2003 IEEE Workshop on Charge-Coupled Devices and
Advanced Image Sensors
Elmau, Germany - May 2003

A Software-Controlled Pixel-Level A-D Conversion Method for Digital Vision Chips

Shingo Kagami, Takashi Komuro, Masatoshi Ishikawa

Department of Information Physics and Computing, University of Tokyo
7-3-1, Hongo, Bunkyo-ku, Tokyo 113-8656, Japan

swk@k2.t.u-tokyo.ac.jp
http://www.k2.t.u-tokyo.ac.jp/

Abstract

A pixel-level analog-to-digital (A-D) conversion method suitable for digital vision chips is proposed. In this method, each quantization boundary is compared by determining whether the signal level integrated for a certain time has reached a certain voltage, where both the time and voltage are properly controlled by software for each boundary. We derive an algorithm to generate an optimal controlling schedule to realize an arbitrary A-D conversion scale subject to the condition that noise is minimized. Experimental results show that images with given A-D characteristics can be obtained with low noise.

1 Introduction

CMOS image sensors that integrate a digital processing element (PE) and a photo detector (PD) directly connected in each pixel are receiving attention as key devices to realize high-speed and general-purpose vision systems [1, 2, 3]. In this paper, we call these devices digital vision chips, or simply, vision chips. Since these vision chips enable high-speed programmable image processing on their focal planes, they are highly effective for many applications such as high-speed target tracking, grasping, gesture recognition, and microscopic target controlling [4].

Pixel-level analog-to-digital (A-D) conversion [5] of incident light is an essential technology for these vision chips because digital operations must be executed in each pixel. Since processing time and local memory capacity are limited, it is desirable to extract minimum necessary information in a flexible and efficient way during imaging, rather than achieve merely high-precision A-D conversion.

The fact that each pixel contains a programmable PE contributes to this requirement since it directly controls the behavior of the PD with software. If A-D conversion characteristics can be controlled arbitrarily by exploiting this feature, we expect that flexible visual sensing and processing adaptive to environments or purposes can be achieved.

In this paper, we propose a pixel-level A-D conversion method that realizes arbitrary conversion scales by controlling the PD operation with software, where each quantization boundary can be freely chosen from an ultra-wide dynamic range. With this method, each quantization boundary is compared to the amount of photocurrent by determining whether the signal level integrated for a certain time has reached a certain voltage. Both the time and voltage are properly controlled by software for each boundary. The availability of these two degrees of freedom, time and voltage, contributes to minimizing image noise.

We summarize related works in section 2. The pixel architecture is described in section 3. An algorithm to generate an optimal PD-controlling schedule is described in section 4. Experimental results are shown in section 5.

2 Related Works

Pixel-level A-D conversion is one of hot topics in the field of CMOS image sensors [5]. It has many advantages such as high signal-to-noise ratio, low power, and ability to integrate various signal processing at the pixel level including widening its dynamic range. It is also a technically challenging problem and many architectures and methods have been proposed [2, 6, 7].

Some of the works pointed out that their A-D conversion characteristics can be controlled. Yang et al. [6] developed an image sensor in which a bit-serial A-D converter is integrated for every four pixels, and realized wide dynamic range imaging by synthesizing several images sampled at different exposure times. This method, however, is unsuitable for real-time applications since it requires multiple sampling.

Forchheimer and Åström [2] proposed a concept of a digital vision chip called NSIP, and adopted a photo detector architecture similar to the design that this paper deals with, where illuminance is transformed into temporal information in each pixel. Åström et al. [8] proposed a method to linearize the A-D conversion

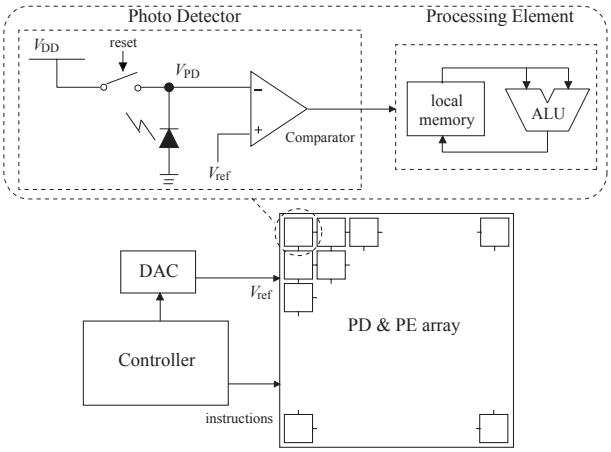


Fig. 1: Structure of the sensor.

scale in the NSIP architecture by either controlling the readout timings of the PD, or controlling the reference voltage supplied for the PD. However, their work is different from ours in that it was not considered to control these two degrees of freedom simultaneously. We will discuss this difference in detail in the end of section 4.

