Frameless, time domain continuous image capture

Henry Gordon Dietz*
University of Kentucky, Lexington, Kentucky, USA

Abstract. *Most image sensors mimic film, integrating light during an exposure interval and then reading the "latent" image as a complete frame. In contrast, frameless image capture attempts to construct a continuous waveform for each sensel describing how the Ev (exposure value required at each sensel) changes over time. This is done using an array of on-sensor nanocontrollers, each independently and asynchronously sampling its sensel to interpolate a smooth waveform.*

Still images are computationally extracted after capture using the average value of each sensel's waveform over the desired interval. Thus, image frames can be extracted to represent any interval(s) within the captured period. Because the extraction of a frame is done using waveforms that are continuous time-varying functions, an Ev estimate is always available, even if a particular sensel was not actually sampled during the desired interval. The result is HDR (high dynamic range) with a low and directly controllable Ev noise level.

This paper describes our work toward building a frameless imaging sensor using nanocontrollers, basic processing of time domain continuous image data, and the expected benefits and problems.

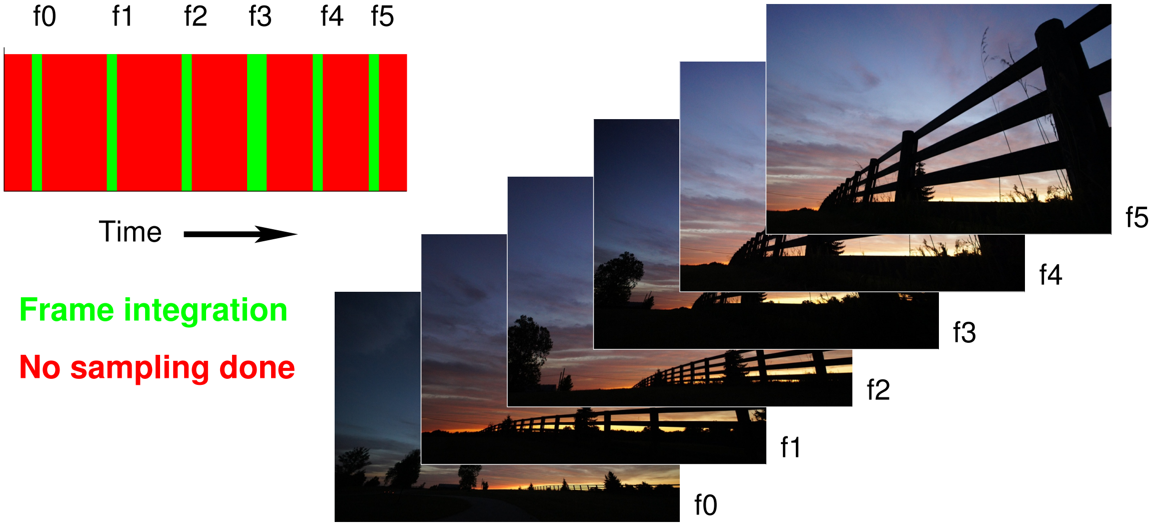

Figure 1. Traditional Image Capture

1. INTRODUCTION

The concept of optically creating a two-dimensional image in a *camera obscura* was known to Aristotle over 2,300 years ago. However, the images created were fleeting. By the 15th century, there is evidence that artists, perhaps including Leonardo da Vinci, were making lasting "copies" of camera obscura images by drawing or painting. In 1558, Giovanni Battista della Porta documented the use of a camera obscura as an aid in drawing.[9] Creation of permanent images using light-sensitive silver compounds coated on various materials was a 19th-century innovation. Film technologies continued to improve through the 20th century, and as images moved into computers, CCD and CMOS sensors began to replace film. Still, all these methods for capturing a permanent two-dimensional representation of a scene generally have shared the property that they capture an entire image at a time. Even movies and video are implemented by capturing a sequence of distinct images.

---

* Professor in the University of Kentucky's Department of Electrical and Computer Engineering, but also a member of the Center for Visualization and Virtual Environments. E-mail: `hankd@engr.uky.edu`

