

A 1920×1080 30-frames/s 2.3 TOPS/W Stereo-Depth Processor for Energy-Efficient Autonomous Navigation of Micro Aerial Vehicles

Ziyun Li, *Student Member, IEEE*, Qing Dong, *Student Member, IEEE*, Mehdi Saligane, Benjamin Kempke, Luyao Gong, Zhengya Zhang, Ronald Dreslinski, Dennis Sylvester, *Fellow, IEEE*, David Blaauw, *Fellow, IEEE*, and Hun-Seok Kim, *Member, IEEE*

Abstract—This paper presents a single-chip, high-performance, and energy-efficient stereo vision depth-estimation processor for micro aerial vehicles (MAVs). The proposed processor implements the state-of-the-art semi-global matching (SGM) algorithm to deliver full high-definition (HD, 1920×1080) stereo-depth outputs with a maximum of 38 frames/s throughput. Algorithm-architecture co-optimization is conducted, introducing overlapping block-based processing that eliminates very large on-chip memory and off-chip DRAM. We exploit inherent data parallelism in the algorithm by processing 128 local disparity costs and aggregating the SGM costs along four paths for all 128 disparities in parallel. A dependence-resolving scan associated with 16-stage deep pipeline is introduced to hide the data dependence between neighboring pixels in the SGM algorithm. Moreover, we propose a customized ultra-high bandwidth dual-port SRAM that utilizes the unique memory access characteristic of SGM to achieve highly energy-efficient memory access at a very high on-chip memory bandwidth of 1.64 Tb/s. The fabricated processor produces 512 levels of depth information for each pixel at full HD resolution with 30-frames/s performance, consuming 836 mW from a 0.75-V supply in TSMC 40-nm GP CMOS. We ported the design on a quadcopter MAV to demonstrate its performance in realistic real-time flight.

Index Terms—8T-SRAM, autonomous navigation, semi-global matching (SGM), stereo vision.

I. INTRODUCTION

PRECISE depth estimation is essential to realize autonomous navigation on micro aerial vehicles (MAVs), robots, and self-driving cars. Depth estimation serves as a key kernel function in simultaneous localization and mapping, 3-D scene understanding and reconstruction, object recognition, and obstacle avoidance, as depicted in Fig. 1. Real-time reliable autonomous navigation requires the depth estimation to be dense, accurate, wide range, and high performance. Emerging mobile platforms such as MAVs introduce additional size, weight, and power (“SWaP”) constraints on depth estimation systems. For mobile applications, the system must be small (e.g., 50 cm³), lightweight (<100 g),

Manuscript received May 1, 2017; revised July 17, 2017; accepted August 30, 2017. Date of publication September 27, 2017; date of current version December 26, 2017. This paper was approved by Guest Editor Muhammad M. Khellah. (*Corresponding author: Ziyun Li.*)

The authors are with the University of Michigan, Ann Arbor, MI 48105 USA (e-mail: liziyun@umich.edu).

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.2017.2751501

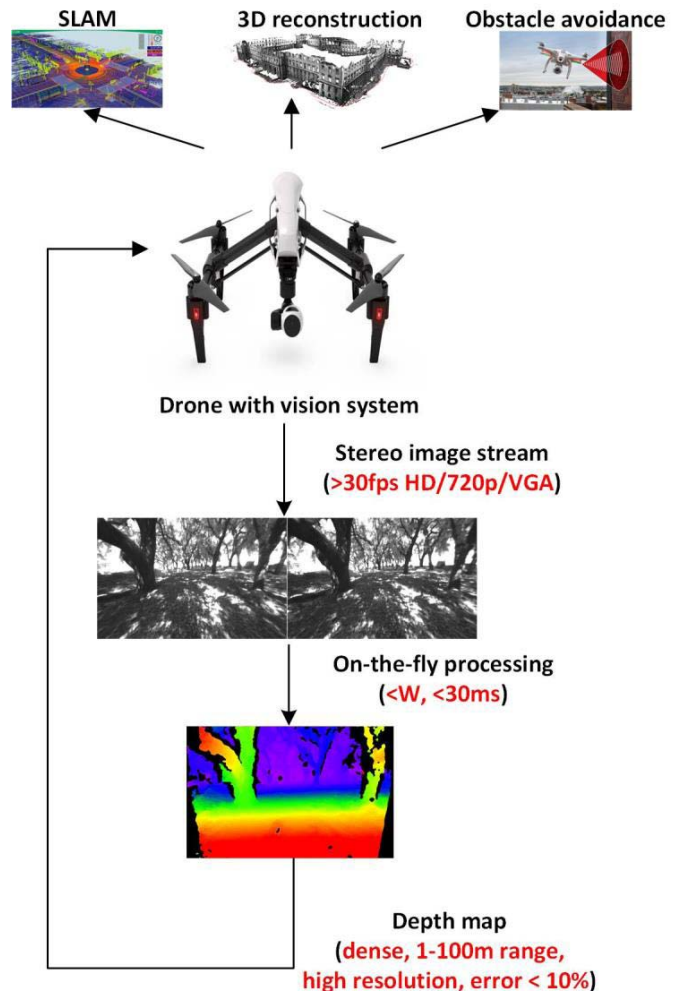


Fig. 1. Drone with stereo vision system and its applications.

fast (~ 30 ms response time), and low power (< 1 W) [1]. Fig. 1 highlights the requirements of real-time stereo-depth estimation for MAV applications. Light detection and ranging (LIDAR) [2], RADAR [3], ultrasonic sensor [4], or IR sensor [5] are conventional approaches for depth sensing. IR sensors typically have low resolution and accuracy and ultrasonic sensors have limited ranging distance [6]. Therefore, they are not widely adopted on autonomous systems. While 24G RADAR is accurate and robust, it has a limited field

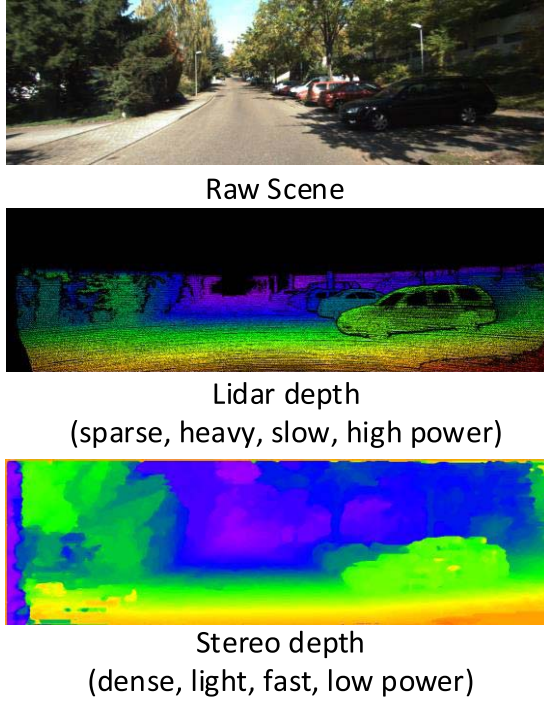


Fig. 2. Comparison between stereo-depth estimation and LIDAR-based depth estimation.

of view ($\sim 30^\circ$ horizontal angle) and will still need other sensors for wide-range, large-scale 3-D scene construction. LIDAR is the most frequently used sensor for 360° 3-D scene construction in autonomous systems. Fig. 2 visualizes the difference between the depth map acquired by LIDAR and that obtained with stereo vision correspondence. Nowadays, even the most advanced LIDAR-based ranging systems suffer from a limited field of view (e.g., $\sim 30^\circ$ vertical view angle, which accounts for the top blackout region of the LIDAR image in Fig. 2), weigh > 600 g, and consume 10 W [2]. In contrast, depth estimation from stereo vision is fast, energy efficient, and lightweight when mounted on an MAV platform.

There are prior application-specified integrated circuit (ASIC) implementations of stereo vision depth estimation based on various algorithms [7]–[11]. These designs [7]–[11] employ hardware-oriented algorithmic optimizations to enable depth estimation in real-time systems, but there are deficiencies associated with these ASIC designs. Some use local matching [7] or aggressively truncated global algorithms [8], which result in inferior quality. Other works limit their disparity range to 32 or 64 pixels and therefore fail to support industry standard automotive scene benchmarks [7]–[11]. Semi-global matching (SGM)-based field-programmable gate array (FPGA) implementation is demonstrated in [12], but it is not applicable for robust navigation of power constrained MAV platforms because of its limited performance (~ 30 frames/s for 320×240 QVGA) and high-power consumption (~ 3 W for QVGA). Prior advanced driver assistance system system-on-chips [5], [13], [14] are not favorable for SGM because of the memory bandwidth bottleneck. Due to the high memory requirement of SGM, prior methods [8]–[10] use external DRAM to store intermediate results, limiting frame rate and power efficiency.

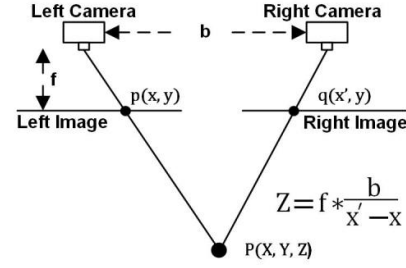


Fig. 3. Illustration of principles in stereo vision.

Several alternative approaches have been explored to reduce the high complexity of global methods [15] using dynamic programming [16], belief propagation [17], [18], flow vector search space pruning [19], and pseudo-random flow candidate selection [20]. However, SGM (and its variations) is clearly one of the most widely used algorithms in industry standard benchmarks, such as KITTI [21].

In this paper (extended from [22]), we first study the computation, memory, and bandwidth bottlenecks of the SGM algorithm and the proposed algorithm–architecture co-optimization techniques that significantly reduce the hardware cost with negligible accuracy degradation. We propose a deeply pipelined hardware architecture with a dependence-resolving scan to handle the critical-path data dependence in the algorithm and to significantly improve throughput. We also introduce a custom-designed dual-port 8T-SRAM that leverages the unique memory access characteristics of the SGM algorithm to enable ultra-high bandwidth (1.64 Tb/s) and energy-efficient on-chip memory access. The fabricated chip employs a standard USB3.0-compliant interface, allowing effortless integration with a wide range of commercial off-the-shelf stereo cameras and general purpose mobile application processors (APs). Integrated with a ZED camera [23] and ODRIOD-5422 mobile AP on the ODRIOD-XU4 [24], the fabricated chip was successfully mounted on a quadcopter MAV for a system demonstration in realistic flight scenarios. The chip delivers 512 levels of stereo depth for each pixel at full high-definition (HD) (1920×1080) resolution with real-time 30-frames/s throughput, while consuming 836 mW.

II. OVERVIEW OF STEREO VISION ALGORITHMS

A. Local Approach

Vision-based depth estimation is computed by stereo correspondence. As shown in Fig. 3, a point P in the real world will be horizontally displaced at the pixel positions \mathbf{p} and \mathbf{q} in stereo images because the left and right cameras are placed apart by distance b . This horizontal displacement between a pixel in the left image $\mathbf{p} = (x, y)$ and its matching pixel in the right image $\mathbf{q} = (x', y)$ is defined as disparity ($x' - x$). The depth Z is inversely proportional to the disparity as follows:

$$Z = f \frac{b}{x' - x} \quad (1)$$

where f represents the focal length of the camera.

