12.4 A 320mW 342GOPS Real-Time Moving Object Recognition Processor for HD 720p Video Streams

Jinwook Oh, Gyeonghoon Kim, Junyoung Park, Injoon Hong, Seungjin Lee, Hoi-Jun Yoo

KAIST, Daejeon, Korea

Moving object recognition in a video stream is crucial for applications such as unmanned aerial vehicles (UAVs) and mobile augmented reality that require robust and fast recognition in the presence of dynamic camera noise. Devices in such applications suffer from severe motion/camera blur noise in low-light conditions due to low-sensitivity CMOS image sensors, and therefore require higher computing power to obtain robust results vs. devices used in still image applications. Moreover, HD resolution has become so universal today that even smartphones support applications with HD resolution. However, many object recognition processors [1][2] and accelerators [3] reported for mobile applications only support SD resolution due to the computational complexity of object recognition algorithms. For example, the prior works [1] and [2] obtain 11f/s and 7f/s processing speed for HD video streams, respectively, which is far below real-time requirements.

This paper presents a moving-target recognition processor for HD video streams. The processor is based on a context-aware visual attention model (CAVAM) (Fig. 12.4.1), comprising 3 architectural features: 1) simultaneous multithreading feature-extraction clusters (SFEC) for high-speed processing of the scale invariant feature transform (SIFT); 2) a keypoint matching processor (KMP) for fast object matching; and, (3) a dynamic resource controller (DRC) with a massively parallel artificial intelligence engine (MP-AIE).

For robust moving object recognition with high processing speed, the CAVAM is applied to analyze the regions-of-interest (ROIs) in video sequences. The previous model [1] generated ROIs solely using spatial familiarity between proto and target objects. Conversely, the CAVAM considers the temporal similarity between successive images, as well as spatial familiarity in an image. The context state buffer (CSB) stores the prior recognition results and calculates the current object location, r_t, velocity, v_t, and acceleration, a_t, as clues to the next object location. Then, the Kalman filter receives r_t, v_t, and a_t to compute a predicted next object location. The temporal familiarity is an elliptic disk whose intensity is given by $e^{-|m_{t+1}-r_{t+1}|}$, where r_{t+1} is the calculated location, and m_{t+1} is the measured location. The familiarity map reflects not only the spatial conspicuity but also temporal continuity so that the obtained ROI can track the target object accurately irrespective of blurring, color distortion, and occlusion. As a result, the CAVAM achieves 81% attention accuracy and reduces recognition complexity by 60% for object recognition in HD video streams.

Figure 12.4.2 shows a block diagram of the processor. For object recognition with 720p resolution, 26 heterogeneous cores are connected together through a hierarchical network-on-chip (NoC) and clustered into the image processing core (IPC) and the DRC. The IPC, comprising 4 SFECs, a KMP, and a MP-AIE, performs the data-intensive recognition operations of the CAVAM, while the DRC reconfigures the heterogeneous IPC tasks dynamically to maximize system throughput [4]. The DRC's global task scheduler (GTS) monitors the number of ROIs and keypoints in the workload, and dynamically modifies 4 resources: the threads allocated to the 4 SFECs, the bandwidth of the global network switch, the operating voltage and the clock frequency.

The 3-stage coarse-grained pipeline architecture realizing the CAVAM algorithm is depicted in Fig. 12.4.3. The MP-AIE extracts the ROIs in the input image based on saliency and familiarity maps. The SFEC calculates SIFT keypoint descriptors from the extracted ROIs. The KMP uses the calculated descriptors for keypoint matching in target object recognition. The DRC uses the MP-AIE to analyze the workload patterns at run-time, allowing for the dynamic reconfiguration of the hardware. For example, even though the CAVAM has at most 45% workload variance in keypoint description and matching, the DRC achieves stable system throughput with sustained 95% utilization of the SFECs and KMP and 140.4GB/s bandwidth in the global network switch, using a weighted round-robin scheme.

