Generalized Mosaicing: Wide Field of View Multispectral Imaging

Yoav Y. Schechner and Shree K. Nayar

Abstract—We present an approach to significantly enhance the spectral resolution of imaging systems by generalizing image mosaicing. A filter transmitting spatially varying spectral bands is rigidly attached to a camera. As the system moves, it senses each scene point multiple times, each time in a different spectral band. This is an additional dimension of the generalized mosaic paradigm, which has recently demonstrated yielding high radiometric dynamic range images in a wide field of view, using a spatially varying density filter. The resulting mosaic represents the spectrum at each scene point. The image acquisition is as easy as in traditional image mosaics. We derive an efficient scene sampling rate, and use a registration method that accommodates the spatially varying properties of the filter. Using the data acquired by this method, we demonstrate scene rendering under different simulated illumination spectra. We are also able to infer information about the scene illumination. The approach was tested using a standard 8-bit black/white video camera and a fixed spatially varying spectral (interference) filter.

Index Terms—Multispectral, hyperspectral imaging, color balance, enhancement, image fusion, physics-based vision, panorama, mosaic, mosaicing, illumination, image-based rendering.

1 Introduction

Image mosaicing is a popular way to obtain a wide field of view (FOV) image of a scene. In this approach, images are captured as a camera moves, and are then stitched to obtain a larger image. Image mosaicing has found applications in consumer photography [7], [8], [9], [18], [20], [28], [34], [38], [39], [43], [50]. It has also been used in various scientific fields, such as astronomy [26], [36], [52], remote sensing [6], [13], [24], [46], [53], and underwater imaging [17]. As depicted in Fig. 1, traditional image mosaicing mainly addresses the extension of the FOV (the spatial resolution may also be somewhat improved by mosaicing [7]), while other imaging dimensions are not improved in the process. In [40], we introduced the notion of *generalized mosaicing* in which much more information about the scene is extracted, given a similar amount of acquired data. The objective of this paper is to detail how mosaicing can be used for enhancement of the spectral information.

A typical video sequence acquired during mosaicing has great redundancy in terms of the data it contains, since each point is observed multiple times. In our approach, we rigidly attach to the camera a fixed filter with spatially varying properties, as in the setup shown in Fig. 2. As the camera moves, each scene point is measured under different optical settings. This simple optical filtering significantly reduces the redundancy in the captured video stream. In return, the filtering embeds in the acquired data more information about each point in the mosaic FOV.

In [3], [40], [41], this observation was exploited to extend the dynamic range of the camera. In this paper, we demonstrate how this concept can be used to easily obtain multispectral information on each scene point with ordinary cameras. This is achieved using a filter in which the transmitted spectral band varies spatially. This reinforces the generality of the approach introduced in [40]. For any given camera, a simple replacement of the mounted filter enables the extension of different imaging dimensions. Multispectral imaging is thus obtained in the same unified framework that obtained high dynamic range and wide FOV imaging. We use the gathered data for obtaining information about the scene illumination and for scene rendering under arbitrary illumination spectra.

In the following sections, we describe multispectral data acquisition in a wide FOV using the generalized mosaicing approach. We compare it to other approaches for multispectral imaging. We detail the configuration we used, and describe a criterion for efficient sampling of the scene using this configuration. Fig. 3 shows prototype systems we developed. We demonstrate multispectral mosaicing by standard 8-bit monochrome video camera and an external spatially varying spectral filter.

2 Spectral Imaging: Previous Approaches

Multispectral imaging has proven extremely useful in numerous imaging applications, including object and material recognition [44], color analysis and constancy [1], [19], [47], remote sensing [30], [44], [48], [55], and astronomy [32]. The fields which find applications for multispectral imaging are growing in number [15], for example, medical imaging, agriculture, archaeology, and art. Multispectral images are typically acquired by specialized systems. This section describes the main methods used.

The most common method for obtaining "spectral" information is by covering the detector array by a mosaic-array of red, green, and blue filters, leading to three samples of the spectrum integrated over very broad bands [35]. This method trades spatial resolution for spectral resolution, as depicted in the spectral-spatial space in Fig. 4. Mosaicing such

• *The authors are with the Department of Computer Science, Columbia University, 1214 Amsterdam Ave., New York NY 10027. E-mail: {yoav, nayar}@cs.columbia.edu.*

Manuscript received 16 Feb. 2001; revised 9 Dec. 2001; accepted 12 Feb. 2002. Recommended for acceptance by P. Anandan.
For information on obtaining reprints of this article, please send e-mail to: tpami@computer.org, and reference IEEECS Log Number 113643.

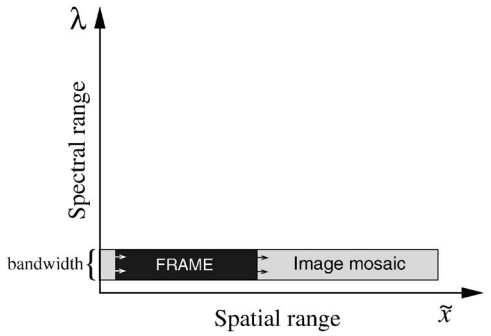

Fig. 1. By stitching partly-overlapping frames a traditional image mosaic extends the field of view of any camera. However, other imaging dimensions, such as the spectral quality, are usually not improved.

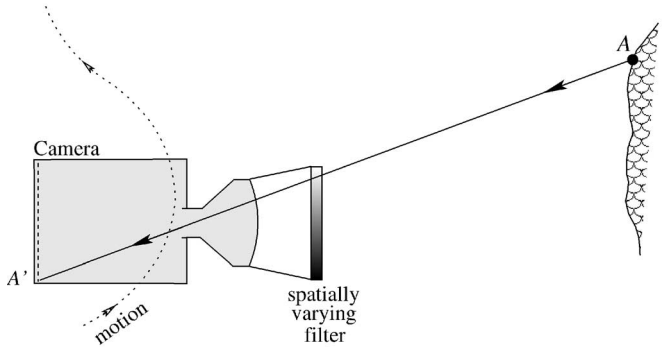

Fig. 2. Scene point A is imaged on the detector at A' through a spatially varying filter attached to the camera. As the imaging system moves, each scene point is sensed through different portions of the filter, thus multiple measurements are obtained under different optical settings.

images extends the FOV, and even enables superresolution in the spatial domain, compensating for the lower sensor spatial resolution [7]. However, the spectral resolution remains low.

Higher spectral resolution is achieved by imaging the scene sequentially with temporally varying optical settings. The images are taken with a static camera, hence with a fixed FOV (see Fig. 5a). Realizations of this approach include changing discrete filters (by a filter wheel) and imaging spectrometers based on tunable filters [15], [32], [23], [54]. The latter are based on interference or on birefringence. This category of methods also includes the use of dichroic beamsplitters (e.g., 3-CCD cameras) [10], [35], [55]. In that method, however, the images are acquired in parallel by several detector arrays. It also includes Fourier-transform

Fig. 3. Two generalized mosaicing systems. (a) A system composed of a Sony black/white video camera and an extended arm which holds the filter. (b) A system that includes a Canon Optura digital camera and a cylindrical attachment that holds the filter. In both cases, the camera moves with the attached filter as a rigid system.

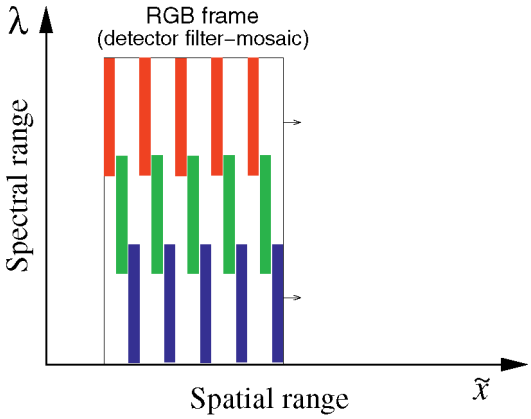

Fig. 4. In most RGB cameras, spatial resolution is traded for the spectral resolution. Traditional mosaicing of such images extends the FOV and enables super *spatial* resolution, but the spectral resolution remains very low.

spectrometers [25], [54] in which after all the measurements have been acquired, they are transformed in the spectral domain to obtain the spectrum at each point. Note that there is an analogy between scanning the FOV and scanning the spectral range. Each is dealing with a different dimension of the spectral-spatial space depicted in Figs. 1, 4, 5, 6, and 7. While traditional mosaicing addresses the spatial dimension, the spectral scanning depicted in Fig. 5a addresses the spectral dimension.

Multispectral data is obtained in an extended FOV using pushbroom [55] imaging spectrographs, which are generally rather complex and expensive. Such systems are used especially in aerial remote sensing. Imaging spectrographs view the scene through a narrow slit, say, perpendicular to the frame's x axis. The light coming through the 1D slit is dispersed by prisms or diffraction gratings [32], [47], [55] onto a 2D detector array, hence high-resolution spectral information is obtainable for a swath (1 pixel wide column) of the scene.[1] The multispectral image is obtained by scanning the scene in 1 pixel increments along the $\tilde{x}$-axis, as depicted in Fig. 5b. This category of methods also includes line-scanners (e.g., desktop RGB scanners), in which the image is scanned with several 1D arrays [31], each sensing the light in a different broad spectral band. A generalization of the pushbroom dispersive method enables the measurement of an arbitrary set of points. This is obtained by first imaging the scene onto a plane. Optical fibers [54] placed on selected image points guide the light to form a "slit" bundle of fibers which is then analyzed by the imaging spectrograph.

There is a way to combine the principles of the pushbroom scan with varying interference filtering. Rather than temporally tune the transmitted band, the spectral filter is spatially varying. Such a filter can be placed on, or optically relayed onto, the detector array, as in [14], [27], [30], [48], [55]. If the variations are along the x axis, each pixel in the camera FOV senses a different spectral band, according to its x coordinate. Multispectral images are obtained as the camera moves. That configuration is similar to ours, but more complicated to implement. It is also much less adaptive since the filter is mounted inside the inner parts of the imaging

1. The pushbroom dispersive approach is a derivative of whiskbroom systems in which each point of the scene is serially measured by a spectrometer.