The most straightforward approach to compute disparity is local matching. As shown in Fig. 4, local matching compares each pixel (e.g., the white pixel in the far left image) with

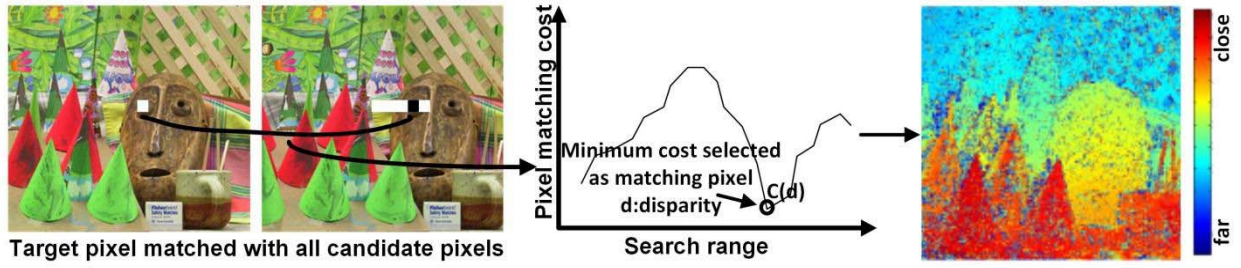


Fig. 4. Local matching method of stereo-depth estimation.

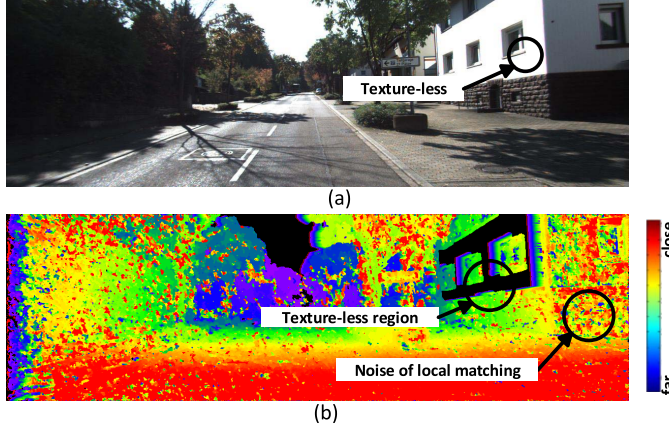


Fig. 5. Problems of local matching. (a) Input image. (b) Local matching.

all its matching candidates and then finds the best matching pixel (e.g., the black pixel in the next image to the right) associated with the minimum matching cost within the search range (depicted by the white bar in Fig. 4). Typically, to enhance robustness, local matching is performed based on a window that consists of a group of pixels surrounding the matching pixel. The same step is applied to determine the disparity of all of the pixels. An example disparity map resulting from using local matching is shown in Fig. 4. In the local approach, every pixel in the image can be processed independently in parallel to improve throughput.

The accuracy of a local approach is unreliable since it typically fails to resolve ambiguities in many challenging but realistic scenarios such as occlusions, texture-less regions, transparency, and repetitive patterns. As shown in Fig. 5, almost all of the pixels of the wall on the right are saturated and texture-less due to strong illumination. The disparity results obtained from a local approach on this texture-less region will be completely incorrect, as shown in Fig. 5, because many matching pixels will appear identical with the same cost. As seen in other less challenging regions in Fig. 5, the disparity map derived from using a local approach has substantial noise that cannot be easily removed by advanced post processing.

B. Semi-Global Matching and Its Complexity

Recently, various global methods [25]–[27] have been proposed to improve accuracy. In these global algorithms, information from neighboring pixels is (semi-)globally propagated to the current processing pixel to enhance the correspondence

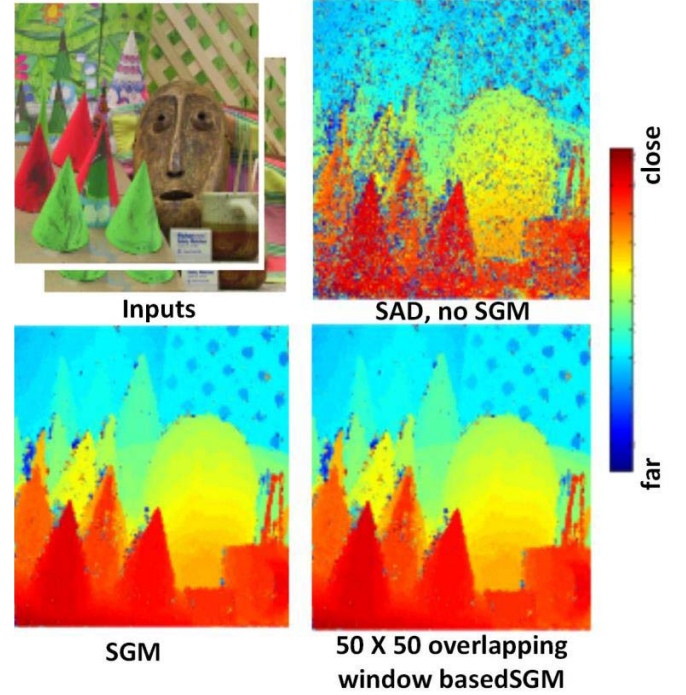


Fig. 6. Comparison between local matching, original SGM, and overlapping block-based SGM.

matching accuracy. The SGM algorithm introduced in [28] is one of the most popular global methods. SGM is favored for its robustness and high accuracy under various scenarios. SGM has been validated to achieve good accuracy in various industry standard benchmarks [21]. In particular, it effectively handles low texture regions with its dynamic programming-based global optimization of the disparity over the entire image. Fig. 6 visualizes the output difference between the local sum of the absolute difference [25] algorithm and SGM, clearly illustrating the higher quality obtained with SGM.

SGM consists of three steps: 1) pixel-wise matching cost computation; 2) semi-global aggregation; and 3) disparity selection. To compute the pixel-wise matching cost, $N \times N$ census transform [29] is performed on both the left and right images. Census transform $I_L(\mathbf{p})$ of a pixel \mathbf{p} is computed by comparing the grayscale intensity of center pixel \mathbf{p} with all of its neighboring pixels within the $N \times N$ window. As a result, each pixel in the image is converted to a bit string of the length $N^2 - 1$. We use a 7×7 census for our design, and each pixel is represented by a 48-bit string as a result of

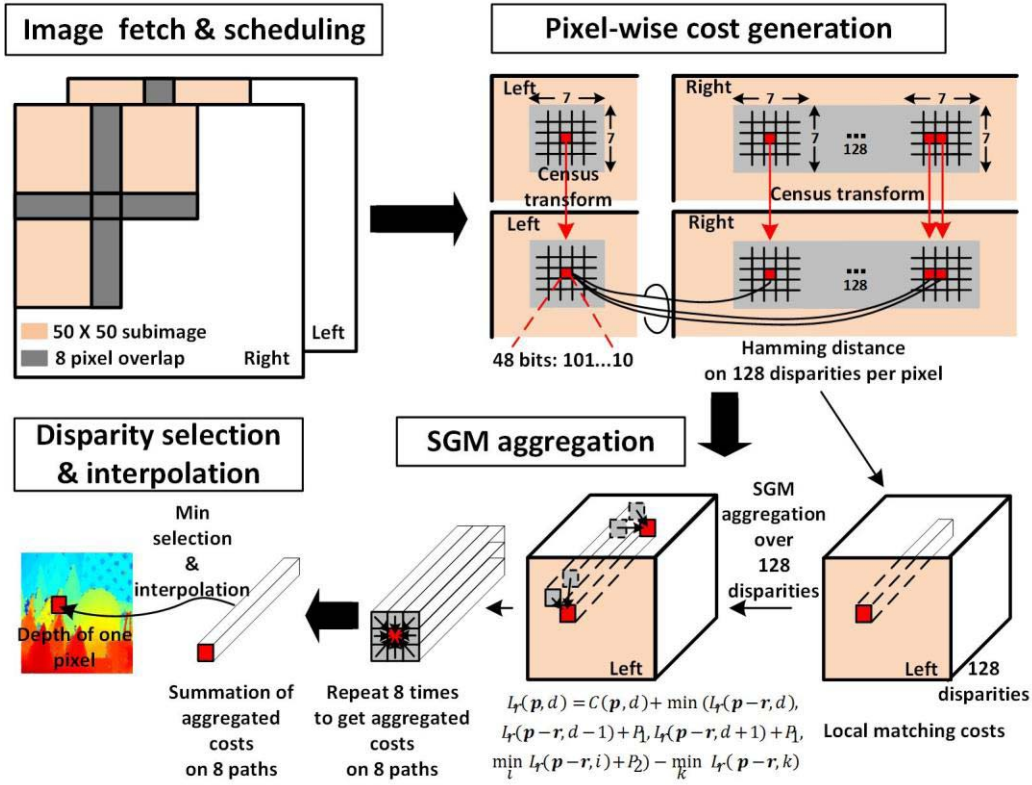


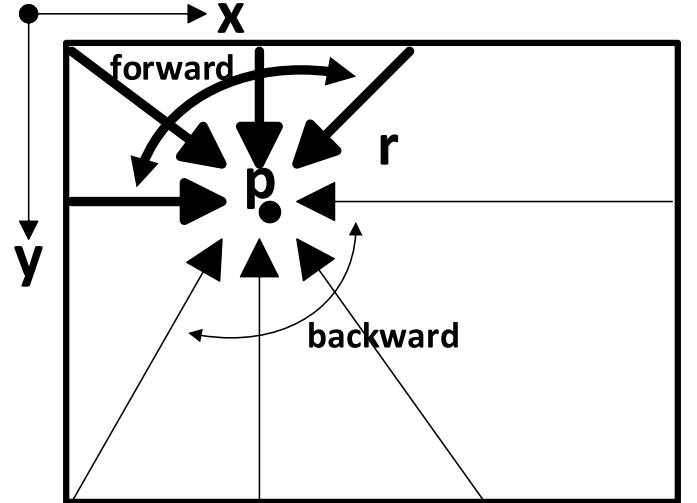
Fig. 7. SGM algorithm processing flow.

the census transform. The pixel-wise matching cost $C(\mathbf{p}, d)$ for a pixel \mathbf{p} with a disparity d is evaluated by the Hamming distance [31] between the census-transformed left image pixel $I_L(\mathbf{p})$ and the right image pixel $I_R(\mathbf{p} - d)$, as shown in (2), where $|\cdot|_H$ denotes the Hamming distance

$$C(\mathbf{p}, d) = |I_L(\mathbf{p}) - I_R(\mathbf{p} - d)|_H. \quad (2)$$

This operation is repeated for all disparity candidates per pixel. Since each pixel will have 128 matching candidates, the local matching costs evaluation results in a cube of dimension $H \times W \times 128$, as shown in Fig. 7, where $H \times W$ is the image height \times width in the number of pixels. For each pixel, a 128-entry depth vector is generated as the local matching cost associated with 128 disparities.

