# A Prism-based System for Multispectral Video Acquisition

Hao Du*† Xin Tong‡ Xun Cao§ Stephen Lin‡

*Fudan University †University of Washington ‡Microsoft Research Asia §Tsinghua University

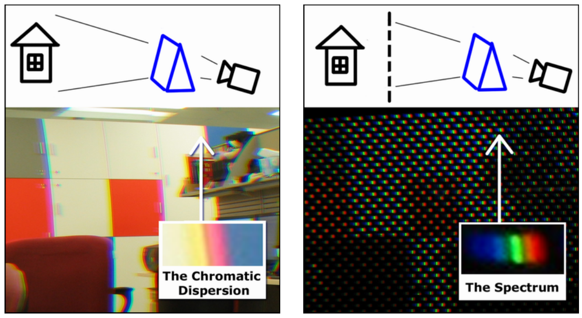
## Abstract

*In this paper, we propose a prism-based system for capturing multispectral videos. The system consists of a triangular prism, a monochrome camera, and an occlusion mask. Incoming light beams from the scene are sampled by the occlusion mask, dispersed into their constituent spectra by the triangular prism, and then captured by the monochrome camera. Our system is capable of capturing videos of high spectral resolution. It also allows for different tradeoffs between spectral and spatial resolution by adjusting the focal length of the camera. We demonstrate the effectiveness of our system with several applications, including human skin detection, physical material recognition, and RGB video generation.*

## 1. Introduction

Light emitted from illumination sources or reflected from scene objects generally spans a broad range of wavelengths. Though trichromatic sensing suffices for the human visual system, multispectral imaging with many more than three spectral measurements per scene point provides greater information about scene colors, which can be used to the advantage of a number of machine vision applications, including illumination analysis, object tracking, and material recognition.

Several methods have been developed for acquiring multispectral images and video of real scenes. Filter-based methods [4, 11] sequentially place band-pass filters in the optical path to measure a set of specific spectral components over multiple shots. These systems tradeoff temporal resolution for spectral resolution, and cannot be used to capture a large number of spectral measurements at video rates. Methods based on computed tomographic imaging spectrometry (CTIS) [2] use diffraction gratings to generate multiple, spectrally dispersed images of the scene on the camera sensor, then reconstruct the final image by solving a linear system. Due to practical difficulties in calibration, only very simple scenes have been successfully acquired in this manner [13]. Other methods for multispectral video acquisition either are limited to scenes with coded illumination [10] or increase the spectral resolution by only one color channel [7]. At present, multispectral video acquisition with high spectral resolution for arbitrary scenes remains a challenging task.




(a) (b)

Figure 1. Role of the occlusion mask in the prism-based multispectral imaging system. (a) Without an occlusion mask, a prism disperses an incoming light ray into its spectrum. The dispersed spectra of neighboring rays overlap one another and become difficult to separate. (b) With an occlusion mask, the spectra of the rays that pass through the mask holes are separated and distinguishable in the recorded frames.

In this paper, we propose a new multispectral imaging system that enables capture of many spectral samples at video rates. In this system, a triangular prism is placed in front of the camera. Since different wavelengths of light have different refractive indices, the prism disperses each incoming light ray into a spectrum of its constituent colors, which is then recorded by the monochrome camera. To avoid overlap among the spectra of neighboring rays, an occlusion mask is put in front of the prism so that only rays passing through the mask holes can be dispersed and projected onto the image plane, as shown in Figure 1.

The prism-based multispectral imaging system consists of low-cost off-the-shelf components and is easy to setup. In contrast to previous techniques based on a sequence of filters [4, 11] or filter-pixel pairings [7], our prism-based system provides *true multispectral measurements of an imaged scene point at a given point in time*. Spatial resolution

* This work was done during Hao Du and Xun Cao's internships at Microsoft Research Asia.

is exchanged for this gain in spectral resolution, and this tradeoff can be easily adjusted by changing the focal length of the camera's lens. Unlike CTIS systems [2] where calibration noise significantly affects image reconstruction, our system directly measures the spectral values of scene points and outputs images of much higher quality. As a result, the prism-based system passively captures multispectral video at high spectral resolutions and potentially at the frame rate of the camera. We illustrate the effectiveness of our system with several applications, including human skin detection, material recognition, and RGB video generation with illumination identification.

## 2. Related Work

Several methods place bandpass filters in front of the sensor to capture specific spectral components of the scene. This may be implemented with a tunable filter [4] or a wheel of filters [14], and produces a set of images captured in sequence. Schechner and Nayar [11] attached a spatially varying filter to a camera. By scanning the scene with this camera, each scene point is sensed by multiple pixels each having a different filter response. These techniques are well designed for capturing static scenes by exchanging temporal resolution for spectral resolution. However, they are not well suited for multispectral imaging at video frame rates. Kidono and Ninomiya [7] developed a multiband camera for capturing four spectral channels (R, G, B and infrared) of the scene, in which half of the green CCD filters are replaced with infrared filters. A drawback of this approach is that the different spectral samples are being measured from different scene points, and this spatial discrepancy becomes more significant when considering greater numbers of different bandpass filters. In all these filter-based solutions, the spectral resolution of captured images is limited by the minimal passband of such filters (10-15nm).

