Research of Fast 3D Imaging Based on Multiple Mode

Shibing Chen[1,2], Huimin Yan[1,2], Xuxiang Ni[1,2], Xiuda Zhang[1,2],Yu Wang[1,2]

[1] State Key Laboratory of Modern Optical Instrument, Zhejiang University

[2] National Engineering and Technology Research Center for Optical Instrument, Zhejiang University, Hangzhou, 310027, P.R.China

Email:yhm@zju.edu.cn

Abstract. Three-dimensional (3D) imaging has received increasingly extensive attention and has been widely used currently. Lots of efforts have been put on three-dimensional imaging method and system study, in order to meet fast and high accurate requirement. In this article, we realize a fast and high quality stereo matching algorithm on field programmable gate array (FPGA) using the combination of time-of-flight (TOF) camera and binocular camera. Images captured from the two cameras own a same spatial resolution, letting us use the depth maps taken by the TOF camera to figure initial disparity. Under the constraint of the depth map as the stereo pairs when comes to stereo matching, expected disparity of each pixel is limited within a narrow search range. In the meanwhile, using field programmable gate array (FPGA, altera cyclone IV series) concurrent computing we can configure multi core image matching system, thus doing stereo matching on embedded system. The simulation results demonstrate that it can speed up the process of stereo matching and increase matching reliability and stability, realize embedded calculation, expand application range.

1. Introduction

Stereo matching is one of the most active research areas in computer vision. Researchers domestic and overseas have presented a large number of excellent property binocular vision and stereo algorithms over the years[1].Among those ways, using transcendental parameter to constrain the search range when comes to stereo matching, has been proved to be an effective way. As without this limit, not only will it cost tremendous calculations (especially for big size images), but also many algorithms will not perform well[2]. An improved variogram analysis of the maximum expected disparity in stereo images, which is based on statistical correlations[3]. One research estimated the maximum and minimum disparity of stereo image depending on image characteristic[4]. And a hierarchical matching method was applied[5], however, it still need choose proper disparity range factitiously.

The use of a calibrated system, composed of a time-of-flight (TOF) camera and of a stereoscopic camera pair, allows data fusion thus overcoming the weaknesses of both individual sensors[6]. This method could greatly improve the resolution of depth map, of course this is a good try using multiple mode. But it's quite complicated to map the point captured by the TOF camera and stereo image, because he spatial resolution of currently available range sensors is lower than high-definition (HD) color cameras and the

two different camera is individual. In order to solve this problem, a dual-mode camera was used to gather binocular vision image and TOF depth maps time-sharing, thus the images own a same spatial resolution[7]. The experimental results demonstrate that it can speed up the process of stereo matching and reduce the error in matching effectively for stereo pairs with large disparity range.

In the meanwhile, a lot hardware system are presented to implement real-time processing. Reference [8] use verilog (a hardware design language) to design sum of absolute differences (SAD) computing units, and dynamic programming algorithm, gaining a fast speed. A preprocessing is applied on each rectified image before looking for the best matching between a left pixel, with a right pixel and different algorithms units are constructed on FPGA with a frequency of 100MHz[9]. Sophisticated local matching algorithms (CA and FLC) that are suitable for FPGA implementation are used to achieve low error rate while maintaining the high processing speed in reference[10] and evaluate the performance of our circuit on Xilinx Vertex-6 FPGAs.

Those method mentioned above have achieved considerably high processing speeds; however, their error rates are not as good as those of software programs. And they also cannot avoid another problems, about the influence of image storage, transport, display on the processing speed, and the restriction of image size. In addition, the configuration of such a hardware system is so complex that it often needs long time to test and the portability is low.

This paper synthesizes those methods advantages, applies the TOF camera and binocular vision with the same spatial resolution as in reference[7]. With the help of TOF depth maps, we could estimate each stereo pixel's disparity range. Simultaneously using field programmable gate array (FPGA, altera cyclone IV series) concurrent computing we can configure multi core image matching system, thus doing stereo matching on embedded system. Our experiment results indicate that it is benefit to use the Nios system to handle stereo imaging.

The remainder of this paper is structured as follows: Section II describes the proposed research algorithm. And the experiment set-up shows in section III. Experimental results on a real dataset and evaluation of the method, are presented in section IV. Finally, section V draws some conclusions.

