Multiple Capture Single Image Architecture with a CMOS Sensor

Brian Wandell[1,2], Peter Catrysse[1], Jeffrey DiCarlo[1], D. Yang[1] and Abbas El Gamal[1]
Electrical Engineering[1] and Psychology[2] Departments
Jordan Hall, Building 420
Stanford University
Stanford, CA 94305

Introduction

We describe a programmable digital camera sensor with pixel-level analog-to-digital conversion (ADC). The sensor, which was designed and implemented by our group, is programmable in the sense that the spatial and temporal pooling of the sensors in the array can be changed from frame-to-frame. This sensor programmability offers the opportunity to re-think the overall design of the digital capture system. In this paper we describe our experiments with a new system architecture, which we call *multiple capture, single image* (MCSI). The experiments illustrate how this architecture may be applied for demanding applications, such as image archiving, where a high-quality record of a scene is desirable.

This paper is organized into three parts. First the principles of the MCSI architecture are described. Second, the programmable CMOS sensor laboratory that implements the MCSI architecture is characterized. Third, applications of the MCSI architecture are presented. The applications illustrate how the MCSI architecture extends the intrinsic dynamic range image, intensity resolution, and spatial contrast sensitivity of the programmable sensor.

Multiple Capture Single Image Architecture

Principles

The principles of image capture used by the human visual system provide some useful examples for the design of electronic image capture systems. One principle of human vision is flexibility: properties of the image capture process can be altered as the viewing conditions change. For example, under low mean illumination levels human vision has relatively low spatial and temporal resolution and relatively high spatial and temporal summation. Under high illumination conditions, the tradeoff between resolution and summation is reversed. A second principle of human vision is multiple capture: in the early stages of image capture the visual pathways segregate image data into a set of parallel pathways. These parallel pathways are initiated within the retina, beginning at the transfer of signals from the photoreceptors to retinal neurons. Each pathway forms a mosaic of neurons that tiles the image plane. The receptive fields of the neurons within a mosaic share similar spatial, temporal and chromatic properties. The individual pathways are specialized to capture accurately different parts of the spatial, temporal and chromatic elements of the image. The outputs of these pathways are communicated in parallel to the brain where they are integrated into a unified percept of the world (Wandell, 1995).

In conventional film and digital imaging, primitive versions of the first principle, flexibility, are implemented. Images are measured with fixed sensor array properties, but over long time scales there is some flexibility in the system settings. For example, the aperture and integration times can be adjusted to coordinate with the imaging conditions. Experienced photographers often take multiple pictures

of the same scene, bracketing the imaging parameters. In this limited sense, conventional camera systems are programmable. Overall, however, conventional imaging includes only a limited degree of flexibility compared to the example of human vision. Each rendered image is acquired using a single set of parameters. Should the photographer bracket an exposure range using a set of pictures, the data are not easily integrated so that usually only the best image is used. Sophisticated efforts to integrate multiple captures are cumbersome to apply (Debevic & Malik, 1997).

Advances in electronic capture have made it is possible to acquire images at video rates. For some architectures the image capture parameters, can be modified quickly, frame-to-frame, effectively bracketing a parameter range. Because the data are in a digital format, the integration of data from multiple frames can happen very quickly, and a single image that integrates the best data from the range of parameters can be produced.

The MCSI architecture is based on the principle of making multiple, rapid measurements with a variety of sensor parameters. The data from the multiple captures are integrated into a single final image. By integrating a collection of such images, it is possible to extend the dynamic range, intensity resolution, and contrast resolution of the image sensor. Ultimately, it should be possible to place much of the multiple capture processing on the sensor itself. Hence, in the capture and integration of multiple frames into a single image should impose very little extra burden on other parts of the imaging system.

The MCSI architecture is analogous to the human visual system's use of parallel pathways. The principal difference is that the digital images are acquired over time rather than in parallel mosaics. It is not uncommon for technology to use rapid processing in cases where the human brain uses large parallel arrays.

