7.3 A 1000fps Vision Chip Based on a Dynamically Reconfigurable Hybrid Architecture Comprising a PE Array and Self-Organizing Map Neural Network

Cong Shi[1,2], Jie Yang[1], Ye Han[1], Zhongxiang Cao[1], Qi Qin[1], Liyuan Liu[1], Nan-Jian Wu[1], Zhihua Wang[2]

[1]Chinese Academy of Sciences, Beijing, China,
[2]Tsinghua University, Beijing, China

A vision chip is a high-speed and compact vision system that integrates an image sensor and parallel image processors on a single silicon die. Nowadays, high-speed vision chips with powerful recognition capabilities are greatly demanded in applications such as: industrial automation, security, entertainment, robotic vision, and human-machine interaction. Some 100-to-1,000fps vision chips have been reported [1-4]. These chips integrate pixel-parallel and row-parallel SIMD array processors to speed up low- and mid-level image processing [1,2]. Recently, microprocessors (MPU) have been embedded to carry out high-level image processing [3,4]. Although excellent in low- and mid-level processing, these systems are poor in high-level feature vector (FV) recognition tasks due to the von Neumann bottleneck of the MPU. As a consequence, these chips can no longer achieve 1,000fps system-level performance, from image acquisition to high-level feature-recognition processing.

This paper reports on a 1,000fps vision chip based on a dynamically reconfigurable hybrid architecture. It integrates a high-speed image sensor, von Neumann-type pixel-parallel and row-parallel array processors, and a non-von Neumann-type Self-Organizing Map (SOM) neural network. The SOM network can significantly speed up the high-level image processing in a vector-parallel fashion [5]. Furthermore, the SOM network can be dynamically reconfigured from the pixel-parallel array processor at negligible costs of ~2% total chip area and 3 clock cycle latency. The reconfiguration technique extends the capability of conventional pixel-parallel array processors from low-level to high-level processing, and avoids significant chip area budget for a standalone SOM network. The chip can achieve 1,000fps system-level performance even when complicated high-level recognition tasks are involved.

Figure 7.3.1 shows the system architecture of the vision chip. It mainly consists of a 256×256 4T-APS rolling-shutter pixel array, a 256 10b cyclic-ADC array with CDS and PGA circuits, a 64×64 pixel-parallel 1b processing element (PE) array processor, a 64 row-parallel 8b row processor (RP) array, a thread-parallel dual-core 32b RISC MPU, and a 16×16 vector-parallel SOM neural network. The pixel-parallel PE array processor can be dynamically reconfigured as the SOM network for high-level image processing, and then back as the PE array for low-level image processing, as demand requires. This reconfiguration method is enabled by the similar mesh grid topology between the two components. As the PE array is smaller than the pixel array in size, dynamically mapping relationships with different pixel sampling intervals can be established between the two arrays (Fig. 7.3.1, bottom). The PE and RP array processors perform low- and mid-level processing, such as image filtering, image segmentation, mathematical morphology, and FV extraction. The SOM network recognizes FVs with a speedup of 16×16, avoiding the serious bottleneck in the high-level recognition processing. The dual-core MPU performs other simpler non-recognition high-level processing tasks, as well as the overall chip management.

The PE array processor can be divided into 16×16 sub-arrays. Each sub-array contains 4×4 PEs and constitutes one neuron along with a small-size condition generator (CG). Figure 7.3.2 illustrates the dynamic reconfiguration scheme from the PE array processor to the SOM network, and vice versa. If R-signal is logic low, the solid paths among the PEs are activated to form a 2D-meshed 64×64 pixel-parallel PE array processor. Otherwise, the dashed paths among the PEs are activated to reconfigure each sub-array with CG into one SOM neuron. Thus the 64×64 PE array processor is reconfigured into the SOM network with 16×16 neurons. The 16 1b PEs in one neuron are chained in a snaky style to form a 16b processing engine. Each PE has a unique bit-position (bp) in the neuron. The CG is dedicated for conditional operations involved in the SOM network training and recognition procedures. The dynamic reconfiguration can be completed in 3 clock cycles by switching the topological connections among the PEs and the CG.