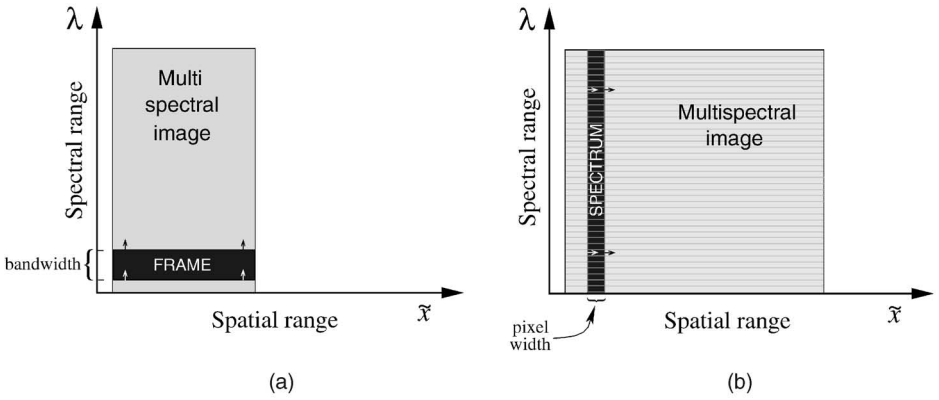

Fig. 5. Previous approaches to multispectral imaging as represented in the spectral-spatial space. (a) Images are taken with a static camera with temporally varying optical settings have enhanced spectral range and resolution in a limited FOV. (b) Pushbroom dispersive systems capture the spectrum for a 1-pixel wide 1D swath (column) of the scene. Moving the system in 1 pixel increments perpendicular to that swath scans the scene in an arbitrarily wide FOV.

system. In contrast, in our system the filter is external, thus can be easily changed and mounted on any camera. Moreover, the user has the flexibility of traditional imaging simply by removing the filter.

3 Mosaicing in the Spectral-Spatial Space

In [40], [41], we used the generalized mosaicing approach for high dynamic range mosaicing. Here, we use the same method to obtain enhanced spectral resolution with an ordinary camera. In the current configuration, we mount on the camera an external filter (as in Figs. 2 and 3), which transmits a spectral band that varies across it. Out of the unlimited theoretical possibilities for such filters, we mainly consider the *linear variable interference filter* [11]. Here, the transmitted band is concentrated around a central wavelength which changes linearly across[2] the filter (its x axis).

Now, let the scene be scanned by the camera. As in traditional mosaicing, the camera can have a general motion. For example, it can be hand held during the scan, or rotated by a motorized turntable (Section 6.1 discusses the implications of parallax in general motion). The moving system filters the light color from any scene point differently in the different parts of the frames. Thus, in each frame, the light coming from any given scene point is filtered differently. For instance, image points are captured through the part transmitting the "red" wavelengths, then through the part transmitting the "green" wavelengths, etc. Therefore, mosaicing images taken with such a filter enables multispectral measurements of each scene point, while extending the FOV. This improves the sampling of the plenoptic function [2]. As seen in Fig. 6, this may be viewed as another dimension of the mosaicing process.

In addition, information becomes available about the periphery of the central region of interest. Images and partial spectral information are obtained on the periphery, though with a gradually narrowing spectral range (Fig. 7). Such a structure is analogous to foveated images, in which the acquisition quality improves from the periphery towards the center of the FOV. The periphery is at most one frame wide and is eliminated in 360^o panoramic mosaics.

2. For simplicity, we assumed filter variations along one spatial dimension. The results can be generalized to 2D filter variations.

4 Illumination at a Glance

When using a spatially varying filter, the spectral information in each raw frame is multiplexed with the spatial features which appear in ordinary images. Consider the images acquired through the filter which are shown in Fig. 8. These images were taken through an off-the-shelf linear variable interference filter [11] with a commercial monochrome camera. The spectral range of the filter spans the visible light[3] wavelengths 400-700 nm. The spatial details of the scene are clearly recognizable (e.g., the computer monitor). Recognizing spatial content in each frame is not possible when using dispersive imaging spectrographs (whose principle of operation is depicted in Fig. 5b). The spatial features are clear because, as depicted in Fig. 2, the system is an imaging device which captures an area of the scene.

At the same time that the spatial features are sensed, the frames are dominated by vertical bands which appear at the same places in all the images. The reason for this phenomenon is as follows: Let the illumination spectrum be $S_{\text{source}}(\lambda)$ and the spectral response of the system be $l_{\text{system}}(\lambda)$. Let (x, y) be the coordinates of a pixel in a frame, as depicted in Fig. 8 and the spectral reflectance of the scene point seen in pixel (x, y) at frame k be $r_k(x, y, \lambda)$. The detected intensity at pixel (x, y) is

$$g_k(x, y) = \tilde{S}_{\text{source}}(\lambda) r_k(x, y, \lambda), \tag{1}$$

where

$$\tilde{S}_{\text{source}}(\lambda) \equiv S_{\text{source}}(\lambda) l_{\text{system}}(\lambda) \tag{2}$$

is the illumination spectrum, modulated by the system response.

Since the central wavelength passed by the filter changes linearly with x, we may make the transition:

$$\lambda \sim x. \tag{3}$$

For simplicity, we omitted shift and scaling factors and assume for the moment that the filter pass band is a δ-function (i.e., having a very narrow bandwidth). Thus, we substitute λ by x:

$$S_{\text{source}}(\lambda) \sim S_{\text{source}}(x) \qquad l_{\text{system}}(\lambda) \sim l_{\text{system}}(x). \tag{4}$$

3. In other filters, the spectral range may include infrared or UV light.

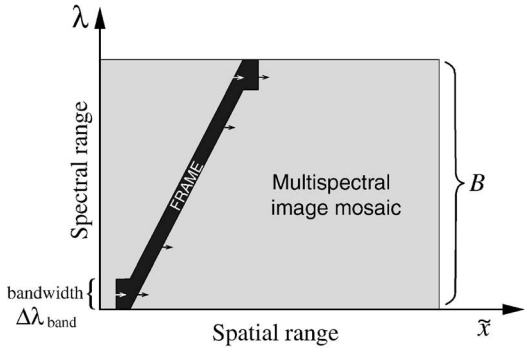

Fig. 6. Image mosaicing coupled with exploiting space-varying spectral filtering yields multispectral image mosaics. Besides the field of view, this process also enhances the spectral resolution of the camera.

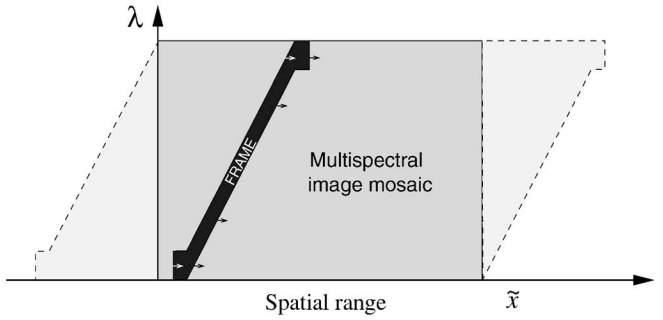

Fig. 7. Outside the main region of interest, the mosaic provides additional information on the scene periphery.

and

$$r_k(x, y, \lambda) \sim r_k(x, y). \tag{5}$$

Therefore,

$$g_k(x, y) = \tilde{S}_{\text{source}}(x) r_k(x, y). \tag{6}$$

Now, the system response l_{system} is the same in all the frames. Also, the illumination spectrum S_{source} is typically similar in all the frames. Therefore, wavelengths (i.e., x-coordinates) at which the illumination energy distribution $\tilde{S}_{\text{source}}$ is low, manifest as dim lines parallel to the y-axis (vertical). Wavelengths where the illumination is strong, typically manifest as bright vertical lines (unless an object greatly absorbs these wavelengths, i.e., $r_k(x, y, \lambda)$ is very small).

The filter pass band is not a δ-function, but has a finite effective width, as we discuss in Section 5. Therefore, the bright/dark lines broaden to bands. The bright bands in Fig. 8 are thus due to the characteristic spectral peaks of the fluorescent lamp [49] which illuminated the scene. On the other hand, these peaks do not appear in the sequence frames taken when the illumination was by incandescent lamps. Samples of that sequence are shown in Fig. 9. The reason is that the incandescent illumination (essentially having a black body spectral distribution) is unimodal and wide band.

A useful implication of this phenomenon is an ability to infer the kind of the scene illuminant. Usually, the illuminant spectrum can be measured if a white object (e.g., a calibration chart) exists in the FOV. Nevertheless, in the absence of such an object, e.g., when the scene is not controlled, one may still get an estimate of the illuminant. By averaging over the y axis and over the imaged frames, we obtain the average horizontal profile

$$M(x) \propto \sum_{k=1}^{\text{frames}} \sum_{y} g_k(x, y). \tag{7}$$

Using (6)

$$M(x) \propto \tilde{S}_{\text{source}}(x) \langle r(x) \rangle, \tag{8}$$

where $\langle r(x) \rangle$ is the average horizontal profile of the reflectance,

$$\langle r(x) \rangle \propto \sum_{k=1}^{\text{frames}} \sum_{y} r_k(x, y). \tag{9}$$

Recalling (3), (4), and (5),

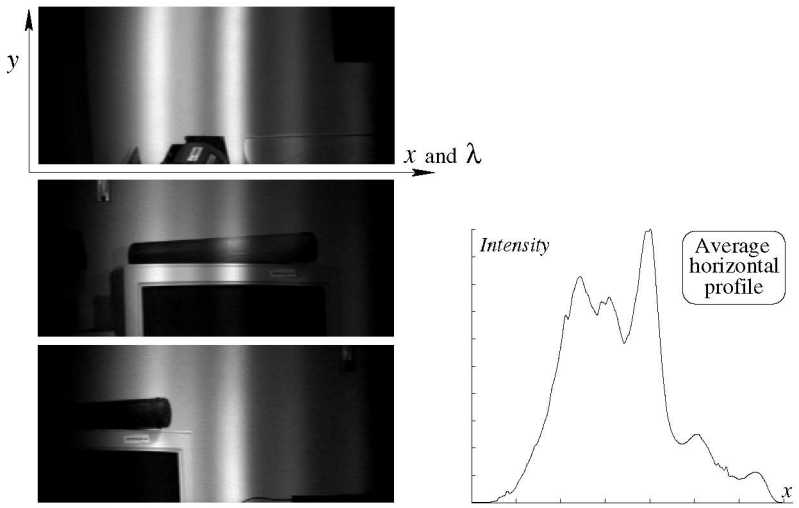

Fig. 8. Images taken with the linear variable interference filter. The left of the FOV senses the energy density at $700nm$, while the right senses it at $400nm$. Besides the spatial features of the scene, its *fluorescent illumination* is revealed by the typical spectral peaks [49], seen as bright columns appearing in the same places in all the images. The peaks appear also in the horizontal profile averaged over five frames in the sequence.

