FPGA Based Disparity Value Estimation

Celal GÜVENDİK
Vestel Elektronik San. ve Tic. A.Ş.
R&D Department
Manisa, Turkey
email:ceguvendik@hotmail.com

Özgür TAMER
Electrical and Electronics Eng. Dept.
Dokuz Eylul University
Izmir, Turkey
email:ozgur.tamer@deu.edu.tr

Abstract- **Depth map of an image provides information about the distance of the objects in the image to the stereoscopic camera. This information is widely employed in three dimensional visualization of stereoscopic images. Depth map of an image can be evaluated by calculating the disparity values of objects in the image. Disparity is the planar difference between corresponding points in an image pair, captured by a stereoscopic camera. In this work, we propose a fast and accurate method for evaluating the disparity values of images captured by a digital stereoscopic camera system. The stereovision system, evaluating the disparity values, runs on a field programmable gate array chip and uses an off the shelf stereo camera module to capture stereo images. The disparity values, evaluated by the system, show good match with the theoretical values.**

Keywords: Computer stereovision; stereoscopic camera; disparity; depth map; image segmentation; FPGA.

I. INTRODUCTION

The ability of the human eye to see in three dimensions and judge the distance of an object is called depth perception [1]. Creating a virtual depth perception by using stereoscopic images has been studied even in the first years of photography. Digital images captured by a stereo camera system can be used to estimate the corresponding depths of objects in the images. This process is called depth map generation and it has become a popular research area in recent years with the development of computers capable of processing digital images in real time [2]. The application areas of evaluating depth information from a stereoscopic image cover a wide range, from movie industry to smart surveillance systems, even smart robots and the video games. Depth maps help us to identify the distance of any point in the image to the camera.

Several methods have been proposed to acquire the depth information of an image including stereo with two or more cameras, triangulating light stripers, millimeter wavelength radar, and scanning and flash Light Detection And Ranging (LIDAR). Radar and Lidar use the time of flight information of an electromagnetic pulse or a laser pulse to estimate the distance of an object to the observation point. Generally, they are used in combination with a camera to evaluate a depth map of an image. Triangulating light stripers project laser or visible light stripes on objects and the shape and the distance of the object is estimated according to the deformation of the stripe [3].

Most widely used method for depth map generation is to employ stereo camera pairs. A stereo camera pair consists of two identical cameras in a line. Construction of depth map using a stereo camera is based on finding pixels corresponding to the identical objects in the images captured by both cameras and evaluating the distance based on the difference of the location of the objects. This difference is called the disparity and it is given in more detail in Section 2.

An automatic depth map estimation technique from a single input image has been proposed by Battiato [2]. Their method is based on first classifying digital images as indoor, outdoor with geometric elements or outdoor without geometric elements and estimating the depth map. They first detect vanishing lines in an image and characterize the depth gradients, which are assigned according to these lines. Holzzman and Hochgatterer presented a solution for mobile devices with a single camera [4]. They use multiple images and the inertial sensors of the phone to generate the depth map. Aldavert and colleagues presented a stereo camera based method to align obstacle maps and matches between features in the stereo image pairs [5]. They use Gauss-Newton algorithm to evaluate the relation and the geometry of the corresponding map. Fua proposed a parallel algorithm for stereoscopic depth map generation [6]. His technique used correlation to calculate the sparse disparity values of pixels. Then these values are interpolated and a dense depth map is evaluated. The performance of the algorithm is significant. A super resolution depth map generation technique assisted by a stereo vision system has been proposed by Yang et. al. [7]. They use a low-resolution depth map as input, we recover a high-resolution depth map and color stereo image pair to evaluate a high-resolution depth map for the corresponding image. A Kinect® sensor and a stereo camera based depth map estimation system is proposed by the same research group [8]. In this method, they employed a kinect sensor instead of a low resolution depth map to acquire a high resolution depth map. The system proposed by Meier et al. produces disparity maps and optical flow field at 127 fps and 376x240 pixels resolution based on block matching [9]. Their system focuses on evaluating the velocity of Micro Aerial Vehicles (MAVs) for a robust operation of navigation control loops which makes estimation speed a critical parameter. Another method using FPGA is proposed by Wang et. al. [10]. Their system gives

 ISBN: 978-1-61208-638-5

the user to adjust the system according to the image resolution, disparity range and parallelism degree for optimization of the performance. Results are superior according to the comparison tables presented in the article.

