Interleaved Imaging: An Imaging System Design Inspired by Rod-Cone Vision

Manu Parmar[a] and Brian A. Wandell[b]

[a]Electrical Engineering Department, Stanford University, Stanford, CA, USA 94305;
[b]Psychology Department, Stanford University, Stanford, CA, USA 94305.

ABSTRACT

Under low illumination conditions, such as moonlight, there simply are not enough photons present to create a high quality color image with integration times that avoid camera-shake. Consequently, conventional imagers are designed for daylight conditions and modeled on human cone vision. Here, we propose a novel sensor design that parallels the human retina and extends sensor performance to span daylight and moonlight conditions. Specifically, we describe an interleaved imaging architecture comprising two collections of pixels. One set of pixels is monochromatic and high sensitivity; a second, interleaved set of pixels is trichromatic and lower sensitivity. The sensor implementation requires new image processing techniques that allow for graceful transitions between different operating conditions. We describe these techniques and simulate the performance of this sensor under a range of conditions. We show that the proposed system is capable of producing high quality images spanning photopic, mesopic and near scotopic conditions.

Keywords: Image sensor, color filter array, low-light imaging

1. INTRODUCTION

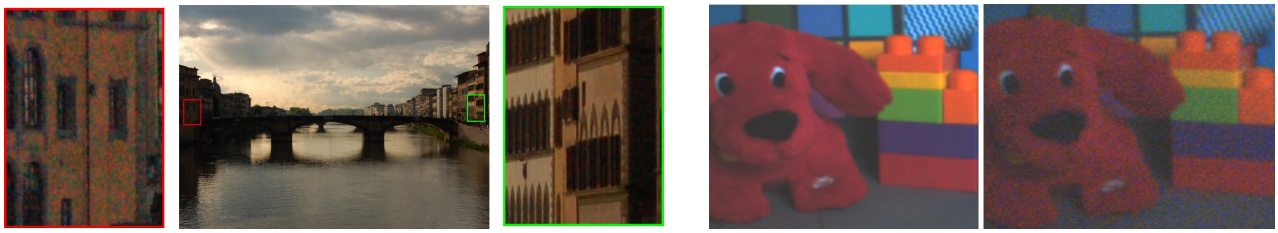
Designers and consumers would like cameras to operate under the same range of conditions as the human visual system. From day to night, scene intensities span an intensity range of roughly 10^8 units and the human visual system adapts to encode images effectively across this enormous range. Multiple mechanisms play a role in adapting to light-levels;[1] these include regulation of the pupil and scaling of the response gain of individual receptors. In addition, the system uses two distinct types of photoreceptors, rods and cones, to encode the broad range of intensity levels. The three types of cones are the principal encoding cells under relatively high levels of illumination (photopic). Under these conditions spatial and temporal resolution are highest, and we experience color. Vision is mediated by rods under the lowest levels of illumination (scotopic). Rod vision has significantly lower spatial and temporal resolution, and since there is only one type of rod, scotopic vision is achromatic. Over the intermediate (mesopic) range, signals encoded by rods and cones both contribute to vision. Experiments show that rod and cone signals interact, with both receptor types influencing both color appearance and sensitivity at mesopic levels.[2,3]

The need to span such a large range intensity range imposes a substantial challenge because there are very few photons available at low levels. For example, under moonlight conditions an $f/5.6$ lens gathers about 10 photons per square micron per second. Hence, in an exposure of 25 ms a 2 μm pixel (100% fill factor, 100% quantum efficiency) will receive on average only one incident photon. Xiao et al.[4] showed that photon noise on a uniform background becomes visible at an SNR $\lesssim$ 30 dB (1000 photons). Even under moderate imaging intensities noise frequently becomes visible. For example, Figure 1(a) was acquired under low photopic levels, but shadows reduce the illumination in a significant portion of the image to scotopic levels. The noise in the shadowed region is apparent and much greater than the noise in the well-lit portion.

