International Journal of Applied Engineering Research ISSN 0973-4562 Volume 12, Number 7 (2017) pp. 1432-1436

Real-time Multiview Autostereoscopic Image Synthesis System

Yongjun Jon[1], Kyunghan Chun[2], Bonghwan Kim[2], Dongin Lee[3], Insoo Lee[4] and Dongkyun Kim[1*]

[1]School of Computer Science and Engineering, Kyungpook National University, Daegu 41566, Korea.
[2]Department of Electronic and Electrical Engineering, Catholic University of Daegu, Gyeongbuk 38430, Korea.
[3]Department of Mobile Information and Communication Engineering, Yeungnam University, Gyeongbuk 38541, Korea.
[4]School of Electronics Engineering, Kyungpook National University, Daegu 41566, Korea

**Corresponding author*

Abstract

This paper proposes the development of a real-time multiview autostereoscopic image synthesis system development. Recently, many autostereoscopic displays have been developed, ranging from experimental displays in university departments to commercial products. This paper uses 2 or 4 cameras as input and applies a multi-level algorithm for parallel processing in a disparity estimation and shows a multi-view glasses-free 3D image generation. A 16 view autostereoscopic system are tested by using GPU CUDA parallel processing and some functions are implemented via DSP/FPGA for real-time application.

Keywords: Autostereoscopy, Disparity Estimation, Real-time Implementation.

INTRODUCTION

Many researchers have developed autostereoscopic 3D displays by using a range of different technologies[1]. The method of creating autostereoscopic flat panel video displays using lenses was mainly developed at the HHI (Heinrich Hertz Institute)[2]. Prototypes of single-viewer displays were already being presented by Sega AM3 and the HHI[3]. Nowadays, this technology has been developed further and one of the best-known 3D displays developed was the Free2C, a display with very high resolution and very good comfort achieved by an eye tracking system and a seamless mechanical adjustment of the lenses. Eye tracking has been used in a variety of systems in order to limit the number of displayed views to just two, or to enlarge the stereoscopic sweet spot. However, as this limits the display to a single viewer, it is not favored for consumer products.

Currently, most flat-panel displays employ lenticular lenses or parallax barriers that redirect imagery to several viewing regions; however, this manipulation requires reduced image resolutions. When the viewer's head is in a certain position, a different image is seen with each eye, giving a convincing illusion of 3D. Such displays can have multiple viewing zones, thereby allowing multiple users to view the image at the same time, though they may also exhibit dead zones where only a non-stereoscopic or pseudoscopic image can be seen, if at all[4].

The proposed system synchronizes 2~4 cameras and uses a high speed image capture module. To improve finding an exact matching point, an image rectification method is applied to align epipolar line exactly. After stereo camera calibration, to estimate a disparity in multi-view by parallel processing, a multi-level algorithm is utilized to captured images (left, right camera) by GPU (Graphic Processing Unit) CUDA (Computer Unified Device Architecture) parallel processing. And to make a multi-view performance better, an approximation algorithm is used to fill a hole between 3D images. For the disparity estimation system realization, DSP/FPGA implementation are used which includes development of multi-channel stereoscopic image frame grabbing board, and high speed disparity estimation module over 20 frames per seconds.

MULTI-VIEW 3D IMAGE CONTENT AND SYNTHESIS

Stereoscopic 3D cinema and television are in the process of changing the landscape of entertainment. However, the necessity to wear glasses is often regarded as a main obstacle of today's mainstream stereoscopic 3D display systems. As a solution, multi-view auto-stereoscopic displays overcome this obstacle. They enable glasses-free stereo viewing by emitting several images at the same time. The multi-view autostereoscopic display technology ensures that each viewer in front of the display sees only that stereo pair which is appropriate for his particular viewing position[5].

