Interactive 3D Contents Generation for Auto-stereoscopic Display based on Depth Camera

Yuki Takaya, François de Sorbier , Yuko Uematsu, Hideo Saito
Graduate School of Science and Technology Keio University
3-14-1 Hiyoshi, Kouhoku-ku, Yokohama, Kanagawa, Japan
{takaya, fdesorbi, yu-ko, saito}@hvrl.ics.keio.ac.jp

Abstract— **In this paper, we propose a system for auto-stereoscopic display based on a color camera and TOF depth camera. Auto-stereoscopic display can render different view images from a single color image with its corresponding depth map. The TOF depth camera does not provide color information and is low resolution. It means that captured data cannot be used directly with the auto-stereoscopic display. A high resolution color camera is added beside the depth camera. The color camera and the depth camera are not located at the same position which means that viewpoints are slightly different. We apply view translation techniques to depth information and generate the corresponding depth map.**

Keywords-component; Auto-stereoscopic 3D display, TOF depth camera

I. INTRODUCTION

3DTV has been recently considered as the next generation of multimedia consumer products. For that purpose, the development of stereoscopic displays has been considerably promoted by several display manufactures.

Some 3D contents which aim at the stereoscopic display are being made centering the entertainment field such as TV, movie, Video games. In general, such contents for stereoscopic display are generated based on computer graphic techniques, because depth information can be easily obtained. Creating 3D content from live real scene for auto-stereoscopic displays [1] remains more difficult because it required to compute depth information in real time. This is still a very important issue.

For that purpose, a depth camera can used to obtain depth information from a real scene. By associating the depth image with the corresponding color image, it is possible to quickly generate the input images for auto-stereoscopic displays.

Our framework for interactive 3D content generation for auto-stereoscopic display consists in a depth camera based on the Time Of Flight technology [2], a color camera and an auto-stereoscopic display.

The color camera is used to capture the texture corresponding to the depth map. The 3D content is generated by merging the depth map and the texture image and sending the result to the auto-stereoscopic display. Since the viewpoints of depth and color cameras are slightly different, the image from one camera should be transferred to the other camera's viewpoint.

This paper is structured as follows. We start giving an overview of related and previous works, then we present a description of our equipment. In the next section we give details about our full system and algorithms. Finally, we present and discuss the results of our approach.

II. RELATED WORK

A common method to generate input images for auto-stereoscopic display is to use as many cameras as the required views. Naemura [et al.] have developed a real time 3D system called "TransCAIP" [3]. It uses 64 images captured by web cameras and generates the new 60 viewpoint images by using image based rendering techniques. Since this system needs many cameras, it is not convenient to apply it in every situation. Moreover, when considering broadcasting or streaming, such huge volume of data is not acceptable.

Another algorithm called 2D plus Depth generates the input images thanks to depth information [4]. In these method, the 2D points of an original image are re-projected into the 3D world coordinate by using depth information. After that, these 3D points are projected back onto the image plane of the virtual camera which is located in left or right side of original camera position and we can get the different view images. Such depth information can be obtained by stereo measurement of the input images [5, 6]. However, computational time is a significant limitation in such case.

The problem of computational time may be solved by applying the high spec LSI and GPUs with low prices. Nozick [et al.] have proposed a pixel matching method using Plane Sweep algorithm on GPU [7]. Since each process is parallelized, the process can be performed in real-time. However, the resolution of images is about VGA or QVGA, because of computational time limitation. For displaying 3D contents on a large monitor, resolution is a very important issue.

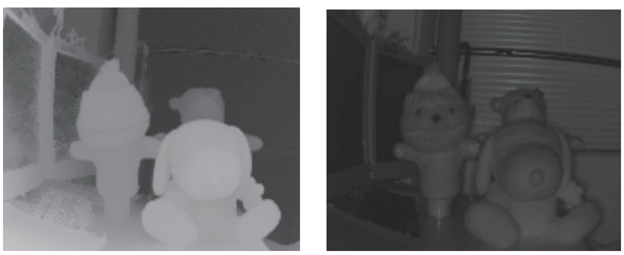
A depth camera can directly capture depth information. Especially, a Time Of Flight (TOF) camera obtains the distance to the object by measuring the time of reflected infrared light. Since the system is simple and the processing speed is fast, it is suit for capturing the dynamic scene. Therefore our system employs such king of camera.

III. OVERVIEW

A. The Time Of Flight Camera

A camera using TOF can measure distances in a scene in real time. The camera emits infrared light that is reflected by the objects and come back onto the camera's sensor. Then, the traveling time of the light is measured for each pixel of the sensor and used for computing the depth of the scene.

Figure 1 presents images captured by the TOF camera. Figure 1(a) is the depth map where bright areas are closer from the camera than dark areas. Figure 1 (b) is the intensity map of the reflected infrared rays that can be used as a roughly gray scale image. However, the resolution of such a TOF camera is only 176 x 144 and it is not possible to retrieve the color information. Therefore, a high resolution color camera should be added and the view point of the depth map should be transferred onto the color camera's view point.



(a) depth map (b) intensity map

Figure 1. An example of images captured from the TOF camera.