A prism is used in spectrometers, which measure the color spectrum of a single light ray [6]. Although a spectrometer provides very high spectral resolution (e.g., $0.1nm$), the limited spatial resolution ($1 \times 1$) precludes its use in general video acquisition. Although a scene can be scanned with a spectrometer [12], this solution would be unsuitable for dynamic scenes. An alternative approach proposed by Mohan *et al*. [8] controls the spectral components arriving at the CCD plane by using an agile optical path. This system, however, has been demonstrated only for efficiently obtaining a few color primaries of static scenes.

Fletcher-Holmes and Harvey [3] developed a device based on optical fiber bundles for real-time multispectral imaging. However, the system is complex to implement, and has been demonstrated for only a 14x14 image resolution due to practical limitations on fiber thickness. Harvey et al. [5] presented a multispectral imaging system based on a generalization of the Lyot filter. This system requires special optical components, and may produce smoothed spectra due to its bell-shaped passbands with significant side lobes. Recently, Park *et al*. [10] captured multispectral video of scenes illuminated with multiplexed coded lighting. This system captures six-band multispectral video at 30fps, but is limited by its need for special light sources.

Computed tomography imaging spectrometers [2][13] obtain multiple, spectrally dispersed images of a scene within a single shot using specially-made gratings. To reconstruct both the scene and spectra from this data, a computational tomography algorithm is used. An advantage of this approach over our prism-based system is its relatively high light throughput, which leads to a higher signal-to-noise ratio in theory. However, in practice this system is difficult to calibrate, and results suggest that the system is sensitive to calibration inaccuracies. As a result, only simulations and snapshots of very simple scenes have been successfully demonstrated [13].

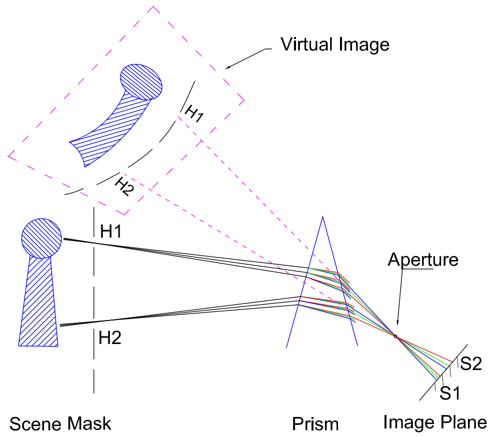


Figure 2. Overview of optical path. Incoming light rays are sampled by the occlusion mask, dispersed by the prism into a spectrum, and projected to different positions on the image plane. Also, all rays undergo a change in direction after refraction.

## 3. Prism-based Multispectral Imaging System

In this section, we discuss the design and configuration of the prism-based imaging system. For ease of understanding, we focus on a 2D imaging system illustrated in Figure 2, which can be regarded as a 2D slice of the 3D system.

### 3.1. System Overview

Figure 2 illustrates the basic configuration of our imaging system. The light that passes through a mask hole ($H1$ or $H2$) is dispersed by the triangular prism into a linear spectrum ($S1$ or $S2$). Since the prism changes the overall direction of incoming rays, the rays projected onto the image plane appear to arrive from directions different from the actual scene, thus forming a "virtual image". We rotate

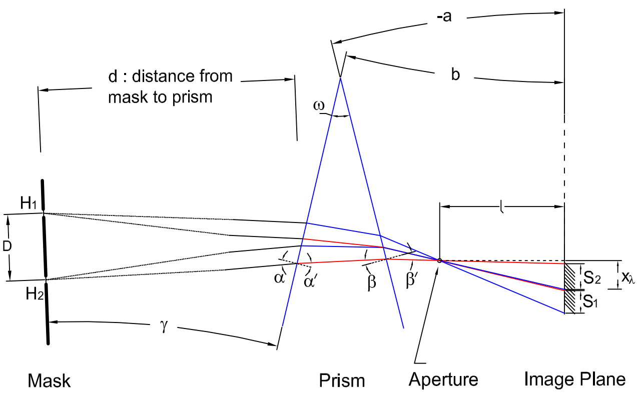



Figure 3. Optical paths of light rays passing through neighboring mask holes $H_1$, $H_2$ with their spectra $S_1$, $S_2$ projected onto the image plane. In this illustration, only rays of maximum or minimum wavelength that pass through the aperture are shown.

the camera so that it faces the virtual image, and calibrate the spatial and spectral distortions caused by ray refraction. The calibration procedure is described in Section 4.

In the remainder of this section, we first discuss the relationship between the system configuration and the spectrum width on the image plane for a single ray, which limits the spectral resolution in the video frames. We then discuss how to decide the distance among holes in the mask.

