International Journal of Computer Vision 53(3), 245–267, 2003

Generalized Mosaicing: High Dynamic Range in a Wide Field of View

YOAV Y. SCHECHNER
Department of Electrical Engineering, Technion—Israel Institute of Technology, Haifa 32000, Israel

SHREE K. NAYAR
Columbia Automated Vision Environment, Department of Computer Science, Columbia University, New York, NY 10027

Received January 29, 2001; Revised January 27, 2003; Accepted January 28, 2003

Abstract. We present an approach that significantly enhances the capabilities of traditional image mosaicking. The key observation is that as a camera moves, it senses each scene point multiple times. We rigidly attach to the camera an optical filter with spatially varying properties, so that multiple measurements are obtained for each scene point under different optical settings. Fusing the data captured in the multiple images yields an image mosaic that includes additional information about the scene. We refer to this approach as generalized mosaicing. In this paper we show that this approach can significantly extend the optical dynamic range of any given imaging system by exploiting vignetting effects. We derive the optimal vignetting configuration and implement it using an external filter with spatially varying transmittance. We also derive efficient scene sampling conditions as well as ways to self calibrate the vignetting effects. Maximum likelihood is used for image registration and fusion. In an experiment we mounted such a filter on a standard 8-bit video camera, to obtain an image panorama with dynamic range comparable to imaging with a 16-bit camera.

Keywords: sensors, inverse problems, image fusion, mosaicing, mosaicking, machine vision, physics based vision, SNR, vignetting, panorama

1. Generalized Mosaics

Image mosaicing[1] is a common method to obtain a wide field of view (FOV) image of a scene (Hsu et al., 2002; Irani et al., 1996; Smolić and Wiegand, 2001). The basic idea is to capture images as a camera moves and stitch these images together to obtain a larger image. Image mosaicing has been applied to consumer photography (Peleg et al., 2001; Sawhney et al., 1998; Shum and Szeliski, 2000), and in optical remote sensing of the Earth (Bernstein, 1976; Hansen et al., 1994) and of other objects in the solar system (Batson, 1987; Soderblom et al., 1978; Vasavada et al., 1998). It has also been used in various other scientific fields, such as optical observational astronomy (Lada et al., 1991; Uson et al., 1990), radio astronomy (Reynoso et al., 1995), remote sensing by synthetic aperture radar (SAR) (Curlander, 1984; Kwok et al., 1990), and underwater research (Ballard, 2001; Eustice et al., 2002; Garcia et al., 2001; Negahdaripour et al., 1998).

As depicted in Fig. 1, traditional image mosaicing mainly addressed the extension of the FOV, while other imaging dimensions were not improved in the process. We show that image mosaicing can be generalized to extract much more information about the scene, given a similar amount of acquired data. We refer to this approach as *generalized mosaicing*. The basic observation is that a typical video sequence acquired during mosaicing has great redundancy in terms of the data it contains; as the camera moves, each scene point is observed multiple times. We exploit this fact to extend

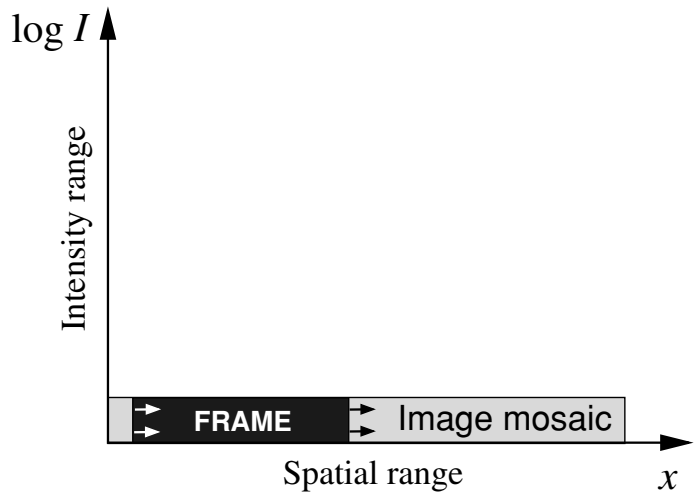

Figure 1. By stitching partly-overlapping frames, a traditional image mosaic extends the field of view of any camera without compromising the spatial resolution. However, other imaging dimensions, such as dynamic range are usually not improved.

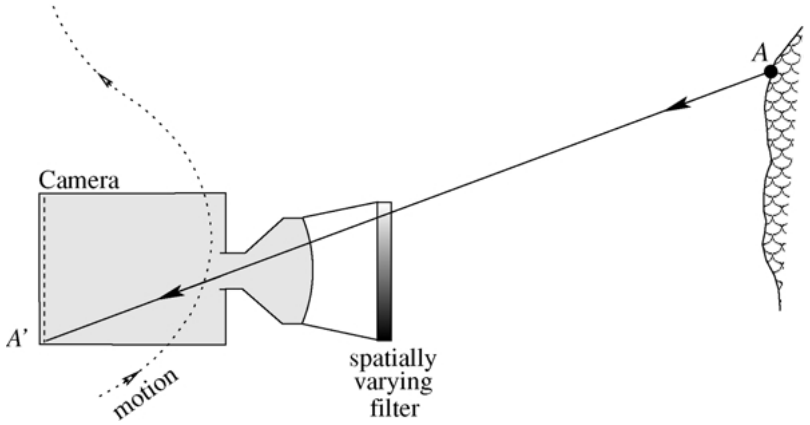

Figure 2. Scene point A is imaged on the detector at A' through a spatially varying filter attached to the camera. As the imaging system moves, each scene point is sensed through different portions of the filter, thus multiple measurements are obtained under different optical settings.

the *radiometric dynamic range* of the camera via its motion. Consider the setup shown in Fig. 2. A fixed filter with spatially varying properties is rigidly attached to the camera. Hence, as the camera moves (or simply rotates), each scene point is measured under different optical settings.[2] This simple modification of the imaging system significantly reduces the redundancy in the captured video stream. In return, more information about each point is embedded in the acquired data. Except for mounting the fixed filter, the image acquisition in generalized mosaicing is identical to traditional mosaicing. When a filter with spatially varying transmittance is attached to the camera, each scene point is effectively measured with different exposures as the camera moves, although the global exposure settings of the system are fixed. These measurements are combined to obtain a high dynamic range (HDR) mosaic.

In the following sections we describe the extension of the dynamic range and the FOV in a unified framework.[3] We review the previous approaches to HDR imaging in this framework, and show that they cover only part of the spatio-intensity space which is introduced here. We then discuss the optimal filter configuration and efficient sampling criteria. Figure 3 shows prototype systems we built. In an experiment, a standard 8-bit black-and-white video camera and a spatially varying transmittance filter were combined to form a

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

mosaicing system with dynamic range comparable to a 16-bit camera. Moreover, the vignetting effects of the system were self calibrated from the image sequence that composed the mosaic.

2. Mosaicing in the Spatio-Intensity Space

In many scenarios, object radiance changes by many orders of magnitude across the FOV. For this reason, there has recently been an upsurge of interest in obtaining HDR image data, as we detail in Section 3, and in their representation (Durand and Dorsey, 2002; Fattal et al., 2002; Larson et al., 1997; Pardo and Sapiro, 2002; Socolinsky, 2000). On the other hand, raw images have a limited optical dynamic range (Ogiers, 1997), set by the limited dynamic range of the camera detector. Above a certain detector irradiance, the images become saturated and their content at the saturated region is lost. Attenuating the irradiance by a shorter exposure time, a smaller aperture, or a neutral (space invariant) density filter can ensure that all the image points will be unsaturated. However, at the same time other information is lost since light may be below the detection threshold in regions of low irradiance.

In our approach to extend the dynamic range, we mount a fixed filter on the camera whose intensity transmittance varies across the filter's extent, as in Figs. 2 and 3. This causes an *intended vignetting*. Including vignetting effects originating from the lens, the overall effect is equivalent to spatially attenuating the image by a mask $M(x)$, where x is the axis along which the mask is changing.[4]

Now, let the scene be scanned by a general motion of the camera. For example, the camera can be rotated manually or using a motorized turntable. The moving system attenuates the light from any scene point differently in each frame. Effectively, the camera captures each point with different exposures during the sequence. Therefore, the system acquires both dark and bright areas with high quality while extending the FOV. It may be viewed as introducing a new dimension to the mosaicing process (Fig. 4) for better describing the plenoptic function (Adelson and Bergen, 1991). This dimension leads to the introduction of the concept of the *spatio-intensity space*. In Fig. 4, the spatio-intensity support of a single frame occupies a diagonal region in the spatio-intensity space. This occurs if the log of the transmittance varies linearly across the FOV.

Let the minimal irradiance that can be sensed by the detector above its noise at darkness (for the given camera specifications) be $I_{\min}^{\text{detector}}$. This determines the minimal irradiance that can be sensed by the entire system for $\max M = 1$ (transparency). Let the maximum irradiance that the detector can measure without saturation be $I_{\max}^{\text{detector}}$. The optical dynamic range of the detector in base 2 (bits) is then

$$DR^{\text{detector}} = \log_2 \frac{I_{\max}^{\text{detector}}}{I_{\min}^{\text{detector}}}. \tag{1}$$

Typically, $DR^{\text{detector}} = 8$ bits. The maximum irradiance that the entire system can sense without being saturated is when the detector yields its maximum output under the strongest attenuation (that is, with the smallest value of the mask M): $I_{\max}^{\text{system}} = I_{\max}^{\text{detector}} / \min M$. Therefore, the optical dynamic range of the system is

$$\begin{aligned} DR^{\text{system}} &= \log_2 \frac{I_{\max}^{\text{system}}}{I_{\min}^{\text{detector}}} = DR^{\text{detector}} - \log_2(\min M) \\ &= DR^{\text{detector}} + \log_2[\max(1/M)]. \end{aligned} \tag{2}$$

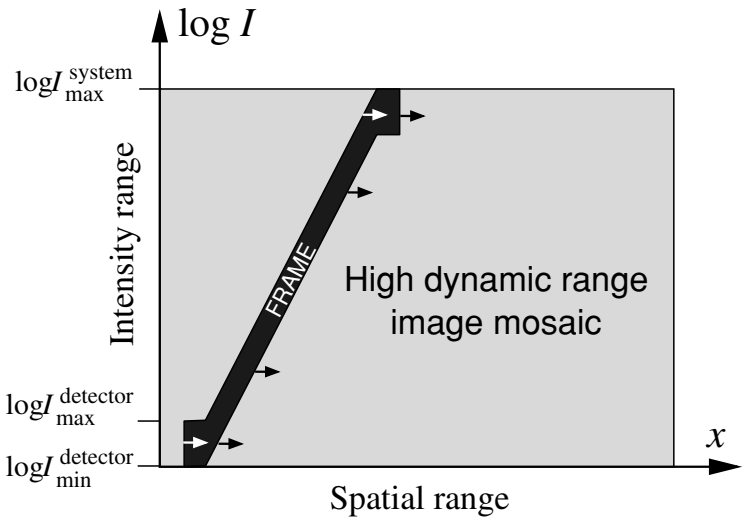

Figure 4. The same procedure of image mosaicing coupled with exploiting vignetting effects yields HDR image mosaics. Besides the FOV, it also extends the camera intensity dynamic range, without compromising the definition (quantization level density) at any intensity range. The dynamic range is extended at all scene points, irrespective of the intensity of their surroundings.

