A 502-GOPS and 0.984-mW Dual-Mode Intelligent ADAS SoC With Real-Time Semiglobal Matching and Intention Prediction for Smart Automotive Black Box System

Kyuho Jason Lee, *Student Member, IEEE*, Kyeongryeol Bong, *Student Member, IEEE*, Changhyeon Kim, *Student Member, IEEE*, Jaeeun Jang, *Student Member, IEEE*, Kyoung-Rog Lee, *Student Member, IEEE*, Jihee Lee, *Student Member, IEEE*, Gyeonghoon Kim, *Student Member, IEEE*, and Hoi-Jun Yoo, *Fellow, IEEE*

***Abstract*— The advanced driver assistance system (ADAS) for adaptive cruise control and collision avoidance is strongly dependent upon the robust image recognition technology such as lane detection, vehicle/pedestrian detection, and traffic sign recognition. However, the conventional ADAS cannot realize more advanced collision evasion in real environments due to the absence of intelligent vehicle/pedestrian behavior analysis. Moreover, accurate distance estimation is essential in ADAS applications and semiglobal matching (SGM) is most widely adopted for high accuracy, but its system-on-chip (SoC) implementation is difficult due to the massive external memory bandwidth. In this paper, an ADAS SoC with behavior analysis with Artificial Intelligence functions and hardware implementation of SGM is proposed. The proposed SoC has dual-mode operations of high-performance operation for intelligent ADAS with real-time SGM in D-Mode (d-mode) and ultralow-power operation for black box system in parking-mode. It features: 1) task-level pipelined SGM processor to reduce external memory bandwidth by 85.8%; 2) region-of-interest generation processor to reduce 86.2% of computation; 3) mixed-mode intention prediction engine for dual-mode intelligence; and 4) dynamic voltage and frequency scaling control to save 36.2% of power in d-mode. The proposed ADAS processor achieves 862 GOPS/W energy efficiency and 31.4-GOPS/mm^2 area efficiency, which are 1.53× and 1.75× improvements than the state of the art, with 30 frames/s throughput under 720p stereo inputs.**

***Index Terms*— Advanced driver assistance system (ADAS), behavior analysis, intention prediction, mixed-mode implementation, p2-less compression, semiglobal matching (SGM), smart automotive black box.**

I. Introduction

ADVANCED driver assistance system (ADAS) has been actively studied for safety and driver convenience with providing forward collision warning [1], lane-departure warning [2], adaptive cruise control [3], advanced emergency braking [4], etc. In addition, autonomous vehicles are animatedly under development with artificial intelligence (AI) technology as the final goal of ADAS [5]–[9]. Both of them are heavily dependent upon the robust image recognition such as vehicle detection, pedestrian detection, and traffic sign recognition. However, they are equipped with several computers with powerful GPUs due to the computationally intensive algorithms for autonomous driving, e.g., behavior analysis and path planning. Since the power consumption of those computers is of the order of 100 W, they are impractical for production thus far. Thus, there exists a huge gap between ADAS and autonomous technology due to the absence of low-power intelligent processors. In history, vehicles mounted GPS navigation systems as a smart device in the 1990s [10], and it led the technology to today's ADAS. Now, a smart automotive black box that exploits ADAS functions will be the next device that bridges the gap because they are installed in most cars for tracking accidents or theft. The smart automotive black box should provide ADAS functions such as forward collision warning and intelligent collision evasion when the vehicle is in D-Mode (d-mode), while it records surveillance video to track theft or damage when the vehicle is in parking-mode (p-mode).

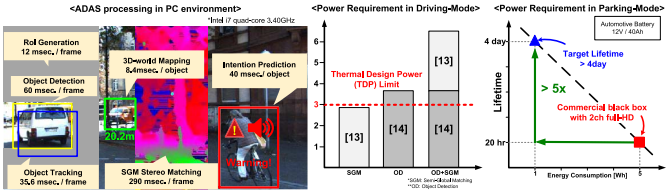


Fig. 1. Examples of ADAS algorithms and different power requirements in d-mode and p-mode.

ADAS functions in d-mode require a high-performance computing engine to meet the real-time constraint (>30 frames/s) for safety, because multiple algorithms should be run simultaneously. Meanwhile, it must ensure energy-efficient computing with <3 W of power consumption as well [11] due to the thermal design power limits without cooling system [12]. Nevertheless, today's CPU cannot fulfill such requirements because of the complexity of each algorithm as shown in Fig. 1. Even with the large power (∼100 W), it cannot satisfy 30 frames/s throughput.

Manuscript received May 1, 2016; revised October 6, 2016; accepted October 8, 2016. Date of publication November 15, 2016; date of current version January 4, 2017. This paper was approved by Guest Editor Dejan Markovic. This work was supported by the MSIP (Ministry of Science, ICT and Future Planning), Korea, under the Information Technology Research Center ITRC) support program under Contract IITP-2016-R7117-16-0163 supervised by the Institute for Information and Communications Technology Promotion (IITP).

The authors are with the School of Electrical Engineering, KAIST, Daejeon 34141, South Korea (e-mail: kyuho.jsn.lee@kaist.ac.kr).

Color versions of one or more of the figures in this paper are available online at http://ieeexplore.ieee.org.

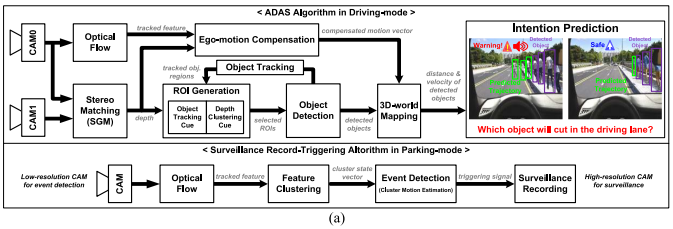
Digital Object Identifier 10.1109/JSSC.2016.2617317

Especially, semiglobal matching (SGM) is essential in most ADAS functions for precise distance estimation due to its highly accurate depth information [13], but its latency (290 ms) is 8.79× than the requirement (33 ms). Therefore, SGM should be accelerated for practical usage, but its system-on-chip (SoC) implementation has not been reported because of its memory bottleneck. Only a field-programmable gate array (FPGA) implementation [13] can be used for a low-power embedded system, but it is not suitable for ADAS applications since it exceeds 3 W when used with object detection [14]. Thus, a dedicated chip implementation is essential but previous ADAS SoCs [14]–[16] were not capable of SGM. Moreover, they provided the driver with several tens of detected results to which not all should be paid attention. In reality, only some objects that suddenly pop out on the driving lane should be given attention, i.e., only meaningful information should be selected intelligently throughout behavior analysis of the objects because providing too much information disturbs the driver. There have been attempts at behavior analysis [17]–[20], but they suffered from intensive computation due to their algorithm complexity. Thus, previous SoCs [14]–[16] focused solely on object detection without behavior analysis.

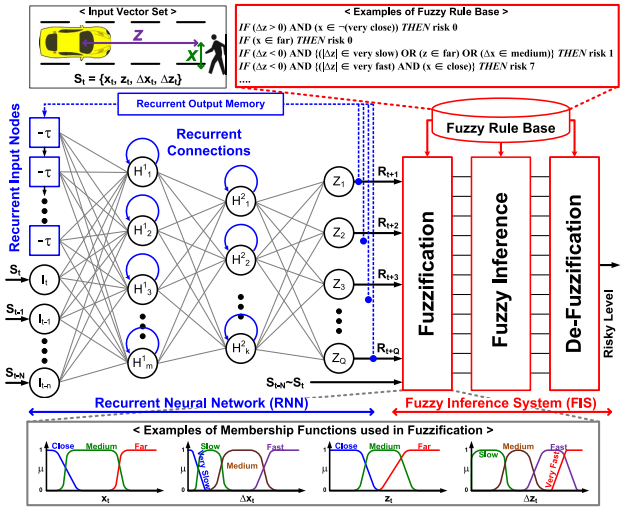
Unlike in d-mode, ultralow-power operation is necessary in p-mode because the black box runs with automotive battery of which typical capacity is 40 Ah with 12-V output. Although it seems to have large capacity, the power budget that surveillance recording can consume drops because of temperature change, powering electronic devices, and engine cranking [21]. However, always-on recording consumes too much power. For instance, present 2ch HD black box consumes 4.8 W in average for recording [22], and it results in only 20 h of continuous recording (Fig. 1). To extend the lifetime of surveillance recording, it should be intelligently turned on only when objects are about to harm the vehicle, or getting closer.