Global aggregation is then performed on local matching costs. SGM aggregation takes the current processing pixel \mathbf{p} and propagates information from neighboring pixels along eight paths r over the entire image using (3) as depicted in Fig. 8. The term $L_r(\mathbf{p}, d)$ denotes the aggregated cost for a given pixel \mathbf{p} with disparity d along path r . The aggregated costs of neighboring pixels associated with the same (d) or similar ($d \pm 1$ and $d \pm 2$) disparities are merged into the aggregated cost for the current pixel with zero or small (P_1, P_2) penalty. Eventually, ‘good’ disparities with low matching costs (when neighboring pixels all see smaller matching costs in general) will propagate from far positions to the current pixel through recursive aggregation (3). This is particularly useful for propagating “good” matching candidates for the center of a low texture region from

Fig. 8. Eight-path aggregation diagram.
texture-rich boundary pixels

$$L_r(\mathbf{p}, d) = C(\mathbf{p}, d) + \min \{ L_r(\mathbf{p} - \mathbf{r}, d), L_r(\mathbf{p} - \mathbf{r}, d - 1) + P_1, L_r(\mathbf{p} - \mathbf{r}, d + 1) + P_1, \min_i L_r(\mathbf{p} - \mathbf{r}, i) + P_2 \} - \min_k L_r(\mathbf{p} - \mathbf{r}, k). \quad (3)$$

The SGM aggregation is performed along eight paths (the size of the r set is eight) separately, as shown in Fig. 8, and the aggregated costs on eight paths are summated together as follows:

$$S(\mathbf{p}, d) = \sum_r L_r(\mathbf{p}, d). \quad (4)$$

TABLE I
LOW EFFICIENCY ON CPU/GPU/FPGA PLATFORMS

	CPU (intel i7)	GPU (TITAN X)	GPU (Tegra)	FPGA
Performance @ FHD	0.05fps	29fps	2.3fps	11fps
Power	~50W	~250W	~11W	~8W
Energy per frame	>35J	8.4J	4.24J	0.8J
External DRAM	yes	yes	yes	yes

TABLE II
SUMMARY OF SGM CHALLENGES AND ALGORITHM-ARCHITECTURE-CIRCUIT OPTIMIZATIONS

Optimization \ Challenges	Massive computation	Inter-pixel Dependency	Large Memory	High memory bandwidth
Parallel processing: Compute 128 local costs in parallel	√			√
Aggregate 128 disparities in parallel Aggregate 4 path in parallel				
16 stage deep pipeline		√		√
Dependency resolving scan		√		
Overlapping block based processing			√	
Overlapping block based processing				√

The disparity d with the minimum summated costs $S(\mathbf{p}, d)$ is eventually selected as the integer level of disparity for the processing pixel \mathbf{p} . To obtain a sub-integer pixel disparity precision, we select three minimums from all of the summated costs $S(\mathbf{p}, d)$ for a given \mathbf{p} and perform a bilinear interpolation [32] on these three minimums. This generates an additional 2-bit sub-integer pixel disparity resolution and eventually generates 512 levels of depth (disparity) for each pixel.

Although SGM provides superior accuracy compared with local approaches, it poses significant hardware challenges. The original SGM requires massive computation ($\sim 2\text{TOP/s}$), extremely high-bandwidth (38.6 Tb/s), and very large memory ($\sim 386\text{ MB}$) for 30 frames/s full HD resolution. Therefore, when realized in general-purpose computing platforms, it leads to very low frame rates and energy efficiency. Specifically, this SGM complexity translates to $\sim 20\text{ s}$ runtimes for a full HD image pair on a 3-GHz CPU with $>35\text{ W}$ power consumption [33]. Although server/mobile GPUs achieve higher energy efficiency, they still consume a few Joules to process a single full HD frame with $\sim 5\text{ frames/s}$ throughput [34]. Table I provides a comparison of the estimated performance, power and memory for different platforms. To resolve these challenges and address “SWaP” requirements of MAVs, we propose a highly optimized ASIC solution attained via a cross-layer optimization conducted across algorithm, micro-architecture, and circuit levels.

III. ALGORITHM, ARCHITECTURE, AND CIRCUIT OPTIMIZATIONS

A high-level summary of our cross-layer optimizations designed to tackle the challenges of SGM is illustrated

in Table II. First, strong data parallelism in the algorithm is exploited so that the processor computes 128 local costs, aggregates 128 disparities, and accumulates four paths all in parallel. Second, instead of processing the whole image frame, we propose an overlapping block-based processing to eliminate the very large on-chip memory requirement and to achieve a single-chip SGM implementation without off-chip DRAM accesses. Moreover, a dependence-resolving scan with a 16-stage deep pipeline is proposed to hide the data dependence and improve throughput by $3\times$. Finally, we also custom designed an ultra-high bandwidth dual-port SRAM that leverages unique memory access patterns of SGM for high-performance and energy-efficient memory access.

A. Algorithm: Overlapping Block-Based SGM Processing

The original SGM algorithm consists of forward and backward scans as shown in Fig. 8, where each scan aggregates costs for each pixel along four paths. This two-scan approach is unavoidable to allow eight-path aggregation over the entire image frame. The partial (four paths) aggregated costs (12 bit each) for every pixel are stored in the memory during the forward scan and then later combined with the backward scan results for the remaining four paths. This two-scan imposes significant on-chip memory and bandwidth requirement for storing 128 (number of disparities) aggregated costs ($\sim 16\text{ bits}$ each) for every (2 M for full HD) pixel in the image. Therefore, the memory requirement of SGM is not scalable to various image resolutions, and a single full HD image pair will require $\sim 386\text{ MB}$ storage for temporary aggregated costs. A prior work [35] reduced the amount of temporary memory usage without significant accuracy degradation by selectively storing sparse aggregated costs. Similarly, we only store three

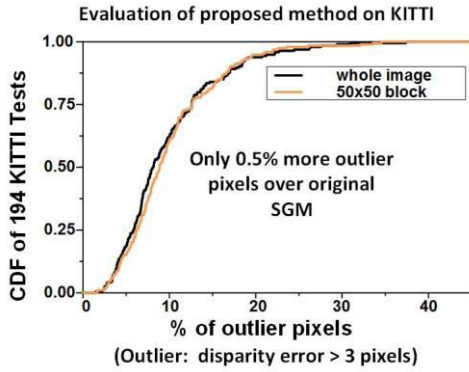


Fig. 9. Quantitative evaluation with overlapping block-based SGM over 194 KITTI test images.

disparities associated with the three minimum summated costs for each pixel. This allows the temporary memory in our SGM implementation to be independent of the disparity search range (the number of disparities evaluated per pixel). However, the temporary memory size still depends on the image size, and storing sparse aggregated costs with their associated disparities for a full HD image would still require ~ 20 MB of memory. This memory requirement will significantly degrade energy efficiency if it is mapped to external DRAM.

To further reduce the on-chip memory requirement and to eliminate the need for external DRAM, we first evaluate the sensitivity of accuracy with different overlapping window size on 194 KITTI cases. While the original SGM achieves 6.5% outlier, overlapping blocks of 200×200 , 150×150 , 100×100 , and 50×50 achieve 6.61%, 6.62%, 6.75%, and 7% outliers, respectively. From the evaluation, we observe that inter-pixel correlation diminishes when pixel pairs are more than 50 pixels apart. Therefore, instead of processing the whole image, the proposed design uses an overlapping block-based processing to partition the input image into units of 50×50 pixel overlapping blocks to minimize on-chip memory size, as shown in Fig. 7 top left. Adjacent blocks are overlapped by 8 pixels to allow cost aggregation across block boundaries. This technique achieves 95.4% memory reduction for storing intermediate aggregation costs for a full HD image. We evaluate this technique with standard Middlebury [36] and KITTI [21] benchmarks. Fig. 6 shows a side-by-side qualitative comparison of this block-based SGM and the original SGM on Middlebury test case, which yields almost identical results. Fig. 9 presents the accumulated density function evaluated on 194 realistic KITTI automotive test cases. The proposed overlapping blocked-based SGM suffers only 0.5% outlier percentage degradation compared with the original SGM throughout 194 KITTI evaluation cases. An outlier is a pixel that has a disparity error of more than three integer levels.

B. Energy-Efficient Hardware Architecture

The proposed block-based SGM processing procedure is shown in Fig. 7, and the chip architecture is shown in Fig. 10. One 32-bit parallel interface streams input image

data and processing instructions into the chip, and the other 32-bit parallel interface is used to stream the final disparity results of the chip. The control registers and on-chip input images are memory mapped and can be accessed with a USB interface through an external USB-to-parallel converter [37]. To maximize the input bandwidth, a streaming mode is supported so that input/output image data are streamed to/from the chip continuously. The block-partitioned left and right images are stored in two on-chip interleaved image (ping-pong) buffers (30 Kb each). Processing is concurrently performed with input/output image blocks streaming to achieve real-time performance.

As the first step, 7×7 census transformations are performed on the processing pixel as well as its matching candidates at 128 different disparity locations using their surrounding (7×7 window) pixels. This census transform computing on-the-fly scheme would result in ~ 6000 compare operations for every processing pixel and thus have poor energy efficiency. We observe that 127 out of 128 census-transformed matching candidates of previous pixels overlap with the census-transformed pixels of the current processing pixel when processing proceeds. Therefore, an on-chip 128-entry circular first-in first-out (FIFO) is employed to eliminate redundant census transforms and to store the pre-computed census results. A total of 127 census-transformed matching candidates (48 bit each) are read directly from the on-chip FIFO, as shown in Fig. 11. At each cycle when a new pixel is pushed into the pipeline, only one new census transform is performed and pushed into the FIFO. This census FIFO eliminates 98% of redundant census transforms. In simulation, storing census transforms in an FIFO and preloading them (5.1 pJ/pixel including memory accesses) achieves $2.8\times$ better energy efficiency compared with on-the-fly re-computing census (14.2 pJ/pixel) for every pixel.

The census-transformed pixel in the left image and the census-transformed pixels in the right image at 128 different disparity locations are then compared in parallel. This produces 128 Hamming distances (6 bits each) for each pixel that represent the ‘local’ pixel-wise matching cost vector for the 128 disparities; $C(p, d)$. These 128 local costs are all sent to four parallel aggregation units for SGM aggregation. Each aggregation unit is equipped with a high-bandwidth buffer and aggregates 128 disparity locations in parallel, accumulating costs over four different paths. Massive parallelism in aggregation shown in Fig. 12 helps us achieve high throughput and energy efficiency. The tree-structured selection unit identifies the best three aggregated costs and disparities. These three-best aggregated costs for each pixel are stored in the on-chip memory. Once the forward scan completes, the backward scan is performed in the similar fashion. Aggregation results that are discarded (except for three-best results) during the forward scan are combined with a constant penalty with the backward scan results. The final best disparity candidates are selected based on the eight-path aggregated costs. Finally, bilinear interpolation is performed, and the 512-level (7-bit integer and 2-bit fractional) disparity results are stored in two interleaved result buffers.

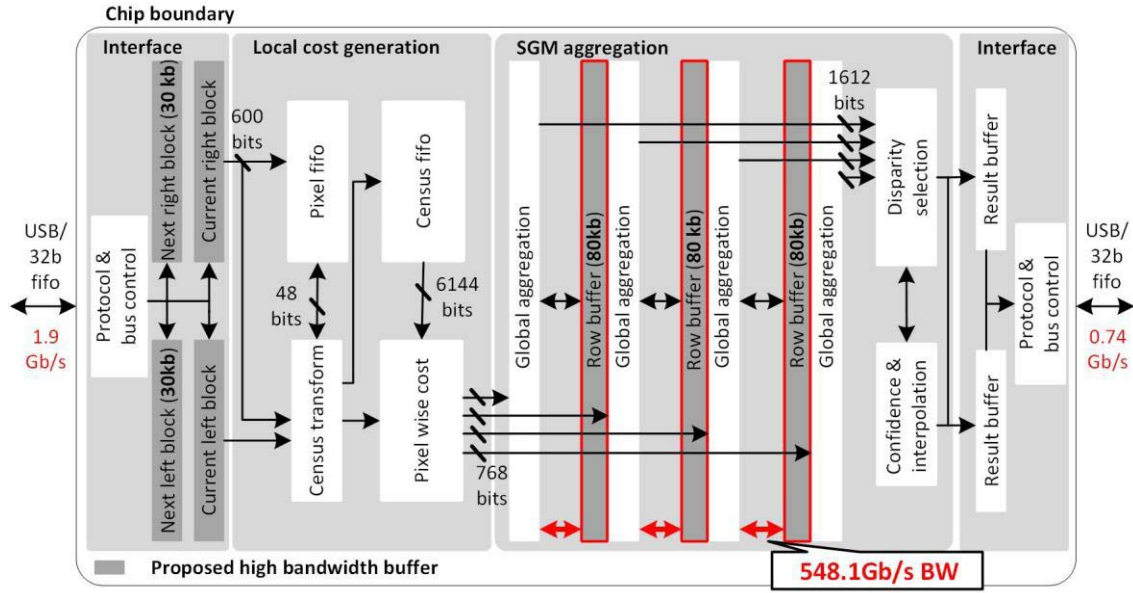


Fig. 10. Hardware architecture of energy-efficient SGM.