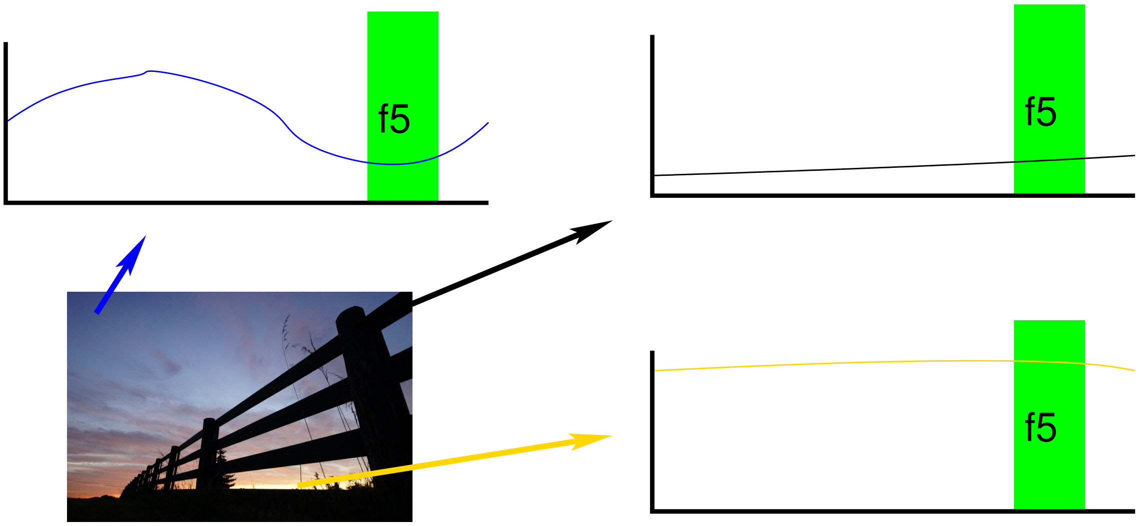

Figure 2. Ideal Time Domain Continuous Imaging (TDCI)

Capture of a sequence of images, or frames in a movie, is normally accomplished as shown in Figure 1. Each image is captured by exposing the film or sensor for a specific photon integration period of a duration often referred to as the *shutter speed* (S) or exposure time, *Tv*. This duration is primarily a function of the luminance *Bv*, sensitivity *Sv*, and aperture of the lens (either f/number or transmission-efficiency-adjusted T number) *Av*. Tv also is constrained by the range of available shutter speeds and cannot be longer than the reciprocal of the desired frame rate. Using a log scale, such as the *additive system of photographic exposure* (APEX) commonly applied inside cameras, exposure value $Ev = \mathrm{Av} + \mathrm{Tv} = \mathrm{Bv} + \mathrm{Sv}$. The problem is that any scene contains a range of luminances, and thus any Ev selected will be a compromise – degrading image quality in areas for which a different Ev would be appropriate. Further, under many circumstances the Ev will necessitate a Tv that is not 1/frame rate, resulting in either underexposure or potentially large temporal gaps between the images.

In contrast, frameless image capture attempts to construct a continuous waveform for each sensel describing how the Ev (exposure value required at each sensel) changes over time, as shown in Figure 2. We call this *Time Domain Continuous Imaging* – TDCI. Still images are computationally extracted after capture using the average value of each sensel's waveform over the desired interval. Thus, image frames can be extracted to represent any interval(s) within the captured period.

Decoupling aperture and shutter settings from scene brightness has been the dream behind "auto ISO" settings in digital cameras, but making significant ISO changes seriously degrades image quality. TDCI's ability to directly manipulate the interval represented by an image, without being directly constrained by the Ev, represents a fundamental change in how one thinks about imaging. For still images, this means that the virtual shutter speed may be selected freely, even after capture. It is also possible to nudge the virtual exposure interval forward or backward in time to capture precisely the intended action with zero "shutter lag."

For movies, the implications are even greater. The movie industry has suffered a variety of incompatible framerates: cinematic 24FPS, PAL 25FPS, and NTSC 59.94 fields/second. Converting between framerates inevitably causes "stutter" or at least a significant loss in quality. In contrast, the continuous waveforms are essentially a framerate-independent encoding from which essentially perfect frames may be extracted at any framerate significantly below the temporal resolution of the waveforms. Not only can the frames be sampled from the waveforms at any framerate, but there need not be any temporal gaps between frames. The "jumping" objects in pans, and discontinuities in motion in general, occur because conventional movie frames often represent exposure intervals that are significantly shorter than 1/framerate.

The direct capture of continuous waveforms is problematic, but there are a variety of methods by which continuous waveforms of varying quality can be synthesized from appropriately structured discrete samples.

Figure 3. NEX-5 on a Burke & James 4x5 with 127mm $f/4.7$ Ektar lens

Because the extraction of a frame is done using waveforms that are continuous time-varying functions, an Ev estimate is always available, even if a particular sensel was not actually sampled during the desired interval. The result is HDR (high dynamic range) with a low and directly controllable Ev noise level. Neither overexposure nor underexposure occur per se; bright sensels simply provide waveforms with potentially finer temporal accuracy than dark sensels.

Having described some of the key benefits of TDCI, the following sections focus on implementation issues. Section 2 discusses the spatial, dynamic range, and temporal goals for a TDCI sensor. Possible sensel designs are discussed in Section 3. The sensor system design is discussed in Section 4. Section 5 summarizes the status of this work.

2. IMAGING SENSOR GOALS

Although the bulk of image sensors are physically small, and there are many reasons that small sensors will continue to dominate the market as traditional camera functionality is replaced by cell phones and imaging sensors become increasingly common as input devices (as they have been for years in optical mice), the more interesting question is how good can an imaging sensor be if size is not constrained by packaging and other "artificially imposed" constraints? That is what a TDCI implementation should be capable of. It is useful to consider separately the properties in spatial, luminance, and temporal terms.

2.1 Spatial resolution

Suppose that resolution is constrained by the total resolution of available lenses. Perhaps the most fundamental limiting factor on lens resolution is diffraction. Fraunhofer diffraction of a uniformly-illuminated circular aperture yields a central bright spot size of approximately 2.44*wavelength*aperture, where aperture is the $f/$number of the lens.[2] The diameter of the aperture and lens focal length do not play any independent role. There is a little complication, however, in that a visible-light image generally is not being formed by a single wavelength. Typical RGB (red-green-blue) color filters have peak sensitivity around wavelengths of 450nm, 530nm, and 600nm respectively, so using the spot size at 450nm provides a reasonable bound. At $f/1$, the spot size is 1.1um, but by $f/5.6$ the spot size grows to 6.2um.

