RGB}) \mathcal{G}_{\sigma_s}(d_k^{xy}) \cdot \rho_k \cdot ms_k^c}{\sum_{k \in \Omega} \mathcal{G}_{\sigma_r}(d_k^{RGB}) \mathcal{G}_{\sigma_s}(d_k^{xy})}$$

(1)

where ms_{ij} denotes the multispectral vector for pixel (i, j), $k \in \Omega$ indexes the orange pixels within a neighborhood centered on (i, j), $\mathcal{G}_\sigma()$ represents the Gaussian operator with zero mean and variance σ, and d_k^{RGB} and d_k^{xy} denote the Euclidean distance between the pixels (i, j) and k in RGB space and (x, y) space, respectively. In computing d_k^{RGB}, the RGB values of k in the original RGB frame are used. The factor ρ_k represents the ratio of a given color channel value at k to the corresponding value at (i, j) (e.g., $\rho_k = R_{i,j}/R_k$ for the red channel) and is included for intensity matching.

In Eq. (1), we propagate multispectral data between pixels with similar RGB values, a strong indicator of similarity in scene radiance spectra. The propagation is done per-channel, so that slight intensity differences between source and target pixels in different parts of the spectrum can be accounted for using factor ρ. For this per-channel propaga-

σ_r	$Acc/\%(\sigma_s = 8)$			$Acc/\%(\sigma_s = 16)$			$Acc/\%(\sigma_s = 32)$		
	min	med	max	min	med	max	min	med	max
4	83.2	93.3	97.5	83.7	93.4	97.8	85.2	93.6	97.8
8	81.7	93.6	97.6	82.7	93.6	97.9	83.3	93.6	97.9
16	81.0	93.7	97.6	82.6	93.8	98.0	82.7	93.5	98.1
32	80.3	93.7	97.6	82.0	93.6	97.9	82.2	93.2	98.0

Table 1. Algorithm evaluation on groundtruth datasets with different propagation parameters, Acc means accuracy

tion, the multispectral data is separated into RGB according to the relative responses of the camera's RGB filters at each wavelength (see Figure 3(b)):

$$\boldsymbol{ms}_k^c = \boldsymbol{ms}_k \otimes \boldsymbol{w}^c \quad for \quad c = r, g, b \qquad (2)$$

where

$$w_\lambda^c = \frac{q_\lambda^c}{\sum_c q_\lambda^c} \qquad (3)$$

and q_λ^c denotes the response of filter c at wavelength λ. After propagating the per-channel multispectral data, the three channels are summed to obtain the full spectrum.

The effectiveness of this propagation scheme is evaluated in Table 1 for different algorithm parameters, using all of the *.mat* and *.aix* images in the Joensuu multispectral dataset [20]. Here, we simulated our system by computing the RGB values of each pixel using the RGB filter responses in Figure 3(b) and by sampling the multispectral data at pixels using the same pattern and density (0.3% of total image pixels) as in our physical system implementation. Accuracy is computed from differences of spectral intensity from the ground truth over all the wavelengths. We found the parameter settings of $\sigma_s = 16$ and $\sigma_r = 16$ to be generally effective, with the radius of neighborhood Ω set to twice the image distance between horizontally neighboring multispectral samples. These values were used for all the remaining experiments in this paper.

An important difference between our propagation algorithm and traditional bilateral filtering methods is our use of intensity modulated spectra (see ρ factor in Eq.(1)). This modulation is critical for our system to accurately handle subtle variations such as shading using the very sparsely sampled multispectral points. In fact, the median accuracy over the datasets of Table 1 drops from 93.8% to 88.7% without the intensity modulation.