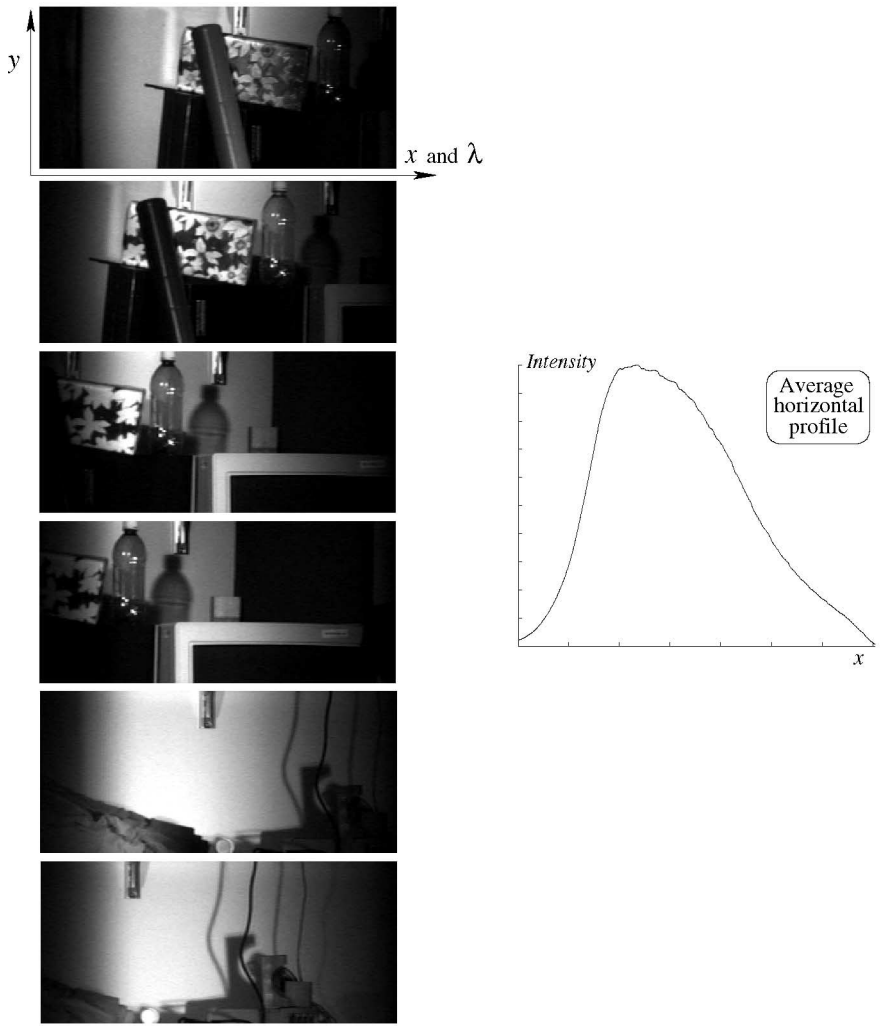

Fig. 9. Frames 29, 32, 38, 40, 60, and 64 of a sequence taken through the linear variable interference filter. As in Fig. 8, the spatial features of the scene are clearly seen. Here, the scene has *incandescent illumination* which has a unimodal, wide band spectral distribution. This leads to a wide bright region appearing in the same area in all the images. This unimodal wide band appears also in the horizontal profile averaged over the sequence.

$$M(x) \sim \tilde{S}_{\text{source}}(\lambda)\langle r(\lambda)\rangle, \tag{10}$$

where $\langle r(\lambda)\rangle$ is the average reflectance of the scene at each wavelength. Note that if the scene is gray *on average*, $\langle r(\lambda)\rangle = Constant$, then the average horizontal profile $M(x)$ is similar to the spectral density of the illumination (modulated by the system response). Indeed, the characteristics of the spectral distributions the fluorescent and incandescent illuminations appear in the average horizontal profiles in Figs. 8 and 9, respectively.

Therefore, just a few frames may imply cues about the illumination type and spectral distribution, while each frame supplies information about the 2D FOV. Had we used a traditional imaging spectrometer based on tunable filers or dichroic mirrors [55], the entire set of frames had to be scanned before global information about the spectrum could be estimated.

5 Efficient Sampling Criterion for Still Images

If still images are acquired rather than a video stream, we cannot assume that the frame displacements are small. The question then is what should the frame displacements be or how many times should each scene point be seen? Let the spectrum be sampled by the moving system at periodic intervals $\Delta\lambda_{\text{sample}}$. Let $B \equiv \lambda_{\max} - \lambda_{\min}$ be the total spectral bandwidth of the filter (see Fig. 6), where $\lambda_{\max}$ and $\lambda_{\min}$ are the maximum and minimum wavelengths that the entire filter passes, respectively. Each point is sampled

$$\#_{\text{samples}} = \lceil B/\Delta\lambda_{\text{sample}} \rceil \tag{11}$$

times.

We estimate the sampling intervals using frequency domain analysis. A sufficient sampling of the spectral intensity distribution depends on how fast it fluctuates as a function of wavelength. This is similar to sampling of other signals: if the samples are too sparse to sense the signal fluctuations, aliasing occurs. The spectral intensity distribution of the light coming from the scene is smoothed by the system since the transmitted band is not a δ-function of wavelength. Let us model the smoothing as a convolution of the spectrum $S(\lambda)$ with a window function $h(\lambda)$ of width $\Delta\lambda_{\text{band}}$. Let $\mathcal{S}(\xi_\lambda)$ and $\mathcal{H}(\xi_\lambda)$ be the Fourier transforms of $S(\lambda)$ and $h(\lambda)$, respectively. $\mathcal{H}$ is a lowpass filter. It has an effective

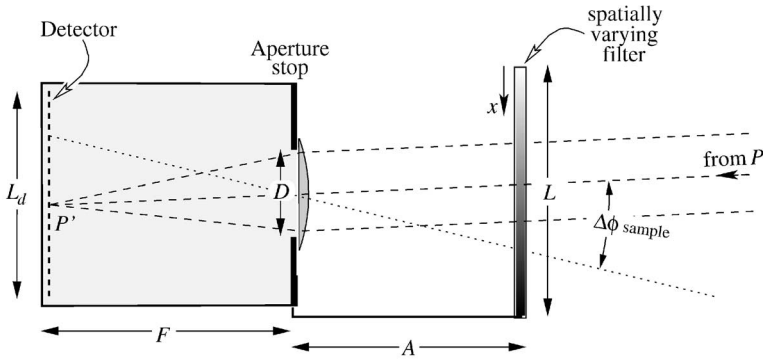

Fig. 10. Light coming from the distant object point P is focused on the detector in image point P'. The detector of length L_d is inside a camera having focal length F. The lens aperture diameter is D. The filter of length L is attached by an arm of length A to the camera. The spectral band transmitted by the filter varies as a function of the position on the filter, x. The camera may change its viewing direction by $\Delta\phi_{\text{sample}}$ between consecutive image frames.

cutoff frequency[4] which is $\approx 1/\Delta\lambda_{\text{band}}$. Therefore, in analogy to the Nyquist theorem, the required sampling interval is

$$\Delta\lambda_{\text{sample}} \approx \Delta\lambda_{\text{band}}/2. \tag{12}$$

Therefore, each point should be sampled $\lceil 2B/\Delta\lambda_{\text{band}} \rceil$ times.

$\Delta\lambda_{\text{band}}$ can be estimated using monochromatic light sources. Nevertheless, to get an intuitive indication, we will express the sampling criterion in terms of the physical dimensions of the system. Consider the system depicted in Fig. 10. If the detector is focused at infinity, then the external filter is blurred by a kernel having a width that is equal to the width of the object light beam that passes through the entrance pupil. For the simple lens camera depicted in Fig. 10, the filter defocus blur has kernel of width

$$\Delta x_{\text{blur}} = D, \tag{13}$$

where D is the diameter of the entrance pupil (the aperture in a simple lens system). For a linear variable filter of length L, the central wavelength of the pass band changes linearly across the filter

$$\lambda(x) = \frac{\lambda_{\max} + \lambda_{\min}}{2} + \frac{B}{L}x, \quad x \in \left[-\frac{L}{2}, \frac{L}{2}\right], \tag{14}$$

Therefore, a geometric window of width Δx_{blur} is equivalent to a spectral window of width

$$\Delta\lambda_{\text{blur}} = \frac{B\Delta x_{\text{blur}}}{L} = \frac{BD}{L}. \tag{15}$$

Even if no defocus blur of the filter exists (as when the lens aperture is very small), the filter pass band is not a δ-function. Approximating the inherent pass band as a Gaussian with standard deviation $\Delta\lambda_0$, and the defocus blur kernel also to be Gaussian, the effective width of the pass band including all effects is

$$\Delta\lambda_{\text{band}} = \sqrt{\Delta\lambda_{\text{blur}}^2 + \Delta\lambda_0^2} = \sqrt{\left(\frac{D}{L}\right)^2 B^2 + \Delta\lambda_0^2}. \tag{16}$$

Combining (16) with (11) and (12), the scene point should be sampled

$$\#_{\text{samples}} \approx \left\lceil \frac{2}{\sqrt{(D/L)^2 + (\Delta\lambda_0/B)^2}} \right\rceil \tag{17}$$

times. This result is intuitively reasonable: If the filter is long, $L \gg D$, the blur-kernel is insignificant. If in addition the filter inherent pass bands are very narrow relative to the entire spectral range, $\Delta\lambda_0 \ll B$, then the fine details of the original spectrum (which may have sharp peaks) remain and, thus, the sampling should be dense. Therefore, the sampling period becomes small and the number of samples (17) becomes large. On the other hand, if the filter is very short and, thus, severely blurred or if the filter has an inherent wide band $\Delta\lambda_0$, then the spectral resolution of the system is very low. Then, just a few samples of each scene point are sufficient to capture the available information.

Although the system motion can be general, we may get a further feeling of the sampling implications by concentrating on the case of rotation about the system's center of projection. Let the filter be situated at a distance A from the rotation axis. Suppose that the system changes its viewing direction by an angle increment of $\Delta\phi_{\text{sample}}$ between image acquisitions (see Fig. 10). Each light ray coming from the scene to the center of projection moves *along the filter* by $\Delta x_{\text{sample}} \approx A\Delta\phi_{\text{sample}}$, assuming the camera FOV to be small. Referring to (14),

$$\Delta\lambda_{\text{sample}} \approx \frac{B}{L}\Delta x_{\text{sample}} \approx \frac{B}{L}A\Delta\phi_{\text{sample}}. \tag{18}$$

Combining (12), (16), and (18),

$$\Delta\phi_{\text{sample}} \approx \frac{D}{2A}\sqrt{1 + \left(\frac{\Delta\lambda_0}{B}\right)^2 \left(\frac{L}{D}\right)^2}, \tag{19}$$

When the interference filter has a very narrow inherent pass band,

$$\Delta\phi_{\text{sample}} \approx \frac{D}{2A} = \frac{F}{2Af_{\#}}, \tag{20}$$

where $f_{\#} \equiv F/D$ and F is the focal length. Interestingly, this result is irrespective of the filter dimensions or its total bandwidth. The intuition behind (20) is that the larger the aperture D, the larger the defocus blur, thus one needs less samples. Therefore, $\Delta\phi_{\text{sample}}$ increases. As seen in Fig. 10, the smaller the distance A between the filter and the center of projection, the smaller is change in λ of the chief ray, for a

4. The support of $\mathcal{H}$ may be infinite. Then, the effective cutoff may be defined as the frequency under which most of the energy of $\mathcal{H}$ is contained.

given rotation increment. This decreases the wavelength sampling period. Therefore, for the desired $\Delta\lambda_{\text{sample}}$, $\Delta\phi_{\text{sample}}$ has to increase as A decreases.

