Applications of Multi-Bucket Sensors to Computational Photography

Gordon Wan* Mark Horowitz† Marc Levoy‡

Stanford University

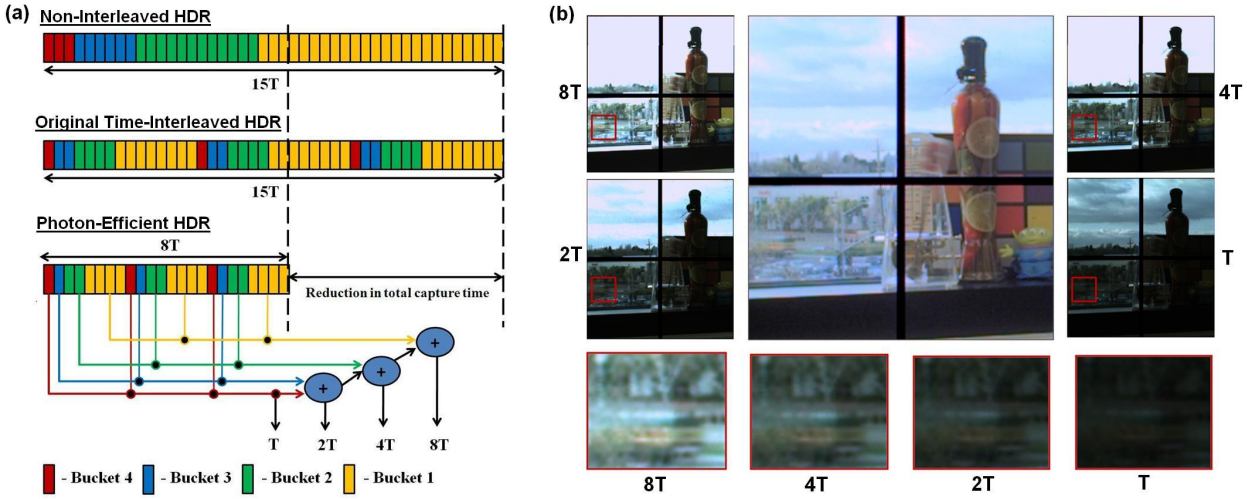

Figure 1: *HDR photography using 4 buckets per pixel. (a) Each of the three colored bars depicts a single exposure. For each of the time slices within the exposure, color denotes which bucket the electrons are stored in at the conclusion of that slice. In non-interleaved HDR (top bar), four images are captured sequentially. In the original time-interleaved HDR (middle bar), four images are captured in a time-interleaved manner. For these two protocols, the four images are read out directly from the buckets at the conclusion of the exposure. In contrast, in photon-efficient HDR (bottom bar), four non-destructive readouts are performed at the conclusion of the exposure, of bucket 4 alone, bucket 4 + bucket 3, and so on, thereby producing images with exposures times, T, 2T, 4T, and 8T. These additions are performed at all pixels in parallel in the analog domain. These images can be combined digitally off-chip to produce exposure times in each pixel that range from T to 8T. The use of non-destructive readout and analog addition allows us to achieve a total capture time of only 8T, by contrast with the first two protocols based on a sequence of exposures of length T, 2T, 4T, and 8T. In the latter case, total capture time is 15T, so motion blur is worse. This is one advantage of our approach. (b) Still life with moving metronome (at center). The images labeled T, 2T, 4T, and 8T are the four images, with crops shown at bottom. At center is the synthesized HDR photograph. The four windows separated by black lines in the images correspond to pixels with slightly different designs. Since capture of the four images are finely interleaved in time, there are no motion differences between them, and no alignment is necessary before HDR synthesis. This is a second advantage of our approach, which can be extended to the capture of HDR video. [Please watch the video].*

Abstract

Many computational photography techniques take the form, "Capture a burst of images varying camera setting X (exposure, gain, focus, lighting), then align and combine them to produce a single photograph exhibiting better Y (dynamic range, signal-to-noise, depth of field). Unfortunately, these techniques may fail on moving scenes because the images are captured sequentially, so objects are in different positions in each image, and robust local alignment is difficult to achieve. To overcome this limitation, we propose using multi-bucket sensors, which allow the images to be captured in time-slice-interleaved fashion. This interleaving produces images with nearly identical positions for moving objects, making alignment unnecessary. To test our proposal, we have designed and fabricated a 4-bucket, VGA-resolution CMOS image sensor, and we have applied it to high dynamic range (HDR) photography. Our sensor permits 4 different exposures to be captured at once with no motion difference between the exposures. Also, since our protocol employs non-destructive analog addition of time slices, it requires less total capture time than capturing a burst of images, thereby reducing total motion blur. Finally, we apply our multi-bucket sensor to several other computational photography applications, including flash/no-flash, multi-flash, and flash matting.

Keywords: Computational Photography, Multi-Bucket Sensors, Time-Multiplexed Exposure, High Dynamic Range Photography, Flash/No-Flash, Multi-Flash, Flash Matting

1 Introduction

In multi-image computational photography, a burst of images exposed under different camera settings are captured. In a single-camera system that uses a conventional image sensor, the images are captured sequentially, then combined to create a final image which is superior in some aspects to any of the component images. Representative examples include multiple exposure high dynamic range (HDR) [Debevec 97][Reinhard 06], flash/no-flash [Eisemann 04] [Petschnigg 04], multi-flash [Raskar 04], color photography using active illumination [Ohta 07], and flash matting [Sun 06].

While the above approach works nicely in a static scene, it is challenging to use in a dynamic scene, because differences can occur

*e-mail: cwan@stanford.edu
†e-mail:horowitz@stanford.edu
‡e-mail:levoy@cs.stanford.edu

between the images. For example, a moving object may appear at different positions, or the captured images may have different amount of handshake blur. Being unpredictable, these differences may cause the subsequent reconstruction algorithms to fail, producing artifacts in the final computed image. Figure 2 shows ghosting due to object motion in a multiple exposure HDR photograph.