Several of the multispectral images used to compute Table 1 are exhibited in Figure 4. Here, the original multispectral data was integrated with the RGB filter curves in Figure 3(b) to generate RGB values for display purposes. In Figure 4(a), we show the image that yielded the minimum accuracy at all parameter settings. The scene contains many spectral variations, even within a single bunch of yarn, and the signal-to-noise ratio is relatively low for this dark image. The accuracy is nevertheless at a reasonable level for this particularly challenging example. While this experiment examines accuracy for multispectral data propagation within a single frame, better results are generally obtainable

(a) (b) (c)

(d) (e) (f) (g)

Figure 4. Propagation results. (a-e) Example images from the Joensuu multispectral dataset [20] used for parameter evaluation in Table 1. (f) Error map for the image in (e) without temporal enhancement. (g) Error map for the image in (e) with temporal enhancement.
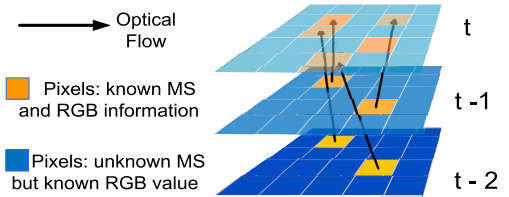

Figure 5. Optical flow for temporal enhancement. For a frame at time instant t, we use temporal correspondences from the two preceding frames $t-1$, $t-2$ to generate more spectral samples.

using multiple frames from a video, as described in the following subsection.

3.3. Temporal Enhancement

Since our system captures video streams, multispectral data can be propagated temporally as well as spatially. The multispectral data recorded for a scene point in one frame is approximately valid for the same scene point in other frames captured close in time. As illustrated in Figure 5, optical flow [3] is adopted to track scene points for which multispectral data has been measured. The multispectral data of these points can then be propagated to other frames, thus increasing the density of multispectral samples used in the spatial propagation described in the previous subsection.

Although spatial propagation is often sufficient for multispectral video generation, temporal enhancement can lead to improved accuracy. This is exemplified using Figure 4(e) with simulated linear motion. Without temporal enhancement, the relative error computed over the wavelengths of propagated colors with respect to the ground truth is 7.8%, with the error map shown in Figure 4(f). When temporal enhancement is used, the relative error drops to 4.3%, with the error map shown in Figure 4(g).
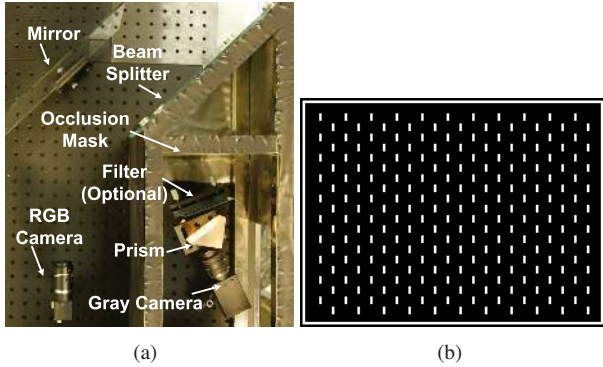
Figure 6. (a) System prototype. (b) Occlusion mask configuration.

To minimize the effects of view-dependent appearance changes and tracking errors, we use only the two immediately preceding frames for temporal enhancement in our implementation. The use of fewer frames also expedites computation. We note that in video analysis, *the objects of interest are generally in motion, and thus benefit from temporal enhancement.*

3.4. Design Tradeoffs

The settings of the hybrid camera system can be adjusted for different spectral resolutions. The spectral resolution depends on several system parameters:

$$Spectral\ Resolution \propto \{f, n, \omega, \frac{1}{\alpha}\} \qquad (4)$$

where the f is the focal length of the camera lens, n is the refraction index of the prism, ω is the prism angle and α is the incident light angle, as shown in Figure 2(a). The most effective way to increase spectral resolution is to use a larger focal length lens with a suitable occlusion mask that prevents overlaps of spectra on the sensor. Unlike the device in [7], higher spectral resolution does not require a corresponding loss of spatial resolution, which is determined by the resolution of the RGB camera in our system. However, when spectral resolution is increased using a larger focal length, our system obtains fewer multispectral sample points per frame, which can reduce propagation accuracy. Another tradeoff when using a large focal length to gain spectral resolution is that less light energy is collected on the grayscale sensor, which may then require an increase in exposure time and a lower frame rate.

