8.3 A 201.4GOPS 496mW Real-Time Multi-Object Recognition Processor with Bio-Inspired Neural Perception Engine

Joo-Young Kim, Minsu Kim, Seungjin Lee, Jinwook Oh, Kwanho Kim, Sejong Oh, Jeong-Ho Woo, Donghyun Kim, Hoi-Jun Yoo

KAIST, Daejeon, Korea

The visual attention mechanism, which is the way humans perform object recognition [1], was applied to the implementation of a high performance object recognition chip [2]. Even though the previous chip achieved 50% gain of computational cost [2], it could recognize only one object in a frame so that it is not suitable for advanced multi-object recognition applications such as video surveillance, intelligent robots, and autonomous vehicle navigation [3].

A real-time multi-object recognition processor is presented based on the bio-inspired visual perception algorithm. The proposed recognition processor has 4 features: 1) 3-stage pipelining with grid-based region-of-interest (ROI) processing for high recognition rate, 2) Neural perception engine (NPE) with three bio-inspired neural and fuzzy processing units for multi-object perception and segmentation, 3) Low latency multi-castable Network-on-Chip (NoC) for high bandwidth integration platform, and 4) Workload-aware power management for low power consumption.

Figure 8.3.1 shows the overall block diagram of the proposed processor. It is composed of 21 IPs on a chip: a NPE, a SPU task/power manager (STM), 16 SIMD processor units (SPUs), a decision processor (DP), and 2 external memory interfaces. The bio-inspired NPE is composed of motion estimator (ME), visual attention engine 2 (VAE2), and object segmentation engine (OSE). It performs global feature extraction and object segmentation using the neural and fuzzy processing to extract ROIs. The 16 SPUs perform complex and data-intensive image processing for the selected ROIs. The detailed block diagram of the SPU is shown in Fig. 8.3.2. Each SPU consists of eight 16b SIMD processing elements (PEs), 1 scalar datapath, 12KB 128b wide data memory with 2 aligners, and 2D DMA. Dual-issue VLIW enables parallel execution of data processing and data transfer operations. A register file with 5-read and 3-write ports enables PE to execute 2-way 8b multiply-and-accumulate, 3-operand 16b min/max compare, and 32b accumulation in a single cycle. For low power consumption, the 16 SPUs are divided into 4 voltage/frequency domains called SPU cluster (SPC). The STM dynamically assigns ROI tasks to 16 SPUs, and controls 4 SPC power domains. The DP recognizes each object using the database search for the generated descriptor vectors by the SPUs.

Figure 8.3.3 shows the 3-stage multi-object recognition with grid-based ROI processing. It is composed of: 1) visual perception, 2) descriptor vector generation, and 3) object decision. The visual perception stage classifies the boundaries of the multiple objects based on the extracted static and dynamic features from the input images. It extracts the ROIs for each object in a 40×40 pixel tile. Extraction of the ROIs in the visual perception stage reduces the workload of the following stages by focusing their operations on only the extracted ROIs. The descriptor vector generation stage calculates descriptor vectors for the selected ROIs. Then, in the object decision stage, the descriptor vectors of the objects are recognized through iterative matching with the database. In the proposed processor, the 3 stages of the object recognition pipeline are directly mapped to the NPE, 16 SPUs, and DP, respectively. The STM controls the processing speed of 16 SPUs according to the workloads coming from the NPE to match the processing time of each stage of the pipeline. As a result, task pipelining with the grid-based ROI processing reduces computation area by 41%, and achieves a 3.8× performance improvement compared to the previous serial object recognition based on column-wise processing [2].

The NPE consists of a 32b RISC processor, cellular neural network based VAE2 [4], ME, OSE, and 24KB shared memory. After the VAE2 and ME generate a saliency map from the 1/8 down-sized 80×60 input images, the OSE finally determines the ROIs for each object by selecting the 10 most salient points and by growing the regions starting from the selected 10 seeds. For robust and human-like object segmentation, the OSE has 4 of the Gaussian fuzzy membership and 4 of the ADALINE neural networks [5]. Each Gaussian fuzzy membership has the 3 evaluation inputs of intensity, saliency, and distance as shown in Fig. 8.3.4. The bell-shaped Gaussian curve, employed to measure the similarities between the seed and target pixel, is simply implemented by the analog differential pair and minimum follower circuit. As shown in measured waveforms of Fig. 8.3.4, the operation of the segmentation processing for 1 pixel is divided into 3 steps. First, 2 digital inputs for the Gaussian function circuit are converted to analog signals. Then, the similarity between the 2 signals is measured by Gaussian function circuit, and, finally, the neural network merges the 3 evaluated similarities using synaptic circuits, and makes the final decision whether the target pixel is joined to the region or not. With the ADALINE learning, the weights are updated to adjust the criteria for the final homogeneity decision. As a result, transistor level analog implementation of Gaussian function circuits and neural synaptic multipliers reduces area by 59% and power consumption by 44% compared to those of a digital implementation, and ROIs for 1 object can be obtained in 7μs at 200MHz.