Figure 1(b) illustrates a second reason for low illumination: the drive to increase spatial resolution by reducing pixel-size. The image on the left simulates a scene with mean illumination 100 cd/m^2, and $f/4$ lens acquired with an image sensor with 6 μm pixels (the simulation was carried out with the ISET imaging pipeline simulator[5]). The image on the right simulates an identical acquisition (similar exposure and noise characteristics) with a

Send correspondence to Manu Parmar. Email: MP (mparmar@stanford.edu), BW (wandell@stanford.edu)

Digital Photography V, edited by Brian G. Rodricks, Sabine E. Süsstrunk,
Proc. of SPIE-IS&T Electronic Imaging, SPIE Vol. 7250, 725008 · © 2009 SPIE-IS&T
CCC code: 0277-786X/09/$18 · doi: 10.1117/12.806367



(a) (b)

Figure 1. **Image sensor noise from low pixel illumination.** (a) An image acquired with a digital camera under low photopic illumination. Portions of the image from a low-light area (left, red border) and a well-lit area (right, green border) are compared. (b) Simulations of the same scene acquired using a sensor with 6 μm (left) and 2 μm (right) pixels.

sensor with 2 μm pixels. The sensor with smaller pixels produces an image with significantly higher noise. The deterioration of signal to noise ratio (SNR) with reduced pixel-size is related to the smaller number of incident photons and the Poisson nature of photon arrival; the SNR declines with the square root of the signal level. In the case illustrated in Fig. 1(b), as the pixel area decreases from 6 μm to 2μm, the photon-dependent SNR decreases by a factor of 3 (10 dB).

The most practical opportunities for increasing SNR have been (a) reducing pixel noise, (b) increasing pixel spectral quantum efficiency. There are very few other opportunities to increase the amount of light incident on the sensor and improve SNR. Lens aperture-widths are often limited by the device form factor. Increasing exposure durations is impractical for video applications because exposure durations are limited by the frame-rate. In still cameras, exposure duration is limited by motion blur and camera shake[6] which is exacerbated by small form-factor.

The modern sensor roadmap, driven largely by pixel-size reduction, is not helpful in accounting for the physical limitations at low light levels; there is a need for additional approaches to sensor design. In this paper we describe and analyze an architecture that contains two interleaved sensor mosaics. One mosaic is optimized to capture color at relatively high light levels. A second mosaic foregoes color information and is optimized to capture images under low-light levels. Because many scenes contain some regions that are adequately illuminated and others that are poorly illuminated, the sensor design requires the development of an image processing framework that can gracefully combine information from these interleaved mosaics.

2. BACKGROUND

The problems caused by the limited number of photons in low-light conditions have been addressed in a number of patents and publications. Several of these rely on methods that increase the number of photons gathered by a color-sensing pixel by using filters with higher transmission efficiencies: say by increasing color filter peak efficiency and by increasing the filter spectral bandwidth. Here we describe salient features of a few such methods most similar in spirit to our technique.

In a U.S. patent granted in 1983, Sato et al.[7] described an imaging system with a CFA with transparent photosites. The transparent pixel serves in lieu of luminance in their application; they do not describe a method for accounting for the sensitivity mis-match between the transparent and color channels.

In 1994 Yamagami et al.[8] were granted a patent for a system design that uses a CFA comprising RGB and luminance-sensitive (denoted by the letter Y) photosites. This RGBY system generates a luminance channel (Y) and two color-difference channels that are derived from the RGB photosites. Yamagami et al. acknowledge the large sensitivity mis-match between RGB and Y pixels. Specifically, the Y channel has significantly higher sensitivity than the other channels and it will saturate when the color channels are well-exposed. They discuss approaches to minimize this problem by (a) using CMY filters rather than RGB, or (b) placing a neutral density filter on the Y channel. In 2002, Gindele and Gallagher[9] were granted a patent that addresses the sensitivity mismatch. They propose a scheme for recovering RGB data from Yamagami's RGBY CFA data in bright conditions when Y photosites are saturated. Gindele's method extends Adams' demosaicking method.[10]

Kijima et al.[11] applied for a patent on image sensors with three color (e.g., RGB) and a fourth wideband photosite that they call *panchromatic* (see also Luo[12]). The signal processing architecture combines panchromatic and color channels at an early stage, and thus faces the sensitivity mis-match problem. Kijima et al. propose that because "the color filter pixels will be significantly less sensitive than the panchromatic pixels" it is "advantageous to adjust the sensitivity of the color filter pixels so that they have roughly the same sensitivity as the panchromatic pixels. (Kijima et al.,[11] page 5, column 2, par 57)."