Different camera settings additionally affect the image distortions introduced by the system optics. For example, a larger aperture increases light throughput but results in greater geometric distortion. Such tradeoffs must also be taken into consideration when selecting a camera design.

4. System Implementation

The hybrid camera system, exhibited in Figure 6(a), was implemented using a PointGrey® GRAS-50S5M as the greyscale camera, which can capture $15\,fps$ video at a maximum resolution of 2448×2048. The focal length of the greyscale camera lens is set according to the spectral resolution required in a given application. The RGB camera is a PointGrey® Flea2-08S2, which has a maximum resolution of 1024×768 at 25 fps. The two cameras are synchronized using the PointGrey® MultiSync program. A half-reflect-half-pass beam splitter provides the two cameras with equal light energy. The occlusion mask is configured as shown in Figure 6(b), with rectangular holes through which the prism disperses light in the horizontal direction. Each hole represents one multispectral sample, and the projected spectrum from the hole is averaged vertically to reduce noise. The prism is made of $BK7$ glass with known refraction indices for wavelengths $400 \sim 1000nm$. An optional filter may be placed in the optical path to isolate a certain band of the spectrum. The spectrum, geometry and radiance distortions in the multispectral imager are calibrated as in [7].

To register the two cameras, we place an LCD screen in front of the capture system and display a calibration video of horizontal and vertical line sweeps. The lines are displayed on the LCD in white, which has a spectrum with peaks similar to those of fluorescent light. When a horizontal line sweeps past a row sampled by the multispectral sensor, it produces a sharp response and its corresponding row position on the RGB sensor is identified. Similar correspondences between columns are obtained using a vertical line sweep. The information from the two sweeps jointly determines the correspondences of individual multispectral samples to pixels on the RGB sensor. This registration and alignment is a one-time operation, and does not need to be repeated when capturing another scene.

5. Experimental Results

In this section, we use the proposed system to capture several videos and demonstrate the utility of the system on various applications. The high-spectral, high-spatial resolution measurements provide considerably more data than traditional RGB or monochrome cameras for identifying and distinguishing different materials and objects. The fast generation of output video from the simultaneously captured streams makes our system practical for real-time video applications such as tracking.

Resolution comparison We demonstrate the increase in spatial resolution that can be gained with our approach over using just a prism-mask multispectral video system as presented in [7]. In this example, the multispectral imager is configured so that each spectral sample spans about 60 pixels on the greyscale sensor over the wavelength range from $400nm$ to $700nm$, giving a spectral resolution of about $5nm$ and a spatial resolution of 116×24. With RGB values computed using the color filter response curves of the RGB
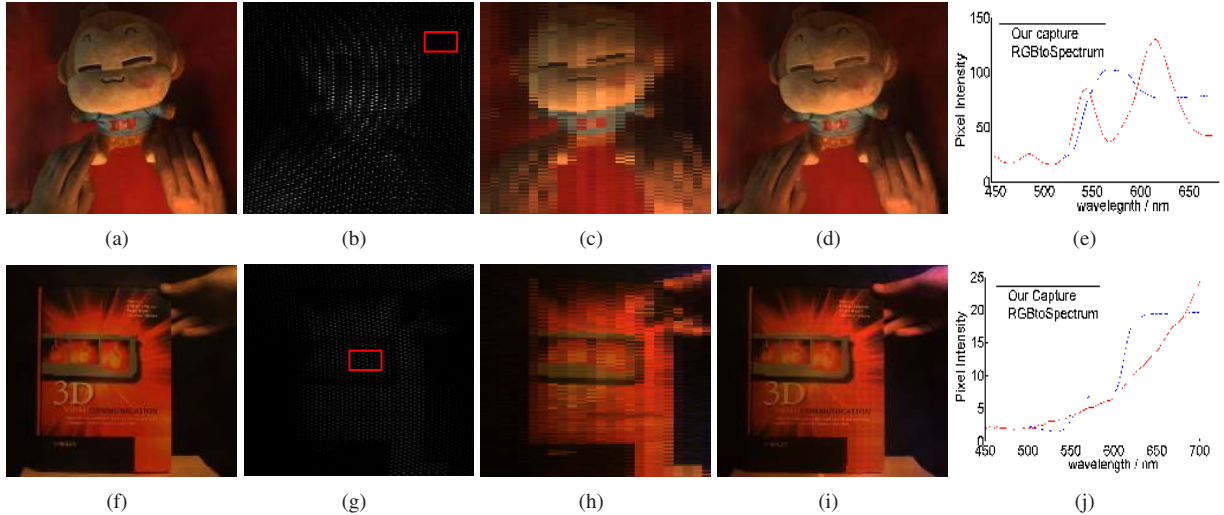
Figure 7. Resolution comparison. (a)(f) Video frames captured by RGB camera. (b)(g) Video frames captured by grayscale camera. (c)(h) Reconstructed RGB in grayscale frames. (d)(i) High resolution RGB video using propagation algorithm. (e)(j) Comparison of spectral detail from an RGB camera and our multispectral system.