Note that the filter should not be placed right next to the lens, because this affects the global aperture properties [12] (i.e., the average spectral transmittance) without producing spatially varying effects in the image. Therefore, the images obtained are panchromatic (no spectral definition). Indeed, when $A \to 0$ then $\Delta\phi_{\text{sample}} \to \infty$ meaning that the available, poor, information about each scene pixel is contained in a single sample.

Equations (17), (19), and (20) express a nice property of this method, that does not exist when laying the filter on the detector: *The user may adjust the system to meet the desired scene scanning rate, while avoiding aliasing* simply by changing the lens aperture or the distance of the given filter from the lens. For instance, if the user would like to scan the scene quickly, thus with a few frames for the given mosaic FOV, then $\Delta\phi_{\text{sample}}$ should be large. In that case, the aperture should be opened widely and/or the filter should come closer to the lens. This ensures that although there are a few samples of the spectral intensity distribution, spectral aliasing is minimized. Note that this capability is coupled with a disadvantage of the method: if the scene radiance is small, hence requiring a large aperture D, (16) implies that the spectral resolution will decrease.[5]

Following (20), the number of frames needed for a 360^o panorama is $\#_{360^o\text{panorama}} = 2\pi/\Delta\phi_{\text{sample}}$. We may require that the filter (of length L) will occupy the entire FOV of the detector, which has a length of L_d. Then, $F/A = L_d/L$. Therefore,

$$\#_{360^o\text{panorama}} \approx \frac{4\pi f_{\#} L}{L_d}. \tag{21}$$

For example, for a $6\ cm$ long filter, a $6\ mm$ long CCD detector and an aperture stop of $f_{\#} = 5.6$, about 700 frames suffice to create panoramic mosaic, which take about 12 seconds to complete at 60 Hz frame rate. This estimate is an upper limit to the actual number of frames needed, since it neglects the inherent bandwidth of the filter, $\Delta\lambda_0$. Because $\Delta\phi_{\text{sample}}$ in (19) is larger than in (20), $\#_{360^o\text{panorama}}$ is smaller.

6 Image Registration

Once the images have been acquired, they need to be registered. A scene point has different coordinates in each image of the sequence. The measurements corresponding to this point should be identified before they can be fused. For generalized mosaicing, the image registration algorithm should cope with the phenomena induced by the filtering. For instance, the scene appearance will usually be different somewhat as the color bands change. On the other hand, this difference may be small if the motion between consecutive frames is small enough, as in most video streams. Therefore, image registration in generalized mosaicing is somewhere between traditional registration and multisensor registration [21], [42].

When still images are taken with large displacements, the spatially varying but temporally static effects of the filter become significant. Consider the images shown in Fig. 9. Although features (e.g., the bottle and monitor) appear to be moving through the camera's FOV, they seem to be modulated by a static "mask" that dominates the images. This mask is similar to the average horizontal profile $M(x)$: bright in the middle and attenuating outwards. The reason for this modulation is explained in Section 4. The apparent static mask is more prominent in Fig. 8, though it is different since the illumination spectrum S_{source} is different: alternating dark and bright vertical bands exist in the same places in all the images. The apparent unmoving mask biases traditional algorithms towards estimating a motion slower than the true one or potentially other local minima. According to our experience, highpass filtering the raw images as in [21], [42] reduced the biasing effect, but did not remove it completely.

Scene points in the spectrally filtered images may become dim due to low-illumination intensity in some spectral bands. This is analogous to measuring points that become darker due to an attenuation mask. These measurements are relatively noisy after quantization and other processes they undergo during image capture. Therefore, instead of matching the original image readouts, we use a transformed version of them that takes into account the "attenuation" dependent uncertainties. We generalized a traditional technique so it can cope with the spatially varying filtering effects and previously used it to construct high-dynamic range mosaics in [41]. We use the same method to register the spectrally filtered images. The method attempts to compensate for $M(x)$, while taking into account the noise amplification associated with this operation. During registration we treat each frame as an ordinary image (gray-level, in particular), ignoring the underlying spectral information. To make the paper self-contained, we hereby describe the principles of the method.

1. We estimate the "apparent attenuation mask" as the average horizontal profile, as in (7). Then, each frame $g_k(x,y)$ is compensated for the mask using $1/M(x)$ to estimate $I_k(x,y) = g_k(x,y)/M(x)$. The uncertainty of I_k is estimated by

$$\Delta I_k = \frac{\Delta g_k}{M(x)}. \tag{22}$$

   The intensity readouts are integers thus we set the readout uncertainty $\Delta g = 0.5$. In case a spatially varying filter is not present, $M \approx 1$, thus ΔI_k is constant.
2. Let I_1 and I_2 be the intensity measurements at candidate corresponding pixels in two images, with respective uncertainties ΔI_1 and ΔI_2. The squared Mahalanobis distance between these measurements is

$$\hat{E}^2_{\text{pixel pair}} = \left(\frac{\hat{I} - I_1}{\Delta I_1}\right)^2 + \left(\frac{\hat{I} - I_2}{\Delta I_2}\right)^2, \tag{23}$$

5. Also in dispersive spectrographs, there is a tradeoff between the light throughput and the spectral resolution, determined by the entrance slit width.

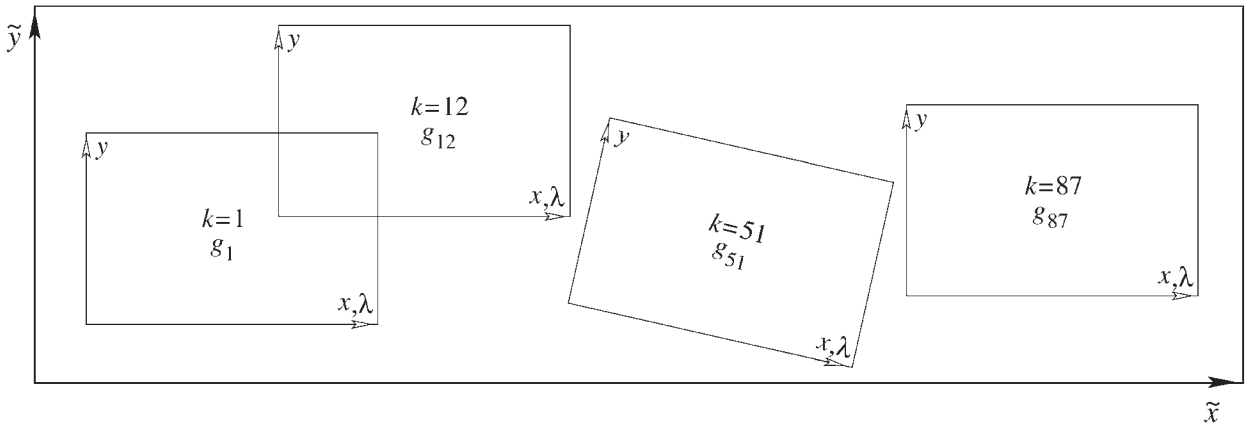

Fig. 11. $(\tilde{x}, \tilde{y})$ are the coordinates of a scene point in the global coordinate system of the mosaic. The mosaic is composed of raw frames indexed by k. The intensity in a raw frame is $g_k(x, y)$, where (x, y) are the internal coordinates in the frame. The central wavelength λ of the transmitted band is a function of x.

where

$$\hat{I} = \widehat{\Delta I}^2 \sum_k \frac{I_k}{\Delta I_k^2}, \tag{24}$$

and

$$\widehat{\Delta I}^2 = \left(\sum_k \frac{1}{\Delta I_k^2} \right)^{-1}. \tag{25}$$

Here, $k = \{1, 2\}$ since only two measurements are involved. The distance measure for the entire images is

$$\hat{E}^2_{\text{total}} = \sum_{\text{all pixels}} \hat{E}^2_{\text{each corresponding pair}}. \tag{26}$$

The best registration between two frames images according to this objective function is the one that minimizes $\hat{E}^2_{\text{total}}$. If the measurements are Gaussian and independent, this match maximizes the likelihood (ML) of the fused intensity values $\hat{I}$. When the spatially varying filter is not present, $\Delta I_1 = \Delta I_2$, hence (26) is proportional to the sum of square difference between the images. Note that $\hat{E}^2_{\text{total}}$ will generally increase with the number of pixels in the overlap area of the images. This may bias the registration towards reducing this overlap. To counter that, (26) is normalized by the number of pixels in the overlap area [41].

3. The registration is done hierarchically, from coarse to fine resolution similar to [21], [39], [42]. We create a *Maximum Likelihood pyramid*, where not only the image value is stored at each scale, but also its uncertainty. The weights used in the construction of the pyramid structure depend also on the uncertainties of the pixels in each neighborhood, so that more reliable pixels contribute more to their coarse representation. If all the uncertainties are the same, the result is the same as in a traditional image pyramid. Details on this structure are given in Appendix A.1. The representation of $I_1, I_2, \Delta I_1, \Delta I_2$ at each scale enables robust image registration by maximizing the likelihood of the match at a coarse scale and gradually in a finer scale.
4. To reduce the accumulation of matching errors, each new sequence frame is registered to the current mosaic [20], [39], and then fused into it. The image and the mosaic are fused using (24). This creates a gray-level image mosaic of the scene.

This algorithm was successful and did not suffer from the biasing problem. However, there may be situations in which it will err since the analogy between the illumination spectrum and an attenuation mask is an approximation. It assumes that the scene has correlation between reflected colors (which is usually the case). Thus, one may think of pathological cases where the scene spectrum exists only in disconnected bands (e.g., far blue and red) and where the spatial texture as seen in any wavelength band do not correlate with the texture in other bands. In such cases, the scene appearance will change wildly as the wavelength is scanned and the registration will fail. Such cases are not typical. The performance may be enhanced by combining the principles of this algorithm with robust statistics methods [4], [38] and principles used in multisensor matching [21].

6.1 A Note on Motion with Parallax

The motion of the camera in generalized mosaicing can be as general (e.g., hand held) as in traditional mosaicing. Naturally, when parallax exists, matching and registration become more challenging and techniques for "deghosting" are needed [43]. Still, if a point can be tracked throughout the sequence, its full spectrum will be measured.

As in any registration problem, occlusions cause ill posed situations. Points which become occluded in some frames of the sequence will miss some of their spectrum samples and, thus, part of their spectral measurements will be lost.

