

Real-time Hybrid Stereo Vision System for HD resolution disparity map

Jiho Chang
changjh@etri.re.kr
Jae-chan Jeong
channij80@etri.re.kr
Dae-hwan Hwang
hdh@etri.re.kr

Intelligent Cognitive Technology Research Department
Electronics and Telecommunications Research Institute
Daejeon, Republic of Korea

Stereo matching is a traditional method used to obtain 3D depth information and has been studied for decades. However, it is still difficult to apply stereo matching algorithms to practical devices due to real-time issues as well as the technique's inability to adequately handle untextured regions. In this paper, we propose a hybrid stereo matching system to remedy the disadvantages of active and passive stereo vision.

Stereo matching algorithm Following Scharstein's taxonomy [5], the stereo matching algorithm divides into four steps: matching cost computation, cost aggregation, disparity computation and disparity refinement. First of all, we calculate raw cost volume using the AD-Census [2, 6]. At this time, we use cost combining with alpha-blending for the AD-Census, and the final raw cost that is the sum of the pattern cost (T_1) and the non-pattern cost (T_2). The information permeability filtering(PF) proposed by Cevahir Cigla *et al.* [2] is an ASW approach that has simple parameters and provides constant operational time for calculating cost aggregation. However, because there is no proximity weight term, PF can encounter problems with images containing large untextured regions. Modified information permeability(MPF), including a proximity weight term, is defined in Section 2.2. Our proposed system uses WTA as disparity computation, because it is very simple algorithm.

Stereo vision system design The proposed system is composed of the stereo head and the stereo emulator. The stereo head includes a LVDS module, an LD/LED projection module and two CMOS sensors. Input from the two CMOS sensors is received in the form of 10-bit monochrome images at a resolution of 1280 x 720 pixels at a rate of 60fps. Image streams from the left & right cameras are transferred to the FPGA board through the LVDS module, which includes control signals. The stereo emulator is based on FPGA modules to obtain disparity maps from stereo pairs. The stereo emulator contains a deserialized module, a USB 3.0 controller and four FPGAs. The USB 3.0 controller transfers the results to the computer and is received parameters for FPGA. Figure 1 (a) represents our entire system. The hybrid system captures a pair of pattern images and an alternating pair of non-pattern images to evaluate disparity. The CMOS sensors are synchronized with LD/LED projection module, so that the images are obtained in operating time with the LD (pattern) in one frame and with the LED (non-pattern) in the next frame.

Implementing stereo algorithm Figure 1 (b) is a block diagram of the hybrid stereo vision system. The stereo matching algorithm consists of four processing elements implemented in each single FPGA. The PrePE is first composed through rectification(based on the Caltech method) and image filtering(based on bilateral filtering). The MPE Left generates a left-referenced disparity that is implemented using the algorithm described in Section 2. Certain modifications are taken into account while implementing the system on the FPGA, and usually appear in the cost aggregation. One of the issues is whether the operations can be processed in a limited clock cycle. In order to solve this problem, we allow as much parallel processing, pipeline insertion and using LUT. Also, we configure to a power of two to use the data shifter instead of the divider. Another issue is whether to use the PF 4way in cost aggregation. When implement-

ing the FPGA in cost aggregation using the PF, data from the entire image must be saved, and hence involves cost volume. If horizontal cost volume data is assumed to be 16-bit, 4-direction processing is needed, which requires at least 570MB. For the above reasons, other research[1, 4] implements FPGA in only the horizontal direction. However, this causes streak noise and thus increases the error due to disparity. Our system is configured to run 3way(including top-to-bottom) while maintaining consistent performance and reducing memory consumption. In this case, we use memories for two lines of aggregated cost and two lines of non-pattern images. In PostPE, we use left-right consistency check and weighted median filtering to eliminate error pixels from the results of disparity calculation. Sub-pixel estimation is implemented as a parabola fitting the costs and is calculated to divide 4 bits more than 256 steps that are separated by a maximum disparity. So, steps of final disparity are 4096(256x16).

Algorithm		error(%)
Not using pattern image (Passive stereo matching)	PF 4way	61.67%
	MPF 4way	54.54%
Using pattern image (Hybrid system)	PF 2way(Horizontal)	11.39%
	MPF 3way for FPGA	3.80%

Table 1: Error pixel rate on non-occlusion regions.

Performance Evaluation We evaluate performance in two ways: quantitative evaluation using the ground truth and qualitative evaluation using Microsoft Kinect. The ground truth is created using a space-time stereo [3]. The performance of the system is then analyzed by using different cost aggregations and comparison between passive stereo matching and hybrid system. When applying the hybrid approach, the MPF has lesser streak noises than the conventional real-time two-way PF. We also see that the MPF 3way for FPGA performs just as well as the MPF 4way, except for the fact that the sharpness of the object is different.

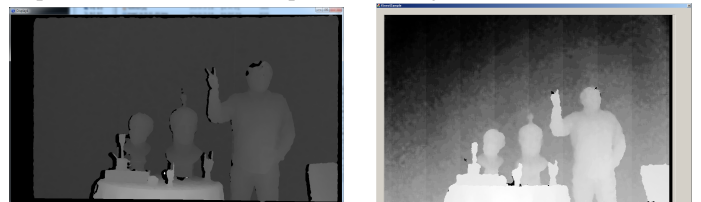


Figure 2: Comparison with Kinect, Left:proposed system, Right:Kinect

We proposed a hybrid stereo matching system that combines active and passive stereo vision. Using the active pattern, our system successfully detected disparity in untextured regions. Through comparison with other algorithms using the ground truth and with Microsoft Kinect, we found that our proposed system shows a significant improvement over current systems in processing untextured regions, and accurately calculates depth for a 1280 x 720 image at 60fps in indoor environments.

- [1] A. Aysu, M. Sayinta, and C. Cigla. Low cost fpga design and implementation of a stereo matching system for 3d-tv applications. In *VLSI-SoC*, 2013.
- [2] C. Cigla and A.A. Alatan. Efficient edge-preserving stereo matching. *ICCV Workshops*, 2011.
- [3] J. Jeong, H. Shin, J. Chang, E. Lim, S. Choi, K. Yoon, and J. Cho. High-quality stereo depth map generation using infrared pattern projection. *ETRI Journal*, 2013.
- [4] S. Mattoccia. Stereo vision algorithms for fpgas. In *CVPR Workshops*, 2013.
- [5] D. Scharstein, R. Szeliski, and R. Zabih. A taxonomy and evaluation of dense two-frame stereo correspondence algorithms. In *SMBV*.
- [6] Xun Sun, Xing Mei, shaohui Jiao, Mingcai Zhou, and Haitao Wang. Stereo matching with reliable disparity propagation. In *3DIMPVT*.

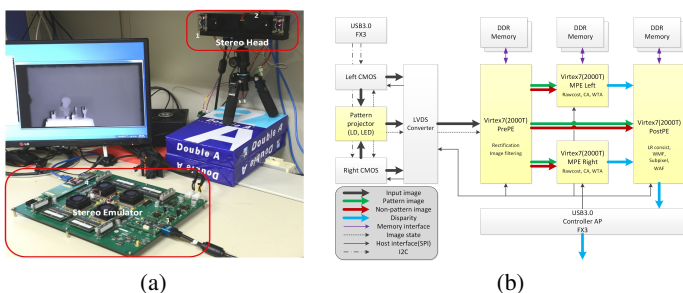


Figure 1: (a) The entire system with stereo head and stereo emulator (b) Block diagram of proposed stereo vision system