The work described here makes two novel contributions. First, earlier work was focused on a single objective, such as super-resolution. Here, we introduce and examine the general imaging architecture. Second, the earlier work integrated a disparate array of technologies to produce the final image. Here, we introduce a laboratory based on a programmable sensor design that uses the MCSI architecture.

The MCSI architecture is applied to integrating images acquired with various integration times and spatial resolutions. We show that the data from multiple integration times both extends the dynamic range and the intensity resolution of the imaging system. By adjusting the spatial resolution improvements in the contrast sensitivity and intensity resolution for certain patterns are obtained. In a separate paper in this symposium, we discuss the topic of analyzing the spectral reflectance information using multiple captures.

Experimental Laboratory

In our laboratory, we have implemented an MCSI architecture to make experimental measurements with a real system. The architecture is designed around a programmable CMOS sensor built by our group (Yang, El Gamal, Fowler, & Tian, 1999). Currently, the sensor is supported by a field programmable gate array (FPGA) that controls the sensor properties and a collection of software tools to manage the sensor programming, read-out, and data analysis. We have implemented large portions of a simulator that describes the sensor characteristics (Catrysse, El Gamal, & Wandell, 1999). In addition, various light sources and spectral radiometric calibration equipment are present.

The CMOS sensor was built using a 0.35 micron technology. It contains a 640x512 array of pixels. Each pixel spans roughly 10 microns on a side. Part of the space within each pixel is devoted to the light sensor (photodiode). The remaining space is devoted to transistor circuitry that implements a pixel-level analog to digital converter (ADC). Each ADC is shared among a block four adjacent pixels. The voltage comparison levels used by all of the ADCs on the chip are delivered by an external signal that can be controlled by a program in the FPGA. By manipulating the signals from the FPGA, the sensor's *transduction function* (sometimes called the sensor "gamma" function) that governs the relationship between light level and digital value can be varied. Hence, the sensor has a programmable transduction function that can be varied from frame-to-frame. For the experiments described here a linear transduction function was used.

An overview of the pixel-level programmable ADC circuitry is shown in Figure 1. Prior to image acquisition, the photodiodes and storage capacitor are reset. During image capture, the switches from some of the photodiodes to the storage capacitor are closed, permitting the charge to accumulate both on the photodiode itself and the storage capacitor. At any point during the image acquisition period, the switch may be opened and the storage capacitor value can be read. (During read-out time, light from the image will accumulate charge on the photodiode's capacitance.) To continue integrating the image, the switch between the photodiode and read-out capacitor is closed and the image capture operation continues. Used in this way, measurements at multiple integration time measurements can be made during a single image acquisition period with no additional measurement time.

The circuit diagram shows that four pixels are connected to each ADC. It is possible to obtain spatial measurements from each pixel separately, yielding a 640x512 image. It is

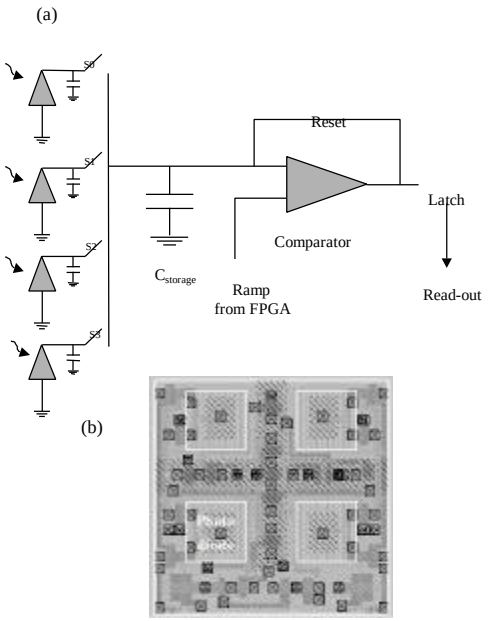

Figure 1. Pixel-level circuitry of the CMOS sensor. An ADC is shared between 2x2 pixel blocks. The transduction function of all pixels can be controlled by reprogramming the voltages sent to the ADC's comparator. The sensor's integration time and spatial summation can also be externally controlled.