In this paper, an FPGA based stereovision technique is employed to evaluate the disparity value of an object in a scene. While evaluating the disparity value, the geometrical and computational parameters of a stereovision system are also examined. The most appropriate disparity values are chosen from possible disparities. By employing the camera system parameters with these disparit values, all the objects at a certain distance in a scene can be segmented. The proposed stereo vision system works on the Digilent's Genesys® board which includes a Virtex® 5 FPGA by Xilinx corporation. The Digilent® Vmod® Cam is used as the stereo camera while the display is a generic computer screen with an HDMI interface.

This paper is organized as follows. A brief introduction about disparity techniques is presented in Section II and FPGA implementation of the proposed method is introduced in Section III. Experimental results of the implemented hardware on the built testbed are presented in Section VI and the results are concluded in Section V.

II. MATERIALS AND METHODS

Disparity can be determined as the dissimilarity of the right and left images that are acquired by a stereo camera system [11]. Each camera of the stereo system captures the same view from a different angle of view as presented in Figure 1. Parameters x and x' in Figure 1 are the distance between points in image plane corresponding to the same scene point three dimensional and their camera center. B is the baseline of the stereo camera and f is the focal length of camera [11]. Disparity is defined as;

$$d = x - x' = \frac{Bf}{Z} \quad (1)$$

Here, d is the disparity, B is the baseline, f is the focal length and the Z is the distance from camera to object captured. Thus, the disparity of a point in a scene is inversely proportional to the difference in distance of corresponding image points and their camera centers. By using this information, the depth of all pixels in a stereoscopic image can be found.

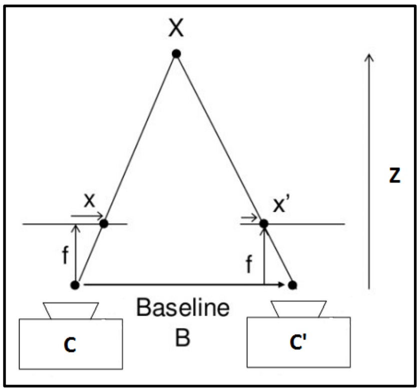


Figure 1. The relation between disparity (x-x'), baseline (B), focal length (f) and the distance (Z) between camera and the object "X"

The scene can be segmented according to the depth information by using stereo cameras. Objects with similar distances with the camera will be segmented separately from the rest of the image.

A. FPGA Implementation

The hardware is implemented using Verilog HDL on Xilinx® ISE 14.5 and implemented on Xilinx® Virtex-5® XC5VLX50T FPGA chip. The FPGA development board is GenesysTM and the stereo camera board is VmodCAMTM. Both of them have the Digilent brand [12] and [13].

In Figure 2, the general structure of the implemented hardware is given as a block diagram. The first block is the camera driver block. The camera board has two independent cameras that are integrated. The frames from left and the right cameras are buffered in the block RAMs.

The second block is the Disparity Map calculator block. In this block, the frames are stored as the left and the right images in the block RAMs. The first block is used to generate the disparity map with the disparity value that is given externally. The generated disparity map is buffered in the block RAM. The third block is the clock generator, which generates necessary clocks for the other blocks by using the 100 MHz system clock [14].

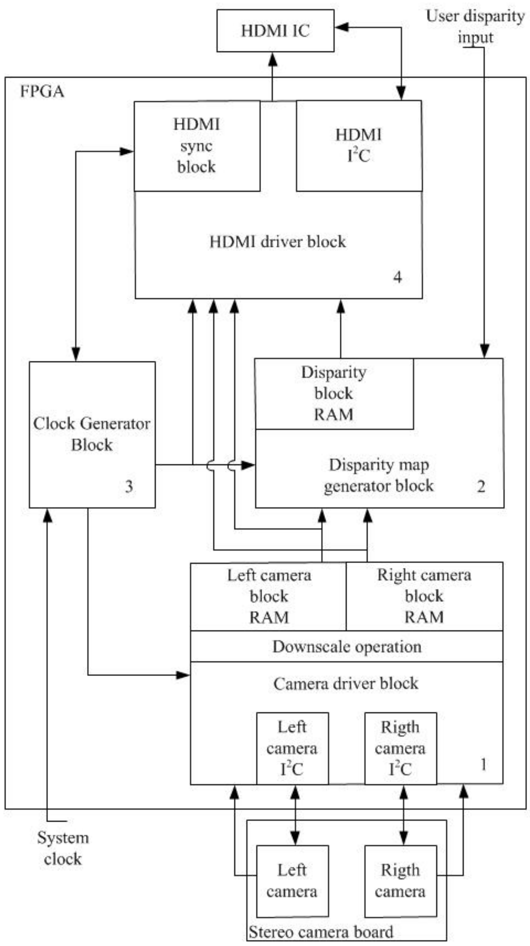


Figure 2. General Structure of hardware

The last block works as the HDMI driver. This block initializes the HDMI transmitter IC and sends the contents of left camera data, right camera data and the disparity map to the HDMI IC. This block also has a pixel clock and digital video synchronous signal generator to drive the HDMI chip. With the help of this block instant test results are viewed.