3 A-D Conversion Architecture

3.1 Pixel Architecture

First, the pixel architecture is described. Figure 1 shows the block diagram of the PD. The photodiode voltage $V_{\rm PD}$ is increased to $V_{\rm DD}$, then left to float. Next, $V_{\rm PD}$ is reduced, discharged by photocurrent. The comparator output holds logical 0 until $V_{\rm PD}$ drops under $V_{\rm ref}$, and then it turns to logical 1. Thus the photocurrent i can be estimated by observing the time t between the reset and the comparator output inversion as

$$i = \frac{C(V_{\rm DD} - V_{\rm ref})}{t} - \bar{i}_{\rm d} \qquad (1)$$

where C is the photodiode capacitance, and $\bar{i}_{\rm d}$ is the expectation of the dark current. The incident light illuminance is assumed to be constant during the exposure time.

The PE in each pixel has a bit-serial ALU and a bit-wise random-access local memory space, and the PEs are controlled in an SIMD manner. The comparator output of the PD is connected to the input port mapped to this local memory space so that the logical value can be read out at any time. Hence the time t can be counted by software. Equation (1) shows that the A-D conversion characteristics can be modified by controlling the counting intervals, or by controlling the voltage $V_{\rm ref}$.

We can conclude from simple observations that the voltage $V_{\rm ref}$ must be dynamically controlled. Since the time available for A-D conversion is limited when used for real-time applications, $V_{\rm ref}$ must be as high as possible in order to lower the detection limit of photocurrent, that is, $i_{\rm lower} = C(V_{\rm DD} - V_{\rm ref})/t_{\rm max} - \bar{i}_{\rm d}$ where $t_{\rm max}$ is the maximum available exposure time.

On the other hand, the exposure time must be as long as possible in order to suppress image noise. Thus the voltage $V_{\rm ref}$ must be as low as possible since lowering $V_{\rm ref}$ increases the time t for a certain photocurrent i, as can be seen from Eq. (1). These two conditions cannot be compatible if $V_{\rm ref}$ is fixed during the exposure time.

3.2 Real-time Control of $V_{\rm ref}$

We can carry out n-level A-D conversion with $V_{\rm ref}$ control as follows. With a PD readout time sequence $\{t_k\}$ and the corresponding $V_{\rm ref}$ sequence $\{V_k\}$ ($k = 1, \cdots, n-1$) predetermined,
1) Initialize the pixel value stored in the local memory with zero in each pixel,
2) Supply V_1 to $V_{\rm ref}$,
3) Reset all the pixel ($t = 0$),
4) For $k = 1, 2, \cdots, n-1$:
 i) Wait until $t = t_k$, then read out the PD output (a 1-bit digital value),
 ii) Supply V_{k+1} to $V_{\rm ref}$,
 iii) Add the PD output that was read out at i) to the pixel value.

This procedure is controlled by a dedicated vision chip controller that guarantees real-time operation with high temporal resolution [9]. Since its latest implementation achieves a resolution of 100 ns, $\{t_k\}$ can be scheduled with the resolution of 1/10000 of the maximum exposure time, when it is assumed to be 1 ms, a typical frame time for our vision chip. The controller is equipped with a D-A converter that supplies $\{V_k\}$ to $V_{\rm ref}$.

For illustration, a timing chart of A-D conversion with $n = 8$ is shown in Fig. 2. Downward-sloping lines that pass through the points (t_k, V_k) correspond to quantization boundaries. In the rest of this paper, we call a pair of $\{t_k\}$ and $\{V_k\}$ an A-D conversion schedule, or simply a schedule. In order to guarantee that the procedure works correctly, each downward-sloping line and the $V_{\rm ref}$ curve must have only one intersection point. This condition can be satisfied if we use a monotonically increasing $V_{\rm ref}$ curve.