One would expect that ever larger sensors would produce steadily increasing total numbers of resolved pixels, but for practical reasons lenses that can cover a larger sensor generally have smaller apertures (large $f/$numbers), somewhat compensating for increases in sensor area. Finding the sensor size with maximum resolved pixel count is further complicated by the fact that most lenses do not achieve the diffraction-limited resolution. Thus, it is necessary to resort to searching measured resolutions of lenses.

Resolution measurements of many lenses are freely available on the Internet for lenses covering up to about 4x5 format film. Perhaps surprisingly, measured line-pairs-per-mm for lenses covering 4x5 are not significantly lower than for lenses designed to cover much smaller sensors – even APS-C – so the highest total resolution with readily-available lenses would be using the 4x5 format. According to the tests published by Perez and Thalmann,[11] while individual 4x5 lenses ranged from 10 to 85 lppmm, the average performance was 54.3 center, 51.2 middle, and 39.3 edge. This roughly corresponds to matching a Bayer-filtered sensor with about 500MP, while the best 4x5 lenses could justify up to about 1.5GP.

The 500MP sensel density is similar to that of the 14MP APS-C sensor in a Sony NEX-5. Thus, a variety of large-format lenses were tested using the NEX-5 mounted where the film holder would normally go on a 4x5 camera, as shown in Figure 3. Contrast was low for most large-format lenses, and one circa 1901 lens was quite poor, but resolution was generally consistent with the the numbers quoted above.

Resolution, however, is not just a spatial measure: the concept of resolution also applies to the dynamic range of brightness that can be recorded for each pixel and the precision with which the time period represented by a reading is known.

2.2 Dynamic range (luminance)

Dynamic range is primarily about the difference between the brightest value that can be recorded and the noise floor, but measurement is not well standardized. According to DxOMark sensor ratings,[1] the sensors in high-end consumer cameras have the ability to record a dynamic range between 9.6 and 14.4 Evs. Even 9.6 Evs is a significant improvement over earlier digital cameras, and the best cameras approximate the instantaneous dynamic range of human eyesight, but most sensors far exceed the dynamic range of most image output devices.[12] Much of the work in *high dynamic range* (HDR) imaging has centered on preserving image properties while mapping very large dynamic ranges into the smaller dynamic ranges supported by display technologies.[7, 8]

However, the dynamic range of naturally-occurring light is clearly huge, with no obvious upper bound, so there is a fundamental problem with sensors in which all sensels attempt to collect charge from photons for the same exposure interval and then digitize the stored charges. It is very difficult to implement analog charge storage, and routing of analog signals, that achieves both the high dynamic range and sensel density desired. Although there are a variety of techniques for capture of high-quality HDR image data, perhaps the most common involves the combination of multiple conventional exposures captured with different exposure times.[12] This technique essentially allows different sensels to have different photon integration times without requiring special hardware, but unfortunately temporally skews the sampling such that scenes that change over time cannot be effectively processed by this method.

2.3 Temporal resolution

The precision with which the time period represented by a sensel's value is known is a matter of temporal resolution. Most imaging devices aim to employ the same integration period for all sensels, but that is actually difficult to achieve.

Mechanical leaf shutters do expose all sensels simultaneously, however the effective aperture size changes during that exposure integration interval, introducing potentially-significant artifacts in out-of-focus portions of the imaged scene. Mechanical focal-plane shutters, which are most common in interchangeable-lens cameras, do not vary the aperture during exposure. However, the moving slit through which they expose takes time to traverse the sensor, such that integration interval is temporally shifted in the direction the slit travels across the sensor (typically either vertically or horizontally). At shutter speeds no faster than the "flash sync" speed, the moving slit is logically longer than the sensor dimension it crosses, and this means that there is a portion of the integration period during which all sensels are integrating photons, but the integration start and end times remain skewed by the time it takes a curtain or blade to travel across the sensor. For example, a focal-plane shutter might have a top speed of 1/4000s, but actually take 1/200s for a curtain to traverse the sensor. Thus, the 1/4000s exposure is actually temporally "smeared" across a sampling period that is 20x as long.

There are some materials that can have their optical properties electrically modified to create a shutter with no moving parts. For example, liquid crystal (LC) polarization can be electrically altered such that, in combination with a fixed polarizer, it can act as a shutter. The fixed polarizer significantly reduces the light reaching the sensor (and only passes light of the correct polarization). Further, crossed polarizers are not entirely opaque. For these and similar reasons, these types of technologies have not found widespread use in cameras.

The types of electronic shuttering commonly used are appealing for several reasons:

- No moving parts to draw power, wear out, cause camera shake, or make noise
- Simpler construction – the shutter is an integrated part of the sensor chip
- Smaller – no space needed in front of the sensor

The primary disadvantage in omitting a shutter in front of the sensor is that without one the integration period is not "stopped" until the sensel value is actually read: photons continue to hit the sensels. Because reading analog charge values essentially requires analog-to-digital conversion, which normally requires too much circuitry for each sensel to have its own converter, there typically is sensel-readout sequencing that creates a temporal skew closely resembling that of a mechanical focal-plane shutter.