2. Research method
We use dual mode camera developed in our lab, which capture the grayscale and depth maps with the same spatial resolution in division time. Then we use the data from depth maps to figure out disparity limit factor when comes to stereo matching.

2.1. TOF camera working principle
TOF camera means its working principle is the time of flight method. As Figure.1(a) shows, it is made up of fast gate functional CCD camera and short pulse laser diode. As shows in Figure.1 (b), laser diode emits rectangular wave and the CCD camera use gate technology[10]. A modulated near infrared light from the camera's internal lighting source is reflected by objects in the scene and travels back to the sensor, where its precise time of flight is measured independently at each of the sensor's pixel by calculating the phase delay between the emitted and the detected wave[6]. In the valid ranging scope, the rate of echo pulse falling into the gate time is linear related to the back time. Considering the effect of the background light and echo pulse power, and capturing two frames of background image and when all echo pulse falling into the gate time, normalization intensity of every pixel that have been eliminated background will be in proportion to target point.
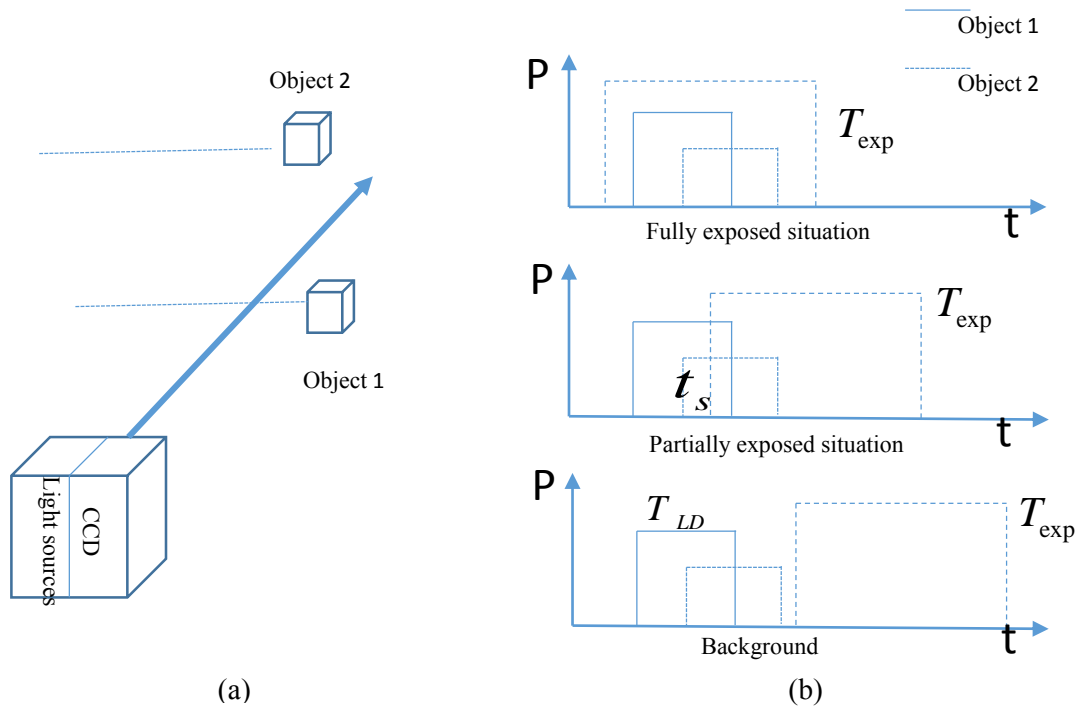
Figure. 1. The description of TOF camera. The image (a): the structure of the system. The image (b): CCD gate switch form

When all the echo pulse fall into the gate time, the grayscale off each CCD pixel is:

$P_1 = \eta(P_{LD}T_{LD} + P_BT_{exp})$ (1)

(η means the optical-to-electrical efficiency of CCD, P_{LD} and P_B means the echo pulse power and background power, T_{LD} is the pulse width of laser diode, T_{exp} is the gate time of CCD)

When part of the echo pulse fall into the gate time: $P_2 = \eta[P_{LD}(\frac{2z}{c} - t_s) + P_BT_{exp}]$ (2)

(where z means target distance, c means the speed of light, t_s the gate time of CCD)

The background grayscale is: $P_0 = \eta P_BT_{exp}$ (3)

From formula (1)~(3) we could know each pixel corresponding target range is: $z = \alpha r + \beta$ (4)

When we know α and β, we can figure out z according to normalization grayscale. When we calibrate system parameters, we can chose target longitudinal distance to obtain z and r relation matrix and transform the TOF depth maps to initial disparity image, for the sake of reducing hardware image processing procedure.