B. Disparity Map Generator Block

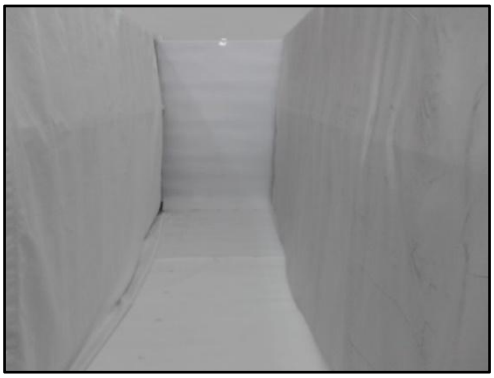
The aim in this block is to provide that the objects in different distance from cameras can be segmented by using stereovision with the disparity information.

The disparity map is generated by keeping one of the frames constant, and shifting the other as much as the given value horizontally. After the shifting operation, the images are subtracted from each other and these subtraction values are stored. This difference frame is the disparity map of the scene for the given disparity value. To calculate the disparities, right image frame is shifted vertically as the value of the disparity register in hardware.

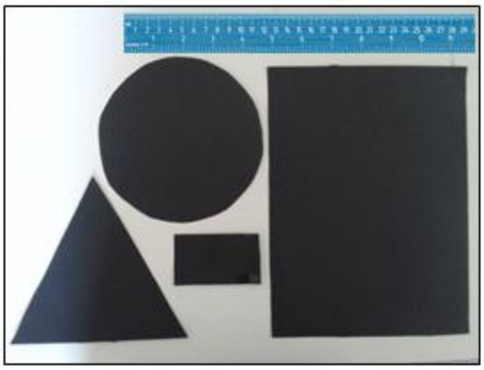
C. Disparity Value EvaluatiOn

To evaluate the disparity value of an object located at a certain distance a test bed is built and some standard test objects are manufactured. Figure 3(a) and 3(b) shows the test bed and the test objects, respectively (in B, the ruler is 30 cm). Here, background of the test bed is made up of white fabric to make the objects visible.

In Figure 4(a),(b), the test setup is shown;



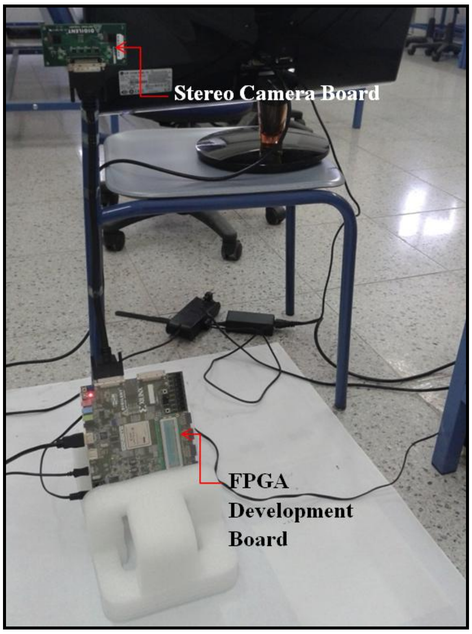
(a)



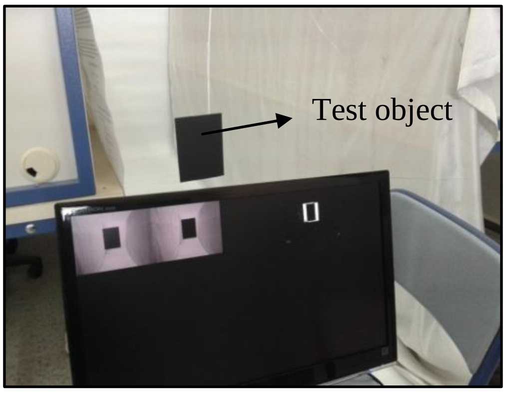
(b)

Figure 3. The test bed(a) and the test objects (b)

During the tests, these objects are positioned at several distances (only result for one meter is presented in this work) from the cameras. Another variable to be evaluated during the experiment is the disparity value, which is entered manually during calibration process. Several experiments were performed with different disparity values at a fixed distance to validate the disparity value. The results are displayed on the LCD screen located at the side of the testbed.



