Design and Implementation of Stereo Vision Systems Based on FPGA for 3D Information

Kuo-Yang Tu[1], Chen-Yu Chiu[1], Shih-An Li[2], and Jacky Baltes[3]

[1] Institute of System Information & Control,
National Kaohsiung First University of Science & Technology,
2 Jouyue Road, Nantsu, 811 Kaohsiung City, Taiwan, R.O.C.
[2] Department of Electrical Engineering, Tamkang University
[3] Department of Computer Science, University of Manitoba
`tuky@nkfust.edu.tw`

Abstract. The purpose of this paper is to utilize Field Programmable Gate Array (FPGA) to perform stereo vision distance detection. However, the stereo vision built by two cameras makes memory space lacking and image process slow under the constraints of FPGA application. In this paper, efficient memory space allocation and hardware calculation for stereo vision detection built in a System on a Programmable Chip (SOPC) based on FPGA are proposed. The hardware for stereo vision distance calculation includes the processing for the images of gray, binary, dilation, erosion, etc, and image geometry method for the vision distance through information of phase differences between two lenses. In addition, the simple hardware algorithm of background image subtraction to capture an object image from a series of image frames is also included. The totally hardware to perform stereo vision distance detection is difficult implementation, but firmware (some calculation in software) is flexible and quick to develop. Therefore, the performance of stereo vision distance detection according to hardware and firmware is compared. Finally, the distance calculation between objects and the lenses is demonstrated by practical experiments.

Keywords: Stereo Vision, FPGA, Epipolar Geometry, Image Processing.

1   Introduction

Computer vision is brisk up in present computer science. In human sensor system, people receive messages almost from vision, so accomplishing artificial intelligence of robotics is critical for computer processing of vision. The computer vision system is primarily divided to flat vision and stereo vision. The difference is whether the depth information, called depth perception as well, is acquired from an objective image. There have been many researches proposing algorithms of stereo vision which uses computer as a platform since the late of 1970s. Yau and Wang built an active stereo vision system to compute relative depth fast [2]. Candocia and Adjouadi developed the algorithm to measure a similarity matching of stereo feature [3].

The general stereo vision system requires two cameras to capture images from one object, and then calculates the depth of the object from features of the object. Moreover, it needs to be in the condition of parameter of cameras being known, it may calculate the stereo depth of the object through asymmetric geometric relationships.

In the field of visual autonomous robot, a robot requires an immediate image processing in order to enable the robot quickly response to the external objects. Thus, the robot needs to be connected to a high speed computing processor, such as laptop, to accelerate the image processing speed. In addition, the visual autonomous robot needs a visual processing system, plenty of sensor and motor systems are required as well. While these installments need numerous IO pins of processor to complete the transmission and reception of data. If a computer is spontaneously used to process visual image and control various sensors and motors, it will have a difficult situation of designing a program and a waste of computer resources due to only one processor implements many actions at the same time.

Hence, an additional controller such as digital signal processor (DSP) or single-chip processor of 8051 is generally used to aid computer processing signals of sensor and controlling the motor unit. However, the use of additional control processor will increase the difficulty of designing a robot, such as a protocol and packet format between computer and external controller. The recent use of FPGA clip is a feasible way to rapidly process the motion control of robot. The advantage of the design of FPGA clip is circuit design is able to be processed in parallel with high speed and be reprogramed repeatedly. Therefore, the core design of process and control of robot being achieved on FPGA clip is a rather practical way. Furthermore, Nios Ⅱ processor is 32-bit Embedded Soft-Core Processor being proposed for System On a Programming Chip (SOPC) by Altera Company; It is a Reduced Instruction Set Computer (RISC) processor being programed by VHDL or Verilog. Nios Ⅱ processor is an Embedded Soft-Core Processor that is able to be put into FPGA clip.