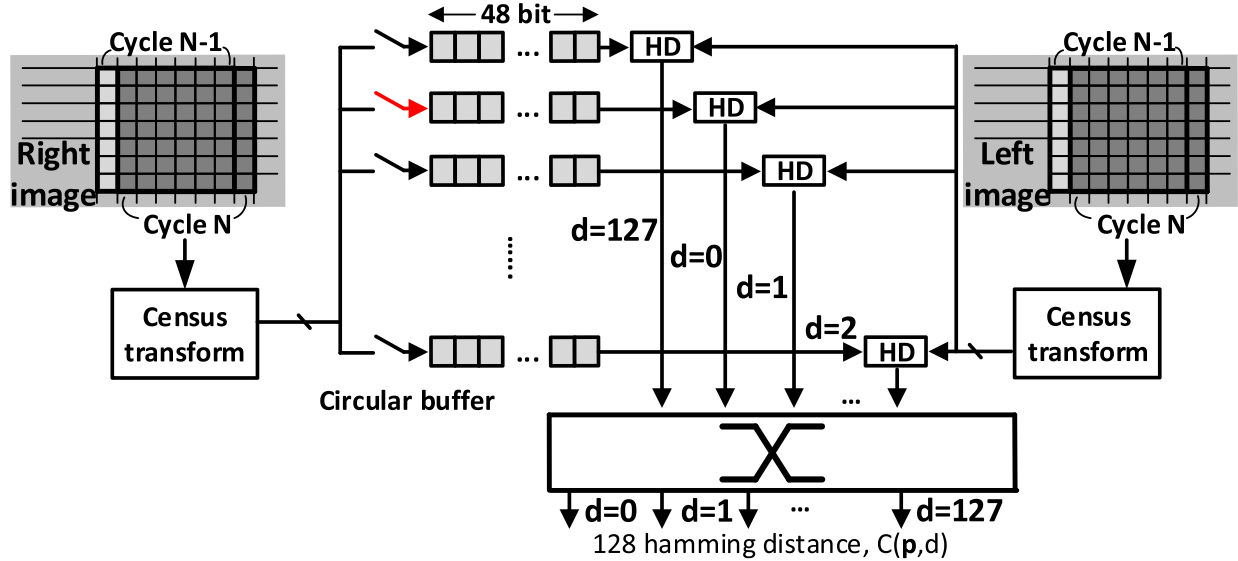


Fig. 11. Rotating FIFO-based local cost generation.

C. Dependence-Resolving Scan, Pipelining, and Forwarding

In the proposed highly parallelized cost aggregation, each aggregation unit has its own ultra-high bandwidth row buffer (marked in dark gray in Fig. 10). During each clock cycle, each aggregation unit reads the 128 aggregated costs from each neighboring pixel (four neighbors as shown in Fig. 13) from the buffer and writes the 128 aggregated costs of the current pixel to the buffer. However, this straightforward implementation would result in data dependence because the aggregation of the current pixel depends on the results of the neighbor that is processed in the previous cycle.

As discussed earlier, SGM is implemented with a forward and a backward raster scan, with each scan performing aggregation along four paths (Fig. 8). However, following this conventional raster scan order results in data dependence because the previous pixel must complete its computation before the

current pixel can be aggregated. As shown in Fig. 13(a), the forward scan aggregates the results from its four neighbors marked with arrows. Aggregation in the top left, top, top right does not lead to dependence because those pixels belong to the last row and are ready before processing the current pixel. Data dependence is from the left neighboring pixel along the raster scan path, and processing of the current pixel must wait until its left neighbor finishes aggregation. This data dependence dominates the critical path, limiting the clock frequency and voltage scalability for low-power operation.

We therefore propose a dependence-resolving scan in which pixel processing proceeds diagonally [Fig. 13(b)]. Now the original single cycle data dependence extends to five cycles because of the diagonal scan. This allows pipelining the aggregation unit with five cycles and resolving inter-pixel dependence in a deep pipeline. Fig. 14 shows the proposed

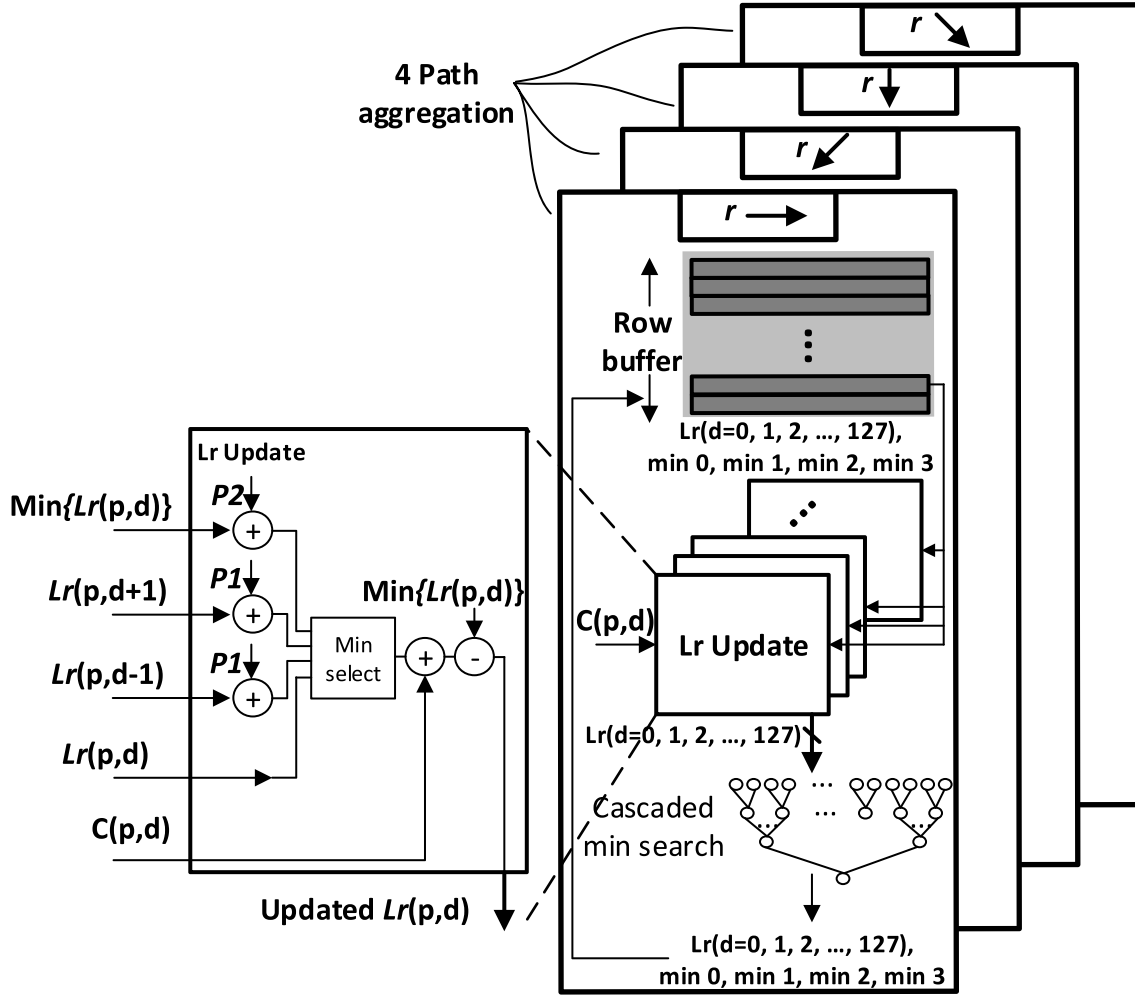


Fig. 12. Implementation of four-path aggregation.

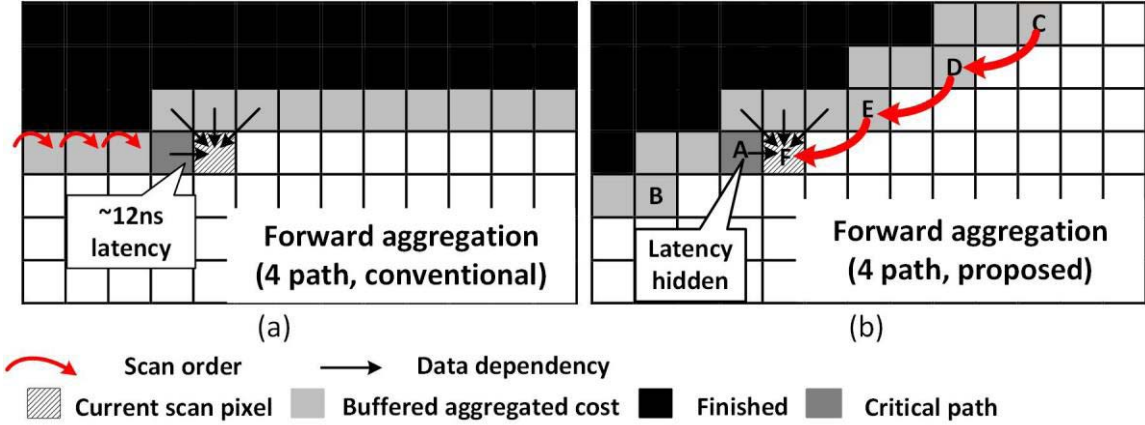


Fig. 13. Illustrations of conventional raster scan and proposed diagonal scan.

16-stage deep pipeline for SGM processing. With the diagonal scan, there are five cycles between pixel A and F during which we can process the other 4 pixels (B, C, D, and E). When F is fetched into the pipeline, the aggregated costs of previous pixels (light gray and dark gray) are already computed and stored in high-bandwidth custom SRAMs. The critical-path data from pixel A is forwarded to pixel F in

the pipeline. This mechanism enables aggressive pipelining with a 4-ns clock frequency, yielding a 3× performance gain compared with that of the conventional raster scan. Moreover, because the data processed for pixel A are forwarded in the pipeline, this successfully eliminates unnecessary row buffer, leading to an extra 25% memory reduction. The 16-stage deeply pipelined design operates at a relatively low

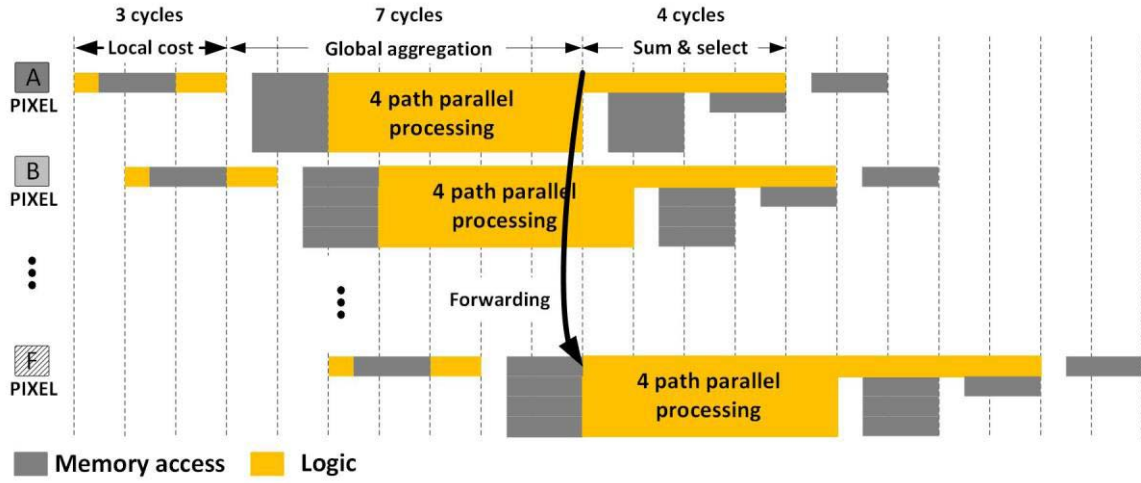


Fig. 14. Pipelining and forwarding in SGM processing.