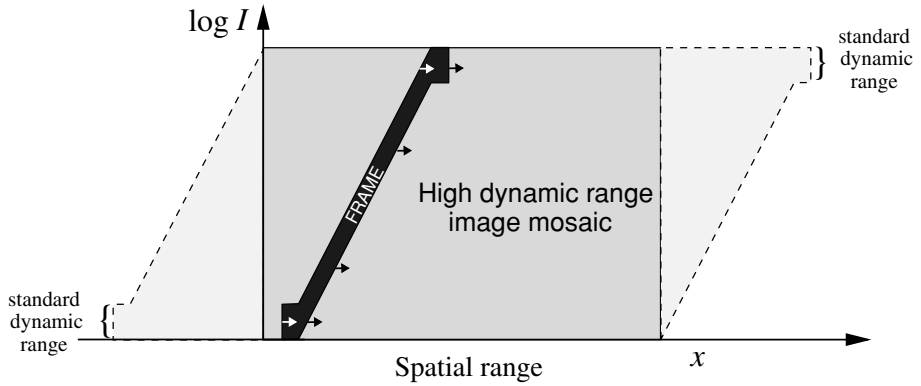

Figure 5. Outside the main region of interest the mosaic provides additional information about the scene periphery, whose quality gradually coincides with that of a single frame. This figure, as in Fig. 4, describes a system where each frame occupies a diagonal region in the spatio-intensity space.

High definition intensity may be obtained for all the pixels in a wide FOV image mosaic. In addition, information becomes available about the periphery of the central region of interest: the periphery has a smaller dynamic range, but at least the standard dynamic range of the detector (Fig. 5). Such a structure is analogous to foveated imaging systems, in which the acquisition quality improves from the periphery towards the center of the FOV. The periphery is at most one frame wide, and it diminishes in 360° panoramic mosaics.

3. High Dynamic Range: Previous Approaches

A common approach to avoid saturation and extend the dynamic range is to take multiple exposures of the scene. Note that there is an analogy between the extension of the FOV and the extension of dynamic range. Each is dealing with a different dimension of the spatio-intensity space depicted in Figs. 1 and 4–6. Traditional mosaicing addresses the spatial dimension. On the other hand, fusion of differently exposed images (Burt and Kolczynski, 1993; Debevec and Malik, 1997;

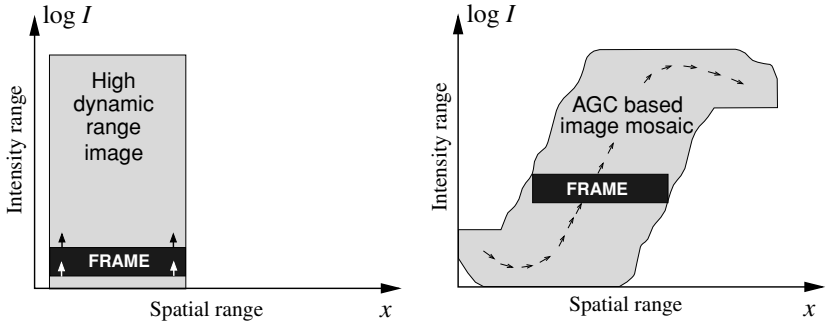

Figure 6. Previous approaches to HDR imaging as represented in the spatio-intensity space. (Left) Fusing differently exposed images taken with a static camera improves the dynamic range in a limited FOV. (Right) If the camera gain adapts to the global scene radiance by AGC, the dynamic range is extended over the global mosaic, but not at each point. Scene points may still be saturated or dark if their intensity resides below or above the local intensity range covered by the mosaic.

Mann and Picard, 1995; Mitsunaga and Nayar, 1999) addresses the intensity dimension (see the left side of Fig. 6). In that method, images are acquired sequentially with a static camera, hence the dynamic range is extended in a fixed FOV. A related approach to HDR is based on specialized hardware, where a mosaic of neutral density filters is placed on the detector (Nayar and Mitsunaga, 2000). This configuration trades the spatial resolution for the extension of the optical dynamic range. Moreover, it requires the modification of the inner parts of the imaging system.

It is possible to enlarge the captured FOV without mosaicing using a wide FOV lens. In analogy, there are ways to have extended dynamic range at each point and time, without combining images taken with different exposures. CMOS detectors may be manufactured to yield an electronic output logarithmic to the light intensity (C-Cam sensors; Davis, 1998; Ogiers, 1997; Schwartz, 1999). Nonlinear transmittance hardware (Khoo et al., 1999; Tabiryan and Nersisyan, 2002) can extend the dynamic range of any given detector by having a lower transmittance for higher light intensities. In these ways, high intensity can be sensed unsaturated. However, just as using a wider FOV lens with a given camera reduces the spatial resolution at which the scene is captured, using nonlinear response reduces the resolution at which the intensity is measured. The intensity information is compressed, since some quantization levels devoted to the lower intensities are traded for sparse samples of the high intensity range. Thus, extending the optical dynamic range in images without compromising the resolution of the intensity measurement requires fusion of several images.

In the realm of mosaicing, simultaneous extension of the FOV and dynamic range by exploiting the automatic gain control (AGC) feature of a camera was proposed in Mann (1996). However, AGC has a global (or, at best regional) effect and does not guarantee the required measurements at all scene points. For instance, a very bright scene point may remain saturated throughout the image sequence if it happens to be surrounded by a large dark area. Similarly, a dim point may remain dark throughout the sequence when it is surrounded by a bright area (see the right side of Fig. 6).

The generalized mosaicing approach extends the FOV and the dynamic range simultaneously, without compromising the spatial or intensity resolution. In addition, the amount of data gathered is similar to that gathered for a traditional image mosaic. Generalized mosaicing can be combined with some of the approaches mentioned above. Cameras that incorporate detectors having a logarithmic response (CMOS cameras), and/or nonlinear transmittance filter (as in Khoo et al. (1999) and Tabiryan and Nersisyan (2002)) placed on the sensor, and/or electronic AGC mechanisms can also exploit the proposed method to further extend the optical dynamic range of the overall system. Generalized mosaicing results in HDR information at each scene point. Therefore, different exposures of the scene can then be artificially *rendered* as if the scene was acquired multiple times in a traditional (not filtered) mosaicing process. In addition, the HDR information can be processed for high fidelity display, by one of the methods described in Burt and Kolczynski (1993), Durand and Dorsey (2002), Fattal et al. (2002), Larson et al. (1997), Pardo and Sapiro (2002), and Socolinsky (2000).

4. Image Acquisition

Although data acquisition in generalized mosaicing is similar to that of traditional mosaicing, the raw data is unlike any other video sequence. This is due to the spatially varying properties of the optical attachment. We are thus confronted with the following questions:

1. Which configuration of the optical filter is most suitable to gather the information?
2. What is the efficient sampling rate of the scene scan to acquire each point at high intensity definition in a minimal number of frames?
3. How can such images be registered?
4. How can the information from all the registered images be fused to give a single, high definition value to each scene point?
5. How can we achieve auto-calibration of the light modulation of the entire system, given the acquired sequence?

We address the first two questions in this section, while the other ones will be addressed in Sections 5–7.

4.1. The Filter Modulation

The filter has to significantly change the attenuation of light across the camera FOV. From the unlimited possibilities for such filters, we mainly consider two configurations: The simple occluder, and the linear variable density filter. The simple occluder is a piece of opaque material that completely blocks the light rays hitting it. Since it doesn't occlude all the rays that enter the lens (Fig. 7), and because of the finite aperture of the lens, its effect on the light coming from the scene to the detector is gradual. Therefore this is a simple way to obtain intended vignetting.

If the occluding edge is along the y axis, the transmittance varies along the x direction. Let us model the occluder's transmittance as a step edge: $f = 1$ for $x < 0$, and $f = 0$ otherwise. Following (Marshall et al., 1996) the scene radiance undergoes a spatially varying attenuation by a mask M that can be modeled[5] as

$$M(x, y) = f(x, y) * h(x, y). \tag{3}$$

Here $h(x, y)$ is the defocus blur point spread function of the camera for objects as close as the filter, when the system is focused at the distant scene. For circularly symmetric point spread functions the mask is practically a one dimensional function of x, and f is convolved with $\tilde{h}$, the Abbel transform of h. For example, if the kernel is modeled by a Gaussian (Rajagopalan and Chaudhuri, 1995; Surya and Subbarao, 1993) then $\tilde{h}$ is a Gaussian of standard deviation Δx, and $M(x) = \mathrm{erf}(-x/\Delta x)$, as plotted at the top of Fig. 8. Since $M(x)$ takes any value between 0 and 1, then in principle any scene point can be imaged unsaturated, no matter how bright it is, if it is seen through the filter at the appropriate location. Therefore, this

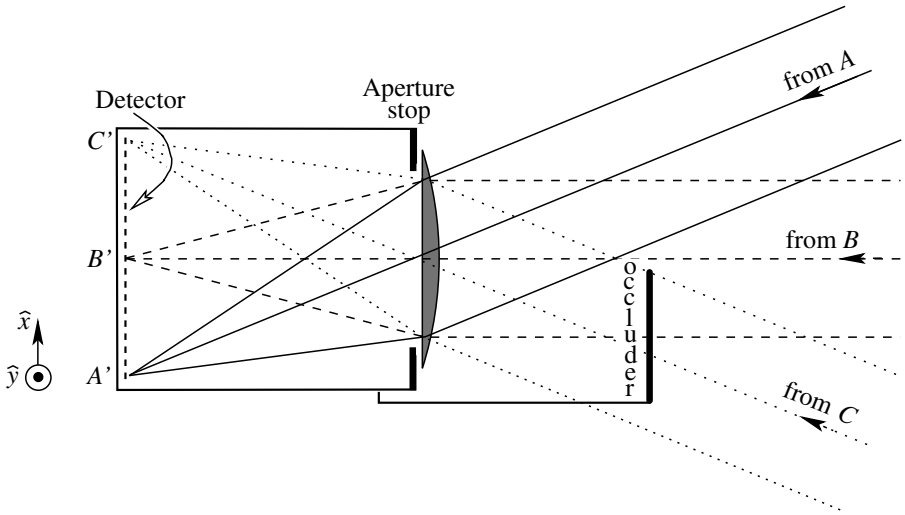

Figure 7. The simple occluder blocks all the light that comes from C, part of the light that comes from B, but doesn't block A. The transition is gradual. Therefore, the system has a strong vignetting effect. This vignetting can be exploited to enhance the dynamic range as the system moves.