To fulfill the requirements, in algorithmic aspect, the proposed chip utilizes: 1) intention-prediction algorithm for intelligence and 2) precise region-of-interest (ROI) generation. The intention-prediction provides intelligence in both modes. In d-mode, it gives alerts only on objects that have potential risk to cut in driving lane for intelligent collision evasion. In p-mode, it intelligently triggers surveillance recording only when objects are getting closer to the car. In hardware aspect, the proposed ADAS processor features: 1) tile-based task-level pipelined SGM processor (SGMP) for high-accuracy depth map extraction; 2) fully digital configuration of the hybrid intention prediction engine (IPE) for high performance operation in d-mode; and 3) mixed-mode configuration of the IPE for ultralow-power consumption in p-mode.

The rest of this paper is organized as follows. Section II describes the overall ADAS algorithm flow as well as the SGM and the proposed intention-prediction in detail. In Section III, the overall architecture of the SoC is explained with detailed core implementations. The dynamic voltage and frequency scaling (DVFS) control for high energy efficiency in d-mode will be explained in Section IV. Sections V and VI show the implementation and evaluation results, respectively, followed by the conclusion in Section VII.



(a)



(b)

Fig. 2. (a) Overall ADAS algorithm flow of the proposed system in each mode. (b) Proposed intention-prediction algorithm.

II. Algorithm

A. Overall ADAS Algorithm

Fig. 2(a) shows the overall ADAS algorithm flow with stereo cameras where CAM0 becomes the reference image for image processing while CAM1 is the target image for SGM. The SGM extracts depth map by finding correspondence between CAM0 and CAM1, and it provides longitudinal distance. Meanwhile, optical flow is performed for feature tracking, and the results of both optical flow and SGM are fed into ego-motion compensation, which discards the movement of the car. At the same time, the ROI generation stage selects small portions of image by using both depth and object tracking information. The object detection is performed only within the selected ROIs to reduce computation cost. Then, the 2-D-pixel coordinates of the detected objects are transformed into the 3-D-world coordinates to interpret distance and velocity. Finally, the intention-prediction decides which object will cut in the driving lane based on trajectory prediction, providing behavioral analysis for intelligent collision evasion.

The system triggers surveillance recording in p-mode only when an *event*, which means something is getting closer to the car, is detected. In recording theft or damage, the objects are not particularly subjected to pedestrians and vehicles but any objects coming closer to the car should be regarded as an event. After optical flow vectors are extracted, neighboring vectors are clustered where each cluster is regarded as an object. Then the state vectors representing size and position of the cluster are used for event detection that decides if the object is getting closer to the car or not by monitoring changes in the clusters. If the object size is getting bigger without

motion in the x-direction, it is an event because it can be interpreted as the object is getting closer. In this manner, the system turns on surveillance recording with high-resolution camera.

B. Semiglobal Matching and the Proposed Intention-Prediction Algorithms

The depth, or distance, information is essential for ADAS applications and it can be estimated by computing the disparities between the stereo camera images. For this purpose, SGM is widely used because it sustains a high depth-map accuracy [13]. It consists of three stages [23]: cost generation, cost aggregation, and disparity computation. The census cost values of each pixel p, $C(p,d)$, are first computed in cost generation stage. Then, the disparity of corresponding pixels is obtained by minimizing global cost function over the whole image along eight directions [0°-to-315°] throughout cost aggregation stage. The cost aggregation along path r, $L_r(p,d)$, is described as

$$\begin{aligned} L_r(p,d) = C(p,d) + \min[&L_r(p-r,d), L_r(p-r,d-1) \\ &+P_1, L_r(p-r,d+1)+P_1, \\ &\min_i L_r(p-r,i)+P_2] - \min_k L_r(p-r,k) \end{aligned} \quad (1)$$

with constant penalty values, P_1 and P_2. Penalty for an increment of one in disparity, d, is P_1 and anything greater is P_2.

$L_r(p,d)$ over all of the eight directions are summed as

$$S(p,d) = \sum_r L_r(p,d) \quad (2)$$

where $S(p,d)$ is the aggregation map. Finally, the corresponding disparity at pixel p, D_p, is calculated by searching the minimum value d among the aggregation map

$$D_p = \min_d S(p,d). \quad (3)$$

in the disparity computation stage. Then, the estimated distances are used in intention-prediction.

The proposed intention-prediction algorithm is shown in Fig. 2(b). It has a concurrent neurofuzzy architecture, which consists of recurrent neural network (RNN) and fuzzy inference system (FIS). The neurofuzzy architecture can take advantages from both RNN and FIS, each of which shows great performance in low-level computation and high-level reasoning by mimicking cognition of human brain [24], respectively. RNN provides high prediction accuracy in time-series prediction because its feedback connection addresses the temporal relationship of the inputs by maintaining an internal state [25], and so does it in real-time video processing in which involves spatiotemporal coherence. Moreover, it can provide up to N-step-ahead predictions in trade of degraded accuracy. In contrast, FIS simplifies complex computations by dealing with fuzzy variables (e.g., close and far) that are transformed by its membership functions (MFs). Hence, deploying neurofuzzy architecture greatly reduces computation cost because it does not require heavy number of complex mathematical models [24]. Moreover, the fuzzy parameters can be adapted to different environments, e.g., urban or highway, through online learning, providing a high-level intelligence.

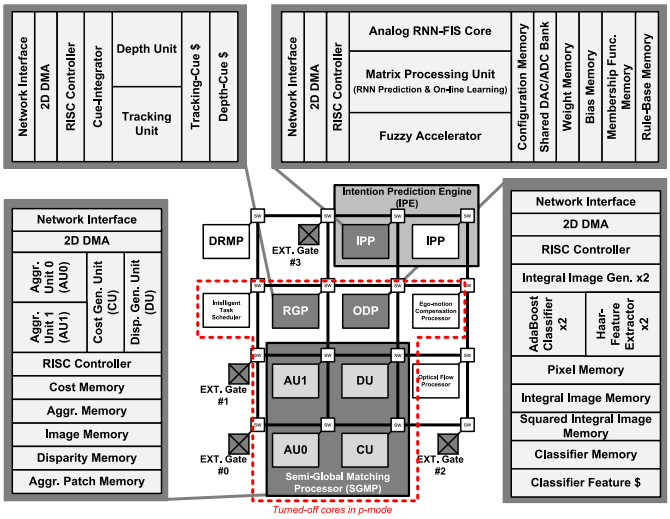


Fig. 3. Overall SoC architecture

In the proposed intention-prediction algorithm, RNN takes sets of input $S_t = \{x, z, \Delta x, \Delta z\}_t$ where x is lateral distance, z is longitudinal distance, and $(\Delta x, \Delta z)$ are their derivatives of the detected object. The input set (S_t-to-S_{t-N}) represents the trajectory history of the detected objects from current to N previous frames. Then it predicts the object's future states for next Q frames, $R_{t+1} \sim R_{t+Q}$, which indicate the predicted trajectories of the detected object. The output becomes the recurrent input set to RNN. For online learning, the actual state S_{t+1} obtained in time $t+1$ becomes the answer for learning R_{t+2}. Then, FIS takes (S_{t-N}-to-R_{t+Q}), as input, fuzzifies them with MFs, and makes decision by using fuzzy rule bases as shown in Fig. 2(b). The fuzzy *IF-THEN* rules are computed by MIN–MAX operations [26]. Finally, the result is defuzzified into the crisp variable that represents the risky level of the detected objects.

III. Chip Architecture

A. Overall Architecture

Fig. 3 shows the overall architecture of the proposed ADAS SoC. It consists of seven accelerators and one IPE, and each of them is dedicated to the different ADAS algorithms in Fig. 2(a). The SoC operates differently in d-mode and p-mode as they have different requirements. The whole processing cores operate in d-mode to achieve high performance and accuracy for safety. Each processing core is mapped to the algorithmic block by its name, while the ROI generation and object tracking are performed in the ROI generation processor (RGP), and ego-motion compensation and 3-D-world mapping are computed in the ego-motion compensation processor. The data resource management processor (DRMP) controls DVFS for each processing core according to its power and clock domain. To reduce much power waste, it turns off some cores by clock gating when they are not necessary. The processing cores are connected via 2-D mesh NoC from [27], which utilizes the congestion-aware flexible router architecture and the intelligent task scheduler to enhance network throughput and energy efficiency.

In p-mode, only the optical flow processor and the IPE operate while other cores are turned off for ultralow-power

consumption, because the exact identification, distance, and 3-D location of the object are not necessary for event detection for record-triggering, as mentioned in Section II. The optical flow processor extracts and clusters optical flow vectors that are directly sent to the IPE. Then, the IPE decides event and outputs surveillance trigger signal.