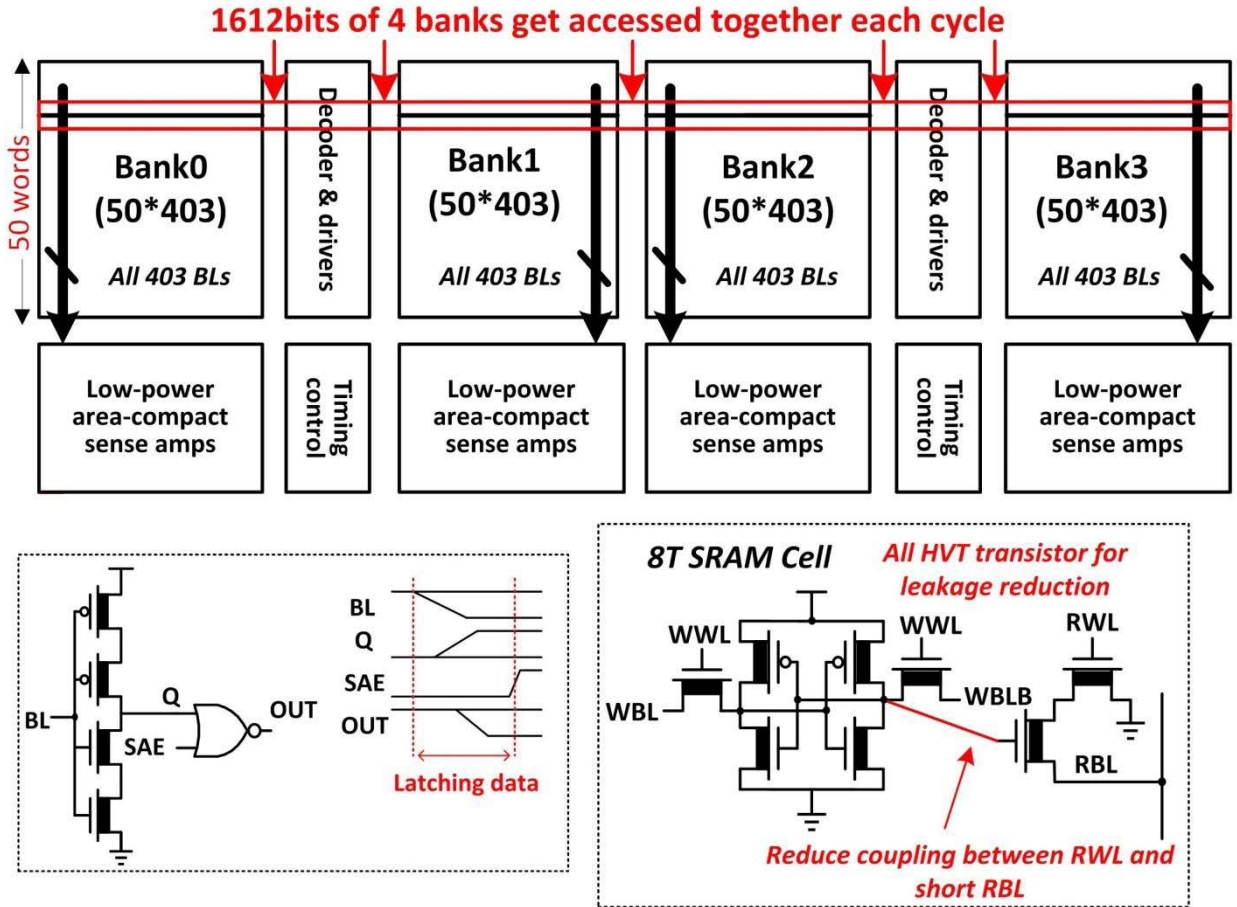


Fig. 15. Block diagram and circuit of the proposed high-bandwidth 8T-SRAM.

frequency (200 MHz) to minimize the energy overhead of tremendous parallel pipeline registers if the design has more pipeline stages with higher frequency. As shown in Fig. 13, our design leverages parallelism in cost aggregation by running four paths in parallel on four aggregation units. Each aggregation unit contains 128 processing elements and 512 selection units, resulting in a total throughput of 1.882 TOP/s.

Each OP is defined as 8-bit integer operation including add, subtract, compare, and memory access.

D. Custom-Designed High-Bandwidth 8T-SRAM

In the proposed design, each aggregation unit has its own ultra-high bandwidth row buffer. For each row buffer, 128 aggregated costs (12 bit each) are read and written

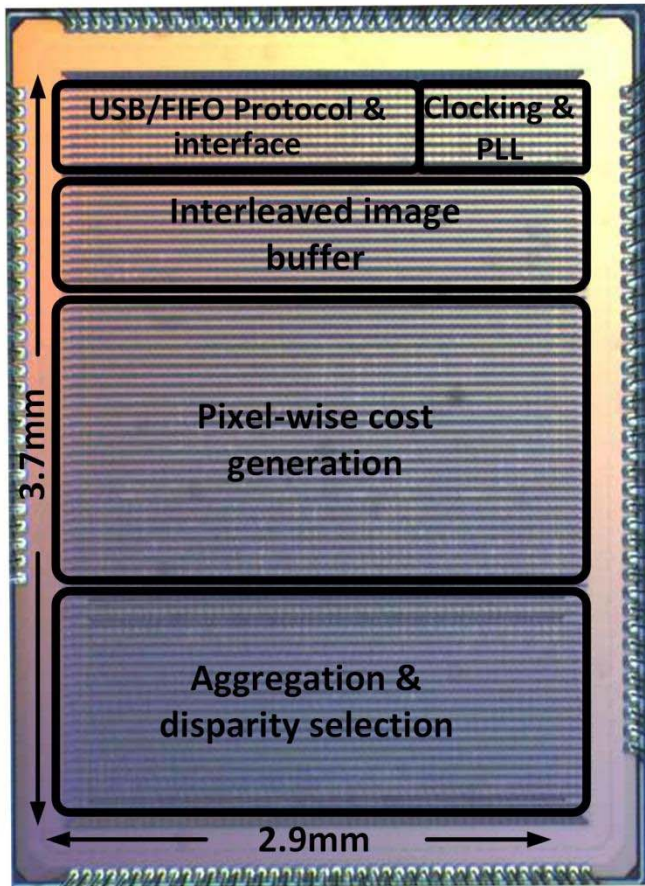


Fig. 16. Die photograph and summary of performance.

simultaneously in a single cycle at 170 MHz. This translates to a total memory bandwidth of 1.64 Tb/s for the three row buffers accessed in parallel. This bandwidth would incur significant chip area and power overhead if realized with compiled SRAMs as a large number of banks and redundant peripherals are unavoidable due to the limited word length of compiled SRAMs. In simulation, instead, we propose a custom-designed dual-port SRAM to cope with this challenging memory access characteristic of SGM.

Because of the design's highly parallelized structure, the row buffer has a very unconventional aspect ratio: there are only 50 words in the buffer, but each word is 1612 bits wide. This motivates the proposed high-bandwidth custom SRAM that provides enhanced area/power efficiency of SGM that was previously unattainable by general-purpose computing platforms. Fig. 15 shows the block diagram of the customized high-bandwidth SRAM. We partition the row buffer into four banks, and each bank has 50 words with a word size of 403 bits. All four banks are accessed in parallel with concurrent read and write functions, realizing 1612-bit dual-port access.

The very unbalance aspect ratio of this custom SRAM results in a massive number of very short bit lines ($\sim 50 \mu\text{m}$ each) and very long word lines ($\sim 380 \mu\text{m}$ each). Therefore, unlike conventional 8T cells, we propose swapping the position of the conventional 8T-SRAM read transistor stack to avoid directly connecting the read access transistor to the read bit line (RBL). This approach effectively reduces

	This work
Algorithm	8 Path SGM
Technology	40nm
Core Area	10.8mm ²
On-chip memory	1064Kb
Frequency	170MHz
Image size & throughput	1920 X 1080 30 fps
Depth	512 (128 + 2 bit fractional)
Benchmark evaluation	7% outlier @ KITTI
Operating voltage	0.75V @ 30 fps HD 0.52V @ 30 fps VGA
Power	836mW @ 30 fps HD 55 mW @ 30 fps VGA
Normalized energy	0.0262nJ @ 30 fps HD 0.0117nJ @ 30 fps VGA

coupling between the read word line (RWL) and the short, low-capacitance RBL. In spite of the simulation with 0.9-V nominal voltage, the coupling from RWL to RBL is reduced from ~ 18 to ~ 2 mV when read access transistor stack is flipped. Fig. 15 also shows the bit cell circuit in the bottom right. To reduce leakage power in the 40-nm technology, the custom 8T bit cell uses HVT transistors. The stacked skewed inverter-based sense amplifier and the timing of the SRAM read operation are shown in the bottom left of Fig. 15. Output latches are transparent during the RBL evaluation phase to ensure the correct memory read operation. Employing conventional sense amplifiers for 1612-bit lines would lead to significant area overhead. Therefore, skewed inverters perform RBL voltage sensing to achieve better area efficiency. Compared with conventional sense amplifiers, skewed inverters reduce the area overhead by $2.8\times$. The low capacitance on the short BL allows the proposed SRAM to reliably operate at 200 MHz with a supply voltage as low as 0.6 V, further improving the energy efficiency. Overall, each 80-Kb SRAM is measured to consume only 6 mW with 548.1 Gb/s bandwidth at 200 MHz. Three banks operate at 200 MHz with concurrent 1612-bit read and write operations, achieving 1.64-Tb/s access bandwidth with 18-mW power consumption.