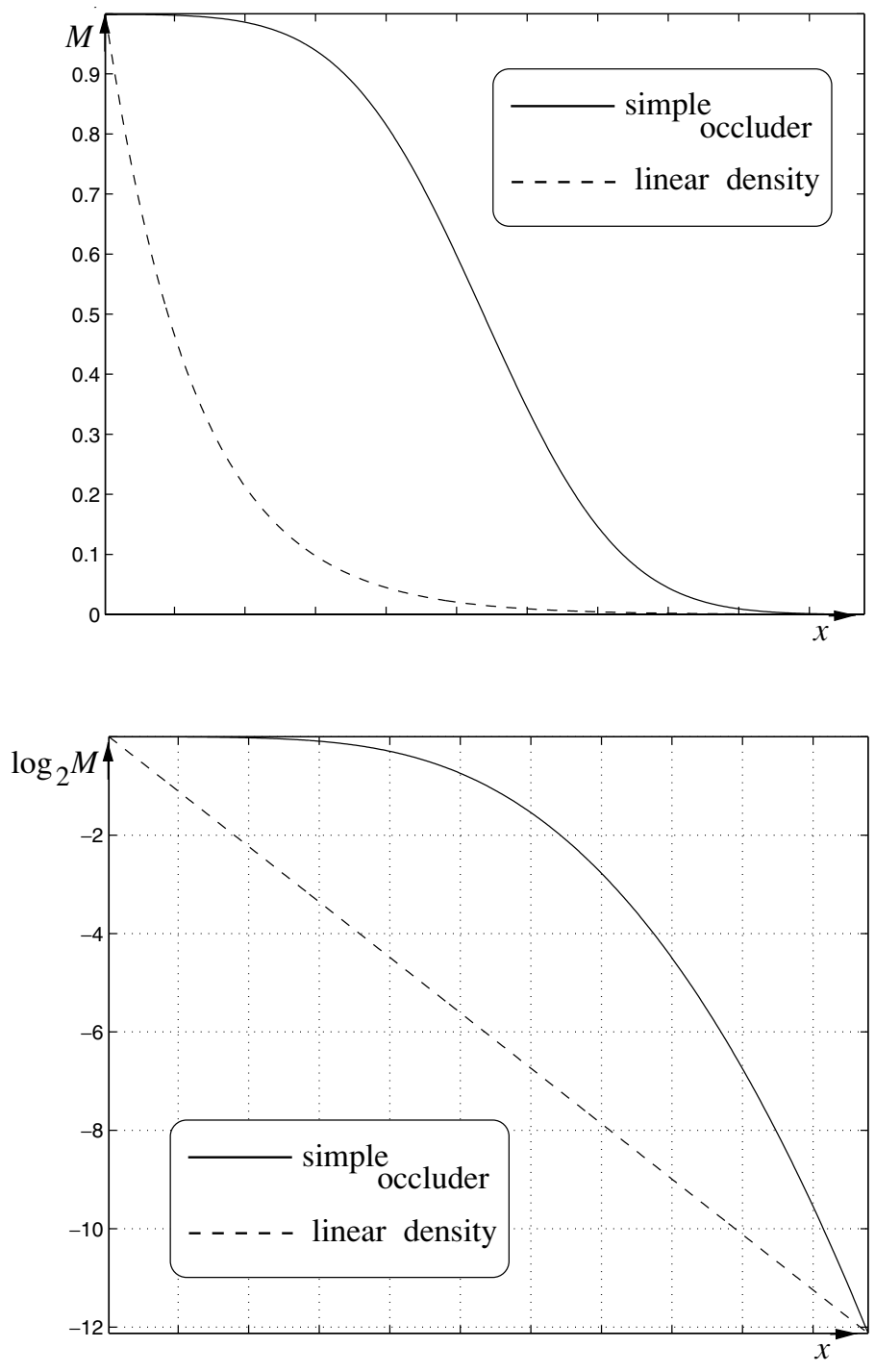

Figure 8. (Top) The effective transmittance mask of a blurred simple occluder (solid line), and the transmittance of a linear variable density filter (dashed). (Bottom) The blurred simple occluder does not change significantly the order of magnitude of the transmittance across a large part of the field of view. The most significant changes occur in a narrow part. In contrast, the order of magnitude of the transmittance changes at a constant rate if a linear density filter is used.

simple system can have, theoretically, an infinite dynamic range.[6]

However, the simple occluder has significant disadvantages. Most notably, the attenuation order of magnitude is not constant across the mask. When we view a scene having high contrast we are interested in the definition of the intensity in *orders of magnitude*. Camera $f_{\#}$ stops are ordered so that each stop increase leads to doubling of the measured intensity. In digital camera detectors, this corresponds to a shift of 1 bit in the binary representation of the measurement: If the measurement of the light intensity in a 8 bit camera is, say, 00011010, then an increase of one stop will lead to a corresponding readout of 0011010(0), where the new least significant bit is the information added by the new image. However, the simple occluder mask does not give us such an equal division of the attenuation by orders of magnitude (see the bottom of Fig. 8). In fact, most of the spatial extent of the mask will yield a very slow change in the order of magnitude of M, while most of the change will occur in a rather narrow region.

The optimal mask for achieving an equal distribution of orders of magnitude is one in which the attenuation changes exponentially across the camera FOV. Then, $\log_2 M(x) \propto x$. In this configuration, a constant scanning increment will yield a constant change in the order of magnitude of the measured intensity, and all intensity ranges are sampled equally (Fig. 8). Such a behavior can be approximately achieved by attaching a *linear variable density filter* (Edmund Industrial Optics, 2002) to the camera at some distance in front of the lens.[7] This is because the filter transmittance (Edmund Industrial Optics, 2002) is $10^{-density}$.

As in the case of the simple occluder, the characteristics of the mask M are a blurred version of those of the filter. This actually adds to the flexibility of the setup: the filter's characteristics can effectively be tuned by changing its smoothing via controlling the lens aperture. Moreover, the linear variable density filter can be approximated by a *stepped density filter* (Edmund Industrial Optics, 2002), since the defocus blur of the steps will make the effective mask have characteristics similar to those of the continuous filter.

4.2. *Efficient Sampling Criteria for Still Images*

Overlapping corresponding areas in different frames can increase the dynamic range, but decrease the rate of FOV expansion. Thus at first sight, it may appear that there is a tradeoff between the FOV and the enhancement of dynamic range. In this section we show that the price paid in FOV expansion is limited. We can distinguish between two cases: video streams and still images. In video, inter-frame motion is typically very small, otherwise motion blur degrades the images. Thus, video cameras move slowly in order to create high quality mosaics. Then, extension of the FOV requires the acquisition of a lot of raw images, even if only a small part of them is actually used for mosaicing. Consequently, generalized mosaicing and dynamic range enhancement do not cause an increase in the number of acquired images.

Raw still images, such as those acquired for remote sensing and astronomy applications, can have arbitrary magnitudes of mutual displacement, or other coordinate transformations. We thus cannot assume that frame displacements are small. Nevertheless, significant areas in the FOV of each frame overlap with other frames in order to facilitate *image registration*. This overlap can then be exploited for extending the radiometric dynamic range. Nevertheless, there may be cases in which the overlap region between frames is insignificant. Then, indeed, extension of the FOV can be achieved with a smaller number of frames, while HDR measurements across the FOV are not performed. In such scenarios we will need more images to extend the dynamic range, than in traditional mosaicing. The question is, how many more frames are needed? Alternatively, we may ask, what should the frame displacements be, or, how many times should each scene point be seen?

Let I be the light intensity that falls on the detector (irradiance) when the transmittance is maximal ($M = 1$). We assume it satisfies $I_{\min}^{\text{detector}} \leq I \leq I_{\max}^{\text{system}}$, i.e., that it is within the dynamic range of the entire system. We need at least one measurement of this point in a state that is not saturated, and at the same time above the detector's threshold (not dark) under the mask operation. We term this state the *effective state*.

We may use a scan for which there is either 1 or 2 effective states for each point in the entire sequence. Consider the spatio-intensity space depicted in Fig. 9. In order to have no point saturated, or too dark at the end of the scan, this space should be covered completely by the spatio-intensity support of the frames. The most efficient way to cover the space is by tiling the frames with minimal spatio-intensity overlap. This yields the *most efficient scan*, where the optical

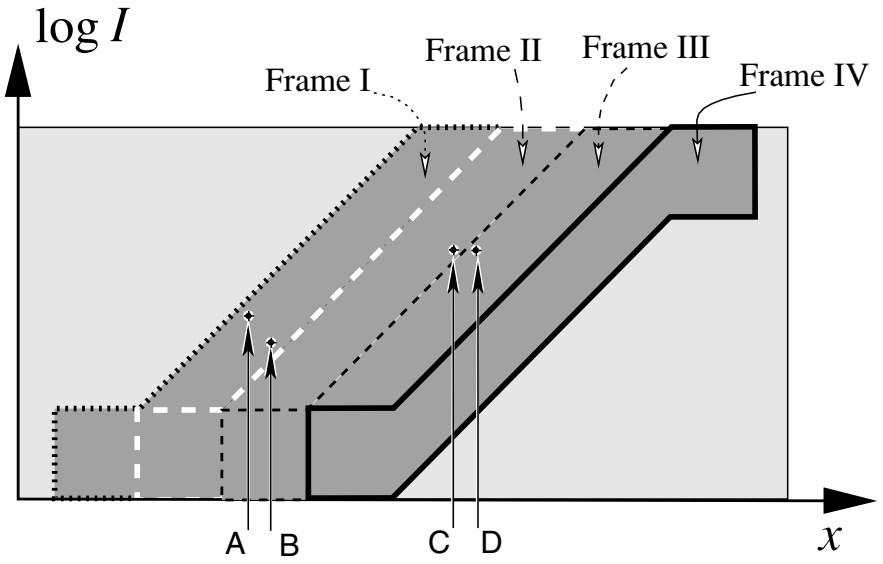

Figure 9. All frames have the same spatio-intensity support. They are tiled with minimal overlap, hence the intensity dynamic range of the measurements is extended maximally between consecutive frames. Most scene point are imaged only once in a state that is unsaturated and at the same time not dark: Points A and B in Frame I, Point C in Frame II, and Point D in Frame III. The other measurements of these points are either too dark or too bright.

dynamic range of the measurements is extended maximally at each increment. Each point should be sampled $\lceil DR^{\text{system}}/DR^{\text{detector}} \rceil$ times.

In Fig. 9, all frames have the same spatio-intensity support as Frame IV. There is some overlap in the tiling at the regions corresponding to max M and min M (the lower and upper intensity ranges). However, for most intensities, each scene point is acquired only once in the effective state. For example, Points A and B are in this state in Frame I, are saturated in Frames II and III, and outside the camera FOV in Frame IV. Point C is in the effective state in Frame II, while being too dark in Frame I and saturated in Frames III and IV.

However, this sampling has obvious shortcomings. A point measured closely above the detector threshold is indeed in an effective state, but is relatively very noisy. This is because the mask attenuated its intensity to be of similar order of magnitude as the detector noise. On the other hand, a point measured closely below the detector saturation (thus also in an effective state[8]) is much more intense than the detector noise. Therefore, this measurement has a high quality, and contains a maximal number of significant bits. For example, in Fig. 9, Points C and D have the same intensity. However, the only effective state measurement of Point C is very dark, hence relatively very noisy. This is in contrast to Point D for which the effective state is bright and thus has high signal to noise ratio. Moreover, the fact that the redundancy between the frames is so minimal may make it hard, if not impossible, to register them. Therefore, it may be better to use a somewhat more dense sampling rate, as described below.