Figure 8.3.5 shows the NoC architecture and its multi-castable crossbar switch. A hierarchical star topology [2] integrates all 21 IPs with low latency interconnection. To reduce the switch hopping latency between neighboring SPCs, the ring connection links 4 SPC local networks. Multi-casting is useful for program kernel or input image distribution to the 16 SPUs. To remove redundant data transactions, multi-castable ports in crossbar switches reserve the desirable output paths until they gather all grants from the requested arbiters. The NoC utilizes a 400MHz operating frequency, which is twice that of the IP clock, with the heterogeneous clock domain converter [6] in each input port. As a result, the NoC guarantees 4-cycle latency for one switch hopping and provides 118.4GB/s aggregated bandwidth.

The STM performs workload-aware power domain management and IP-level clock gating as shown in Fig. 8.3.6 to reduce the power consumption of the 16 SPUs. The STM determines the number of activated SPC power domains by measuring the per-frame workload from the NPE, and schedules ROI tasks to the activated SPCs. In the activated power domains, the clock of each SPU is gated by a software request for a further reduction of SPU dynamic power. Through the domain management and the clock gating, the power dissipation of 16 SPUs is reduced by 38% at a 60fps sustained frame rate.

Figure 8.3.7 shows the chip micrograph and summarizes its features. It is fabricated in 0.13μm CMOS technology and occupies 49mm^2 containing 36.4M transistors with 3.73M gates and 396KB on-chip SRAM. The 1.2V processor achieves 60fps object recognition for a maximum of 10 objects with 496mW power consumption at 200MHz IP clock and 400MHz NoC clock frequency. Its 290GOPS/W power efficiency is the highest among the previously reported parallel processors [2, 7, 8] as shown in Fig. 8.3.7.

References:

[1] M.I. Posner and S.E. Petersen, “The attention system in human brain,” *Annual Review of Neuroscience*, vol. 13, pp. 25-42, 1990.

[2] Kwanho Kim, et al., “A 125GOPS 583mW Network-on-Chip Based Parallel Processor with Bio-inspired Visual Attention Engine,” *ISSCC Dig. Tech. Papers*, pp. 308-309, Feb. 2008.

[3] Yutaka Hirano, et al., “Industry and Object Recognition: Applications, Applied Research and Challenges,” LNCS 4170, pp. 49-64, 2006.

[4] Seungjin Lee, et al., “The Brain Mimicking Visual Attention Engine: An 80x60 Digital Cellular Neural Network for Rapid Global Feature Extraction,” *IEEE Symp. VLSI circuits*, pp. 26-27, 2008.

[5] Bernard Widrow, et al., “Layered neural networks for pattern recognition,” *IEEE Trans. Acoustics, Speech and Signal Processing*, vol. 36, pp. 1109-1118, July 1988.

[6] Jakov N. Seizovic, “Pipeline synchronization,” *Proc. Of International Symposium on Advanced Research in Asynchronous Circuits and Systems*, pp. 87–96, Nov. 1994.

[7] B. Khailany, et al., “A Programmable 512 GOPS Stream Processor for Signal, Image, and Video Processing,” *ISSCC Dig. Tech. Papers*, pp. 272-273, Feb. 2007.

[8] Sumito Arakawa, et al., “A 512GOPS Fully-Programmable Digital Image Processor with full HD 1080p Processing Capabilities,” *ISSCC Dig. Tech. Papers*, pp. 312-313, Feb. 2008.