Figure 2: *Ghosting artifact due to motion in multiple exposure HDR photography. This HDR photo was taken by an iPhone 4 using an image sensor running at a maximum of 15fps. Two frames, one long and another short, were taken by the phone to synthesize the photo. A time gap of roughly 1/15s exists between the two frames due to the limited frame rate of the sensor, giving rise to the observed motion.*

Many algorithms have been devised to avoid the artifacts described above [Kang 03] [Ward 03] [Eden 06] [Gallo 09] [Mills 09]. In particular, image alignment or motion compensation is usually performed prior to blending the images. However, the effectiveness of these algorithms is scene and sensor dependent and will not work all the time [Szeliski 10]. For example, when a scene has little texture or contains a region that is out of focus or blurred by handshake or object motion, then image alignment can fail. If a significant portion of the scene undergoes a non-rigid motion, then alignment can also fail. Finally, if the images have widely different exposures such as in flash/no-flash, the task of aligning them becomes very challenging [Eisemann 04] [Petschnigg 04]. As a result, most reported multi-image computational photography techniques can only be used in limited situations.

The algorithms cited thus far assume a conventional image sensor. In this paper, we remove this assumption and demonstrate time-multiplexed exposure as an alternative imaging approach. Instead of capturing multiple frames sequentially, time-multiplexed exposure partitions the frames into time slices and interleaves them in a desired way and at high rate (up to kilohertz). This interleaving equalizes unpredictable changes in the scene such as motion among the frames, eliminating the need for error-prone image alignment or motion compensation algorithm. This simplification of reconstruction algorithm eliminates many artifacts that currently plague multi-image computational photography.

To implement time-multiplexed exposure, we have designed and fabricated multi-bucket sensors that contain multiple analog memories per pixel. In such sensors, photo-generated charges in a photodiode can be transferred and accumulated in the in-pixel memories in any chosen time sequence during an exposure. Therefore, intermediate sub-images captured under different camera settings can be transferred and accumulated inside the pixels before readout.

Multi-bucket pixels are not new. For example, pixels with two memories, commonly known as lock-in or demodulation pixels, have been used to detect amplitude modulated light [Yamamoto 06], including time-of-flight 3D imaging [Kawahito 07] [Kim 10] [Stoppa 10], HDR imaging, motion detection [Yasutomi 10], etc. However, there is no discussion in the literature of applying multi-bucket sensors to computational photography.

In this paper, we describe several such applications. Foremost among these is a new high dynamic range (HDR) imaging technique we call photon-efficient HDR photography. A unique feature of this technique is that it employs non-destructive addition in the analog domain, allowing us to use images with shorter exposures to synthesize the next longer exposure. Consequently, as we will see, this technique uses the shortest possible exposure time to acquire multiple time-interleaved images, thereby incurring the minimal amount of motion blur. In particular, it requires strictly less total capture time than frame-sequential burst-mode photography as shown in Figure 1, so there is less total object motion. We also show that multi-bucket sensors can be applied to other multi-image computational photography problems, including flash/no-flash, multi-flash, and flash matting. In these applications as well, we avoid artifacts that would normally be observed when a conventional sensor is used.

2 Time-Multiplexed Exposure

In this section, we first review how multiple images are captured by conventional image sensors. Then, we describe the principle of time-multiplexed exposure, and we analyze the interleaving frequency needed in time-multiplexed capture protocols.

2.1 Limitations of Sequential Image Capture

Conventional image sensors capture images sequentially, so their frame rate determines how fast multiple images can be taken successively. Figure 3 shows an example of three images captured with different exposure times (e.g. in HDR photography) by conventional rolling shutter sensors with different frame rates. Figure 3 (a) illustrates the point that even though exposure times of all frames are shorter than the frame time, the next frame cannot start immediately, due to the limited readout speed of the sensor, which limits its frame rate. For example, taking three images with 1/125s, 1/250s, and 1/500s back-to-back would require an image sensor with a frame rate of 500 frames per second (fps) in order to avoid idle time. This idle time exacerbates inter-frame object motion.

Inter-frame time gaps, together with rolling shutter artifacts, can in principle be reduced by increasing the frame rate of the sensor (Figure 3 (b)). However, despite the fact that frame rates of image sensors have been improving steadily over times, practical constraints such as circuit speed and power consumption limit the maximum achievable frame rate especially for sensors with high resolutions. In fact, even in the ideal situation where a sensor has an infinite frame rate (Figure 3 (c)), the captured frames can still have different amount of motion blur or moving objects appearing at different locations because their exposures start and end at different times. As a result, multi-image computational photography needs to post-process the captured frames (e.g. image alignment or motion compensation) before computing a final image.

2.2 Time-Interleaved Image Capture

In this alternative approach, an image is not constrained to be captured in a contiguous block of time. Instead, each exposure is partitioned into time slices, which are interleaved with those of other exposures. Figure 4 illustrates this concept, again using the example that three images with different exposure times are to be captured. Under time-multiplexed exposure, frames 1, 2, and 3 now

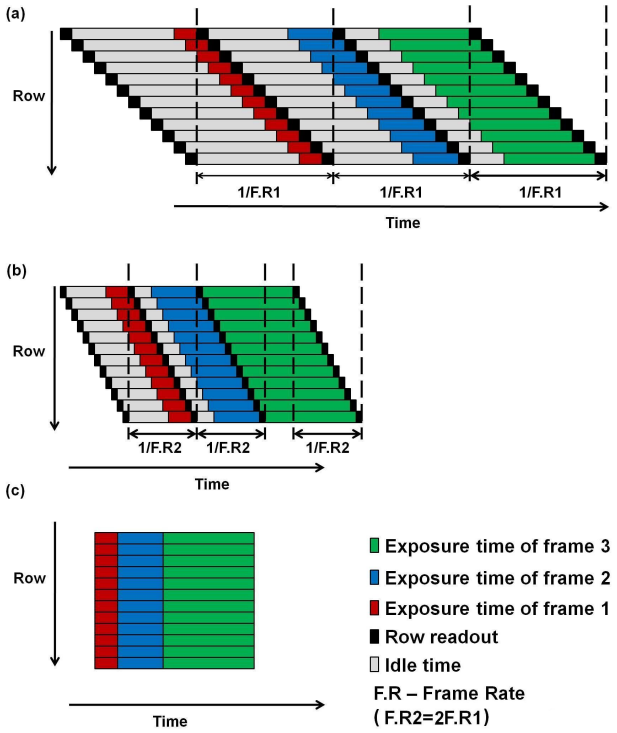

Figure 3: *Capturing three images with different exposures using conventional CMOS rolling shutter sensors with (a) low (b) high (c) infinite frame rates. If the exposure time is significantly shorter than the readout rate (top image), then the requirement that readout of frame N cannot begin until readout of frame N-1 has completed (this constraint is represented by the dashed vertical black lines) leads to a high percentage of idle time (gray boxes). This percentage decreases as frame rate rises, but higher frame rates consume more power, shortening battery life.*

correspond to the sum of the sub-images captured in the red, blue, and green time slots, respectively.

