9.8 A 646GOPS/W Multi-Classifier Many-Core Processor with Cortex-Like Architecture for Super-Resolution Recognition

Junyoung Park, Injoon Hong, Gyeonghoon Kim, Youchang Kim, Kyuho Lee, Seongwook Park, Kyeongryeol Bong, Hoi-Jun Yoo

KAIST, Daejeon, Korea

Object recognition processors have been reported for the applications of autonomic vehicle navigation, smart surveillance and unmanned air vehicles (UAVs) [1-3]. Most of the processors adopt a single classifier rather than multiple classifiers even though multi-classifier systems (MCSs) offer more accurate recognition with higher robustness [4]. In addition, MCSs can incorporate the human vision system (HVS) recognition architecture to reduce computational requirements and enhance recognition accuracy. For example, HMAX models the exact hierarchical architecture of the HVS for improved recognition accuracy [5]. Compared with SIFT, known to have the best recognition accuracy based on local features extracted from the object [6], HMAX can recognize an object based on global features by template matching and a maximum-pooling operation without feature segmentation. In this paper we present a multi-classifier many-core processor combining the HMAX and SIFT approaches on a single chip. Through the combined approach, the system can: 1) pay attention to the target object directly with global context consideration, including complicated background or camouflaging obstacles, 2) utilize the super-resolution algorithm to recognize highly blurred or small size objects, and 3) recognize more than 200 objects in real-time by context-aware feature matching.

Figure 9.8.1 shows the multi-classifier object recognition system. In order to realize object recognition in real-time with low energy consumption, the proposed processor adopts 3 architectural features: 1) a 5-stage fine-grained pipeline for high throughput in a tile-based SIFT operation, 2) a mixed-mode intelligent hierarchical perception engine (IHPE) for fast HMAX operation, and 3) context-aware dynamic resource management (DRM) for real-time operation with low energy consumption. The HMAX and SIFT blocks shown in Fig. 9.8.1 generate global and local features, respectively, in parallel, and iteratively interact via back-and-forth feedback. Label A in Fig. 9.8.1 represents the global contextual top-down attention map, discarding features from unrelated objects. Label B represents super-resolution feedback produced for tiles with insufficient features by running a super-resolution algorithm on an "upper scale", i.e. a larger tile containing more information. Label C corresponds to the context-aware feature matching based on global features obtained from HMAX. The previous models [2-3] utilize SIFT features alone, without the complex segmentation process for top-down attention. However, their object recognition achieves less than ~80% attention accuracy. In the proposed system, the HMAX engine extracts the global feature descriptors of the scene and provides prior probability to the local features for high attention accuracy. The prior probabilities are used for feature matching; low probability features are discarded as unrelated. The super resolution is performed on ambiguous regions with insufficient local features. As a result, the accuracy of attention is increased to 83.6%, while the previous study achieved only 60.0% accuracy of "object-related" attention.

The proposed MCS many-core processor of Fig. 9.8.2 incorporates 21 IP cores: a Scale-Space Engine (SSE), 4 Feature Detection (FD) Clusters containing 8 vector cores and 4 shared-bus Vector Processing Elements (VPE), 4 Description Generation (DG) clusters, including scalar cores with special ALUs, an Intelligent Hierarchical Perception Engine (IHPE), a Feature Matching Processor (FMP), a Dynamic Resource Manager (DRM), and a filter accelerator. The cores are connected to one another by a global hierarchical star NoC [7] and the vector cores are connected through a local network to share the VPEs in the feature detection cluster.

Fig. 9.8.3 shows the task-level partitioned architecture for SIFT-based object recognition. It is composed of: (1) Gaussian Filtering (GF), (2) Difference of Gaussian (DoG), (3) Local Maximum (LM), (4) Feature Description (FD), and (5) Feature matching (FM). The first stage, GF, is implemented by a massively parallel SSE for simultaneous multi-tile operations. The SSE reduces GF processing time by 66.8% and 33.7% compared to a single-threaded 20-way SIMD processor, and a SMT-based SIMD implementation, respectively [3]. Since only 46% of instructions activate the SIMD datapath in the two stages, the DoG and the LM cores share the 16b 8-way VPE through an inter-bus scheduler. The shared VPE is operated in a superscalar manner, providing 2 operation modes: 1) coarse-grained sharing which allows a core to occupy all the buses, 2) fine-grained sharing which allows two cores to perform different operations on the same bus. As a result, the overall processing time for local feature extraction is reduced by 57.2% compared to the SMT-SIMD-based object processor.

