A General-purpose Vision Processor with 160x80 Pixel-Parallel SIMD Processor Array

Alexey Lopich Piotr Dudek
School of Electrical and Electronic Engineering
University of Manchester
Manchester, UK
{a.lopich, p.dudek}@manchester.ac.uk

Abstract— **In this paper we present a vision processor, which incorporates a 160×80 SIMD array of pixel-processors. The processor operates with a 100MHz clock and 1.8V supply. The device provides 640 GOPS (binary) and 23 GOPS (greyscale) consuming 0.5 W. The chip occupies 50mm^2 and is fabricated in a standard 0.18 µm CMOS process. The I/O interface supports 200 MPixels/s (greyscale), 1.6 GPixels/s (binary) and 40 MPixels/s (address-event readout) data rate, and PE-parallel image sensing mode for embedded high-speed vision applications. Experimental results indicate that the performance of the presented chip approaches the efficiency of recently reported application-specific vision processors, while providing full programmability and thus being adjustable to a wide range of applications.**

I. INTRODUCTION

Computer vision applications demand high computational power, due to the large amount of raw image data required to be processed at real-time speeds. Low- and medium-level image processing algorithms are characterised by a high degree of data parallelism and computational locality, i.e. every value in the result image is a function of the original pixel value and its bounded neighbourhood. Conventional processor architectures have to cope with massive data-flows and frequent memory access, which result in increased power consumption and performance limitations. At the same time, massively parallel processor arrays are ideally suited to regular, local computations with inherent data parallelism. Massively parallel fine-grain computing has been explored for several decades. Particular application benefits have been realised in image pre-processing, where regularity of image data and full-scale parallelism of processing algorithms naturally map onto fine-grain processor-per-pixel architectures. Many examples of such arrays have been recently presented in literature [1-6]. Typically, the architecture of such devices is based on a cellular processor array (CPA) operating according to the SIMD paradigm. Progress in CMOS fabrication technology has led to integration of image sensing and processing on a single device, a so called "vision chip". This approach eliminates the I/O data-transfer bottleneck for high frame-rate applications as sensory data is fed directly into corresponding processing elements. The output of the vision chip can be a simplified representation of the scene in the form of abstract descriptors (e.g. object coordinate, presence indicator, etc.), thus significantly increasing useful data-throughput. A number of application-specific [7-8], and general-purpose [1-6] vision chip architectures have been recently proposed. While application-specific devices often offer better performance to area/power ratio, we concentrate our efforts on programmable general-purpose chips that can be applied in a wide range of tasks.

In this paper we present a 160×80 digital vision chip that operates both in synchronous (local pixel operations) and asynchronous (global operations) modes. The presented work builds on our previous design of a vision chip a with pixel-parallel SIMD array [6]. The new device introduces further improvements in processor functionality and performance, combined with reduction in cell area and an increase in fill-factor. The chip facilitates high frame-rate processing, by enabling the output of global image descriptors (e.g. pixel coordinates, global summation, etc.), thus reducing the required output data rate down to several bytes per frame. Real-time feedback control systems can then be created based on sensed image data. Basic object segmentation and coordinate extraction takes 1.62 µs at 100 MHz operating frequency. The device is implemented in a standard 180 nm CMOS process and provides 46 GOPS/W and 650 MOPS/mm^2, approaching the efficiency of recently reported application-specific vision processors [7-8], while providing a general-purpose programmable processor architecture.

II. ARCHITECTURE

The outline of the presented chip architecture is depicted in Fig. 1. It is based on a massively parallel fine-grain cellular processor array that operates according to the SIMD paradigm, i.e. all cells receive and execute identical instructions issued by a central controller (located off chip), and operate on local data. All cells are locally interconnected with 4 nearest neighbours. The I/O interface between every PE and its four neighbouring cells consists of four 8-bit input ports for greyscale values, four 1-bit input ports for binary asynchronous propagations, one 8-bit output port for greyscale data (connects to corresponding input ports of four neighbours) and one 1-bit input and one 1-bit output port in