B. Semiglobal Matching Processor

Among the three stages of SGM mentioned above, the cost aggregation stage is the memory bottleneck because it requires (*Width*) × (*Height*) × (*Disparity Range*) × (bit width) of memory size. For the case of 64-range disparity under wide-360p resolution, the aggregation map requires 14.1 MB, which is too large to be stored on chip. Because the cost aggregation is divided into two scanning substages of downward aggregations ($r \in 0°$-to-$135°$) and upward aggregations ($r \in 180°$-to-$315°$) [13] as shown in Fig. 4(a), 14.1 MB of the downward aggregation map must be stored in the external memory after the first scan stage and fetched back into the chip in the second scan stage to complete the computation of (2). Therefore, the conventional pixel-based processing required 28.1 MB/frame of external bandwidth. Moreover, the left/right consistency check for occlusion handling [23] approximately doubles the required external accesses. Thus, it results in 1.65 GB/s of external bandwidth only for the cost aggregation. To resolve the massive external bandwidth, we propose; 1) tile-based processing; 2) P_2-less data compression; and 3) the task-level pipelined SGMP architecture.

Fig. 4(a) shows the proposed 8 × 8 tile-based processing method. Unlike the conventional pixel-based processing where the complete downward aggregation map is fetched into the chip in the second scan stage, only the first of every eighth row of the whole downward aggregation map for those directions is written-to and read-from the external memory in the proposed tile-based processing. In the second scan stage, it reconstructs the remaining 7 × 8 data by performing downward aggregation within the tile again as in the first scan stage. Therefore, the proposed tile-based processing obtains the whole 8 × 8 data without any loss. As a result, the external bandwidth is reduced by 62.3% in trade of 43.8% increment in on-chip computation.

In addition to the tile-based processing, the P_2-less compression is proposed to much reduce the external access. Substituting (4) into (1) leads $L_r(p,d)$ equal to (5), where $T_r(p,d)$ indicates the minimal difference with constants (P_1, P_2). The left diagram in Fig. 4(b) shows the intermediate data distribution of $L_r(p,d)$ and $T_r(p,d)$. The values of $T_r(p,d)$ have limited range from 0 to P_2, where P_2 is 100 in this case. Because >70% of it is saturated to P_2, only the unsaturated values ($\neq P_2$) are used to reduce the amount of data to be stored. The right diagram in Fig. 4(b) shows the proposed P_2-less compression. Among the original aggregation data (A_{original}), the unsaturated values are stored as $A_{\text{compressed}}$ together with A_{tag} that indicates the corresponding location of the values in A_{original}. The original data can be completely recovered by a simple masking operation between $A_{\text{compressed}}$ and A_{tag}. As Fig. 4(c) depicts, exploiting the proposed P_2-less compression with the tile-based processing reduces the external bandwidth and the required on-chip memory size for

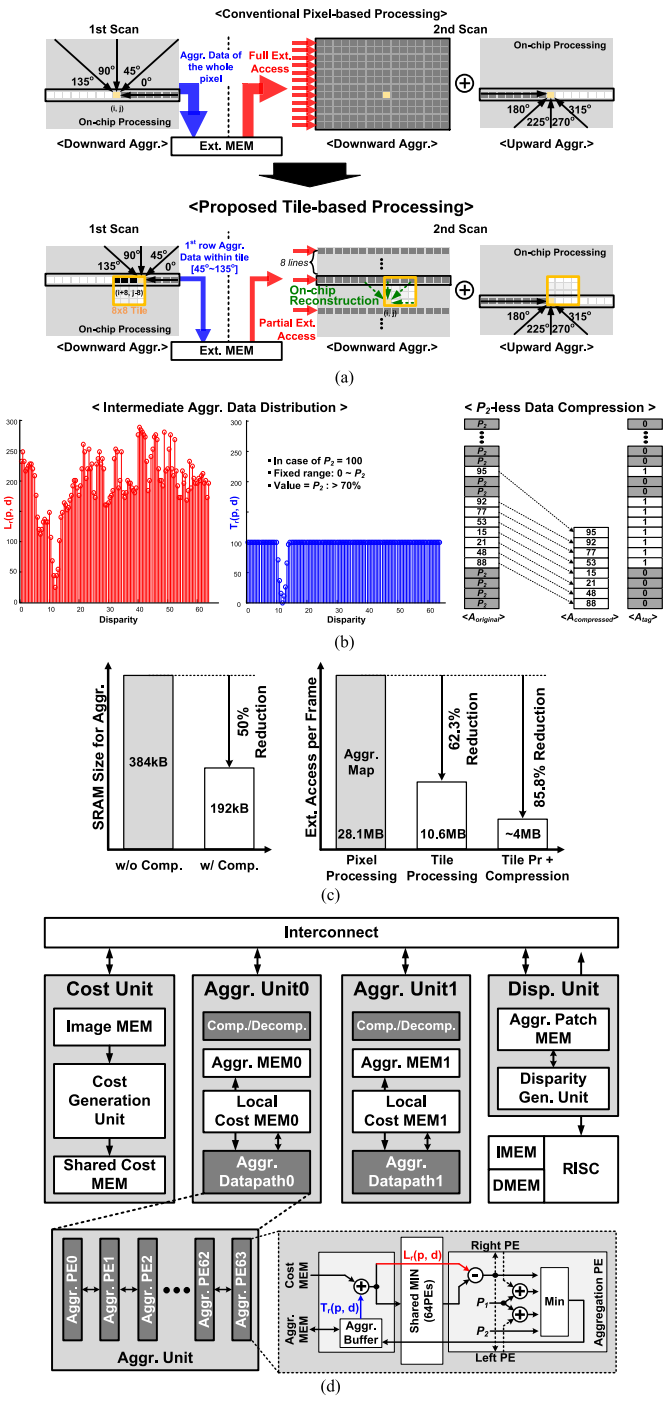


Fig. 4. (a) Proposed tile-based SGM processing. (b) Internal data distribution of SGM and the proposed data compression. (c) Performance improvement of the SGMP. (d) Hardware architecture of three-stage task-level pipelined SGMP.

aggregation by 85.8% and 50%, respectively:

$$\begin{aligned} T_r(p,d) &\equiv T_r(p,d,P_1,P_2) \\ &= \min[L_r(p-r,d), L_r(p-r,d-1)+P_1, \\ &\quad L_r(p-r,d+1)+P_1, \min_i L_r(p-r,i)+P_2] \\ &\quad - \min_k L_r(p-r,k) \qquad (4) \\ L_r(p,d) &= C(p,d)+T_r(p,d). \qquad (5) \end{aligned}$$

TABLE I
SGM PERFORMANCE COMPARISON TABLE

	ISVC 2008 [28]	CVPRW 2010 [29]	ICVS 2009 [13]	IV 2013 [30]	This work
Implementation	GPU	CPU	FPGA	GPU	ASIC
Resolution	320x240	640x320	340x200	640x480	640x360
Processing Type	Pixel-based	Pixel-based	Pixel-based	Pixel-based	Tile-based
Compression	-	-	-	-	P2-less Comp. (Lossless)
Disparity Range	64	128	64	64	64
Latency [ms]	76	224	40	85	25
Performance [px/s]	64.7 M	117 M	109 M	231 M	590 M

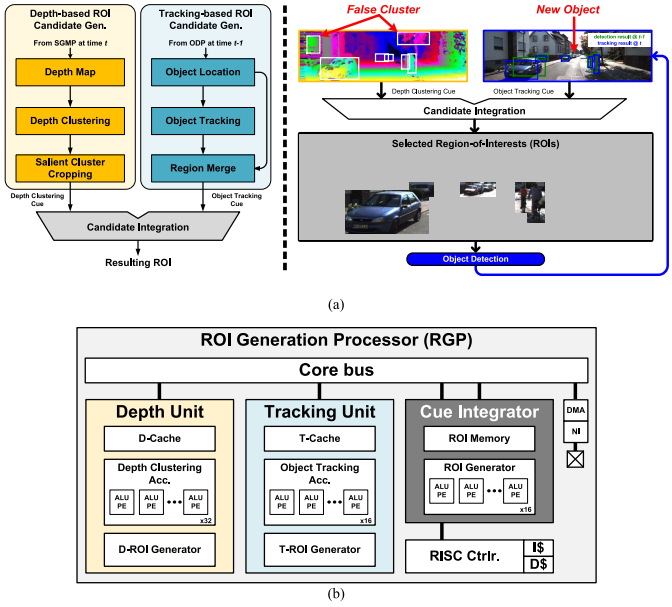


Fig. 5. (a) ROI generation algorithm flowchart and an example of resulting ROI with KITTI database. (b) Hardware architecture of the RGP.

Fig. 4(d) explains the hardware architecture of the proposed SGMP. It has three-stage task-level pipelined architecture to increase throughput. The census cost generated in the cost unit is stored in the shared cost memory, and then distributed to the local cost memories of each aggregation unit. In the aggregation datapath, 64 PEs are integrated to calculate 64-range disparities in parallel, and two aggregation units execute the upward and the downward cost aggregation simultaneously to reduce processing time. As a result, the proposed SGMP achieves 590 Mp/s performance. Table I shows the performance comparisons to previous SGM works [13], [28]–[30]. The proposed SGMP achieves 70.6% reduction in latency while improving the performance by 2.55× than the state of the art [30].