There is a significant lost opportunity in all of these methods. Combining the wideband and RGB color signals at an early stage limits the ability to create a high-dynamic range sensor. The proposed solutions - using CMY filters, adding a neutral density to the Y channel - reduce the effective sensor dynamic range. Thus, they fail to exploit fully the wideband channel.

In contrast to these approaches, the interleaved imaging system proposed here has a wideband channel with peak quantum efficiency that can be very high relative to RGB channels. The wideband channel provides high SNR and spatial detail information in low light conditions. When light levels increase, the RGB signal has sufficient SNR, and the system smoothly reduces its use of the wideband channel. The interleaved imaging design parallels the biological rod-cone design and frees us to maintain the full dynamic range of the two types of channels.

3. INTERLEAVED IMAGING

We simulate an image sensor CFA with two kinds of photosites: RGB photosites and and wideband photosites (W) specialized for low-light sensing (Figure 2). The wideband channel maximizes the number of photons gathered and has approximately 6x the SNR of the G channel and 10x higher than the R and B channels. There are many possible spatial arrangements of these photosites; we illustrate one. A single acquisition provides two interleaved images: the RGB image encodes color information while the wideband image encodes an achromatic high SNR representation. The key problem is to find a method of combining the two images without compromising the advantages inherent to the two interleaved images.

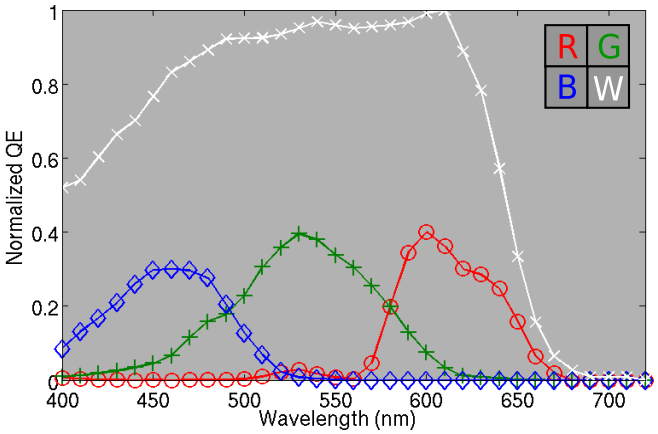


Figure 2. **Color filter array.** The curves show the transmittances of the RGBW filters used in the simulation. The legend shows the spatial arrangement of these filters.

Figure 3 illustrates the general operation of the imaging pipeline. The RGBW channels are decomposed into two images. The images at the center of the pipeline in Fig. 3 illustrate conditions where the interleaved imaging system is most effective (low-light). The RGB photosites gather only a limited number of photons and provide a noisy but colored image. This is the image that would be acquired by a conventional imaging system in such conditions. The W channel output is shown at center-bottom. This achromatic image has high SNR and carries reliable spatial information. The interleaved image processing combines the two images to yield a single final output, shown at the right.