Figure 7.3.3 shows the PE and RP circuit schematics. The reconfiguration multiplexers (gray colored) in the PE circuit can switch the topological connections between neighboring PEs for different modes. For PE array processor mode, the PE circuit can not only accomplish single-bit operations but also multiple-bit operations by bit-serial processing. For SOM network mode, the 1b ALUs and 1b-wide memories of the 16 PEs in one neuron equivalently constitute a 16b ALU and a 16b-wide memory. The inter-PE connections of *LL*, *LH, AH* support I/O operations and multiplier-free multiplication/division operations based on a bit-shifting method for the SOM neuron. The RP processor contains an 8b ALU and can support some nonlinear and *skip* chain operations.

The SOM network is trained by an online LVQ training procedure [5]. First, the random reference vectors (RV) are inputted to all of the neurons. After a sample FV with known class label is broadcasted to the SOM network, the Manhattan distances between the FV and RVs are computed simultaneously in all the neurons. Then the RP array processor determines the neuron with the minimum distance as the winner, which represents the recognized class of the FV. Finally, RVs of the neurons within the winner neighborhood are updated based on the consistency of the recognized class with the known class. After the network is trained, it can be used to recognize the FV of the real-time sensor image by the above procedure without the RV updating stage.

The vision chip is fabricated in a 0.18μm 1P5M CIS process. Figure 7.3.4 shows the experimental results of >1,000fps hand gesture recognition and face recognition based on 16D and 32D PPED [6] FVs. The high processing speed and ~86% recognition accuracy for hand gesture and face recognition can facilitate many natural human-machine interactions. Figure 7.3.5 compares the image-recognition speed of the SOM network with that of the 32b dual-core MPU at 50MHz on the fabricated chip. The SOM network reduces the processing time by >98% for the hand gesture and face recognitions. The SOM network achieves 45.4 and 31.6 kilo recognitions per second (KRPS) for 16D and 32D vectors, respectively. While the speed of dual-core MPU is only 0.84 and 0.46 KRPS, respectively. The performance improvement of the SOM network over the dual-core MPU for recognition tasks becomes more impressive with the increase of the FV dimension.

Figure 7.3.6 compares the vision chip with the state of the art. Our chip avoids the high-level image-processing bottleneck in previous chips and makes the high-level processing speed well matched with that of the low- and mid-level processing. It exhibits the best system-level performance of >1000fps from image acquisition to high-level feature-recognition processing. Figure 7.3.7 shows the chip micrograph and summarizes the chip specifications.

Acknowledgement:
The authors would like to thank Quanliang Li, Zhe Chen and Yongxing Yang for their technical suggestions. This work was supported by the National Natural Science Foundation of China (Grant No. 61234003), and Special Funds for Major State Basic Research Project of China (No. 2011CB932902).

References:
[1] W. Miao, *et al.*, "A Programmable SIMD Vision Chip for Real-Time Vision Applications", *IEEE J. Solid-State Circuits*, vol. 43, no. 6, pp. 1470-1479, June 2008.
[2] W. Jendernalik, *et al.*, "An Analog Sub-Miliwatt CMOS Image Sensor with Pixel-Level Convolution Processing", *IEEE Trans. Circuits and Systems—I: Regular Papers*, vol. 60, no. 2, pp. 279-289, Feb. 2013.
[3] C. Cheng, *et al.*, "iVisual: An Intelligent Visual Sensor SoC with 2790fps CMOS Image Sensor and 205GOPS/W Vision Processor", *ISSCC Dig. Tech. Papers*, pp. 306-307, Feb. 2008.
[4] W. Zhang, *et al.*, "A Programmable Vision Chip Based on Multiple Levels of Parallel Processors", *IEEE J. Solid-State Circuits*, vol. 46, no. 9, pp. 2132-2147, Sept. 2011.
[5] D. C. Hendry, *et al.*, "IP Core Implementation of a Self-Organizing Neural Network", *IEEE Trans. Neural Networks*, vol. 14, no. 5, pp. 1085-1096, Sept. 2003.
[6] H. Yamasaki, *et al.*, "A Real-Time Image-Feature-Extraction and Vector-Generation VLSI Employing Arrayed-Shift-Register Architecture", *IEEE J. Solid-State Circuits*, vol. 42, no. 9, pp. 2046-2053, Sept. 2007.