Compared to sequential exposure, time-multiplexed exposure reduces the differences in centroid locations and lengths between the multiple frames in time. Figure 5 (a) shows the case of using a sensor with an infinite frame rate to capture three frames sequentially. The centroid locations of frames 1, 2, and 3 (shown as black dots in the figure) are at T1/2, T1+T2/2, and T1+T2+T3/2, respectively. The difference between the centroid location of frame 1 and frame 2 is (T1+T2)/2 and that between frame 2 and frame 3 is given by (T2+T3)/2. Consider capturing the images using time-multiplexed exposure as shown in Figure 5 (b). Assume each image is partitioned into N time slices and let P_{i1}, P_{i2}, and P_{i3} be the centroid locations of the i^{th} sub-image of frame 1, 2, and 3 respectively. We have P_{i2}-P_{i1} = (T1+T2)/2N and P_{i3}-P_{i2} = (T2+T3)/2N. The difference between the centroid location of frame 1 (C_1) and frame 2 (C_2) is then given by:

$$C_2 - C_1 = (1/N \sum_{i=1}^{N} P_{i2}) - (1/N \sum_{i=1}^{N} P_{i1}) \quad (1)$$

$$= 1/N[\sum_{i=1}^{N} (P_{i2} - P_{i1})] \quad (2)$$

$$= 1/N[N((T1 + T2)/2N)] \quad (3)$$

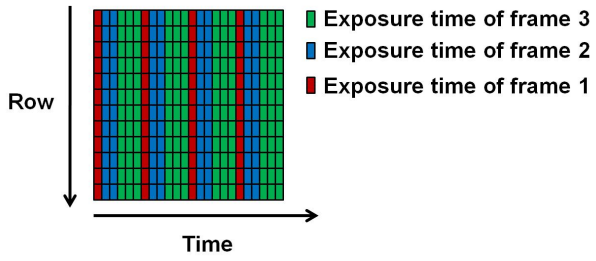

Figure 4: *Illustration of time-multiplexed exposure. Frames 1, 2, and 3 correspond to the sum across rows of the sub-images captured in the red, blue, and green time slots, respectively.*

$$= (T1 + T2)/2N \quad (4)$$

Similarly, the difference between that of frame 2 and frame 3 (C3) is given by:

$$C_3 - C_2 = (T2 + T3)/2N \quad (5)$$

Therefore, the centroid location of the three frames are made N times closer using time-multiplexed exposure. Additionally, the difference between the total duration of frames 1 and 2 becomes (T2-T1)/N and that between frames 2 and 3 becomes (T3-T2)/N, again N times smaller than those using sequential exposure.

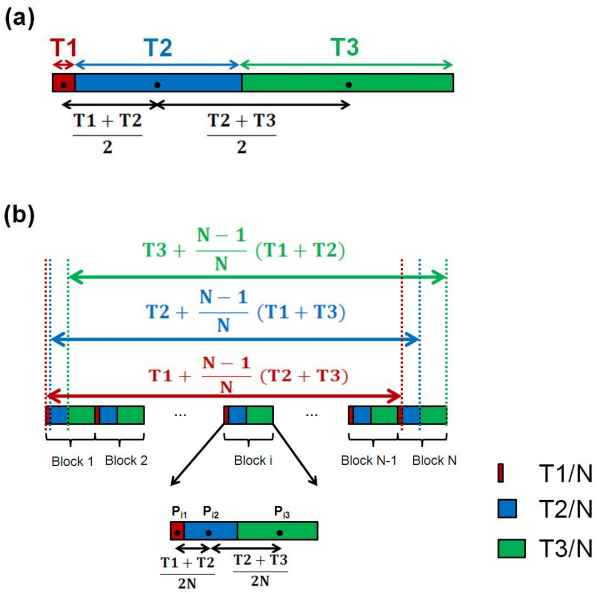

Figure 5: *Time-multiplexed exposure reduces the difference in total duration and centroid location of captured frames in a multi-frame protocol. In (a), three frames are captured back-to-back. This protocol corresponds to sequential capture using a sensor with an infinite frame rate. In (b), time-multiplexed exposure partitions each frame into N equal time slices, and the pieces of different frames are interleaved periodically into N blocks. The black dot inside a block indicates the block's centroid location in time. Compared with (a), the protocol in (b) reduces difference in total duration and centroid location of the captured frames by a factor of N.*

The implication of these results is that by using time-multiplexed exposure, and by increasing N (i.e. the number of time slices),

we can make the multiple frames tightly interleaved and represent virtually the same span of time. As a result, undesired changes in the scene, such as motion, become more evenly distributed between the frames. In particular, all frames captured using this strategy have the same handshake or object motion blur, and moving objects are in the same position. Therefore, this interleaving eliminates the need to align the frames or perform motion compensation after capture.

2.3 Analysis of Interleaving Frequency

Given a total exposure time T_0, it is fairly obvious that we should interleave each exposure condition as frequently as possible. In fact, if N (the number of time slices) is not large enough, ghosting artifacts like those shown in Figure 6 can occur.

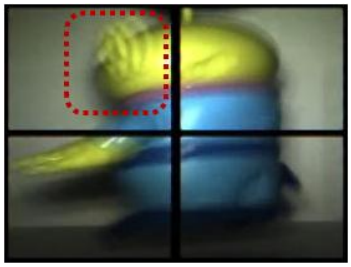

Figure 6: *Potential ghosting artifacts resulting from time-multiplexed exposure when object is moving and the interleaving frequency is too low.*

To prevent these artifacts and mimic natural motion blur, an image must not move more than the pixel pitch p within the interleaving period T. Assume the image is moving linearly with a speed v, the criteria to prevent ghosting is then given by:

$$vT \quad << \quad p \tag{6}$$

In our experiments, we use an interleaving period (T) of 3.2ms and so we can handle up to 300 pixels per second of image motion. In fact, our system, described in the next section, allows a much smaller T (e.g. 200us) but we find the current setting sufficient for most natural scenes.