### 3.2. Determining Spectrum Width

As shown in Figure 3, consider the light that passes through mask hole $H_2$ and the aperture (illustrated as a pinhole), and is projected into its spectrum $S_2$ on the image plane. The incidence angle of the rays at the prism is $\alpha$, and is assumed to be equal for all rays that pass through a hole, given the large distance $d$ from the mask to the prism. According to Snell's law of refraction, we have

$$n_\lambda = \frac{\sin(\alpha)}{\sin(\alpha'(\lambda))} = \frac{\sin(\beta'(\lambda))}{\sin(\beta(\lambda))}, \tag{1}$$

where $n_\lambda$ is the refraction index of the prism for wavelength $\lambda$. $\alpha'(\lambda)$ and $\beta(\lambda)$ are the angles of the refraction ray inside the prism relative to the normals of the two prism edges, and $\beta'(\lambda)$ is the angle of refraction of the outgoing ray, as shown in Figure 3.

From Eq. 1, we have

$$\beta'(\lambda) = \arcsin\Big(n_\lambda \cdot \sin\big(\omega - \arcsin(\frac{\sin\alpha}{n_\lambda})\big)\Big), \tag{2}$$

where $\omega = (b - a)$ is the prism angle, and $a$ and $b$ denote the angles of the prism's two surfaces with the image plane.

As shown in Figure 3, for a specific wavelength $\lambda$ of light that passes through a given mask hole (e.g., $H_2$), its displacement $x_\lambda$ relative to the projection of the aperture on the image plane is

$$x_\lambda = l \cdot \tan(a + \beta'(\lambda)), \tag{3}$$

where $l$ is the camera's focal length.

For incoming light with wavelengths ranging from $\lambda_s$ to $\lambda_e$ that passes through a mask hole (e.g., $H_2$), the width of its spectrum on the image plane $w(S_2)$ can be computed as

$$\begin{aligned} w(S_2) &= x_{\lambda_e} - x_{\lambda_s} \\ &= l \cdot \big(\tan(a + \beta'(\lambda_e)) - \tan(a + \beta'(\lambda_s))\big) \end{aligned} \tag{4}$$

Eqs. (4) and (2) show that increasing either the camera's focal length $l$ or the refractive indices of the prism $n_\lambda$ will increase the width of the projected spectrum on the image plane and thus increase the spectral resolution measurable in the captured images. Also, note that the distance between the mask and prism has no effect on the spectral resolution.

We can also conclude from Eq. (3) that the projected spectrum is not uniformly distributed on the image plane with respect to wavelength. This nonuniform distribution is called *spectral distortion* and also needs to be calibrated in our system.

### 3.3. Determining Distance Between Mask Holes

To maximize spectral and spatial sampling rates in a single shot, the dispersed spectra of all rays passing through the mask holes should tightly tile the entire image plane. Thus the minimal distance $D$ between the mask holes should be determined such that the spectra of rays passing through neighboring mask holes are positioned closely on the image plane without overlap between them.

As illustrated in Figure 3, consider two mask holes $H_1$ and $H_2$. With a minimal distance $D$ between them, light of maximum wavelength $\lambda_e$ (shown in red) through $H_1$ and light of minimum wavelength $\lambda_s$ (shown in blue) through $H_2$ would follow the same path after refraction and be projected to the same position (between $S_1$ and $S_2$) on the image plane. To compute $D$, we shoot a ray from the common position between $S_1$ and $S_2$ on the image plane toward the occlusion mask. Its $\lambda_s$ component goes to $H_2$ and $\lambda_e$ component goes to $H_1$.

Let $d$ be the distance from the occlusion mask to the prism measured along the normal of the mask. Since the size of the prism is small relative to the distance between the mask plane and prism, $d$ can be considered as approximately the distance from the mask plane to the intersection of the two components ($\lambda_e$ through $H_1$, and $\lambda_s$ through $H_2$) within the prism. Similar to computing $w(S_2)$ as described in Section 3.2, we obtain $\alpha(\lambda)$ from a given $\beta'$ (note that $\beta'(\lambda)$ here has the same value for the two examined components) as

$$\alpha(\lambda) = \arcsin\Big(n_\lambda \cdot \sin\big(\omega - \arcsin(\frac{\sin\beta'}{n_\lambda})\big)\Big). \tag{5}$$

Similar to Eq. (4), here we have

$$D = d \cdot \big(\tan(\gamma + \alpha(\lambda_e)) - \tan(\gamma + \alpha(\lambda_s))\big), \quad (6)$$

where $\gamma$ is the angle between the occlusion mask and the edge of the prism that is closer to the mask.

Eq. (6) illustrates that the minimal distance between mask holes is independent of the spectrum width on the image plane and the camera's focal length. From the above analysis we can also conclude that the minimal distance between mask holes is not uniform on the occlusion mask plane.