C. ROI Generation Processor and Object Detection Processor

The proposed RGP exploits hierarchical information by integrating both pixelwise depth clustering cue and objectwise tracking cue for precise ROI generation as shown in Fig. 5(a). With the depth map generated from the SGMP, the RGP

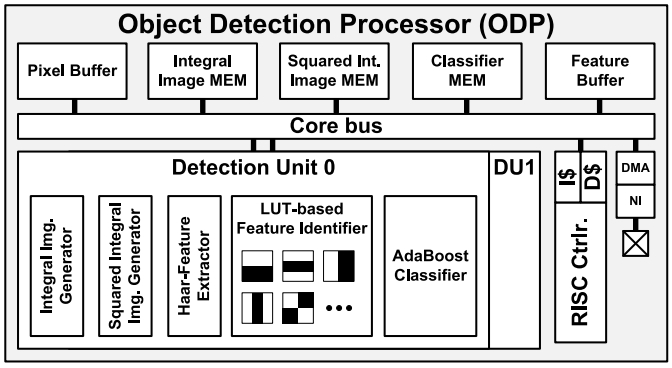


Fig. 6. Object detection processor architecture.

generates the depth clustering cue ROI candidates by clustering depth value of neighboring pixels with $< \pm\varepsilon$ differences and cropping the salient clusters. However, dealing with pixelwise depth information results in incorrect clusters due to the limited depth range, light condition, object size, and occlusion. As in Fig. 5(a), the depth clustering cue contains false-positive clusters (*window, tree*) where interested objects (*vehicle, pedestrian*) are not present, as well as misdetection of the occluded car. At the same time, the RGP receives the regions of objects detected from previous frame and performs Kalman filtering [31] to track the regions to estimate the location of the regions at current frame. Unlike the depth clustering cue, the object tracking cue can detect ROI for occluded objects, but it fails for newly appeared objects. Therefore, the RGP integrates the both cues to overcome the drawbacks of each and generates precise ROIs (see Section VI). Then, objects are searched only within the ROIs to reduce computation.

Fig. 5(b) shows the hardware architecture of the proposed RGP. Because depth clustering should be performed over the whole image, the depth unit contains 32 parallel SIMD PEs of depth clustering accelerator for fast operation. The resulting depth clustering cue is stored in D-cache. The tracking unit accelerates Kalman filtering with 16 parallel SIMD PEs and stores the result in T-cache. Then, the cue integrator generates the final ROIs using both cues.

Fig. 6 shows the object detection processor architecture, which consists of memory banks, RISC controller, and two detection units. The detection unit has integral image generator, squared integral image generator, Haar-Feature [32] extractor, feature identifier, and AdaBoost [32] classifier. It takes ROIs as input and generates both integral and squared integral images. Then, the features are extracted, identified, and classified as an object. To hide off-chip access latency, the classifier features of further stages are fetched into feature buffer from the external memory before they are used. The object detection processor has MIMD architecture to enhance throughput, which means each detection unit is operated with different ROIs simultaneously. With the proposed RGP and the object detection processor, the computation for object detection is reduced by 86.2% compared with it without ROIs, and 20 objects are detected in 325 μs.

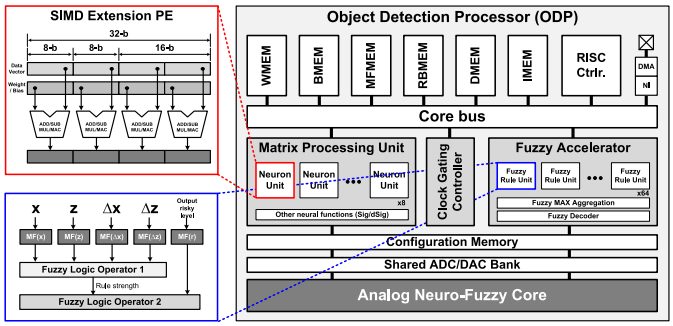


Fig. 7. Hardware architecture of the intention prediction processor in the IPE.

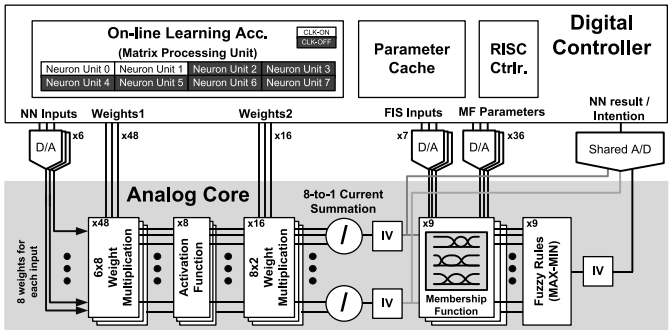


Fig. 8. Mixed-mode configuration of the intention prediction processor.

D. Mixed-Mode Intention Prediction Engine (IPE)

The proposed IPE has mixed-mode architecture for dual-mode operation to fulfill different requirements in different operating modes as mentioned above. The IPE operates in fully digital configuration for fast and robust intention-prediction in d-mode. On the other hand, it operates in mixed-mode configuration for ultralow-power consumption in p-mode.

E. Fully Digital Configuration With Massively Parallel PE Architecture for D-Mode

The detailed architecture of intention-prediction processor, which comprises the IPE, is shown in Fig. 7. In d-mode, only digital parts are fully used and the analog core is turned off. The intention prediction processor consists of massively parallel matrix processing unit and fuzzy accelerator, memory banks, RISC controller, and clock-gating controller. Since the online learning and feed-forward prediction of an RNN is mainly composed of MAC operations, a dedicated matrix processing unit is used to accelerate RNN. It consists of eight neuron units of SIMD extension PE that can selectively perform 32 8-b, 16 16-b, or 8 32-b operations in parallel. It truncates the results when needed.

FIS operation is accelerated by the fuzzy accelerator. It consists of 64 fuzzy rule units, fuzzy aggregator, and fuzzy decoder to compute 64 different fuzzy rules in parallel according to various road conditions. The fuzzy rule unit contains LUTs for different MFs and fuzzy logic operators. The example categories of MFs are shown in Fig. 2(b). After the rule strength calculation, it computes fuzzy logic operation with the output risky level MF. The fuzzy IF-THEN rules stored in the RBMEM are used for the inference, and the fuzzy decoder defuzzifies the resulting risky level. Thanks to the massively parallel matrix processing unit and fuzzy accelerator, the IPE predicts the intention of 20 objects within 1.24 ms.

F. Mixed-Mode Configuration of the IPE for P-Mode

For ultralow-power consumption in p-mode, the IPE operates in mixed-mode configuration as shown in Fig. 8. It consists of analog core, shared ADC/DACs, and digital controller. The analog core is designed with current-mode circuits to take advantages of ultralow-power consumption as well as natural parallelism of neural networks due to the parallel processing of current-mode and simple summation by Kirchhoff's law [33]. It performs feed-forward neurofuzzy prediction replacing complex computations such as MAC and LUT operations, and the digital controller performs the online learning. To save power, clock-gating controller turns off the fuzzy accelerator and six neuron units of the matrix processing unit, and only two neuron units perform the online learning of the analog core. The RISC controller assigns the digital parameters, i.e., weights and MFs, to DACs to set the parameters in the analog core. The output values of RNN and FIS are fed into the digital controller through a shared ADC to save area. The weight multiplication utilizes current-steering multiplying-DACs instead of current multipliers. Thus, it reduces the number of DACs for weight multiplication by 56.6%.

The detailed circuits of the analog core and their measured waveforms are shown in Fig. 9. In the analog core, sigmoid function is used as the activation function for neurons. Fig. 9(a) shows the sigmoid circuit. It uses a simple sigmoidal differential amplifier to reduce area, and it consumes <1 μA to save power. The shape of sigmoid function is controlled by changing resistance between the differential input pair, R_{var}.

Fig. 9(b) shows the highly controllable transconductance MF circuit. It changes MF shape with five control parameters: switched-current source I_s, two bias voltages V_{ref}, PMOS switch *B[4:0]*, and MUX p where each of them determines height, width and center, slope, and up/down phase of the MF, respectively. The bottom of Fig. 9(b) shows the measured controllability waveforms. V_{ref1} and V_{ref2} shift the transient points of each corner independently, and thus, a pair of (V_{ref1}, V_{ref2}) determines the center and width of MFs. The slope of transient curve is controlled by *B[0:4]* by differing the g_m, and I_s determines the height of MF. With this highly controllable MF circuit, various shapes of MF can be made as depicted as *small*, *medium*, and *big* in Fig. 9(b).

Fig. 9(c) shows the MIN circuit and its measured waveform. The circuit compares two inputs by their strength. When $I_y < I_x$, current through M1 becomes positive so $I_{\text{OUT}} = I_y$. When $I_y > I_x$, I_b becomes 0 so $I_{\text{OUT}} = I_x$. The MIN circuit successfully works with small current <1 μA. The analog core adopts MAX circuit from [34] and it is also designed to operate with <1 μA as in Fig. 9(d). The final output is fuzzy variable and it is defuzzified in the digital controller.