also possible to close several switches at once and thus adjust the spatial resolution of the sensor array. For example, if an acquisition is made with all four switches closed, the spatial resolution is 320x256. The light sensitivity is correspondingly increased by about a factor of four. The switches can be opened and closed in any configuration, making it possible to adjust sensitivity and improve spatial sensitivity to patterns in horizontal, vertical or diagonal orientations. This integration takes place prior to quantization so that for a fixed integration time intensity levels below the quantization threshold of a single pixel can be detected by the spatial pooling.

Simulations and Measurements

Multiple Integration Times: Dynamic Range and Intensity Resolution

The dynamic range of a capture system is the light level that produces a response just below system saturation divided by the light level needed to produce a response just above the dark noise. Notice that the dynamic range is an input-referred measurement, independent of the signal quantization. Thus, a system that quantizes the output signal to 12 bits can have the same dynamic range as a system that quantizes the output to 8 bits. Two 8-bit systems can have very different dynamic ranges.

A straightforward method for extending the dynamic range of the capture device is to combine images obtained with different integration times. Light transduction functions (light intensity to digital count) from our CMOS sensor for measurements made with varying integration times are shown in Figure 2a. This figure suggests several properties that follow from combining multiple integration times.

First, adjusting the integration time produces valid responses (above dark noise, below saturation) over a much larger range, but the range does not increase indefinitely. Notice that for long duration integration times, the dark current noise shrinks the output range and limits the dynamic range. Notice that the measurements using an integration time of 256 ms classifies intensities near 2 cd/m^2 and saturates near 300 cd/m^2, so the dynamic range is roughly 2.5 log units. The transduction function for longer durations have decreasing dynamic ranges. The 2048 ms integration time has a dynamic range of less than 1.5 log units.

Second, the data from multiple integration times include several independent measurements from 10 to 100 cd/m^2. These

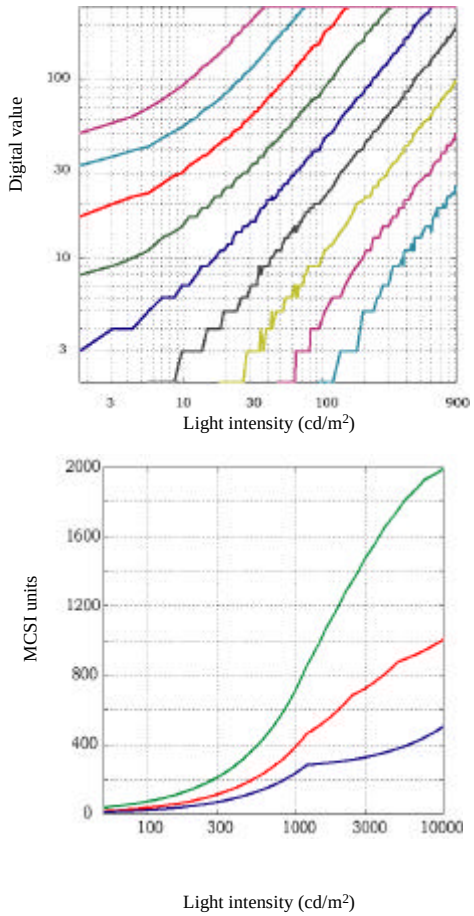

Figure 2. Extending the dynamic range and intensity resolution using multiple integration times. (a) The transduction function measured at single integration times of 8, 16, 32, 64, 128, 256, 512, 1024 and 2048ms. The curves continue to the right, but are not shown. (b) The transduction function of the MCSI architecture combines the data across integration times. See text for details.

measurements can be combined into a single intensity estimate, thereby reducing the system noise and improving the ability to resolve intensities in this region.
The data from the separate integration times can be combined into a single image as follows. Suppose there are N measurements at each pixel. Determine which of the measurements are within the sensor's compliance range and remove the others. The calibration data in (a) specify a relationship between r and the true input intensity, I, that is well-fit (smooth curves)

by a function r = I s t + d, where d is dark noise, t is integration time, and s is fixed pattern noise. From the measured digital value, r, estimate the signal intensity. Combine the data from different integration times by averaging the separate intensity estimates. Because the input data are digital, the estimates will also fall into an ordered list of discrete categories that we call the MCSI digital value. The number of digital values will depend on the sensor ADC and the number of integration times used in the capture process.