camera, video frames captured by the multispectral sensor are shown in Figure 7(c) and 7(h). At this resolution, details on the objects are unclear. The characters on the monkey and book become visible at the 1024×768 resolution produced with our hybrid camera system as shown in Figure 7(d) and 7(i). These images also show the RGB reproduction from propagated multispectral data to be consistent with the captured RGB frames. Figure 7(e) and 7(j) exemplifies the additional spectral detail that can be acquired with our device in comparison to the coarse spectra reconstructed from RGB values using the method of [19].

Automatic White Balance High spectral resolution can benefit techniques for automatic white balance. Figure 8 shows an example of a warm colored object under fluorescent lighting, captured by a regular RGB camera and by our system. The warm colors of the object can mislead a white balance algorithm to infer a warm illumination color like tungsten light, giving the incorrect result in Figure 8(b). Figure 8(c) is the white balance result using the grey world hypothesis [4], a standard approach that assumes the average scene radiance to be achromatic. In reality, the scene is illuminated by a fluorescent light, obvious from the spectra in 8(e) captured by our system. Fluorescent lighting is detected by the two characteristic peaks at wavelengths $546nm$ and $611nm$. Figure 8(d) shows the white balance result with the correct illumination. Since the background color is white in this scene, we can evaluate the white balance result based on the average color of the background, shown at the bottom-left of each figure. In the supplementary material, we additionally show a video that demonstrates the white balance capability of our system under temporally varying illumination, a challenging case for current automatic white balance algorithms.

In Figure 9, we show another example of white balancing for a scene with hybrid illumination, where a fluorescent light source is on the left and a tungsten light source is on the right. Such scenes also present great difficulty to white balance algorithms because of blended light colors. Here, we use a simple heuristic to estimate the proportion α of fluorescent light intensity at each point, according to the height of its characteristic spectral peaks (at 546nm and 611nm) in relation to the spectral intensity at a wavelength in between (578nm). Specifically, we compute α from $k(p(546) + p(611) - 2 * p(578))/max(p(546), p(578))$, where $p(\lambda)$ denotes spectral intensity at wavelength λ, and k is a scale factor. Then the white balance matrix is computed by a combination of the matrices for fluorescent light M_f and tungsten light M_t as $\alpha M_f + (1 - \alpha)M_t$. For the input frame exhibited in Figure 9(a), we obtain the spatially-variant white balance result shown in Figure 9(d) using our simple method. For comparison, global white balancing with fluorescent and tungsten lighting is displayed in Figure 9(b) and Figure 9(c) respectively. In Figure 9(e), we show the differences in spectra among points with different light blending. The high spectral and spatial resolution of our device allows for spatially-variant white balancing with high spatial precision.