The maximal current in the whole analog core is <8 μA, and its power consumption is 156 μW with $<$ 50 ms network latency. With the mixed-mode circuit implementation, the

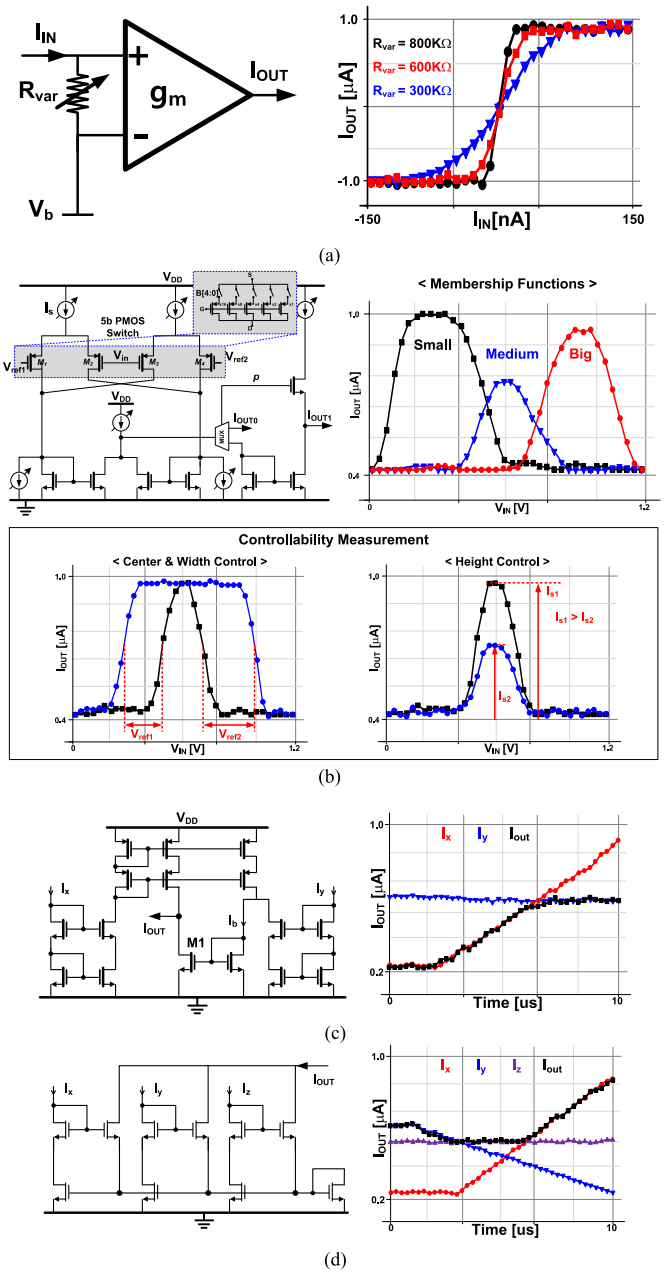


Fig. 9. Current-mode circuit implementations and measured waveforms of (a) sigmoidal activation function, (b) MF, (c) fuzzy minimum function, and (d) fuzzy maximum function.

total area and power consumption of the IPE are reduced by 64.1% and 39.1%, respectively, compared with the fully digital implementation.

IV. Chip Control With Workload-Prediction DVFS in D-Mode

The DRMP is used to control DVFS for high energy efficiency. The SoC has three dynamic power-clock domains as it involves with three different data hierarchies: 1) high pixel-parallel processing; 2) moderate pixel-/task-parallel processing; and 3) complex task-parallel processing. The SGMP and the optical flow processor have high pixel-level parallelism due to the massively repeated pixel-intensive computation, hence

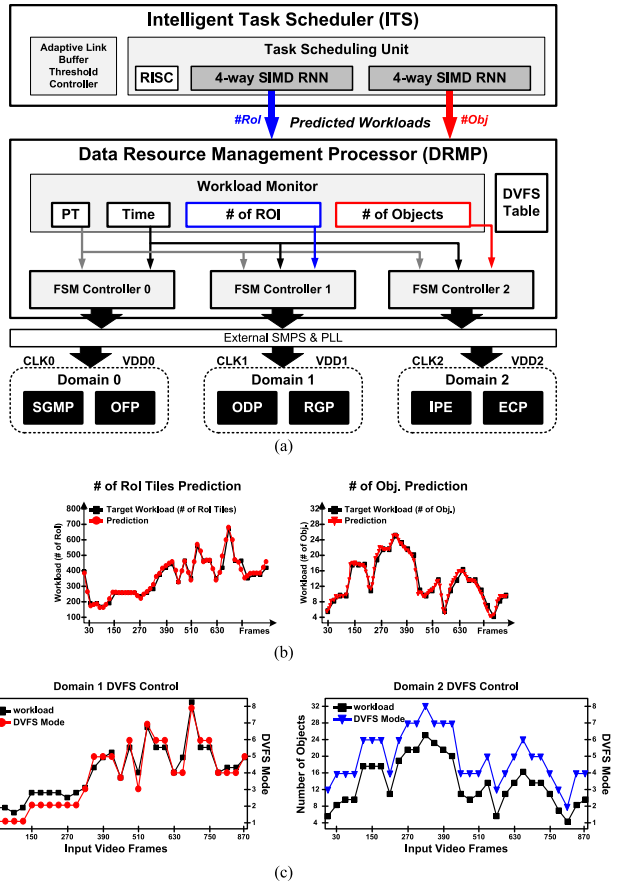


Fig. 10. (a) DRMP architecture for workload-prediction DVFS. (b) Workload prediction results of the intelligent task scheduler used for DVFS control. (c) Measured DVFS control result of domain 1 and 2 with respect to the corresponding predicted workloads in (b).

they share DVFS *domain 0*. The RGP and the object detection processor involve with moderate pixel-/task-parallel processing and share *domain* 1 of which the workload is dependent on the number of ROIs. Finally, the IPE and the ego-motion compensation processor belong to *domain* 2 with the complex task-parallel processing, where its workload varies according to the number of objects. Fig. 10(a) shows the proposed DVFS system. The intelligent task scheduler predicts the workload of *domain* 1 and 2 of next frame, namely, the number of ROIs and objects, respectively. The prediction results are fed into the DRMP that consists of workload monitor, DVFS table, and controllers that decide power and clock modes for each domain from DVFS table.

The *domain* 0 monitors power, thermal, and processing time headroom while *domain* 1 and 2 monitor additional workloads. Fig. 10(b) shows the predicted workloads of the intelligent task scheduler that are used for controlling each domain, and the DRMP adjusts DVFS mode in advance based on the predictions. Fig. 10(c) shows the measurement results of the workload-prediction DVFS for *domain* 1 and 2. With the proposed DRMP, the average power consumption of the SoC is reduced by 36.2%. The total power reduction ratio varies according to the driving environment due to the workload dependence of each domain. In rural area where the number

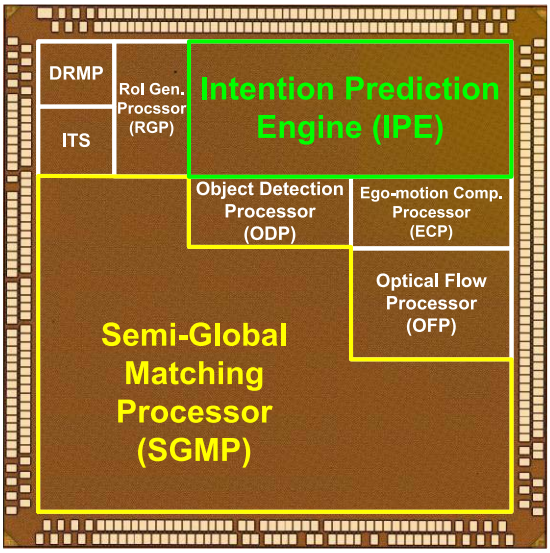


Fig. 11. Chip micrograph.

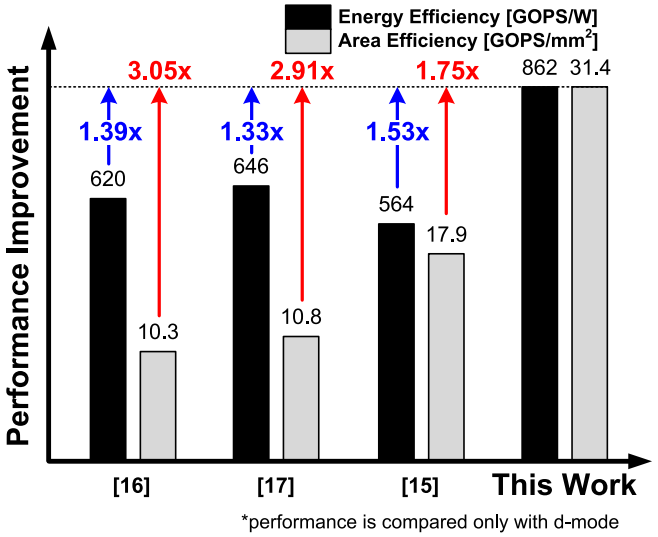


Fig. 12. Comparison of energy efficiency and area efficiency.