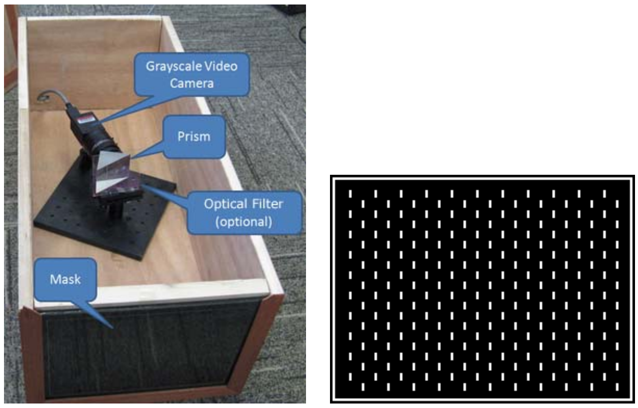
In practice, instead of measuring these parameters to compute the minimal distance between each pair of neighboring mask holes on the occlusion mask plane, we apply the same minimal distance for all pairs of neighboring mask holes (as shown in Figure 4(b)) to simplify the design. Although this solution may not place the spectra as tightly as possible on the image plane to obtain the maximal spatial resolution, the reduction of the spatial resolution with uniform hole spacing is small. The correspondence between mask holes and captured spectra is then found in the calibration process described in Section 4.2.

## 4. System Implementation and Calibration

We implemented the prototype shown in Figure 4(a). The prism is made of $BK7$ glass with known refraction indices for wavelengths $400 \sim 1000nm$. The monochrome camera used in our system is the PointGray GRAS-50S5M, which can capture $2448 \times 2048$ images at $15fps$. The focal lengths for the camera are 16, 25 or 50mm, selected according to the spectral resolution needed in a given application.

A possible design for the occlusion mask, which we used in our system, is shown in Figure 4(b). The prism disperses light along one direction (e.g., horizontally), so we created mask holes as vertical slits instead of circular holes, which could be used to obtain spectra from vertically adjacent scene points that may be averaged to reduce noise. The shift of the hole configuration between consecutive rows is intended for more uniform spatial sampling. The width of the mask holes presents a tradeoff between spectral resolution and light ray intensity. In our implemented system, the hole width is set to 0.2mm, height is set to either 1mm or 2mm, and the size of the entire occlusion mask is 26*20cm. The distance between the neighboring holes is 5mm.

We go through three steps to calibrate the system. *Spectrum calibration* is to rectify the nonlinearity between the displacement $x_\lambda$ of a spectral component on the image plane and its wavelength $\lambda$. *Geometry calibration* is implemented to undo the global distortion caused by the prism. *Spectral radiance recovery* is required to obtain the true spectral radiance according to the CCD sensitivity curves with respect to wavelength. These steps are discussed in the following subsections.




(a) (b)

Figure 4. (a) Our system prototype. (b) A mask sample.

### 4.1. Spectrum Calibration

As described in Section 3.2, the dispersed spectrum is not linearly distributed on the image plane with respect to wavelength. Although theoretically the exact projection of each spectral wavelength on the image plane can be computed, it is difficult to obtain all influencing parameters from a real device setup. We thus approximate the spectrum distortion for this calibration and then verify its accuracy.

Since the refractive index of $BK7$ glass for each $\lambda$ in the visible to near-infrared range $[400nm, 1000nm]$ is known, we study the properties of Eq. (3) by randomly setting the parameters $a$, $b$, $c$ and $l$ in the following ranges $[0, \pi/5]$, $a + 0.1 + [0, \pi/4]$, $b + 0.1 + [0, \pi/4]$, $3000 + [0, 2000]$ and simulating the *spectrum distortion curves* of $(x_\lambda - x_{400})$ for $\lambda \in [400, 1000]nm$, (six sampled simulation curves are plotted in red in Figure 5). We found all of these distortion curves to be roughly consistent up to a scale factor. So by detecting two points of known wavelength on the image plane, we can accurately approximate the spectrum distortion curve. Specifically, with two known wavelengths $\lambda_0$ and $\lambda'_0$, we can observe their displacements $x_{\lambda_0}$ and $x_{\lambda'_0}$. We generate distortion curves based on random parameters, and learn the relative ratios, $r_{\lambda_i}$, for $m$ other control points with displacements $x_{\lambda_i}$ $(i = 1, 2, ..., m)$. The relative ratios are defined as

$$r_{\lambda_i} = \frac{x_{\lambda_i} - x_{\lambda_0}}{x_{\lambda'_0} - x_{\lambda_0}}. \quad (7)$$

These $m$ points together with the originally observed two points serve as control points that are fit with a B-spline to approximate the true spectrum distortion curve.

To verify how well we can recover the *spectrum distortion* using this method, we synthesize 100 *spectrum distortion curves*, based on two displacements $x_\lambda$ for $\lambda_0 = 546.5nm$ and $\lambda'_0 = 611.6nm$, which are wavelengths of two strong peaks easily obtainable from fluorescent light, and learn from these curves the

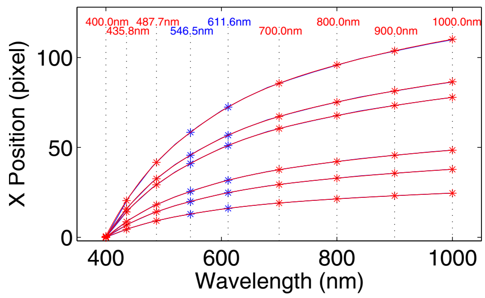



Figure 5. Spectrum distortion approximation. The red curves show the simulated ground-truth position of the spectrum ranging from 400nm-1000nm. Based on only two observations (546.5nm, and 611.6nm), the other seven control points are inferred via learned ratios $r_{\lambda_i}$. Blue curves are the approximated curve which are very similar to the genuine positions shown by the red curves.