978-1-4673-6146-0/13/$31.00 ©2013 IEEE

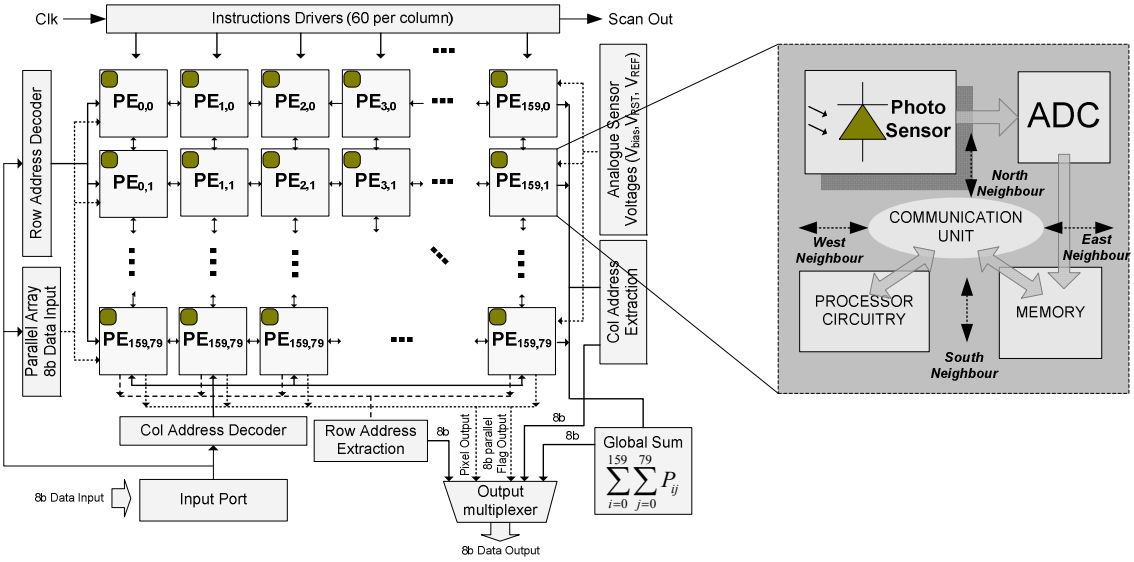

Figure 1. Block diagram of the ASPA2 architecture

the ALU for global asynchronous addition. The latter port is connected only to the east neighbour, so that global addition is enabled row-wise. Altogether, each cell has 37 digital inputs (nine from every neighbour and one sum input from the west) and 10 digital outputs. The local autonomy of each cell is provided by flags, which specify whether the cell executes an instruction or omits it. The flags can also be used to constrain the topology of the asynchronous network for continuous-time operation so that the link between neighbours is defined in each pixel individually. Instructions are latched inside the chip at a 100MHz rate and distributed by periphery drivers placed at the boundary of the array. Pixels can be addressed for I/O individually in a random-access fashion, as a block or in a flexible pattern, providing simultaneous writing to many locations and global logic operations on readout. The chip can output various types of data, which are multiplexed onto a single 8-bit output port. This port can be configured to output single 8-bit greyscale pixel value, eight binary values from 8 cells in parallel, a global sum value, X/Y pixel coordinates (enabling Address Event Representation (AER) readout mode [9]) and finally, the result of logic OR operations of selected pixels. The I/O interface supports 200 MPixels/s (greyscale) and 1.6 GPixels/s (binary) data rate.

III. PROCESSING CELL

Each processing element (PE) operates as a simple bit-serial digital microprocessor (Fig. 2). The chip is programmable and software development is based on a custom