Each change of a full stop (attenuating by 2) is equivalent to a 1 bit shift in the binary representation of the measurement within DR^{detector}. An increment of a factor of 2 between consecutive measurements yields, for each scene point, one measurement relatively close to saturation. The representation of this bright measurement contains a maximal number of significant bits. This leads to a "high quality scan".

To clarify, consider an example in which $DR^{\text{detector}} = 8$ bits and $\min M = 1/64$, hence $DR^{\text{system}} = 14$ bits. Let a scene point yield the binary value 11111111 when $M = 1$ and 10111001 when $M = 1/2$. Then the former measurement is saturated while the latter is not. Obviously for $M = 1/64$ (shift of 5 more bits to the right) the measurement will yield 00000101. The 14 bits representation of the irradiance is thus 0000010111001(0), where the unknown least significant bit is in parenthesis. On the other hand, in the most efficient scan in

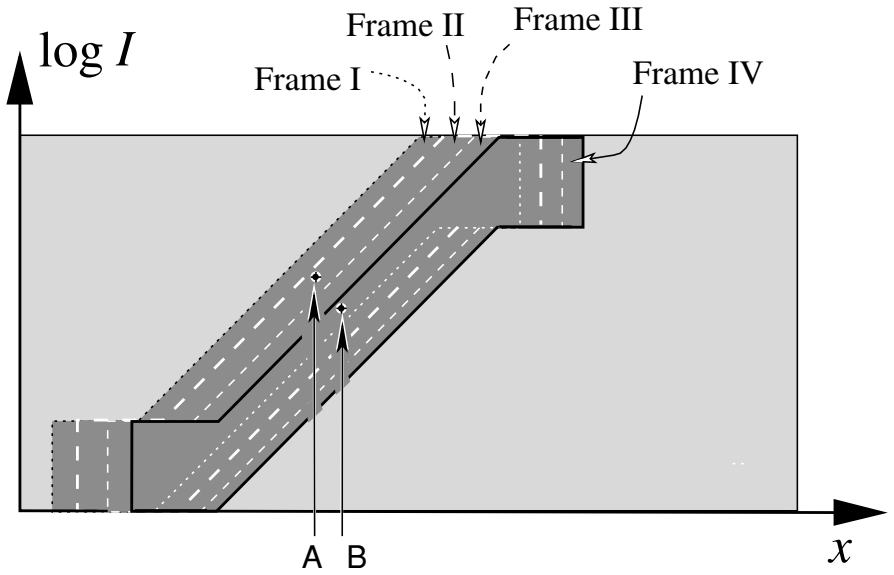

Figure 10. The frames are similar to those in Fig. 9. However, their spatial displacement is smaller, leading to smaller increments of exposures for each scene point. For each point, the best frame is that at which it is brightest, but not saturated. For Point A it is Frame II. For Point B it is Frame IV, since it is sensed darker in the other frames.

which the optical dynamic range is extended maximally at each increment, only the measurements corresponding to $M = 1$ and $M = 1/64$ are used. Then, the resulting 14 bits representation is 00000101(000000) with many more unknown bits (in parenthesis), thus with much lower quality than the result obtained using the high quality scan.

This high quality scan of the spatio-intensity space is depicted in Fig. 10. It is similar to Fig. 9, but the spatial displacement between the frames is smaller, leading to a denser sampling of the intensity range. Point A is in an effective state both in Frame I (dotted circumference) and in Frame II. However, in Frame II it appears bright and thus in high quality. Point B is in an effective state in all frames. However, it is best captured in Frame IV (solid circumference), less so in Frames II and III (dashed circumference) and least in Frame I.

The maximal number of valuable bits is obtained with factor 2 increments of the transmittance. Using the linear variable density filter the high quality sampling increment can be linearly translated to the actual displacement between consecutive frames. When $\max M = 1$ each scene point is imaged at least

$$1 - \log_2(\min M) \tag{4}$$

times. In practice it is desirable to regard Eq. (4) just as a rule of thumb, and use a somewhat denser sampling rate. Redundancy will not only enable less noisy irradiance estimation but is essential for stable registration of image sequences. Moreover, we stress that the motion is general (i.e., not limited to rotation about the camera's center of projection) thus the transformation between frames is generally not periodic translation.

Finally, we would like to point out the possibility for "super-resolution" in the intensity domain. The mask acts as an analogue computational device, and can therefore provide better quantization accuracy than a single measurement in many cases. For example, the image values 96.8 and 97.2 are indistinguishable by the integer quantizer of the camera, and both will yield the output value of 97. However, using $M = 1/2$ the camera yields the integer values 48 and 49, respectively, hence giving the required discrimination. In another example, 95.8 and 96.2 can be distinguished using the result of sensing with $M = 2^{-6}$, that is, using 7 images. In general, the more images used, the better the intensity resolution of the fused estimate. This is another advantage of dense sampling of the scene intensity. In the fusion method which we describe in Section 6, all the multiple measurements of each scene point are indeed fused to improve the accuracy of the estimated image value.

5. Self Calibration of the Effective Mask

5.1. Crude Estimation

Even if the properties of the mask are not calibrated before the image sequence has been acquired, it is possible to calibrate it "on the fly", from the sequence itself, even during the acquisition. We assume that the intensity readout is linearly related to the intensity, and discuss the implications of nonlinear radiometric response in Appendix A.2. Let I be the intensity of light that falls on the detector (irradiance) had the mask been totally transparent ($M = 1$). In the presence of the mask the intensity readout[9] is

$$g(x, y) = M(x, y)I(x, y). \tag{5}$$

Imaging a very large ensemble of scenes, the expectation of the readout at each pixel (x, y) is

$$\langle g(x, y)\rangle = M(x, y)\langle I(x, y)\rangle, \tag{6}$$

where $\langle I(x, y)\rangle$ is the expectation of the intensity at pixel (x, y). We assume that $\langle I(x, y)\rangle$ is a constant across the camera FOV. This is because spatial radiometric effects such as foreshortening and vignetting are accounted for by the mask $M(x, y)$. Therefore the estimated mask is

$$\hat{M}(x, y) \propto \langle g(x, y)\rangle. \tag{7}$$

Since we have a finite sequence of frames $g_k(x, y)$, the expectation is estimated by averaging over the sequence of frames at each pixel. We may get more stable results by incorporating the assumption that the mask variations are mainly along the x axis. Hence the mask is estimated by the average horizontal profile of the frame readouts:

$$\hat{M}(x) \propto \sum_{k=1}^{\text{frames}} \sum_{y} g_k(x, y). \tag{8}$$

We demonstrated this method in an experiment. We used an off-the-shelf linear variable density filter (Edmund Industrial Optics, 2002), 3 inches long, rigidly attached to a CCD camera $\approx$30 cm in front of its 25 mm lens. The filter has a maximum density of 2 (attenuation by 100). We expected the effective mask to have a wider dynamic range due to additional vignetting effects in the system. According to Eq. (4) we had to sample each point 8 times across the range of change of M. Using that as a guideline, the camera was rotated about its center of projection so that each point was actually imaged 14 times[10] across the camera FOV. Some images of this 36 frames sequence are presented in Fig. 11. The radiometric response of the CCD camera was linear.

The frames were simply averaged using Eq. (8), without the need for registration. The estimated mask is shown as a dashed line in Fig. 12. It is apparent in the logarithmic plot, that the estimated mask values are very low in the right (dark) regions of sequence. These values should be taken with caution. The reason is that when $g < 0.5$, the intensity is less than the detector's threshold, and the readout is rounded to $g = 0$ by the detector's integer quantizer. Since many such points exist in the right hand side of the camera FOV, this method tends to underestimate the relative transmittance in this area. Problems are also introduced by saturated areas which clearly violate the linearity expressed in Eq. (5).

5.2. Estimation by Consistency

We now show how to estimate the transmission mask, without relying on the assumption of a large ensemble of frames. The method avoids the problems related to saturated or too dim measurements, encountered with the method described in Section 5.1. We assume for a moment that the images are registered (automatic image registration is discussed in Section 7). Let a scene point be seen in frame k at image point x_k, with unsaturated intensity readout

$$g_k = IM(x_k). \tag{9}$$

Then, this same scene point is measured without saturation in frame p at image pixel x_p, with intensity readout

$$g_p = IM(x_p). \tag{10}$$

Assuming scene radiance is constant between frames, these points should satisfy

$$M(x_k)g_p - M(x_p)g_k = 0. \tag{11}$$

Tracking some of the scene points in several images provides many such linear equations, which the mask should satisfy at each image pixel x. This set of equations can be written as $\mathbf{F}M = 0$. For example, $\mathbf{F}$ may

Figure 11. Frames 4, 9, 11, 15, 17, 23, and 31 from a sequence taken with a linear variable density filter. Scene features become brighter as they move leftwards in the frame. Bright scene points gradually reach saturation. Dim scene points, which are not visible in the right hand side of the frames, become visible when they appear on the left.