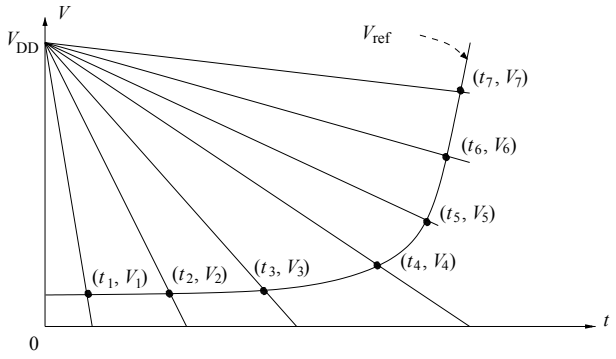


Fig. 2: Timing chart of A-D conversion. ($n = 8$)

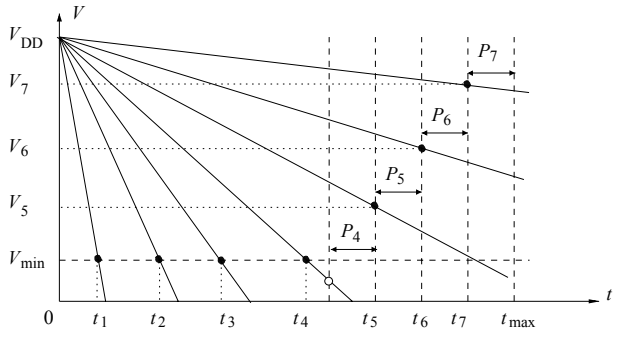


Fig. 3: Illustration for the scheduling algorithm. ($n = 8$)

4 Scheduling Algorithm

We have seen so far what kind of A-D conversion scale is achieved when a certain schedule is given. This section described the inverse procedure, that is, an algorithm to generate a schedule when a certain A-D conversion scale is given. The proposed algorithm takes a desired photocurrent quantization scale $\{i_k\}$ as its input, where digital numbers $(n-k)$ are to be assigned to the analog quantities of the photocurrent i if $i_k \le i < i_{k-1}$. And it yields the schedule $\{t_k\}$ and $\{V_k\}$ that achieves the specified $\{i_k\}$. Here, i_0 and i_n are fixed to ∞ and 0, respectively. The indices of i_k are assigned so that they correspond to those of t_k and V_k.

In order to realize the specified $\{i_k\}$, the downward-sloping line that corresponds to i_k, whose slope is $-(i_k + \bar{i}_\mathrm{d})/C$, must cross the V_ref curve at $t = t_k$. Since we still have latitude in determining (t_k, V_k), we can choose a schedule that maximizes t_k for any k so that image noise is minimized. The algorithm can be described as follows:

Algorithm 1 For $k = n-1, n-2, \cdots, 1$:

- If $V_{k+1} \neq V_\mathrm{min}$ and
  $\Delta V \left\lfloor \left\{ V_\mathrm{DD} - \dfrac{(i_k + \bar{i}_\mathrm{d})(t_{k+1} - P_k)}{C} \right\} / \Delta V \right\rfloor \ge V_\mathrm{min}$:
  $t_k = t_{k+1} - P_k$
  $V_k = \Delta V \left\lfloor \left\{ V_\mathrm{DD} - \dfrac{(i_k + \bar{i}_\mathrm{d})(t_{k+1} - P_k)}{C} \right\} / \Delta V \right\rfloor$
  Terminate abnormally if $V_k > V_{k+1}$.
- Otherwise:
  $V_k = V_\mathrm{min}$
  $t_k = \Delta t \left\lfloor \dfrac{C(V_\mathrm{DD} - V_\mathrm{min})}{(i_k + \bar{i}_\mathrm{d})\Delta t} \right\rfloor$
  Terminate abnormally if $t_{k+1} - t_k < P_k$. □

where V_max and V_min are the maximum and minimum values for V_ref, and Δt and ΔV are the resolution of the control of t_k and V_k, respectively. t_n and V_n are treated as $t_n = t_\mathrm{max}$ and $V_n = V_\mathrm{max}$. $P_k\,(k = 1, \cdots, n-1)$ is the time needed to execute the process between t_k and t_{k+1}, which depends on how the vision chip and its peripherals are designed and implemented. The algorithm terminates abnormally when the specified $\{i_k\}$ cannot be achieved by any feasible schedules.