IV. CHIP MEASUREMENT RESULTS

Fig. 16 shows a die photograph with a summary of the test chip performance. This work is fabricated in TSMC 40-nm GP process with 10.8-mm² chip area. TSMC 40-nm GP

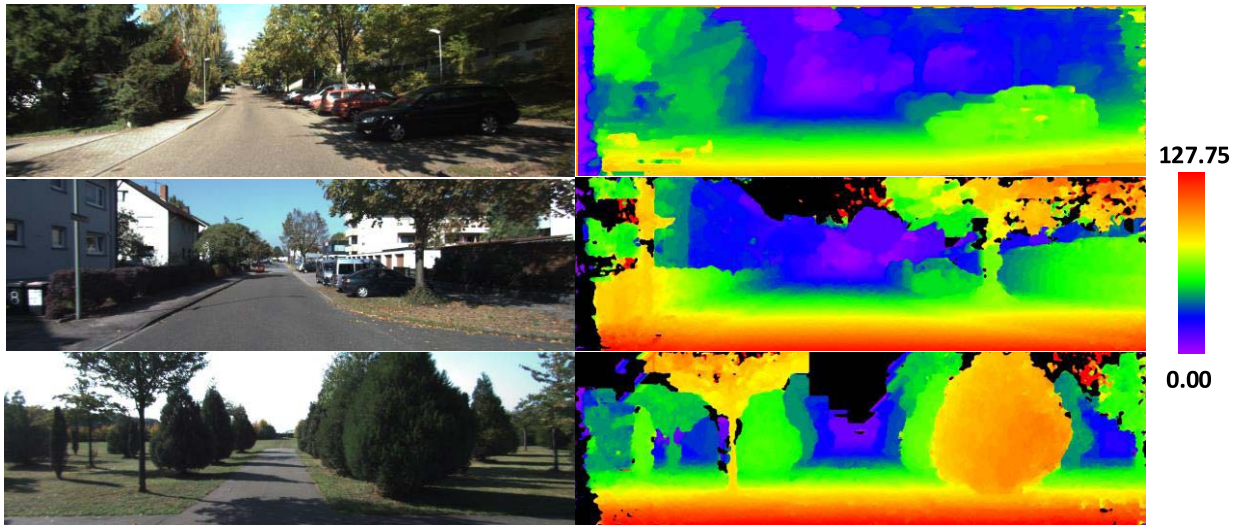


Fig. 17. Measured result with KITTI tests.

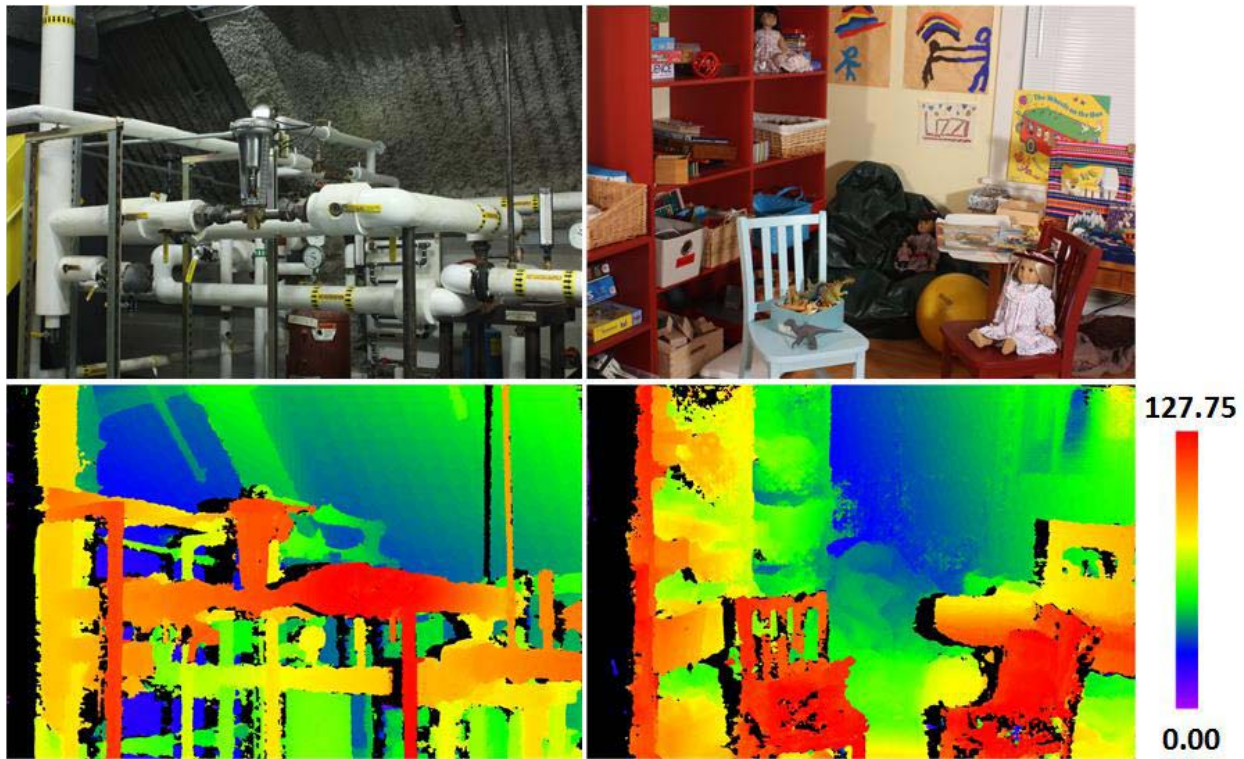
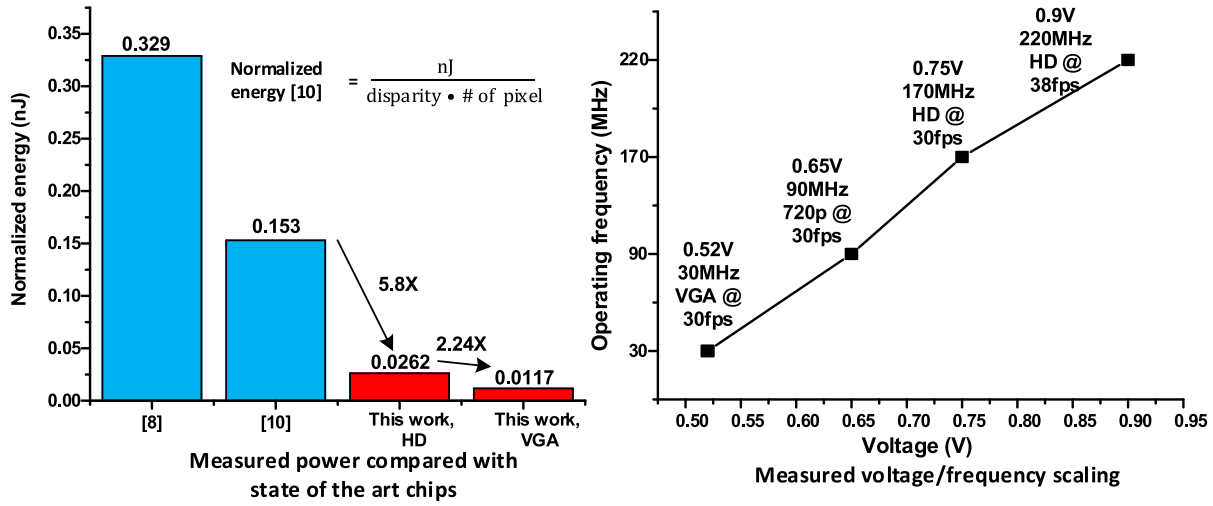


Fig. 18. Measured result with Middlebury tests.

process has low nominal voltage (0.9 V), and high performance (low V_{th}), which meets our design target. The fabricated chip successfully produces 512 levels of depth in full HD (1920×1080) resolution with real-time 30 frames/s performance with 170 MHz core frequency and consumes 836 mW from a 0.75-V supply. The depth image outputs produced by the chip using KITTI [21] automotive scenes are shown in Fig. 17. Notice that the depth information of the cars in the shadow is successfully obtained. Large (>100 pixels) disparity frequently occurs at close distances, and the proposed processor is able to generate an

accurate depth map over the entire image due to its 512 levels of resolution. The proposed chip achieves 7% outlier pixels running 194 KITTI evaluation images. Fig. 18 shows the typical chip measurement results from Middlebury indoor scenes.

Fig. 19 shows the measured voltage and frequency scaling of the chip and provides a comparison with prior works. Compared with other state-of-the-art chips, this paper implements SGM depth with 512 disparity levels, resulting in $8\times$ improvement. It exhibits only 7% outliers in the KITTI benchmark, whereas other chips have limited



	VLSI 2004[7]	ISSCC 2015[10]	ISSCC 2016[8]	This work
Method	SAD	5-View BP	Truncated SGM	SGM
Technology	180nm	40nm	65nm	40nm
Chip area	25mm ²	22mm ²	16mm ²	10.8mm ²
On-chip memory	N.A	352Kb	3946.9Kb	1064Kb
Frequency	125MHz	215MHz	250MHz	170MHz
Throughput & image size	320 X 240 30fps	1920 X 1080 30fps	1280 X 720 30fps	1920 X 1080 30fps
Depth level	N.A	64 (32 + 1 bit fractional)	64	512 (128 + 2 bit fractional)
Accuracy (Outlier % on KITTI)	Not reported due to limited depth range			7%
Operating voltage	1.8 V	0.9 V	1.2V	0.75V @ 30 fps HD 0.52V @ 30 fps VGA
Power	N.A	1019mW (includes DRAM power)	582mW (excludes DRAM power)	836mW @ 30 fps HD 55 mW @ 30 fps VGA
Normalized energy*	N.A	0.153nJ	0.329nJ	0.0262nJ @ 30 fps HD 0.0117nJ @ 30 fps VGA

* Normalized energy [10] - equation shown in top left

Fig. 19. Voltage and frequency scaling of the design and comparison with state-of-the-art chips.

disparity search ranges that are insufficient to run the industrial standard KITTI benchmark. The chip is programmable and supports various frame rates and image resolutions. It consumes 836 mW at 30-frames/s full HD. Power scales to 55 mW at 30-frames/s VGA at low voltage (0.52 V). Normalized energy is an FoM used in [10] and [8]. This paper achieves 5.8× better FoM (energy per pixel per disparity) compared with other state-of-the-art works at 30-frames/s full HD resolution. Normalized energy scales to 0.0117 nJ at 30-frames/s VGA resolution, yielding 2.2× higher efficiency. Fig. 19 top right shows the frequency and voltage scaling. The maximum chip performance is 38 frames/s for full HD resolution.

V. SYSTEM INTEGRATION AND EVALUATION

To demonstrate a complete system, the chip is integrated with a camera general processing system and mounted on a real-time quadcopter platform. A small, light custom board is designed and fabricated to satisfy the “SWaP” requirements for system integration on MAVs. Fig. 20 provides the board specs. Our system consists of the stereo daughterboard with the chip on top (covered with black epoxy) and a motherboard with two Cypress USB bridges where USB signals are converted to the 32-bit parallel interface. Fig. 21 shows

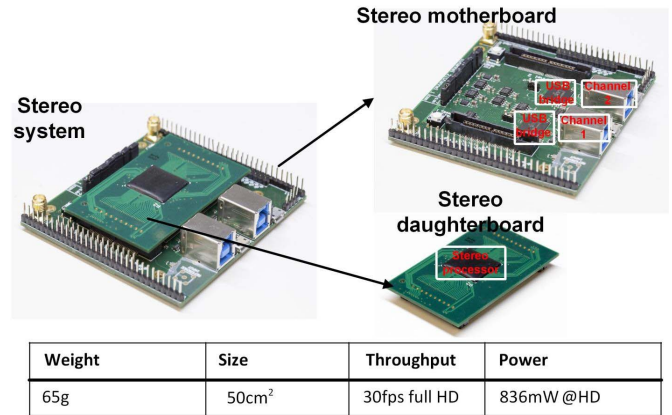


Fig. 20. Stereo system setup and summary.

the measurement setup and complete stereo system. The real-time image streams captured by the ZED stereo camera [23] are rectified, block partitioned into 50 × 50 blocks by a Samsung Exynos-5422 processor on the ODRIOD-XU4 [24] board, and then transmitted to the stereo processor through the input USB3.0 interface. Instructions are also sent to the chip with the same input USB3.0 interface that sustains the total

TABLE III
MEASURED SYSTEM POWER BREAK DOWN

	Stereo vision processor	2 Cypress FX3 USB bridge	Odroid XU4	ZED stereo camera
Power(W)	836mW	288mW	3.4W	1.9W*

*estimated from ZED camera [23]

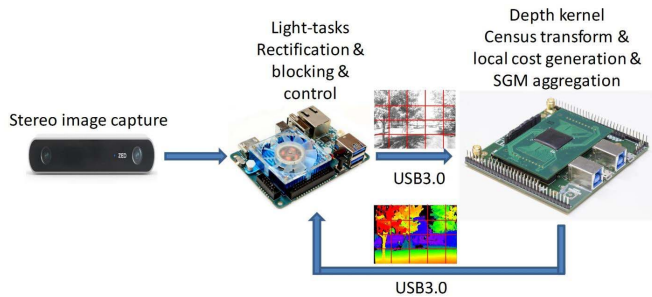


Fig. 21. Chip measurement setup with the stereo system.