Figure 12.4.4 shows the SFEC architecture which performs simultaneous multithreading (SMT) for high-throughput SIFT operations. Thanks to the dual-thread capability, 4 SFECs can process ~6.8 ROIs simultaneously for HD images. Each SFEC consists of 1 dual-thread vector processing element (DVPE) for feature detection and 4 scalar processing elements for keypoint description. The DVPE has 16 units of 16b SIMD. Each unit is composed of Gaussian filtering (GF), the difference of Gaussian (DoG), local maximum extraction (LME) and a SIMD controller. Since the GF, DoG and LME are pipelined (3 stages), the DVPE reduces processing delay by 35% compared to a single-threaded SIMD unit for feature detection. Compared to a scenario without the dual-thread capability or the ARM Cortex-A9 SIMD, the DVPE achieves 42% and 65% higher throughput, respectively, for the SIFT operations, with an 11% area overhead.

Figure 12.4.5 shows the KMP architecture for keypoint matching with zero-filtered least-sensitive hashing (ZLSH). The collision equalizer, which is the hash index generator of KMP, adopts a random permutation algorithm for the 16 signatures of the least-sensitive hashing (LSH) to reduce the redundant zeros in the signature, reducing DB size by 24% vs. the original LSH using the 30b ZLSH index. There are two keypoint matching approaches: cache-based matching (CBM) and DB-based matching (DBM). CBM uses the keypoints from a prior matching, stored in the cache, to generate the nearest neighbor. If a keypoint is matched, the matching ends with a 98% reduction in external accesses. Otherwise, if the keypoint is not matched, an additional DBM is performed, with the ZLSH index used to access the candidate keypoints of the DB, reducing accesses by 86% relative to a brute force approach.

The measurement results for the DRC operation are shown in Fig. 12.4.6 for 30 continuous HD video frames. The DRC adjusts the learning rate, α, within 4ms for target objects with different characteristics. For high-speed on-line learning, a latch-based fast true random number generator circuit is integrated. Since the hardware-oriented weight perturbation needs a series of random bits for state parameters, fast and accurate random numbers are needed to obtain accurate learning in the DRC. With the help of FIFO-based post processing, the circuit provides high randomness to the DRC and achieves a 1.3× higher frame rate than if the IPC were statically allocated. 650GOPS/W power efficiency is achieved.

A 4×8mm^2 test chip is fabricated using 0.13μm 6-metal CMOS technology, integrating 1.4M gates and 382kB of SRAM (Fig. 12.4.7). It achieves 342GOPS peak performance, with 320mW average power, 1.2V, 200MHz. For moving object recognition in a 720p video stream severely corrupted by environmental noise, the processor sustains 83% recognition accuracy – a ~2× improvement over the 45% attained by the previous processor. The chip achieves 650GOPS/W power efficiency and 10.69GOPS/mm^2 area efficiency, representing 17% and 83% improvements over a state-of-the-art recognition processor [1], respectively.

References:

[1] 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.

[2] 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.

[3] J. Oh, et al., "A 57mW Embedded Mixed-Mode Neuro-Fuzzy Accelerator for Intelligent Multi-core Processor," *ISSCC Dig. Tech. Papers*, pp. 130-131, 2011.

[4] E. Ipek, et al., "Self-Optimizing Memory Controllers: A Reinforcement Learning Approach," *IEEE International Symp. On Computer Architecture*, pp. 39-50, 2008.

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

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

 978-1-4673-0377-4/12/$31.00

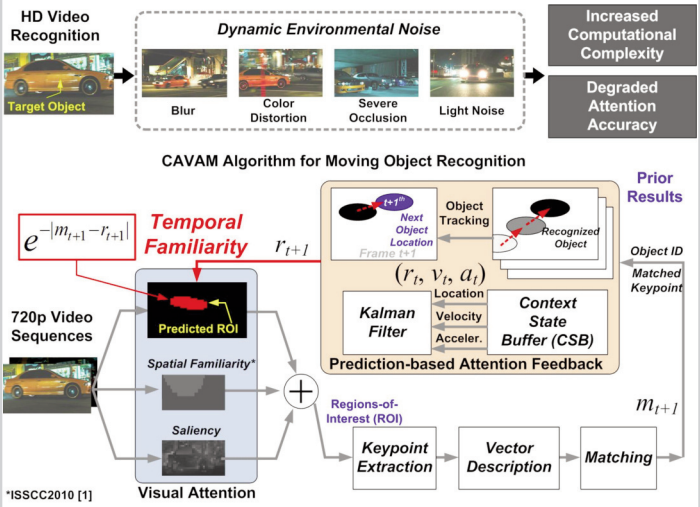


Figure 12.4.1: Moving object recognition system with CAVAM.