2.2. Matching Step
    The matching process mainly contains three parts, census transformation, cost aggregation and disparity calculation, as we can see in Figure.2. For the purpose of paying key attention to matching algorithm, we ignore the image capture course and camera calibration, and we implement image pre-

processing on PC for getting initial disparity maps. On the Nios II embedded system constructed by altera FPGA, we try to run the stereo matching algorithm and finish 3D reconstruction.
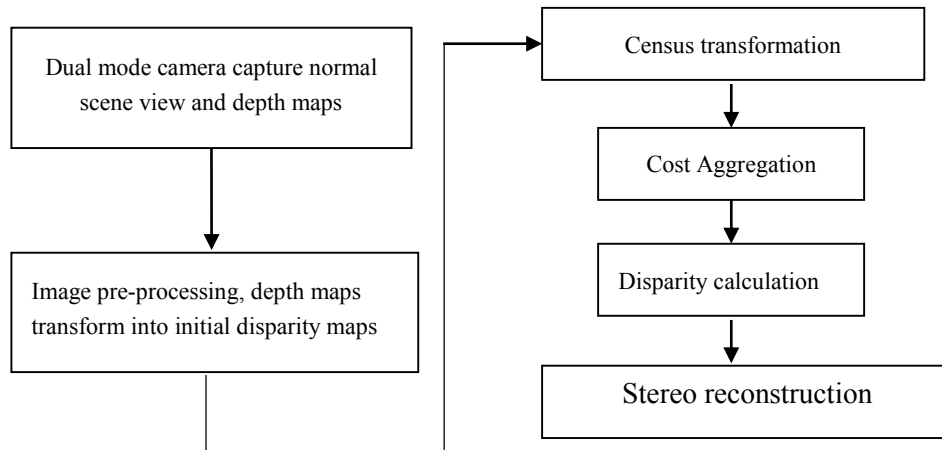
Figure.2 The description of the stereo vision algorithm procedure

2.3. Stereo algorithm

2.3.1. Single point match cost calculation

We chose census algorithm that has been confirmed owning obvious strength in local matching algorithm, as the similarity measurement function with regard to single point match cost calculation. This function make a bit string to express the structure characteristics of the matching points of local area (census transform neighboring area) by using census transform.

And the hamming distance of two match point's bit string represents the match cost. The census method single point match cost C_{Census} calculation procedure is below:

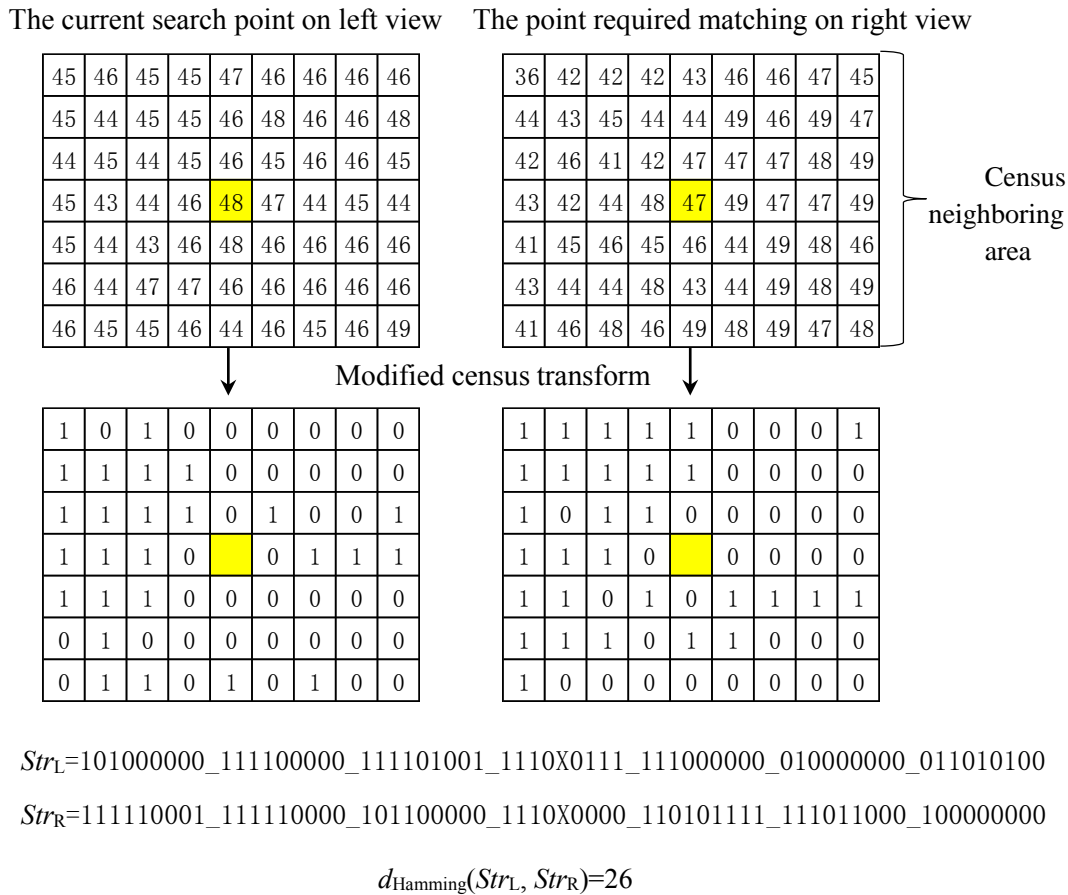$$C_{Census}(P,d) = d_{\text{Hamming}}(Str_R(P), Str_L(P+d)) \quad (5)$$

where d_{Hamming} represents the Hamming distance between two bit string (the number bits of different value at the same bit. $P=(x,y)$ represents the matching pixel coordinate in the reference view, d is the disparity value, Str represents the bit string gained from modified ways based on the foundation of traditional census transform:

$$Str(P) = \underset{[i,j]\in N}{\otimes} \xi(\bar{I}(P), I(P+[i,j])) \quad (6)$$

$\bar{I}(P)$ is the average grayscale in the census transform window N(P), and

$$\xi(I,I') = \begin{cases} 1 & I' < I \\ 0 & \text{otherwise} \end{cases}, \otimes \text{ means bit connection}$$

And Figure.3 indicates the single point match cost calculation course based on real experiment data, and the census window is rectangle window with a size of 9×7

The current search point on left view   The point required matching on right view

45	46	45	45	47	46	46	46	46
45	44	45	45	46	48	46	46	48
44	45	44	45	46	45	46	46	45
45	43	44	46	48	47	44	45	44
45	44	43	46	48	46	46	46	46
46	44	47	47	46	46	46	46	46
46	45	45	46	44	46	45	46	49

36	42	42	42	43	46	46	47	45
44	43	45	44	44	49	46	49	47
42	46	41	42	47	47	47	48	49
43	42	44	48	47	49	47	47	49
41	45	46	45	46	44	49	48	46
43	44	44	48	43	44	49	48	49
41	46	48	46	49	48	49	47	48

Census neighboring area

Modified census transform

1	0	1	0	0	0	0	0	0
1	1	1	1	0	0	0	0	0
1	1	1	1	0	1	0	0	1
1	1	1	0		0	1	1	1
1	1	1	0	0	0	0	0	0
0	1	0	0	0	0	0	0	0
0	1	1	0	1	0	1	0	0

1	1	1	1	1	0	0	0	1
1	1	1	1	1	0	0	0	0
1	0	1	1	0	0	0	0	0
1	1	1	0		0	0	0	0
1	1	0	1	0	1	1	1	1
1	1	1	0	1	1	0	0	0
1	0	0	0	0	0	0	0	0

Str_L=101000000_111100000_111101001_1110X0111_111000000_010000000_011010100

Str_R=111110001_111110000_101100000_1110X0000_110101111_111011000_100000000

$$d_{Hamming}(Str_L, Str_R)=26$$

Figure.3 The bit string of census transform calculation diagram

2.3.2. Cost Aggregation

For accurate local stereo matching, it is important to decide an appropriate local support region for each pixel adaptively. In principle, this local support region should contain only the neighboring pixels from the same depth with the pixel under consideration. A common assumption is that pixels with similar intensity within a constrained area are likely from the same image structure, therefore having similar disparity [11].

