A 100,000 fps Vision Sensor with Embedded 535GOPS/W 256x256 SIMD Processor Array

Stephen J. Carey, Alexey Lopich, David R.W. Barr, Bin Wang and Piotr Dudek

The University of Manchester, School of Electronic and Electrical Engineering
Manchester, M13 9PL, United Kingdom
Tel: 44(0)1613064721; p.dudek@manchester.ac.uk

Abstract

A vision chip operating with 1.9pJ/OP efficiency has been fabricated in 0.18µm CMOS. Each of the 256x256 pixel-processors (dimensions 32x32µm), contains 14 binary and 7 analog S2I registers coupled to a photodiode, an arithmetic logic unit, diffusion and asynchronous propagation networks. At the chip's periphery, facilities exist to allow pixel address extraction, analog or digital readout. The chip has been exploited to conduct real-time image processing operations at 100,000fps, locating a closed-shape object from amongst clutter.

Keywords: SIMD, vision-chip, parallel, processor, cellular and array

Introduction

A conventional embedded vision system, consisting of an image sensor, followed by A/D converters and image processor, expends considerable energy acquiring and transmitting digital images. A vision chip, using dedicated circuits to process data at a pixel site, can significantly reduce the power and external bandwidth requirements, providing pre-processed images, and/or abstract information from the scene (e.g. location of salient objects, features, etc), instead of raw video frames.

In this work, we describe a general purpose vision chip with an array of pixel-parallel mixed-signal processing elements (PEs). Integration of a software-programmable processor into a 32µmx32µm silicon area provides greater functionality and versatility than algorithm-specific vision sensors (e.g. [1,2]) with similar cell size. At a 10MHz instruction rate, the chip achieves peak computational performance of 655GOPS (gray-scale arithmetic operations) and energy efficiency of 1.9pJ/OP (5x better than [3] and 2.5x better than [4]) in a low-cost 180nm CMOS technology. A flexible readout system enables data reduction on the focal plane. As an example, an application that discriminates and tracks an object at 100,000 fps is demonstrated.

Chip Architecture

Fig. 1 shows the block diagram of the IC. The image sensor array is tightly integrated with an array of analog/digital processing elements programmed as an SIMD computing system. A sequence of 79-bit instruction words, determines the algorithms executed by the array. Each PE incorporates a pixel circuit, 7 analog storage registers, a comparator, a squarer and a neighbor communication/diffusion network. The analog system operates in sampled-data current mode – the programmable connectivity between the storage registers (S2I memory cells), analog bus and functional nodes enables microcode-controlled computations in the analog domain [5]. Summation, inversion, subtraction, division, squaring and spatial diffusion (low pass filter), operate directly on analog data samples (e.g. gray-level pixel values) without the need for A/D conversion. The application of masks within the

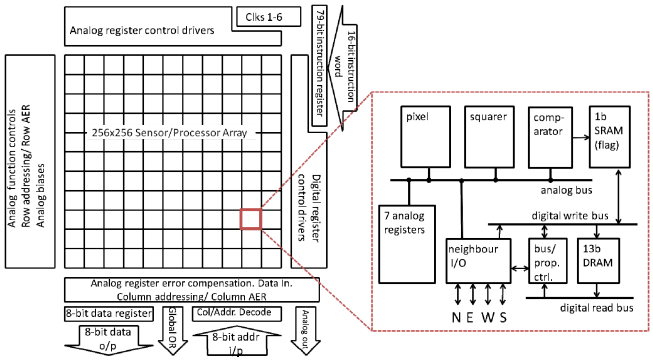

Figure 1. Architecture of the 256x256 vision chip

communication network allows arbitrary patterns of PEs to be joined, creating a macro cell within which currents can be summed and stored. A comparator circuit provides an interface between analog and digital domains. Within the digital sub-system, 13 bits of DRAM and 1 bit of SRAM are included with OR and NOT logic operations.

Readout can be in the form of a binary or 8-bit digital data, analog frames, or global image-wide data. Through controlling PE addressing, readout can locally OR binary data from multiple PEs (for digital output) or sum analog current data (analog readout). A direct readout of an ordinate of an active (foreground) pixel can be obtained from the asynchronous address extraction circuitry, providing rapid ("event-based" [2]) identification of pixels of interest. Readout-based computations are critical to operations at high frame rates, whereby processed results, and no images, are output from the device.