Due to FPGA clip provides numerous IO pins that are required as robot processes and control sensors and motor units and the IO pins on the FPGA clip are mostly able to be programmed as input pins or output pins by user, it is relatively flexible for the design and use of FPGA clip. Thus, it is suitable for the control and design of robot. Besides, the design trends of FPGA clip forward the development of large capacity and small size. Masrani and MacLean uses FPGA to measure a real-time large disparity range stereo-system [4]. Hamblen used FPGA-based SOC project for stereo vision system [5]. In recent years, Altera Company is continuing to develop and produce various FPGA clip of large-capacity or rapid processing of signals, such as Flex10k, MAX II, Cyclone II, Cyclone III, Stratix, Stratix II and so forth. Therefore, it will be suitable to use FPGA clip as a core control chip of robot in the future.

The stereo-vision systems need complicated computation. And FPGA is a good solution for complicated computation. Therefore, in this paper FPGA-based stereo-vision system is developed. However, the full hardware FPGA design is hard to exactly implement the stereo-vision system. In this paper, the firmware combining FPGA hardware and some software is proposed. Thus the performance of full

hardware and firmware FPGA are compared. The comparison tries to figure out if it's valuable to spend much time and more resource to develop the full FPGA stereo-vision system.

2   Stereo Vision

The most important part of stereo vision is to use a pair of image to calculate the depth of the object distance from a pair of images. Depth perception of computer vision systems is known as stereo vision system. In 1980s the United States, Massachusetts Institute of Technology Artificial Intelligence Laboratory, Marr proposed a visual computing theory and applied in the two eyes. Let two parallax images produce depth of three-dimensional graphics. It laid on the theoretical foundation of the development of binocular stereo vision.
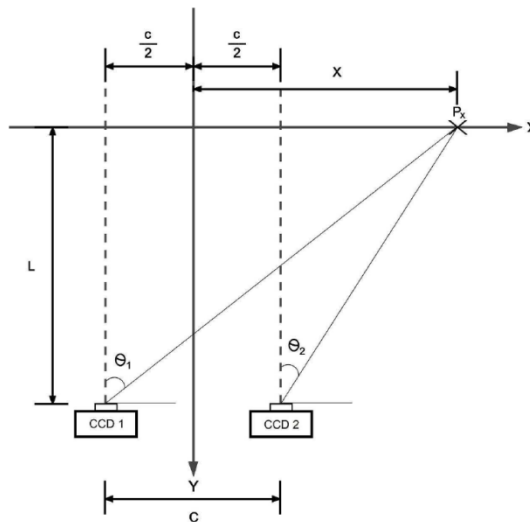
Fig. 1. Target and the geometric relationship of the Camera center

The stereo visual observation method is mainly divided into two types; one is crossing method, the other is parallel method as shown in Fig. 1 [1]. The crossing method is similar to the human eye to see objects. Using two eyes directly focus on the target to keep on the center of object vision, so it needs two cameras to follow object. This system is based on real-time observation as the goal not the tracking object. The crossing method will spend much time in object tracking. The other main reason is that the crossing method makes use of the CMOS Capture Module in constant (as shown in Fig.1). So, the parallel method is adopted. The benefit of parallel method is to save the tracking object control, a fixed geometric relationship and camera parameters to simplify the calculation process.

Let P_1 and P_2 are the distance between target pixels and center lines on left and right cameras, respectively. Then

$$\rho = \frac{P_1}{P_2} = \frac{x + C/2}{x - C/2} \tag{1}$$

where C is the distance between the centers of two cameras.