To achieve these advanced functionalities, multi-view autostereoscopic displays require not two but many different views as input. Typical multi-view auto-stereoscopic displays which are on the market today require 8-view[6], 9-view[7], or even 28-view[8] as input[9-16]. To use least cameras as input, this paper uses 2- or 4-cameras as in Figure 1. And to generate in real-time 16 view autostereoscopic images from 2 or 4 camera inputs, DSP (Digital Signal Processing) / FPGA (Field-Programmable Gate Array) is applied to implement the 3D image synthesis.

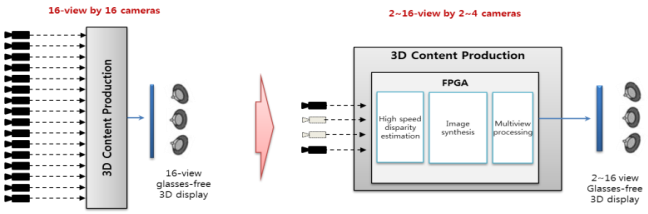
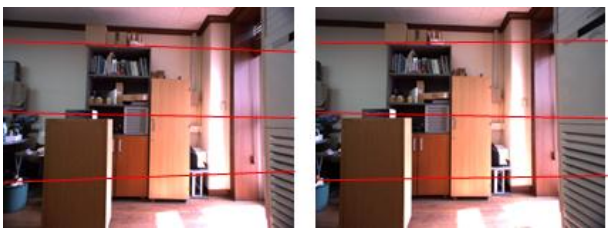


Figure 1. 3D content production.

In a multi-view autostereoscopic image synthesis, it is important to estimate the disparity from two images (left image and right image) for 3D content production and to get a good performance, matching point process should be done accurately for finding an exact matching point from two images. So first of all, by calibrating cameras, the image rectification method aligns the epipolar line to enhance the exactness of a match point in a camera image and to minimize the searching range in finding a match point as in Figure 2.

Figure 2. Stereo camera calibration.

By using calibration patterns for applying mathematical solutions of two cameras positions and optical equation, intrinsic parameters of the camera such as a focal length, a principal point can be computed and extrinsic parameters such as rotation, translation and distortion coefficient vectors also can be computed. As a result, the image rectification by these parameters can make an exact match of two camera images as shown in Figure 3.



(a) before

(b) after

Figure 3. Image rectification (a) before and (b) after.

The multi-level depth map methods are composed of 5 levels which are from 1st level to 5th level and utilized for parallel processing in CUDA implementation.

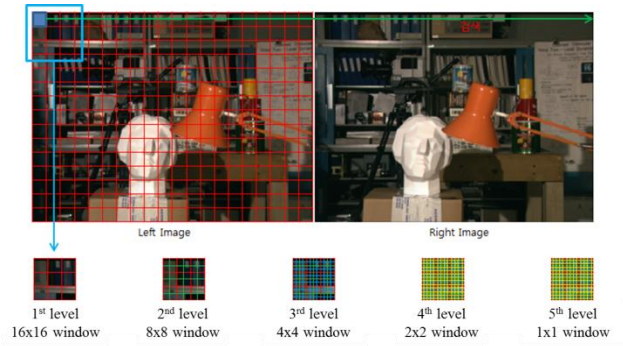


Figure 4. Windows for multi-level method.

Each level of the multi-level depth map method is possible to accomplish independently, so the process is applied for utilization of parallel processing of CUDA in GPU. Result of each level is used for the next level process. The multi-processing flow is shown in Figure 5 and the multi-level method makes it possible to process fast