Note that starting integration with electronic shuttering does not have the same problem as ending it. Simply shunting charge to ground effectively zeros all sensels, and that easily can be accomplished with any timing properties desired. Thus, in cameras that use a live view from the sensor, it has become common to provide an *electronic first curtain* mode in which the start of an exposure is done by electronically resetting, but the exposure is ended by closing the focal-plane shutter. This is particularly desirable because during live view electronic shuttering is used (providing lower image quality, but avoiding a rapid sequence of mechanical shutter firings). Without the electronic first curtain, the focal-plane shutter would first have to close so that it could re-open to start a high-quality capture exposure.

From a broader perspective, the problem with all the above schemes is that temporal sampling is not just often skewed, but is generally sparse. In capture of a sequence of images, the periods between captures are significantly long.

3. SENSEL DESIGN

The fundamental sensor property needed for this new model is that individual sensels must be able to function as somewhat independent photon counters. More precisely, the exposure integration periods must be able to be different for sensels with significantly different photon arrival rates. Conventional integrate-then-digitize sensors are not ideal for this model. Thus, let us consider several basic alternative concepts: single-pixel compressive sensing, single-photon sensing, threshold sensing, and a method for placing control logic under each sensel.

3.1 Single-pixel compressive sensing

One of the fundamental principles of digital signal processing is the concept of Nyquist rate sampling: that a signal cannot be recovered unless sampled with at least twice the frequency of the highest component frequency naturally occurring in the the signal. However, this requirement has been increasingly challenged as researchers have been able to apparently reconstruct signals, and images, from far less data. This controversial process of deliberately sampling at much lower than Nyquist rate, generally with a randomized sampling pattern, has come to be known as compressive sensing.[3]

Most research involving use of compressive sensing with image sensors seems to center on the use of a single photodetector (a so-called single pixel) to sample the sum of light from a random set of pixel locations. A DMD (digital micromirror device) is used to reflect light from a random set of pixel locations toward a single photodetector, and repeating this single sampling significantly fewer times than there are pixel locations on the DMD allows the values for each of the pixel locations to be determined.

However, the engineering of a DMD system accurately targeting arbitrary combinations of 500MP is a daunting challenge. As is discussed in Section 4.2, compressive sensing still has useful insights to offer, but existing design concepts simply do not appear to scale to meet the requirements outlined in this paper.

3.2 Single-photon sensing

In theory, it should be possible for each sensel to directly record the arrival time and wavelength of each individual photon that hits it. Such a sensor would capture all the information the lighting of the scene provides, and in that sense would provide perfect data from which images could be computationally constructed.

There have been a number of efforts to construct sensors capable of sensing arrival times for individual photons. Although there are a number of alternative sensel structures, perhaps the most commonly discussed is a Geiger-mode avalanche photodiode.[5] These structures generally do not sense the wavelength of each photon, so color information would have to be recovered using another mechanism, such as a traditional Bayer color filter array or perhaps Foveon-style color-by-penetration-depth stacking.

The motivation for single-photon detectors primarily has been construction of time-of-flight depth-map sensors. A light source on the camera is pulsed so that the time taken for a photon to arrive directly represents the distance that photon traveled from the pulsed source to the corresponding spot in the scene and back to the sensor. For example, a 128x128 single-photon sensor with 97ps temporal resolution within a 100ns exposure window was constructed by Niclass et al.[10] For time-of-flight detection, only the arrival of the first photon at each sensel is significant in determining the distance; later arrivals are presumed to be photons that took longer paths by reflecting off something else before hitting the relevant portion of the scene. This, combined with temporal gaps between exposure windows, results in specialized readout circuitry that would not suit our needs. Further, the readout circuitry consumes an area comparable to that of the sensels and requires high I/O bandwidth off-chip, making it unclear how a similar architecture could scale to the gigapixel resolution we target.

As conceptually attractive as single-photon detection is, the truth is that under normal scene lighting conditions photon shot noise makes the amount of statistically-significant information carried by individual photons minimal. Using multiple bits of digital data to resolve timing information that is essentially noise would complicate the readout circuitry and magnify I/O bandwidth problems without sufficient benefit.

3.3 Threshold detection

If the information carried by the arrival time of a single photon is too small to warrant recording and transmitting, the correct approach would be to detect when statistically significant numbers of photons have arrived. Using a photon-count threshold that is greater than 1 offers many benefits:

- Photon shot noise averages out, thus increasing the statistical accuracy of the photons sampling the scene – our goal is accurately sampling the scene luminance and color, not properties of photons
- Other noise sources have reduced impact with larger charges being accumulated
- Choice of threshold essentially sets the sample luminance accuracy, which will be roughly constant independent of how long it takes to reach the threshold; clearly, to get a luminance reading that is accurate to 10 bits requires sampling at least 1024 photons even if there was no noise
- Detection events become less frequent without increasing the amount of data to record per event, thus dramatically reducing the bandwidth required
- The reference time clock, and everything it controls, can run much slower, reducing signal routing issues and power consumption
- A much wider range of easily-fabricated sensel architectures is viable

The reduction in sample rate of course implies a loss of temporal accuracy. The single-photon detector cited earlier[10] has a temporal resolution of 97ps, or about 1/10000000000s. Conventional cameras rarely have shutter speeds faster than 1/4000s. Even if the sample rate was reduced by a factor of 10000X from the single-photon rate, samples would still have 1/1000000s temporal resolution – 250X better than the fastest shutter speeds commonly available on high-end still cameras.