Operations such as flood-fill are accelerated using an asynchronous propagation network that is formed from interconnecting adjacent PE cells to allow direction-specific binary signal triggered propagation in a combination of 4 orthogonal directions.

Implementation and Results

The speed of asynchronous propagation is 3.2ns from cell to cell, or <1µs to propagate across the entire array. This allows a 62x speed-up over equivalent synchronous operations (requiring two clock cycles per pixel, with 100ns clock) to conduct image-wide commands such as hole-filling or binary object reconstruction.

While the analog datapath makes compact PEs viable, it is important to achieve acceptable levels of accuracy to ensure robust processing outcomes, competitive with conventional digital computation. For registers A and B of a PE at position i,j in the array, register B is copied to A with errors given by (1).

$$A_{i,j} = B_{i,j} + k_1 B_{i,j} + k_2 + \varepsilon_{i,j}(t) + \delta_{i,j} \qquad (1)$$

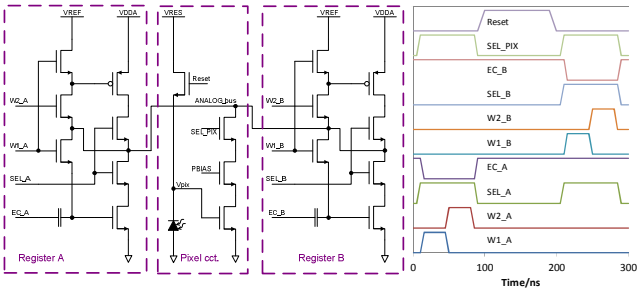

Fig. 2. Simplified pixel schematic and image capture routine. Light level is sampled by register "A" then pixel is reset. Register "B" stores sum of register "A" (inverted) and reset pixel level. (W1, W2 are write phases; SEL selects pixel or register; EC is register error correction signal, Reset sets the pixel potential Vpix to VRES).

The fixed error k_2 tends to be unimportant since a simple constant error correction can remove it, while k_1, the signal dependence of error is non-correctable. At a 10MHz clock, and considering a nominal register range of 0 to 100, k_1 is 0.07. $\varepsilon_{i,j}(t)$ indicates the random error associated with a register transfer; averaged across the array the RMS value is 0.09. $\delta_{i,j}$ (at 0.05 RMS average of array) is the error due to fixed pattern noise i.e. a constant error particular to a PE location and registers being copied. The clock can be increased beyond 10MHz up to 16MHz, at the expense of increased noise. Typically for analog memories, the registers incorporated into each PE degrade with time. This varies according to value, with a worst case rate of 2.8% per second, and typically <1% per second; with frame times of ≤50ms, this is sufficient for most applications requiring inter-frame storage, e.g. motion detection. For longer term storage, in-PE analog to digital conversions can be made, allowing data to be stored digitally within the DRAM/SRAM. Most of the 3T DRAM registers require refresh every 50ms.

Each PE includes an N-well photodiode (with fill factor 6.2%) and a current-mode pixel sensor circuit (Fig. 2). In-PE image capture generally utilizes differential double sampling. Pixel current is subtracted from the pixel-reset current. The typical overall FPN is 1.6% rms. Through dark current and leakage, the pixel value decays to half its reset level in 13s, allowing realistic integration times as long as 2s.

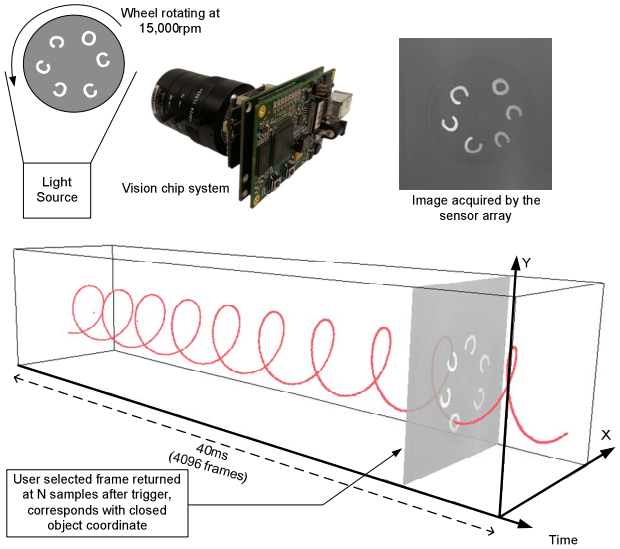

Figure 3. Object identification and tracking setup; single captured frame of the spinning wheel is shown, alongside the measured coordinates of the tracked object forming a helix on the {x,y,time} plot; object location is obtained at 10µs intervals.