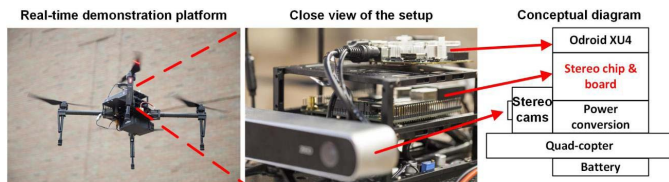


Fig. 22. Real-time quadcopter demonstration platform.

1.8 Gb/s bandwidth. The processed real-time depth images along with the “confidence” side information on each pixel are streamed back via the other (output) USB3.0 interface exhibiting 0.8 Gb/s bandwidth. Each 50×50 block is processed concurrently when the next block is being transmitted and stored on the on-chip interleaved image buffers. This technique minimizes camera-chip-depth latency. The real-time demonstration platform mounted on a quadcopter is shown in Fig. 22. At 0.9-V nominal voltage, the real-time VGA (full HD) frame processing latency of the stereo processor is 4.1 ms (26 ms), which is sufficient for real-time flight control. Table III shows the measured system power breakdown. The stereo vision board consumes $\sim 20\%$ of the system power. Fig. 23 shows the qualitative results measured from our own quadcopter scene. As seen in the left image, strong illumination on a sunny day leads to saturation of the sky and grass, however, the chip still generates accurate depth maps for navigation control.

VI. CONCLUSION

This paper presents a single-chip, accurate, high-performance, energy-efficient depth-estimation processor using the SGM algorithm for autonomous MAV applications. The fabricated processor generates 512 levels of depth in full HD (1920×1080) resolution with real-time 30-frames/s throughput consuming 836 mW from a 0.75-V supply in 40-nm CMOS. The chip reports 7% outlier on industry standard KITTI evaluation. The overlapping block-based processing achieves 95.4% memory reduction, eliminating the need

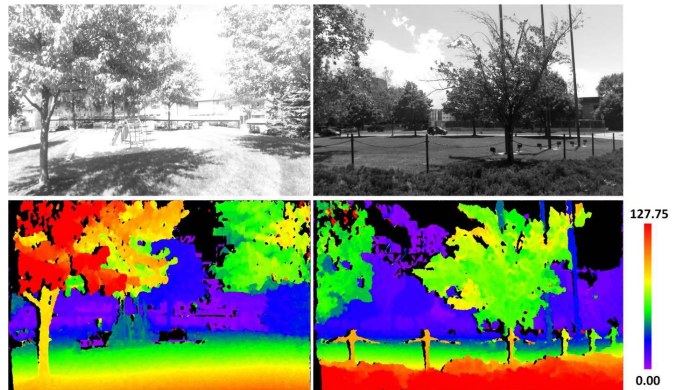


Fig. 23. Measured HD result with quadcopter.

for external DRAM at the cost of only 0.5% accuracy degradation. The proposed image-scanning stride with 16-stage deeply pipelined implementation yields $3\times$ performance gain, 25% additional memory reduction and enables processing 512-level depth output at 30 frames/s for full HD resolution. Customized ultra-wide SRAM enables 1.64-Tb/s on-chip memory access bandwidth with 18-mW power consumption. The chip is measured with industry standard benchmarks. A complete stereo system is built and demonstrated on a quadcopter for realistic real-time operations.

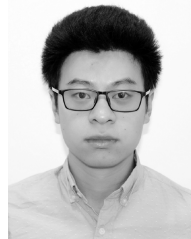
ACKNOWLEDGMENT

The authors would like to thank TSMC University Shuttle Program for chip fabrication.

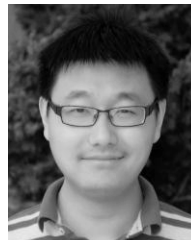
REFERENCES

- [1] W. R. Davis, B. B. Kosicki, D. M. Boroson, and D. Kostishack, “Micro air vehicles for optical surveillance,” *Lincoln Lab. J.*, vol. 9, no. 2, pp. 197–214, 1996.
- [2] *Velodyne LIDAR*. [Online]. Available: http://velodynelidar.com/docs/datasheet/63-9229_Rev-F_Puck%20Spec%20Sheet_Web.pdf
- [3] J. Wenger, “Automotive radar—Status and perspectives,” in *Proc. Compound Semiconductor Integr. Circuit Symp. (CSIC)*, 2005, pp. 21–24.
- [4] T. A. Krouskop, D. R. Dougherty, and F. S. Vinson, “A pulsed Doppler ultrasonic system for making noninvasive measurements of the mechanical properties of soft tissue,” *J. Rehabil. Res. Devised*, vol. 24, no. 2, pp. 1–8, 1987.
- [5] H. C. Wikle, III, S. Kottilingam, R. H. Zee, and B. A. Chin, “Infrared sensing techniques for penetration depth control of the submerged arc welding process,” *J. Mater. Process. Technol.*, vol. 113, nos. 1–3, pp. 228–233, 2001.
- [6] H. Cho, Y.-W. Seo, B. V. K. V. Kumar, and R. R. Rajkumar, “A multi-sensor fusion system for moving object detection and tracking in urban driving environments,” in *Proc. IEEE Int. Conf. Robot., Autom. (ICRA)*, Jun. 2014, pp. 1836–1843.

- [7] M. Hariyama and M. Kameyama, "VLSI processor for reliable stereo matching based on window-parallel logic-in-memory architecture," in *Symp. VLSI Circuits Dig. Tech. Papers*, Jun. 2004, pp. 166–169.
- [8] K. J. Lee *et al.*, "A 502 GOPS and 0.984 mW dual-mode ADAS SoC with RNN-FIS engine for intention prediction in automotive black-box system," in *IEEE Int. Solid-State Circuits Conf. (ISSCC) Dig. Tech. Papers*, Feb. 2016, pp. 256–257.
- [9] K. J. Lee *et al.*, "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," *IEEE J. Solid-State Circuits*, vol. 52, no. 1, pp. 139–150, Jan. 2017.
- [10] H.-H. Chen, C.-T. Huang, S.-S. Wu, C.-L. Hung, T.-C. Ma, and L.-G. Chen, "A 1920×1080 30 fps 611 mW five-view depth-estimation processor for light-field applications," in *IEEE Int. Solid-State Circuits Conf. (ISSCC) Dig. Tech. Papers*, Feb. 2015, pp. 1–3.
- [11] J. Park, S. Lee, and H.-J. Yoo, "A 30 fps stereo matching processor based on belief propagation with disparity-parallel PE array architecture," in *Proc. IEEE Int. Symp. Circuits Syst. (ISCAS)*, Jun. 2010, pp. 453–456.
- [12] S. K. Gehrig, F. Eberli, and T. Meyer, "A real-time low-power stereo vision engine using semi-global matching," in *Proc. Int. Conf. Comput. Vis. Syst.*, 2009, pp. 134–143.
- [13] J. Tanabe *et al.*, "A 1.9 TOPS and 564 GOPS/W heterogeneous multicore 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.
- [14] J. Park *et al.*, "A 646 GOPS/W multi-classifier many-core processor with cortex-like architecture for super-resolution recognition," in *IEEE Int. Solid-State Circuits Conf. (ISSCC) Dig. Tech. Papers*, Feb. 2013, pp. 168–169.
- [15] M. Z. Brown, D. Burschka, and G. D. Hager, "Advances in computational stereo," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 25, no. 8, pp. 993–1008, Aug. 2003.
- [16] J. Kim, S. Park, C. Chen, and H. Jeong, "Real-time architecture of stereo vision for robot eye," in *Proc. 8th Int. Conf. Signal Process.*, vol. 1, Nov. 2006.
- [17] C.-C. Cheng, C.-T. Li, C.-K. Liang, Y.-C. Lai, and L.-G. Chen, "Architecture design of stereo matching using belief propagation," in *Proc. IEEE Int. Symp. Circuits Syst.*, Jun. 2010, pp. 4109–4112.
- [18] Y.-C. Tseng, N. Chang, and T.-S. Chang, "Low memory cost block-based belief propagation for stereo correspondence," in *Proc. IEEE Int. Conf. Multimedia Expo*, Jul. 2007, pp. 1415–1418.
- [19] J. Xiang, Z. Li, H. S. Kim, and C. Chakrabarti, "Hardware-efficient neighbor-guided SGM optical flow for low power vision applications," in *Proc. IEEE Int. Workshop Signal Process. Syst.*, Oct. 2016, pp. 1–6.
- [20] J. Xiang, Z. Li, D. Blaauw, H. S. Kim, and C. Chakrabarti, "Low complexity optical flow using neighbor-guided semi-global matching," in *Proc. IEEE Int. Conf. Image Process.*, Sep. 2016, pp. 4483–4487.
- [21] A. Geiger, P. Lenz, and R. Urtasun, "Are we ready for autonomous driving? The KITTI vision benchmark suite," in *Proc. IEEE Int. Conf. Comput. Vis. Pattern Recognit.*, Jun. 2012, pp. 3354–3361.
- [22] Z. Li *et al.*, "A 1920×1080 30 fps 2.3 TOPS/W stereo-depth processor for robust autonomous navigation," in *IEEE Int. Solid-State Circuits Conf. (ISSCC) Dig. Tech. Papers*, Feb. 2017, pp. 62–63.
- [23] Y. Bi *et al.*, "An MAV localization and mapping system based on dual realsense cameras," *Int. Micro Air Vehicles, Conf. Competitions*, Nat. Univ. Singapore, Singapore, Tech. Rep., 2016.
- [24] J. Ivković, A. Veljović, B. Randelović, and V. Veljović, "ODROID-XU4 as a desktop PC and microcontroller development boards alternative," in *Proc. 6th Int. Conf. (TIO)*, Zrenjanin, Serbia, 2016.
- [25] Y. Ohta and T. Kanade, "Stereo by intra- and intra-scanline search using dynamic programming," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. PAMI-7, no. 2, pp. 139–154, Mar. 1985.
- [26] Y. Boykov, O. Veksler, and R. Zabih, "Fast approximate energy minimization via graph cuts," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 23, no. 11, pp. 1222–1239, Nov. 2001.
- [27] J. Sun, H.-Y. Shum, and N.-N. Zheng, "Stereo matching using belief propagation," in *Proc. Eur. Conf. Comput. Vis.*, 2002, pp. 510–524.
- [28] 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.
- [29] D. I. Barnea and H. F. Silverman, "A class of algorithms for fast digital image registration," *IEEE Trans. Comput.*, vol. C-100, no. 2, pp. 179–186, Feb. 1972.
- [30] P. Viola and W. M. Wells, III, "Alignment by maximization of mutual information," *Int. J. Comput. Vis.*, vol. 24, no. 2, pp. 137–154, Sep. 1997.
- [31] R. Zabih and J. Woodfill, "Non-parametric local transforms for computing visual correspondence," in *Proc. Eur. Conf. Comput. Vis.*, 1994, pp. 151–158.
- [32] C. D. Pantilie and S. Nedevschi, "SORT-SGM: Subpixel optimized real-time semiglobal matching for intelligent vehicles," *IEEE Trans. Veh. Technol.*, vol. 61, no. 3, pp. 1032–1042, Mar. 2012.
- [33] H. Hirschmüller, "Accurate and efficient stereo processing by semiglobal matching and mutual information," in *Proc. IEEE Int. Conf. Comput. Vis. Pattern Recognit.*, vol. 2, Jun. 2005, pp. 807–814.
- [34] D. Hernandez-Juarez, A. Chacón, A. Espinosa, D. Vázquez, J. C. Moure, and A. M. López, "Embedded real-time stereo estimation via semi-global matching on the GPU," *Procedia Comput. Sci.*, vol. 80, pp. 143–153, Dec. 2016.
- [35] H. Hirschmüller, M. Buder, and I. Ernst, "Memory efficient semi-global matching," *ISPRS Ann. Photogram., Remote Sens. Spatial Inf. Sci.*, vol. 3, pp. 371–376, Aug. 2012.
- [36] S. Baker, D. Scharstein, J. P. Lewis, S. Roth, M. J. Black, and R. Szeliski, "A database and evaluation methodology for optical flow," *Int. J. Comput. Vis.*, vol. 92, no. 1, pp. 1–31, 2011.
- [37] *USB Bridge Chip, FX3*. [Online]. Available: <http://www.cypress.com/file/140296/download>