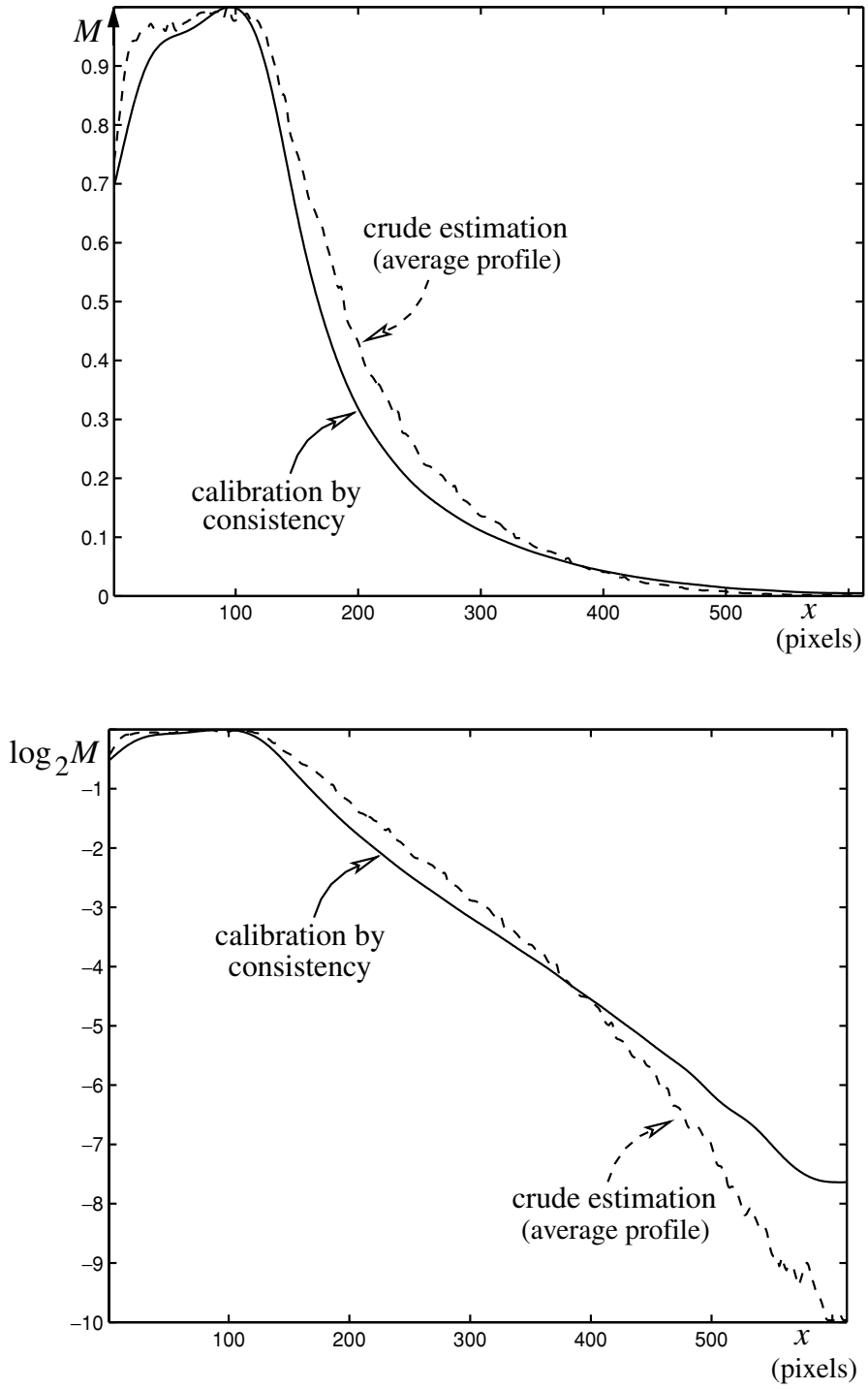

Figure 12. (Top) The effective transmittance mask, self calibrated from the image sequence. The mask function drops on the left due to vignetting by system parts other than the filter, which was constant there. Over most of the field of view, the transmittance approximately decreases exponentially. (Bottom) Solid line: The order of magnitude of the mask changes approximately linearly over most of the camera FOV. It extends the dynamic range of the camera by about 8 bits (factors of 2) beyond that of the detector. Dashed line: Crude estimation by the average horizontal profile yields similar results but underestimates the transmittance at the areas of strong attenuation.

look like

$$\mathbf{F}=\begin{bmatrix} 0 & 0 & g_{p=1}^{x=50} & 0 & .. & .. & .. & 0 & g_{p=2}^{x=3} & 0 & .. & .. & 0 & 0 & 0 \\ 0 & .. & 0 & .. & 0 & g_{p=1}^{x=87} & 0 & .. & .. & .. & 0 & g_{p=2}^{x=40} & 0 & .. & 0 \\ \vdots & & & & & & & & & & & & & & \vdots \\ g_{p=15}^{x=144} & 0 & .. & .. & & & & & & & .. & .. & 0 & g_{p=18}^{x=1} & 0 \end{bmatrix}, \tag{12}$$

where the frame number is indexed by p.

In addition, we impose smoothness on the estimated mask, by penalizing for 2nd order variations in M (e.g., the Laplacian). The smoothest mask would satisfy $\mathbf{L}M = 0$, where

$$\mathbf{L}=\begin{bmatrix} 1 & -2 & 1 & 0 & .. & .. & .. & .. & .. & .. & 0 \\ 0 & 1 & -2 & 1 & 0 & .. & .. & .. & .. & .. & 0 \\ 0 & 0 & 1 & -2 & 1 & 0 & .. & .. & .. & .. & 0 \\ \vdots & & & & & & & & & & \vdots \\ 0 & .. & .. & .. & .. & .. & .. & 0 & 1 & -2 & 1 \end{bmatrix}. \tag{13}$$

The result of Eqs. (12) and (13) is an overconstrained system of equations. The least squares solution of this system of equations is

$$\hat{M} = \arg\min_{M}(M^{\mathrm{t}}\mathbf{A}^{\mathrm{t}}\mathbf{A}M), \tag{14}$$

where

$$\mathbf{A}=\begin{bmatrix} \mathbf{F} \\ \beta\mathbf{L} \end{bmatrix}, \tag{15}$$

and β is a parameter that weights the penalty for unsmooth solutions relative to the penalty for disagreement with the data.

Singular value decomposition yields the nontrivial solution up to a scale factor. The scale is set by letting $\max \hat{M} = 1$. These equations also enable the estimation (Chapra and Canale, 1998) of the covariance matrix of M:

$$Cov(M) = (\mathbf{A}^{\mathrm{t}}\mathbf{A})^{-1}\hat{M}^{\mathrm{t}}\mathbf{A}^{\mathrm{t}}\mathbf{A}\hat{M}(n+1-l)^{-1}, \tag{16}$$

where l is the number of elements of M, and n is the number of rows in $\mathbf{A}$. This can be viewed as a weighted least squares problem: rows that belong to $\mathbf{L}$ are weighted by β, while the rows that belong to $\mathbf{F}$ are generally more accurate for more intense pixels (larger g). This is equivalent to using normalized rows, and then weighting each row r by $\sqrt{\sum_c \mathbf{A}^2(r,c)}$, where c is the column index. To accommodate this, we actually used $n = \sum_{r,c} \mathbf{A}^2(r,c)$, i.e., summing the squared weights of each row.

The variance of M given by the diagonal of $Cov(M)$ leads to the uncertainty estimates $\Delta\hat{M}(x)$ of the mask $\hat{M}(x)$. Note that this formulation is not in the $\log M$ domain. Thus, it does not penalize strongly for relative disagreements with the data at very low M, or fluctuations which may be *relatively* significant at low M. So, a final post-processing, smoothing of $\log \hat{M}$ is also performed.

This self-calibration method was demonstrated on the same sequence, samples of which are shown in Fig. 11. We registered the sequence of frames using the method described in Section 7. There are millions of corresponding pixel pairs in the sequence. From them we obtained about 50,000 equations as Eq. (11) based on randomly picked corresponding pairs, to determine the mask. To avoid the problems encountered in Section 5.1, each image point used for this estimation was unsaturated and also non-dark (i.e., in an effective state). The self-calibrated mask is plotted by the solid line in Fig. 12. The mask enables the extension of dynamic range by about 8 bits beyond the intrinsic dynamic range of the detector. Therefore, using an ordinary 8 bit camera we can obtain image mosaics with dynamic range close to that produced by a 16 bit camera.

6. Fusing the Measurements

We now describe the method we used to estimate the intensity at each mosaic point, given its multiple corresponding measurements. As in Section 5.2, this is done after the images have been registered. Let a measured intensity readout at a point be g_k with uncertainty Δg_k, and the estimated mask be $\hat{M}$ with uncertainty $\Delta\hat{M}$. Compensating the readout for the mask, the scene

point's intensity is

$$I_k = \frac{g_k}{\hat{M}} \tag{17}$$

and its uncertainty is[11]

$$\Delta I_k = \sqrt{\left(\frac{\partial I_k}{\partial g_k}\Delta g_k\right)^2 + \left(\frac{\partial I_k}{\partial \hat{M}}\Delta \hat{M}\right)^2}. \tag{18}$$

We assumed the readout uncertainty to be $\Delta g_k = 0.5$, since the intensity readout values are integers. Any image pixel considered to be saturated (g_k close to 255 for an 8 bit detector) is treated as having high uncertainty, thus its corresponding Δg_k is set to be a very large number.

Assuming the measurements I_k to be Gaussian and independent, the log-likelihood for a value I behaves like $-E^2$, where

$$E^2 \equiv \sum_k \left(\frac{I - I_k}{\Delta I_k}\right)^2. \tag{19}$$

The maximum likelihood (ML) solution[12] for the intensity I of this scene point is the one that minimizes E^2:

$$\hat{I} = \widehat{\Delta I}^2 \sum_k \frac{I_k}{\Delta I_k^2}, \tag{20}$$

where

$$\widehat{\Delta I} = \left(0.5 * \frac{\mathrm{d}^2 E^2}{\mathrm{d}I^2}\right)^{-1/2} = \left(\sqrt{\sum_k \frac{1}{\Delta I_k^2}}\right)^{-1}. \tag{21}$$

Although Eq. (20) suffices to determine the value at each point, annoying seams may appear at the boundaries of the frames that compose the mosaic. At these boundaries there is a transition between points that have been estimated using somewhat different sources of data. Seams appear also at the boundaries of saturated areas, where there is an abrupt change in the uncertainty Δg, while the change in g is usually small. These seams are removed by feathering techniques discussed in Appendix A.3.

The images from the sequence, of which samples are shown in Fig. 11, were fused into a mosaic using this method. The histogram equalized version of $\log \hat{I}$ is shown in Fig. 13. Contrast stretching of $\hat{I}$ in selected regions shows that the mosaic is not saturated anywhere, and details are seen wherever $I \geq 1$. The HDR of this mosaic is expected due to the fact that it was created by a generalized mosaicing system with dynamic range of 16 bits.

The periphery parts of the mosaic are left of the $\mathcal{L}$ mark and right of the $\mathcal{R}$ mark, having a width of a single frame. These parts were not exposed at the full range of attenuation. Still, in most of these areas there is no noticeable deterioration in the dynamic range. HDR is

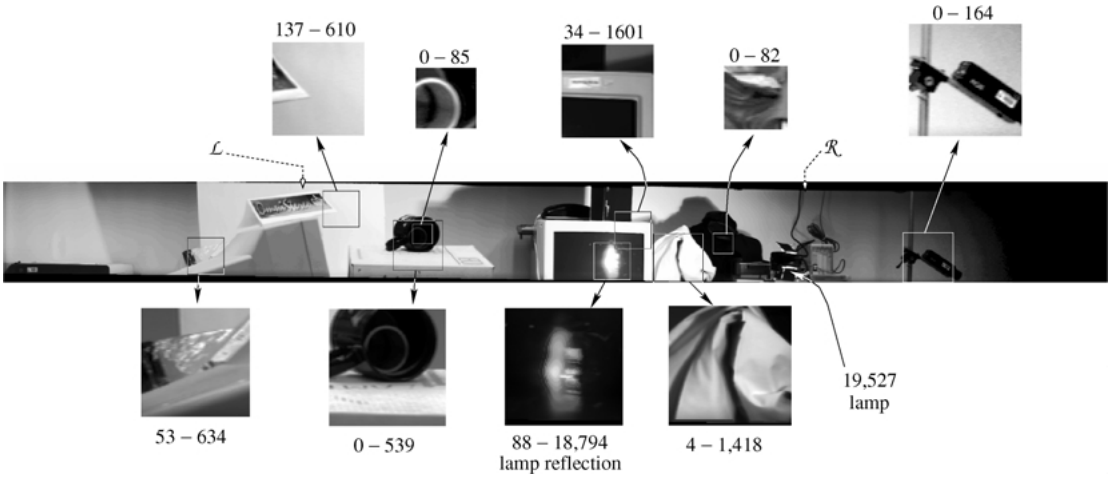

Figure 13. An image mosaic of 51° horizontal FOV, created with a generalized mosaicing system having dynamic range of 16 bits. It is based on a single rotation about the center of projection of an 8 bit video camera. Contrast stretching in the selected squares reveals the details that reside within the computed mosaic. The numbers near the squares are the actual (unstretched) brightness ranges within the squares. Note the shape of the filament of the lamp in its reflection from the computer monitor. The periphery regions are left of the $\mathcal{L}$ mark and right of the $\mathcal{R}$ mark.

observed there as well, as the range gradually decreases to DR^{detector} towards their outer boundaries. Only right of the $\mathcal{R}$ mark it is possible that points with $I \geq 1$ are not detected.

