

# Hardware implementation of a census-based stereo matching using FPGA

Jiho Chang, Seung Min Choi, Eul-Gyoon Lim, Jae-il Cho

*Electronics and Telecommunication Research Institute, Korea*

*(Tel : 82-42-860-6131; Fax : 82-42-860-6796)*

*{changjh, ccsmm, eglim, jicho}@etri.re.kr*

**Abstract:** The real-time stereo vision is becoming important increasingly in the field of Robotics. It is very difficult to implement the real-time stereo vision system, because it requires very powerful processors. For solving this problem, we present the real-time hardware architecture of the census-based stereo matching IP (Intellectual Property) comprising support-weight and trellis dynamic programming structure. We use census-based cost that is robust for difference in brightness. Using Cost aggregation with support-weight window, our stereo vision system has more robustness for real environment. The stereo matching processor is implemented on a FPGA with the pre-processing part for the rectification and post-processing for reduction of noise. This stereo vision IP is implemented with HDL language and performed up to 30 FPS.

**Keywords:** Stereo matching, Dynamic programming, Census transform

## I. INTRODUCTION

The stereo vision is becoming important increasingly in the field of virtual interface, gesture recognition, robot navigation, depth measurements and environment reconstruction as well as in many other aspects of security, robotics, and entertainment[1, 7]. In order to adapt 3D depth on these various applications, many different algorithms have been developed and materialized in various systems.

Especially in robots, the real time is a very important issue since the circumstances are changing continuously and many of studies have been progressed in various ways to indicate the stereo vision in real time base[7]. However, due to the complexity of matching algorithm, it is very difficult to materialize in real time base and this may require very powerful processors.

Normally, many real time systems use the local method. Although it has low complexity, there are some local problems where it fails to match due to occlusion, uniform texture, ambiguity of low texture etc. Also, in the popular local matching method, the block matching skill blurs disparity data in the object boundary[3, 8]. Through these results, though it is practically usable in simple applications, they bring many restrictions on the applications which require more accurate data. Recently, a few real-time global matching methods have been implemented through GPU in the graphic card or MMX of CPU[6, 9]. But, in the mobile robot which was composed with the small embedded system, since the size and power usage is the important issue, it is

difficult to materialize using the PC or GPU. If there were cases of materialization using high efficiency embedded processors, but, the real time performance (30fps) cannot be guaranteed as well, since it is the use of the resources of main processor, other applications in embedded system cannot be performed.

In our previous work, census-based stereo matching IP which is based on trellis dynamic programming[2]. For the case of previous work, since it uses census transform, is robust for difference in brightness. But there are some steak noises, because it is based on dynamic programming and there are some calibration errors.

In this paper, our stereo matching algorithm has support-weighted cost aggregation and census-based stereo matching IP. For the case of Trellis DP algorithm, since it is strong for the occlusion by generating center referenced disparity, it is very effective under the practical circumference[4, 5]. And cost function of the Stereo matching IP use census transform, so that is robust for difference in brightness. Also we aggregate the matching cost by adaptive Support weight that is used for decreasing steak noise.

The stereo matching IP has three parts. First, Pre-processing part is composed with the Rectification, Calibration and controlling the brightness which affect to the result a lot in stereo matching. A second part is main-processing which is consisted of census transform, support weight and trellis DP algorithm. At last, post-processing which reduces the noise that is being generated in Dynamic programming.

In this paper, we described for the total block diagram and the performance and conclusion through the result gained from this system.

## II. Stereo matching IP

In this paragraph, we described the stereo matching IP which has three part - pre-processing, main-processing, post-processing.

### 1. Pre-processing part

Pre-processing part can improve the performance of stereo matching by minimizing the distortion of camera using the functions of individual brightness control and rectification of left and right images. The constraint of one point in one side image in stereo image geometry corresponds on epipolar line in other image. Therefore, due to the constraint of this epipolar line, the more epipolar line of the two left and right images matches each other accurately, the more the accuracy is increased. Since the stereo vision processing technology in this paper composed the algorithm subject to this epipolar line constraint, input after matching the left and right image accurately before it is input on the stereo matching part affects on the performance a lot.

In order to match this epipolar line, extract the rectification parameter in various ways. Thus save the pre computed parameters in the pre-processing register files and rectify the images by paralyzing the epipolar line using the above parameters. Also, in order to control the difference of physical characteristics of left and right camera or brightness of left and right images occurred by the strong light input in one camera only, the brightness control function is applied. Above functions are set to input the revised information continuously matching to the characteristic and circumstances through the host interface.

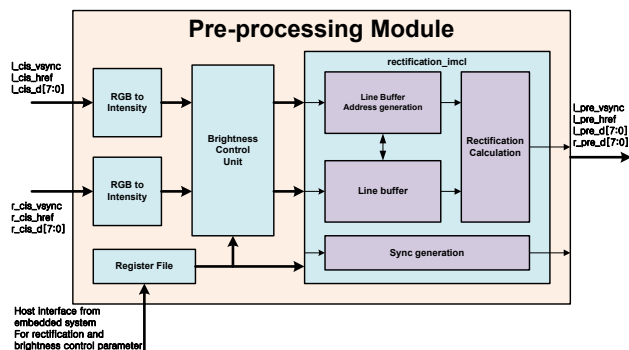


Fig.1. Block diagram of pre-processing part.

