An 81.6 GOPS Object Recognition Processor Based on NoC and Visual Image Processing Memory

Donghyun Kim, Kwanho Kim, Joo-Young Kim, Seungjin Lee and Hoi-Jun Yoo
Semiconductor System Laboratory, Department of Electronic Engineering and Computer Science
Korea Advanced Institute of Science and Technology (KAIST)
Daejeon, Republic of Korea
donghyun53@eeinfo.kaist.ac.kr

Abstract–**An 81.6 GOPS object recognition processor is developed by using NoC and Visual Image Processing (VIP) memory. SIFT (Scale Invariant Feature Transform) object recognition requires huge computing power and data transactions among tasks. The chip integrates 10 SIMD PEs for data/task level parallelism while the NoC facilitates inter-PE communications. The VIP memory searches local maximum pixel inside a 3x3 window in a single cycle providing 65.6 GOPS. The proposed processor achieves 15.9fps SIFT feature extraction at 200 MHz.**

I. INTRODUCTION

Recently, many recognition processors for vehicle or face detection have been reported. IMAP-CE [1] and Vision Instruction Processor [2] implemented around 50 GOPS recognition performance by exploiting data level parallelism in 2D image data. Even though these processors integrated many number of Processing Elements (PEs) for high computing power, their architecture is not suitable for task level parallelism. Because the PEs have tightly coupled pipeline tailored for a single main processor, simultaneous execution of multiple independent task is not supported. On the other hand, Visconti [3] supports task level parallelism as well as data level parallelism based on 3 VLIW processors. However, the number of processors is too small to provide sufficient computing power for more sophisticated object recognition than the vehicle or obstacle detection. In addition, large amount of data transactions among tasks are not supported efficiently because of bus based interconnection architecture.

Compared to the vehicular applications, object recognition in intelligent robot is more complicated and Scale Invariant Feature Transform (SIFT) [4] based object recognition is widely used [5, 6]. The SIFT can be efficiently computed by exploiting 2 levels of parallelism. The one is data level parallelism in the process of 2D image filtering. And the other is task level parallelism which is applicable to descriptor vector generation and repeated Gaussian filtering with varying coefficients. In contrast to the vehicular algorithms, large amount of intermediate data transactions among tasks are the other distinguishing feature of the SIFT object recognition.

In addition to high computing power, low power consumption is especially important for intelligent robots due to their limited power supply. In the perspective of power consumption, conventional processor is not appropriate because it dissipates ~13W, which is more than 50% of total robot power [7]. Therefore, dedicated object recognition

processor which provides high GOPS/W is demanded for intelligent robot application.

In this paper, we propose an 81.6 GOPS object recognition processor based on Visual Image Processing (VIP) memory and Network-on-Chip (NoC). The proposed processor comprises 10 SIMD processing elements (PEs) for data and task level parallelism. Data transactions between PEs are efficiently supported by hierarchical star topology NoC [8] and 8 VIP memories. The benefit of the proposed architecture is low overhead inter-PE communication to facilitate PE level task pipelining. In addition, active utilization of the VIP memory which accelerates key-point localization stage of the SIFT leads to 65.6 GOPS performance gain.

The rest of the paper is organized as follows. In section 2, important tasks of the SIFT computation and their characteristics are described. Then, section 3 details the proposed processor architecture and its operation. Performance evaluation and implementation results will be shown in section 4. Finally, summary and conclusion of the paper is made in section 5.

II. SIFT BASED OBJECT RECOGNITION

SIFT is the process of generating descriptor vectors from input image data. These vectors represent features of object and used by subsequent classifiers to perform object recognition [4].