Tracking In tracking, features of the target object need to be distinguished from the background. This problem is particularly challenging when the foreground object and background have a similar color. Figure 10 illustrates such an example, with a moving foreground object of blue poster color and part of the background in blue water color. With a traditional RGB camera, the two colors are difficult to dis-
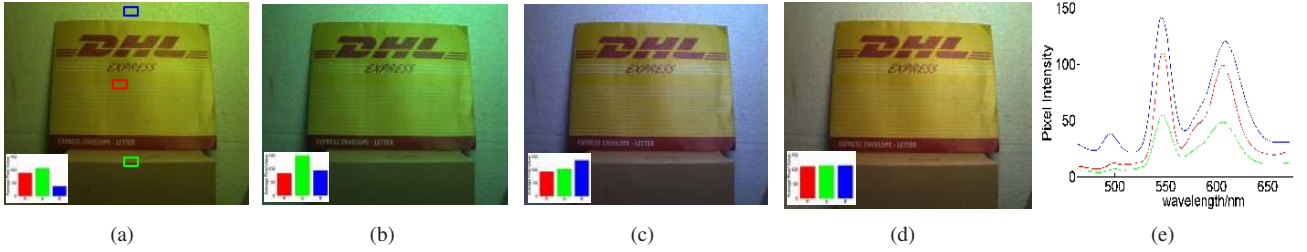
Figure 8. Automatic white balance. (a) Raw frame captured by the RGB camera. (b) White balance result based on tungsten light. (c) Result of grey world white balance algorithm. (d) Result using correct illumination information (fluorescent light). (e) Multispectral curves of the marked areas in (a), which were captured by the proposed system. The peaks of these curves reveal the actual illumination to be fluorescent light.

Figure 9. White balance with hybrid illumination. (a) Raw frame captured by the RGB camera. (b) Global white balancing based on fluorescent light. (c) Global white balancing based on tungsten light. (d) Spatially-variant white balancing with our system. (e) Multispectral curves of points with different light blending.

tinguish. Tracking with a state-of-the-art algorithm [16] fails when the object moves over the blue background. With the hybrid camera system, the spectral differences between poster color and water color become evident, allowing for accurate tracking. Here, the tracking algorithm uses responses at wavelengths $550nm$, $625nm$ and $700nm$ as color features for the multispectral input video.

We also demonstrate tracking of human skin, a task important in many surveillance applications. A special characteristic of human skin spectra is a small hump centered on wavelength $559nm$, a physical feature caused by melanin and hemoglobin in skin [1]. Detection of this hump using a multispectral camera can distinguish human skin from other materials with similar colors in RGB space. Figure 11 shows an example in which we capture a real human hand in front of a fake hand printed on paper. In this experiment, we identify true skin pixels by thresholding the quantity $\big(r(559)-0.5(r(540)+r(580))\big)$ to detect the $559nm$ peak, where $r(\lambda)$ denotes the spectral radiance at wavelength λ. We use a $50mm$ lens and a $400 \sim 620nm$ bandpass filter in this example, which gives a spectral resolution of about $1nm$. We capture a video clip at $5fps$ with the hand performing different gestures. One frame is shown in Figure 11(a). Figure 11(b) shows the corresponding spectral frame captured by the grayscale camera. Figure 11(c) plots the spectral curves of samples from the genuine human hand and fake hand. The spectral hump around $559nm$ is evident in measurements of the real hand. Figure 11(e) shows the detection result in a video frame, with the detected skin pixels highlighted. In comparison to the low resolution frame

in Figure 11(d), the high spatial resolution of our system in Figure 11(e) allows for a clearer analysis of hand gestures.