 978-1-4799-0920-9/14/$31.00

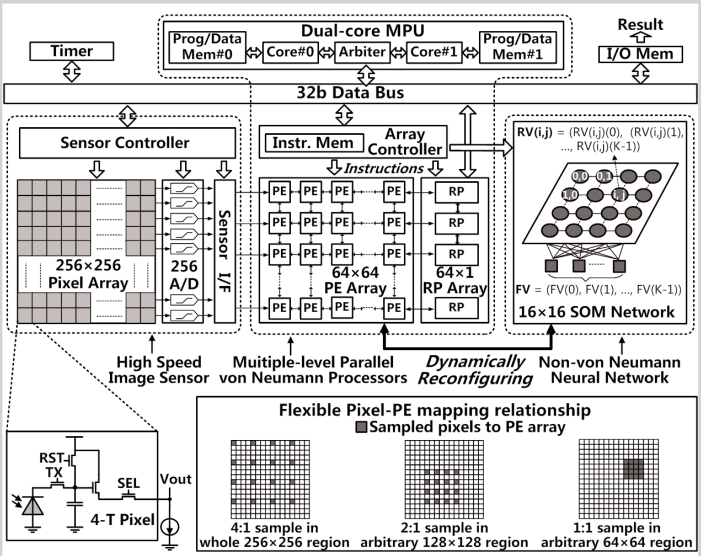


Figure 7.3.1: Vision chip architecture.

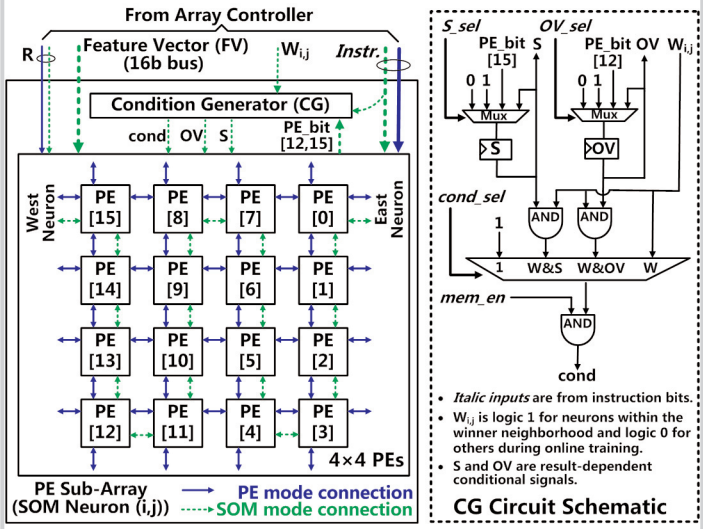
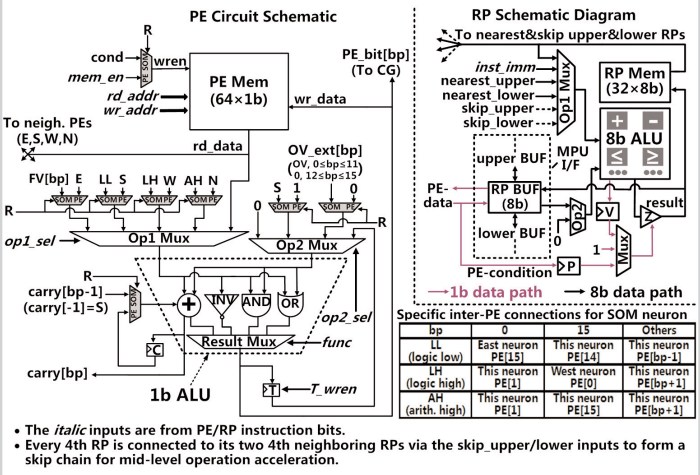


