Bulletin of Networking, Computing, Systems, and Software – www.bncss.org, ISSN 2186–5140
Volume 2, Number 1, pages 59–61, January 2013

Real-time Stereo for Embedded Vision System

Soon Kwon, Hyeon-woo Kim, Jaewook Ha, Jang woo Lee
Division of IT-Convergence
Daegu-Gyeongbuk Institute of Science and Technology
Daegu, Republic of Korea
soonyk@dgist.ac.kr

***Abstract*—This paper describes a real-time stereo vision system for mobile embedded vision applications. An NCC based cost function with a rectangular mask is used to evaluate corresponding pixels, which minimize matching cost energy. We propose a hardware-friendly architecture to reducing the computation complexity in a matching algorithm and implement it on a field programmable gate array (FPGA). The proposed data reuse technique makes it possible to be 20 times faster than a simple parallel processing method and it also occupies less hardware resources than previous works. The proposed stereo vision system is expected to be an efficient solution for a low cost and real-time embedded vision system.**

Keywords— stereo vision, embedded vision system, NCC, FPGA

I. INTRODUCTION

The recent stereo vision based 3D recognition technology is widely applied to embedded systems for next-generation advanced research fields, such as robot navigation, mobile based augmented reality, automotive pre-crash warning system, and intelligent surveillance[1]. Stereo vision has the advantage of more accurate object detection by utilizing depth information. However, computational overhead in a stereo matching process, which finds corresponding positions between left and right image pixels, has to be overcome. A lot of literatures have showed algorithmic or architectural improvements for speeding up the stereo matching process. Compared to algorithmic approaches[2, 3] that deal with computation efficiencies, architectural approaches[4, 5] are more adequate for a real-time computation goal by considering hardware-friendly algorithm architectures and achieving considerable computation performances.

We illustrate here a dedicated stereo vision processor for a real-time embedded vision system. Among the several stereo correspondence solvers, a normalized cross correlation(NCC) based local stereo matching method is chosen because of its robustness for illumination errors and sensor noises. In particular, this kind of characteristic is required for the mobile embedded vision system to prevent degradations in its performance due to the lack of system stability.

Our main contribution is hardware friendly stereo matcher design maximizing data reuse techniques. The paper is organized as follows. Section 2 introduces an NCC based stereo matching model. In section 3, the proposed data reuse techniques and hardware implementation is described in detail. Finally, Section 4 concludes the paper.

II. STEREO MATCHING MODEL

A stereo matching algorithm is similar to an energy minimization problem and can be defined as follows:

$$E(d) = \arg\min(D_p(f(p)) + V(f(p), f(q)) \qquad (1)$$

From equation (1), f(p) and f(q) are disparity levels in which we want to estimate them at each pixels. $D_p(f(p))$ is the cost of an assigning lebel f_p to pixel p and $V(f_p,f_p)$ is the cost of assigning labels f_p and f_q to two neighbor pixels, and is generally referred as the discontinuity cost. Local search based pixel-wise data-cost functions, such as Sum of Absolute Difference(SAD), Sum of Squared Difference(SSD), Normalized Cross Correlation(NCC), and Census transform are widely used due to its computational simplicity[6]. Among them, a correlation based cost function shows more robust matching results to some errors due to lightning difference or sensor noise that may occurs severe disparity estimation error. In this paper, an NCC based cost model is used as a stereo matcher robust to real-scene as a following equation:

$$C_{(x,y,d)} = \frac{\sum_{j=1}^{l}\sum_{i=1}^{k}[I_{L(x+i,y+j)} - \mu_L] \cdot [I_{R(x+d+i,y+j)} - \mu_R]}{\sigma_{L(x,y)} \cdot \sigma_{R(x+d,y)}}$$

$$\text{where, } \begin{cases} \mu_{(x,y)} = \dfrac{1}{k \cdot l}\sum_{j=1}^{l}\sum_{i=1}^{k} I_{(x+i,y+j)} \\ \sigma_{(x,y)} = \sqrt{\sum_{j=1}^{l}\sum_{i=1}^{k}[I_{(x+i,y+j)} - \mu_{(x,y)}]^2} \end{cases} \qquad (2)$$

Then, we applied the discontinuity cost as a simple truncated linear function that can improve the disparity estimation accuracy in boundary conditions.