Fig. 1 shows the overall flow of the SIFT computation which is divided into (a) key-point localization and (b) descriptor vector generation stages. For key-point localization, Gaussian filtering with varying coefficients is repeatedly performed on the input image. Then subtractions among the filtered images are carried out to yield Difference of Gaussian (DoG) images. Locations of key-points are decided by finding local maximum pixels using a 3x3 search window over entire DoG images. The pixels holding local maximum value greater than given threshold become key-points. The number of key-points detected in a single input image is a few hundreds. In this case, it is advantageous to use both data and task level parallelism for fast key-point localization. SIMD data level parallelism is applicable to each Gaussian filtering while task level parallelism is appropriate to reduce processing time of numerous filtered image calculations. Large data transactions occur in the subsequent tasks because each DoG and local maximum pixel search task requires the results from multiple previous tasks as shown in Fig. 1 (a).
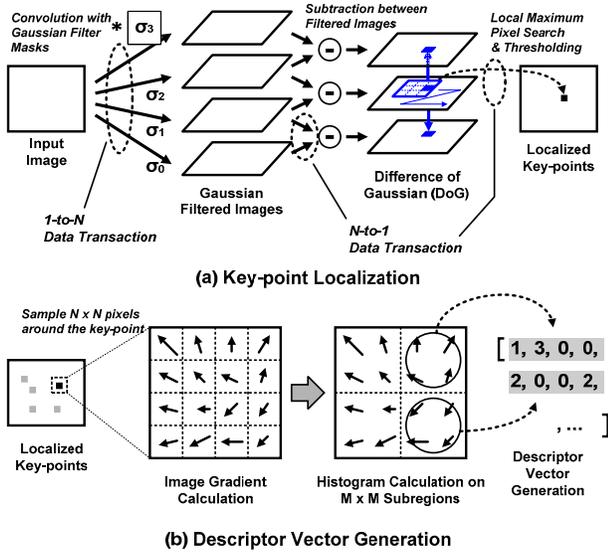
(a) Key-point Localization

(b) Descriptor Vector Generation

Figure 1. Overall Flow of SIFT Computation

The next stage of key-point localization is descriptor vector generation. For each key-point location, N x N pixels of input image are first sampled, then the gradient of the sampled image is calculated. The sample size N is decided according to the DoG image where the key-point is selected. Finally, descriptor vector is generated by computing orientation and magnitude histograms over M x M sub regions of the sampled image. The dimension of descriptor vector is decided in accordance with M. As N varies at each key-point, independent control flows for each PE are essential. Therefore, task level parallelism is suitable for parallel generation of descriptor vectors. In the next section, we propose processor architecture that supports both data/task level parallelism and simultaneous data transactions among parallel tasks.

III. PROCESSOR ARCHITECTURE AND OPERATION

Fig. 2 shows architecture of the proposed object recognition processor. Because Gaussian filtering involves intensive computation, special instructions for image filtering are implemented in the PE. The instructions are SDP (Sum of Dot Product) and LE (Load Extension). The SDP instruction calculates 8 bit 4-way SIMD multiplications and subsequent 4 additions including accumulation in a single cycle. The LE instruction is combination of 8 bit shift and byte load operation, which is designed to supports seamless filtering window movement over image data. By performing the SDP instruction, the PE performs 8 operations in one cycle, where its operation frequency is 200 MHz. As a result, the 10 PEs contributes to 16 GOPS of the total performance.

It is noticeable that tasks in object recognition are tightly related to each other as described in the previous section. Therefore, providing simultaneous data transactions between tasks is essential for exploiting task level parallelism of the 10 PEs. In the proposed processor, simultaneous data transactions between tasks are provided by 8 Visual Image Processing (VIP) memories and their hierarchical star topology Network-
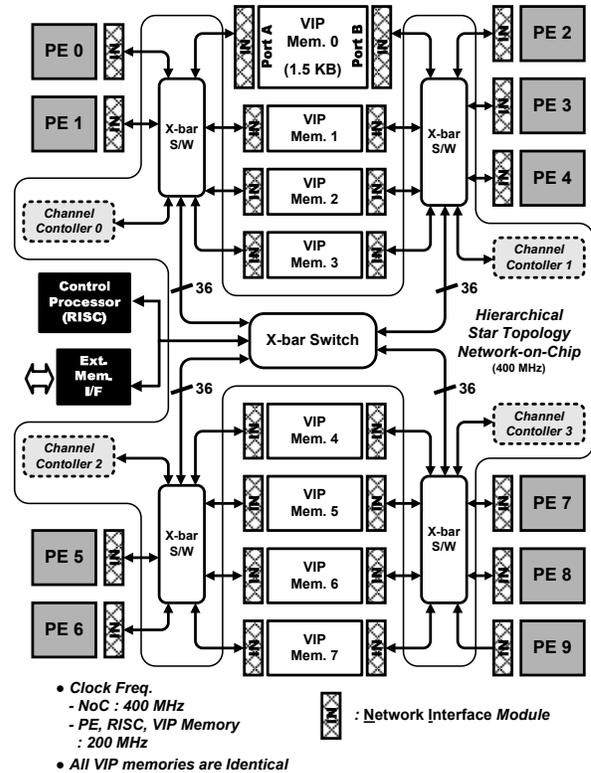

Figure 2. Architecture of the Proposed Object Recognition Processor

on-Chip (NoC). Details of the NoC and the VIP memory are described in the following subsections.