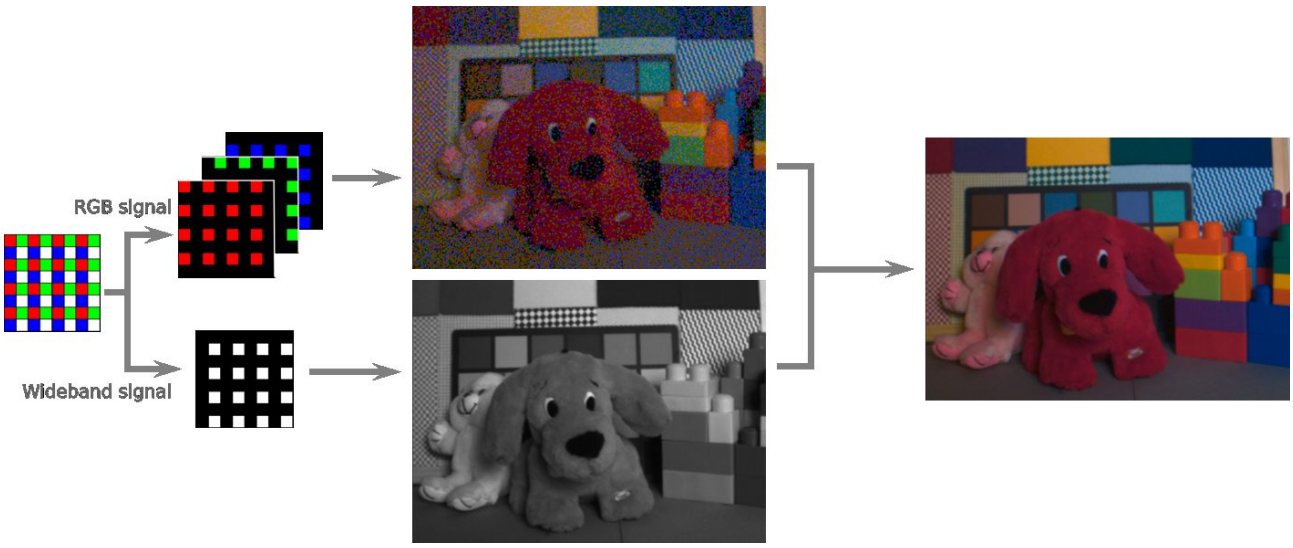


Figure 3. **Interleaved imaging.** The data from the RGB and W photosites are decomposed into a color (top) and wideband (bottom) image. The color image is noisy and comparable to the quality that one would obtain with a conventional camera. The interleaved image processing system combines the color and wideband images to improve the final output, which is a single high SNR color image shown at the right.

3.1 Interleaved imaging signal processing

Interleaved image processing confronts many of the same issues as conventional digital image processing, such as demosaicking and white balancing. In addition, interleaved imaging has one further challenge: managing the large sensitivity mis-match between the wideband and color images. In this section we describe our approach to this problem.

Interleaved imaging operates in several different regimes that parallel the scotopic, mesopic, and photopic ranges. Under very low-light the SNR in the RGB channels is so low that little useful information can be derived. In such conditions spatial information is derived completely from the W channel. In mesopic conditions, there is useful information in both the W image and the RGB image. In this regime it is useful to combine the information from the W and RGB images. Finally, in photopic conditions the W channel is saturated and information is derived from the RGB image. Below we describe how we manage the transition between these domains.

The interleaved imaging system acquires a 2D CFA ($m \times n$) image that measures one intensity level at each spatial location. This image can be expanded to an $m \times n \times 4$ array with zeros in locations that are not sampled by the CFA. The first three bands ($m \times n \times 1, 2, 3$) of this array contain the RGB image. This demosaicking process is analogous to conventional demosaicking, but it has the additional problem of the sensitivity mis-match between the channels. Hence, demosaicking the CFA data requires special considerations. We use adaptive smoothing based on bilateral filtering and non-local means to recover RGB from the CFA data.

We denote each band of the interleaved image and measurements by f_i and g_i, $i = \mathrm{R}, \mathrm{G}, \mathrm{B}, \mathrm{W}$, respectively. The estimated value of a pixel $g_i(x)$ is found as a weighted sum of the pixels in its neighborhood. The weight associated with each pixel in the neighborhood of x depends on two factors: a) its distance from x, and b) its similarity with respect to $g_i(x)$. In the examples below, we used a neighborhood size of 21×21 pixels.

3.1.1 Bilateral filter structure

Tomasi and Manduchi introduced the term *bilateral filter*[13] to describe the idea of selecting filter weights based on geometric and photometric similarities. In the original implementations, photometric similarity was based entirely on pixel intensity. The bilateral filter adapts to local image content and is able to perform smoothing while preserving edges. Buades et al.[14] in the *non-local means filter* use an inter-pixel similarity measure based on the similarity of image patches surrounding x and the neighborhood pixel. This inter-pixel similarity measure captures the closeness of image features and is successful at preserving textures.

