A 1920 × 1080 25fps, 2.4TOPS/W Unified Optical Flow and Depth 6D Vision Processor

for Energy-Efficient, Low Power Autonomous Navigation

Ziyun Li, Jingcheng Wang, Dennis Sylvester, David Blaauw, and Hun-Seok Kim

University of Michigan, Ann Arbor, MI, liziyun@umich.edu

Abstract

This paper presents a unified 6D vision (3D coordinate + 3D motion) processor that is the first to achieve real-time dense optical flow and stereo depth for full HD (1920×1080, FHD) resolution. The proposed design is also the first ASIC that accelerates real-time optical flow computation. It supports a wide search range (176×176 pixels) to enable dense optical flow on FHD images with real-time 25fps/30fps (flow/depth) throughput, consuming 760mW in 28nm CMOS.

Introduction

Autonomous navigation of micro aerial vehicles (MAVs) and selfdriving cars (Fig. 1) requires real-time accurate and dense perception of 3D coordinates (depth) and 3D apparent motion (optical flow). The semi-global matching (SGM) [1] algorithm is widely used [2,3] in real-time stereo depth estimation and achieves high accuracy by applying a semi-global optimization over the entire image. However, directly applying SGM to optical flow is impractical for real-time applications because each pixel requires evaluation of d^2 (= 31k for d = 176 in our chip) flow candidates as well as SGM on each matching candidate. Recently, a variant of SGM for optical flow, NG-fSGM [4], was proposed to avoid the quadratic complexity by performing neighbor-guided aggressive candidate pruning. However, NG-fSGM still requires a large memory footprint (~4MB) with massive memory access bandwidth (2.6Tb/s). Because of aggressive neighbor guided pruning and randomly added candidates to the search space [1], the memory access pattern is highly irregular and inter-pixel dependency has variable latency. Addressing these challenges, the proposed design uses: 1) four new custom designed 16×16 (136b/access) coalescing crossbar (xbar) switches that efficiently eliminate redundant memory accesses with built-in arbitration at each cross point for 4.08Tb/s peak on-chip memory bandwidth; 2) a new rotating buffer to maximize on-chip memory reuse for a wide 2D search range; 3) a diagonal image-scanning stride to resolve the variable length inter-pixel dependency in the deep pipeline.

Proposed Processor Architecture

As shown in Fig. 2, the processor streams previous and current image blocks into 64 on-chip rotating image buffers (20 Kb each, Fig. 3). Pixel matching is performed using census transformed pixels (80b each) of the current vs. previous (optical flow) or left vs. right frames (depth). Rather than exhaustively evaluating all candidates, each pixel selectively evaluates 60 candidates guided by previously processed neighboring pixels plus 4 random positions (64 total). This produces 64 'local' matching costs (7b each) for the 64 optical flow/depth vectors. The processor then aggregates the local matching costs along 8 paths. A subset of optical flow/depth vectors with the minimum aggregated cost for each path are passed to next processing pixel as flow/depth candidates. The minimum of combined aggregated costs is selected as the final optical flow/depth.

To eliminate the need for off-chip DRAM and to ensure scalability to different image resolutions, the proposed design processes the input in units of 88×88 overlapping blocks. Adjacent blocks are overlapped by 8 pixels to allow candidate propagation across block boundaries [2,4]. However, block-based processing incurs significant memory inefficiency on optical flow computation because, unlike 1D disparity matching, the same blocks are read multiple times for 2D matching to process different parts of the image. Therefore, we propose a rotating on-chip image buffer scheme with 64 SRAMs (Fig. 4) to maximize on-chip memory reuse and reduce interface bandwidth by $2 \times$ at the cost of 28% larger on-chip memory size.

Furthermore, processing the image in the conventional raster scan order results in a severe dependency for NG-fSGM where the previous pixel processing must be completed to pass candidates to the adjacent pixel (Fig. 5). This issue is even more critical in NG-fSGM where pixel processing has variable latency. Inspired by [2], we process

pixels diagonally to hide the variable length dependency in the pipeline, achieving better frequency and voltage scalability from a shorter critical path.