With this reduction in data rate and less stringent requirements on the sensor and related circuitry, there are many viable sensel implementation methods. The key is that, as shown in Figure 4, each sensel waveform is

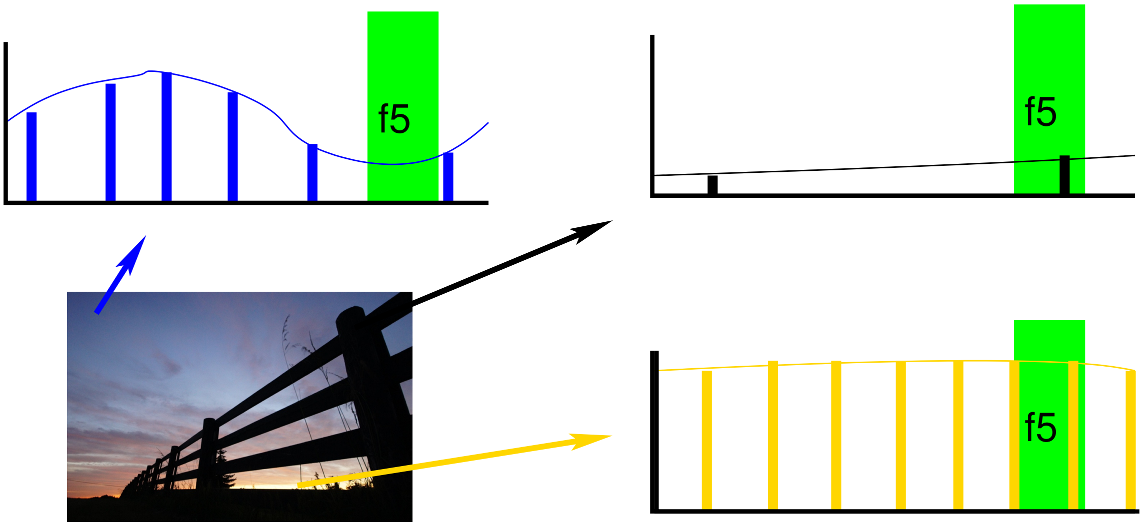

Figure 4. Practical Time Domain Continuous Imaging

updated no more than once each time the threshold is reached. This process can be viewed as the input version of the frameless output rendering proposed by Bishop et al.[4] As discussed in the following sections, it is even possible to reduce off-sensor bandwidth further by only updating the waveform when the time taken to reach threshold is significantly different from the *expected* time-to-threshold.

3.4 Logic under a segmented solar cell

Using threshold detection, it is easy to imagine a modest amount of digital control logic being associated with each sensel, but dedicating valuable potentially light-sensing area for that purpose is difficult to justify. Unfortunately, placing logic and wiring under conventional CCD or CMOS sensels is problematic. There are a variety of technologies that would allow sensels to be fabricated on top of circuitry. The exotic processing that produces rear-illuminated sensors could be used, but there are great technical difficulties in making large rear-illuminated sensors. Instead, why not build a segmented solar cell?

Solar cells are commonly fabricated using semiconductor process technologies that are two or three generations old, which is equivalent to saying that they are conservative, well tuned, processes with acceptable yields even for wafer-scale systems – which solar cells commonly are. The solar cell itself can be a thin or thick film applied to the surface of the wafer after circuitry has been imposed upon it. In fact, segmented solar cells have been fabricated on top of circuitry before, for the purpose of more efficient power management.

Solar cells in general have sensitivity profiles that are not too different from those of conventional CCD and CMOS sensors, although different materials have different responses. In fact, the obvious way to recover color might be to pattern use of several different materials rather than imposing a traditional Bayer filter on the sensor. Because the solar cells make-up the entire top layer of the chip, the fill-factor of a solar cell segment can be quite high, so it is probably not necessary to employ a microlens array to increase efficiency. If the sensel density is selected to match a Nyquist sampling of the intended lenses, it also should not be necessary to incorporate an anti-alias filter. Thus, the sensor filter stack might consist of only a protective cover glass and perhaps a NIR-blocking filter or coating.

The solar cell segments would be deposited on an insulating passivation layer covering most of the already-fabricated digital circuitry on the wafer, with gaps in the insulation to make connections to each solar cell segment and for bonding pads. The charge from each solar cell segment is passed through a diode to the circuitry below such that no charge is transferred until the designed threshold is reached. This essentially creates a digital sensel threshold detection circuit in which the output is 0 until the threshold is reached, at which time it transitions to 1 and can be detected and reset.

4. SENSOR SYSTEM DESIGN

Given a technology is which threshold-detection sensels can be fabricated above digital logic and wiring, the question becomes one of how to most efficiently control the sensor. There are several possibilities. The first is simulation of the threshold sensing using a conventional sensor; this is not expected to yield high performance, but it does allow testing algorithms for processing the expected type of data streams. More practical implementation alternatives include a form of compressive sensing and use of a massively-parallel nanocontroller array.