Figure 7.3.2: The reconfiguration between PE array and SOM network.



- The *italic* inputs are from PE/RP instruction bits.
- Every 4th RP is connected to its two 4th neighboring RPs via the skip_upper/lower inputs to form a skip chain for mid-level operation acceleration.

Figure 7.3.3: The PE and RP circuit schematics.

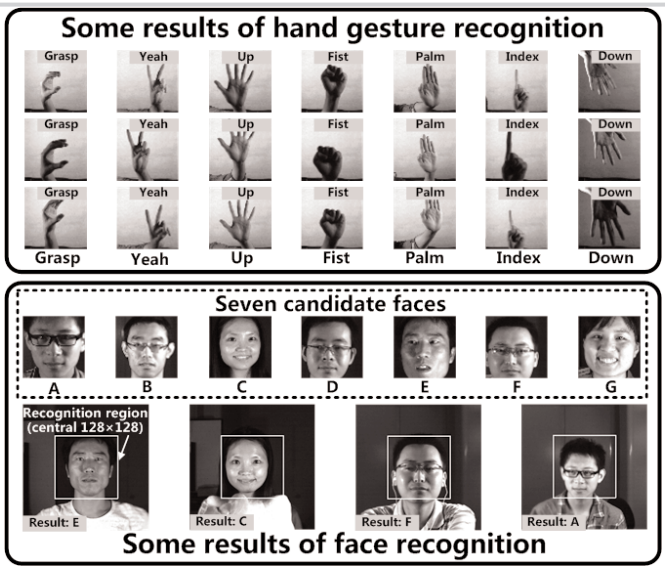


Figure 7.3.4: Experimental results of the fabricated vision chip.

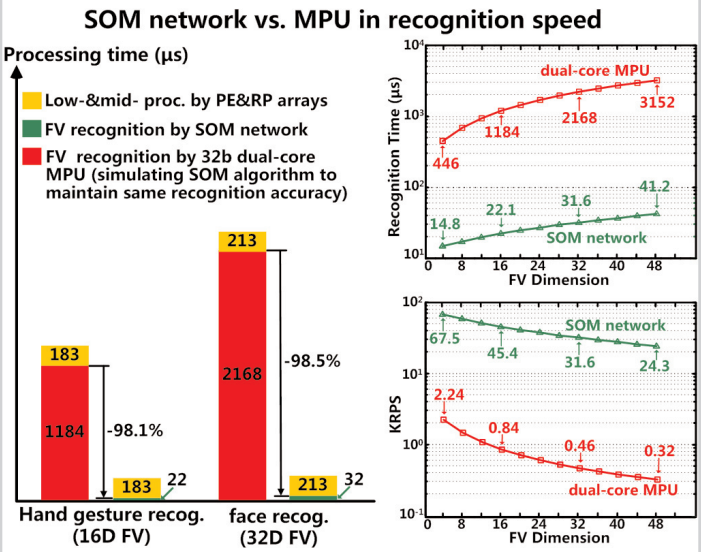


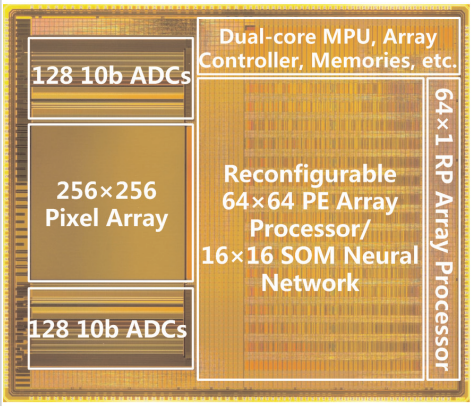
Figure 7.3.5: Performance comparison: SOM network vs. MPU.