Table 1. Summary of chip characteristics.

Process	0.18µm 1P6M CMOS
Chip size	10 mm x 10 mm
Power supplies	1.8V (digital), 1.5V (analog)
Resolution/No. of processors	256 x 256 / 65536
Pixel (PE) size	32.26µm x 32.26µm
PE complexity	176 transistors
Storage per PE	7 (analog) + 14 bits (digital)
Analog register variation with time	2.6%/s (max), <1%/s (typ)
Peak performance	655 GOPS @ 10MHz clock
Power consumption	1.23 W (peak), 0.2 mW (idle)
Efficiency	1.9 pJ/OP

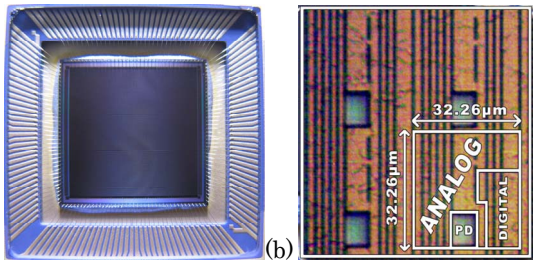

Figure 4. (a) Die photo of the chip with a single PE shown in (b)

The chip can be used in a range of embedded computer vision applications which require low power operation and/or high-speed operation.

Applications

At a 10MHz clock, Sobel edge detection requires 5.8µs, median filter 20.1µs, locally adaptive thresholding 1.2µs and skeletonization 572µs to complete. Fig. 3 shows an application in high-speed tracking. The chip detects and tracks an object (letter 'o') amongst distractors (letters 'u') placed on a disc moving at 15,000 rpm. All image processing is done on chip at a rate of 100 kfps, with the only output per frame being the coordinates of the tracked object.

Conclusions

The chip is implemented in 180nm 1P6M CMOS technology and measures 10mm x 10mm (Figure 4). Chip parameters are summarized in Table 1. Each PE contains 176 transistors. The chip operates with 1.8V digital and 1.5V analog supplies. The energy consumption of the chip is primarily related to the rate of usage of the analog registers, with typically 3 being used in any operation. At an instruction rate of 10MHz (performing 655GOPS), the power consumed is 1.23W. The quiescent consumption is 0.2mW; this state being entered by executing a NOP and removing clocks.

References

[1] T. Morie, Y. Kim, "A Subjective-Contour Generation LSI System with Expandable Pixel-Parallel Architecture for Vision Systems", ISSCC Dig. Tech. Papers, pp.478-479, Feb 2009

[2] C. Posch, D. Matolin, R. Wohlgenannt, "A QVGA 143dB Dynamic Range Asynchronous Address-Event PWM Dynamic Image Sensor with Lossless Pixel-Level Video Compression", ISSCC Dig. Tech. Papers, pp.400-401, Feb 2010

[3] W. Zhang, Q. Fu, and N. Wu, "A Programmable Vision Chip Based on Multiple Levels of Parallel Processors", IEEE Jnl.of Solid-State Circuits, vol. 46, pp. 2132-2147 (2011)

[4] C.-C. Cheng, C.-H. Lin, C.-T. Li, S. Chang, C-J. Hsu, and L-G. Chen, "iVisual: An intelligent visual sensor SoC with 2790 fps CMOS image sensor and 205GOPS/W vision processor," ISSCC Dig. Tech. Papers 2008, pp. 306–307

[5] P. Dudek and P. J. Hicks, "A CMOS General-Purpose Sampled-Data Analogue Processing Element", IEEE Transactions on Circuits and Systems - II: Analog and Digital Signal Processing, vol. 47, no. 5, pp. 467-473, May 2000