978-1-4244-3457-2/09/$25.00 ©2009 IEEE

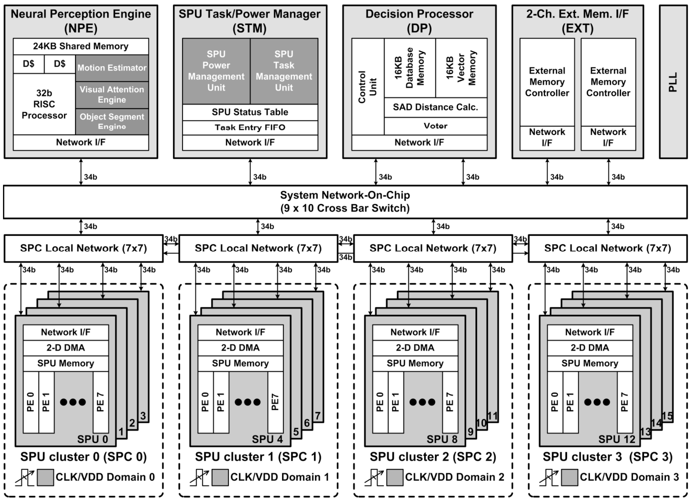


Figure 8.3.1: Overall block diagram of the real-time multi-object recognition processor.

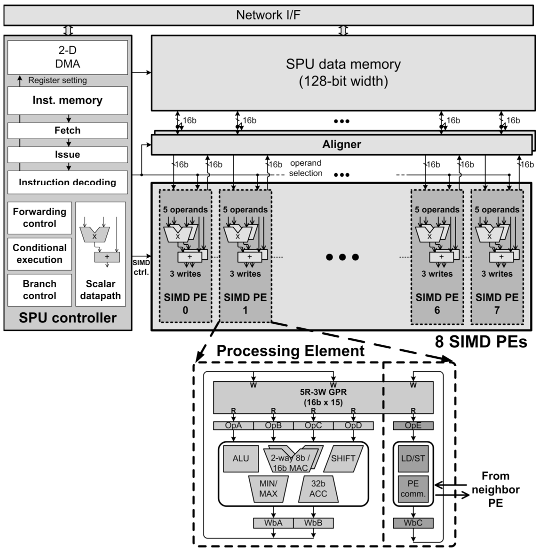


Figure 8.3.2: Block diagram of SIMD Processing Unit (SPU).

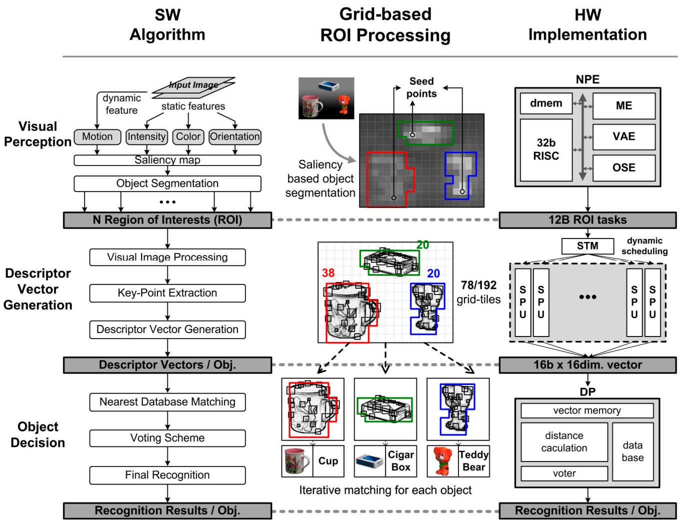


Figure 8.3.3: 3-stage multi-object recognition with grid-based ROI processing.

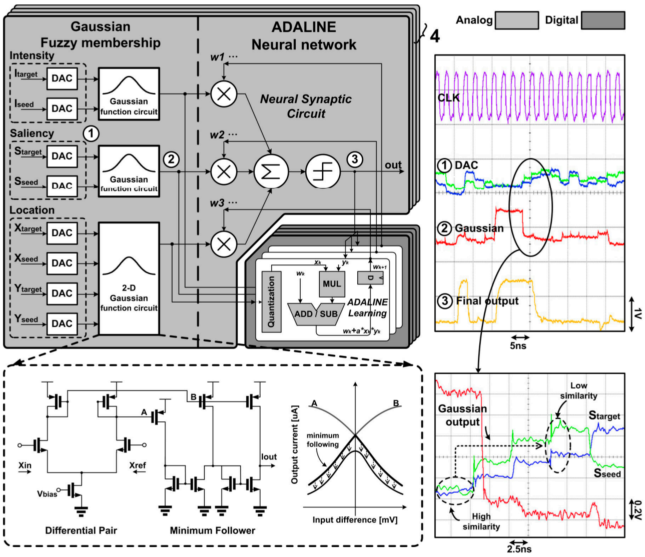


Figure 8.3.4: Analog-digital mixed OSE and measured waveforms.