In this paper we chose Cross-based aggregation window. The determine ways of the cross-based window shows in Figure.4 First, take the pixel required matching as start point to find the up and down boundry based on disparity similarity code. Then use determined point in the 1D vertical window as start point to search right and left directions in order to find the boundry.

Take the up boundry as an example, search every one pixel every step until the point does not satisfy this three disparity similarity code

$$\begin{cases} 1. D_c(P',P) < \tau_1 \text{ and } D_c(P',P'-(0,1))) < \tau_1 \\ 2. D_s(P',P) < L_1 \\ 3. D_c(P',P) < \tau_2 \text{ if } L_2 < D_s(P',P) < L_1 \end{cases}$$

P′ represent current searching point, Dc(P′,P)=|I(P′)-I(P)| means the grayscale difference value between two point. $\tau1$、$\tau2$、L1 and L2 are constant. We make $\tau1$、$\tau2$、L1 and L2 are 20、6、10 and 5 on the basis of reference[11]
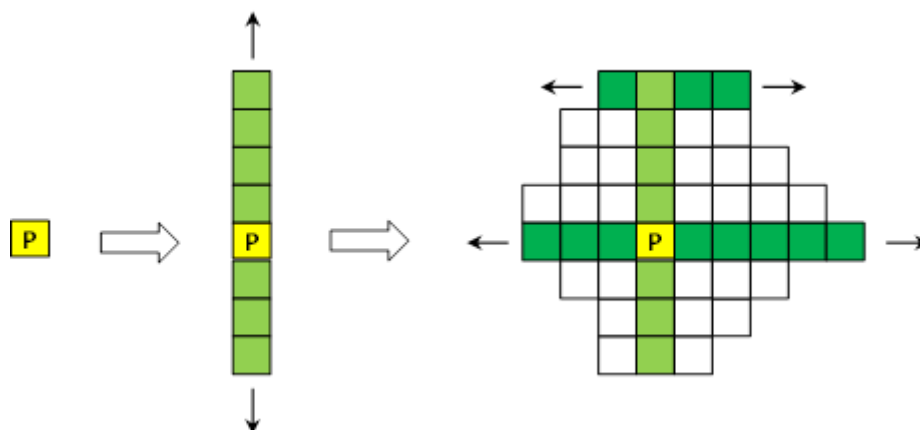

Figure.4 Cross-based aggregation determination method

2.3.3. Disparity Calculation
Now we use WTA (Winner-Takes-All) optimization thoughts for the most optimal operating on every point required matching and chose disparity of single matching point which has the smallest match cost. If we know the TOF depth maps disparity image, we can get the disparity constraint range.
D1= BF * Disp / (BF + DETA * Disp) - Offset + i;
D2 = BF * Disp / (BF - DETA * Disp) - Offset + i;
Where B represents the two camera's distance and F represents the camera focus, Disp represent the TOF disparity value and DETA represents the TOF camera precision. Offset means a rectified constant, and i represents any pixel.

3. Experiment Setup
Our hardware system is based on Altera Cyclone IV E and we can see some key parameters as shows in table.1 The actual working frequency is 100MHz, and the UART is used transport image data to the system. On account of enormous data, SDRAM is taken into consideration. Of course, Matlab is a useful tool to display the processed image.