3 Design and Fabrication of a 4-Bucket Sensor

The key difference between a conventional and multi-bucket sensor is the addition of several memory nodes per pixel. In the course of this project, we have designed and fabricated both 2 and 4-bucket sensors. Figure 7 shows a conceptual view of a multi-bucket pixel. In addition to a photodiode, the pixel has multiple memories to accumulate photo-generated charges, and switches that are programmable by the user so that we can control which light goes into which bucket. In particular, photo-generated charges in the photodiode can be transferred and accumulated in the buckets in any chosen time sequence during an exposure. Since the buckets are in close proximity to the photodiode and no signal processing is involved, sub-images can be transferred and accumulated in these buckets rapidly. Therefore, this architecture can achieve high interleaving frequency. To reduce the number of control signal lines, our implementation switches all pixels together, i.e. charges in photodiodes are transferred to the same bucket at once for all pixels. Fortunately, this design decision does not limit the applications that we will present in Sections 4 and 5.

Similar to most CMOS imager pixels, our sensor includes an additional output bucket called a floating diffusion [Nakamura 06] inside the multi-bucket pixel. Charges accumulated in the other buckets are transferred to this output bucket, converted into voltages, and read out. There are also programmable switches between the accumulation buckets and this output bucket, so that the image stored in each bucket can be read out separately. Since this readout is non-destructive, the output bucket can be used to compute sums of buckets. We will employ this ability to implement the novel "photon-efficient" HDR protocol described in the next section.

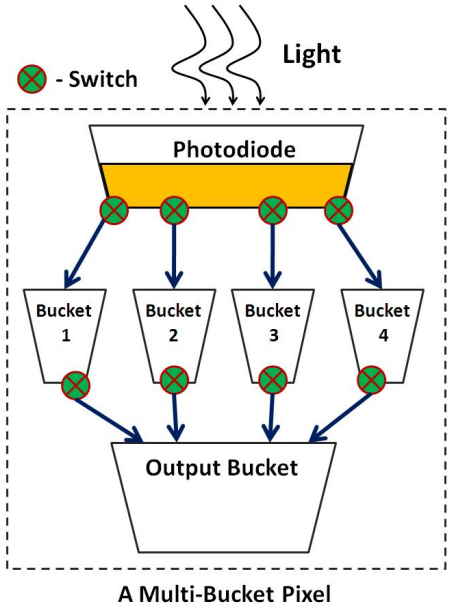

Figure 7: *A conceptual view of a multi-bucket pixel. The pixel consists of a photodiode, which converts incoming light into electrical charges, multiple buckets to accumulate photo-generated charges, and switches that are programmable by the user, such that charges in the photodiode can selectively go into the chosen buckets. Like most CMOS imager pixels, there is an output bucket called a floating diffusion inside the pixel. Charges accumulated in the numbered buckets are transferred to this output bucket, converted to voltages, and read out by an external circuit (not shown). Since this readout is non-destructive, the output bucket can form the sum of any number of the numbered buckets.*

In this paper, we perform our experiments using the quad-bucket sensor reported in [Wan 12]. This sensor comprises 640_Hx512_V array of 5.6μm pixels and each pixel contains four analog memories. Figure 8 shows the physical layout of our quad-bucket pixel.

4 Photon-Efficient High Dynamic Range Photography Using Multi-Bucket Sensors

An example timing diagram for the use of our 4-bucket sensor in HDR photography is shown in Figure 9. Since the images are interleaved in time, similar motion blurs appear in the captured images as argued earlier in this paper. Using this approach, the authors of [Wan 12] were able to synthesize HDR photographs without ghosting or color artifacts, and without performing any image alignment or motion compensation algorithms.

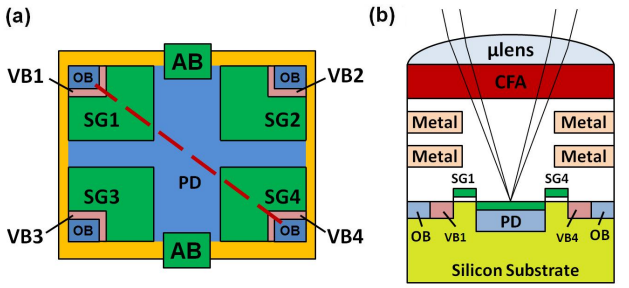

Figure 8: *A physical view of our quad-bucket pixel [Wan 12]. (a) Top view (b) Cross-sectional view across the red dashed line in (a). Four storage gates (SG) and two anti-blooming (AB) gates are connected to a photodiode (PD). The AB gates serve to reset the PD and provide protection against charge leakage to adjacent pixels. The SGs implement both the buckets and the corresponding switches to the PD. The Virtual barrier (VB) represents the switch between the accumulation bucket and the output bucket (OB). All the OBs are connected electrically (not shown). Light falling on the pixel opening is converged by the microlens (µlens) sitting on the top of the pixel. The light then passes through a color filter array (CFA) and a dielectric stack before being collected by the PD.*

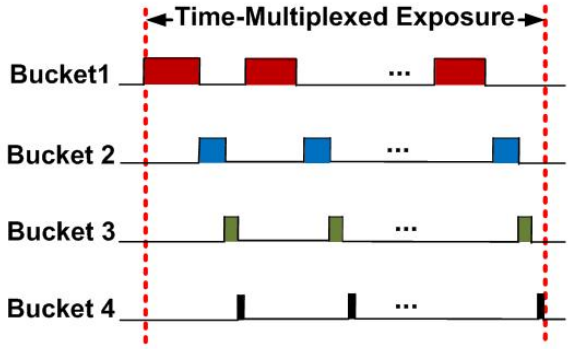

Figure 9: *One possible timing diagram for time-interleaved quad-exposure HDR photography. The duty cycles of the buckets can be programmed to set the desired exposure ratio of the captured images. However, we can improve on this timing, as described in Section 4.*

Although this approach is effective in overcoming motion and color artifacts, it does suffer from a serious drawback. Each captured image spans a longer absolute time, and therefore is more susceptible to motion blur, than the longest individual exposure in the protocol. To see why, consider the case of time-interleaved quad-exposure HDR. Assume T_1, T_2, T_3, and T_4 are the desired exposure times and that $T_1 > T_2 > T_3 > T_4$. Let us further assume that each of the four exposures is partitioned into N pieces, which are interleaved periodically. The four captured images would then span T_1+(N-1/N)(T_2+T_3+T_4), T_2+(N-1/N)(T_1+T_3+T_4), T_3+(N-1/N)(T_1+T_2+T_4), and T_4+(N-1/N)(T_1+T_2+T_3). Therefore, especially when N is large or the exposure ratio (i.e. T_i/T_{i-1}) is small, the increases in time spans of the images make them more susceptible to motion blur.