III. Architecture Design and Implementation

Computational overhead of the conventional NCC algorithm described in equation (2) increases proportional to the window size. A data reuse technique, such as box-filtering [7], is well known method to reduce computation overhead of window sum calculation. In the case of hardware design, we can further maximize the rate of data reuse by parallelizing the box-filtering process to be processed multi-pixel simultaneously [8]. For example, in the case of a 2 x 2 pixel group as shown in Fig. 1, the blue area, which is shared by neighbor pixels, can be calculated only once during a box-filtering process. Considering a conventional stereo matching process, however, the efficiency of the box-filtering process is severely degraded when the disparity search range is too short. Instead, looking at the data-cost space image(DSI) which means the set of initial data-costs to be fed disparity estimation process as shown in Fig 2, the benefit using box-filtering can be maximized by processing the unit of each data-cost slice. Data-cost slice means a set of data-cost to the disparity value 'D' along the disparity axis of the DSI cube.

In addition, the proposed process can solve the problem of the box-filtering method that is hard to be easily applied to the numerator part of the NCC. The proposed architecture of the stereo matcher reflecting data reuse technique consists of several blocks, such as window-sum block, NCC calculation block, and disparity estimation block. In detail, the window-sum block consists of three different kinds of sub-blocks, such as sum(x), sum(x2), and sum(r(x)*l(x)). In the case of the NCC calculation block, we applied a look-up table based root square module for the fast execution. The proposed architecture was designed using a Verilog hardware language and was implemented to a field programmable gate array(FPGA). In the multiplier and divider modules, we used the Xilinx DSP function block [9]. The implemented stereo matching hardware accomplished a real-time requirement of 30 fps depth-map generation to search up to a 64 disparity range for VGA(640*480) sized images. By comparing the computation complexity using the data reuse technique, the multi-pixel based parallel box-filtering method shows that the total execution time is 20x faster than a simple parallel processing method. From Table 1, we found that the required memory cost can be reduced up to 27% by applying the proposed architecture with compromising small increases in logic resources. Note that the memory resource is the major bottleneck in a low cost embedded system.