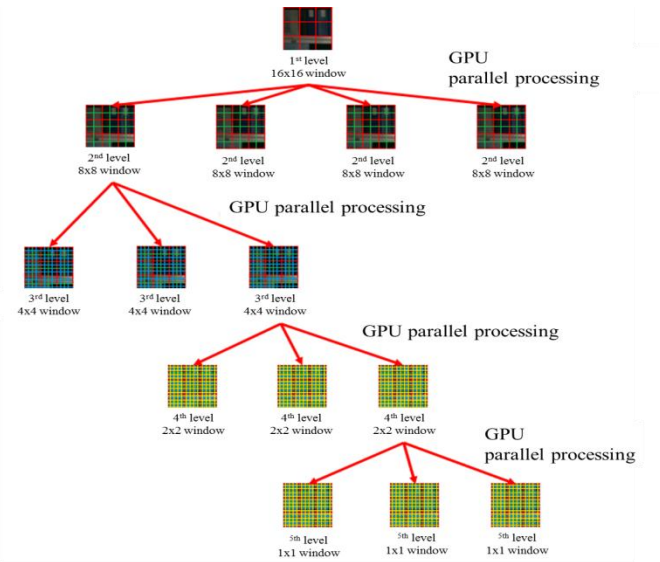


Figure 5. Parallel processing flow by using GPU CUDA.

DISPARITY ESTIMATION

This section describes the algorithm of the disparity estimation which is developed for the enhanced performance. The improved algorithm includes cost measurement method, modified census transform method, cost aggregation method and occulusion handling method. The cost measurement is composed of three combined measurement functions. The first comes from the color difference of the pixel R, G, B and the second comes from census transform and the last comes from the gradient distance.

The census transform uses binary value (0/1) for measuring the difference between the central pixel and the surrounding pixels in the window. And this is a rough approximation in the image information, which means that there is a possibility to lose some image information. So the proper way to save more

image information is required. The modified census transform uses 2 bits to save the difference of image information. The values of the previous frame are saved for the reuse of next frame parallel processing and also the modified census transform have a parameter 'alpha' in the inequality which can consider slight change of neighbor pixels. The modified method sets the boundary range by changing the value of alpha parameter.

The cost aggregation is applied for reducing errors in the boundary process. In the disparity estimation, the method usually relies on color value of the pixel to find the boundary and this may result an error. To improve this error, Gestalt theory is introduced and Gestalt in German means form or shape. The image is divided into figure and background and the close part of the image is applied to reduce an error.

Occlusion handling is applied by using the previous frame. Occluded area comes out by the distance change and this should be handled to estimate the exact disparity. For occlusion handling, if there is no occluded area in the previous frame, the corresponding occluded area in the frame is reserved by keeping the values of the previous frame.

16 VIEW 3D IMAGE SYNTHESIS

A 16 view 3D image synthesis is developed and for the exact image synthesis, the approximation method is applied to fill a hole. From 2 cameras as input, the 16 view image synthesis is accomplished as follows in Figure 6.

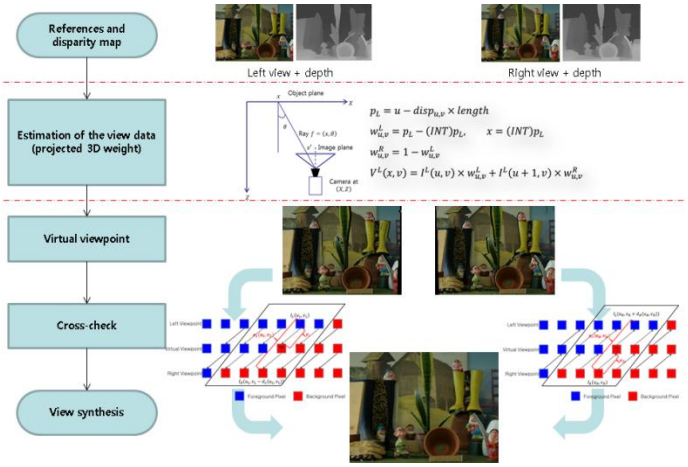


Figure 6. 16 view 3D image synthesis.

The 3D image synthesis means new disparity generation based on the estimated results. For a chosen view, the image can be generated from the computed camera position and the computation is as follows in Figure 7.

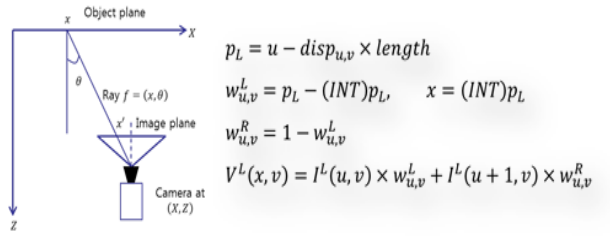


Figure 7. 16 view position computation.