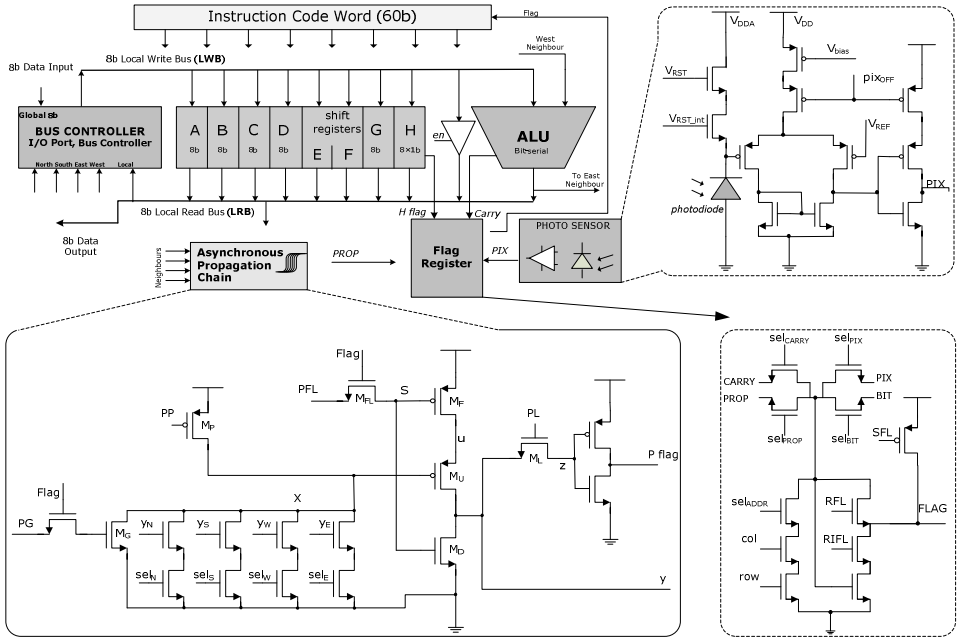

Figure 2. Schematic diagram of the processing cell with detailed schematic of photo sensor, asynchronous propagation chain and flag register.

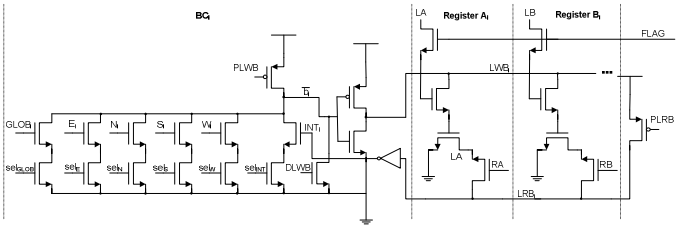

Figure 3. Bus controller and memory bit-slice.

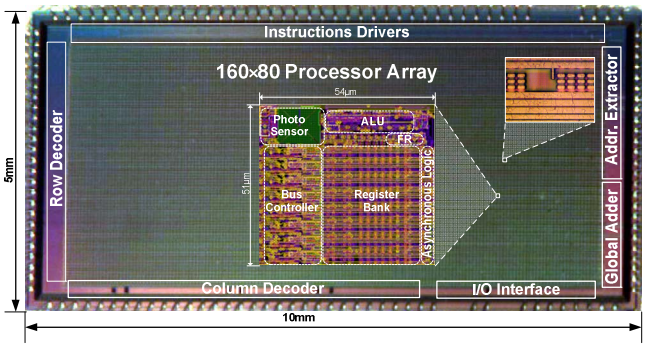

Figure 4. Chip photomicrograph, with inset showing the layout of a single PE with highlighted functional blocks and a photomicrograph of the PE.

built Assembly language. Every processor executes various instructions ranging from basic logic operations to multiplication/division and global data-flow processing. Auxiliary program flow aspects such as global variables, function calls and loops are handled by an off-chip central controller.

A. Memory and data-path

Every PE comprises eight 8b general purpose registers (GPR), two of which are shift registers. Because image pre-processing algorithms constantly deal with both greyscale and binary data, the PE was set to support fast access to binary image descriptors as well as logic and conditional operations on this data. Therefore in addition to GPRs there are also eight 1b registers, which can be used either separately for storing intermediate binary values or as a single 8b register (register *H*). The value of this register can either be transferred to the Local Read Bus (LRB) as an 8-bit greyscale value, or a selected bit can be passed as a single binary value (*H flag*) to the flag register. The binary output is the result of a logic OR operation on bits, selected in the instruction word. The total PE memory capacity is therefore 72 bits. All local memory is based on 3-transistor dynamic registers. Inputs of memory cells are connected to Local Write Bus (LWB) while outputs are connected to a precharged LRB. LRB and LWB are connected via the Bus Controller (BC), responsible for multiplexing input data (local, neighbour or global) to the LWB. The BC also performs the logic OR operation on input data. One bit-slice of the 8-bit BC and two registers is presented in Fig. 3. In the current implementation, the retention time of such dynamic registers ranges from 100 µs to 1 ms. Such a large range appears to be the result of unmatched coupling between the storage node, ground plane and instruction lines. It results in the necessity to refresh values algorithmically for temporal inter-frame processing.