$r_{\lambda_i}$ for the other seven arbitrarily chosen control points ($\{400, 435.8, 487.7, 700, 800, 900, 1000\}nm$). Then we synthesize another 100 curves, 'measure' the displacements of $\lambda_0 = 546.5nm$ and $\lambda_0' = 611.6nm$, estimate the displacements of control points, and approximate the curve using B-spline fitting. From this experiment, the RMS error is found to be small, $0.62\%$. Six of the approximations are shown by blue curves in Figure 5, which closely match the synthesized ground truths.

In practice, we capture a white surface under fluorescent light, and detect the two strong peaks at known wavelengths: $546.5nm$ and $611.6nm$. Figures 6(a) and 6(c) show one captured spectrum. Using the above technique, we undistort the captured spectrum to obtain the linear spectrum shown in Figures 6(b) and 6(d), which closely matches the ground truth spectrum of fluorescent light. Note that this experiment involves real capture using our system with a $50mm$ lens, and demonstrates the ability of our system to recover high resolution spectra.

From a single image of a white surface under fluorescent lighting, we can calibrate the spectrum distortion curve of each mask hole. After this one-time calibration, the system can be used to capture arbitrary scenes with arbitrary illumination.

### 4.2. Geometry Calibration

Geometric distortion also exists due to the prism. Instead of analyzing the underlying optical geometry of this distortion, we calibrate and recover the distortion, then unwarp its effects. Figure 7(a) shows a sample distorted image of a white surface illuminated by fluorescent light, captured by our system using the mask as shown in Figure 4(b). Since the mask has a known regular pattern and the precise locations of each mask hole in the captured image can be found by detecting the two fluorescent light peaks, we can simply unwarp the captured image to match the known mask.

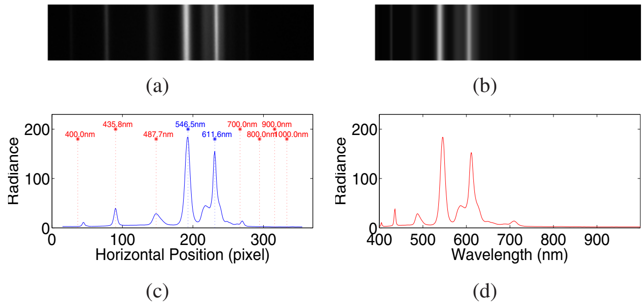



Figure 6. Spectrum Calibration. (a) Captured spectrum of fluorescent light on a white surface; (c) The radiance-displacement curve with control points detected (blue) and inferred (red); (b) Warped result of (a); (d) the intensity-position curve of (b), which is consistent with the ground truth spectrum of fluorescent light.

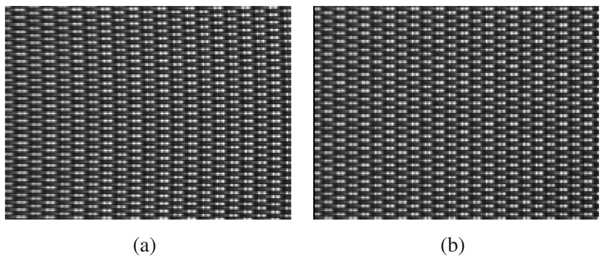



Figure 7. Geometric Calibration. (a) Captured image of a white surface illuminated by a fluorescent light, where the regular mask pattern (the scene) is slightly distorted by the prism of our camera system; (b) The resultant image after unwarping based on the correspondence between regular known mask-hole positions and the captured spectra peaks that correspond to wavelength $546.5nm$ of the fluorescent light.

Specifically in our implementation, we unwarp using a triangular mesh with the detected spectrum peaks as control points. This is also a one-time calibration and the recovered unwarping function can be applied to any frame captured in the future. Figure 7(b) shows the geometrically unwarped image of Figure 7(a).

### 4.3. Spectral Radiance Recovery

Once the spectrum and geometry distortion are calibrated, we then recover the true spectral radiance at each wavelength up to a constant scale factor. Let $c(\lambda)$ be the CCD spectrum sensitivity at wavelength $\lambda$, which is generally obtainable in CCD documentation, and let $f()$ be the CCD response curve, which is linear for our sensor but can be calibrated if unknown. Consider a pixel $x$ that covers a spectral range $[\lambda_a, \lambda_b]$ according to the distortion curve we obtained from spectrum calibration. The relationship between pixel intensity $I(x)$ at $x$ and the spectral radiance density $l(\lambda)$ is

$$I(x) = f\Big(\int_{\lambda_a}^{\lambda_b} c(\lambda) l(\lambda) d\lambda\Big). \quad (8)$$

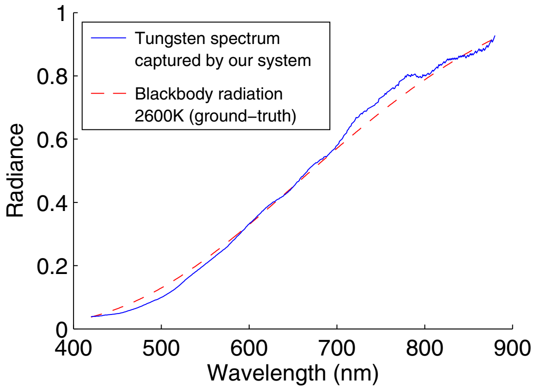



Figure 8. Spectrum radiance calibration.