Table.1 Basic FPGA system parameters used in this paper

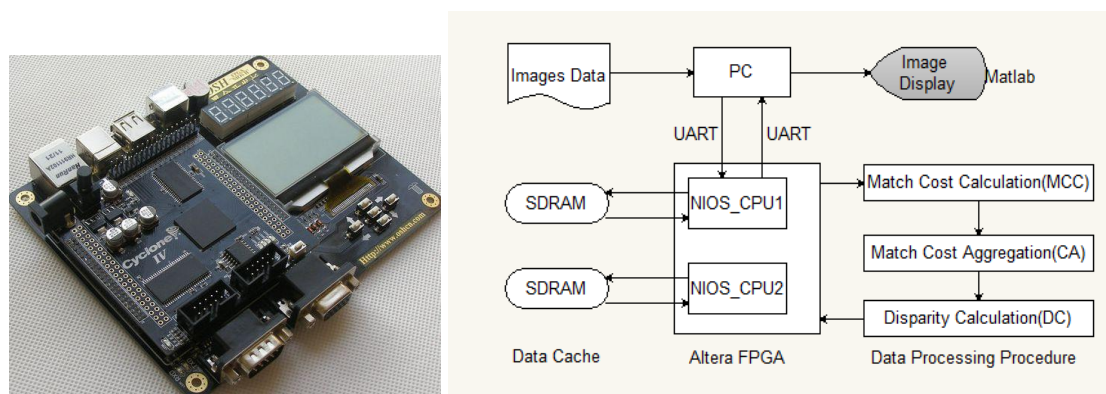| Altera	EP4CE15F17C8	Cyclone IV E
	RAM	504kbit
	LE	15408
CPU	Max frequency	200MHz
	Actual frequency	100MHz
	M9K	Six modes
SDRAM	256M	Hynix SDRAM
	(16M*16bit)	Memory
UART	115200bit/s	--
Sub-image Size	32*32（on FPGA)	640*480（on PC)

Figure.5 The new-designed FPGA-center platform. The left image: Physical prototype of our FPGA-center platform. On the left below a RS232 interface to communicate with PC. The right image: Block diagram of the system components

4. Results

In the experiment, images required matching were cut into small size sub-image, with a size of 40*30, 96*96, or 160*120, which the original image size is 498*672 as shows in Figure.6. And the stereo match results display in Figure.7. Since the sub-image is so small that we cannot judge the stereo effect macroscopically, we just need to know how long it take to finish the stereo match procedure. And we make a comparison between this system and on PC.
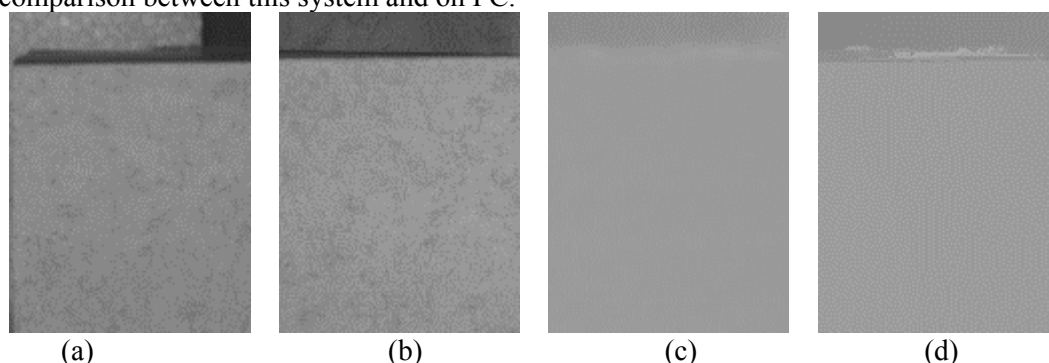

| (a) | (b) | (c) | (d) |

Figure.6 Results on Sub-image size of 160*120 (a) the left view, (b) the right view, (c) the coarse TOF depth map (d) the disparity image. The (a)···(b) represent the same meaning in Figure.7 and Figure.8
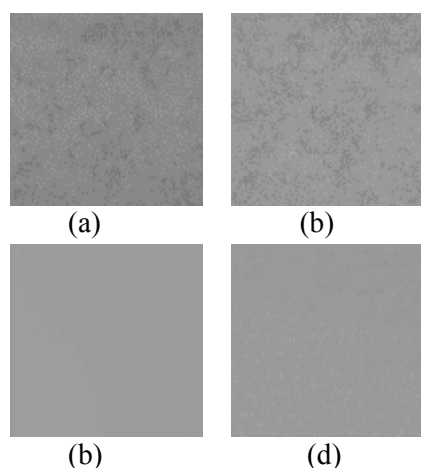

| (a) | (b) |

| (b) | (d) |

Figure.7 Results on Sub-image size of 96*96

| (a) | (b) | (c) | (d) |

Figure.8 Results on Sub-image size of 40*30

Match cost time		
Sub_image size	Altera FPGA NiosII 100Mhz	Personal Computer 2.0GHz
40*30	2.1s	1.1s
96*96	5.3s	2.6s
160*120	9.6s	5.2s