In 3D image synthesis, the occluded area comes out by the distance change and in the left and right images, there are areas which include the occluded area in one side, and not in other side. At that time, cross-check is applied to decide how to handle those areas. By comparing data of the occluded area from the estimated image and true image, the similar value is chosen and the image synthesis is fulfilled.

PERFORMANCE EVALUATION

PSNR (Peak Signal to Noise Ratio) is applied to evaluate the performance of the developed 16 view 3D image synthesis system. PSNR is a value to show the difference between two images in the quantitative manner and is the ratio of peak signal power to noise power.

$$PSNR = 10\log_{10}\left(\frac{MAX_I^2}{MSE}\right) = 20\log_{10}\left(\frac{MAX_I}{\sqrt{MSE}}\right) = 20\log_{10}(MAX_I) - 10\log_{10}(MSE) \quad (1)$$

where

$$MSE = \frac{1}{mn}\sum_{i=0}^{m-1}\sum_{j=0}^{n-1}[I(i,j) - K(i,j)]^2 \quad (2)$$

m : horizontal resolution of the image

n : vertical resolution of the image

$I(i,j)$: (i,j) pixel value of the reference image

$K(i,j)$: (i,j) pixel value of the compared image

MSE: Mean Square Error

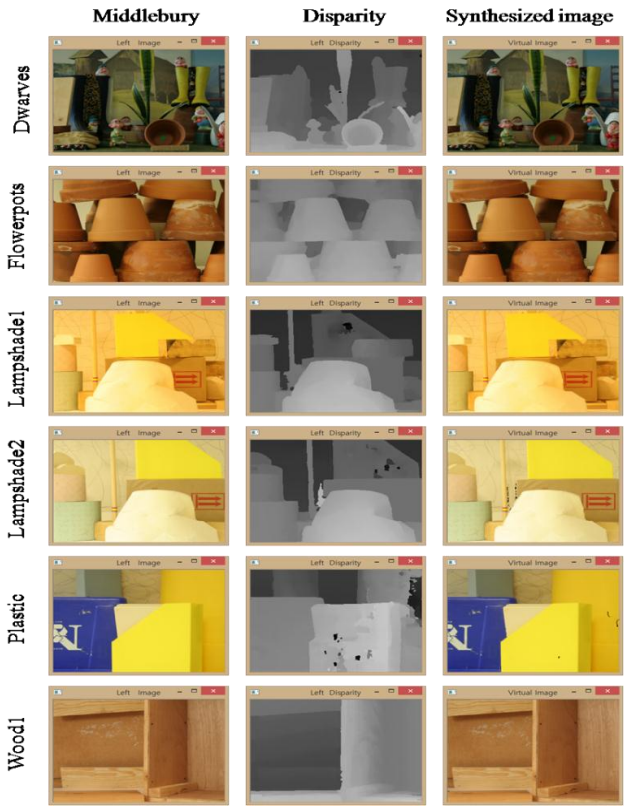


Figure 8. Images for PSNR Test.

The test image from the Middlebury is used to measure the performance because the image is recognized as a sample of objectivity in the disparity estimation[9]. Test procedure is as follows:

Step 1. Disparity estimation from the left and right images of the Middlebury.

Step 2. Multi-view 3D image synthesis based on the estimated disparity.

Step 3. PSNR computation between synthesized image and Middlebury image.

Step 4. 6 Middlebury image sets are tested for the accuracy.

The Middlebury and synthesized image are shown in Figure 8 and the test results are in Table 1 and show that the average PSNR is 31.711 dB.

Table 1. Transient Responses.