The bilateral filter is applied for demosaicking by updating each pixel as:

$$f_i(x) = \frac{1}{W_y} \sum_{y \in \Omega} G(\beta_d, \sigma_d) G(\beta_s, \sigma_s) S_i^{\Omega} g_i(y), \tag{1}$$

where S_i^{Ω} is a mask that has ones at locations where the CFA samples the i^{th} channel and zeros everywhere else; $G(\beta, \sigma)$ is the 2D Gaussian kernel

$$G(\beta, \sigma) = \frac{1}{2\pi\sigma^2} \exp\left(-\frac{\beta^2}{2\sigma^2}\right) \tag{2}$$

and W_y is a normalization factor

$$W_y = \sum_{y \in \Omega} G(\beta_d, \sigma_d) G(\beta_s, \sigma_s) S_i^{\Omega}, \tag{3}$$

and Ω is the neighborhood.

In the next section we define the Gaussians for the distance-weight, $G(\beta_d, \sigma_d)$, and the pixel-similarity weight, $G(\beta_s, \sigma_s)$.

3.1.2 Similarity functions

The distance similarity measures the pixel distance between two points:

$$\beta_d = \|x - y\|_2. \tag{4}$$

We illustrate the pixel-similarity measure with an example in Fig. 4. Consider the pixel at location x. It has a neighborhood, Ω, indicated by the white circle and an associated image patch, $h(x)$, shown on the right. Two other pixels in the neighborhood, y and z and their image patches are also marked. The pixel-similarity between x and y is measured by comparing their image patches, $h(x)$ and $h(y)$. In this example $h(x)$ is more similar to $h(y)$ than to $h(z)$. The similarity is greater because the edge orientation around x and y are similar, but the edge around z is not similar.

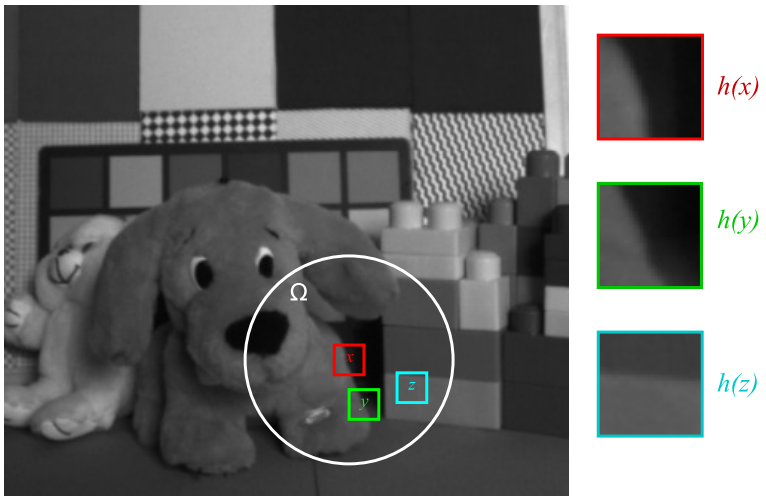


Figure 4. **Pixel-similarity.** Each pixel is associated with a small surrounding patch. The pixel-similarity between two pixels is determined by the similarity of their associated patches. The image patches for three pixels are shown. The pixel-similarity between x and y is higher than the pixel-similarity between x and z.

Under low and moderate luminance levels, the W channel has more reliable spatial information than the RGB channels. As the illumination increases the W channel saturates. Hence, under low illumination we prefer

to judge the image patch similarity based on the W channel and under high illumination we prefer to use the RGB channels. To transition gracefully between these regimes, we define the pixel-similarity as a weighted sum of the similarity in the W channel and RGB channels:

$$\beta_s = \alpha(y)\|h_i(x) - h_i(y)\|_2 + (1 - \alpha(y))\|h_W(x) - h_W(y)\|_2, \tag{5}$$

The weight, α, is determined by the fraction of saturated pixels in the image patch near y.

$$\alpha(y) = \frac{N_{\text{saturated}}(h_W(y))}{N_{\text{total}}(h_W(y))}. \tag{6}$$

The adaptive pixel-similarity computation is illustrated in Fig. 5. Panel (a) is the R channel image for a simulated acquisition with an interleaved sensor. Panel (b) is the corresponding W channel image. Note that several areas in the W channel image are saturated and lose all spatial detail. In such conditions, the pixel-similarity measure relies mainly on R channel data. In dark regions of the image, where R channel spatial detail is unreliable and the W channel is unsaturated, the pixel-similarity measure relies mainly on W channel information data. Panel (c) shows the resulting RGB image.