Although NG-fSGM aggressively prunes the search space, it still requires a tremendous memory bandwidth of >2Tb/s with irregular and unpredictable memory access patterns (Fig. 2). This motivates a new multi-bank memory access with a coalescing crossbar discussed below. Once pixel data is fetched, NG-fSGM computation can be highly parallelized. Therefore, the proposed 6D vision processor is partitioned into 3 clock domains and 2 voltage domains to optimize and balance compute vs. memory access performance and power (Fig. 3). Image access, census transform, and local cost computation are in the (Fhigh, Vhigh) domain to improve throughput, while cost aggregation and flow selection are performed with (Flow, Vlow) to improve energy efficiency. Fig. 7 shows that crossbar and memory accesses account for ~62% of overall power (simulation).

High Bandwidth, Coalescing Xpoint Crossbar

Fig. 6 (top) shows the proposed architecture and timing of the custom-designed high-bandwidth coalescing xbar. Fig. 6 (bottom) details the circuits of a cross point (Xpoint). In the proposed design, computing census transform of 64 flow candidates on-the-fly requires 2.6Tb/s random memory access. To provide such a high throughput with an area-/power-efficient solution, we use 4 custom 136b/word, 2-cycle pipelined coalescing xbars that enable 4.08Tb/s peak memory bandwidth. NG-fSGM memory accesses are highly irregular but neighboring pixels tend to generate overlapping accesses. Only ~55 out of 162 accesses are unique on average (simulation). Our crossbar detects and arbitrates memory access (query) collisions at each Xpoint, and it broadcasts memory data and address to queries that lost arbitration. In the first cycle, each Xpoint decides which query wins or loses arbitration by examining neighboring arbitration lines and it launches the access address for the winning query. In second cycle, both winning query and losing queries receive the broadcasted data and address. If a query in the queue has a matching address, the read data is consumed and the query is removed from the queue (to avoid redundant memory accesses). This coalescing yields 54% performance improvement. Fig. 8 compares different xbar design choices. The proposed four 16×16 coalescing xbars achieve an access throughput comparable to that of a single 64×64 xbar while achieving $3.1 \times$ reduction in power and $3.4 \times$ reduction in area. Overall, this approach enables 2.6Tb/s average bandwidth from 4 xbars.

Measurement Results and Comparisons

The processor was fabricated in TSMC 28nm CMOS (Fig. 9). Realtime images are block-partitioned and streamed to the processor through a USB3.0 interface. Figs. 10 and 11 show the measured optical flow and depth results for the KITTI automotive and JPLcaptured MAV benchmarks, respectively. Fig. 12 shows the measured throughput and energy efficiency trade-off with varying Fhigh. The optimal energy point occurs at Fhigh≈2.7×Flow. Fig. 13 shows the voltage and frequency scaling behavior of the chip. Table 1 and Fig. 15 provide comparison to prior works. The proposed processor consumes 760mW to process optical flow/depth on 25/30fps FHD images. Normalized energy [6] (Fig. 15) on depth @ 30fps FHD images is 0.069 nJ/pixel, marking >1.5× improvement over prior arts specialized for depth processing only. Normalized energy on optical flow @ 25fps FHD is 0.0048, vielding additional 14.3× improvement while all prior work is unable to support optical flow computation.

Acknowledgement

We thank TSMC for chip fabrication, and JPL for MAV benchmark.

References

Hirschmuller, CVPR, pp.807-814, 2005
Li, ISSCC, pp. 62-63, 2017 [5] Xiang, SiPS, pp. 1-6, 2016
Lee, ISSCC, pp. 256-257, 2016 [6] Chen, ISSCC, pp. 422-423, 2015
Xiang, ICIP, pp. 4483–4487, 2016