The mixed-mode IHPE is shown in Fig. 9.8.4. The digital FSM controller extracts global features from the scale-space image input, and an analog radial basis function network (RBFN) classifies the input into the corresponding scene category. The 3-tuple, (C_i, R_i, $V_{LT,i}$), representing the center of the RBF, radius of the RBF, and logical threshold voltage of Sigmoid function, respectively, are learned for each scene S_i. The C_i and R_i are adjusted by controlling V_{REF1}, V_{REF2} and the parallel transistors B as shown in Fig. 9.8.4(c), and $V_{LT,I}$ is adjusted by controlling V_{BP}. For example, the RBFN learning gives (C_1, R_1, V_{LT1}) of indoor images (S_1) as (0.20, 0.22, 0.60V) and (C_2, R_2, V_{LT2}) of highway images (S_2) as (0.56, 0.54, 0.31V) as Fig. 9.8.4. When a new highway scene is input, the confidence value of S_2's RBF is higher than the confidence value of S_1's RBF and classifies the input as S_2. As a result, the HMAX recognition can be performed within 16ms for 25-category scenes. The area and power consumption of the mixed-mode RBFN are 68.4μm² and 0.723mW, respectively. This represents an 87% and 96% reduction, respectively, compared to the equivalent digital implementation.

Fig. 9.8.5 shows the timing chart of the proposed processor performing the MCS in real-time. Because the HMAX recognition takes 16ms regardless of the scene content, the SIFT recognition should be performed within about 17.3ms by adjusting the processing speed of the SSE, FD cluster, and DG clusters. Fig. 9.8.6 shows the DRM operation for continuous video frames. The DRM monitors the number of features, the number of ROI (region of interest) tiles from the FD clusters, and power-thermal headroom from the external power management IC. Based on these parameters, the DRM controls the power-mode of two domains separately (i.e. the FD domain and the DG domain) to reduce the average power consumption of the processor by 24% compared to when the SIFT is performed at nominal voltage for a normal scene with medium feature density. As a result, the proposed processor achieves 646GOPS/W power efficiency and 9.4nJ/pixel energy efficiency, which is competitive with state-of-the-art object recognition processors.

The proposed processor is implemented with 0.13μm 8-metal CMOS technology and occupies 25mm² with 1.8M equivalent gates and 200kB of on-chip SRAM (Fig. 9.8.7). The 21 IP cores consume 260mW on average at 200MHz, 1.2V. Peak performance is 271.4GOPS, while peak power efficiency is 646 GOPS/W, using DVFS operation at 50MHz, 0.65V. As a result, 96% recognition accuracy and 9.4 nJ/pixel energy efficiency are achieved on 30fps HD video of highly blurred and small-size objects in complicated backgrounds with applications in UAV surveillance.

References:

[1] J.-Y. Kim, *et al.*, "A 201.4GOPS 496mW Real-Time Multi-Object Recognition Processor with Bio-Inspired Neural Perception Engine," *ISSCC Dig. Tech, Papers*, pp. 150-151, 2009.

[2] S, Lee, *et al.*, "A 345mW Heterogeneous Many-Core Processor with an Intelligent Inference Engine for Robust Object Recognition," *ISSCC Dig. Tech, Papers*, pp. 332-333, 2010.

[3] J. Oh, *et al.*, "A 320mW 342GOPS Real-Time Moving Object Recognition Processor for HD 720p Video Streams," *ISSCC Dig. Tech. Papers*, pp. 220-221, 2012.

[4] Ho, T.K., Hull, J.J., Srihari, S.N., "On multiple classifier systems for pattern recognition," Pattern Recognition, vol. 2. pp. 84-87, 1992.

[5] M. Riesenhuber and T. Poggio, "Hierarchical models of object recognition in cortex," *Nature Neuroscience*, vol. 2, pp. 1019-1025, 1999.

[6] D.G. Lowe, "Distinctive image features from scale-invariant keypoints," Intenational Journal of Computer Vision, vol. 60, no. 2, pp. 91-110, 2004.

[7] H.-J.Yoo, *et al.*, "Low-Power NoC for High-Performance SoC Design," *CRC Press*, 2008.

[8] Y. Tanabe, *et al.*, "A 464GOPS 620GOPS/W Heterogeneous Multi-Core SoC for Image-Recognition Applications," *ISSCC Dig. Tech. Papers*, pp. 222-223, 2012.

[9] T. Kurafuji, *et al.*, "A Scalable Massively Parallel Processor for Real-Time Image Processing," *ISSCC Dig. Tech. Papers*, pp. 334-335, 2010.