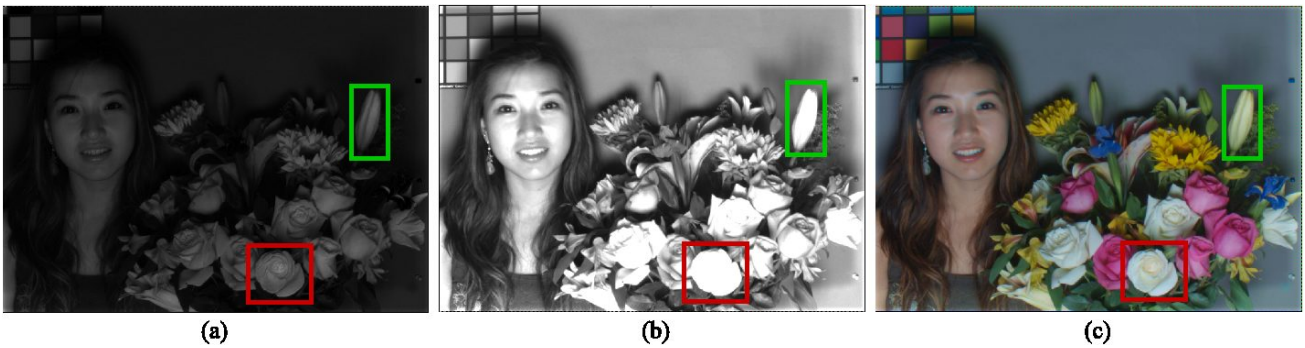


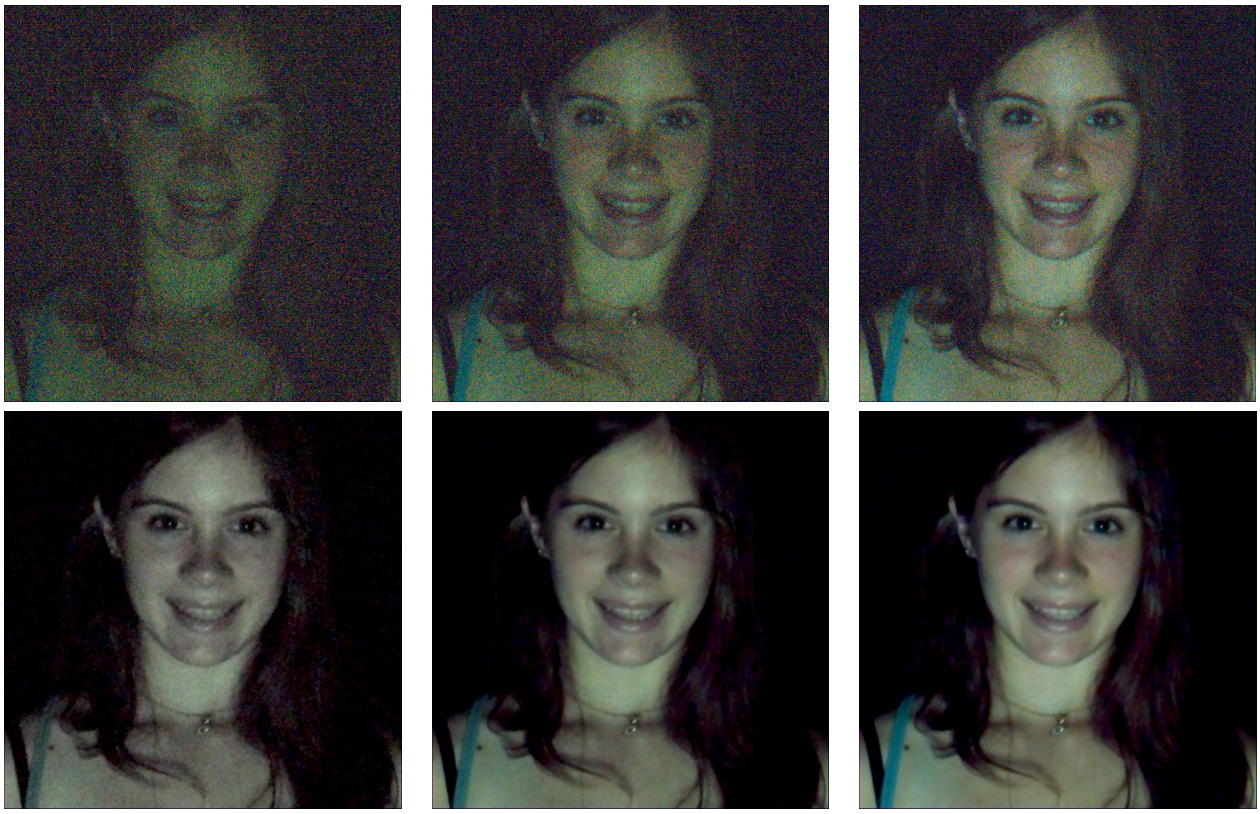
Figure 5. **Channel sensitivity mis-match.** Simulated acquisition of a scene with an interleaved image sensor (mean illumination 200 cd m^2, 66 ms exposure time,3 μm pixels). Under these conditions the RGB channels are adequately exposed in some regions but not others. The W channel is saturated in some regions. The images show the (a) R channel (b) W channel and (c) The output image after applying interleaved image processing to the data.