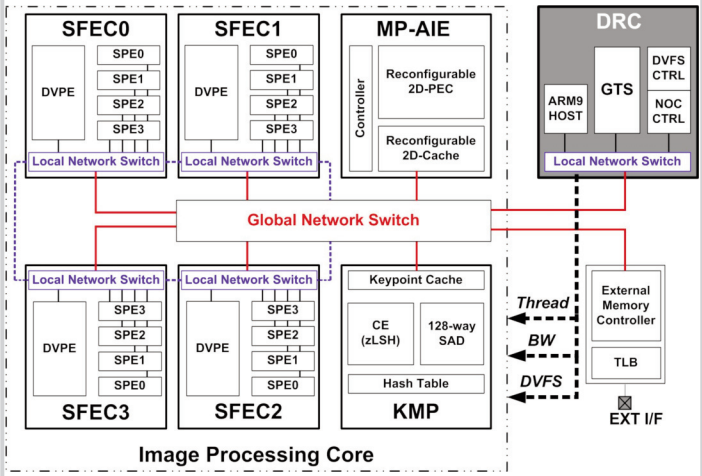


Figure 12.4.2: Overall architecture of the object recognition processor.

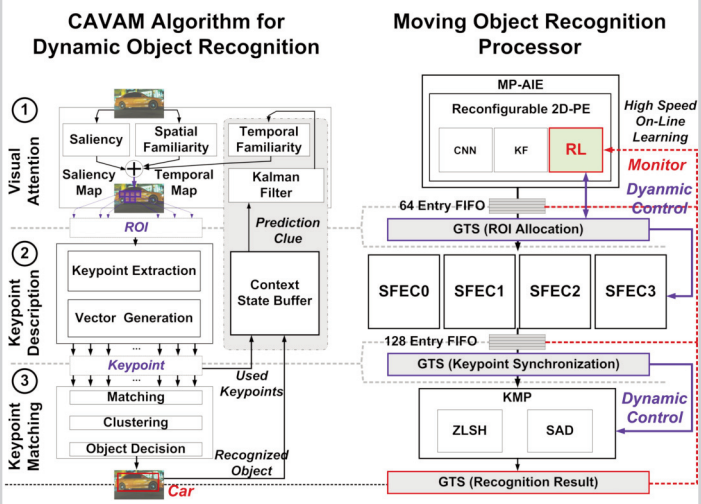


Figure 12.4.3: Dynamic resource management of 3-stage CAVAM.

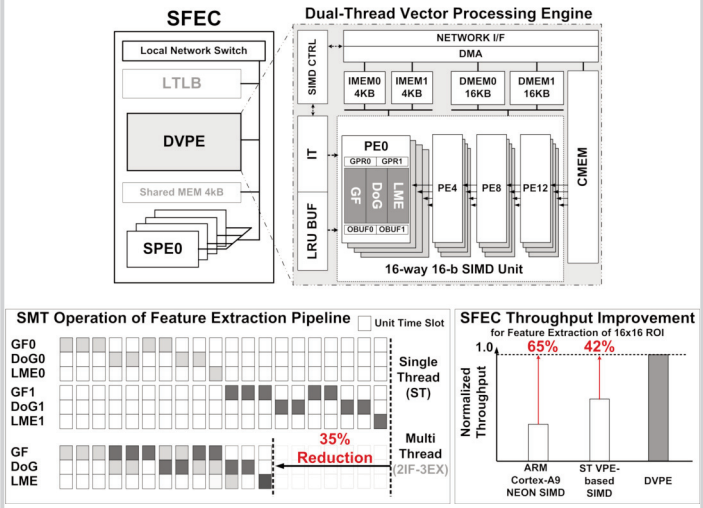


Figure 12.4.4: Block diagram of SFEC and its multi-threading.