(a)



(b)

Figure 4. Stereo camera and FPGA development board in (a), and disparity demonstration in (b)

The rectangular object is positioned at 1 meter from the stereo camera. Firstly, the registers for the disparity is zero so; the result stream is not at the correct disparity level. This case can be observed in Figure 5(a) by observing the white pixels around the rectangular object on the screen. Each white pixel means that the corresponding pixels do not match and difference is high.

The next operation is to find the correct disparity level of the object at one meter from the cameras. By using equation (1), it can be calculated that the disparity value of (0010110)2=(22)10 is the appropriate value for one meter distance. In Figure 6, it can be seen that the slide switch values are set to the appropriate disparity value evaluated theoretically.

As seen in Figure 6, if the disparity value is set to the correct value for the object distance, the left and the right images of the object almost overlap totally in the disparity map and the white surrounding around the rectangular image almost disappears. This means that, the adjusted disparity value is true? In order to validate the result, the disparity values (0010111)2=(21)10 and (0010101)2=(23)10 are also tested as shown in Figure 7 and Figure 8, respectively. It can be observed that when these disparity levels are entered to the switch, white surrounding around the rectangle object starts to appear.

Figure 5. The disparity map for a zero disparity level at 1 meter distance. from the cameras

As a result for the 1 meter test, we are able to determine the correct disparity value by finding the one with minimum number of white pixels around any object.

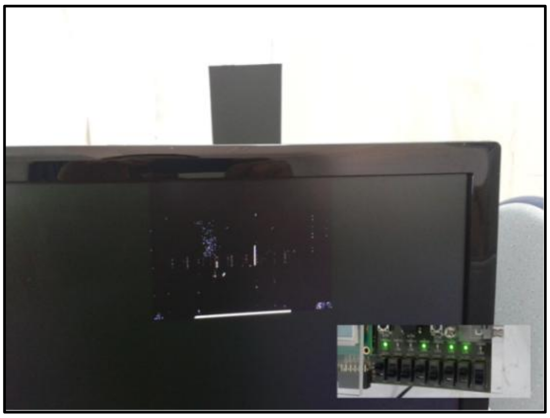


Figure 6. The appropriate disparity value of (0010110)2=(22)10 for one meter distance from the cameras

Figure 7. Disparity map for (0010101)2=(21)10 disparity level at one meter from the cameras

Figure 8. Disparity map for (0010111)2=(23)10 disparity level at one meter from the cameras

III. RESULTS

This setup is tested from 25 cm to 3 m by 25 cm steps and the appropriate disparity levels are observed and compared with the calculated disparity levels. For each distance disparity values from 0 to 127 are entered as the disparity value are entered to the system and the one with minimum number of white pixels is determined as the correct disparity value for the corresponding distance.

The theoretical value of the disparity values are also evaluated by using the disparity formula given in equation (1). Here, the cameras has 3.81 mm lens (f in equation (1)) and distance between cameras is 6.3 cm (similar to human eyes). The formula gives the disparity result in cm when the other parameters are in cm. These disparity values are changed to pixel value by dividing them to pixel width of the image sensor. The image sensors are 2.2 µm x 2.2 µm and the pixels were downscaled because of the limited amount of memory inside the FPGA chip. Thus, the pixel width is assumed here, 4 times of real value (2.2×4=8.8 µm) because of the 1/16 representation of pixels.

As it can be seen from Figure 9 and Table 1, the observed and the calculated disparity values have no difference bigger than 8 pixels. This situation can be named as the system tolerance (max. 4 pixels from right and 4 pixels from left image).

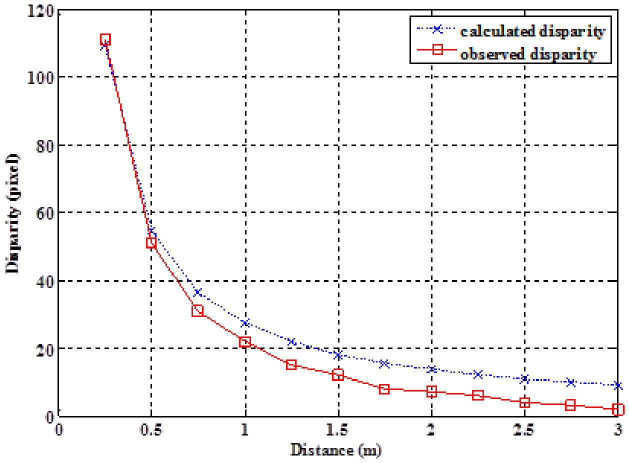


Figure 9. Comparison of the observed and calculated disparity values

TABLE 1. VALUES OF THE GRAPH IN FIGURE 12

Distance from cameras (m)	Calculated Disparity (pixels)	Observed Disparity (pixels)
0.25	109	111
0.50	54	51
0.75	36	31
1.00	27	22
1.25	22	15
1.50	18	12
1.75	15	8
2.00	13	7
2.25	12	6
2.50	11	4
2.75	10	3
3.00	9	2