7. Image Registration

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. Traditional registration methods are typically based on optimizing the sum of square or absolute difference (Coorg and Teller, 2000; Irani et al., 1996; Irani and Anandan, 1998b), correlation (Duplaquet, 1998; Hansen et al., 1994; Kwok et al., 1990) or mutual information (Thevenaz and Unser, 2000; Viola and Wells III, 1997) between frames, over the general motion parameters. For generalized mosaicing, the image registration algorithm should cope with the phenomena induced by the filtering.

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. 11. Although features appear to be moving through the camera's FOV, the static mask clearly dominates the images. This would bias traditional algorithms towards estimating a motion slower than the true one, as demonstrated in Fig. 14(a). The mask varies gradually across the image, thus high-pass filtering the raw images as in Irani and Anandan (1998a) and Sharma and Pavel (1997) reduces the biasing effect. Nevertheless, even then it isn't completely removed, according to our experience.

Measurements of scene points that become too dark due to strong attenuation are relatively noisy after quantization and other processes undergone 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 adapted a traditional technique so it can cope with the spatially varying filtering effects. The algorithm maximizes the likelihood of the matched data, and does not suffer from the biasing problem. Similarly to traditional algorithms, the optimization can be done over any number of motion parameters (which depends on the complexity of motion we wish to describe). We note that registering ordinary (not filtered) images by minimizing their mean squared difference is obtained as a special case of this algorithm. The following is a brief description of the aspects of the algorithm, that were different from those of traditional registration methods.

1. Each frame $g(x, y)$ is roughly flat fielded using $1/\hat{M}(x)$ to estimate $I(x, y)$ as in Eq. (17). This is done given a rough estimate of the mask $\hat{M}$. We estimated M using Eq. (8) with the method described in Section 5.1, and initially set the mask uncertainty to be $\Delta\hat{M} = 0.01$. Alternatively, M can be grossly estimated by the information supplied by the filter's manufacturer, or by interpolating a few measurements beforehand. Given $\Delta\hat{M}$ and the readout uncertainty Δg, ΔI is estimated using Eq. (18). In case a spatially varying filter is not present (no vignetting exists), $M \equiv 1$, thus ΔI is constant. Note that if we attempt to register the images $I(x, y)$ by minimizing their mean squared difference, we may fail. This is demonstrated in Fig. 14(b). To counter that, we should incorporate the spatially varying uncertainties as follows.
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. As in Eq. (19), the squared distance between this pair of pixel measurements is

$$\hat{E}^2_{\text{pixel pair}} = \left(\frac{\hat{I} - I_1}{\Delta I_1}\right)^2 + \left(\frac{\hat{I} - I_2}{\Delta I_2}\right)^2. \tag{22}$$

where

$$\hat{I} = \widehat{\Delta I}^2 \sum_{k=1,2} \frac{I_k}{\Delta I_k^2}, \tag{23}$$

and

$$\widehat{\Delta I}^2 = \left(\sum_{k=1,2} \frac{1}{\Delta I_k^2}\right)^{-1} \tag{24}$$

as in Eqs. (20) and (21). The distance measure for the entire images is

$$\hat{E}^2_{\text{total}} = \sum_{\text{all pixels}} \hat{E}^2_{\text{each corresponding pair}}. \tag{25}$$

The best registration between two frames (or between a new frame and an existing mosaic) according to this objective function is the one that minimizes $\hat{E}^2_{\text{total}}$. If the measurements are Gaussian and independent, this is the *most likely match*. When the spatially varying filter is not present, $\Delta I_1 = \Delta I_2$ and

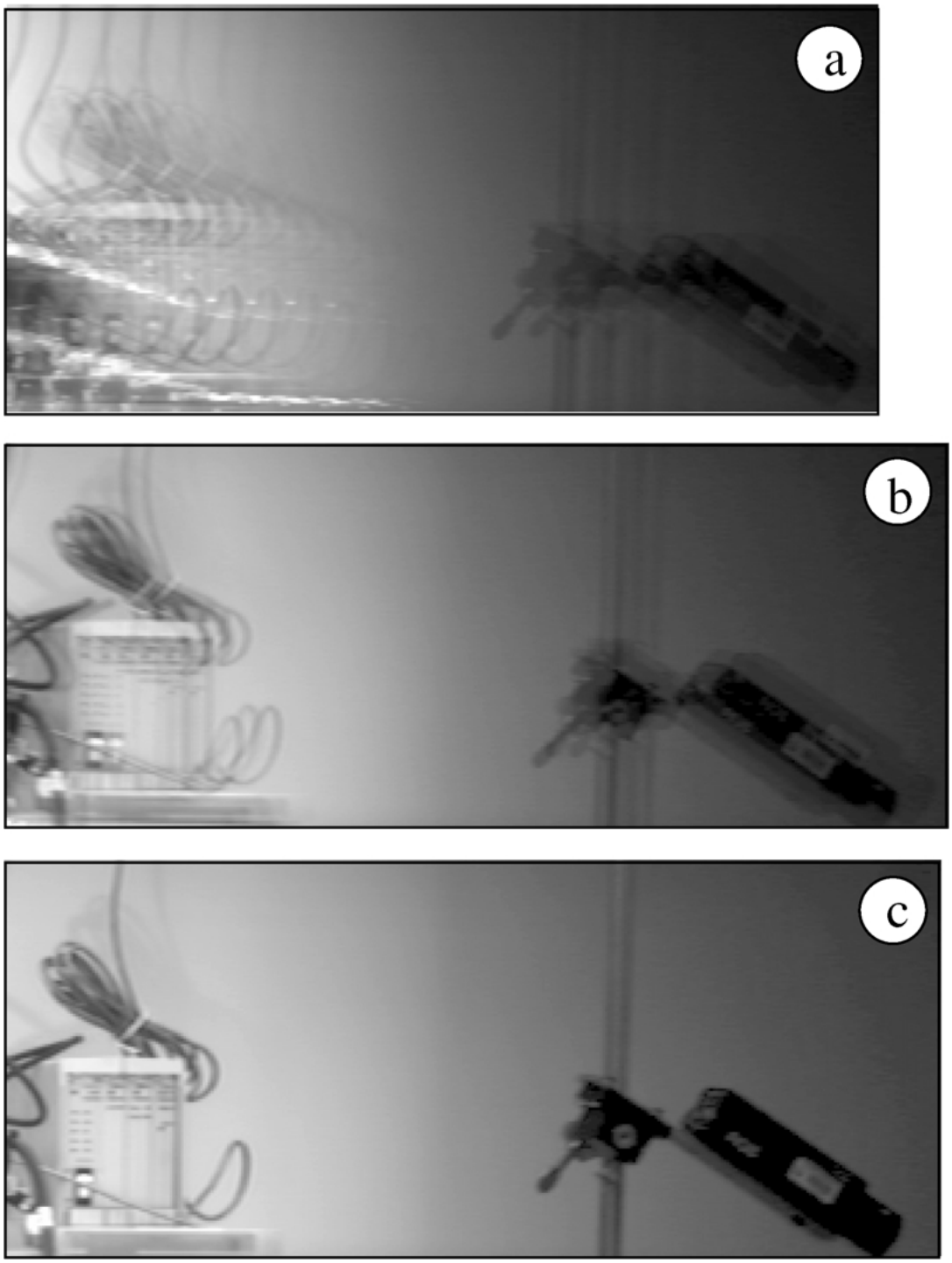

Figure 14. Corresponding mosaic regions. (a) Attempting to register the raw images by minimizing their squared difference fails. (b) Compensating for the spatially varying attenuation without accounting for uncertainty amplification does *not* lead to a good registration. (c) Accounting for spatially varying uncertainties due to the attenuation compensation leads to successful image registration.

Eq. (22) becomes proportional to the squared difference between the measurements, hence Eq. (25) 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. This may bias the registration towards minimizing the number of pixels in the sum, hence reducing the overlap between the images. To counter that, Eq. (25) may be normalized by the number of pixels in the overlap area. However, also when defining this normalization, it is worth noting that some pixels in the overlap are significant, while others make negligible contribution to $\hat{E}^2_{\text{total}}$ due to their high uncertainty. Thus we may normalize $\hat{E}^2_{\text{total}}$ by dividing it by $\Delta I^{-2}_{\text{total}}$, where we define

$$\Delta I^{-2}_{\text{total}} \equiv \sum_{\text{all pixels}} \frac{1}{\widehat{\Delta I}^2_{\text{each pixel pair}}}, \tag{26}$$

while for each corresponding pixel pair $\widehat{\Delta I}^2$ is given by Eq. (24). When the spatially varying filter is not present, $\widehat{\Delta I}^2$ is the same for all the pixel pairs, so $\Delta I^{-2}_{\text{total}}$ is proportional to the number of pixels in the overlap area. In this case this normalization makes Eq. (25) proportional to the *mean* square difference between the images.

If the statistical dependence between different measurements or between different pixels cannot be neglected, then Eqs. (22) and (25) can be generalized to use the covariance matrix of the measurements, rather than just their variances.

3. The registration is done hierarchically, from coarse to fine resolution similar to Hansen et al. (1994), Irani and Anandan (1998a, 1998b), Sawhney et al. (1998), and Sharma and Pavel (1997). 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, ΔI_1, ΔI_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. Registering only pairs of frames leads to the accumulation of errors of the estimated image positions in the global coordinate system. To reduce the accumulation of matching errors, each new sequence frame is registered to the current mosaic (Irani et al., 1996; Sawhney et al., 1998), and then fused into it (see Section 6). The consecutive frame is registered to the updated mosaic.
5. In HDR data, we prefer to penalize for relative errors rather than absolute ones. To do that, the log of each measured scene radiance is calculated:

$$s(x, y) = \log[I(x, y)] \tag{27}$$

with the uncertainty

$$\Delta s = \left|\frac{\mathrm{d}s}{\mathrm{d}I}\right| \Delta I = \frac{\Delta I}{I}. \tag{28}$$

Therefore, we applied the above algorithm to the $s(x, y)$ images rather than the intensity images $I(x, y)$.

Figure 14(c) shows that the registration by our ML algorithm leads to much better results, than cases where the spatially varying uncertainties are not accounted for.

8. Discussion

We propose generalized mosaicing as a framework for capturing additional information about the scene, while requiring a similar or even the same amount of data as in the case of traditional mosaicing. Here we used this framework to compute HDR mosaics. However, generalized mosaicing is not limited to this dimension of imaging but is a general concept that can be applied to other valuable dimensions. In particular (Schechner and Nayar, 2002) used a spatially varying spectral filter to obtain wide FOV multispectral mosaics. The information gathered during the acquisition of images for a mosaic can be done in additional dimensions (e.g., polarization, focus) using other kinds of spatially varying filters.