Since the CCD samples a narrow spectrum at each pixel, we approximate $c(\lambda)$ and $l(\lambda)$ as being locally constant for $\lambda \in [\lambda_a, \lambda_b]$:

$$l(\lambda_s) \approx \frac{f^{-1}(I(x))}{c(\lambda)(\lambda_b - \lambda_a)}. \tag{9}$$

Figure 8 verifies the accuracy of the spectral radiance recovery. We captured the spectrum of a tungsten light using our system. By going through the spectrum calibration and spectrum radiance recovery procedure, we obtained the measured spectrum shown as the blue solid curve. The red dashed curve shows the blackbody radiation curve of 2600K, which serves as the ground truth for tungsten light. In this prototype, we did not consider the spectral effects of the glass that holds the mask and the lens, and conjecture that the slight differences between the recovered spectrum and the ground truth spectrum could be caused by them.

## 5. Results and Applications

To test the prism-based multispectral video imaging system, we captured multispectral videos of several scenes, and used these videos in different applications including human skin detection, material discrimination, and RGB video generation with illumination identification. The high resolution spectral measurements from our device give a detailed representation of surface colors, which provide more reliable cues for identifying scene objects/lighting than from traditional grayscale and color cameras.

**Human Skin Detection** Figure 9(a) shows frames from a dynamic scene that includes a genuine human hand and a fake one. The fake hand shown on the left is a color image printed on paper. Although both hands exhibit similar intensities in both the RGB and grayscale domains, in actuality the real hand exhibits a distinctive spectral feature of human skin, specifically a peak of 30nm width centered at the 559nm in the spectrum [1]. By contrast, the inks of the printed skin do not have this characteristic. Here, we exploit this special feature of human skin to distinguish the real hand from the fake. Effective identification of this spectral feature requires fine spectral sampling around 559nm. To obtain this fine sampling without greatly sacrificing spatial resolution, we place a band-pass filter (500nm-620nm) in front of the prism and use a 25mm lens. With this configuration, the spatial resolution of each video frame is $52 \times 27$, and video is captured at 7.5 fps. Each imaged scene point yields about 60 spectral samples that are approximately 2nm apart in the filtered band.

Figures 9(c) and 9(e) illustrate two sample frames captured by our system. Figure 9(g) shows the spectra of genuine (left) and fake (right) hand pixels marked in Figures 9(c) and 9(e), and the differences due to the special skin feature are clearly visible. To detect skin pixels based on the existence of the characteristic peak, we employ a simple method that thresholds the quantity $\big(r(559) - (r(544) + r(574))/2\big)$, where $r(\lambda)$ denotes the spectral radiance at wavelength $\lambda$. Figures 9(d) and 9(f) indicate the detected skin pixels in the two video frames.

**Material Discrimination** Our device can also capture high resolution spectral samples in the near-infrared band (700nm-1000nm), unlike existing infrared cameras [9, 7] that obtain only a single measurement over the entire infrared spectrum. These spectral samples outside the range of visible wavelengths are often useful for identifying materials that are difficult to distinguish in the visible spectrum or with a single infrared intensity.

Figure 10(a) displays a scene consisting of strokes painted with either water color or poster color. All of the strokes have similar RGB colors and are difficult to distinguish based on their appearance in the visible spectrum. As shown in Figure 10(b), the strokes drawn with the two different materials are still indistinguishable when also considering the overall infrared intensities captured by a typical infrared camera. To capture high resolution spectral samples in both the visible and infrared bands, we placed an infrared filter in front of the prism and captured multispectral video using a 25mm lens. This results in a $50 \times 44$ video at 7.5fps, where each imaged scene point has about 100 spectral samples in the range 400nm-1000nm, which yields a 6nm resolution in the spectral domain. Figure 10(c) shows a video frame captured by our multispectral camera. Figure 10(e) illustrates the infrared reflectance curves from samples of the two materials as marked in 10(c), where the spectral differences between the water color and poster color in the infrared band are obvious. Figure 10(d) highlights strokes in the scene that are identified as being water color, based simply on thresholding infrared spectral samples using $\big(r(750) - (r(650) + r(800))/2\big)$, where $r(\lambda)$ is the spectral radiance at wavelength $\lambda$.