7 Color and Multispectral Mosaics

Let $(\tilde{x}, \tilde{y})$ be the coordinates of the scene points in the global coordinate system of the mosaic (see Fig. 11). Once the raw frames are registered, we have for each scene point in the mosaic FOV a set of wavelength samples $\lambda(\tilde{x}, \tilde{y}, k)$, and corresponding intensity measurements $g(\tilde{x}, \tilde{y}, k)$, where k is the index of the individual frames that compose the mosaic. This raw data structure is converted to an "image cube"

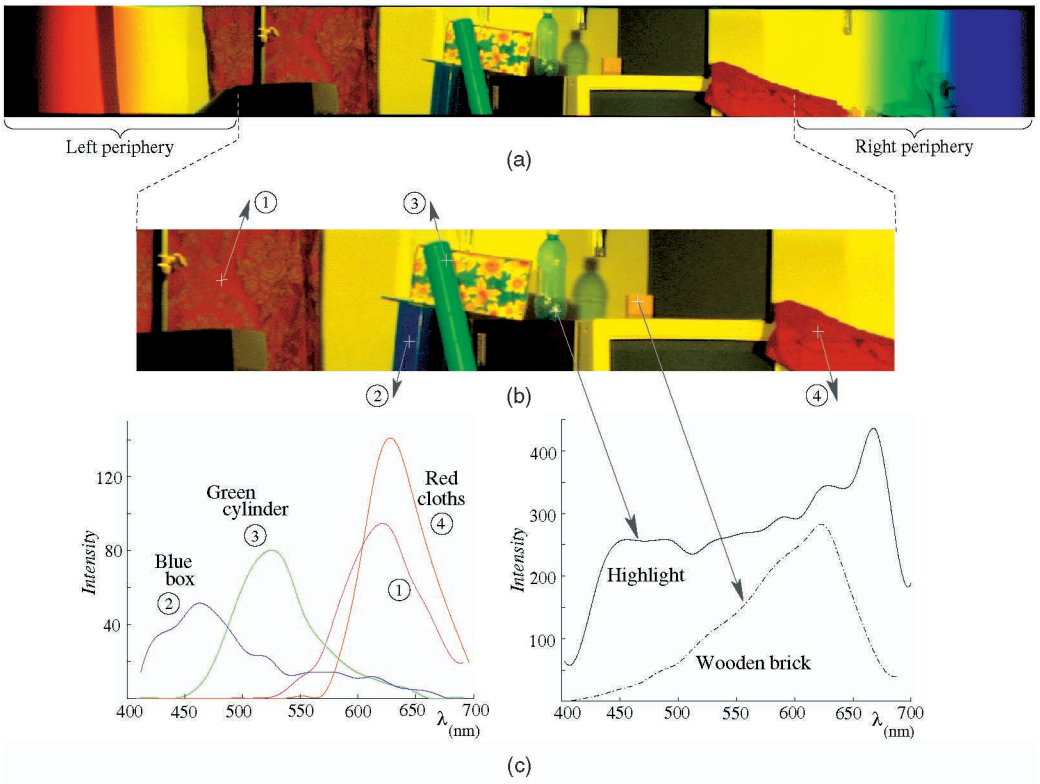

Fig. 12. (a) A color image mosaic rendered using the spectral data acquired at each point in its FOV, based on a single pass (rotation about the center of projection) of an ordinary black/white camera with a single fixed filter. The scene was illuminated by incandescent lamps. In the mosaic periphery, the spectral range becomes narrower towards the outer boundaries, thus gradually deteriorating the color rendering. (b) The mosaic's central region of interest contains the full spectral range the filter can scan. (c) The spectrum is plotted for selected points.

$g(\tilde{x}, \tilde{y}, \lambda)$, which is more convenient to work with (see Appendix A.2).

The spectral data is now available at each scene point, and can be used in multispectral imaging applications (see Section 2). If the response l_{system} is known[6] and not too small, it can be compensated for to yield the spectrum

$$I(\lambda) = \frac{g(\lambda)}{l_{\text{system}}(\lambda)} \tag{27}$$

at each scene point $(\tilde{x}, \tilde{y})$.

The human response to the spectral-spatial distribution can be estimated using the CIE-1931 tristimulus values for a standard observer:

$$\begin{aligned} X &= \int_{\lambda} I(\lambda) X(\lambda) d\lambda, \\ Y &= \int_{\lambda} I(\lambda) Y(\lambda) d\lambda, \\ Z &= \int_{\lambda} I(\lambda) Z(\lambda) d\lambda, \end{aligned} \tag{28}$$

where $X(\lambda)$, $Y(\lambda)$, and $Z(\lambda)$ are the tristimulus functions [16]. For display purposes, it is possible to convert the multispectral data into RGB images. We calculated the display RGB values from the XYZ values, based on the NTSC receiver primary system [22].

6. The response can be obtained by calibration or by the manufacturer's specifications.

8 Mosaicing Experiments

8.1 Multispectral Mosaicing

In an experiment, we used a commercial monochrome (black/white) video camera, with an off-the-shelf [11] linear variable interference filter.[7] Our experiment parameters were: $A = 30cm$, $F = 25mm$, and $f_{\#} = 5.6$. According to (20), $\Delta\phi_{\text{sample}} \approx 0.4^o$. The effective camera FOV subtended by the filter was $\approx 10^o$. Hence, about 21 spectral samples were acquired for each scene point.[8]

The scene was illuminated by incandescent lamps. A sequence of 65 still frames was taken, samples of which are shown in Fig. 9. The grabbed images were compensated for camera vignetting effects that were computed before hand. We registered the images using the method discussed in Section 6. The registration yielded a 43^o wide multispectral image mosaic, where the spectrum can be computed for each point. The multispectral mosaic was then converted to the RGB mosaic[9] shown in Fig. 12. This mosaic is based on the intensity measurements $g(\lambda)$. For more accurate results, the system response l_{system} needs to be accounted for.

As described in Section 3, the full range spectrum is obtained for the central region of interest. This region seems yellowish because the light coming from the incandescent

7. The characteristics of the interference filter somewhat depend on the angle of incidence [29]. However, in the experiment, the FOV angle of the camera was narrow so we neglected this effect.

8. Some were acquired less times because the displacement between frames was not strictly uniform.

9. To better show dim features the intensity in the mosaics shown in this paper was γ-contrast stretched, while the hue and saturation at each pixel were untouched.

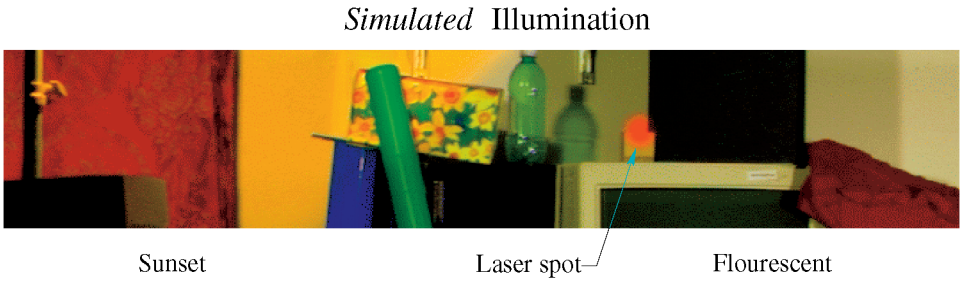

Fig. 13. Using known spectra, the scene was rendered *as if* its left part is illuminated by a setting sun [33], while its right part is illuminated by a "cool-white" fluorescent lamp [49]. In addition, a simulated red HeNe laser beam spot ($\lambda \approx 633nm$) "illuminates" the top edge of the wooden brick.

lamps is rather yellow, though the human visual system adapts to it (color constancy). Other than that, the estimated colors were consistent with the appeared colors of the objects.

Information is obtained also about the periphery, though with decreasing spectral range. Fig. 12a indeed shows the periphery regions, one frame wide, on both sides of the mosaic central part. Since the spectral range in these regions changes gradually from the central part, there is no abrupt decrease of quality in the periphery. However, towards the outer boundaries of the left and right periphery, the image becomes red and blue, respectively. This is due to the absence of data on the complimentary wavelengths in these regions. Even there, substantial information is still available for vision algorithms that make do with partial spectral data, or that do not rely on color but on spatial features. For example, the objects in the right periphery clearly appear in the raw frame shown at the last photograph in Fig. 9. It shows loose dark cables hanging down through the frame and their shadows on the wall behind. This can indicate the number and spatial distribution of the light sources in the scene, as in [37]. In addition, other objects (shaped bricks) can be recognized in this region. Therefore, the peripheral regions are not wasted data, but can be useful for vision.

One can get the spectrum at any point in the mosaic FOV, using a simple interactive tool we created, simply by clicking on the point with the computer's mouse, as shown in Figs. 12b and 12c.[10] We note that the system that we used to create this multispectral image mosaic was the same system used in [41] to create high-dynamic range mosaics. The only difference is that in [41], we attached a spatially varying density filter, while in this work, we attach a spatially varying spectral filter. This demonstrates the great flexibility and generality of the generalized mosaicing approach.

8.2 Rendering the Real Scene with Simulated Illumination

Given the spectral data, we may render the scene under any given illumination spectra. To do this, we first need to estimate the spectrum of the illumination that existed during the acquisition. This can be done easily if an object with calibrated reflectance exists in the mosaic FOV [15] (Otherwise, color constancy cues may be used). Let the acquired image cube be

$$g(\tilde{x},\tilde{y},\lambda) = \tilde{S}_{\text{source}}(\lambda)\, r(\tilde{x},\tilde{y},\lambda), \tag{29}$$

where $r(\tilde{x},\tilde{y},\lambda)$ is the reflectance, and $\tilde{S}_{\text{source}}(\lambda)$ is given by (2). We assume that the illumination spectrum S_{source} is constant across the mosaic FOV. The measurements of a white patch yield

$$g_{\text{white}}(\lambda) = \tilde{S}_{\text{source}}(\lambda). \tag{30}$$

Therefore, we estimate the reflectance by

$$\hat{r}(\tilde{x},\tilde{y},\lambda) = \frac{g(\tilde{x},\tilde{y},\lambda)}{g_{\text{white}}(\lambda)}. \tag{31}$$

Note that the system response $l_{\text{system}}(\lambda)$ is not needed to estimate (31). Given the spectrum of the new illuminant, $S^{\text{new}}_{\text{source}}$, which may be spatially varying, the rendered spectrum is

$$\begin{aligned} I_{\text{rendered}}(\tilde{x},\tilde{y},\lambda) &= S^{\text{new}}_{\text{source}}(\tilde{x},\tilde{y},\lambda)\hat{r}(\tilde{x},\tilde{y},\lambda) \\ &= S^{\text{new}}_{\text{source}}(\tilde{x},\tilde{y},\lambda)\frac{g(\tilde{x},\tilde{y},\lambda)}{g_{\text{white}}(\lambda)}. \end{aligned} \tag{32}$$

The rendered multispectral mosaic can be transformed to a rendered RGB image.

In our example, consider the label on the computer monitor appearing in the mosaic in Fig. 12. We assumed it to be white with black letters on it. Therefore, we averaged the measured spectra $g(\tilde{x},\tilde{y},\lambda)$ associated with the label's pixels,[11] to estimate $g_{\text{white}}(\lambda)$. Given models of the spectra of fluorescent lamps [49], sunset [33], and a HeNe laser, we rendered the image shown in Fig. 13.