Basic register-transfer operations comprise two clock cycles: precharge and transfer. During the precharge cycle the LRB and an internal node b_i (Fig. 3) in the BC are precharged so that values on the LRB and the LWB are 0xFF and 0x00 correspondingly. During the transfer cycle, the data is read from any register(s) or accumulator to the LRB, transferred through the BC to the LWB and then loaded into another register(s) or accumulator. The data appears inverted on LRB and node b_i of the BC. At the end of the transfer cycle, when the load signal (e.g. LA and LB in Fig. 3) goes to logic '0', the data is latched in the dynamic register. The communication between local neighbours is identical to local register-transfer operations and also performed within two clock cycles. Additionally, each cell contains an 8-bit parallel three-state buffer, in the form of two pass transistors per bit, which connects LWB and LRB (see Fig. 2). It is used to perform the NOT operation on binary data. This operation is bit-controlled, so it enables inversion of only specified bits.

B. ALU

The 18-transistor bit-serial ALU performs NOT, ADD, SUB, XOR and XNOR logic operations. The design of the adder is based on compact XOR/XNOR gates coupled with dynamic pull-up/pull-down transistors and a transmission gate multiplexer, resulting in an area of 3.1×8.9 µm^2. It has a single non-accumulative V_{th} loss on the CARRY output and full voltage swing on data output, which enables an unlimited number of such arithmetic cells to be used in an asynchronous addition/subtraction chain. ALUs in neighbouring cells can be chained together to form a single multi-bit ripple-carry adder. This enables asynchronous row-wise global summation. The final column-wise addition is performed at the periphery of the array and the result of the global summation is stored in a dedicated register at the periphery of the chip, whereas row-wise sub-sums are stored in the PEs in the most right column. Global addition is extensively used to gather various image statistics, e.g. pixel count, histograms, etc.

C. Data I/O

In addition to digital I/O, the chip supports in-cell sensing in each PE, so that it can operate as a 'vision chip'. The sensor is based on a n+/p-subst. photodiode followed by a simple voltage comparator (as shown in Fig. 2). The photodetector works in integration mode. The voltage across photodiode V_{PD} is compared to the reference voltage V_{ref}. The inverted value of the comparator output is transferred to the PIX input of the flag register. Based on this flag indicator a digital value that corresponds to light intensity can be written to a corresponding GPR. For greyscale conversion the binary threshold is performed at different times. The A/D conversion is set, but not limited, to 8-bit resolution. Every cell can locally control its integration time.

D. Cell layout

The microphotograph of the chip with highlighted functional block is depicted in Fig. 4. A summary of measured characteristics of the chip is provided in Table I. Compared to the previously reported chip [6], the area of the PE cell is decreased by 76% while the overall functionality of the cell

TABLE I. ASPA2 VISION CHIP CHARACTERISTICS

Parameter	Value
Fabrication process	0.18μm, 1P6M CMOS
Power supply	1.8V
Array size	160×80
PE Size	54×51μm^2
Performance	640 GOPS (binary) 23 GOPS (greyscale) 2.9 GOPS (mult/div)
Clock frequency	100MHz
Power efficiency	46 GOPS/W
Area efficiency	650 MOPS/mm^2
PE Memory	72b

has been improved. With 5 routing layers placed above active circuitry, less than 1% of the silicon space is occupied solely by routing, therefore further increase in the number of routing layers will not result in area optimization. As compared with [6], The area of the photodiode increased by 40% and the fill-factor has improved from ~2% to 5.6%. The "complexity" of the cell (number of transistors) has also increased by 27%.

IV. PERFORMANCE

Because the chip has a software-controlled architecture it is capable of executing a wide range of image processing algorithms. The support of logic and arithmetic operations in every cell enables straightforward implementation of basic convolutions and filters. Examples of pre-processing algorithms implemented on the fabricated device are illustrated in Fig. 5 (a)-(c).

A significant number of useful morphological image processing algorithms deal with images represented in binary format (binary dilation, erosion, skeletonization). The chip can execute many of these tasks in an efficient way, as some of the logic operations are performed on data buses during data transfers, i.e. at no additional cost. In addition to local operation, the support for global binary trigger-wave propagations enables fast geodesic reconstructions and hole filling, which can be a part of more complex algorithms. An example of using these operations for object segmentation and coordinate extraction is presented in Fig. 5 (d)-(i).

The performance of the chip on a set of several common image processing tasks is presented in Table II.