6. Conclusion

Spectral imaging at video rates has long been a challenging problem. In this paper, we address this problem with a hybrid camera system that generates video with both high spectral and high spatial resolution. Our experiments demonstrate the utility of such video in different computer vision applications.

A limitation of this system is the relatively large depth of field needed to reduce optical distortions from the prism and avoid blur of the occlusion mask. While this can be accomplished with a small aperture size, this limits the light throughput of our system. In future work, we plan to extend the bilateral filtering formulation of the propagation algorithm to a trilateral filter that includes a time dimension, where multispectral samples from temporally closer frames are more heavily weighted in the temporal enhancement. As trilateral filtering is a standard function supported in modern graphics cards, real-time processing should remain possible with such a propagation scheme.
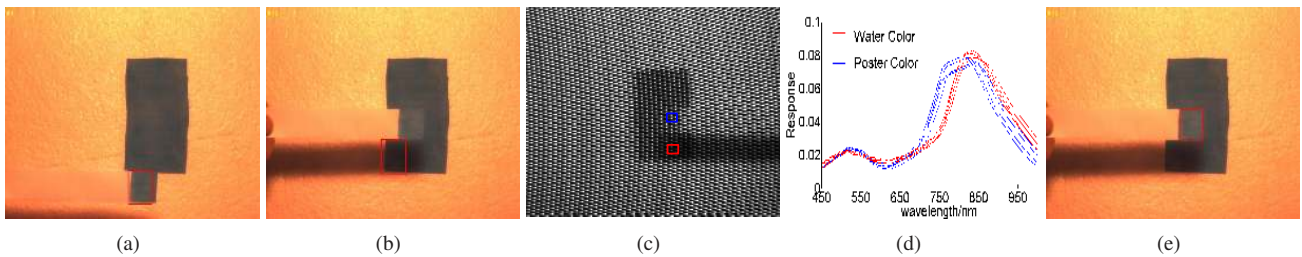
Figure 10. Object tracking. (a) Tracking target, highlighted by a red rectangle. (b) Unsuccessful tracking using RGB camera. (c) Multi-spectral frame from grayscale sensor. (d) Spectra of poster color and water color in areas marked in (c). (e) Successful tracking using the hybrid camera system.
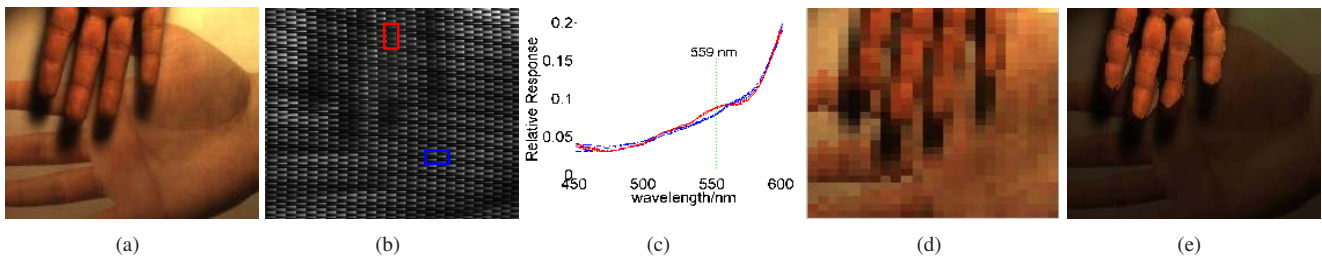

Figure 11. Human skin tracking. (a) RGB frame containing a genuine hand and fake hand. (b) Multispectral frame captured by grayscale sensor. (c) Spectral curves from real human skin in red and printed skin in blue. (d) Low resolution frame without multispectral propagation, where image details are unclear. (e) High resolution capture by our system, where the detected human skin pixels are highlighted.

References

[1] E. Angelopoulou. Understanding the color of human skin. In *SPIE Conference on Human Vision and Electronic Imaging*, volume 4299, pages 243–251, 2001.