As shown in Fig. 3, this algorithm finds the largest feasible t_k by subtracting P_k's from t_max, and then specifies V_k with t_k fixed at the point. After it iterates this process and comes across the point where V_k falls below V_min, it calculates t_k with V_k fixed at V_min. With this algorithm, arbitrary quantization boundaries chosen from an ultra-wide dynamic range can be achieved while image quality is optimized, that is, t_k is maximized for any k subject to the physical constraints of $0 \le t_k \le t_\mathrm{max}$, $P_k \le t_{k+1} - t_k$, $V_\mathrm{min} \le V_k \le V_\mathrm{max}$ and $V_k \le V_{k+1}$.

Since i, t and V_ref are mutually constrained by Eq. (1), once two of them are fixed, the other immediately follows. The method proposed by Åström et al. [8] utilizes this relation to realize given A-D conversion scales while keeping the reference voltage constant, or while keeping the PD readout interval constant. Our method, on the other hand, manipulates two degrees of freedom, time and voltage, to control one degree of freedom of quantizing photocurrent. This means that we have a redundant degree of freedom left, and it is exploited to minimize image noise.

5 Experimental Results

This section describes the experiments applying the proposed method to a vision chip [10] that was fabricated in a 0.35 μm CMOS process. It integrates 64×64 pixels and V_DD is 3.3 V. Dark current is ignored in the

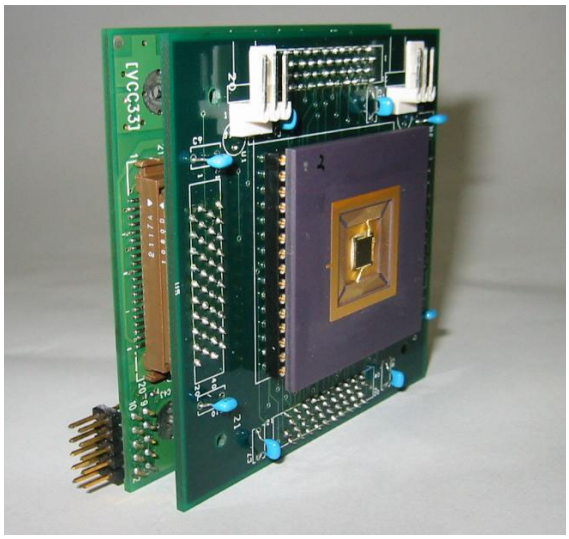


Fig. 4: Photograph of the vision chip and its controller.

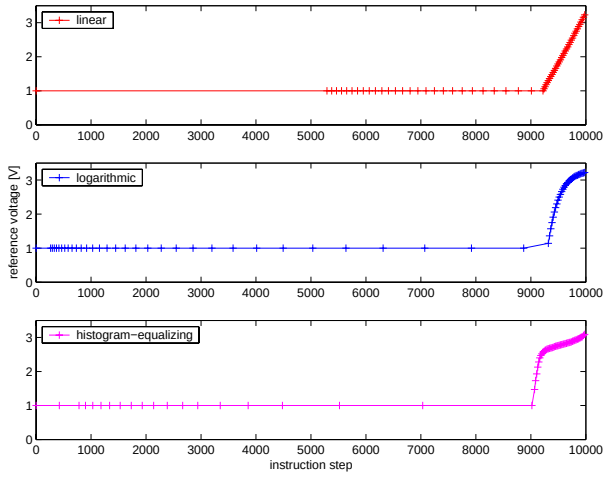


Fig. 5: A-D conversion schedules for several characteristics.

rest of this paper. Figure 4 shows the photograph of the vision system used in these experiments where the vision chip and its controller are implemented.