3.1.3 Luminance substitution mode

The adaptive methods based on bilateral filtering and non-local means produce a complete RGB image. Under typical or even fairly dark conditions this image is the final output of the system. When the illumination is extremely low, however, this image may still be quite noisy. There is one additional mode that we can call upon to attempt to rescue images under these very low conditions. Specifically, we can use the W image as a substitute for the luminance component of the RGB image. This substitution will alter the chromatic appearance of the image slightly, but the additional SNR in the W image compared to the luminance component of the RGB image makes the substitution worthwhile. The decision to make this substitution is based on the SNR of the RGB image.

4. EXPERIMENTS

We used the ISET Digital Camera Simulator[5] to simulate the the proposed interleaved imaging system and compared its performance with a conventional imaging system with a similar size sensor. ISET is a software package that offers a system approach to modeling and simulating the image processing pipeline of a digital camera. The ISET simulation begins with scene data (a physical description of radiance); these are transformed by the imaging optics into the optical image, an irradiance distribution at the image sensor array. The irradiance is transformed into an image sensor array response and finally, the image sensor array data are processed to



(a) 5 ms (b) 10 ms (c) 20 ms

Figure 6. **Interleaved imaging compared with conventional imaging.** Top row: simulations of an image acquired using a conventional Bayer CFA of a scene with average illumination 25 cd/m^2. Each image was acquired at a different exposure duration (5, 10, 20 ms). Bottom row: simulations of images reconstructed with the proposed interleaved imaging system at the same exposure durations. The W channel replaced the RGB luminance information (luminance substitution mode) in the 5ms image, but not in the other images. Simulation parameters are listed in Table 1.

generate a display image. ISET includes the ability to simulate a variety of visual scenes, imaging optics, sensor electronics and image processing pipelines.

We show results for a simulated scene in Fig. 6. The average illuminant level for this scene was set at 25 cd/m^2. The values of sensor parameters, such as, read noise, dark voltage, photoresponse non-uniformity, etc. were approximated from values obtained for similar sensors from experiments and specifcation sheets and are listed in Table 1. The image processing pipeline for the conventional imager used for comparison is based on bilinear demosaicking and Gray World color balancing. The image processing pipeline for the interleaved imager relies on the signal processing steps described in Section 3.1 and subsequent color balancing.