[2] D. J. Brady and M. E. Gehm. Compressive imaging spectrometers using coded apertures. *SPIE*, Vol.6246, 2006.

[3] T. Brox, A. Bruhn, N. Papenberg, and J. Weickert. High accuracy optical flow estimation based on a theory for warping. In *ECCV*, 2004.

[4] G. Buchsbaum. A spatial processor model for object colour perception. *Journal of the Franklin Institute*, 310, 1980.

[5] C.Volin. *MWIR Spectrometer Operating Theory*. University of Arizona Press, 2000.

[6] M. Descour and E. Dereniak. Computed-tomography imaging spectrometer: experimental calibration and reconstruction results. *Applied Optics*, 1995.

[7] H. Du, X. Tong, X. Cao, and S. Lin. A prism-based system for multispectral video acquisition. In *Proc.ICCV*, 2009.

[8] D. W. Fletcher-Holmes and A. R. Harvey. Real-time imaging with a hyperspectral fovea. *J. Opt. A: Pure Appl. Opt.*, 7:S298–S302, 2005.

[9] A. Gupta, P. Bhat, M. Dontcheva, B. Curless, O. Deussen, and M. Cohen. Enhancing and experiencing spacetime resolution with videos and stills. In *International Conference on Computational Photography*, 2009.

[10] N. Hagen and E. L. Dereniak. Analysis of computed tomographic imaging spectrometers. i. spatial and spectral resolution. *Applied Optics*, 47:F85–F95, 2008.

[11] J. James. *Spectrograph Design Fundamentals*. Cambridge University Press, 2007.

[12] W. R. Johnson, D. W. Wilson, and G. Bearman. Spatial-spectral modulating snapshot hyperspectral imager. *Applied Optics*, 45:1898–1908, 2006.

[13] J. M. Mooney, V. E. Vickers, M. An, and A. K. Brodzik. High throughput hyperspectral infrared camera. *J. Opt. Soc. Am. A*, 14:2951–2961, 1997.

[14] S. Paris and F. Durand. A fast approximation of the bilateral filter using a signal processing approach. In *ECCV*, 2006.

[15] J.-I. Park, M.-H. Lee, M. D. Grossberg, and S. K. Nayar. Multispectral imaging using multiplexed illumination. In *ICCV*, 2007.

[16] D. Ross, J. Lim, R.-S. Lin, and M.-H. Yang. Incremental learning for robust visual tracking. *IJCV*, 77, 2008.

[17] Y. Y. Schechner and S. K. Nayar. Generalized mosaicing: wide field of view multispectral imaging. *IEEE PAMI*, 24(10):1334–1348, 2002.

[18] E. Shechtman, Y. Caspi, and M. Irani. Increasing space-time resolution in video. In *ECCV*, pages 753–768, 2002.

[19] B. Smits. An rgb to spectrum conversion for reflectances. *Journal of Graphics Tools*, 4(4):11–22, 1999.

[20] Spectral Database, University of Joensuu Color Group. http://spectral.joensuu.fi/.

[21] Spectral Imaging LTD. http://www.specim.fi.

[22] A. Wagadarikar, R. John, R. Willett, and D. Brady. Single disperser design for coded aperture snapshot spectral imaging. *Applied Optics*, 47:B44–B51, 2008.

[23] A. A. Wagadarikar, N. P. Pitsianis, X. B. Sun, and D. J. Brady. *Optics Express*, (8).

[24] M. Yamaguchi, H. Haneishi, H. Fukuda, J. Kishimoto, H. Kanazawa, M. Tsuchida, R. Iwama, and N. Ohyama. High-fidelity video and still-image communication based on spectral information: Natural vision system and its applications. *SPIE/IS&T Electr. Imag.*, Vol.6062, 2006.

[25] Q. Yang, K.-H. Tan, and N. Ahuja. Real-time O(1) bilateral filtering. In *CVPR*, 2009.