Figure 2b shows the relationship between input intensity and MCSI digital value for three system configurations. The horizontal axis measures the logarithm of the light intensity, and the vertical axis shows the MCSI digital value. The leftmost curve shows the MCSI values when eight integration times, spanning 8 to 64 ms (logarithmic spacing), are combined. The middle curve shows MCSI values for four integration times (8, 16, 32, 64) and the rightmost curve combines only 8 and 64 ms times. All three curves have the same dynamic range, but increasing the number of captures increases the slope of the curve (the intensity resolution). Hence, the MCSI architecture increases the intensity resolution beyond the conventional summary that uses only the image sensor ADC gain.

The images in Figure 3 show a high dynamic range scene captured using four different integration times. The peak intensity in the image was 1075 cd/m^2 and the darkest point we could measure was limited by the spectrophotometer (0.5 cd/m^2). With this dynamic range, none of the individual images accurately captures the scene. The problem of combining the four images into one is possible to solve numerically, as described above. But, to display the high dynamic range result on a low dynamic range medium, such as the printed page, requires additional processing. We have developed and implemented an algorithm to achieve this goal. The algorithm preserves the local image contrast

Figure 3. Four images of a high dynamic range scene acquired at integration times of 16, 64, 256, and 1024 ms. The original scene spanned more than 3.5 orders of magnitude of luminance intensity. None of the individual images is satisfactory, as they each contain either saturated regions or regions that are too dark to make out structure.

Figure 4. The four images in Figure 3 are combined into a single image using a method that preserves local contrast while reducing dynamic range (see (DiCarlo & Wandell, in press)).

but compresses the image intensity range to that of the display device (DiCarlo & Wandell, in press). The image produced by combining the four images in Figure 3 is shown in Figure 4. Such rendering algorithms would be an important part of an imaging system that includes how dynamic range measurements.

Multiple Spatial Resolutions: Spatial Contrast Sensitivity

The ability to adjust spatial resolution between frames, integrating the charge over blocks of four pixels prior to quantization, is a feature of the programmable CMOS sensor array. The MCSI architecture can use images captured at fine and coarse spatial resolution to improve estimates of image intensity.

A linear method for combining data obtained at two resolutions is this. Suppose that the intensity pattern falling on a block of four pixels is given by the four dimensional column vector, **p**. Suppose further that the switch settings are set in a pattern specified by the four dimensional column vector $\mathbf{s}_i$. Then, charge accumulated will be $\langle \mathbf{s}_i, \mathbf{p} \rangle$ where the angle brackets mean dot product (inner product). In the case that, say, only the first switch is closed, $\mathbf{s}_i$ will be (1,0,0,0). A similar unit basis vector describes the spatial image capture for each of the four switches. When all four switches are closed simultaneously, $\mathbf{s}_i$ = (1,1,1,1).

To integrate the data from multiple captures, create a matrix, **S**, whose rows are the switch settings, $\mathbf{s}_i$ .The data acquired at the i^{th} pixel will be $\mathbf{d}_i = \mathbf{S}\,\mathbf{p}$. We estimate the picture data, **p**, from the multiple image capture values, $\mathbf{d}_i$, using the least-squares calculation, $\mathbf{p}$ = pinv(**S**) $\mathbf{d}_i$ where pinv() is the pseudo-inverse operator. For the case of integrating one high resolution picture and one low resolution picture, the pseudo-inverse is

0.80	-0.20	-0.20	-0.20	0.20
-0.20	0.80	-0.20	-0.20	0.20
-0.20	-0.20	0.80	-0.20	0.20
-0.20	-0.20	-0.20	0.80	0.20

With these coefficients, the image intensity estimate can be obtained with very simple hardware operations. Hence, multiscale image captures can be combined to form a single high quality image.