This paper shows a novel and simple way to create image mosaics while significantly extending the optical dynamic range of cameras. The dynamic range is extended at each scene point, irrespective of the brightness of its surroundings. The system has no internally moving parts, and its rigid motion is the same as that required to create a traditional image mosaic. It can be applied also to cameras that use other means to enhance

the dynamic range, such as active CMOS detectors or AGC.

Since any video or still camera can have its dynamic range extended simply by attaching such a filter in front of the lens (or exploiting vignetting effects) along with automatic image analysis, we expect this to have direct applications in amateur and professional photography. The usefulness of the approach is especially relevant due the increasing interest in image mosaics (Hsu et al., 2002; Irani et al., 1996; Peleg et al., 2001; Shum and Szeliski, 2000). Note that mosaicing is used extensively in various scientific fields (Ballard, 2001; Hansen et al., 1994; Kwok et al., 1990; Lada et al., 1991; Reynoso et al., 1995; Vasavada et al., 1998). Therefore, also these fields can benefit from this generalization. For example, mosaicing is a common technique in observational astronomy. However, the stars are very bright while their surrounding matter is very faint. It is therefore beneficial if the brightness dynamic range of the imaging system can be extended by the motion used to scan the sky for the mosaic creation.

Generalized mosaicing can also be used in conjunction with structure from motion (Tomasi and Kanade, 1992) methods. In current algorithms for structure from motion, the imaged points are not attenuated during the motion, thus there exists a significant redundancy in the sequence. Using a spatially varying filter will remove some of this redundancy, extending the dynamic range of the sensed scene in conjunction with its structure estimation. Therefore, better data is made available for later texture mapping on the computed depth map to render new views of the scene. The vignetting effects may be further exploited in conjunction with the geometric self-calibration of the camera (Kang and Weiss, 2000).

The method described in this paper is very flexible. If the user wants to change the characteristics of the filtering, he may simply change the external filter. For example, using a filter with a wider range of transmittance order of magnitude can easily extend further the dynamic range. Hence if $M \in [10^{-4}, 1]$ the dynamic range of the camera is enhanced by $\approx$13 bits beyond the detector's intrinsic range (equivalent to 80 dB improvement). Note that neutral density is an additive quantity, so stacking two identical filters to one another can double their density (Edmund Industrial Optics, 2002). For example, the filter we used in the experiment had a maximum density of 2. Two such filters stacked have a maximum density of 4, and since the transmittance is $10^{-density}$ such a dynamic range is easy to obtain.

The filtering characteristics may also be easily changed by inclining the filter relative to the optical axis. Due to the projective transformation that the inclined filter plane undergoes, we may vary the part of the camera FOV that corresponds to one end of the filter, relative to the other parts. For example, a larger area in the frames can be dedicated to the stronger attenuation region, hence more information can be gathered on the brighter light sources than on the dim ones. Since the filter is external it can be set at any azimuthal angle around the optical axis. Note that the spatially varying filter may reflect the light into the camera rather than transmitting it, which may find use in catadioptric systems. More complicated implementations of the concept presented in this paper can be made by mounting the filter inside the camera or among its imaging optics. However, this may limit the flexibility of changing the generalized mosaicing system characteristics.

Appendix

A.1. The Maximum Likelihood Pyramid

To make image registration more robust and efficient, it is done hierarchically, from coarse to fine resolution. For the algorithm described in Section 7 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 (Burt and Adelson, 1983b) by sub-sampling 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[13]:

$$I = \frac{\sum_k \omega_k I_k}{\sum_k \omega_k}. \tag{29}$$

In a conventional pyramid, the weights ω_k are equal to the values of a Gaussian kernel a_k (Burt and Adelson, 1983b).

Suppose, however, that we let the weight ω_k assigned to a pixel decrease linearly both as its Gaussian weight a_k decreases, and as its squared uncertainty ΔI_k^2 increases. Thus

$$\omega_k = \frac{a_k}{\Delta I_k^2} \tag{30}$$

and we set

$$\Delta I = \left(\sqrt{\sum_k \omega_k} \right)^{-1}. \tag{31}$$

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. Thus Eq. (29) yields the traditional Gaussian pyramid representation for the pixels. On the other hand, in the special case for which $a_k \equiv 1$ for all pixels k, Eqs. (29) and (31) become equal to Eqs. (20) and (21), respectively. Thus Eq. (29) yields the ML representation of the pixels, assuming them to be Gaussian and independent.

Therefore, Eqs. (29) and (30) 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. Since points that have smaller uncertainties get more weight in this pyramid, the uncertain areas (such as the saturated regions) shrink at the coarser levels. The representation of such areas is more influenced by adjacent stable points. Thus, the representation of a region by one value is more reliable.

A.2. Nonlinear Response of the Detector

The analysis in this paper was based on the assumption that the intensity readout g is proportional to the irradiance that falls on the detector. The following are some modifications in case the response of the detector is nonlinear and known. Let the response function be R. This response can be estimated a priori by one of the methods described in Debevec and Malik (1997), Grossberg and Nayar (2002), Mann and Picard (1995), and Mitsunaga and Nayar (1999). Then, the image readout at the pixel is

$$\tilde{g} = R(g). \tag{32}$$

Linearizing the response, the estimated intensity readout at the detector is

$$\hat{g} = R^{-1}(\tilde{g}), \tag{33}$$

and thus

$$\Delta \hat{g} = \left| \frac{\mathrm{d}R^{-1}}{\mathrm{d}\tilde{g}} \right| \Delta \tilde{g}. \tag{34}$$

In section 6 we assumed the readout uncertainty to be $\Delta g = 0.5$ since the intensity readout values are integers. In the nonlinear response case, we set $\Delta \tilde{g} = 0.5$ for unsaturated points, and use $\Delta \hat{g}$ in Eq. (18). Saturation is determined by the vicinity of $\tilde{g}$ to the saturated readout. We use $\hat{g}$ in Eqs. (9)–(12) and (17).

Note that in Item 5 of Section 7 we actually did not use the intensity but its log in the registration process. Thus, if the response R of the detector is logarithmic as in C-Cam sensors, Davis (1998), Ogiers (1997), and Schwartz (1999), we may set $s = \tilde{g}$ and $\Delta s = \Delta \tilde{g}$.

A.3. Removal of Seams

At the boundaries of the frames that compose the mosaic there is an abrupt transition between the data sources of the computed mosaic. This causes seams to appear in the mosaic. There are numerous ways to remove the seams. One approach is based on searching for an optimal seam line, as in Duplaquet (1998). The other approach is to use feathering, as in Burt and Adelson (1983a) and Shum and Szeliski (2000), in which the images are weighted according to the pixel position from the image centers or boundaries. This weighting fits easily into our ML estimation; the uncertainty ΔI_k is multiplied by a factor that smoothly increases to ∞ towards the image boundaries.

Seams appear also at the boundaries of saturated areas, where there is an abrupt change in the uncertainty Δg, while the change in g is usually small. These seams may also be removed by feathering. For this purpose we needed a smooth transition towards areas considered saturated. Therefore, we created a fuzzy definition of the saturated areas. The areas treated for this phenomena were either saturated or in the neighborhood of such points. The principle we used is as follows: We define "low" and "high" saturation values, L and H, respectively. If $g > H$ it is saturated. If $H \geq g > L$ and this point is connected (a neighbor) to a saturated point, it is considered a "saturation-associated". It is also considered a "saturation-associated" if it is a neighbor of another "saturation-associated" point and $H \geq g > L$. Eventually, points that are "saturation-associated" are formed in groups that are always connected to

a saturated point. On the other hand, intense points ($g > L$) that are not related to saturated points are not treated for saturation feathering, just as less intense points ($g \leq L$). In our experiment we used an 8-bit camera, so we set $H = 250$ and $L = 230$.

After all the "saturation-associated" points have been found, their uncertainty is multiplied by $(H - L)/(H - g)$. This gradually makes a transition from a regular uncertainty (multiplied by a factor of 1, if $g \leq L$ or if not connected to a saturated area), to a very large uncertainty as g reaches the saturated value H.

Acknowledgments

Yoav Schechner is a Landau Fellow—supported by the Taub Foundation, and an Alon Fellow. This work was supported in parts by a National Science Foundation ITR Award, IIS-00-85864, a DARPA/ONR MURI Grant, N00014-95-1-0601, the Louis Morin Fellowship, a David and Lucile Packard Fellowship, and the Ollendorff Center in the Department of Electrical Engineering at the Technion.

Notes

1. In different communities the terms *mosaicing* (Capel and Zisserman, 1998; Peleg et al., 2001) and *mosaicking* (Batson, 1987; Duplaquet, 1998; Eustice et al., 2002; Garcia et al., 2001; Kwok et al., 1990) are used.
2. The filter is not placed right next to the lens, as this would only affect the aperture properties (Farid and Simoncelli, 1998), without producing spatially varying effects in the image.
3. We presented partial results of this work in a conference paper (Schechner and Nayar, 2001), which was published in parallel to Aggarwal and Ahuja (2001).
4. For simplicity we assumed filter variations along one spatial dimension. The results can be generalized to 2D filter variations. In the experiment, vignetting effects along the y axis were negligible due to the filter's aspect ratio.
5. For simplicity we ignore the reflection transformation the object points undergo when projected to the image plane. Inverting the coordinate system is straightforward.
6. As the dimensions of the light bundle allowed into the camera decrease, diffraction effects eventually become dominant. This significantly complicates the later image analysis. Thus, practically the optical dynamic range of the system is limited.
7. Due to vignetting in the lens, perspective and lens distortions, the linearity of $\log f(x)$ will not be accurately conserved in $\log M(x)$.
8. Assuming it is not under the effect of blooming from adjacent pixels.
9. We make the intensity I unit-less by normalizing it by the minimal irradiance that the detector can sense above its noise level, $I_{\min}^{\text{detector}}$. Thus also the readout g is unit-less and has the same normalization.
10. Some measurements were redundant since the filter's transmission was constant across part of the camera FOV.
11. This estimate for ΔI_k assumes that the estimated mask is independent of the signal. Even if the mask is estimated using many points from the sequence, its dependence on a single intensity measurement is small.
12. An improved solution may be obtained using maximum a posteriori (MAP) estimation. However, such estimation requires the prior probability of an intensity value I. We are not aware of a prior reliable enough.
13. 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. However, usually pyramids are constructed iteratively and then the pixels in the intermediate levels are not statistically independent. If accuracy is sought in the iterative process, the weighting should rely not only on the pixel variance in the intermediate levels, but on their full covariance matrix with their neighbors. This matrix should be thus be propagated up the pyramid as well.

References

Adelson, E.H. and Bergen, J.R. 1991. The plenoptic function and the elements of early vision. In *Computational Models of Visual Processing*, M. Landy and J.A. Movshon (Eds.). MIT Press: Cambridge, MA, pp. 3–20.