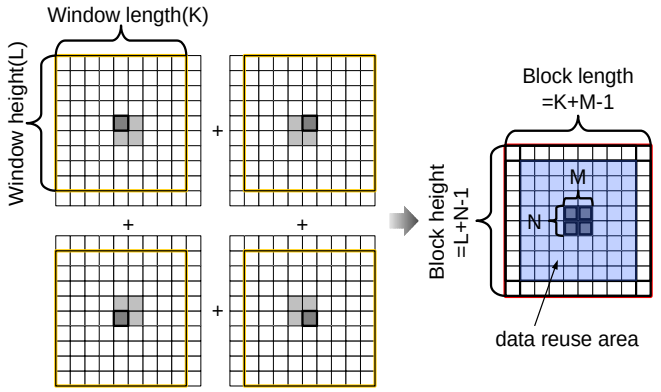

Figure 1. Multi-pixel based parallel box-filtering method

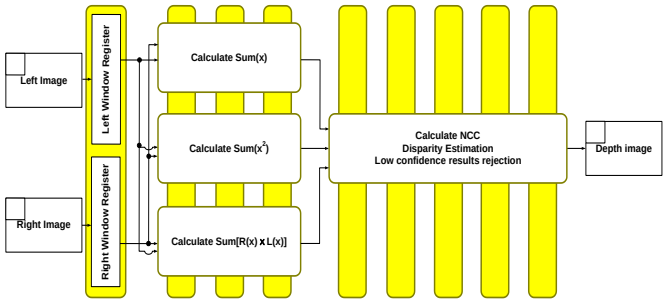

Figure 3. Proposed stereo matching hardware architecture

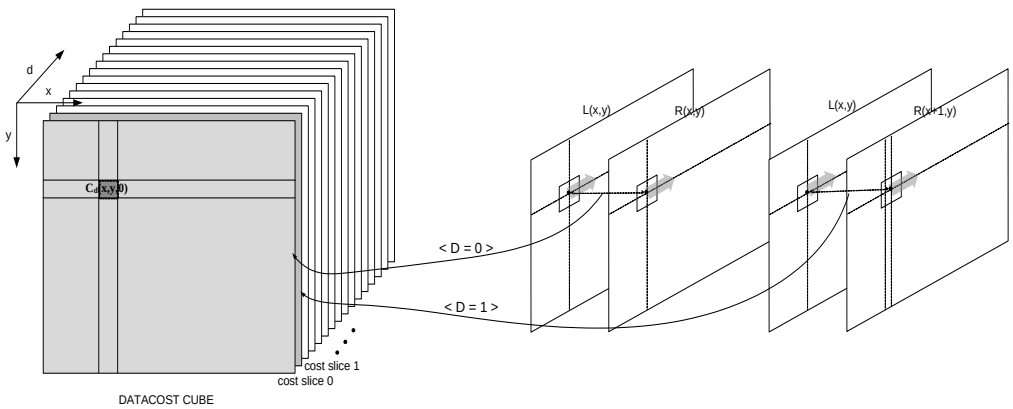

Figure 2. Data-cost slice generation in DSI cube

TABLE I. COMPARISON OF IMPLEMENTATION COSTS BETWEEN PROPOSED METHOD AND [8]'S METHOD (USED DEVICE : XILINX VERTEX-5 LX330)

Resource used	Proposed method	[8]'s method
LUT used	44,405/207,360 (21%)	37,735/207,360 (18%)
Block-ram used (Kbits)	4,932/10,368 (47%)	6,804/10,368 (65%)

IV. CONCLUSIONS

We have presented a NCC based efficient stereo matching architecture and its implementation. A correlation based cost function is considered for a robust stereo matcher. Our hardware-friendly matching architecture is focused on maximizing data reuse efficiency and reducing memory cost. We expect that the proposed stereo matching hardware becomes an efficient solution to a real-time embedded vision application.

ACKNOWLEDGMENT

This work was supported by the DGIST R&D Program of the Ministry of Education, Science and Technology of Korea (12-BD-0202).

REFERENCES

[1] K. Huh, J. Park, J. Hwang, D. Hong, "A stereo vision-based obstacle detection system in vehicles", Optics and Lasers in Engineering, vol. 46, no. 2, pp. 168—178, 2008

[2] D. M. Tsai, and C. T. Lin, "Fast normalized cross correlation for defect detection", Pattern Recognition Letters, vol. 24, no. 15, pp. 2625--2631 , 2003

[3] S.-D. Wei, and S.-H. Lai, "Fast template matching based on normalized cross correlation with adaptive multilevel winner update", IEEE Transactions on Image Processing, vol. 17, no. 11, pp. 2227--2235 , 2008

[4] S. Perri, D. Colonna, P. Zicari, and P. Corsonello, "SAD-Based stereo matching circuit for FPGAs", Proc. of the 13th IEEE International Conference on Electronics, Circuits, and Systems, pp. 846—849, 2006

[5] C. Geogoulas, L. Kotoulas, G. Sirakoulis, Ch. I. Andreadis, and A. Gasteratos, "Real-time disparity map computation module", Microprocessors and Microsystems, vol. 32, no. 3, pp. 159—170, 2008

[6] D. Scharstein, R. Szeliski, "A taxonomy and evaluation of dense two-frame stereo correspondence algorithms. International Journal of Computer Vision", vol. 47, pp. 7—42, 2002

[7] M. J. McDonnell, "Box-filtering techniques", Computer Graphics and Image Processing, vol. 17, pp. 65—70, 1981

[8] K. Bae, S. Kwon, Y.-H. Lee, J.-H. Lee, and B. Moon, "A hardware architecture based on the NCC algorithm for fast disparity estimation in 3D shape measurement systems", Journal of the Korean Sensors Society, vol. 19, no. 2, pp. 99 –111, 2010

[9] Xilinx. Inc. , Virtex-5 FPGA XtreamDSP Design Considerations, http://www.xilinx.com, 2010