We now present an alternative time-interleaved HDR approach that removes this drawback. Figure 1 illustrates how the new approach works, by considering the example of capturing four frames with exposure times 8T, 4T, 2T, and T, where T is an arbitrary unit of time. Instead of setting the relative exposure of the buckets to be 8:4:2:1 as in the original time-interleaved HDR [Wan 12], in this new approach they are set to be 4:2:1:1. As shown in the figure, the image corresponding to exposure time T is read out directly from the bucket 4, while that corresponding to 2T is obtained by summing the signals captured by buckets 4 and 3. Similarly, the image corresponding to 4T is the sum of the signals captured by buckets 4, 3 and 2. Finally, the image corresponding to 8T is the sum of the signals captured by all four buckets. As we can see from the figure, the total capture time is shortened from 15T to 8T in this particular example. The image that corresponds to the longest exposure, i.e. the one captured by bucket 1, now spans an absolute time of 8T instead of 8T+(N/N-1)(7T). Since the longest exposure itself requires 8T, this new approach takes the theoretical shortest time to capture the data needed to synthesize the multiple images. As a result, this approach also incurs the minimal amount of motion blur.

The key to our improved protocol is "re-use" of photo-generated charges to reduce the amount of photons (i.e. exposure time) needed in forming the multiple images. We call this new approach ***photon-efficient high dynamic range photography***. Besides having the same benefits as the original time-interleaved HDR in removing motion and color artifacts, this approach has an additional benefit that the interleaving frequency of each exposure is effectively increased, as we can see from the figure. Consequently, this approach is more robust to the ghosting artifact discussed in Section 2.3.

4.1 Signal-to-Noise Ratio Analysis

Figure 10 shows a simplified signal chain that converts electrons in a photodiode to digital values at the analog-to-digital converter (ADC) output. A sensor's read noise, denoted by R, is defined to be the total noise generated by circuits in the signal chain. This read noise is added to an image every time it is read out. Therefore, if signals need to be added, as in the case of photon-efficient HDR photography, it is desirable to do so before these circuits to improve signal-to-noise ratio (SNR). Our multi-bucket sensor accomplishes this task by taking advantage of the fact that analog addition at the output bucket is noiseless. Using the previous example, let us assume S, S, 2S, and 4S are the signals acquired in buckets 4, 3, 2, and 1, respectively. Figure 11 then shows that the SNR of four images obtained by adding the signals before the circuits are higher than those when the signals are added after they are read out.

5 Other Applications

In this section, we present other computational photography applications that would benefit from our multi-bucket sensor.

5.1 Time-Interleaved Flash/No-Flash Photography

Flash/no-flash photography [Petschnigg 04] [Eisemann 04] requires a relatively static scene and a fixed camera. Otherwise, a good image alignment is required. However, registering flash and no-flash images is hard, because the two lighting conditions are different [Petschnigg 04] [Eisemann 04].

For the case of LED-based flash, our sensor can overcome this limitation by alternating between flash and no-flash and synchronizing the flash with one of the buckets. Thus, one of the buckets captures a scene illuminated by flash, while the other captures the scene under ambient light only. Compared to a conventional sensor, our multi-bucket sensor thereby produces two images representing the same span of time and having roughly the same motion. Figure 12 shows an experimental demonstration. The letter S attached to

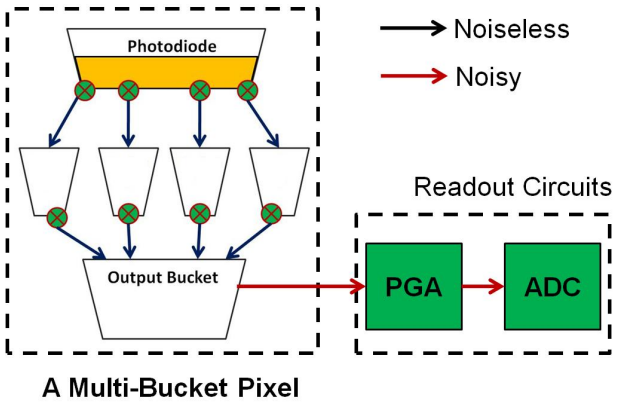

Figure 10: *A simplified signal chain that shows how electrons in a photodiode are converted to digital values at the analog-to-digital converter (ADC) output. Electrons in a photodiode are transferred and accumulated in a numbered bucket. They are then transferred to an output bucket which performs electron-to-voltage conversion. The voltage is amplified by a programmable gain amplifier (PGA). Eventually, the ADC converts this amplified voltage to a digital value. The total noise added to the signal by the PGA and ADC is defined to be the read noise of the sensor.*

(a)

	I_1	I_2	I_3	I_4
Signal	S	S	2S	4S
Noise	$\sqrt{S+R^2}$	$\sqrt{S+R^2}$	$\sqrt{2S+R^2}$	$\sqrt{4S+R^2}$
	$J_1 = I_1$	$J_2 = I_1+I_2$	$J_3 = I_1+I_2+I_3$	$J_4 = I_1+I_2+I_3+I_4$
Signal	S	2S	4S	8S
Noise	$\sqrt{S+R^2}$	$\sqrt{2S+2R^2}$	$\sqrt{4S+3R^2}$	$\sqrt{8S+4R^2}$
SNR	$\frac{S}{\sqrt{S+R^2}}$	$\frac{2S}{\sqrt{2S+2R^2}}$	$\frac{4S}{\sqrt{4S+3R^2}}$	$\frac{8S}{\sqrt{8S+4R^2}}$

(b)

	I_1	I_2	I_3	I_4
Signal	S	2S	4S	8S
Noise	$\sqrt{S+R^2}$	$\sqrt{2S+R^2}$	$\sqrt{4S+R^2}$	$\sqrt{8S+R^2}$
SNR	$\frac{S}{\sqrt{S+R^2}}$	$\frac{2S}{\sqrt{2S+R^2}}$	$\frac{4S}{\sqrt{4S+R^2}}$	$\frac{8S}{\sqrt{8S+R^2}}$

Figure 11: *Comparison of the signal-to-noise ratio (SNR) of two ways to add signals acquired in the numbered buckets. For simplicity, we consider only read noise and photon shot noise, with the later defined to be the square root of signal. Let I_1, I_2, I_3, and I_4 be the four images that are read out from the sensor. (a) Adding images acquired in the numbered buckets after they are read out. Here J_1, J_2, J_3, and J_4 represent the final four images that we are interested in. (b) Adding images at the output buckets inside a multi-bucket pixel. Comparing the last rows of the two tables, we see that adding signals before they are read out results in higher SNR of the final images.*

the oscillating metronome needle has exactly the same blur in both flash/no flash images. This would prevent artifacts when the two images are combined (not done here).

