A 95pJ/label Wide-Range Depth-Estimation Processor for Full-HD Light-Field Applications on FPGA

Li-De Chen, Yu-Ta Lu, Yu-Ling Hiao, Bo-Hsiang Yang, Wei-Chi Chen, and Chao-Tsung Huang
Department of Electrical Engineering
National Tsing Hua University, Hsinchu, Taiwan
Email: ldchen@gapp.nthu.edu.tw

***Abstract*—High-resolution and wide-range depth maps are the key to enable novel light-field applications, such as digital refocusing, view synthesis, and 3D reconstruction. In this paper, we present an energy-efficient depth-estimation processor on FPGA to meet this purpose. There are two major contributions. First, image-guided depth inference and upsampling is adopted and implemented to provide accurate depth maps while lowering the working frequency from 215MHz to 54MHz. Second, octave search range sampling is proposed to efficiently allocate depth labels for wide-depth-range scenes to save computation and maintain accuracy. Finally, the implementation result on Xilinx ZC706 shows ASIC-comparable energy efficiency—95pJ/label—for Full-HD five-view light fields at 30fps.**

Keywords—light field, depth estimation, FPGA, Full HD

I. INTRODUCTION

Multi-view stereo matching can provide high-quality depth maps for high-resolution light-field applications. However, its demanding computation complexity could introduce high implementation cost and high power consumption and thus defy its usage on embedded devices. For example, [1] requires 1.5M gates in 40nm CMOS for Full-HD five-view depth estimation and consumes 611mW, i.e. 153pJ per depth label. In addition, as image resolution goes up the corresponding depth range will become wider. Accordingly, conventional designs will need to increase the number of depth labels and therefore demand more hardware resources, such as computation logics, on-chip SRAM size, and off-chip DRAM bandwidth.

In this paper, we address these issues mainly by two novel design approaches. One is image-guided depth inference and upsampling, which allows us to perform stereo matching using half-resolution views to save computation and then recover the depth resolution back by referencing the full-resolution center view. The other one is octave search range sampling, which smartly allocates the limited number of depth labels for wide-range scenes to save hardware resources. In the following, we will present the details of their designs and also the final implementation on Xilinx ZC706 FPGA.

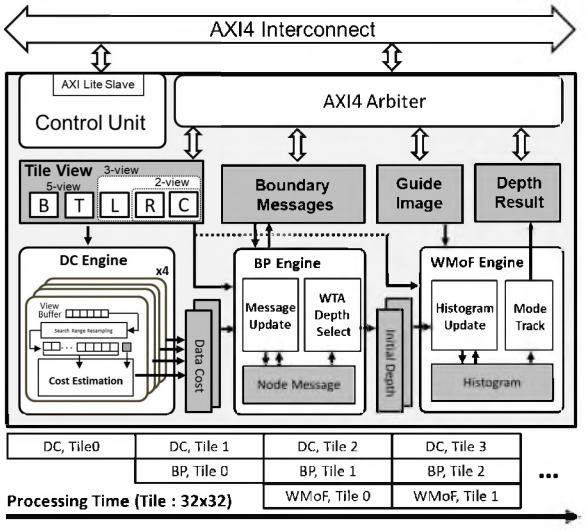

Fig. 1. System block diagram and tile-based pipelining.

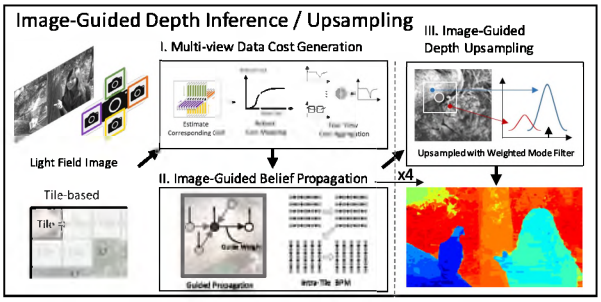

Fig. 2. Computation flow of the image-guided inference and upsampling.

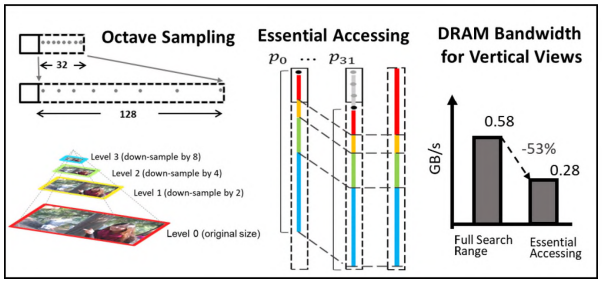

Fig. 3. Octave search range sampling and its bandwidth-efficient accessing.

II. Depth Estimation Processor

Fig.1 shows the system diagram of the processor. It follows a 32x32 tile pipeline and consists of three major engines: data cost (DC), belief propagation (BP), and weighted mode filter (WMoF). The DC estimates and aggregates data costs from four half-resolution surrounding views for each depth label. The BP performs intra-tile belief propagation and estimates a half-resolution depth map based on minimum costs. Finally, the WMoF upsamples and refines the depth map back to the full resolution using the high-quality center view.