The differences are almost stable after 0.5 m from the cameras, in the near zone (in 0.25 m and 0.5 m) they are smaller than the others.

IV. CONCLUSION

In this paper, the relationship between disparity and the distance of an object in a scene is examined experimentally and the images are segmented according to the depth information. The stereovision technique is used to obtain the disparity values.

The system is implemented on an FPGA and the relation between the disparity and the distance of an object from the cameras is measured in a test bed and the test object is segmented according to the distance.

The results in the near zone seems better than the results in the far zone depending on the inverse ratio between the disparity and the distance. In the near zones the disparity values are bigger which brings out that the distances are represented with more pixels than they are represented in far zones from the cameras. Thus, the disparity values shows better match with theoretical values in the near zones than they are in the far zones. On the other hand, it must be noted that the observed values depend on the observer and they can vary only a few pixels.

One remarkable property of this implementation is that, no external components rather than the FPGA chip are used. This makes it a tiny application and applicable to any Virtex 5 board independent of the board design while it can still be applied to other FPGA chips with minor changes in the code.

A. Future Work

Most remarkable future work is to adapt the system for an automatic depth map generation of a scene. This could be accomplished by first segmenting the objects visually and acquiring the disparity values automatically from the proposed system. This would be a promising solution for real time depth map generation on a streaming video file.

The design can also be developed to process more complex task and big images by using onboard DRAMs or FPGA chips with larger resources with an appropriate hardware design.

REFERENCES

[1] I. P. Howard and B. J. Rogers, *Seeing in Depth: Depth perception*. I. Porteous, 2002.

[2] S. Battiato, S. Curti, M. Cascia, E. Scordato, “Depth-Map Generation by Image Classification,” in *Proceedings of SPIE*, 2004, pp. 95--104.

[3] A. Klette, Reinhard Schluns, Karsten Koschan, *Computer Vision*. Springer, 1998.

[4] C. Holzmann and M. Hochgatterer, “Measuring Distance with Mobile Phones Using Single-Camera Stereo Vision.,” in *ICDCS Workshops*, 2012, pp. 88–93.

[5] D. Aldavert, A. Ramisa, R. Toledo, and L. Ramon, “Obstacle Detection and Alignment Using an Stereo Camera Pair,” Technical report, Artificial Intelligence research institute, Spain, 2008.

[6] P. Fua, “A parallel stereo algorithm that produces dense depth maps and preserves image features,” *Mach. Vis. Appl.*, vol. 6, no. 1, pp. 35–49, 1993.

[7] Y. Yang, M. Gao, J. Zhang, Z. Zha, and Z. Wang, “Depth map super-resolution using stereo-vision-assisted model,” *Neurocomputing*, vol. 149, pp. 1396–1406, 2015.

[8] S. Zhang, C. Wang, and S. C. Chan, “A New High Resolution Depth Map Estimation System Using Stereo Vision and Kinect Depth Sensing,” *J. Signal Process. Syst.*, vol. 79, no. 1, pp. 19–31, 2015.

[9] D. Honegger, P. Greisen, L. Meier, P. Tanskanen, and M. Pollefeys, “Real-time velocity estimation based on optical flow and disparity matching,” in *Intelligent Robots and Systems (IROS), 2012 IEEE/RSJ International Conference on*, 2012, pp. 5177–5182.

[10] W. Wang, J. Yan, N. Xu, Y. Wang, and F.-H. Hsu, “Real-time high-quality stereo vision system in fpga,” *IEEE Trans. Circuits Syst. Video Technol.*, vol. 25, no. 10, pp. 1696–1708, 2015.

[11] S. Mattoccia, “Stereo vision: algorithms and applications,” 2011.

[12] “Digilent, VmodCAM Reference Manual.” 2011.

[13] “Digilent, Genesy Board Reference Manual.” 2013.

[14] “Xilinx Virtex-5 Libraries Guide for HDL Design.” 2009.

 ISBN: 978-1-61208-638-5