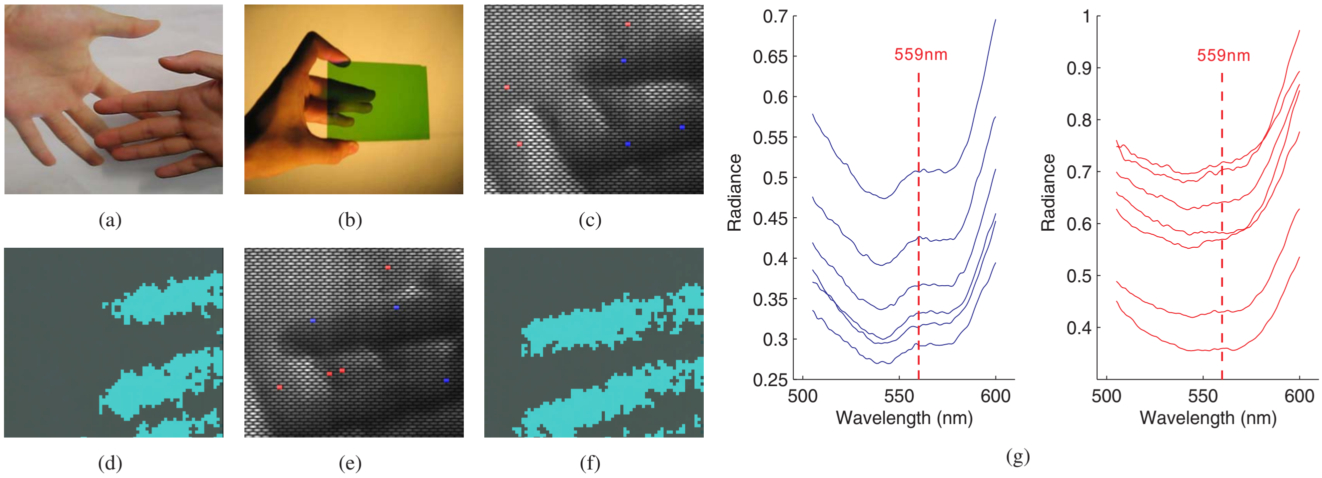



Figure 9. Human skin detection in video based on its characteristic peak at 559nm. (a) The scene includes a printed hand (left) and a real hand (right). From the displayed RGB image, it is difficult for computer vision methods to identify which hand is genuine. (b) The band-pass filter (500-620nm) used with our system in this application. (c,e) Two original video frames captured by our system. (d,f) The detected genuine human hand. (g) Left: spectra of the blue pixels in (c,e) that are sampled from the genuine human skin; Right: spectra of the red pixels in (c,e) that are sampled from the fake hand.

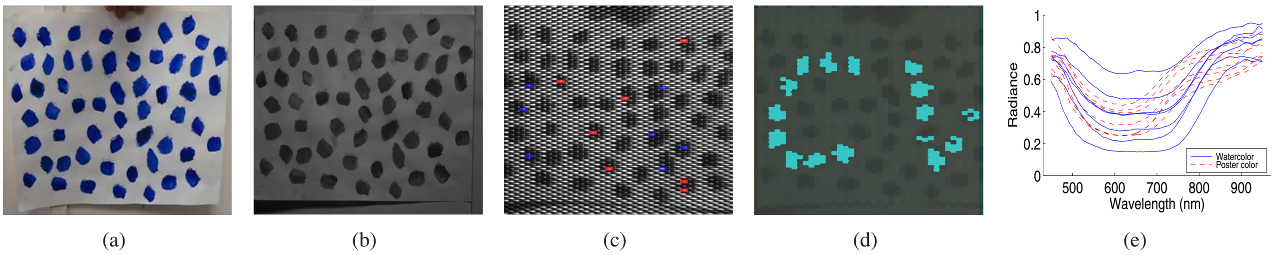



Figure 10. Discriminating physical materials based on infrared spectrum. (a) A scene consisting of strokes painted with either blue water color or poster color, which are indistinguishable in the visible spectrum. (b) An image captured by the camera with an infrared filter, where the two different color materials are still indistinguishable based on only a single overall infrared intensity. (c) An image captured by our multispectral device. The infrared spectrum is sampled at a high resolution. (d) Identified water color strokes detected by their infrared spectra. (e) The infrared spectra for samples of the two materials marked in (c). The spectral differences of the materials are clearly visible.

**RGB video generation and illumination detection** From the multispectral video captured with our camera, it is straightforward to synthesize normal RGB video according to a given set of RGB filter responses. Meanwhile, the multispectral information can be used to identify the type of illumination in the scene, which can aid in processing the RGB video such as for automatic white balance. Figure 11(a) and 11(f) display scenes illuminated with fluorescent and tungsten light respectively. To capture multispectral video of these scenes with higher spatial resolution, we used a 16mm lens and obtained a $121 \times 98$ video at 4fps (for fluorescent illumination) and 12fps (for tungsten illumination). This gives about 50 spectral samples per scene point in the visible band (400nm-700nm).

Although the spectral resolution in this case is relatively small with a short focal length of 16mm, it is sufficient for generating RGB video and identifying the illumination type. We recover the RGB values of each pixel by integrating its spectral samples according to the spectral sensitivity curves of a standard color CCD (CCD with RGB filters in a Bayer pattern), and then adjusting the exposure and white balance as done in a normal RGB camera. The resulting RGB video frames are shown in Figures 11(b), 11(c), 11(d), 11(g), 11(h) and 11(i).

Figures 11(e) and 11(j) illustrate the average spectrum of all pixels in the color rectangles marked in the resulting sample frames. Note that the average spectrum of the scene under fluorescent illumination contains sharp peaks, which is a characteristic of fluorescent light. By contrast, the spectra of the scene under tungsten illumination are relatively smooth.