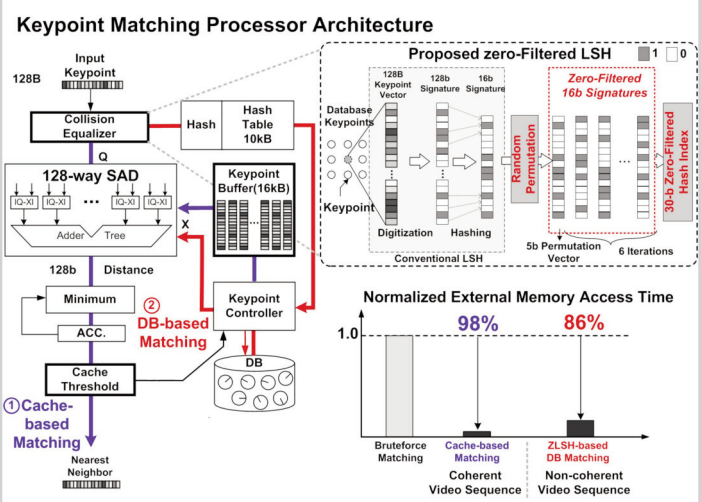
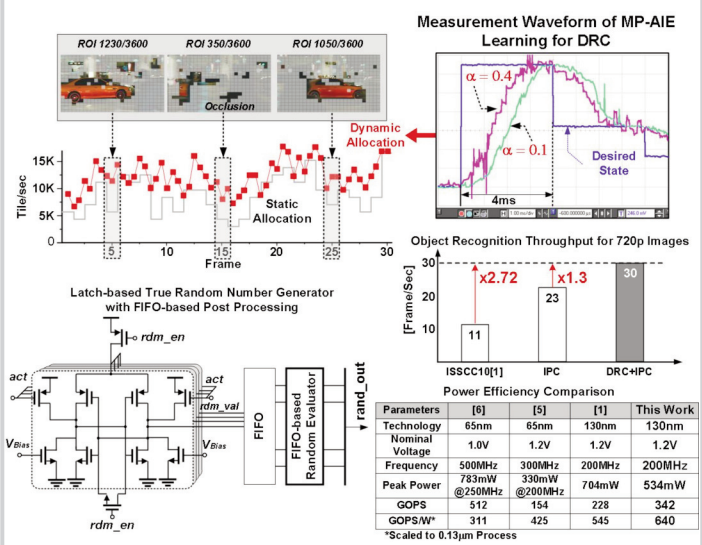


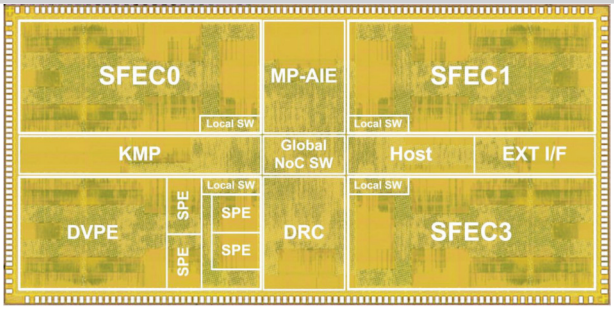
Figure 12.4.5: KMP architecture and throughput improvement.



Parameters	[6]	[5]	[1]	This Work
Technology	65nm	65nm	130nm	130nm
Nominal Voltage	1.0V	1.2V	1.2V	1.2V
Frequency	500MHz	300MHz	200MHz	200MHz
Peak Power	783mW @250MHz	330mW @200MHz	704mW	534mW
GOPS	512	154	228	342
GOPS/W*	311	425	545	640

*Scaled to 0.13μm Process

Figure 12.4.6: Measurement results of learning-based DRC.



Technology	0.13μm 1P6M Logic CMOS	
Chip Size	4.0mm x 8.0mm	
Gate Count	1.4M	
SRAM	382kB	
Supply Voltage	Nominal Voltage	1.2 V
	DVFS Domain	0.7 ~1.2 V
Clock Frequency	Nominal Frequency	200MHz (90FO4)
	DVFS Domain	50~200MHz (90FO4)

Power Consumption	534mW (Peak) 320mW (Average)	
Peak Performance	4 SFECs	102.4 GOPS (SIMD) 9.6GOPS (MIMD)
	KMP	98 GOPS
	MP-AIE	128 GOPS
	DRC	4 GOPS
	Total	342 GOPS
Area Efficiency	10.69 GOPS/mm^2	
Power Efficiency	640GOPS/W	

Figure 12.4.7: Die micrograph and feature summary.

978-1-4673-0377-4/12/$31.00