1) Memory Centric NoC

The proposed NoC is designed to provide low overhead parallel data transactions among PEs to facilitate the SIFT computation. As shown in Fig. 1 (a), most of the data transactions are occurred in the form of 1-to-N and N-to-1 source/destination relations. However, traditional inter-processor communication schemes such as message passing and shared memory have their own drawbacks supporting these data transactions. Message passing scheme has overhead of explicit message parsing and redundant data transfer occurs at source PE in case of 1-to-N data transaction. On the other hand, shared memory scheme is appropriate for both 1-to-N and N-to-1 data transactions but unrealistically high speed memory design is required as the number of parallel data transactions increases. Data coherency management is another overhead of shared memory communication.

To resolve disadvantages of the traditional schemes, we devised a hybrid NoC which incorporates advantages of shared memory and hierarchical star topology NoC. Distinctive feature of the proposed NoC is performing inter-PE communications by accessing the VIP memory assigned to the involved PEs. By using the assigned VIP memory as temporary shared memory, efficient inter-PE communication is possible. Memory bottleneck of shared memory communication is resolved by using 8 dual ported VIP
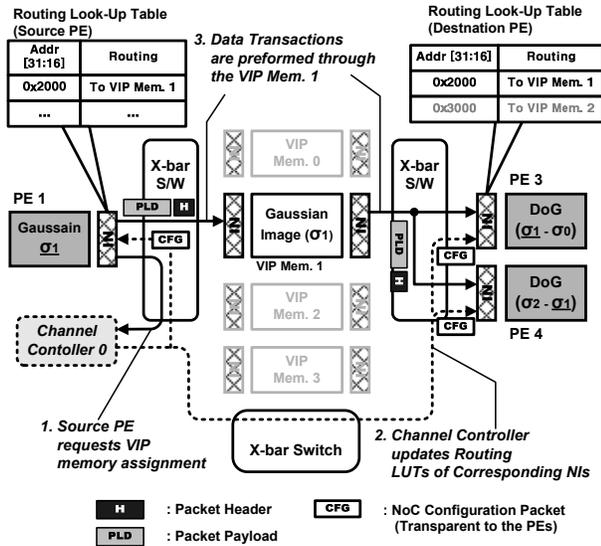
Figure 3. Detailed Operation of the Proposed NoC

Figure 4. Cycle Counts of Local Maximum Location Search Operation

memories simultaneously. The hierarchical star topology NoC provides concurrent interconnections among 8 VIP memories and 10 PEs.

Fig. 3 shows overall operation of the proposed NoC in case of 1-to-N data transaction. In the proposed NoC, assignment of the VIP memory is managed by the channel controllers. (1) In response to the request of source PE, (2) channel controller updates/invalidates packet routing Look-Up Tables (LUT) which are located in Network Interface (NI) modules of source and destination PEs. As a result, (3) accesses to specified address at each PE are directed to the assigned VIP memory.

In the proposed NoC, coherency manage scheme is also supported for unidirectional data transactions. Each pixel entry in the VIP memory has valid bit so that invalid data is not fetched by destination PEs until source PE writes valid data. While the destination PE is waiting for valid data, the NI automatically stalls the PE operation by checking the valid bit status. Therefore, data coherency management is transparent to programs running at each PE. This technique reduces wasted power in polling valid data from source PE.

Operation frequency of the NoC is twice the other part of the chip, which is 400MHz, to reduce overhead of packet switching latency. Aggregated internal bandwidth provided by the proposed NoC is 16 GB/s.