First, a linear scale was given to the scheduling algorithm, that is

$$i_k^{\text{linear}} = 0.03 I_0 \times (64 - k) \quad (k = 1, \cdots, 63) \quad (2)$$

where I_0 is the amount of photocurrent that inverts the PD output just at the moment $t = t_{\max}$ with V_{ref} fixed at $V_{\min}$, $V_{\max} = 3.3$ [V], $V_{\min} = 1.0$ [V], $t_{\max} = 8$ [ms], $\Delta V = 0.01$ [V], and $\Delta t = 1/10000\, t_{\max}$. P_k was set to $(3\lfloor \log_2 k \rfloor + 7)\Delta t$, which is actually required by the implementation [10, 9]. The generated schedule and the realized characteristics are shown as 'linear' in Figs. 5 and 6.

A 6-bit digitized image of a grayscale chart obtained using this schedule is shown in Fig. 7 (a). For comparison, images with the same characteristics obtained with V_{ref} fixed at 3.0 V, 2.0 V and 1.0 V are shown in (b), (c) and (d), respectively. When V_{ref} is fixed high, we can see that the image has a great deal of noise. Although the noise is reduced as V_{ref} is decreased, it makes the detection limit high so that dark scenes cannot be imaged. The results show that the proposed method overcomes these drawbacks.

Figure 8 shows images of a scene obtained with different scales. The scene includes a lighted incandescent lamp and a stuffed dog. An image with the linear scale of Eq. 2 is shown in (a), and one with a logarithmically compressing scale of

$$i_k^{\log} = 0.03 I_0 \times 1.12^{(64-k)} \ (k = 1, \cdots, 63) \quad (3)$$

is shown in (b). An image with scale that performs histogram equalizing is shown in (c). Their schedules and characteristics are also shown in Figs. 5 and 6 as 'logarithmic' and 'histogram-equalizing.'

6 Conclusion

We have proposed a method to control the operation of pixel-level A-D conversion with software. The method can achieve any given A-D conversion characteristics with image noise minimized. It can expand the applicable scope of the programmability of digital vision chips so that it includes not only processing but also the image sensing process, which is an important feature to achieve flexible real-time vision systems with adaptability to environments or purposes.

References

[1] Thierry M. Bernard, Y. Zavidovique, and Francis J. Devos, "A programmable artificial retina," *IEEE Journal of Solid-state Circuits*, vol. 28, no. 7, pp. 789–798, 1993.

[2] Robert Forchheimer and Anders Åström, "Near-sensor image processing: A new paradigm," *IEEE Transactions on Image Processing*, vol. 3, no. 6, pp. 736–746, 1994.

[3] Takashi Komuro, Idaku Ishii, and Masatoshi Ishikawa, "Vision chip architecture using general-purpose processing elements for 1ms vision system," in *4th IEEE International Workshop on Computer Architecture for Machine Perception (CAMP'97)*, 1997, pp. 276–279.

[4] Masatoshi Ishikawa and Takashi Komuro, "Digital vision chips and high-speed vision systems," in *Digest*

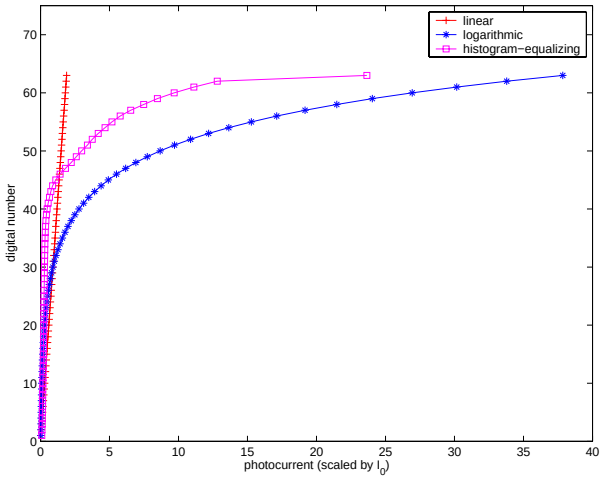


Fig. 6: Realized A-D Conversion characteristics.

of Technical Papers of 2001 Symposium on VLSI Circuits, 2001, pp. 1–4.

[5] Abbas El Gamal, David Yang, and Boyd Fowler, "Pixel level processing – why, what, and how?," in *Proceedings of the SPIE Electronic Imaging '99 Conference*, 1999.