5.2 Color Photography using Active Illumination

The most common approach for color photography is to superimpose a color filter array (CFA) organized in a Bayer pattern atop the sensor. An alternative approach, suitable for controlled environments, would be to illuminate the scene sequentially using three light sources - Red (R), Green (G), and Blue (B) - and take three corresponding pictures. These can then be combined to form a final color photograph [Ohta 07]. This approach improves color fidelity, because it reduces inter-color cross-talk. It also improves light sensitivity by eliminating the CFA. Thus, this approach is attractive in light-limited applications such as capsule endoscopy [Ohta 07].

Unfortunately, such a frame-sequential approach suffers from color artifacts due to motion between the three exposures [Xiao 01]. Figure 13 shows how our quad-bucket sensor can overcome this artifact. Three time-interleaved RGB light sources are used to illuminate the scene while three of the four buckets are synchronized with the light sources. Three time-interleaved RGB images are obtained and are combined to form a final color picture. Color artifacts are avoided in this approach due to the time-interleaved nature of the captured images. In this protocol, the 4th bucket is not used, but it could have been used to image the scene as illuminated by a different light source, such as ultraviolet or white.

5.3 Time-Interleaved Multi-Flash Photography

Multi-flash photography [Raskar 04] has been proposed for depth edge detection and non-photorealistic rendering. However, like flash/no-flash, this idea cannot be applied to moving scenes. Figure 14 shows the experimental setup that demonstrates our quad-bucket sensor's capability in eliminating this limitation. Four groups of white LEDs, located at the top, bottom, left, and right of the sensor, are turned on sequentially and repeatedly during an exposure, with each group of LEDs being synchronized to one of the buckets. The resulting four images, which are illuminated from different directions, and therefore contain different shadows, can be used to compute a shadow-free image, as shown in Figure 15. Once again, because the quad-bucket sensor time-interleaves the captures, it is robust to motion in the scene.

5.4 Flash Matting

Flash matting utilizes the fact that a flash brightens foreground objects more than the distant background to extract mattes from a pair of flash/no-flash images [Sun 06]. One of its assumptions is that the input image pair needs to be pixel aligned. Although the technique was later improved by combining flash, motion, and color cues in a MRF framework [Sun 07], only moderate amounts of camera or subject motion can be handled.

Our multi-bucket sensor, when combined with a flash, can also be used to perform flash matting for a dynamic scene. Two of the buckets are used to record an image when a flash is illuminating foreground objects, while the other two buckets capture the scene when the flash is off. Alternatively, since the flash image is brighter, we can instead use three buckets to store the flash image and the remaining bucket for the no-flash image. Again, because our sensors allow interleaving of the captures, motion artifacts or other changes in the scene are effectively suppressed. By simple arithmetic operations on the captured images, we can compute a final image that shows only the foreground objects, as shown in Figure 16.

From the various applications described, we can see that the multi-bucket sensor, through enabling time-multiplexed exposure, eliminates the need for image alignment when combining images in multi-image computational photography and therefore avoids artifacts that would potentially arise when a conventional sensor is used.

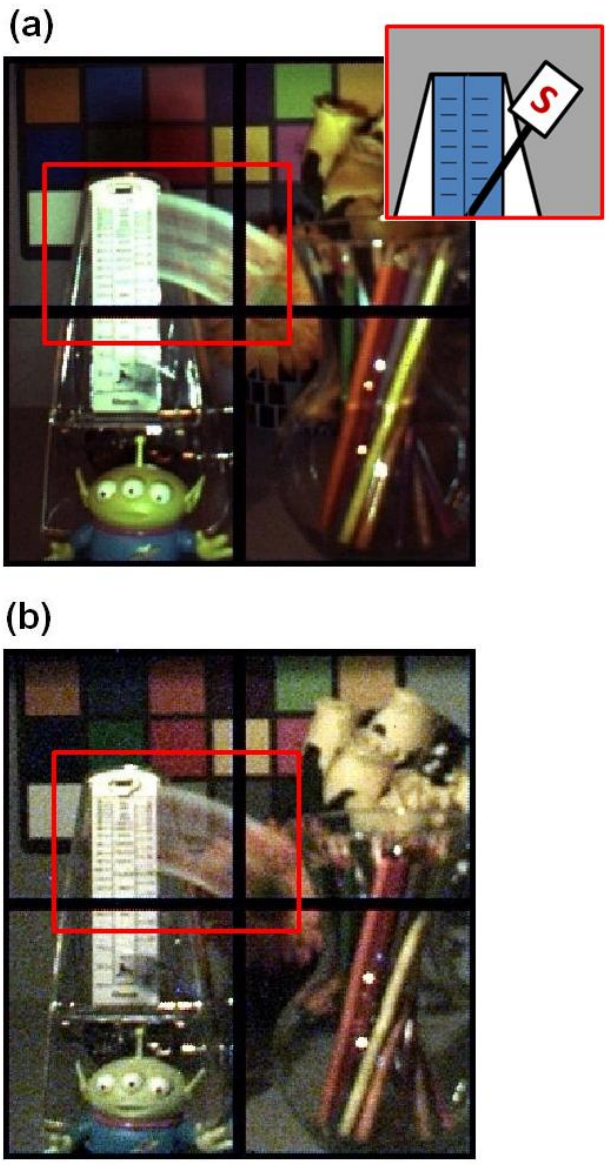

Figure 12: *Time-Interleaved Flash/No-Flash Photography. The scene is illuminated by a pulsing LED flash. A letter S is attached to the tip of a metronome needle, as illustrated in the cartoon (inset). The metronome needle is oscillating when the flash and no-flash images are taken. The letter S shows the same motion blur in both images. (a) Flash image (b) No-flash image.*