Aggarwal, M. and Ahuja, N. 2001. High dynamic range panoramic imaging. In *Proc. ICCV*, Vancouver, Canada, vol. I, pp. 2–9.

Ballard, R.D. 2001. Ancient Ashkelon. *National Geographic*, 199(1):61–93.

Batson, R.M. 1987. Digital cartography of the planets: New methods, its status, and its future. *Photogrammetric Engineering and Remote Sensing*, 53:1211–1218.

Bernstein, R. 1976. Digital image processing of earth observation sensor data. *IBM Journal of Research and Development*, 20(1):40–57.

Burt, P.J. and Adelson, E.H. 1983a. A multiresolution spline with application to image mosaics. *ACM Trans. on Graphics*, 2:217–236.

Burt, P.J. and Adelson, E.H. 1983b. The Laplacian pyramid as a compact image code. *IEEE Trans. on Communications*, 31:532–540.

Burt, P.J. and Kolczynski, R.J. 1993. Enhanced image capture through fusion. In *Proc. Int. Conf. Comp. Vis.*, Berlin, Germany, pp. 173–182.

Capel, D. and Zisserman, A. 1998. Automated Mosaicing with super-resolution zoom. In *Proc. Comp. Vis. Patt. Rec.*, Santa Barbara, CA, pp. 885–891.

C-Cam sensors. http://www.vector-international.be/C-Cam/sensors.html

Chapra, S.C. and Canale, R.P. 1998. *Numerical Methods for Engineers with Programming and Software Applications*, 3rd ed., WCB/McGraw-Hill: Boston, pp. 443, 444, 463–467.

Coorg, S. and Teller, S. 2000. Spherical mosaics with quaternions and dense correlation. *Int. J. Comp. Vision*, 37:259–273.

Curlander, J.C. 1984. Utilization of spaceborne SAR data for mapping. *IEEE Trans. on Geoscience and Remote Sensing*, 22:106–112.

Davis, A. 1998. Logarithmic pixel compression. *TechOnLine Review* 2(3).

Debevec, P.E. and Malik, J. 1997. Recovering high dynamic range radiance maps from photographs. In *Proc. SIGGRAPH*, Los Angeles, CA, pp. 369–378.

Duplaquet, M.L. 1998. Building large image mosaics with invisible seam lines. In *Proc. SPIE Visual Information Processing VII*, vol. 3387, Orlando, FL, pp. 369–377.

Durand, F. and Dorsey, J. 2002. Fast bilateral filtering for the display of high-dynamic-range images. *ACM Transactions on Graphics*, 21(3):257–266.

Edmund Industrial Optics 2002. *Optics and Optical Instruments Catalog*, p. 89, Stock # E41-960, E32-599.

Eustice, R., Pizarro, O., Singh, H., and Howland, J. 2002. UWIT: Underwater Image Toolbox for optical image processing and mosaicking in MATLAB. In *Proc. IEEE Int. Sympos. on Underwater Technology*, Tokyo, Japan, pp. 141–145.

Farid, H. and Simoncelli, E.P. 1998. Range estimation by optical differentiation. *J. Opt. Soc. Amer. A*, 15:1777–1786.

Fattal, R., Lischinski, D., and Werman, M. 2002. Gradient domain high dynamic range compression. *ACM Transactions on Graphics*, 21(3):249–256.

Garcia, R., Batlle, J., Cufi, X., and Amat, J. 2001. Positioning an underwater vehicle through image mosaicking. In *Proc. IEEE Int. Conf. on Robotics and Automation*, Seoul, South Korea, part 3, pp. 2779–2784.

Grossberg, M.D. and Nayar, S.K. 2002. What can be known about the radiometric response from images? In *Lecture Notes in Computer Science, Proc. ECCV*, vol. 2353, Copenhagen, Denmark, part IV, pp. 189–205.

Hansen, M., Anandan, P., Dana, K., van der Wal, G., and Burt, P. 1994. Real-time scene stabilization and mosaic construction. In *Proc. 2nd IEEE Workshop on Applications of Computer Vision*, Sarasota, FL, pp. 54–62.

Hsu, S., Sawhney, H.S., and Kumar, R. 2002. Automated mosaics via topology inference. *IEEE Computer Graphics and Application*, 22(2):44–54.

Irani, M. and Anandan, P. 1998a. Robust multi-sensor image alignment. In *Proc. Int. Conf. Comp. Vis.*, Mumbai, India, pp. 959–966.

Irani, M. and Anandan, P. 1998b. Video indexing based on mosaic representations. *Proceedings of the IEEE*, 86:905–921.

Irani, M., Anandan, P., Bergen, J., Kumar, R., and Hsu, S. 1996. Efficient representations of video sequences and their application. *Signal Processing: Image communication*, 8:327–351.

Kang, S.B. and Weiss, R. 2000. Can we calibrate a camera using an image of a flat, textureless Lambertian surface? In *Proc. ECCV*, Dublin, Ireland, part 2, pp. 640–653.

Khoo, I.C., Wood, M.V., Shih, M.Y., and Chen, P.H. 1999. Extremely nonlinear photosensitive liquid crystals for image sensing and sensor protection. *Optics Express*, 4:432–442.

Kwok, R., Curlander, J.C., and Pang, S. 1990. An automated system for mosaicking spaceborne SAR imagery. *Int. J. Remote Sensing*, 11:209–223.

Lada, C.J., DePoy, D.L., Merrill, K.M., and Gatley, I. 1991. Infrared images of M17. *The Astronomical Journal*, 374:533–539.

Larson, G.W., Rushmeier, H., and Piatko, C. 1997. A visibility matching tone reproduction operator for high dynamic range scenes. *IEEE Trans. on Visualization and Computer Graphics*, 3:291–306.

Mann, S. 1996. 'Pencigraphy' with AGC: Joint parameter estimation in both domain and range of functions in same orbit of the projective-Wyckoff group. In *Proc. Int. Conf. Imag. Process.*, Lausanne, Switzerland, vol. 3, pp. 193–196.

Mann, S. and Picard, R.W. 1995. On being 'Undigital' with digital cameras: Extending dynamic range by combining differently exposed pictures. In *IS&T 48th Annual Conference*, Washington, DC, pp. 422–428.

Marshall, J.A., Burbeck, C.A., Ariely, D., Rolland, J.P., and Martin, K.E. 1996. Occlusion edge blur: A cue to relative visual depth. *J. Opt. Soc. Amer. A*, 13:681–688.

Mitsunaga, T. and Nayar, S.K. 1999. Radiometric self calibration. In *Proc. Comp. Vis. Patt. Rec*, Fort Collins, CO, vol. I, pp. 374–380.

Nayar, S.K. and Mitsunaga, T. 2000. High dynamic range imaging: Spatially varying pixel exposures. In *Proc. Comp. Vis. Patt. Rec.*, Hilton Head Island, SC, vol. I, pp. 472–479.

Negahdaripour, S., Xu, X., Khemene, A., and Awan, Z. 1998. 3-D motion and depth estimation from sea-floor images for mosaic-based station-keeping and navigation of ROV's/AUV's and high-resolution sea-floor mapping. In *Proc. IEEE Workshop on Autonomous Underwater Vehicles*, Cambridge, MA, pp. 191–200.

Ogiers, W. 1997. Survey of CMOS imagers. In *IMEC report* P60280-MS-RP-002, Issue 1.1, part 1.

Pardo, A. and Sapiro, G. 2002. Visualization of high dynamic range images. In *Proc. IEEE Int. Conf. Image Processing*, vol. 1, pp. 633–636.

Peleg, S., Ben-Ezra, M., and Pritch, Y. 2001. Omnistereo: Panoramic stereo imaging. *IEEE Trans. Patt. Analys. Machine Intell.*, 23:279–290.

Rajagopalan, A.N. and Chaudhuri, S. 1995. A block shift-invariant blur model from defocused images. In *Proc. Int. Conf. Imag. Process.*, Washington DC, vol. 3, pp. 636–639.

Reynoso, E.M., Dubner, G.M., Goss, W.M., and Arnal, E.M. 1995. VLA observations of neutral hydrogen in the direction of Puppis A. *The Astronomical Journal*, 110:318–324.

Sawhney, H.S., Kumar, R., Gendel, G., Bergen, J., Dixon, D., and Paragano, V. 1998. VideoBrush^{TM}: Experiences with consumer video mosaicing. In *Proc. IEEE Workshop on Applications of Computer Vision*, Princeton, NJ, pp. 56–62.

Schechner, Y.Y. and Nayar, S.K. 2001. Generalized mosaicing. In *Proc. IEEE Int. Conf. on Computer Vision*, Vancouver, Canada, vol. I, pp. 17–24.

Schechner, Y.Y. and Nayar, S.K. 2002. Generalized mosaicing: Wide field of view multispectral imaging. *IEEE Trans. Patt. Analys. Machine Intell.*, 24:1334–1348.

Schwartz, J. September 1999. CMOS camera won't be blinded by the light. *Photonics Spectra*.

Sharma, R.K. and Pavel, M. 1997. Multisensor image registration. In *Society for Information Display*, vol. XXVIII, pp. 951–954.

Shum, H.Y. and Szeliski, R. 2000. Systems and experiment paper: Construction of panoramic image mosaics with global and local alignment. *Int. J. of Computer Vision*, 36:101–130.

Smolić, A. and Wiegand, T. 2001. High-resolution image mosaicing. In *Proc. IEEE Int. Conf. Imag. Process.* Thessaloniki, Greece, vol. 3, pp. 872–875.

Socolinsky, D.A. 2000. Dynamic range constraints in image fusion and realization. In *Proc. IASTED Int. Conf. Signal and Image Process.*, Las Vegas, NV, pp. 349–354.

Soderblom, L.A., Edwards, K., Eliason, E.M., Sanchez, E.M., and Charette, M.P. 1978. Global color variations on the Martian surface. *Icarus*, 34:446–464.

Surya, G. and Subbarao, M. 1993. Depth from defocus by changing camera aperture: A spatial domain approach. In *Proc. Computer Vis. and Patt. Rec.*, New York, pp. 61–67.

Tabiryan, N. and Nersisyan, S. 2002. Liquid-crystal film eclipses the sun artificially. *Laser Focus World*, 38(5):105–108.

Thevenaz, P. and Unser, M. 2000. Optimization of mutual information for multiresolution image registration. *IEEE Trans. Imag. Process.*, 9:2083–2099.

Tomasi, C. and Kanade, T. 1992. Shape and motion from image streams under orthography: A factorization method. *Int. J. of Comp. Vision*, 9:137–154.

Uson, J.M., Boughn, S.P., and Kuhn, J.R. 1990. The central galaxy in Abell 2029: An old supergiant. *Science*, 250:539–540.

Vasavada, A.R., Ingersoll, A.P., Banfield, D., Bell, M., Gierasch, P.J., and Belton, M.J.S. 1998. Galileo imaging of Jupiter's atmosphere: The great red spot, equatorial region, and white ovals. *Icarus*, 135:265–275.

Viola, P. and Wells III, W.M. 1997. Alignment by maximization of mutual information. *Int. J. of Computer Vision*, 24:137–154.