TABLE II. ASPA2 PROCESSING PERFORMANCE

Operation	Frame processing time @ 100 MHz
Sobel Edge Detection	5.2 μs
3×3 Erosion(Dilation)	60 ns
3×3 Median Filter	9.3 μs
Frame min/max value	1.1 μs
Binary object reconstruction	180 ns
Closed contour detection	200 ns
Object bounding box extraction (4 coordinates)	1.42 μs
Output rate	greyscale: 200 MPixel/s binary: 1.6 GPixel/s [x,y] pair: 40 MPixel/s

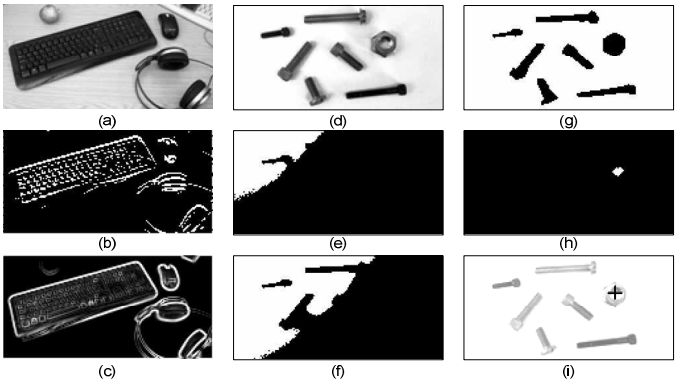

Figure 5. Image processing in ASPA2: a – original; b – high pass filter with threshold; c – sobel edge detection; d – original; e-g – asynchronous propagation triggered in the top left corner; h – closed curve detection; i – coordinate extraction of the closed area (nut).

V. CONCLUSIONS

In this paper we presented a general-purpose vision chip, based on fine grain cellular processor-per-pixel array that operates in the focal-plane. The 160×80 chip is fabricated in 0.18 μm standard CMOS technology and measures 10×5 mm^2. The overall architecture is scalable to larger dimensions. The chip operates at 100MHz with V_{DD} 1.8V. At a 100MHz clock, each cell provides 50 MOPS (binary), 1.8 MOPS (greyscale) and 0.23 MOPS (unsigned products and quotients). The chip is suitable for low- and medium-level image processing and its main application areas include high-speed industrial inspection, robotics and control systems.

ACKNOWLEDGMENT

This work has been supported by the EPSRC under UK grant no EP/D029759/1.

REFERENCES

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

[2] A. A. Abbo, R. P. Kleihorst, et,. al., "Xetal-II: A 107 GOPS, 600 mW Massively Parallel Processor for Video Scene Analysis," IEEE Journal of Solid-State Circuits, vol. 43, pp. 192-201, 2008

[3] D. Ginhac, J. Dubois, M. Paindavoine, and B. Heyrman, "A high speed programmable focal-plane SIMD vision chip", Analog Integrated Circuits and Signal Processing, vol 65 (3), pp. 389-398, 2010.

[4] C.-C. Cheng, C.-H. Lin, C.-T. Li, et.al.: "iVisual: An Intelligent Visual Sensor SoC with 2790fps CMOS Image Sensor and 205GOPS/W Vision Processor" in ISSCC Dig.Tech.Papers, pp. 306-615, 2008

[5] S.J.Carey, A.Lopich, D.R.W.Barr, P.Dudek, "A 100,000 fps Vision Sensor with Embedded 535 GOPS/W 256x256 SIMD Processor Array", VLSI Circuits Symposium, Kyoto, 12-14 June 2013

[6] A. Lopich and P. Dudek, "A SIMD Cellular Processor Array Vision Chip With Asynchronous Processing Capabilities," IEEE Transactions on Circuits and Systems I, vol 58 (10), pp. 2420-2431, 2011.

[7] Jinwook Oh, Junyoung Park, Gyeonghoon Kim, Seungjin Lee, et.al, "A 57mW Embedded Mixed-Mode Neuro-Fuzzy Accelerator for Intelligent Multi-core Processor", ISSCC Dig.Tech.Papers, pp. 130-131, 2011

[8] Jae-Sung Yoon, et.al. " A Graphics and Vision Unified Processor with 0.89μW/fps Pose Estimation Engine for Augmented Reality", ISSCC Dig.Tech.Papers, pp. 336-337, 2010

[9] T. Serrano-Gotarredona, and B.Linares-Barranco, "A 128x128 1.5% Contrast Sensitivity 0.9% FPN 3 μs Latency 4 mW Asynchronous Frame-Free Dynamic Vision Sensor Using Transimpedance Preamplifiers", IEEE J. Solid-State Circuits, vol. 48, pp. 827-838, 2013