2) Visual Image Processing Memory

The VIP memory is specially designed to find location of local maximum pixel inside 3x3 search window in a single cycle. As shown in Fig. 4, searching local maximum pixel location needs 29~53 cycles on ARM based RISC processor. By replacing this time consuming computation with single read operation, huge performance gain is obtained for key-point localization task. Since local maximum pixel search operation takes 41 cycles on average, 8 VIP memories operating at 200MHz give 65.6 GOPS performance gain.
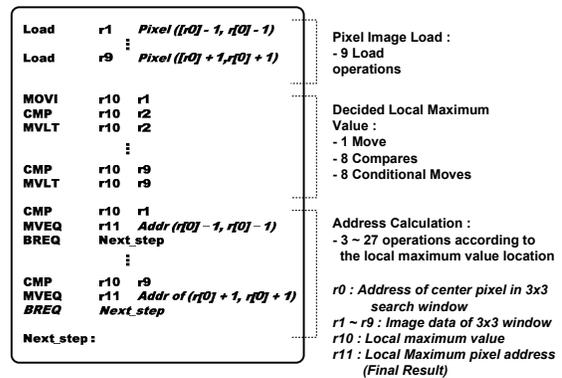
Fig. 5 shows overall architecture of the VIP memory. In the VIP memory, 12 rows by 32 columns of 32 bit pixels are stored which result in total 1.5KB capacity. To compare 9 pixel values in one cycle, every row is interleaved into 3 banks so that bank number assigned for each row is decided by Mod3 operation. Three pixels in the same row are first compared inside the bank, then results from 3 banks are compared again to find local maximum pixel among 9 pixels. Address of local maximum pixel is automatically generated according to the comparison result by address generation unit. At each banks, 3 Comparison Amplifiers (CA) are integrated into every 4 bit line pairs to read 3 pixel values simultaneously. The transistor size of the CA is smaller than normal sense amplifier because it does not drives capacitive long DB lines. To reduce area overhead of comparison logic, Bit Competition Logic (BCL) is also devised. By using priority coded sequential pull down logic, transistor count of the BCL is reduced from 2400 to 536 when compared to the conventional adder based comparator.
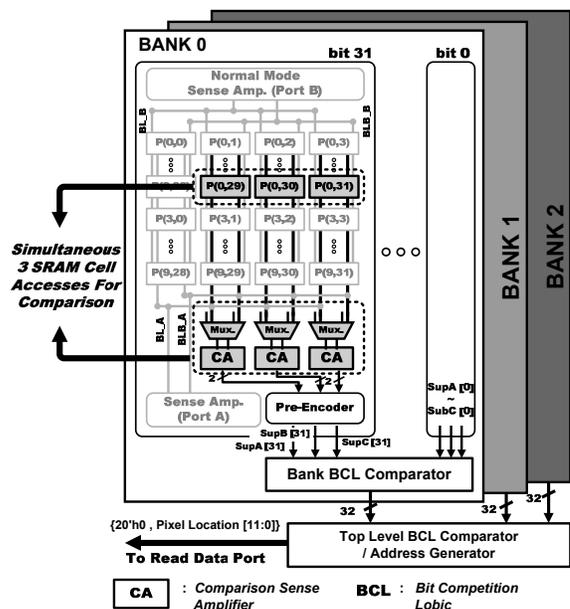

Figure 5. Overall Architecture of the VIP memory

IV. PERFORMANCE EVALUATION AND RESULTS

To evaluate performance, we mapped key-point localization tasks of the SIFT to the proposed processor. Input image, whose size is 320 x 240 pixels, is divided into 32 x 12 pixel sub-images to fit into the VIP memory size. Fig. 6 illustrates cycle count required to perform key-point localization on a sub-image. The SDP/LE instructions and the VIP memory contribute to reduce execution cycles from 94,204 to 30,904. Considering iterations required for computing whole input image and subsequent descriptor vector generation tasks, the proposed object recognition processor generates SIFT features at 10.1 ~ 15.9 fps. The variation of the performance comes from varying number of key-points which is decided by input image context. Table 1 depicts performance comparisons with the previous works. When we compare GOPS/W, the proposed processor shows at least 3.2 times improvement over previous implementations. This power efficiency is obtained from the VIP memory, which removes power consumed for numerous memory accesses and computing cycles by replacing them with a single VIP memory read operation.

Fig. 7 summarizes implementation results. The proposed object recognition processor is implemented with 0.18um 1-poly 6-metal standard CMOS process. Operation frequency of the chip is 400 MHz for the NoC and 200 MHz for the other part of the chip. Size of the chip is 7.7mm x 5mm and its peak power consumption is 1.4W. Evaluation board is also implemented to verify chip operation.

TABLE I. PERFORMANCE COMPARISON WITH PREVIOUS WORKS

	IMAP-CE [1]	VIP [2]	Visconti [3]	This Work
Peak GOPS	51.2	53	18	**81.6**
Power	4W (1.8V)	8W (3.3V)	1W (1.5V)	1.4W (1.8V)
Area (mm²)	121	506	48.7	38.5
GOPS/W	12.8	6.63	18	**58.3**
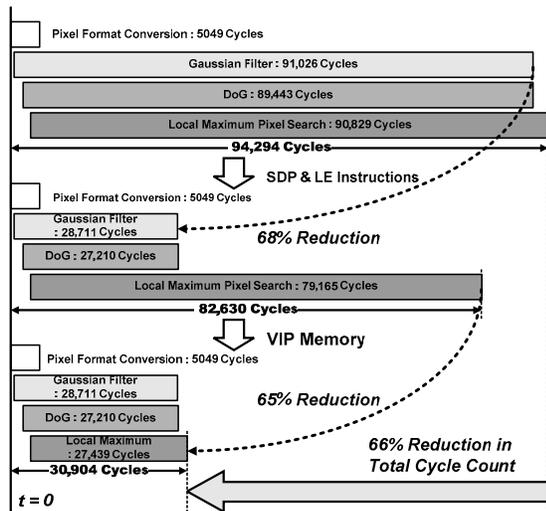
Figure 6. Performance Evaluation Results of Key-point Localization