TABLE II
CHIP SPECIFICATION

*Temp=25℃		Driving-mode	Parking-mode
Process		65nm 1P8M Logic CMOS	
Chip Size		4.0 x 4.0 mm^2	
Gate Count		5.42M	
SRAM		493 kByte	
Supply Voltage	Nominal	1.2 V	0.8 V (Digital) 1.2 V (Analog)
	DVFS	0.65-1.2 V	
Clock Frequency	Nominal	250 MHz	20 MHz
	DVFS	50-250 MHz	
Power	Average	330 mW	0.984 mW
	Peak	582 mW	1.96 mW
Peak Performance		502 GOPS	1.80 GOPS
Energy Efficiency		862 GOPS/W	918 GOPS/W
Area Efficiency		31.4 GOPS/mm^2	0.113 GOPS/mm^2

of objects ranges from 0 to 2, the SoC achieves 53.1% power reduction.

V. IMPLEMENTATION

The chip micrograph and specification of the proposed ADAS SoC are shown in Fig. 11 and Table II, respectively. It is fabricated in 65-nm CMOS technology and occupies 16 mm^2. It integrates 5.42-M gates and 494 kB of on-chip SRAM, and it operates at 30 frames/s throughput with up to 720p stereo images. The average power consumption in d-mode is 330 mW running at 250 MHz and 1.2-V supply voltage, while that of p-mode is 984 μW running at 20 MHz and 0.8 V. Fig. 12 shows the performance improvement. The SoC achieves 862 GOPS/W of energy efficiency and 31.4 GOPS/mm^2 of area efficiency, which are 1.53× and 1.75× improvements over the state of the art [14]. Table III is the comparison table. Unlike previous works were only capable of object detection, this paper also provides intelligence in different operation modes. It provides intention-prediction as well as object detection in d-mode, and intelligent surveillance record-triggering function in p-mode. Moreover, it is the first chip implementation of real-time SGM, while others were only capable of local stereo matching algorithms.

Fig. 13 shows the system implementation. The stereo cameras are mounted behind the windshield and the proposed system is located on the dashboard. Camera inputs are delivered to FPGA by PCIe network. The FPGA converts the input images to corresponding NoC packet protocol and controls the overall chip operation.

VI. SYSTEM EVALUATION

The SGMP computation is the most burdensome among the processors because it needs to calculate pixel-by-pixel over the whole image with short processing time to fulfill 30 frames/s throughput. Therefore, only the SGMP uses one-fourth downsized images while the other processors operate with the original images for high object detection accuracy, even though the maximal performance of SGMP is 40 frames/s with 640 × 360 resolution as shown in Table I. However, using downsized SGM does not affect object detection and intention-prediction accuracies, because SGM is only used for ROI generation and distance (z) estimation. Fig. 14(a) shows the original and downsized SGM results of Daimler-Stereo Database [35] with VGA resolution. As in the figure, using downsized image does not alter the actual distance to the object, but the object boundary is blurred. This can result in wide ROI regions that increases search time of object detection, but it is compensated by integrating object tracking cue as shown in Fig. 14(b). As described in Section III, depth clustering cue results in false-positive ROIs (*poles, sign, tree*) while tracking cue fails to provide ROI region for newly appeared objects. The proposed ROI generation successfully

TABLE III
HARDWARE COMPARISON TABLE WITH STATE-OF-THE-ART ADAS PROCESSORS

	ISSCC 2012 [15]	ISSCC 2013 [16]	ISSCC 2015 [14]	This Work	
				Driving-mode	Parking-mode
Functions	Object Detection	Object Detection	Object Detection	Object Detection + Intention Prediction	Surveillance Record-Triggering
Resolution	N/A	HD (720p)	HD (720p)	HD (720p)	
Throughput	30 fps	30 fps	20 fps	30 fps	
Process	40 nm	130 nm	40 nm	65 nm	
Gate Count	N/A	1.8M	50.3M	5.42M	
SRAM [kB]	N/A	200	8460	493	
Area [mm^2]	44.5	25	106	16	
Operating Frequency [MHz]	266 / 180	200	266 / 180	250	20
Peak Power [mW]	749	420	3370	582	1.96
Performance [GOPS]	464	271	1900	502	1.80
Energy Efficiency [GOPS/W]	620	646	564	862	918
Area Efficiency [GOPS/mm^2]	10.3	10.8	17.9	31.4	0.1125
Stereo Matching	Local + Sparse	Not Supported	Local + Sparse	Global + Dense (SGM)	Not Used
Intelligence	X	X	X	O	O

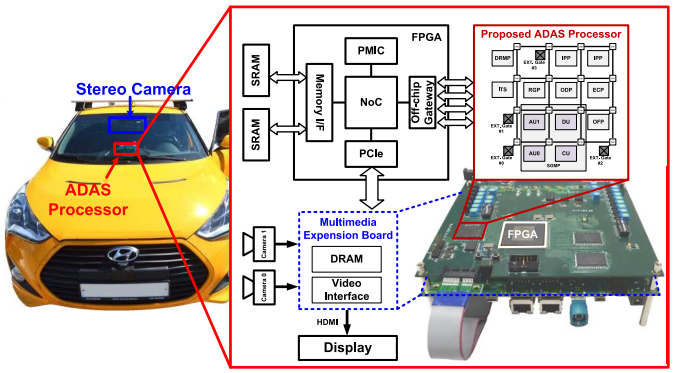


Fig. 13. System implementation.

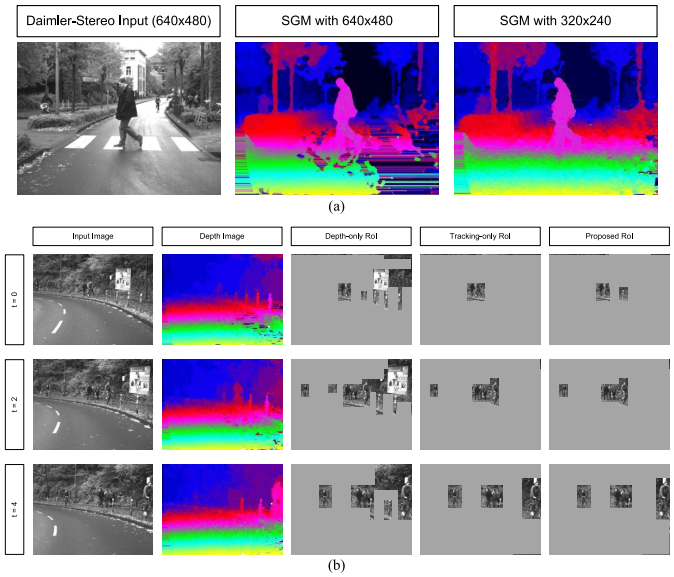


Fig. 14. (a) Comparisons of SGM depending on the image resolution. (b) Performance comparison of tracking-only, depth-only, and the proposed ROI generation.

achieves 91.6% of ROI generation accuracy by combining both cues.

The proposed system is tested under various databases of manually recorded campus database as well as the most widely used ADAS databases, KITTI [36] and Daimler-Stereo databases. Fig. 15(a) shows the object detection and intention-prediction results of each database. Unlike the object recognition returns every detected objects, only pedestrians and vehicles that pop out on the driving lane or that are getting closer to the car get attention in the proposed intention-prediction.