The spectrum of fluorescent light is dominated by peaks related to light emission by the mercury vapor inside the fluorescent bulbs [49]. If the wavelength samples in the cube $g(\tilde{x},\tilde{y},\lambda)$ are too sparse, these peaks may be missed when calculating $I_{\text{rendered}}(\tilde{x},\tilde{y},\lambda)$. This may cause a noticeable error in the estimated color. Therefore, it is important that the spectral reflectances will be properly interpolated during the rendering. Of course, the denser the interpolation, the larger the number of calculations needed. The complexity of the rendering is therefore inversely related to the spectrum "feature size." Specifically, the data we had on the fluorescent peaks [49] was given in $10nm$ resolution. Therefore, to calculate the image under fluorescent illumination we cubic-interpolated the acquired multispectral mosaic to $5nm$. A more efficient way would be to use an adaptive interpolation scheme, where the wavelength samples are not equally spaced, but become dense only around the spectral peaks.

10. The compensation for vignetting enabled measurements of intensities higher than 255 with the 8-bit camera, as seen in Fig. 12.

11. Only pixels for which the intensity did not saturate the detector at any measurement were used in this estimation.

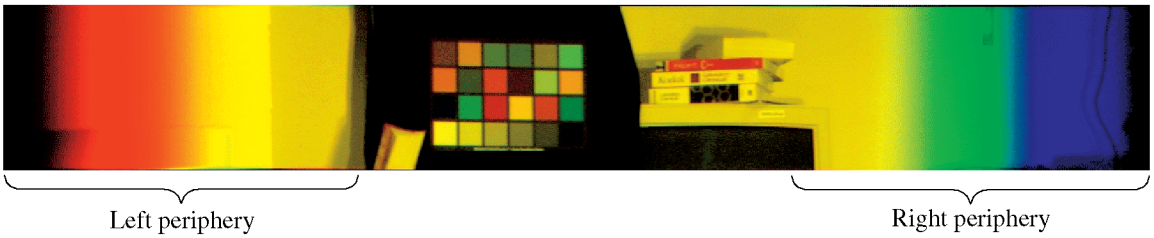

Fig. 14. The central region of interest of the color mosaic represents multispectral data on the Macbeth ColorChecker. The spectral range in this region spans the $400-700nm$ band. The range gradually narrows in the periphery.

8.3 Consistency Validation

The system can be calibrated using a calibrated light source or spectrometer. In the following experiment, we are mainly interested in validating the consistency of the intensity measurements with a target of known reflectances, rather than a source of known intensity distribution. We created a multispectral mosaic of the Macbeth ColorChecker (See Fig. 14) chart. For this chart, the reflectances $r_{\text{patch}}(\lambda_b)$ in narrow bands λ_b are known for each of the chart's 24 patches [16]. Since we used the same system parameters as in Section 8.1, 20 wavelength samples could suffice to approximate the measured spectrum. Nevertheless, since this is a verification experiment, we preferred to have denser samples. Thus, we obtained intensity measurements at hundreds of wavelength samples $g_{\text{patch}}(\lambda)$ per patch.

For each band b

$$g_{\text{patch}}(\lambda_b) = \tilde{S}_{\text{source}}(\lambda_b) r_{\text{patch}}(\lambda_b). \tag{33}$$

Thus, the average intensity distribution (a global property of the chart) is

$$\langle g(\lambda_b)\rangle = \frac{1}{24}\sum_{\text{patch}=1}^{24} g_{\text{patch}}(\lambda_b) = \tilde{S}_{\text{source}}(\lambda_b)\langle r(\lambda_b)\rangle, \tag{34}$$

where

$$\langle r(\lambda_b)\rangle = \frac{1}{24}\sum_{\text{patch}=1}^{24} r_{\text{patch}}(\lambda_b) \tag{35}$$

is the average spectral reflectance of the chart. Therefore, we may estimate the illumination source spectrum (modulated by the camera response) as

$$\widehat{\tilde{S}}_{\text{source}}(\lambda_b) = \frac{\langle g(\lambda_b)\rangle}{\langle r(\lambda_b)\rangle}. \tag{36}$$

Note that, although it is possible to estimate the illumination spectrum using any one of the patches (especially the white one), we preferred to use all of them for better stability.

Given the estimated illumination source spectrum and the known reflectances, we may expect to measure the spectrum

$$\hat{g}_{\text{patch}}(\lambda_b) = \widehat{\tilde{S}}_{\text{source}}(\lambda_b) r_{\text{patch}}(\lambda_b) = \frac{\langle g(\lambda_b)\rangle}{\langle r(\lambda_b)\rangle}\, r_{\text{patch}}(\lambda_b), \tag{37}$$

for each patch. The plots of the expected spectrum $\hat{g}_{\text{patch}}$ and the measured one, g_{patch} were in strong agreement for all the patches. The results for two patches are shown in Fig. 15. This test roughly means that the behavior of the measured spectra as a function of wavelength is consistent with the behavior of the known spectral reflectances.

To get a quantitative measure of the consistency, we measured the correlation coefficient between the measured spectra $g(\lambda_b)$ and the known reflectances $r(\lambda_b)$. Equation (33) shows that for each band, $g(\lambda_b)$ is linearly related to $r(\lambda_b)$. Thus, given different measurements at this band, the correlation coefficient should be close to 1. We have 24 different measurements per wavelength band—the number of patches. The correlation coefficient at a band b is thus estimated by

$$Corr(r,g)_b = \frac{Cov(r,g)_b}{\sqrt{Var[r(\lambda_b)]\; Var[g(\lambda_b)]}}. \tag{38}$$

$Var[r(\lambda_b)]$ and $Var[g(\lambda_b)]$ are the variances of the reflectance and the intensity of the band b, respectively:

$$Var[r(\lambda_b)] = \frac{1}{24}\sum_{\text{patch}=1}^{24}\left[r_{\text{patch}}(\lambda_b) - \langle r(\lambda_b)\rangle\right]^2, \tag{39}$$

$$Var[g(\lambda_b)] = \frac{1}{24}\sum_{\text{patch}=1}^{24}\left[g_{\text{patch}}(\lambda_b) - \langle g(\lambda_b)\rangle\right]^2, \tag{40}$$

while

$$Cov(r,g)_b = \langle \left[r_{\text{patch}}(\lambda_b) - \langle r(\lambda_b)\rangle\right]\left[g_{\text{patch}}(\lambda_b) - \langle g(\lambda_b)\rangle\right]\rangle. \tag{41}$$

The correlation was calculated for each of the 61 bands for which we had data on r in the $400-700nm$ range [16]. It

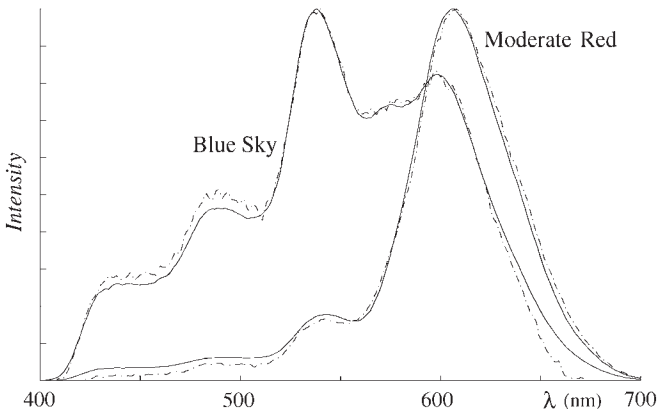

Fig. 15. (Solid) Expected spectra (normalized) for two patches of the Macbeth ColorChecker, based on the chart's average spectra and known reflectances. Each patch spectrum is the multiplication of its reflectance by the estimated illumination spectrum. (Dashed) The measured spectra.

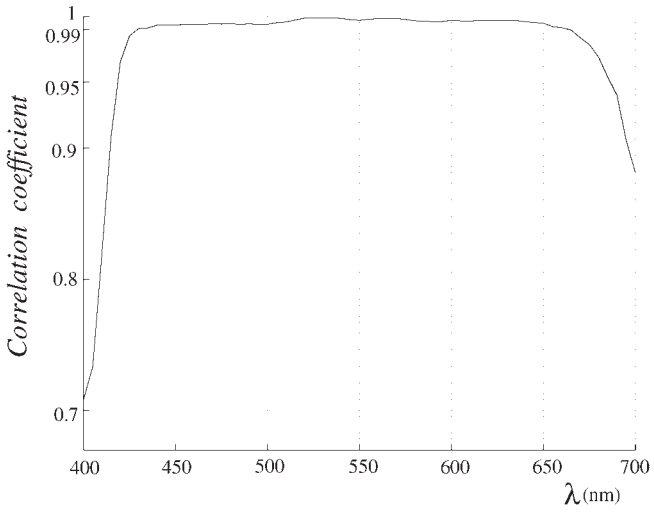

Fig. 16. The correlation between the measured intensities and known reflectances of the Macbeth ColorChecker is high for most wavelength bands. It decreases towards the boundaries of the filter's spectral range, where the intensity readouts were low.

is plotted as a function of wavelength in Fig. 16. The correlation is indeed very high. On average, the correlation is ≈ 0.98. Toward the boundaries of the spectral range, the correlation is lower. At these bands, the intensity readouts were very low and, thus, relatively noisy.

Although the correlation is high, this indication is not sufficient because we may generally expect that different wavelength bands will be somewhat correlated. For example, an object which has a high intensity in the blue band will probably (though not necessarily) be bright in the green band. Therefore, we may have an error in the correspondence between true wavelengths and assumed wavelengths and still get a positive correlation. To study that, we measured the correlation between $r_{\text{patch}}(\lambda_b)$ and $g_{\text{patch}}(\lambda_q)$ for *randomly* picked bands b, q:

$$Corr(r_b, g_q) = \frac{Cov(r_b, g_q)}{\sqrt{Var[r(\lambda_b)]\, Var[g(\lambda_q)]}}, \tag{42}$$

where

$$Cov(r_b, g_q) = \langle \left[r_{\text{patch}}(\lambda_b) - \langle r(\lambda_b) \rangle \right] \left[g_{\text{patch}}(\lambda_q) - \langle g(\lambda_q) \rangle \right] \rangle. \tag{43}$$

We repeated the calculation of $Corr(r_b, g_q)$ numerous times, and the average correlation between randomly picked bands is ≈ 0.65. Hence, there is indeed a positive correlation between different wavelengths. However, it is clear that the correlation $Corr(r, g)_b$ in the predetermined correspondence between assumed bands and measured bands is much higher than in a random correspondence. According to this test, our intensity measurements are therefore highly correlated with the known reflectances.