 978-1-4673-4516-3/13/$31.00

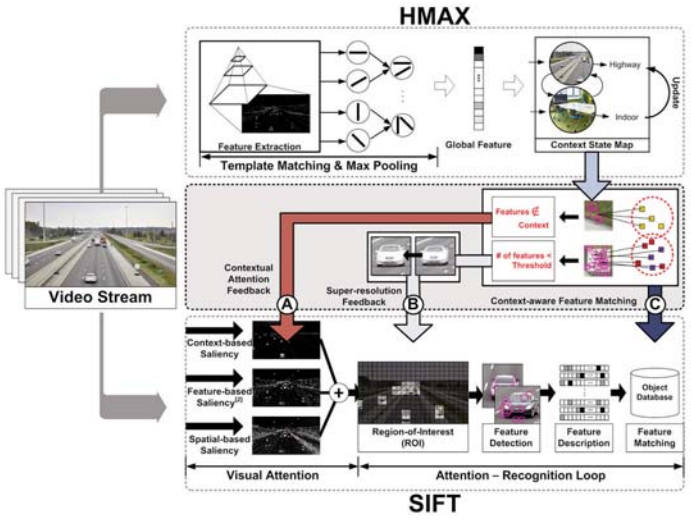


Figure 9.8.1: Multi-classifier object recognition system.

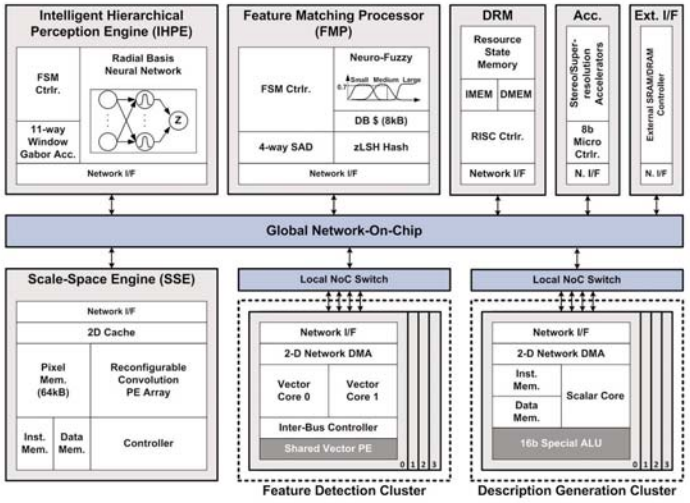


Figure 9.8.2: Overall architecture of the MCS many-core processor.

9

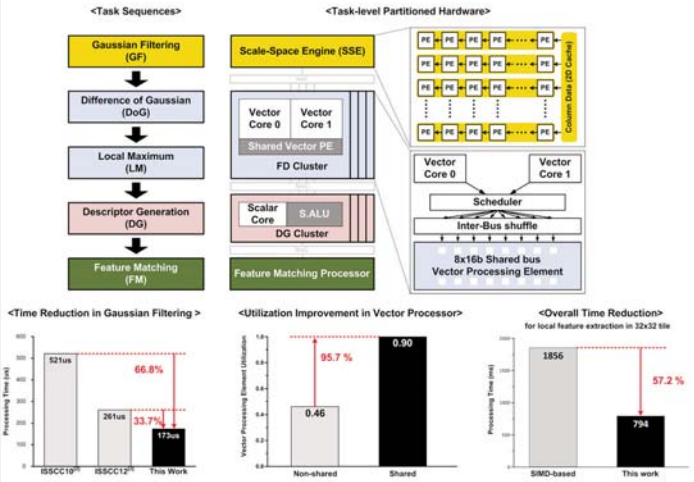


Figure 9.8.3: Task-level partitioned hardware for SIFT object recognition.

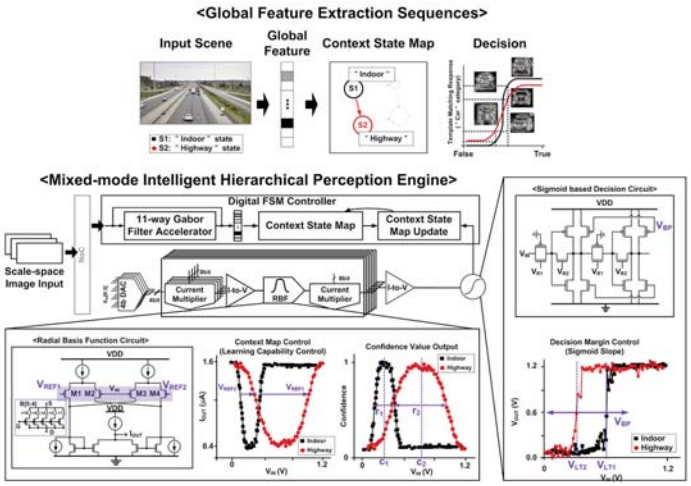


Figure 9.8.4: Mixed-mode intelligent hierarchical perception engine and measurement results.