Table.2 Comparisons of the performances on PC and FPGA system

From table.2 we can see, it took a little more time to deal a sub-image on our hardware system. This is of course right and reasonable because of the shortage parallel Nios II cpu. Only one cpu has been tested to calculate the sub-image data. However, take the sub-image 40*30 for an example, it spent about 2.1s to calculate, we could judge that it will spend 35.28s at least to deal with big image owning the size of 672*498, in comparison to at least 84 second on PC, which indicates that using the FPGA and Nios II system can do speed up the matching time.

5. Conclusions

Applying the combination of binocular vision and TOF camera, using FPGA to construct embedded match method, this paper accomplish a effective transplant which bring the same algorithm on PC to Nios II. Although, the simulation results demonstrate that it cannot speed up the process of stereo matching only using one nios cpu to undertake the hug image with the enormous calculation. Under the single Nios CPU, fast importing the Sub-image data divided from original image required matching to this system, we gain a evident speed up for matching. From the comparison between altera FPGA and PC, the speed can be fasten evidently and the effect will be fine. And this paper develop a method to quickly realize stereo match on mobile device, which also avoid complicated hardware programmer to construct algorithm unit.

Of course what's most efficient method is to construct the calculation module using verilog language and implement real parallel computing.

Acknowledgment

References
[1] Scharstein, Daniel, and Richard Szeliski. "A taxonomy and evaluation of dense two-frame stereo correspondence algorithms." International journal of computer vision 47.1-3 (2002): 7-42.
[2] Kostkova, J., and R. Sara. "Automatic disparity search range estimation for stereo pairs of unknown scenes." Proc. CVWW (2004): 1-10.
[3] Cyganek, Bogusław, and Jan Borgosz. "An improved variogram analysis of the maximum expected disparity in stereo images." Image Analysis. Springer Berlin Heidelberg, 2003. 640-645.
[4] Min, Dongbo, et al. "Disparity search range estimation: enforcing temporal consistency." Acoustics Speech and Signal Processing (ICASSP), 2010 IEEE International Conference on. IEEE, 2010.
[5] Li, Rueihung, et al. "Disparity search range estimation based on dense stereo matching." Industrial Electronics and Applications (ICIEA), 2013 8th IEEE Conference on. IEEE, 2013.
[6] Gandhi, Vineet, Jan Čech, and Radu Horaud. "High-resolution depth maps based on TOF-stereo fusion." Robotics and Automation (ICRA), 2012 IEEE International Conference on. IEEE, 2012.
[7]Wei Shaopeng, Yan Huiming, Zhang Xiuda. "A Stereo Matching Method Based on the Combination of Depth Camera and Stereo Vision" [J].Opto-Electronic Engineering, 2015.
[8]John Kalomiros and John Lygouras. "Comparative study of local SAD and dynamic programming for stereo processing using dedicated hardware." (2009).
[9]Michel Devy and Jean-louis Boizard. "Stereovision algorithm to be executed at 100hz on a FPGA-based architecture. In Advances in Theory and Applications of Stereo Vision". Asim Bhatti Ed., Chapter 17,2011
[10] Jin, Minxi, and Tsutomu Maruyama. "Fast and accurate stereo vision system on FPGA." ACM Transactions on Reconfigurable Technology and Systems (TRETS) 7.1 (2014): 3.

[11] Xiuda, Zhang, Yan Huimin, and Jiang Yanbing. "Pulse-shape-free method for long-range three-dimensional active imaging with high linear accuracy." Optics letters 33.11 (2008): 1219-1221.

[12] Zhang, Ke, Jiangbo Lu, and Gauthier Lafruit. "Cross-based local stereo matching using orthogonal integral images." Circuits and Systems for Video Technology, IEEE Transactions on 19.7 (2009): 1073-1079.