Define $\gamma = P_2/P_{max}$. The target distance L in the image can solved by

$$L = \frac{C}{\tan\theta_{max}\,\gamma(\rho - 1)} \tag{2}$$

3   Firmware and Hardware Design

In the research, a lot of problems arose such as whether it was hardware or software design. We introduced what kind of problems we have in the research and describe the solution. The biggest problem was memory allocation because we needed double memory space for using two cameras and background subtraction. DE2-70 only had 64-Mbyte SDRAM, so saving memory space was the most important thing.
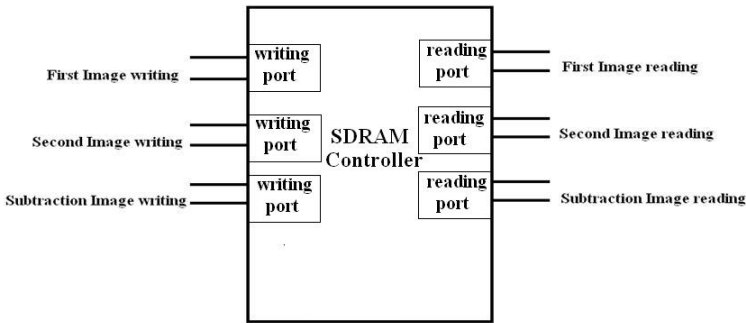

Fig. 2. 6 ports for background subtraction in SDRAM

In the system design before, we analyze the requirement of stereo vision via FPGA. The problems arose by the requirement can be separated into hardware and firmware. At least the hardware development must solve three problems. The first problem is how to take the RGB (Red, Green and Blue) bits from CMOS sensor into SDRAM. Usually the camera CMOS sensor makes use of 32 bits for the RGB data of each pixel. The 32 bits are too long for the 16-bit data bus of SDRAM. Normally, the SDRAM needs two ports to access 32-bit RGB (RGB32). It's inefficient for the hardware usage designed for the SDRAM. Instead of RGB32, the RGB16 defined as 5 bits for R, 6 bits for G and 5 bits for B express the image data for more efficient in hardware usage.

The second problem is how to execute in the background subtraction of two image frames to capture the motion target. The background subtraction requires four more ports in SDRAM. The block diagram is shown in Fig. 2. To solve this problem we use a smart skill that in image storage registers, the last address of the first image accesses the next address from the first address for second image.

The third problem is that switch is not enough for threshold and exposure value during color segmentation. The problem is solved by using one switch key to read the value.
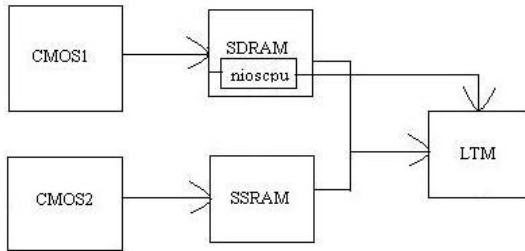

Fig. 3. The flow of firmware

In addition to hardware, the firmware still has problem for stereo vision. The firmware problem is about memory allocation. FPGA is parallel processing CPU, but SDRAM is first in first out. Therefore, master and slave design is very important, which one is first to access date from SDRAM. The SDRAM was not enough in firmware design for the research. Fig. 3 shows the situation of firmware. Memory control is too hard. In the implementation we use extra hardware to control memory.
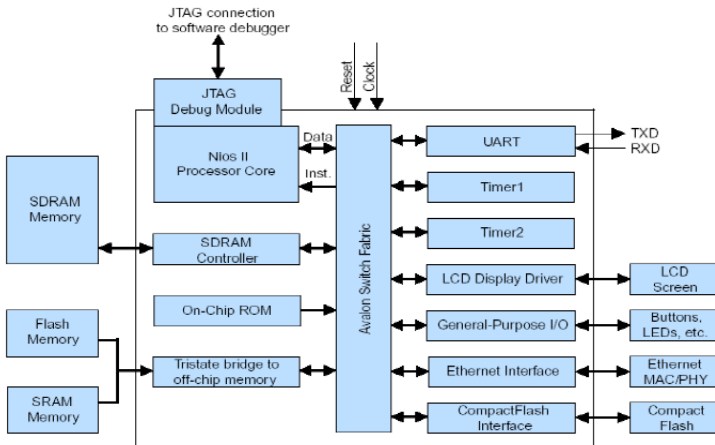