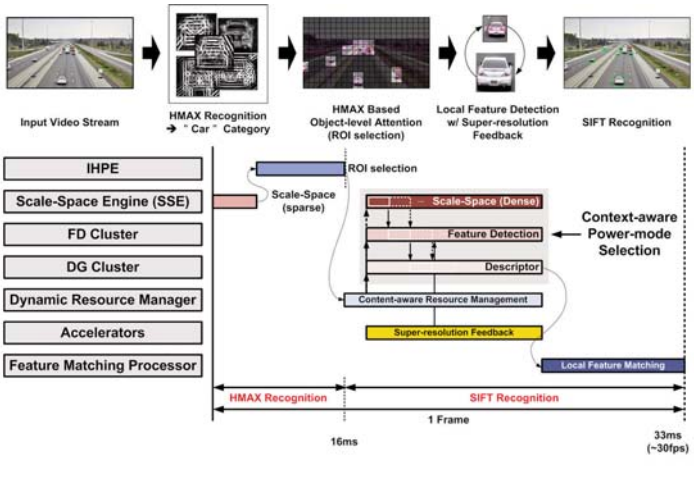


Figure 9.8.5: Timing chart of multi-classifier execution in the proposed processor.

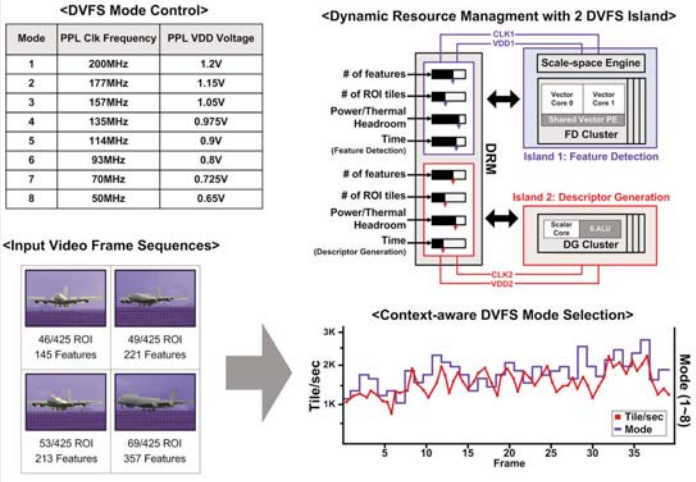
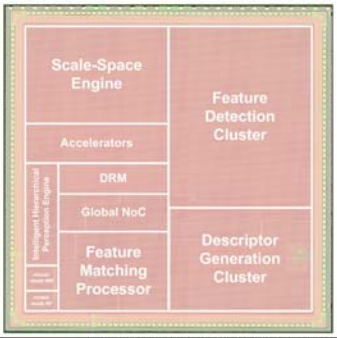


Figure 9.8.6: Context-aware dynamic resource management.



	ISSCC 2010[2]	ISSCC 2010[6]	ISSCC 2012[3]	ISSCC 2012[5]	This Work
Process	130nm	65nm	130nm	40nm	130nm
Area	50mm^2	24.2mm^2	32mm^2	44.5mm^2	25mm^2
Power Supply (V)	0.65-1.2	1.0	0.65-1.2	1.1	0.65-1.2
Frequency (MHz)	50-200	200-560	50-200	180, 266	50-200
Power (mW)	704	330	534	749	420
GOPS/W*	324	310	640	619.4	646
Per-pixel Energy	37.1 nJ	-	10.5 nJ	-	9.4 nJ

Process	0.13μm 1P8M Mixed-mode CMOS	
Chip Size	5.0 x 5.0 mm^2	
Frequency	Nominal	200MHz(90FO4) (Digital) 20MHz (Mixed)
	DVFS	50 – 200MHz
Voltage	Nominal	1.2V
	DVFS	0.65 – 1.2V
Gate / SRAM	1.8M / 200kBytes	
Power Dissipation	420mW (Peak) / 260mW (Avg.)	
Peak Performance	SSE	119.4 GOPS
	FDC	24 GOPS
	DGC	2.4 GOPS
	FMP	109.4 GOPS
	IHPE	10.8 GOPS
	DRM / Acc.	5.4 GOPS
	Total	271.4 GOPS
Area Efficiency	10.86 GOPS/mm^2	
Power Efficiency	646 GOPS/W	
Per-pixel Energy	9.4 nJ/pixel	

Figure 9.8.7: Chip micrograph and performance summary

978-1-4673-4516-3/13/$31.00