## 6. Conclusion

In this paper, we proposed a prism-based multispectral video acquisition system. In this system, light rays arriving from the scene are sampled by the occlusion mask and then

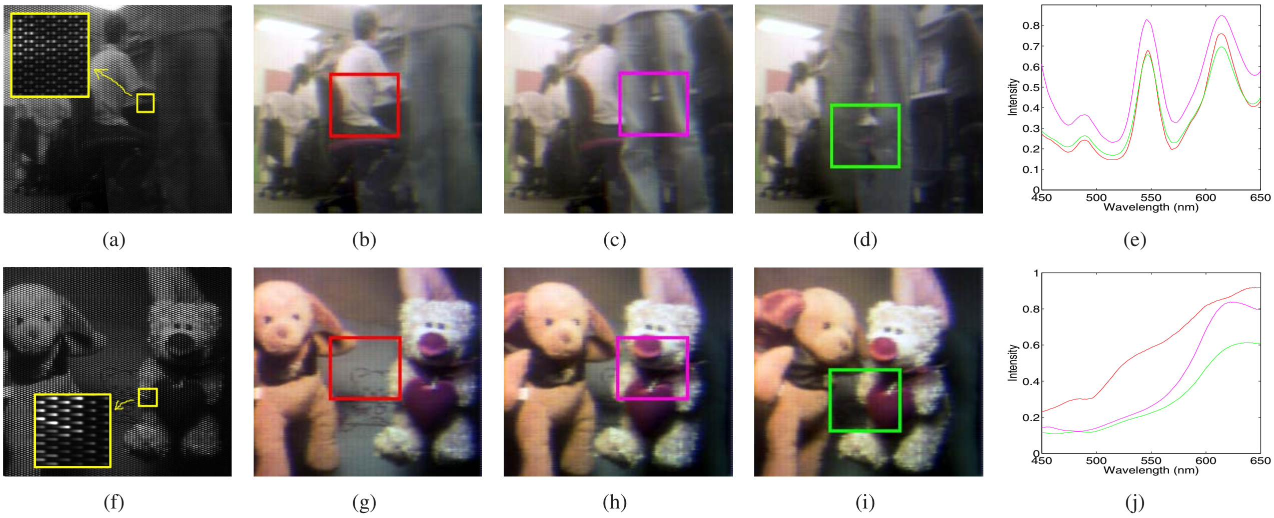



Figure 11. RGB video generation and illumination identification. The first row shows frames captured from the scene under fluorescent illumination. The second row shows frames captured under tungsten illumination. (a)(f) Rectified multispectral video frames captured by our device. (b,c,d)(g,h,i) Resulting RGB video frames generated from the multispectral video. (e)(j) The average spectra of pixels within the rectangles marked in (b)(c)(d).

dispersed by the prism. The spatially dispersed spectra are captured by a monochrome camera. The tradeoff between spatial resolution and spectral resolution can be easily modified by adjusting the focal length of the monochrome camera. Combined with bandpass filters, the new system can provide high resolution samples in the spectral band of interest while limiting the loss of spatial resolution.

One limitation of our system is that the small aperture and occlusion mask limits light flow and thus reduces the signal-to-noise ratio in the captured frames. This issue could partially be alleviated by averaging the spectra of neighboring pixels that share the same mask hole. Another limitation of our system is that the aggregate spatial and spectral resolution is limited by the CCD resolution, which makes it difficult to obtain multispectral video with both high spatial resolution and high spectral resolution. We hope that this problem will diminish in the future with the rapid increase in CCD and CMOS sensor resolutions.

## Acknowledgements

The authors thank Yue Xu for assistance in experimentation, Moshe Ben-Ezra for helpful discussions on implementation issues, and the reviewers for their constructive comments.

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

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

[4] N. Gat. Imaging spectoscopy using tunable filters: A review. In *SPIE Wavelet Applications VII*, volume 4056, 2000.

[5] A. R. Harvey, D. W. Fletcher-Holmes, and A. Gorman. Spectral imaging in a snapshot. In *SPIE*, volume 5694, 2005.

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

[7] K. Kidono and Y. Ninomiya. Visibility estimation under night-time conditions using a multiband camera. In *IEEE Intelligent Vehicles Symposium*, 2007.

[8] A. Mohan, R. Raskar, and J. Tumblin. Agile spectrum imaging: Programmable wavelength modulation for cameras and projectors. *Computer Graphics Forum*, 27(2):709–717, 2008.

[9] N. J. W. Morris, S. Avidan, W. Matusik, and H. Pfister. Statistics of infrared images. In *CVPR*, 2007.

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

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

[12] Spectral Imaging LTD. Spectral cameras. [online] http://www.specim.fi.

[13] C. Vandervlugt, H. Masterson, N. Hagen, and E. L. Dereniak. Reconfigurable liquid crystal dispersing element for a computed tomography imaging spectrometer. In *SPIE*, volume 6565, 2007.

[14] 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. In *SPIE*, volume 6062, 2006.