9 Conclusions

Generalized mosaicing is a framework for capturing information along multiple imaging dimensions. It is based on acquisition of similar or even the same amount of data as in the case of traditional mosaicing. Thus far, we have used this framework to compute high-dynamic range [40], [41] and multispectral mosaics. However, generalized mosaicing is not limited to these dimensions; it is a general concept that can be applied to other valuable dimensions (e.g., polarization, focus, depth), and perhaps to multiple dimensions simultaneously. This has implications for several aspects of computer vision. Generalized mosaicing can be applied to most of the current applications of image mosaicing [6], [7], [8], [13], [28], [39], [43], [26], [50], [53]. Moreover, using a spatially varying spectral filter, generalized mosaicing can enable sensing of scene spectrum in conjunction with scene structure by structure-from-motion [51] methods. Therefore, better data is made available for later texture mapping on the computed depth map to render new views of the scene, and views under simulated illumination spectra. In addition, the numerous scientific fields that rely on multispectral imaging [1], [15], [19], [30], [32], [44], [47], [48], [55] can benefit from this approach. The simplicity of the technique suggests that it can also add significant value to digital photography.

The method described in this paper is very flexible. If the user wants to change the characteristics of the filtering, he or she may simply change the external filter. This simple hardware modification is accompanied with a change in the postprocessing parameters. For example, rather than using a bandpass filter (with a varying band), we may use a short-pass or long-pass filter, with a spatially varying cutoff wavelength [14]. In that case, the intensity in narrow bands is obtained by differentiating the spectral data. Alternatively, the filter may change its pass band in steps, as is the case of a filter array [45]. Due to defocus blur of the filter, the effective filtering will vary continuously. Note that a spatially varying filter may reflect the light into the camera rather than transmitting it, which may find uses in catadioptric systems.

Appendix A

A.1 The Maximum-Likelihood Pyramid

To make the registration more robust and efficient, it is done hierarchically, from coarse to fine resolution. For the algorithm described in Section 6, we need an estimate of the intensity I and its uncertainty ΔI at each pixel and each scale. A coarse representation of an image at a specific pyramid level can be obtained [5] by subsampling it after it is lowpass filtered with a kernel whose width depends on the level (the higher/coarser the level, the wider is the kernel that operates on the *original* image). The value of a pixel in this representation is a weighted sum of the measured pixels:[12]

$$I = \frac{\sum_k \omega_k I_k}{\sum_k \omega_k}. \tag{44}$$

In a conventional pyramid, the weights ω_k are equal to the values of a Gaussian kernel a_k [5].

12. Please note that, we refer here to the construction of the pyramid levels from the original, full resolution image, where the pixels may be considered as independent. This is done to keep the derivation simple. When the pyramid is constructed iteratively, the pixels in the intermediate levels are not statistically independent.

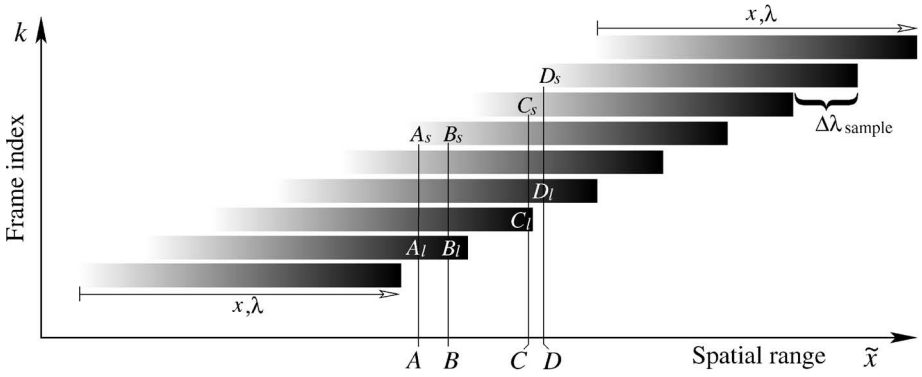

Fig. 17. Each frame k is displaced relative to its previous one. Each scene point $\tilde{x}$ is sampled at wavelength intervals $\Delta\lambda_{\text{sample}}$. However, different scene points generally have different wavelength samples, according to their respective coordinates x in the internal coordinate systems of the frames.

Here, we set the weight ω_k assigned to a pixel such that it decreases linearly both as its Gaussian weight a_k decreases and as its uncertainty ΔI_k increases:

$$\omega_k = \frac{a_k}{\Delta I_k^2} \tag{45}$$

and we set

$$\Delta I = \left(\sqrt{\sum_k \omega_k} \right)^{-1}. \tag{46}$$

In the special case for which the uncertainty ΔI_k is the same for all pixels k, ω_k is proportional to the Gaussian coefficients a_k. On the other hand, in the special case for which a_k is the same for all pixels k, (44) and (46) become equal to (24) and (25), respectively. Thus, (44) yields the ML representation of the pixels, assuming them to be Gaussian and independent.

Therefore, (44) and (45) generalize both the pyramid structure and the ML estimation. The influence of a pixel in a patch on its coarse representation increases the closer it is to the patch center (as in usual pyramids), and the more reliable it is. In this kind of a *Maximum-Likelihood pyramid*, the representation at each resolution level consists not only of the weighted-averaged value at each pixel, but also of its uncertainty.

A.2 The Raw Samples Structure

As the raw images are acquired, different scene points have different wavelength samples. Consider Fig. 17. The wavelength samples of each scene point $\tilde{x}$ depend on its coordinate x in the internal system within each frame. Hence, adjacent points, e.g., A and B have different wavelength samples. In particular, A_s and A_l, the shortest and longest wavelength samples A are different than B_s and B_l, the shortest and longest wavelength samples of B.

The raw multispectral data structure consists of intensity measurements $g(\tilde{x}, \tilde{y}, k)$ and corresponding wavelength samples $\lambda(\tilde{x}, \tilde{y}, k)$, where $(\tilde{x}, \tilde{y})$ are the coordinates of the scene points in the global coordinate system of the mosaic (see Fig. 11). In many applications, this structure is not convenient and an image cube $g(\tilde{x}, \tilde{y}, \lambda)$ is preferable. This requires the estimation of the spectrum at a regular, space invariant grid of wavelengths. In practice, it means that the spectrum is interpolated between the raw samples. Note that to create a color mosaic, one can use directly the raw images: Each raw frame $g_k(x, y)$ can be converted to color images $R_k(x, y)$ $G_k(x, y)$ $B_k(x, y)$, since we know which wavelength band is sensed at each pixel. Then, these color images are mosaiced. Creating a color mosaic this way demands much less memory than using the entire image cube. However, it is accompanied by annoying seams for the reason described in the following.

The wavelength samples usually change gradually along $\tilde{x}$. However, the change is abrupt at the boundaries of the frames that compose the mosaic. For example, in Fig. 17, points C and D are close to one another. However, their respective shortest wavelength samples, C_s and D_s are very different. The same applies to their respective longest wavelength samples, C_l and D_l. The discontinuity equals to the wavelength sampling interval $\Delta\lambda_{\text{sample}}$. As long as points C and D are on two opposite sides of a frame boundary, this discontinuity remains, no matter how close their mosaic spatial coordinates $\tilde{x}(C)$ and $\tilde{x}(D)$ are.

This phenomenon is usually not crucial from a quantitative point of view. The reason is that when the sampling rate is in accordance to that derived in Section 5, the wavelength discontinuity $\Delta\lambda_{\text{sample}}$ is smaller than the typical bandwidth (blur) of any sample. However, it can cause annoying seams in the color image mosaic at the positions of the frame boundaries. The seams can be removed by feathering techniques as in [5], [43]. In those methods, the RGB frames are multiplied by a weight that decreases towards their boundaries, before they are fused. Feathering has a disadvantage, though. Recall that the "red" and the "blue" wavelengths are the ones which are measured near the boundaries of each frame $g_k(x, y)$. Hence, the spatially varying weighting attenuates these wavelengths relative to the central "green" wavelengths. Therefore, the resulting mosaic is a little bit more greenish than what it should be.

This problem was avoided altogether when the color mosaic was constructed from the image cube $g(\tilde{x}, \tilde{y}, \lambda)$. The reason is that the grid of wavelength samples in the image cube is space invariant.

ACKNOWLEDGMENTS

This work was supported in parts by a US National Science Foundation ITR Award, IIS-00-85864, a DARPA/ONR MURI Grant, N00014-95-1-0601, the Louis Morin Fellowship, and a David and Lucile Packard Fellowship.

REFERENCES

[1] A. Abrardo, L. Alparone, V. Cappellini, and A. Prosperi, "Color Constancy from Multispectral Images," *Proc. IEEE Int'l Conf. Image Processing,* vol. 3, pp. 570-574, 1999.

[2] E.H. Adelson and J.R. Bergen, "The Plenoptic Function and the Elements of Early Vision," *Computational Models of Visual Processing,* M. Landy and J.A. Movshon, eds., pp. 3-20, MIT Press, 1991.

[3] M. Aggarwal and N. Ahuja, "High Dynamic Range Panoramic Imaging," *Proc. Int'l Conf. Computer Vision,* vol. I, pp. 2-9, 2001.

[4] M.J. Black and A. Rangarajan, "On the Unification of Line Processes, Outlier Rejection, and Robust Statistics with Applications in Early Vision," *Int'l J. Computer Vision,* vol. 19, no. 1, pp. 57-91, 1996.

[5] P.J. Burt and E.H. Adelson, "A Multiresolution Spline with Application to Image Mosaics," *ACM Trans. Graphics,* vol. 2, pp. 217-236, 1983.

[6] R. Bernstein, "Digital Image Processing of Earth Observation Sensor Data," *IBM J. Research and Development,* vol. 20, no. 1, pp. 40-57, 1976.

[7] D. Capel and A. Zisserman, "Automated Mosaicing with Super-Resolution Zoom," *Proc. IEEE Computer Vision and Pattern Recognition,* pp. 885-891, 1998.

[8] S. Coorg, N. Master, and S. Teller, "Acquisition of a Large Pose-Mosaic Dataset," *Proc. IEEE Computer Vision and Pattern Recognition* pp. 872-878, 1998.

[9] M.L. Duplaquet, "Building Large Image Mosaics with Invisible Seam Lines," *Proc. SPIE Visual Information Processing VII,* vol. 3387, pp. 369-377, 1998.

[10] D. Dykaar and G. Allan, "Capturing Color Information with Linear CCDs: The Case for Three-Chip Cameras," *Photonics Spectra,* vol. 34, no. 5, pp. 183-186, 2000.

[11] *2000 Optics and Optical Instruments Catalog,* Edmund Industrial Optics, stock nos. K45-645, p. 67, 2000.

[12] H. Farid and E.P. Simoncelli, "Range Estimation by Optical Differentiation," *J. Optical Soc. Am. A,* vol. 15, pp. 1777-1786, 1998.