Fig. 4. NIOS II processor system

In this paper, NIOS II is used to implement stereo vision system. Fig. 4 shows the block diagram of NIOS II processor system. In addition, Fig. 5 shows all of the functions for the stereo vision detection.
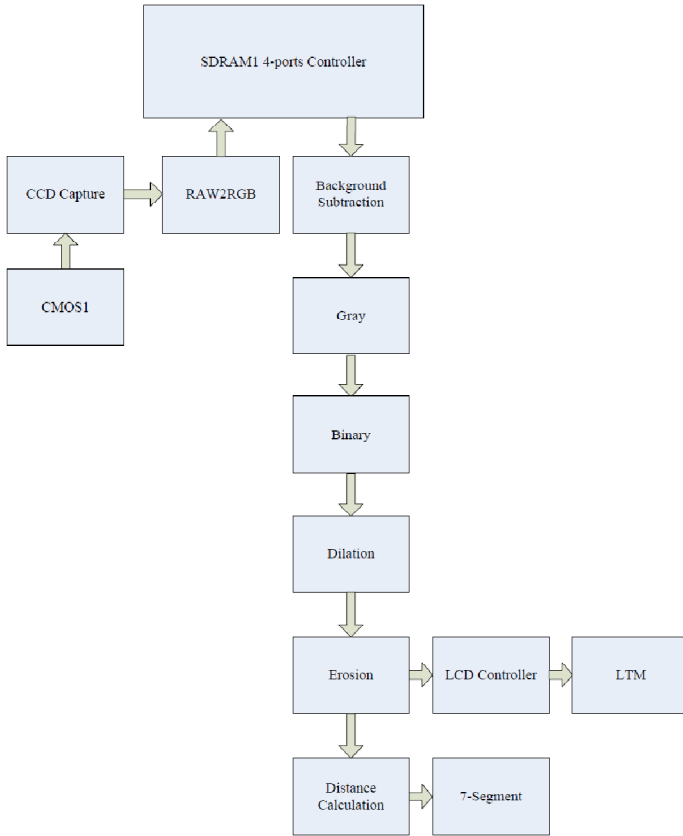
Fig. 5. Function of hardware design

4   Experiments

SOPC Builder integrated user interface into the Avalon Bus in the Nios II processor of hardware architecture through SOPC Builder designing own interface connected to the Avalon Bus, becoming one of the NIOS II processor hardware function to strengthen execution ability of the NIOS II processor.

An illustration which is the experimental results of the hardware and firmware is shown in Fig. 6. Red boxes show speed rate between the LTM and the CMOS before doing image processing. Using two different methods (hardware and firmware) to implement the same function, the velocity rate is faster than doubled. It means hardware can save more memory space and prove changing to another method is correct, so we have enough memory space to do the image processing. The table which is acquisition time of the hardware and firmware is shown in Table 1.

(a)   (b)
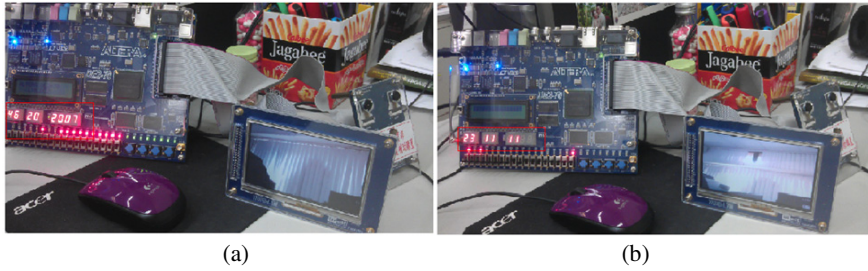
Fig. 6. The image processing speed (a) using hardware (b) using firmware

Table 1. The acquisition time of the hardware and firmware

	Hardware	Firmware
LTM	46 fps	23 fps
CMOS1	20 fps	11 fps
CMOS2	20 fps	11 fps

(a)   (b)