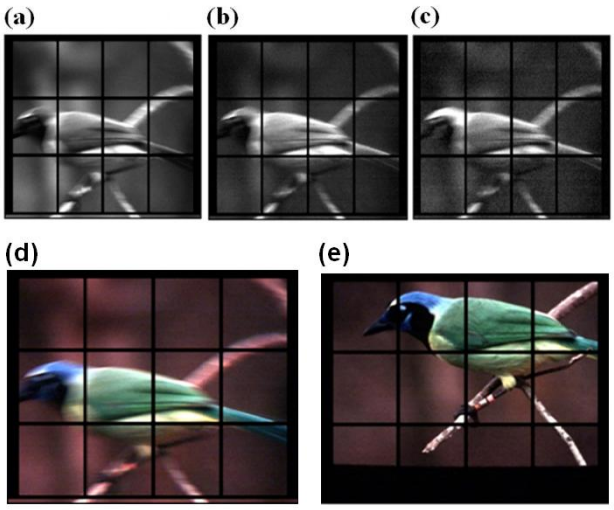

Figure 13: *Color Photography using Active Illumination. Three time-interleaved RGB light sources are used to illuminate a picture of a bird. To mimic camera shake, the picture and the light sources are deliberately moved during an exposure. (a) R (b) G (c) B images captured by three different buckets synchronized separately with the R, G, and B light sources. (d) Synthesized color image. Note the lack of motion artifacts. (e) Reference static image.*

6 Conclusions and Future Work

Although computational photography promises a paradigm shift in photography, existing efforts have focused mainly on modifying the optics or introducing novel reconstruction algorithms; there has been little research in image sensor technology, at least in the graphics and vision communities. The work described in this paper taps into this less-explored area.

Our multi-bucket sensor does have several limitations. For example, all pixels must switch at once. While it does not limit the applications presented in this paper, there may be applications that would benefit from individually controllable pixels. Also, the number of exposure conditions is limited by the number of buckets inside the pixels. Since a multi-bucket pixel needs to accommodate extra memories, it is in general larger than a conventional pixel, thereby resulting in a lower sensor resolution. Finally, all the buckets need to be read out before the next round of time-multiplexed exposure can start. Since our prototype sensor also has a low frame rate (e.g. 3fps), the gap in acquiring two sets of time-multiplexed frames can produce slight temporal aliasing in videography. Therefore, our sensor is more suitable for photography. However, this is not a fundamental limitation. Currently, the speed of our sensor is low because we use an off-chip analog-to-digital converter (ADC). A future design can have an on-chip ADC converter. In this case, the frame rate of our sensor can be significantly improved.

Although this paper focuses on photography, our sensor can also be used in 3D capture methods. For example, when applied to 3D triangulation using structured light, three of the buckets can be used to capture a scene illuminated by three different patterns while the last bucket records the scene due to ambient light only. By subtracting this background image from the images captured by the other three buckets, the effect due to spatial variation of ambient light is suppressed. Also, due to time-interleaving, temporal variation in ambient light is simultaneously eliminated. Therefore, 3D imaging using structured light can become robust against spatio-temporal variation in ambient light, if a multi-bucket sensor is used.

Instead of performing time-interleaved imaging, a multi-bucket sensor can alternatively be used to capture multiple frames of equal length back-to-back. This mode of operation is useful in low-light situation in which up to 4 handshake-free images can be captured, and subsequently aligned-and-averaged to improve the SNR of the final image. Since the images are captured back-to-back without losing time to image readout, image alignment is made much easier.

For computational photography researchers, a multi-bucket sensor is hardware with new functionality. We hope that this new functionality will stimulate development of new algorithms, enabling

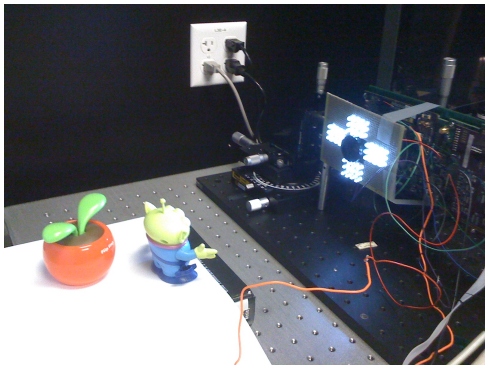

Figure 14: *Experimental setup of time-interleaved multi-flash photography. Four groups of white LEDs, located at the top, bottom, left, and right of the sensor, are turned on sequentially and repeatedly during an exposure with each group of LEDs being synchronized with one of the buckets in the quad-bucket sensor. This photograph of our experimental setup was taken by a conventional sensor, so all 4 banks of LEDs appear to be lit simultaneously.*

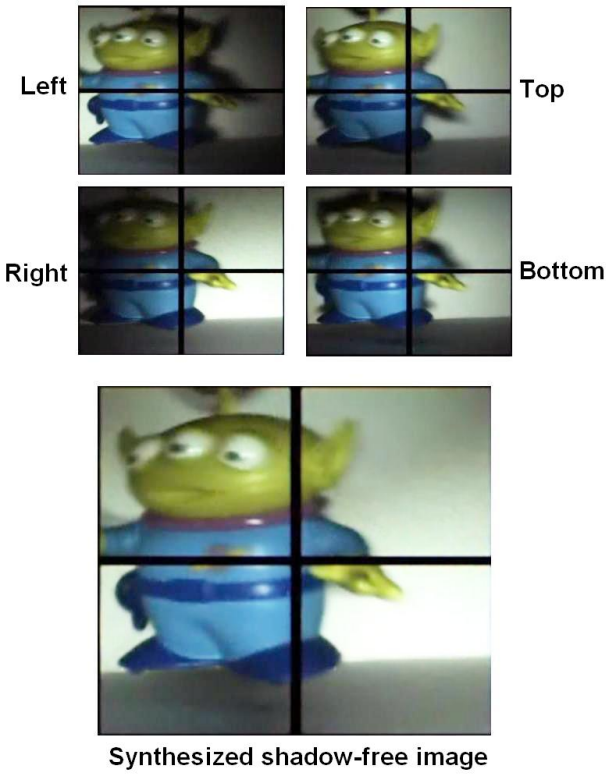

Figure 15: *Time-Interleaved Multi-Flash Photography. Images captured when top, bottom, left, and right LED groups are illuminating the scene and the subsequently computed shadow-free image.*

more applications. For example, our multi-bucket sensor can perform flutter shutter [Raskar 06] without throwing away 50% of the light.

Finally, besides new application development, it is hoped that this paper will trigger further research in image sensors and sensing protocols.