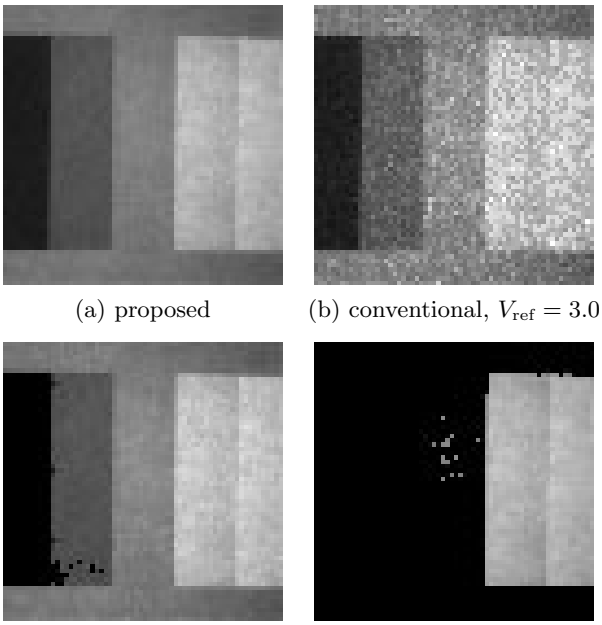
[6] David X. D. Yang, Abbas El Gamal, Boyd Fowler, and Hui Tian, "A 640 × 512 CMOS image sensor with ultrawide dynamic range floating-point pixel-level ADC," *IEEE Journal of Solid-state Circuits*, vol. 34, no. 12, pp. 1821–1834, 1999.

[7] Shinichi Yoshimura, Toshinobu Sugiyama, Kazuya Yonemoto, and Kazuhiko Ueda, "A 48kframe/s CMOS image sensor for real-time 3-D sensing and motion detection," in *Digest of Technical Papers of 2001 IEEE International Solid-State Circuits Conference (ISSCC 2001)*, 2001, pp. 94–95.

[8] Anders Åström, Robert Forchheimer, and Per-Erik Danielsson, "Intensity mappings within the context of near-sensor image processing," *IEEE Transactions on Image Processing*, vol. 7, no. 12, pp. 1736–1741, 1998.

[9] Shingo Kagami, Takashi Komuro, Idaku Ishii, and Masatoshi Ishikawa, "A real-time visual processing system using a general-purpose vision chip," in *Proceedings of the 2002 IEEE International Conference on Robotics and Automation*, 2002, pp. 1229–1234.

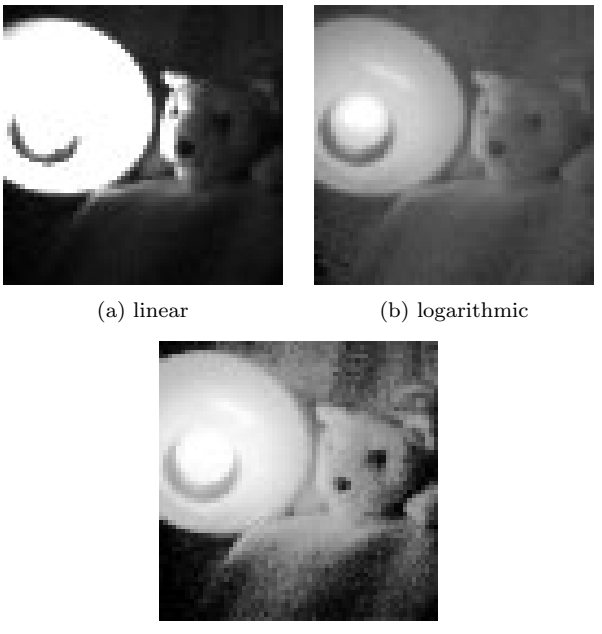
[10] Takashi Komuro, Shingo Kagami, and Masatoshi Ishikawa, "A new architecture of programmable digital vision chip," in *Digest of Technical Papers of 2002 Symposium on VLSI Circuits*, 2002, pp. 266–269.



(a) proposed (b) conventional, $V_{\mathrm{ref}} = 3.0$

(c) conventional, $V_{\mathrm{ref}} = 2.0$ (d) conventional, $V_{\mathrm{ref}} = 1.0$

Fig. 7: Images of a grayscale chart with linear characteristics.



(a) linear (b) logarithmic

(c) histogram-equalized

Fig. 8: Images obtained with different characteristics.