(c)   (d)

(e)   (f)
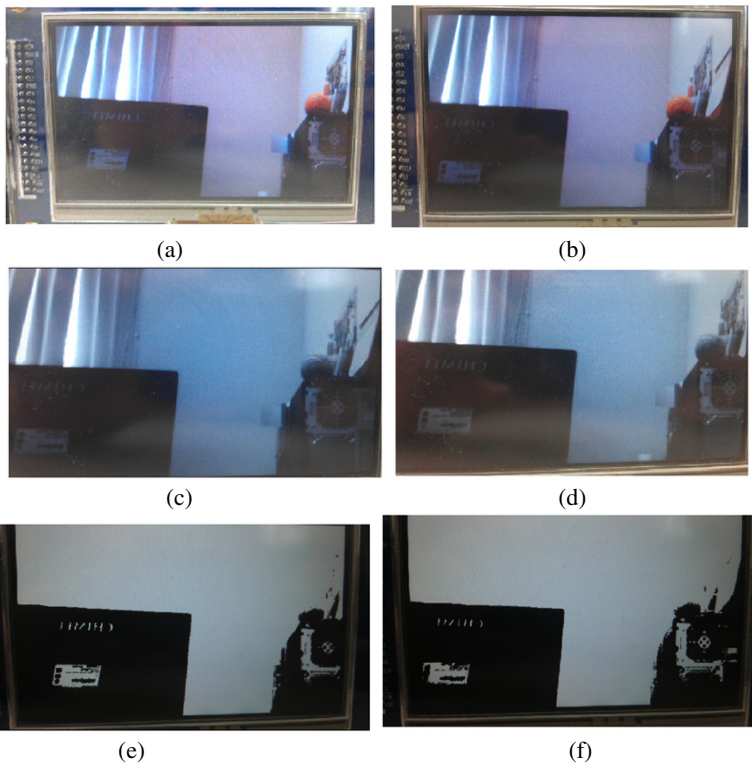
Fig. 7. The experimental results: (a) original left image (b) original right image (c) Gray left image (d) Gray right image (e) Binary left image (f) Binary right image

Fig. 7 shows six images which are image processing between right image and left image. An illustration (e) is binary right image and an illustration (f) is binary left image. These images are presented after dilation and erosion, it does not only use binary process, therefore, it makes the images look good and also reduce noise.

The background subtraction has two methods, static background subtraction and dynamic background subtraction. Fig. 8 is the static background subtraction. One method is the static background subtraction. It is based on a fixed background to grab the target, and the white zone is a target. The other one is the dynamic background subtraction. It subtracts last image, so the white part identifies the difference on last image. If a picture is whole black, it means target stationary in one place stays more than two seconds. Fig. 9 is process of the dynamic background subtraction.

The experimental result of distance computation is shown in Fig. 10. The blue circle is the distance from orange ball to camera module. The value 1F is hexadecimal (hex) because we only have two empty 7-segment displays in the DE2-70 board. It can save one more 7-segment displays. The camera module was placed 15 cm from the ground. Camera height of the placement of the relationship, so the ball only 33 of the 80 cm position, otherwise the target will be beyond the screen area.
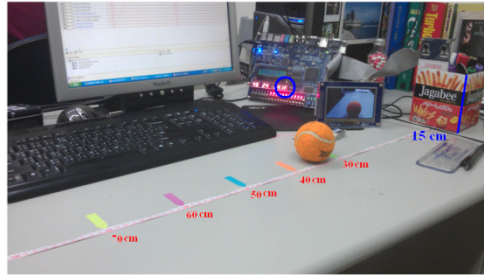