Figure 1. shows pre-processing block diagram. There are n-lines buffer memory, rectification parameter register and Calculation unit. This line buffer size depends on user requirement and mechanical structure of stereo camera. Calculation unit computes using a rectification formula described on Luping An et al.'s paper[12].

### 2. Main processing part

Main-processing part calculates the disparity from the rectified left and right input image which was processed at pre-processing module and makes disparity information as the intensity data. Following is the algorithm for practical performance at stereo matching module.

One of the deciding methods of the standard image of the 3 dimensional images in stereo vision is to select one of the left or right images and the other case is to define a virtual one(center reference disparity) in between those two. When express the depth map of image, there is a balance of these expressions if use the center referenced disparity while there are much of image loss for discontinuity of the other side if show the information using one of the images only. Also, it has many good points in interpreting the discontinuous space and finding the solution.

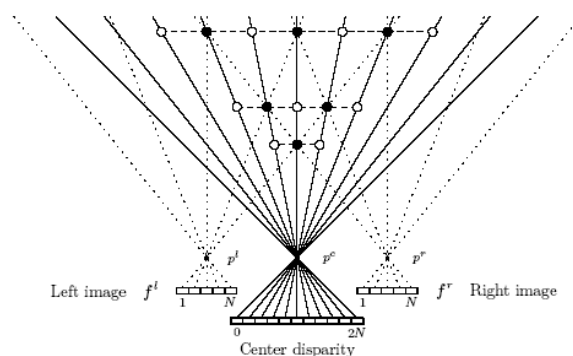


Fig. 2. Center referenced coordinate systems.

Figure 2. shows how the center referenced disparity which became the base of the used algorithm in stereo matching module can be formed. The coordinates of center referenced disparity includes the territories that can be generated by both of the left and right images and if it is composed with each of image N columns, it is verified that the center disparity image has  $2N + 1$  indexes.

In Figure 2., white/black dots represent stereo matching element that is consisted of cost computation

and cost aggregation. Also, the trellis structure is connected to those dots, has optimal path minimized energy. And we find the optimized route in the space using Dynamic Programming.

The block dots combines matching cost computation and cost aggregation. We use Census transform as matching cost computation and fixed weight window for cost aggregation. The Census Transform(CT) is a stereo matching cost algorithm with high robustness to illumination variations and exposure difference. The CT transforms the intensity data into feature data, before the matching costs of each disparity level are calculated. For a higher accuracy of the algorithm, the calculation of the matching costs can be followed by a further aggregation

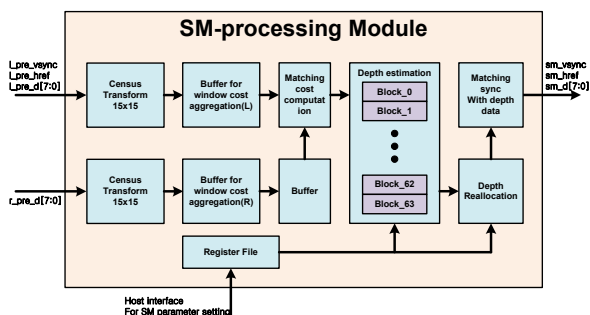


Fig. 3. Block diagram of main-processing method A.

Figure 3. is one method of stereo matching processing module which calculates disparity. Main-processing part is consist of census transform module to create raw costs, cost aggregating for support weight window computation and depth estimation module with trellis DP. But, this structure of block diagram use lots of Flip-Flops and memory for saving census data.

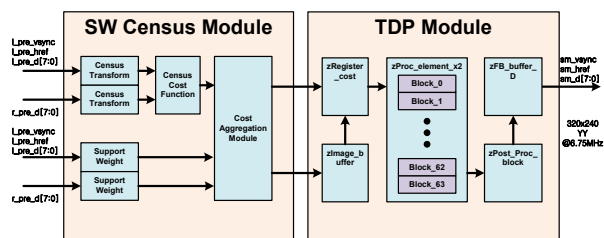


Fig. 4. Block diagram of main-processing method B.

Figure 4. is another method for SW census-based trellis dynamic programming. This method is used more memory than method A for saving Census raw cost. But this structure uses less Flip-Flops than method A, So

this method is efficient for implementing hardware IP. Even if a larger size of the window



Fig. 6. Result images from main-processing part.

Figure 6. are sample results by using Normal Dynamic programming(left) and SW Census DP(right). This results represents that SW census cost function make robust system in real world.

### 3. Post-processing part

Post-processing part is designed to remove the noise from the disparity map which was acquired from Stereo matching part and distinguish the objects and output each of them. Especially, in the stereo matching using Dynamic programming technology, the steak noise in horizontal direction will be generated. For this kind of noise that is generated to outside of object cases, they can be treated at backward application step. However, for the case it is generated to inside of the object, it makes the object to be shown as disconnected. In order to prevent these phenomena, it tried to reduce with 3 steps of filtering on the result from the Main-processing part.