Mr. Li was a recipient of the Best Paper Award at the 2016 IEEE Workshop on Signal Processing Systems.



Mr. Dong was a recipient of the Best Paper Award at the 2012 IEEE International Conference on Solid-State and Integrated Circuit Technology, the 2015 IEEE International Symposium on Circuits and Systems, and the 2016 IEEE Symposium on Security and Privacy.



During 2010–2015, he was a Research Engineer with STMicroelectronics Central R&D, Crolles, France, where he was involved in developing new adaptive solutions and ultra-low power digital design. In 2015, he joined the MICL as a Research Investigator, where he has been a Research Fellow since 2017. His current research interests include on-chip monitoring, adaptive techniques for variability tolerant designs, and near/sub-threshold energy efficient systems.

Ziyun Li (S'14) received the B.S. degree in electrical and computer engineering from the University of Michigan, Ann Arbor, MI, USA, in 2014, where he is currently pursuing the Ph.D. degree with the Michigan Integrated Circuit Laboratory.

His current research interests include high-performance, energy-efficient computer vision/machine learning processing units to enable next generation intelligent, autonomous navigation system.

Qing Dong (S'14) received the B.S. and M.S. degrees in microelectronics from Fudan University, Shanghai, China, in 2010 and 2013, respectively. He is currently pursuing the Ph.D. degree with the University of Michigan, Ann Arbor, MI, USA.

His current research interests include memory circuits design, monitoring circuits design for process variation and BTI.

Mr. Dong was a recipient of the Best Paper Award at the 2012 IEEE International Conference on Solid-State and Integrated Circuit Technology, the 2015 IEEE International Symposium on Circuits and Systems, and the 2016 IEEE Symposium on Security and Privacy.

Mehdi Saligane received the B.S. and M.S. degrees in electrical engineering systems and control from Ecole Polytechnique de Grenoble, Grenoble, France, in 2009, the M.S. degree in electrical engineering from Grenoble University, Grenoble, in 2011, and the Ph.D. degree in electrical engineering and computer science from the University of Aix-Marseille (IM2NP), Marseille, France, in 2016.

He was a Visiting Researcher at the Michigan Integrated Circuit Laboratory (MICL), University of Michigan, Ann Arbor, MI, USA, for two years. During 2010–2015, he was a Research Engineer with STMicroelectronics Central R&D, Crolles, France, where he was involved in developing new adaptive solutions and ultra-low power digital design. In 2015, he joined the MICL as a Research Investigator, where he has been a Research Fellow since 2017. His current research interests include on-chip monitoring, adaptive techniques for variability tolerant designs, and near/sub-threshold energy efficient systems.



Benjamin Kempke received the B.S.E., M.S.E., and Ph.D. degrees in computer engineering from the University of Michigan, Ann Arbor, MI, USA, in 2010, 2011, and 2017, respectively.

Since 2017, he has been with Kempke Engineering, Ann Arbor, providing consulting services in the areas of software-defined radio, low-power sensor design, and embedded engineering. His current research interests include the design of low-power and high-accuracy indoor RF localization technologies.



Luyao Gong is currently pursuing the Ph.D. degree in electrical and computer engineering with the University of Michigan, Ann Arbor, MI, USA.

Her current research interests include computer vision, low-power image compression, and machine learning.



Zhengya Zhang received the B.A.Sc. degree in computer engineering from the University of Waterloo, Waterloo, ON, Canada, in 2003, and the M.S. and Ph.D. degrees in electrical engineering from the University of California at Berkeley (UC Berkeley), Berkeley, CA, USA, in 2005 and 2009, respectively.

He has been a Faculty Member of the University of Michigan, Ann Arbor, MI, USA, since 2009, where he is currently an Associate Professor with the Department of Electrical Engineering and Com-

puter Science. His current research interests include low-power and high-performance very large scale integration (VLSI) circuits and systems for computing, communications, and signal processing.

Dr. Zhang was a recipient of the National Science Foundation CAREER Award in 2011, the Intel Early Career Faculty Award in 2013, the David J. Sakrison Memorial Prize for Outstanding Doctoral Research in Electrical Engineering and Computer Sciences at UC Berkeley, and the Best Student Paper Award at the Symposium on VLSI Circuits. He has been an Associate Editor of the IEEE TRANSACTIONS ON VERY LARGE SCALE INTEGRATION (VLSI) SYSTEMS since 2015. He was an Associate Editor of the IEEE TRANSACTIONS ON CIRCUITS AND SYSTEMS—PART I: REGULAR PAPERS from 2013 to 2015 and the IEEE TRANSACTIONS ON CIRCUITS AND SYSTEMS—PART II: EXPRESS BRIEFS from 2014 to 2015.



Ronald Dreslinski received the Ph.D. degree in computer science and engineering from the University of Michigan, Ann Arbor, MI, USA, in 2011.

He is currently an Assistant Professor with the EECS Department, University of Michigan. His current research interests include near-threshold computing, architectural simulator development, and high-radix on-chip interconnects.

Dr. Dreslinski was a Winner at the ISSCC'2011 student design contest for the design and test of

Centip3De, a 3-D stacked NTC processor. He received the Young Computer Architect Award from the Technical Committee on Computer Architecture of the IEEE Computer Society.



Dennis Sylvester (S'95–M'00–SM'04–F'11) received the Ph.D. degree in electrical engineering from the University of California, Berkeley, CA, USA, where his dissertation was recognized with the David J. Sakrison Memorial Prize as the most outstanding research in the UC-Berkeley EECS Department.

He has held research staff positions with the Advanced Technology Group of Synopsys, Mountain View, CA, USA, Hewlett-Packard Laboratories, Palo Alto, CA, USA, and visiting professorships

at the National University of Singapore, Singapore, and Nanyang Technological University, Singapore. He also serves as a Consultant and a Technical Advisory Board Member for electronic design automation and

semiconductor firms. He co-founded Ambiq Micro, a fabless semiconductor company developing ultra-low power mixed-signal solutions for compact wireless devices. He is currently a Professor of electrical engineering and computer science with the University of Michigan, Ann Arbor, MI, USA, and the Director of the Michigan Integrated Circuits Laboratory, a group of ten faculties and more than 70 graduate students. He has authored or co-authored over 450 articles along with one book and several book chapters. He holds 38 U.S. patents. His current research interests include the design of millimeter-scale computing systems and energy efficient near-threshold computing.

Prof. Sylvester was a recipient of an NSF CAREER Award, the Beatrice Winner Award at ISSCC, an IBM Faculty Award, an SRC Inventor Recognition Award, and ten best paper awards and nominations. He was named one of the Top Contributing Authors at ISSCC, most prolific author at the IEEE Symposium on VLSI Circuits, and was awarded the University of Michigan Henry Russel Award for distinguished scholarship. He serves on the technical program committee of the IEEE International Solid-State Circuits Conference and previously served on the executive committee of the ACM/IEEE Design Automation Conference. He has served as an Associate Editor of the IEEE JOURNAL OF SOLID-STATE CIRCUITS, the IEEE Transactions on CAD, and the IEEE Transactions on VLSI Systems and is an IEEE Solid-State Circuits Society Distinguished Lecturer for 2016–2017.



David Blaauw (M'94–SM'07–F'12) received the B.S. degree in physics and computer science from Duke University, Durham, NC, USA, in 1986, and the Ph.D. degree in computer science from the University of Illinois at Urbana-Champaign, Champaign, IL, USA, in 1991.

He was with Motorola, Inc., Austin, TX, USA, until 2001, where he was the Manager of the High Performance Design Technology Group. Since 2001, he has been a Faculty Member of the University of Michigan, Ann Arbor, MI, USA, where he is

currently a Professor. His research has a threefold focus. He has investigated adaptive computing to reduce margins and improve energy efficiency using a new approach he pioneered, called Razor, for which he received the Richard Newton GSRC Industrial Impact Award and the IEEE Micro Annual Top-Picks Award. He has extensive research in ultra-low-power computing using subthreshold computing and analog circuits for millimeter sensor systems and for high-end servers, his research group and collaborators introduced so-called near-threshold computing, which has become a common concept in semiconductor design. This work led to a complete sensor node design with record low-power consumption, which was selected by the MIT Technology Review as one of the year's most significant innovations. Most recently, he has pursued research in cognitive computing using analog, in-memory neural-networks. He has authored or co-authored over 500 papers and holds 60 patents.

Prof. Blaauw was a recipient of the 2016 SIA-SRC faculty award for lifetime research contributions to the U.S. semiconductor industry, and the Motorola Innovation award, and has received numerous best paper awards and nominations. He was the General Chair of the IEEE International Symposium on Low Power, the Technical Program Chair of the ACM/IEEE Design Automation Conference, and serves on the technical program committee of the IEEE International Solid-State Circuits Conference.



Hun-Seok Kim (S'10–M'11) received the B.S. degree in electrical engineering from the Seoul National University, Seoul, South Korea, and the M.S. and Ph.D. degrees in electrical engineering from the University of California, Los Angeles (UCLA), Los Angeles, CA, USA.

He was a Technical Staff Member of Texas Instruments Inc., from 2010 to 2014, where serving as an industry liaison for multiple university projects funded by the Semiconductor Research Corporation and Texas Instruments Inc. He is

currently an Assistant Professor with the University of Michigan, Ann Arbor, MI, USA. He holds nine granted patents and has over ten pending applications in the areas of digital communication, signal processing, and low-power integrated circuits. His current research interests include system novel algorithms and efficient very large scale integration architectures for low-power/high-performance signal processing, wireless communication, computer vision, and machine learning systems.

Dr. Kim was a recipient of multiple fellowships from the Ministry of Information and Telecommunication, South Korea, Seoul National University, and UCLA.