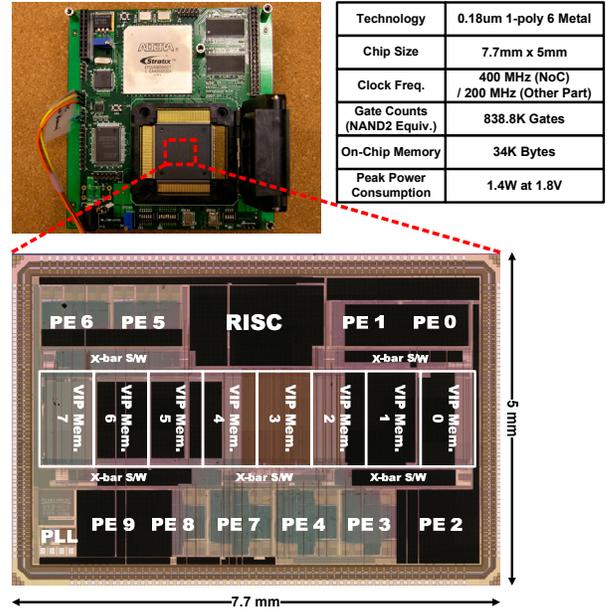| Technology	0.18um 1-poly 6 Metal
Chip Size	7.7mm x 5mm
Clock Freq.	400 MHz (NoC) / 200 MHz (Other Part)
Gate Counts (NAND2 Equiv.)	838.8K Gates
On-Chip Memory	34K Bytes
Peak Power Consumption	1.4W at 1.8V

Figure 7. Chip Micorgraph, Specificaiton and Evaluation Board

V. COLCLUSION

In this paper, we proposed an 81.6 GOPS object recognition processors based on NoC and the VIP memories. The 10 PEs and 8 VIP memories contributes 16GOPS and 65.6 GOPS performance respectively. This huge performance comes from the special image filtering instructions and active utilization of the VIP memory, while the NoC facilitates data transactions among parallel tasks. High power efficiency of the proposed processor is suitable for intelligent robots equipped with limited power supply.

REFERENCES

[1] Shorin Kyo, et. al., "A 51.2 GOPS Scalable Video Recognition Processor for Intelligent Cruise Control Based on a Linear Array of 128 4-Way VLIW Processing Elements," Digest of Technical Papers, IEEE Intl. Solid-State Circuits Conf., Vol. 1, pp. 48-477, 2003

[2] Wolfgang Raab, et. al., "A 100 GOPS Programmable Processor for Vehicle Vision Systems," IEEE Design & Test of Computers, Vol. 20, Issue 1, Jan.-Feb., pp. 8-15, 2003

[3] Jun Tanabe, et. al., "Visconti: Multi-VLIW Image Recognition Processor based on Configurable Processor," Proceedings of the IEEE Custom Integrated Circuits Conf., pp. 185-188, 2003

[4] David G. Lowe, "Distinctive Image Features from Scale-Inavriant Keypoints," ACM Intl. Jounral of Computer Vision, Vol. 60, Issue 2, pp. 91-110, 2004

[5] Sunghwan Ahn, et. al., "Data Assocation Using Visual Object Recognition for EKF-SLAM in Home Environment," Proceedings of IEEE Intl. Conf. on Intelligent Robots and Systems, pp. 2760-2765, 2006.

[6] Patric Jensfelt, et. al., "Augmenting SLAM with Object Detection in a Service Robot Framework," IEEE Intl. Symposium on Robot and Human Interactive Communication, pp. 741-746, 2006

[7] Yongguo Mei, et al., "A Case Study of Mobile Robot's Energy Consumption and Conservation Techniques," Proceedings of the IEEE Intl. Conf. of Advanced Robotics, pp. 492-497, 2005.

[8] Kangmin Lee, Se-Joong Lee, and Hoi-Jun Yoo, "Low-Power Networks-on-Chip for High-Performance SoC Design", IEEE Transactions on Very Large Scale Integration (VLSI) Systems, vol. 14, no.2, pp.148-160, February 2006.