In any case, reduction of the bandwidth requirements is vital. Our goal of a 500MP HDR sensor with at least 16 bits per sensel results in highly impractical data rates using conventional control. For example, control logic that read an image at a time from that sensor would result in TB/s data rate for a relatively slow 1000FPS. Several orders of magnitude reduction in off-chip data rate are needed; fortunately, most scenes don't change all that fast, so substantial compression is possible.

4.1 Simulation using a conventional sensor

A conventional sensor is not well-matched to our stated goals. None the less, especially on cameras that use electronic shuttering, it is possible to approximate the desired behavior using a rapid sequence of raw captures. In that sequence, it may be appropriate to vary exposure somewhat so that HDR processing can extend the dynamic range for brighter areas of the scene; stacking of data from multiple exposures will naturally increase dynamic range for darker areas.

Some cameras are capable of a modest number of raw captures at frame rates up to about 10FPS, but fundamentally the issue is write speed to the storage medium. To obtain higher frame rates, or a longer sequence of frames without interruption, the raw images must be aggressively compressed. Cameras supporting video recording generally employ very effective compression schemes, such as H.264 – the *advanced video coding* (AVC) specified in MPEG-4 part 10. However, such a sophisticated algorithm requires significant computational resources for encoding, and very few consumer cameras offer video at greater than 1080P spatial resolution at 60FPS.

To create a continuous waveform for each sensel, it should be sufficient to record sensel data only when the waveform differs from predicted behavior. The simplest prediction is that values in the raw buffer are roughly the same as they were in the previous capture. This approach was tested using both USB webcams and implementation inside a Canon PowerShot A4000 under CHDK.

The logic under CHDK is something like:

```
for (each block b in the raw buffer) {
  h=hash(block b);
  if (h==oldh[b]) {
    runlength=runlength+1;
  } else {
    oldh[b]=h;
    output(runlength);
    output(block);
    runlength=0;
  }
}
```

where the hash function serves two purposes. First, it reduces the memory required: a copy of the A4000's 12-bit 16MP raw buffer would not fit in the 3MB of free memory available. Second, it embodies the simple prediction that values will not change beyond a noise threshold by ignoring the least-significant bits of each sensel value when computing the hash. There was a significant reduction in the data volume to be output using this method, but the raw capture rate was not increased.

The USB webcam version was quite different, simulating threshold sensel logic. For each sensel, there are four counters: the time since the level changed, number of times threshold was hit since then, the remaining sub-threshold value, and the time to accumulate the sub-threshold value. As a new image is obtained from the webcam, every sensel updates its sub-threshold value and time. If that value hits or exceeds the threshold, the virtual time it took and number of times the threshold was hit are computed. This pair of values is either merged with the previous values or, if the previous values embody a significantly different rate, the previous values are output and the new values replace them. This simulation achieved reasonable compression rates, but is not really practical as anything more than a testbed for development of software processing TDCI streams. For example, we have created software prototypes for extraction of images from these TDCI streams.

4.2 Adaptive compressive sensing

The developing standard concept of compressive sensing is that nonadaptive linear projections that preserve the structure of the signal are captured, and then an iterative reconstruction process is applied.[3] However, it is actually quite easy to make the sampling of threshold sensels be adaptive.

Suppose a sensor is constructed such that each sensel is independently accumulating charge until the threshold is reached, at which time the sensel changes from reporting "0" to reporting "1." Further, assume that the act of reading a sensel value of "1" dumps the accumulated charge and resets the output to "0." Clearly, the average photon arrival rate during the time taken from reset to output of "1" is threshold / time.

For each sensel, it is possible to externally track the expected time at which sufficient charge will have been accumulated to next reach the threshold. If the expected times for all sensels are correct, there is no error introduced by simply reporting that each sensel reached its threshold on schedule and, more importantly, there is no need to sample the sensel to obtain this report. Thus, the concept is to randomly sample sensels such that the probability of sampling a particular sensel is approximately proportional to the uncertainty with which its time-to-threshold may have already occurred. In other words, sensels with either a longer expected time to threshold or more predictable behavior can be sampled less frequently than other sensels.

For example, a sensel which has taken approximately 5000us to reach threshold for every sampling over the last 100000us is not a high priority to sample until a significant fraction of 5000us has passed since the last time the threshold was reached. In fact, entire 5000us intervals could be skipped without causing serious harm, especially if neighboring sensels also have given consistent readings over a relatively long period. The reasoning is simple: light levels at individual sensels may differ from each other in a small area (and with Bayer filters they will depending on color), but significant changes in the light level over time will rarely alter the value of only one sensel. Small changes are most likely due to photon shot noise, and can be smoothed without degrading quality of the sampling. Because sensel sampling is randomized both spatially and temporally, but changes in both dimensions should be relatively smooth continuous functions, the quality of spatial and temporal data can be effectively increased by reconstruction after capture.

It is even possible to use the "single pixel" concept to implement this type of compressive sampling without the complexity of a DMD. Rather than sampling individual sensels, the number of sensels at threshold in an arbitrary group of sensels could be electrically summed as either an analog or digital quantity. In fact, approximating the summation using either a simple OR or XOR (inclusive or exclusive OR) of the threshold status of the group might be effective.