Fig. 7. Results of post-processing part.

## III. Testing FPGA

Following picture shows the Test board set for verification. This test board set operates itself without connection between any computer. Stereo vision system is implemented with HDL using Xilinx virtex-5 XC5VLX330 FPGA. The performance of system is 30

FPS(dependent on a CMOS camera) when size of images is 320x240(when census size is 15x15, support weight window size is 5x5 and maximum disparity range is 128), which is enough performance for autonomous robots and the other applications.

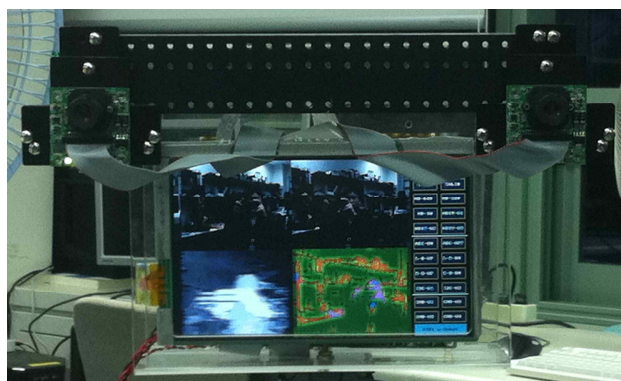


Figure 8: Example of the stereo vision system which is presented in this paper

Slice Logic Utilization	Used	Available	Usage
Number of Slice Registers	52,804	207,360	25%
Number of Slice LUTs	171,836	207,360	82%
Number used as logic	152,121	207,360	73%
Number used as Memory	12,507	54,720	22%
Number of route-thrus	15,953		
<b>Slice Logic Distribution</b>			
Number of occupied Slices	49,571	51,840	95%
Number of LUT Flip Flop pairs used	178,718		
Number with an unused Flip Flop	125,914	178,718	70%
Number with an unused LUT	6,882	178,718	3%
Number of fully used LUT-FF pairs	45,922	178,718	25%
Number of unique control sets	1,222		
Number of slice register sites lost to control set restrictions	1,510	207,360	1%
<b>Specific Feature Utilization</b>			
Number of Block RAM/FIFO	281	288	97%

Table 1. Device utilization summary on XC5VLS330

Table 1. shows the results reported by Xilinx ISE tool after synthesis and place-and-route.

#### IV. CONCLUSION

We developed the adaptive support-weighted census-based stereo matching IP using trellis dynamic programming structure and this includes the pre-processing part for the rectification and post-processing for reduction of noise. Also, in order to use this Stereo vision IP on various applications, we developed the test

system and this makes stereo matching result as images(left, right, disparity map and others), they can be used in embedded system itself.

#### Acknowledgement

This work was supported by the R&D program of the Korea Ministry of Knowledge and Economy (MKE) and the Korea Evaluation Institute of Industrial Technology (KEIT). [Project KI001813, Development of HRI Solutions and Core Chipsets for u-Robot].

#### REFERENCES

- [1] Scharstein, Daniel and Richard Szeliski. A taxonomy and evaluation of dense two-frame stereo correspondence algorithms. International Journal of Computer Vision. 2002
- [2] Jiho Chang, Sung-Min Choi, Eul-Gyoon Lim, Jaeil Cho, Dae-Hwan Hwang. Hardware implementation of a census-based stereo vision for real environment. The 7th International Conference on Ubiquitous Robots and Ambient Intelligence. 2010
- [3] Murray, Don and James J. Little, Using real-time stereo vision for mobile robot navigation. Autonomous Robots. 2000
- [4] Jeong, H. and S. Park. Generalized Trellis Stereo Matching with Systolic Array. In Parallel and Distributed Processing and Applications, Berlin-Heidelberg: Springer Verlag. 2004.
- [5] Park, Sungchan and Hong Jeong. Real-time stereo vision FPGA chip with low error rate. Proceedings of International Conference on Multimedia and Ubiquitous Engineering. 2007.
- [6] Wang, Liang, Miao Liao, Minglun Gong, Ruigang Yang, and David Nister. High-quality real-time stereo using adaptive cost aggregation and dynamic programming. Proceedings of Third International Symposium on 3D Data Processing, Visualization, and Transmission . 2006.
- [7] Nalpantidis Lazaros, Georgios Christou Sirakoulis, and Antonios Gasteratos. REVIEW OF STEREO VISION ALGORITHMS: FROM SOFTWARE TO HARDWARE. International Journal of Optomechatronics. 2008
- [8] Chris Murphy, Daniel Lindquist, Ann Marie Rynning, Thomas Cecil, Sarah Leavitt, Mark L. Chang. Low-cost stereo vision on an FPGA, International Symposium on Field-Programmable Custom Computing Machines. 2007
- [9] Gong, Minglun and Yee-Hong Yang. Near real-time reliable stereo matching using programmable graphics hardware. Proceedings of IEEE Computer Society Conference on Computer Vision and Pattern Recognition. 2005