A. Energy-Efficient Design with Image-Guided Upsampling

Fig. 2 shows the computation flow of the image-guided depth inference and upsampling for the three engines. To preserve object boundaries in full resolution, we adopt WMoF for depth upsampling. In contrast to the universal 4x upsampling in [2], we implemented 2x upsampling (in side length) in this processor to preserve more texture details. As a result, we can use the same amounts of processing elements in DC and BP as [1] while reducing the working frequency to 25% for saving power consumption. To increase the half-resolution depth quality, we also apply image-guided BP, i.e. conditional random fields, to enhance depth sharpness, and the hardware overhead is negligible. In addition, the DC engine is also optimized by only using the luma channel to save 78% of gates.

B. Memory-Efficient Design for Octave Sampling

Conventional stereo matching samples depth labels based on equally spaced disparities; however, this will overly sample near regions with unnecessarily high depth accuracy and thus waste lots of computation power. To address this issue, we proposed the octave search range sampling. It samples the depth labels with equally spaced depth (inverse disparity) to enable wide-depth-range stereo matching.

However, the memory access, including on-chip SRAM and off-chip DRAM, for such wide depth range becomes a problem for multi-view stereo matching. To resolve this memory bottleneck, we adopted the image pyramid and devised an essential accessing method as shown in Fig. 3. Combined with an on-the-fly downsampling unit, only the essential regions of the search range are required to be stored on SRAM. As a result, we save 53% of on-chip SRAM size and external DRAM bandwidth for accessing surrounding views.

III. Implementation Results and Conclusion

The proposed design is implemented on Xilinx ZC706 platform. It occupies 167K LUTs, 67K Registers, and 329K bytes of BRAM. Table I shows the resource utilization of the whole demo system and the depth estimation core. The depth estimation processor operates at 54 MHz and consumes 384 mW based on Xilinx Power Estimator. It equivalently achieves 95pJ per depth label for Full-HD 30fps.

A depth display system for the proposed processor is also implemented for verification and demonstration purpose.

Table I. Utilization Report

	LUT	LUTRAM	FF	BRAM	DSP
Available	277400	108200	554800	755	2020
Whole System	175423 (63.2%)	710 (0.7%)	74376 (13.4%)	240.5 (31.9%)	274 (13.6%)
Depth-Estimation Core	167206	48	67653	240.5	274

Table II. Comparison

	ISSCC 2015[1]	VLSI2017[3]	TCSVT2015[4]	This work
Technology / FPGA	40nm	28nm	Altera 5SGSMD5K2	Xilinx ZC706
Clock	215 MHz	300 MHz	180 MHz	54 MHz
Logic or [LUT, Registers]	1.5M gates	3.2M gates	[222K, 149K]	[167K, 67K]
Memory	352K bytes	582.5K bytes	2M bytes	329K bytes
Algorithm	MRFI	SGM	SGM	MRFI + Upsampling
Stereo type	5	3	2	5
Depth label / Range	64 / 32	128 / 128	128	64 / 128
Throughput	1920 x 1080, 30fps	2048 x 1080, 32fps	1600 x 1200, 42fps	1920 x 1088, 30fps
Power	611 mW	380 mW	-	384 mW*
Energy Efficiency**	153 pJ	42 pJ	-	95 pJ

* The Power result is estimated from Xilinx Power Estimator
** Energy Efficiency = Core Power / (Frame Rate * Depth Image Resolution * number of depth labels)

Integrated with ARM CPU and other peripheral ICs on ZC706 platform, we provide a demo of depth estimation for real-world light-field scenes. The tested light field is captured by Lytro Illum.

Table II shows the performance comparison between the proposed work with previous works [1,3,4]. The proposed processor improves resource and BRAM utilization by 63% and 84% respectively compared to [4]. It also achieves comparable energy efficiency on FPGA to the ASIC implementation [1].

In conclusion, an energy-efficient and cost-effective depth estimation processor is proposed. It is implemented on FPGA with real-world light-field demo. The proposed framework using image-guided depth inference and upsampling enables Full-HD depth throughput at merely 54MHz. The proposed octave search range sampling enables wide-range depth estimation for high-resolution light fields and also alleviates hardware cost for multi-view stereo matching. This work demonstrates ASIC-comparable energy efficiency, which facilitates the use of multi-view Full-HD depth estimation on FPGA and also ASIC.

Acknowledgement

This work was supported by Novatek Microelectronics Corp.

References

[1] H. Chen, et al., "A 1920x1080 30fps 611mW five-view depthestimation processor for light-field applications", in IEEE ISSCC Dig. Tech. Papers, pp. 422-423, 2015.

[2] B.-H. Yang, et al.,"A 320M Pixel/s VLSI Architecture Design of Weighted Mode Filter for 4K Ultra-HD Depth Upsampling", in IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP), 2018.

[3] Lee, Jinsu, et al. "A 31.2 pJ/disparity· pixel stereo matching processor with stereo SRAM for mobile UI application." VLSI Circuits, 2017 Symposium on. IEEE, 2017.

[4] Wang, Wenqiang, et al. "Real-time high-quality stereo vision system in FPGA." IEEE Transactions on Circuits and Systems for Video Technology 25.10 (2015): 1696-1708.