The top row in Fig. 6 shows the results obtained for the Bayer sensor for various acquisition times. The bottom row shows corresponding results for the interleaved sensor. In the 5 ms interleaved acquisition, the W channel was used to replace luminance information as described in Section 3.1.3.

5. CONCLUSIONS

We propose and simulate an interleaved imaging system. The system is designed to expand the effective operating range of the imaging sensor. The system is based on capturing two images that parallel the rod (scotopic) and cone (photopic) photoreceptors in the retina. The imaging pixels can be operated in very different intensity ranges, and the interleaved image processing smoothly combines the spatial and chromatic information captured by each of the images.

Table 1. Sensor parameter values used for simulations.

Sensor parameter	
Pixel width (μm)	2.2
Pixel height (μm)	2.2
Fill factor	0.9
Dark voltage (V)	0.0
Read noise (mV)	4.58
Dark signal nonuniformity (DSNU) (mV)	6
Photoreceptor nonuniformity (PRNU) (%)	1.7
Conversion Gain (μV/e)	30
Voltage swing (V)	1.08
Analog gain	7.98
Mean scene luminance (cd/m^2)	25
Lens f-number	$f/2.8$
Well capacity (electrons)	15000

ACKNOWLEDGMENTS

We thank Dr. Joyce Farrell for help with simulations and Dr. Peter Catrysse and Steve Lansel for helpful advice. We thank Dr. Susumu Kikuchi and Dr. Takashi Kondoh and the Olympus Corp. Tokyo, Japan for their support.

REFERENCES

[1] B. A. Wandell, *Foundations of Vision*, Sinauer Associates, Inc., 1995.

[2] A. Stockman and L. T. Sharpe, "Into the twilight zone: the complexities of mesopic vision and luminous efficiency," *Ophthalmic and Physiological Optics* **26**(3), pp. 225–239, 2006.

[3] R. Knight and S. L. Buck, "Rod influences on hue perception: Effect of background light level," *Color Research & Application* **26**(S1), pp. S60–S64, 2001.

[4] F. Xiao, J. E. Farrell, and B. A. Wandell, "Psychophysical thresholds and digital camera sensitivity: the thousand-photon limit," *Digital Photography* **5678**(1), pp. 75–84, SPIE, 2005.

[5] J. E. Farrell, F. Xiao, P. B. Catrysse, and B. A. Wandell, "A simulation tool for evaluating digital camera image quality," in *Image Quality and System Performance, Proceedings of the SPIE*, **5294**, pp. 124–131, Jan. 2004.

[6] F. Xiao, A. Silverstein, and J. Farrell, "Camera-motion and effective spatial resolution," in *Proc. International Congress on Imaging Science*, pp. 33–36, 2006.

[7] I. Sato, K. Ooi, K. Saito, Y. Takemura, and T. Shinohara, "Color image pick-up apparatus," U.S. Patent 4390895 , 1982.

[8] T. Yamagami, T. Sasaki, and A. Suga, "Image signal processing apparatus having a color filter with offset luminance filter elements," U.S. Patent 5323233 , June 1994.

[9] E. Gindele and A. Gallagher, "Sparsely sampled image sensing device with color and luminance photosites," U.S. Patent 6476865 B1 , Nov. 2002.

[10] J. E. Adams, Jr. and J. F. Hamilton Jr., "Adaptive color plane interpolation in single sensor color electronic camera," U.S. Patent 5652621 , 1997.

[11] T. Kijima, H. Nakamura, J. T. Compton, J. F. Hamilton, and T. E. DeWeese, "Image sensor with improved light sensitivity," U.S. Patent Application No. 20070268533 , 2007.

[12] G. Luo, "A novel color filter array with 75% transparent elements," in *Proceedings of the SPIE*, **6502**, p. 65050T, 2007.

[13] C. Tomasi and R. Manduchi, "Bilateral filtering for gray and color images," in *Proc. Sixth International Conference on Computer Vision*, pp. 839–846, 4–7 Jan. 1998.

[14] A. Buades, B. Coll, and J.-M. Morel, "A non-local algorithm for image denoising," in *Proceedings of the 2005 IEEE Computer Society Conference on Computer Vision and Pattern Recognition*, **2**, pp. 60–65, 2005.