		This work	[1]	[2]	[3]	[4]
Technology		0.18μm 1P5M	0.18μm 1P6M	0.35μm 2P4M	0.18μm 2P4M	0.18μm 1P6M
Chip area		82.3mm²	2.3mm²	9.8mm²	70.5mm²	13.5mm²
Power consumption		630mW	8.72mW	0.28mW	455mW	450mW
Clock frequency		50MHz	20MHz	N/A	50MHz	100MHz
FOM[1] (with 16D FV)		0.1834	~0	~0	0.0380	0.0273
FOM (with 32D FV)		0.1675	~0	~0	0.0192	0.0137
High speed image sensor	Sensor resolution	256×256	16×16	64×64	128×128	128×128
	Pixel fill factor	60%	3%	23%	N/A	58%
	Sensitivity	8.1V/Lux·s	N/A	1.4V/Lux·s	N/A	N/A
	Dynamic Range	48dB	N/A	58dB	N/A	N/A
	ADC resolution	10b	1b	N/A	8b	8b
Parallel image processors	Parallelism	Pixel-, row-, thread-, vector-	Pixel-, row-	Pixel-	Row-	Pixel-, row-
	Processor reconfigurability	Dynamically between PE array processor and SOM network	No	No	No	Statically among different-grained PE array processors
	Neural network	16×16 SOM network	No	No	No	No
	MPU	32b dual-core	No	No	32b single-core	8b single-core
	Low-level proc.	Fast	No	Moderate	Moderate	Fast
	Mid-level proc.	Fast	Fast	No	Fast	Moderate
	High-level proc.	Fast	No	No	Slow	Slow
	Performance in low-,mid-levels	12GOPS	0.2GOPS	0.1ms, 3x3conv. (~0.45GOPS)	76.8GOPS	44GOPS
	Performance in high-level recog.	45.4KRPS[2]@16D 31.6KRPS@32D	0	0	1.24KRPS@16D 0.60KRPS@32D	0.16KRPS@16D 0.08KRPS@32D
	system-level frame rate	1340fps@face recognition	N/A	N/A	360fps@posture recognition	76fps@face recognition

[1]FOM = (GOPS^-1+KRPS^-1)^-1/Area(mm²)·Power(W).

[2]KRPS = kilo recognitions per second.

Figure 7.3.6: Comparison with the state of the art.



Technology	0.18μm 1P5M CIS
Chip area (pad incl.)	9.8mm×8.4mm
Package	BGA 312pin
clock frequency	50MHz
Supply voltages	1.8V for digital and ADC 3.3V for I/O and pixel
Power consumption	630mW
Sensor resolution	256×256
Max. sensor frame rate	1500fps@256×256 3000fps@128×128 6000fps@64×64
Pixel size	10μm×10μm
Pixel fill factor	60%
Pixel Sensitivity	8.13V/Lux·s
Pixel Dynamic Range	48dB
ADC resolution	10b
PE array size	64×64
Memory per PE	64×1b
RP array size	64×1
SOM network size	16×16 neurons
Count of MPU cores	2
Processor performance	PE array processor: 12GOPS @8b gray img 108GOPS @1b binary img RP array processor: 1.6GOPS @8b data Dual-core MPU: ~0.1GOPS @32b data SOM network: 45.4KRPS* @16D vector 31.2KRPS* @32D vector

Figure 7.3.7: Chip micrograph and specifications.

978-1-4799-0920-9/14/$31.00