Fig. 15(b) shows the measured RNN result used for intention-prediction from KITTI. In reality, the motion of preceding object can be fluctuated in small distance even though it is following the driving line. Therefore, predicting risky level of an object by observing the object's state of just one frame ahead is not sufficient to predict its global motion, such as lane change. Utilizing RNN facilitates N-step-ahead prediction in trade of degraded prediction accuracy. Its one-step-ahead prediction tends to be sensitive to dynamic changes while it has high prediction accuracy. On the other hand, N-step-ahead prediction becomes less sensitive but the accuracy is degraded. To increase the intention-prediction accuracy

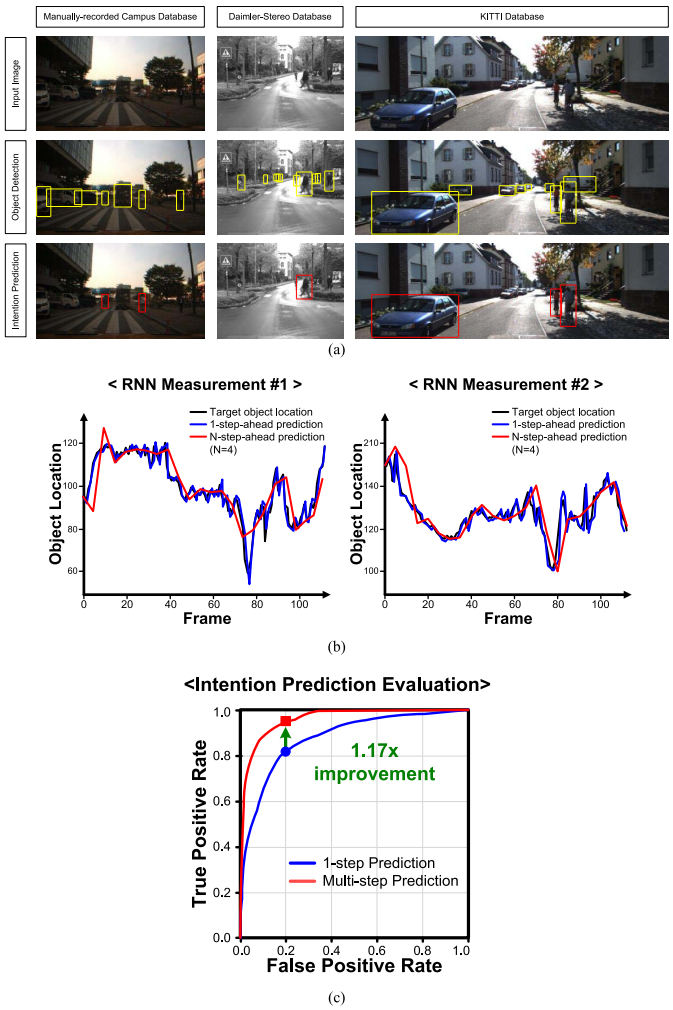


Fig. 15. (a) Evaluation results of object detection and intention prediction with different databases. (b) Measurement result of RNN prediction for different object locations. (c) Performance evaluation of the proposed intention-prediction algorithm.

and predict risky level of an object based on interpretation of its local to global motion prediction, not only the one-step-ahead but up to N-step-ahead predictions are utilized as Fig. 15(c) depicts. As a result, the intention-prediction accuracy is improved by 17% at false-positive rate of 0.2, and it achieves 98.1% accuracy.

VII. Conclusion

In this paper, we present an intelligent dual-mode ADAS SoC for a smart automotive black box system. To enable real-time SGM by reducing external memory bandwidth, we proposed a tile-based processing and P_2-less data compression with three-stage task-level pipelined SGMP. For intelligent behavior analysis, an intention-prediction algorithm and its dedicated hardware accelerator are proposed. The hybrid architecture of the IPE operates in fully digital configuration with massively parallel PEs for high-speed operation in d-mode. For ultralow-power consumption in p-mode, it operates in mixed-mode configuration. As a result, the proposed ADAS processor achieved 30-frames/s system throughput with 720p stereo camera, with 98.1% of intention-prediction accuracy. It achieves 862 GOPS/W of energy efficiency and 31.4 GOPS/mm^2 of area efficiency, which are 1.53× and 1.75× improvements compared with the state-of-the-art vision processors for ADAS applications.

References

[1] E. Dagan, O. Mano, G. P. Stein, and A. Shashua, "Forward collision warning with a single camera," in *Proc. IEEE Intell. Veh. Symp.*, Jun. 2004, pp. 37–42.

[2] J. McCall and M. M. Trivedi, "Video-based lane estimation and tracking for driver assistance: Survey, system, and evaluation," *IEEE Trans. Intell. Transp. Syst.*, vol. 7, no. 1, pp. 20–37, Mar. 2006.

[3] A. Vahidi and A. Eskandarian, "Research advances in intelligent collision avoidance and adaptive cruise control," *IEEE Trans. Intell. Transp. Syst.*, vol. 4, no. 3, pp. 143–153, Sep. 2003.

[4] K. D. Kusano and H. C. Gabler, "Safety benefits of forward collision warning, brake assist, and autonomous braking systems in rear-end collisions," *IEEE Trans. Intell. Transp. Syst.*, vol. 13, no. 4, pp. 1546–1555, Dec. 2012.

[5] S. Kato, E. Takeuchi, Y. Ishiguro, Y. Ninomiya, K. Takeda, and T. Hamada, "An open approach to autonomous vehicles," *IEEE Micro*, vol. 35, no. 6, pp. 60–68, Nov./Dec. 2015.

[6] J. Ziegler *et al.*, "Making Bertha drive—An Autonomous journey on a historic route," *IEEE Intell. Transp. Syst. Mag.*, vol. 6, no. 2, pp. 8–20, Summer 2014.

[7] A. Geiger *et al.*, "Team Annieway's entry to the 2011 grand cooperative driving challenge," *IEEE Trans. Intell. Transp. Syst.*, vol. 13, no. 3, pp. 1008–1017, Sep. 2012.

[8] K. Jo, J. Kim, D. Kim, C. Jang, and M. Sunwoo, "Development of autonomous car—Part II: A case study on the implementation of an autonomous driving system based on distributed architecture," *IEEE Trans. Ind. Electron.*, vol. 62, no. 8, pp. 5119–5132, Aug. 2015.

[9] J. Wei, J. M. Snider, J. Kim, J. M. Dolan, R. Rajkumar, and B. Litkouhi, "Towards a viable autonomous driving research platform," in *Proc. IEEE Intell. Veh. Symp.*, Jun. 2013, pp. 763–770.

[10] H. Abbott and D. Powell, "Land-vehicle navigation using GPS," *Proc. IEEE*, vol. 87, no. 1, pp. 145–162, Jan. 1999.

[11] F. Forster, "Heterogeneous processors for advanced driver assistance systems," *Atz Elektronik Worldwide*, vol. 9, no. 1, pp. 14–18, Feb. 2014.

[12] F. Stein, "The challenge of putting vision algorithms into a car," in *Proc. IEEE Comput. Soc. Conf. CVPR Workshops*, Jun. 2012, pp. 89–94.

[13] S. Gehrig, F. Eberli, and T. Meyer, "A real-time low-power stereo vision engine using Semiglobal matching," in *Proc. Int. Conf. Comput. Vis. Syst. (ICVS)*, 2009, pp. 134–143.

[14] J. Tanabe *et al.*, "A 1.9TOPS and 564GOPS/W heterogeneous multi-core SoC with color-based object classification accelerator for image-recognition applications," in *IEEE Int. Solid-State Circuits Conf. (ISSCC) Dig. Tech. Papers*, Feb. 2015, pp. 1–3.

[15] Y. Tanabe *et al.*, "A 464GOPS 620GOPS/W heterogeneous multi-core SoC for image-recognition applications," in *IEEE Int. Solid-State Circuits Conf. (ISSCC) Dig. Tech. Papers*, Feb. 2012, pp. 222–223.

[16] J. Park *et al.*, "A 646GOPS/W multi-classifier many-core processor with cortex-like architecture for super-resolution recognition," in *IEEE Int. Solid-State Circuits Conf. Dig. Tech. Papers (ISSCC)*, Feb. 2013, pp. 168–169.

[17] M. Liebner, F. Klanner, M. Baumann, C. Ruhhammer, and C. Stiller, "Velocity-based driver intent inference at Urban intersections in the presence of preceding vehicles," *IEEE Intell. Transp. Syst. Mag.*, vol. 5, no. 2, pp. 10–21, Summer 2013.

[18] D. Kasper *et al.*, "Object-oriented Bayesian networks for detection of lane change maneuvers," *IEEE Intell. Transp. Syst. Mag.*, vol. 4, no. 3, pp. 19–31, Fall 2012.

[19] A. Barth and U. Franke, "Tracking oncoming and turning vehicles at intersections," in *Proc. IEEE Int. Conf. Intell. Transp. Syst.*, Sep. 2010, pp. 861–868.

[20] C. Hermes, C. Wohler, K. Schenk, and F. Kummert, "Long-term vehicle motion prediction," in *Proc. IEEE Intell. Veh. Symp.*, Jul. 2009, pp. 652–657.

[21] E. Meissner and G. Richter, "The challenge to the automotive battery industry: The battery has to become an increasingly integrated component within the vehicle electric power system," *J. Power Sources*, vol. 144, no. 2, pp. 438–460, 2005.

[22] *BlackVue DR550GW-2CH Car Camera Tech Specs*. accessed on Oct. 26, 2016. [Online]. Available: http://blackvueshop.co.uk/product/dr550gw-wifi-gps/

[23] H. Hirschmüller, "Stereo processing by semiglobal matching and mutual information," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 30, no. 2, pp. 328–341, Feb. 2008.