(a)   (b)
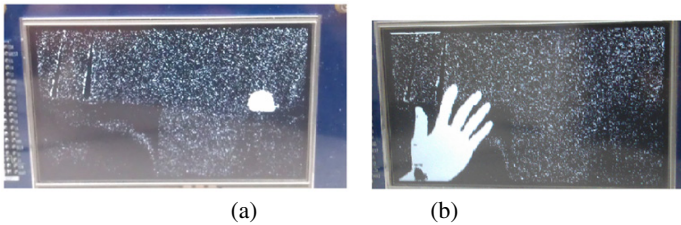
Fig. 8. The static background subtraction

(a)   (b)

(c)   (d)
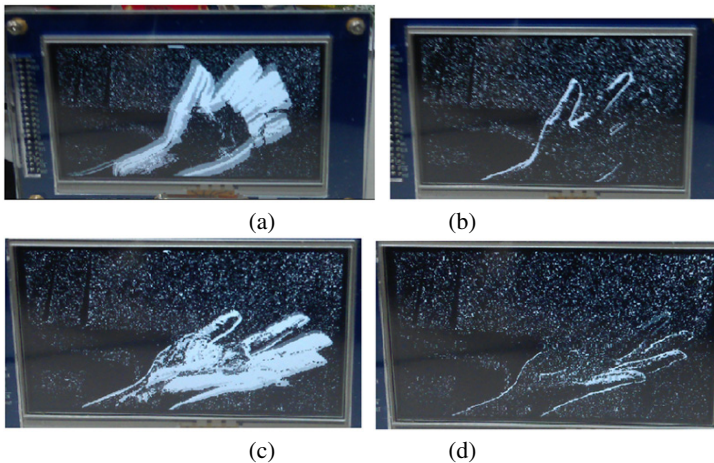
Fig. 9. The dynamic background subtraction

Fig. 10. The experimental result of distance computing

Table 2. The experimental result of percentage error

Actual Distance	Calculative Distance	Percentage Error
37 cm	31 cm	16.2%
70 cm	63 cm	10%

5   Conclusions

As a result of this paper, it can be stated, that stereo vision is a good modality for three-dimensional information. Computer stereo vision is a broad application of science. In the previous application, using the PC with the programming language achieves this project, but the end result is a time limit. FPGA implementation of a stereo vision to derive 3D information system used in this study, it improves the PC carried out the image processing, but it cannot achieve real-time features. FPGA not only accomplish real-time computing, but also save hardware space to do more function in this research.

At this point the work can be developed into two methods. One is firmware design, which can be easy in image processing, but there is a requirement of more memory space. In this research, the memory space is insufficient for the DE2-70 board because two cameras and one LCD require large memory space. So, memory space would be not enough for this research. The other is hardware design, which not only saves memory space, but also can do real-time computing.

In this study, in addition to the distance computation, the image processing is also very substantial. The image processing includes gray, binary, dilation, erosion and background subtraction. The background subtraction is in order to target more clearly and make compute easier after a string of image processing program. Distance is calculated by epipolar geometry algorithm when the screen only has a target. Whether it is image processing or distance computation is very important program in this study.

In the future of visual system application, I hope it can be applied more real-time image processing system on the FPGA. Using FPGA to achieve image processing system not only miniaturizes the system, but also reduces electricity consumption.

References

1. Lee, J.S., Seo, C.W., Kim, E.S.: Implementation of opto-digital stereo object tracking system. Optics Communications 200, 73–85 (2001)
2. Yau, W.-Y., Wang, H.: Fast relative depth computation for an active stereo vision system. Real-Time Imaging 5(3), 189–202 (1999)
3. Candocia, F., Adjouadi, M.: A similarity measure for stereo feature matching. IEEE Transaction on Image Processing 6, 1460–1464 (1997)
4. Masrani, D.K., MacLean, W.: A Real-Time Large Disparity Range Stereo-System using FPGAs. In: Fourth IEEE International Conference on Computer Vision Systems (ICVS 2006), p. 13 (2006)
5. Hamblen, J.O.: Using an FPGA-base SOC approach for senior design projects. In: IEEE International Conference on Microelectronic System Education (2003)