[13] E. Fernandez, R. Garfinkel, and R. Arbiol, "Mosaicking of Aerial Photographic Maps via Seams Defined by Bottleneck Shortest Paths," *Operations Research,* vol. 46, no. 3, pp. 293-304, 1998.

[14] N. Gat, "Spectrometer Apparatus," US Patent 5166755, 1992.

[15] N. Gat, "Imaging Spectroscopy Using Tunable Filters: A Review," *Proc. SPIE Wavelet Applications VII,* vol. 4056, pp. 50-64, 2000.

[16] A.S. Glassner, "Appendix G.4," *Principles of Digital Image Synthesis.* Morgan-Kaufmann, 1995.

[17] R. Gore, "Ancient Ashkelon," *Nat'l Geographic,* vol. 199, no. 1, pp. 66-93, 2001.

[18] S. Gumustekin and R.W. Hall, "Mosaic Image Generation on a Flattened Sphere," *Proc. IEEE Workshop Applications of Computer Vision,* pp. 50-55, 1996.

[19] M. Hauta-Kasari, K. Miyazawa, S. Toyooka, and J. Parkkinen, "A Prototype of the Spectral Vision System," *Proc. 11th Scandinavian Conf. Image Analysis,* vol. 1, pp. 79-86, 1999.

[20] M. Irani, P. Anandan, J. Bergen, R. Kumar, and S. Hsu, "Efficient Representations of Video Sequences and Their Application," *Signal Processing: Image Comm.,* vol. 8, pp. 327-351, 1996.

[21] M. Irani and P. Anandan, "Robust Multi-Sensor Image Alignment," *Proc. Int'l Conf. Computer Vision,* pp. 959-966, 1998.

[22] A.K. Jain, *Fundamentals of Digital Image Processing,* Prentice Hall, Englewood Cliffs, p. 67, 1989.

[23] R.B. Kerr, "Filters Expand Capabilities of Infrared Imaging," *Laser Focus World,* vol. 37, no. 1, pp. 151-158, 2001.

[24] R. Kwok, J.C. Curlander, and S. Pang, "An Automated System for Mosaicking Spaceborne SAR Imagery," *Int'l J. Remote Sensing,* vol. 11, pp. 209-223, 1990.

[25] S.G. Lipson, H. Lipson, and D.S. Tannhauser, *Optical Physics,* third ed. pp. 305-311, Cambridge: UK, 1998.

[26] C.J. Lada, D.L. DePoy, K.M. Merrill, and I. Gatley, "Infrared Images of M17," *The Astronomical J.,* vol. 374, pp. 533-539, 1991.

[27] C. Mahoney, "Thermal Infrared Imaging Spectrometer: An Advanced Optics Technology Instrument," *Proc. SPIE Imaging Spectroscopy of the Terrestrial Environment,* vol. 1298, pp. 87-92, 1990.

[28] S. Mann, "Joint Parameter Estimation in both Domain and Range of Functions in Same Orbit of the Projective-Wyckoff Group," *Proc. IEEE Int'l Conf. Image Processing,* pp. 193-196, 1996.

[29] *Melles Griot Catalog,* pp. 13.27-13.29, 2000.

[30] A.M. Mika, "Linear-Wedge Spectrometer," *Proc. SPIE Imaging Spectroscopy of the Terrestrial Environment,* vol. 1298, pp. 127-131, 1990.

[31] R. Miskelly and C. Seymour, "Capturing Color Information with Linear CCDs: The Case for Trilinear Cameras," *Photonics Spectra,* vol. 34, no. 5, pp. 183-184, 2000.

[32] G. Monnet, "3D Spectroscopy with Large Telescopes: Past, Present and Prospects," *Tridimensional Optical Spectroscopic Methods in Astronomy,* vol. 71, pp. 12-17, 1995.

[33] M.R. Nagel, H. Quenzel, W. Kweta, and R. Wendling, *Daylight Illumination—Color-Contrast Tables,* pp. 78-87, Academic Press, 1978.

[34] S. Peleg and J. Herman, "Panoramic Mosaics by Manifold Projection," *Proc. Computer Vision and Pattern Recognition,* pp. 338-343, 1997.

[35] S.F. Ray, *Applied Photographic Optics: Lenses and Optical Systems for Photography, Film, Video and Electronic Imaging,* second ed. pp. 559-563, Oxford: Focal Press, 1994.

[36] E.M. Reynoso, G.M. Dubner, W.M. Goss, and E.M. Arnal, "VLA Observations of Neutral Hydrogen in the Direction of Puppis A," *The Astronomical J.,* vol. 110, pp. 318-324, 1995.

[37] I. Sato, Y. Sato, and K. Ikeuchi, "Illumination Distribution from Brightness in Shadows: Adaptive Estimation of Illumination Distribution with Unknown Reflectance Properties in Shadow Regions," *Proc. Int'l Conf. Computer Vision,* pp. 875-883, 1999.

[38] H.S. Sawhney and S. Ayer, "Compact Representations of Videos through Dominant and Multiple Motion Estimation," *IEEE Trans. Pattern Analysis and Machine Intelligence,* vol. 18, pp. 814-830, 1996.

[39] H.S. Sawhney, R. Kumar, G. Gendel, J. Bergen, D. Dixon, and V. Paragano, "VideoBrush^{TM}: Experiences with Consumer Video Mosaicing," *Proc. IEEE Workshop Applications of Computer Vision,* pp. 52-62, 1998.

[40] Y.Y. Schechner and S.K. Nayar, "Generalized Mosaicing," *Proc. Int'l Conf. Computer Vision,* vol. I, pp. 17-24, 2001.

[41] Y.Y. Schechner and S.K. Nayar, "Generalized Mosaicing: High Dynamic Range in a Wide Field of View," pending publication.

[42] R.K. Sharma and M. Pavel, "Multisensor Image Registration," *Soc. for Information Display,* vol. XXVIII, pp. 951-954, 1997.

[43] H.Y. Shum and R. Szeliski, "Systems and Experiment Paper: Construction of Panoramic Image Mosaics with Global and Local Alignment," *Int'l J. Computer Vision,* vol. 36, pp. 101-130, 2000.

[44] D. Slater and G. Healey, "Material Classification for 3D Objects in Aerial Hyperspectral Images," *Proc. Computer Vision and Pattern Recognition,* vol. 2, pp. 268-273, 1999.

[45] A. Smajkiewicz, "An Argument for a Filter Array vs. Linear Variable Filter in Precision Analytical Instruments Applications," Barr Assoc. Inc., Document no. P95081.

[46] L.A. Soderblom, K. Edwards, E.M. Eliason, E.M. Sanchez, and M.P. Charette, "Global Color Variations on the Martian Surface," *Icarus,* vol. 34, pp. 446-464, 1978.

[47] H.M.G. Stokman, T. Gevers, and J.J. Koenderink, "Color Measurement by Imaging Spectrometry," *Computer Vision and Image Understanding,* vol. 79, no. 2, pp. 236-249, 2000.

[48] X. Sun and J.M. Anderson, "A Spatially Variable Light-Frequency-Selective Component-Based, Airborne Pushbroom Imaging Spectrometer for the Water Environment," *Photogrammetric Eng. & Remote Sensing,* vol. 59, no. 3, pp. 399-406, 1993.

[49] *Osram Sylvania Corp.,* "Fluorescent Lamps,"Eng. Bulletin 0-341, Danvers, Mass.

[50] R. Szeliski, "Image Mosaicing for Telereality Applications," *Proc. IEEE Workshop Applications of Computer Vision,* pp. 44-53, 1994.

[51] C. Tomasi and T. Kanade, "Shape and Motion from Image Streams under Orthography: A Factorization Method," *Int'l J. Computer Vision,* vol. 9, pp. 137-154, 1992.

[52] J.M. Uson, S.P. Boughn, and J.R. Kuhn, "The Central Galaxy in Abell 2029: An Old Supergiant," *Science,* vol. 250, pp. 539-540, 1990.

[53] A.R. Vasavada, A.P. Ingersoll, D. Banfield, M. Bell, P.J. Gierasch, and M.J.S. Belton, "Galileo Imaging of Jupiter's Atmosphere: The Great Red Spot, Equatorial Region, and White Ovals," *Icarus,* vol. 135, pp. 265-275, 1998.

[54] A. Villemaire, S. Fortin, C. Lafond, M.A. Soucy, J.F. Legault, J. Giroux, S.M. Goodrich, R.S. Bauldree, and R.J. Rapp, "A High Resolution Airborne Imaging Spectrometer," *Proc. SPIE Infrared Technology and Applications XXIV,* vol. 3436, pp. 924-930, 1996.

[55] J.B. Wellman, "Multispectral Mapper: Imaging Spectroscopy as Applied to the Mapping of Earth Resources," *Proc. SPIE Imaging Spectroscopy,* vol. 268, pp. 64-73, 1981.

Yoav Y. Schechner received the BA and MSc degrees in physics and the PhD degree in electrical engineering from the Technion-Israel Institute of Technology in 1990,1996, and 1999, respectively. His research is focused on the use of optics and physics in imaging and computer vision. He is also interested in geographic photography. He is a research scientist in the Department of Computer Science at Columbia University. He was the recipient of the Wolf Foundation Award for Graduate Students in 1994, the Gutwirth Special Distinction Fellowship in 1995, the Israeli Ministry of Science (Eshkol) Distinction Fellowship and the Ollendorff Award in 1998, and the Scwartz Foundation Award in 1999. He is currently a recipient of the Morin Fellowship.

Shree K. Nayar received the PhD degree in electrical and computer engineering from the Robotics Institute at Carnegie Mellon University in 1990. He is a professor in the Department of Computer Science at Columbia University. He currently heads the Columbia Automated Vision Environment (CAVE), which is dedicated to the development of advanced computer vision systems. His research is focused on three areas, namely, the creation of novel vision sensors, the design of physics based models for vision, and the development of algorithms for scene understanding. His work is motivated by applications in the fields of digital imaging, computer graphics, human-machine interfaces, robotics, and image understanding. Dr. Nayar has authored and coauthored papers that have received the Best Paper Honorable Mention Award at the 2000 IEEE Computer Vision and Pattern Recognition Conference (CVPR '00), the David Marr Prize at the 1995 International Conference on Computer Vision (ICCV '95) held in Boston, Siemens Outstanding Paper Award at the 1994 IEEE Computer Vision and Pattern Recognition Conference (CVPR '94) held in Seattle, 1994 Annual Pattern Recognition Award from the Pattern Recognition Society, Best Industry Related Paper Award at the 1994 International Conference on Pattern Recognition (ICPR '94) held in Jerusalem, and the David Marr Prize at the 1990 International Conference on Computer Vision (ICCV '90) held in Osaka. He holds several US and international patents for inventions related to computer vision and robotics. Dr. Nayar was the recipient of the David and Lucile Packard Fellowship for Science and Engineering in 1992, the National Young Investigator Award from the US National Science Foundation in 1993, and the Excellence in Engineering Teaching Award from the Keck Foundation in 1995.