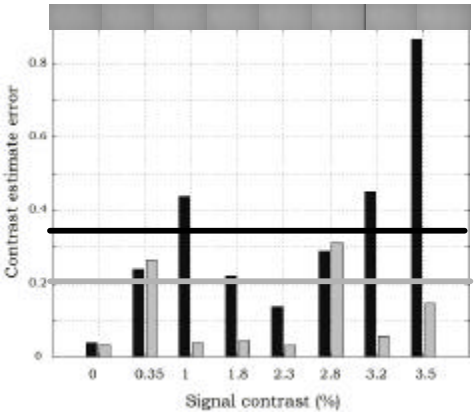

Figure 5. Image contrast estimates are improved by combining high and low resolution measurements. On average, the errors from the single image (black bars) were higher than the errors using the combined images (light gray bars).

Figure 5 shows empirical measurements of how combining images at two spatial resolutions can improve estimates of low contrast signals. The input image was the step pattern shown at the top of the graph. The contrast difference between the first step and each of the others was estimated using a single frame acquisition (black bars) or a two frame (high and low resolution) acquisition (gray bars). The integration time of the low resolution image is one quarter that of the high resolution image. The bars measure the error in the contrast estimate. The true contrast was measured using a spectrophotometer. The average error using the single frame is 34%. The error is reduced to 21% by adding the low resolution measurement. The improvement, roughly a square root of two, occurs only in low intensity regions because higher intensity areas saturate. The improvement in the estimate occurs for several reasons. First, additional frames are always helpful. Second, the low spatial resolution measurements are made at higher quantum well levels that have better SNR. Third, the summation occurs prior to quantization and thus eliminates some of the quantization noise at low intensity levels. All of these effects can be predicted using the software

simulator we have developed to model the sensor (Catrysse et al., 1999).

Conclusions

The development of programmable image sensor technology opens up new possibilities for imaging architectures. Programmable control of the spatial and temporal parameters of the image sensor suggests architectures that expand the basic sensor capabilities by combining multiple captures. We have shown that using this architecture the system as a whole performs better than a single sensor image. An imaging architecture that improves the quality of the peripheral components is a hallmark of human vision, where the quality of the optics and the performance of individual neurons are far below what we expect in commercial equipment. Despite the poor peripheral components, human vision achieves a spectacular level of overall performance. Much of the power of the system must be in the properties of the system, and these system properties are what we have set out to explore.

Here, we have illustrated how a programmable device that discriminates among 256 intensity levels, and has a dynamic range of 2.5 log units, can be used to create images that discriminate among 1600 levels over a 4.5 log unit dynamic range. The creation of novel imaging architectures that take advantage of the programmable sensors should be particularly advantageous for systems involved in digital image archiving.

Acknowledgements

We thank R. Erdmann, M. Khoury, H. Tian for their contributions to this work.

References

Catrysse, P., El Gamal, A. E., & Wandell, B. (1999). *Comparative analysis of color architectures for image sensors.* Paper presented at the Proc. of the SPIE, San Jose.

Debevic, P., & Malik, J. (1997). Recovering High Dynamic Range Radiance Maps from Photographs. *SIGGRAPH, August, 1997*.

DiCarlo, J., & Wandell, B. (in press). Rendering high dynamic range images. *Proc of the SPIE*.

Wandell, B. A. (1995). *Foundations of Vision*. Sunderland, MA.

Yang, D. X. D., El Gamal, A., Fowler, B., & Tian, H. (1999). *A 640*512 CMOS image sensor with ultra wide dynamic range floating-point pixel-level ADC.* Paper presented at the 1999 IEEE International Solid-State Circuits Conference, San Francisco, CA, USA.