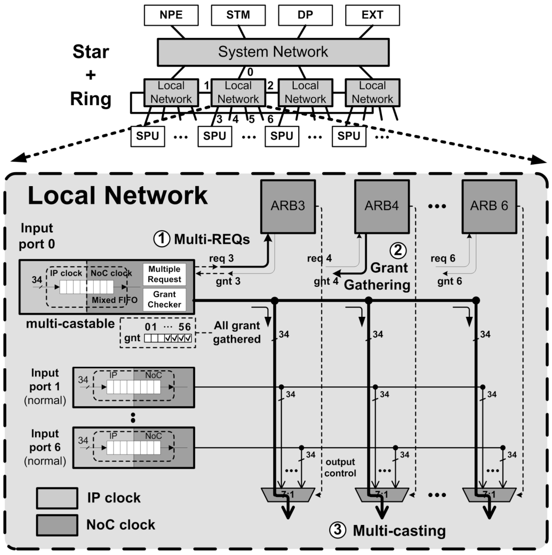


Figure 8.3.5: Proposed NoC architecture and multi-castable crossbar switch.

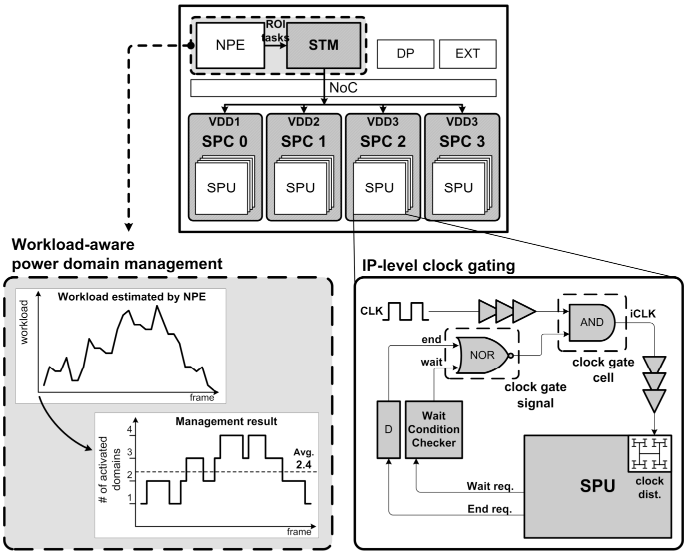
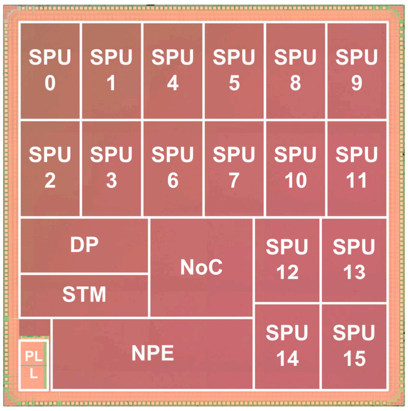


Figure 8.3.6: Workload-aware power domain management and IP-level clock gating.



Technology	0.13mm 1P 8M CMOS	
Die Size	7mm x 7mm	
Power supply	1.2V core, 2.5 I/O	
Operating freq.	400MHz(45FO4) for NoC / 200MHz (90FO4) for IPs	
# of transistors (gates, memory)	36.4M transistors 3.73M gates / 396KB SRAM	
Power consumption	**496mW** (full application running)	
	695mW (peak)	
Peak Performance	16 SPUs	128 GOPS
	NPE	54 GOPS
	DP	19.4 GOPS
	Total	**201.4 GOPS**
Power efficiency	**290 GOPS/W**	
Recognition rate	**60 fps**	
Multi-objects captured	**Up to 10**	
Input image	VGA(640x480 pixels) video image	

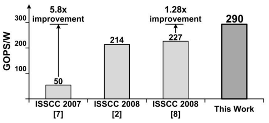


Figure 8.3.7: Chip micrograph and summary.