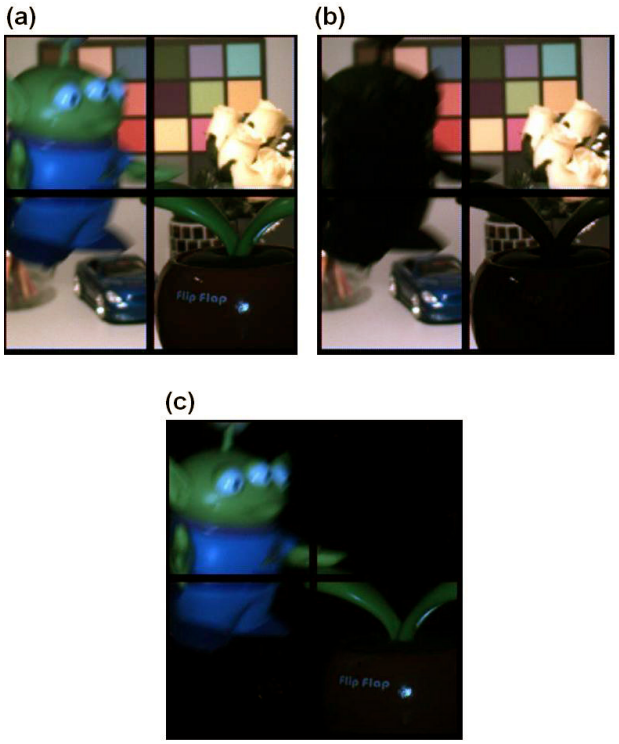

Figure 16: *Flash Matting. In this scene, the toy alien is moving while the leaves of the toy tomato are waving up and down. (a) Background + foreground image when the flash is on (b) Background only when the flash is off (c) Extracted foreground objects.*

Acknowledgements

The authors would like to thank Gennadiy Agranov, Hirofumi Komori, and Jerry Hynecek for their helpful suggestions on pixel design.

References

DEBEVEC, P. E., AND MALIK, J. 1997. Recovering high dynamic range radiance maps from photographs. In Proceedings of SIGGRAPH, 369-378.

EDEN, A., UYTTENDAELE, M., AND SZELISKI, R. 2006. Seamless image stitching of scenes with large motions and exposure differences. Proceedings of CVPR, 3, 2498-2505.

EISEMANN, E. and DURAND, F. 2004. Flash Photography Enhancement via Intrinsic Relighting. ACM Transactions on Graphics, 23, 3, 670-675.

GALLO, O., GELFAND, N., CHEN, W., TICO, M., AND PULLI, K. 2009. Artifact-free high dynamic range imaging. In IEEE ICCP.

KANG, S. B., UYTTENDAELE, M., WINDER, S., AND SZELISKI, R. 2003. High dynamic range video. ACM Transactions on Graphics 22, 3, 319-325.

KAWAHITO, S. HALIN, I., USHINAGA, T., SAWADA, T., HOMMA, M., AND MAEDA, Y. 2007. A CMOS time-of-flight range image sensor with gates-on-field-oxide structure. In IEEE

Sens. J., 7, 12.

KIM, S.J., HAN, S.W., KANG, B., LEE, K., KIM, J.D.K., AND KIM, C.Y. 2010. A three-dimensional time-of-flight CMOS image sensor with pinned-photodiode pixel structure. In IEEE Electron Device Lett., 31, 11 (Nov) 1272-1274.

MILLS, A., AND DUDEK, G. 2009. Image stitching with dynamic elements. Image and Vision Computing, 27, 10, (Sept) 1593-1602.

NAKAMURA, J. 2006. Image Sensors and Signal Processing for Digital Still Cameras. CRC Press.

OHTA, J. 2007. Smart CMOS Image Sensors and Applications. CRC Press.

PETSCHNIGG, G., AGRAWALA, M., HOPPE, H., SZELISKI, R., COHEN, M.F., AND TOYAMA, K. 2004. Digital Photograph with Flash and No-Flash Pairs. ACM Transactions on Graphics, 23, 3, 661-669.

RASKAR, R., TAN, K.-H., FERIS, R., YU, J., AND TURK, M. 2004. Nonphotorealistic camera: depth edge detection and stylized rendering using multi-flash imaging. ACM Trans. Graph. 23, 3, 679-688.

RASKAR, R., AGRAWAL, A., AND TUBMLIN, J. 2006. Coded exposure photography: Motion deblurring using fluttered shutter. ACM Transactions on Graphics, SIGGRAPH 2006 Conference Proceedings, Boston, MA 25, 795-804.

REINHARD, E., WARD, G., PATTANAIK, S., AND DEBEVEC, P. 2006. High Dynamic Range Imaging: Acquisition, Display and Image-Based Lighting. Morgan Kaufmann Publishers.

STOPPA, D., MASSARI, N., PANCHERI, L., MALFATTI, M., PERENZONI, M., AND GONZO, L. 2010. An 80 x 60 range image sensor based on 10μm 50 MHz lock-in pixels in 0.18μm CMOS. In IEEE ISSCC (Feb) 406-407.

SUN, J.,KANG, S. B., AND SHUM, H. Y. 2006. Flash matting. In Proceedings of the International Conference on Computer Graphics and Interactive Techniques. ACM SIGGRAPH. 361-366.

SUN, J., KANG, S.B., XU, Z., TANG, X., AND SHUM, H.Y. 2007. Flash cut: Foreground extraction with flash/no-flash image pairs. In IEEE CVPR.

SZELISKI, R. 2010. Computer Vision: Algorithms and Applications. Springer.

WARD, G. 2003. Fast, robust image registration for compositing high dynamic range photographs from hand-held exposures. Journal of Graphics Tools 8, 2, 17-30.

WAN, G., Li, X., AGRANOV, G., LEVOY, M., AND HOROWITZ, M. 2012. CMOS Image Sensors With Multi-Bucket Pixels for Computational Photography. Solid-State Circuits, IEEE Journal of, 47(4), 1031-1042

XIAO, F., DICARLO, J., CATRYSSE, P., AND WANDELL, B., 2001. Image analysis using modulated light sources. In Proc. SPIE Electronic Imaging Conf., San Jose, CA, pp. 22-30.
YAMAMOTO, K., OYA, Y., KAGAWA, K., NUNOSHITA, M., OHTA, J., AND WATANABE, K. 2006. A 128 x 128 Pixel Complementary Metal Oxide Semiconductor Image Sensor with an Improved Pixel Architecture for Detecting Modulated Light Signals. In Opt. Rev. 13, 64.

YASUTOMI, K., ITOH, S., AND KAWAHITO, S. 2010. A 2.7e- temporal noise 99.7% shutter efficiency 92dB dynamic range CMOS image sensor with dual global shutter pixels. In IEEE ISSCC (Feb) 398-399.