4.3 Nanocontrollers

The problem with threshold detection is that each sensel operates independently. Simulating that independent operation by synchronously polling all the sensels is not really feasible. Consider polling each of the 500MP in our target design at the above quoted 1/1000000s temporal resolution. Instead, we propose to use an array of 500M nanoprocessors, each of which only has to poll the single sensel above it every 1/1000000s. This 1MHz polling rate is feasible with nanoprocessors running at approximately 1GHz.

The key idea is to construct a massively-parallel nanoprocessor computer on a single die which then has the light-sensitive elements fabricated on top of it. The primary task of each nanoprocessor is simply to count how long it takes for the sensel above it to reach threshold – which does not take a very powerful processor. If additional processing time and local memory space are available, then the nanoprocessor also can implement intelligent compression of the sensel waveform by tracking expected time-to-threshold and only generating output

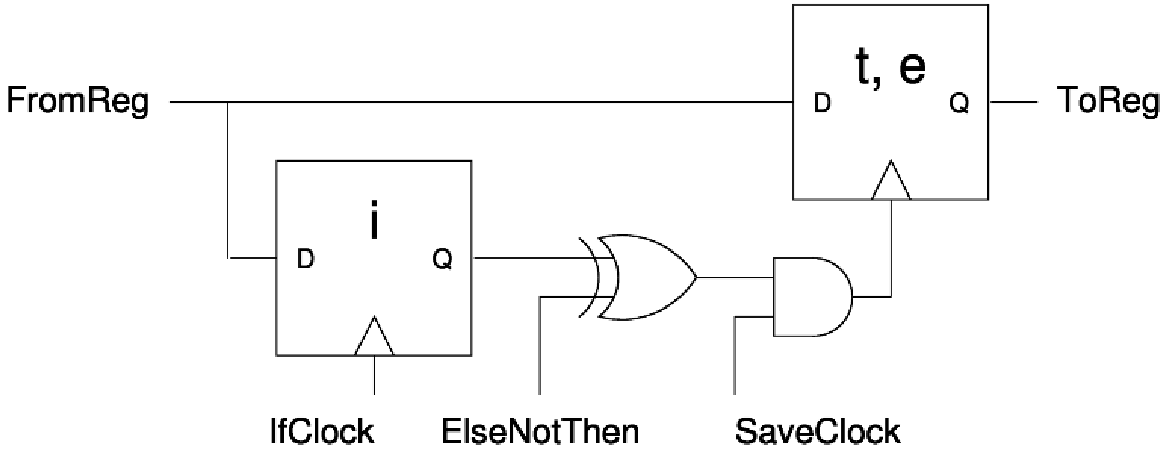

Figure 5. Nanoprocessor Architecture

when the threshold is not reached at the expected time (or within photon shot noise). It would even be possible to improve temporal accuracy of the waveform and compress it further by communicating with nearby nanocontrollers: changes in neighboring sensel values suggest this sensel's value will change at the same time. As circuitry continues to get smaller, but sensels reach a limit determined by optics, this massively-parallel computer can grow to take on increasingly complex processing tasks to improve the waveform quality and reduce the off-chip bandwidth required.

4.3.1 Nanoprocessor architecture

For over a decade, we have been developing nanocontroller technology for use in applications like this. The nanoprocessor architecture[13] implements only a single instruction, a 1-of-2-input 1-bit multiplexer instruction called SITE (store if-then-else) – as shown in Figure 5.

The nanocontroller array is a very simple, but massively-parallel, SIMD (single instruction, multiple data) computer with only a single program memory and a control hierarchy that fetches, decodes, and broadcasts control signals. The result is that the circuit complexity per processor is fewer than 300 transistors, including at least 20 bits of registers per processor. Thus, even using the somewhat outdated fab technology commonly applied to build solar cells, each nanocontroller might just fit underneath a 5um sensel – approximately the space available under each of 500MP using a 4x5 sensor.

4.3.2 Nanoprocessor software

Despite that hardware simplicity, we have developed compiler technology that allows the nanocontrollers to be programmed in a C dialect called BitC.[6] The compiler designs and optimizes a gate-level circuit to implement the program and then serializes the gates to translate them into the nanocontroller program. It takes many instructions to perform even a simple operation, but the nanocontrollers can easily run at a local clock rate in the GHz range. For example, the basic conditional counting operation with a 16-bit counter:

```
unsigned int:1 sensel;
unsigned int:16 counter;
counter=counter+!sensel;
```

takes a total of 110 instructions. The bit-level operations that implement this example code are shown graphically in Figure 6, which the reader might recognize as being equivalent to a ripple-carry increment circuit.

How sophisticated the nanocontroller software can be is partly a function of clock rate, but is more limited by how much memory space is available. At this time, placing circuitry under a segmented solar cell seems straightforward, but efficiently placing memory cells there may be problematic.