Middlebury	Test No.	PSNR	
		[dB]	Average [dB]
Dwarves	1	30.093188	30.09319
	2	30.093188	
	3	30.093188	
Flowerpots	1	29.855463	29.85547
	2	29.855463	
	3	29.855485	
Lampshade1	1	34.985725	34.98584
	2	34.985886	
	3	34.985898	
Lampshade2	1	34.211554	34.22152
	2	34.221474	
	3	34.221527	
Plastic	1	30.914659	30.92068
	2	30.914659	
	3	30.932734	
Wood1	1	30.193223	30.19196
	2	30.190667	
	3	30.191976	
Average			31.71144

CONCLUSION

In this paper, the real-time multi-view 3D image synthesis system is developed. The proposed system just uses 2 or 4 cameras for generating 16 view 3D images. DSP/FPGA implementation is utilized for the real-time processing and GPU CUDA parallel processing is used by applying the multi-level depth map method. To enhance the disparity estimation performance, cost measure, modified census transform, cost aggregation and occlusion handling method are utilized. The performance evaluation is done by PSNR test and used the Middlebury image for PSNR test. The average result of PSNR test is 31.711 dB and to fully implement the whole system by DSP/FPGA, the research is still on-going.

ACKNOWLEDGMENTS

This research was supported by the Basic Science Research Program through the National Research Foundation of Korea (NRF) funded by the Ministry of Education (No. NRF-2015R1D1A1A01059060). This research was also supported by the Basic Science Research Program through the National Research Foundation of Korea (NRF) funded by the Ministry of Education (No. NRF- 2016R1D1A3A03919627).

REFERENCES

[1] N.S. Holliman, Three-Dimensional Display Systems, Handbook of Optoelectronics, ISBN 0-7503-0646-7, 2 (2006) 1067-1100.

[2] R. Boerner, 3D-Bildprojektion in Linsenraster schirmen (in German), (1985).

[3] Electronic Gaming Monthly, 93 (1997) 22.

[4] https://en.wikipedia.org/wiki/Autostereoscopy

[5] N. Stefanoski, O. Wang, M. Lang, P. Greisen, S. Heinzle, A. Smolic, Automatic View Synthesis by Image-Domain-Warping, IEEE Transactions on Image Processing, 22(9) (2013) 3329-3341.

[6] Aliscopy 3D [Online], 2013. Available: http://www.alioscopy.com

[7] Toshiba 55ZL2 [Online], 2013. Available: http://www.toshiba.com

[8] Dimenco Displays [Online], 2013. Available: http://www.dimencodisplays.com

[9] H. Hirschmüller, D. Scharstein, Evaluation of cost functions for stereo matching, in Proceedings of IEEE Computer Society Conference on Computer Vision and Pattern Recognition (CVPR 2007), Minneapolis, MN, (2007).

[10] M. Werner, B. Stabernack, C. Riechert, Hardware Implementation of a Full HD Real-time Disparity Estimation Algorithm, IEEE Transactions on Consumer Electronics, 60(1) (2014) 61-73.

[11] K. Muller, P. Merkle, T. Wiegand, 3-D video representation using depth maps, in Proceedings of IEEE, 99(4) (2011) 643-656.

[12] A. Smolic, 3D video and free viewoint video - From capture to display, Pattern Recognition, 44(9) (2011) 1958-1968.

[13] P. Greisen, S. Heinzle, A. Burg, M. Gross, An FPGA-based processing pipeline for high definition stereo video, in Proceedings of EURASIP J. Image Video Process, 1 (2011) 1-13.

[14] S. Jin, J. Cho, X.D. Pham, K.M. Lee, S.K. Park, M. Kim, J. Jeon. FPGA design and implementation of a real-time stereo vision system, IEEE Transactions on Circuits and Systems for Video Technology, 20(1) (2010) 15-26.

[15] C. Riechert, F. Zilly, M. Muller, P. Kauff, Real-time disparity estimation using line-wise hybrid recursive matching and cross-bilateral median up-sampling, in Proceedings of the IAPR 21st International Conference on Pattern Recognition, (2012) 3168-3171.

[16] C. Riechert, P. Kauff, O. Schreer, Fully automatic stereo-to-multiview conversion in autostereoscoic displays, The Best of IET and IBC, 4 (2012) 8-14.