[24] C.-F. Juang and C.-T. Lin, "An online self-constructing neural fuzzy inference network and its applications," *IEEE Trans. Fuzzy Syst.*, vol. 6, no. 1, pp. 12–32, Feb. 1998.

[25] C. L. Giles, S. Lawrence, and A. C. Tsoi, "Noisy time series prediction using recurrent neural networks and grammatical inference," *Mach. Learn.*, vol. 44, no. 1, pp. 161–183, 2001.

[26] E. H. Mamdani, "Application of fuzzy logic to approximate reasoning using linguistic synthesis," *IEEE Trans. Comput.*, vols. C–26, no. 12, pp. 1182–1191, Dec. 1977.

[27] K. Lee, J. Park, I. Hong, and H.-J. Yoo, "Intelligent task scheduler with high throughput NoC for real-time mobile object recognition SoC," in *Proc. IEEE Eur. Solid-State Circuits Conf. (ESSCIRC)*, Sep. 2015, pp. 100–103.

[28] I. Ernst and H. Hirschmüller, "Mutual information based Semiglobal stereo matching on the GPU," in *Proc. Int. Symp. Adv. Vis. Comput. (ISVC)*, 2008, pp. 228–239.

[29] S. K. Gehrig and C. Rabe, "Real-time Semiglobal matching on the CPU," in *Proc. IEEE Comput. Soc. Conf. Comput. Vis. Pattern Recognit. Workshops (CVPRW)*, Jun. 2010, pp. 85–92.

[30] M. Michael, J. Salmen, J. Stallkamp, and M. Schlipsing, "Real-time stereo vision: Optimizing Semiglobal matching," in *Proc. IEEE Intell. Veh. Symp.*, Jun. 2013, pp. 1197–1202.

[31] D. Comaniciu, V. Ramesh, and P. Meer, "Kernel-based object tracking," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 25, no. 5, pp. 564–577, May 2003.

[32] P. Viola and M. J. Jones, "Robust real-time face detection," *Int. J. Comput. Vis.*, vol. 57, no. 2, pp. 137–154, 2004.

[33] M. Yildiz, S. Minaei, and I. C. Goknar, "A CMOS classifier circuit using neural networks with novel architecture," *IEEE Trans. Neural Netw.*, vol. 18, no. 6, pp. 1845–1850, Nov. 2007.

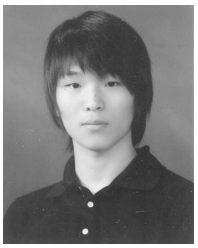
[34] J. L. Huertas, S. Sanchez-Solano, I. Baturone, and A. Barriga, "Integrated circuit implementation of fuzzy controllers," *IEEE J. Solid-State Circuits*, vol. 31, no. 7, pp. 1051–1058, Jul. 1996.

[35] C. Keller, M. Enzweiler, and D. M. Gavrila, "A new benchmark for stereo-based pedestrian detection," in *Proc. IEEE Intell. Veh. Symp.*, Jun. 2011, pp. 691–696.

[36] A. Geiger, P. Lenz, and R. Urtasun, "Are we ready for autonomous driving? The KITTI vision benchmark suite," in *Proc. IEEE Comput. Vis. Pattern Recognit. (CVPR)*, Jun. 2012, pp. 3354–3361.

Kyuho Jason Lee (S'12) received the B.S. and M.S. degrees in electrical engineering from Korea Advanced Institute of Science and Technology, Daejeon, South Korea, in 2012 and 2014, respectively, where he is currently pursuing the Ph.D. degree.

His current research interests include the development of analog/digital mixed-mode neural network system-on-a-chip (SoC) design, object matching processor and its algorithm for computer vision, energy-efficient network-on-chip-based SoC design for mobile devices, and intelligent vision processor for advanced driver assistance system.

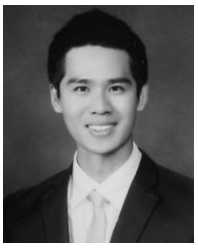


Kyeongryeol Bong (S'12) received the B.S. and M.S. degrees in electrical engineering from Korea Advanced Institute of Science and Technology, Daejeon, South Korea, in 2012 and 2015, respectively, where he is currently pursuing the Ph.D. degree.

His current research interests include intelligent vision system-on-a-chip using parallel architecture and functional CMOS image sensors.

Changhyeon Kim (S'16) received the B.S. degree in electrical engineering from the Korea Advanced Institute of Science and Technology, Daejeon, South Korea, in 2014, where he is currently pursuing the M.S. degree in electrical engineering.

His current research interests include low-power vision system-on-chip design using CMOS image sensors.



Jaeeun Jang (S'13) received the B.S. and M.S. degrees in electrical engineering from the Korea Advanced Institute of Science and Technology, Daejeon, South Korea, in 2014 and 2016, where he is currently pursuing the Ph.D. degree.

He was involved with developing a low-power wireless transceiver and low-power sensor front-end. His current research interests include low-power transceiver design for body-area-networks and low-power biomedical system-on-chip design.

Kyoung-Rog Lee (S'15) received the B.S. degree in electrical engineering from the Korea Advanced Institute of Science and Technology, Daejeon, South Korea, in 2015, where he is currently pursuing the M.S. degree.

His current research interests include low-power bio-medical system-on-a-chip (SoC) for wearable healthcare system and body channel communication SoC for low-power communication system.

Jihee Lee (S'15) received the B.S. degree in electrical engineering from the Korea Advanced Institute of Science and Technology, Daejeon, South Korea, in 2015, where she is currently pursuing the M.S. degree.

Her current research interests include high-efficientcy wireless power transfer system-on-a-chip (SoC) for wearable healthcare system and body channel communication SoC for low-power communication system.

Gyeonghoon Kim (S'10) received the B.S., M.S., and Ph.D. degrees in electrical engineering from Korea Advanced Institute of Science and Technology, Daejeon, South Korea, in 2009, 2011, and 2015, respectively.

His current research interests include low-power digital processors with dynamic resource management for computer vision and network-on-chip based system-on-a-chip-design.

Hoi-Jun Yoo (M'95–SM'04–F'08) received the B.S. degree in electronics engineering from Seoul National University, Seoul, South Korea, in 1983, and the M.S. and Ph.D. degrees in electrical engineering from the Korea Advanced Institute of Science and Technology (KAIST), Daejeon, South Korea, in 1985 and 1988, respectively.

From 2001 to 2005, he was the Director with the Korean System Integration and IP Authoring Research Center, South Korea. From 2003 to 2005, he was a full-time Advisor to the Korean Ministry of Information and Communication, South Korea, and the National Project Manager for system-on-a-chip and computer. In 2007, he founded the System Design Innovation and Application Research Center, KAIST. Since 1998, he has been with the Department of Electrical Engineering, KAIST, where he is currently a Full Professor. He has coauthored *DRAM Design* (Hongrung, 1996), *High Performance DRAM* (Sigma, 1999), *Future Memory: FRAM* (Sigma, 2000), *Networks on Chips* (Morgan Kaufmann, 2006), *Low-Power NoC for High-Performance SoC Design* (CRC, 2008), *Circuits at the Nanoscale* (CRC, 2009), *Embedded Memories for Nano-Scale VLSIs* (Springer, 2009), *Mobile 3D Graphics SoC from Algorithm to Chip* (Wiley, 2010), *Bio-Medical CMOS ICs* (Springer, 2011), *Embedded Systems* (Wiley, 2012), and *Ultra-Low-Power Short-Range Radios* (Springer, 2015). His current research interests include computer vision system-on-a-chip, body-area networks, and biomedical devices and circuits.

Dr. Yoo has been serving as the General Chair of the Korean Institute of Next Generation Computing since 2010. He was a member of the Executive Committee of ISSCC, the Symposium on Very Large Scale Integration, and A-SSCC, the TPC Chair of A-SSCC 2008 and ISWC 2010, an IEEE Distinguished Lecturer from 2010 to 2011, the Far East Chair of ISSCC from 2011 to 2012, the Technology Direction Sub-Committee Chair of ISSCC'13, the TPC Vice Chair of ISSCC'14, and the TPC Chair of ISSCC'15. He was a recipient of the Electronic Industrial Association of Korea Award for his contribution to DRAM technology in 1994, the Hynix Development Award in 1995, the Korea Semiconductor Industry Association Award in 2002, the Best Research of KAIST Award in 2007, the Scientist/Engineer of this month Award from the Ministry of Education, Science, and Technology of Korea in 2010, the Best Scholarship Awards of KAIST in 2011, and the Order of Service Merit from the Ministry of Public Administration and Security of Korea in 2011. He was a corecipient of the ASP-DAC Design Award 2001, the Outstanding Design Awards of 2005, 2006, 2007, 2010, 2011, 2014 A-SSCC, and the Student Design Contest Award of 2007, 2008, 2010, 2011 DAC/ISSCC.