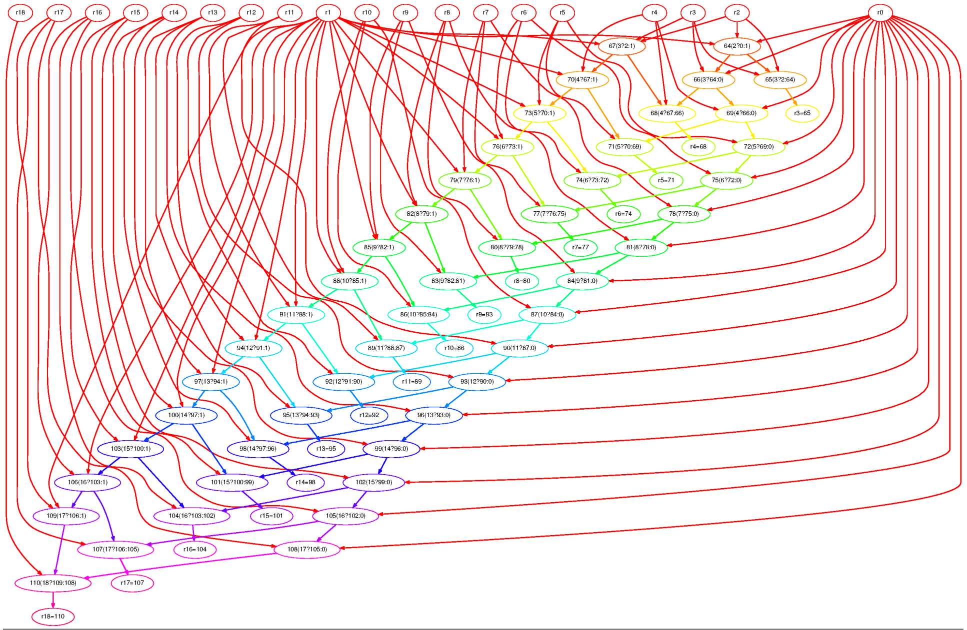

Figure 6. Nanoprocessor operations to conditionally increment a 16-bit counter

5. CONCLUSION

This paper has introduced the concept of frameless, time domain continuous imaging: TDCI. It has discussed the potential advantages, the implementation goals, and possible sensel structures and sensor architectures which could be used to implement TDCI. There have even been a number of tests, technologies, simulations, and software developed – in part building on a decade of work developing nanocontrollers. However, there is not yet a TDCI sensor or camera. Needless to say, most of TDCI is still future work.

The primary purpose of this paper is to get people thinking about a new model for imaging sensors – a model in which the sensor captures time-varying waveforms, not sequences of images.

Underlying that model is the basic truth that cameras are not about recording light, but the properties of a scene revealed to us by light. Individual photons reveal very little information about a naturally-lit scene, and thus are of little concern. On the other hand, changes in the scene over time are highly meaningful to us, so the goal should be to retain as much as possible of that information: spatial, dynamic range (luminance), and temporal.

REFERENCES

1. Camera Sensor Ratings by DxOMark. *http://www.dxomark.com/Cameras/Ratings*, 12 2013.
2. Diffraction effects. *CVI Melles Griot Technical Guide*, 2(2), 2013.
3. R.G. Baraniuk. Compressive sensing [lecture notes]. *Signal Processing Magazine, IEEE*, 24(4):118–121, 2007.
4. Gary Bishop, Henry Fuchs, Leonard McMillan, and Ellen J. Scher Zagier. Frameless rendering: Double buffering considered harmful. In *Proceedings of the 21st Annual Conference on Computer Graphics and Interactive Techniques*, SIGGRAPH '94, pages 175–176, New York, NY, USA, 1994. ACM.
5. S Cova, M Ghioni, A Lacaita, C Samori, and F Zappa. Avalanche photodiodes and quenching circuits for single-photon detection. *Applied optics*, 35(12):1956–1976, 1996.
6. Henry G Dietz, Shashi D Arcot, and Sujana Gorantla. Much ado about almost nothing: Compilation for nanocontrollers. In *Languages and Compilers for Parallel Computing*, pages 466–480. Springer, 2004.

7. Frédo Durand and Julie Dorsey. Fast bilateral filtering for the display of high-dynamic-range images. In *Proceedings of the 29th Annual Conference on Computer Graphics and Interactive Techniques*, SIGGRAPH '02, pages 257–266, New York, NY, USA, 2002. ACM.
8. Raanan Fattal, Dani Lischinski, and Michael Werman. Gradient domain high dynamic range compression. *ACM Trans. Graph.*, 21(3):249–256, July 2002.
9. H. Gernsheim. *A Concise History of Photography.* Dover photography collections. Dover Publications, 1986.
10. Cristiano Niclass, Claudio Favi, Theo Kluter, Marek Gersbach, and Edoardo Charbon. A 128 x 128 single-photon image sensor with column-level 10-bit time-to-digital converter array. *IEEE J. Solid-State Circuits*, 43(12):2977–2989, 2008.
11. Christopher M. Perez and Kerry L. Thalmann. Test Results – Large Format photographic lenses. *http://www.hevanet.com/cperez/testing.html*, 2004.
12. Erik Reinhard, Wolfgang Heidrich, Paul Debevec, Sumanta Pattanaik, Greg Ward, and Karol Myszkowski. *High dynamic range imaging: acquisition, display, and image-based lighting.* Morgan Kaufmann, 2010.
13. Akshay Vummannagari. Verilog design and fpga prototype of a nanocontroller system. 2010.