B. Auto-stereoscopic Display

Auto-stereoscopic displays offer the ability for one or several users to see 3D-images without wearing any glasses. Depending on their characteristics, auto-stereoscopic displays require from 5 to 64 images to display a single 3D frame. A filter, made of small lenses or a fine grating, is overlaid on the surface of the screen and ensures to emit each image in a specific direction. So, if the user is well located in front of the display, each eye can see a single specific image.

Some auto-stereoscopic displays, like Philips one [8], employ the 2D plus depth [9] algorithm to save data transfer bandwidth between the screen and the computer. It requires a single color image associated with its corresponding depth map. The other views are generated directly by the screen using these information as depicted in Figure 2.

IV. PROPOSED SYSTEM

A. System overview

As explained in section III. A., the depth camera does not provide color information and is low resolution. It means that captured data cannot be directly used with the auto-stereoscopic display by applying the 2D plus depth algorithm. To fit the requirement of the screen, a high resolution color camera is added beside the depth camera. Such a system is presented in Figure 3.

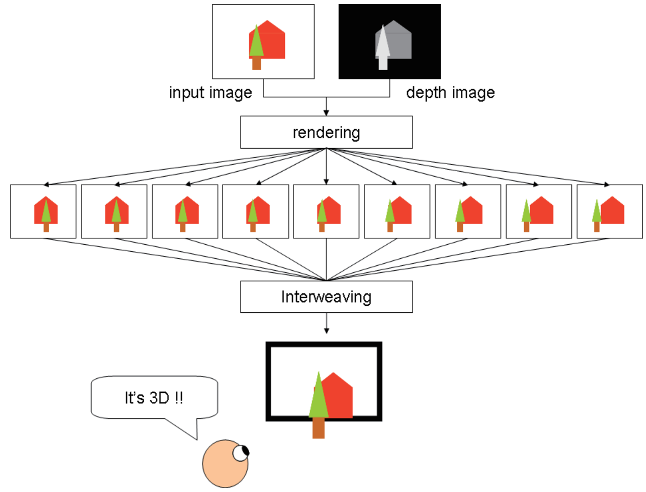


Figure 2. Overview of 3D image rendering.process

Since the color camera and the depth camera are not located at the same position, their viewpoints are slightly different. Therefore, a transformation is required to match the depth map with the color image.

When mapping color and depth images, a lack of data occurs because of occlusions. Since color is the most relevant visual information perceived by humans, the transformation is then applied on the depth map.

The Process of our system is presented in Figure 4. The process consists in two parts. First, considering that both cameras are fixed, we pre-compute the intrinsic and extrinsic parameters of the cameras. Second, for each input frame from the depth camera, we apply the view transformation technique to match the color image by using the computed camera parameters. The result is then converted according to the requirement of our auto-stereoscopic screen input format.

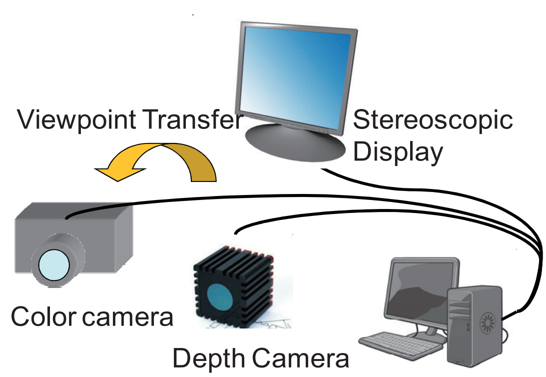


Figure 3. Depth camera with color camera for video capturing to display onto auto-stereoscopic display.

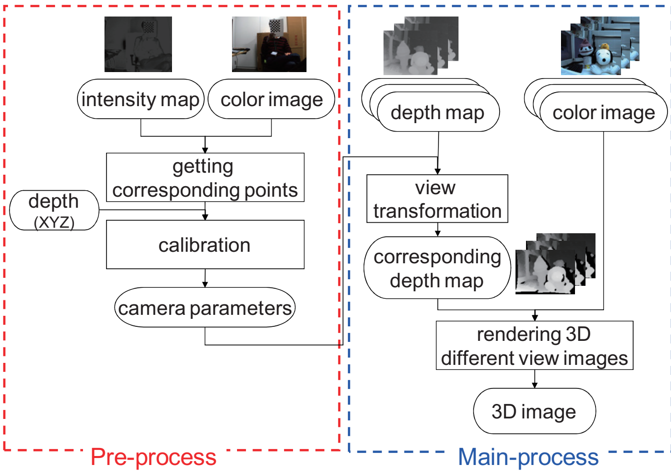


Figure 4. Flow chart of our method.

B. Camera calibration

The calibration phase consists in evaluating the intrinsic and extrinsic parameters of the color and depth cameras.

Intrinsic parameters are computed using the Zhang's method [10] with a chessboard.

Considering that the depth camera is positioned at the origin of the system, only extrinsic parameters of the color camera have to be computed. Since we can get correspondences between the both cameras by manual click as shown Figure 5 and obtain the 3D coordinates corresponding to these points by depth values, it's possible to compute the matrix that projects 3D points into color image space. By using this matrix, the depth map is transformed to the color camera's viewpoint at every frame. This calibration is necessary only once unless we move one of the cameras.

In our current work, we click the corresponding points by manual, although some feature extraction methods like SIFT [11] can be applied to this process. Therefore it will be automatically performed in future work.

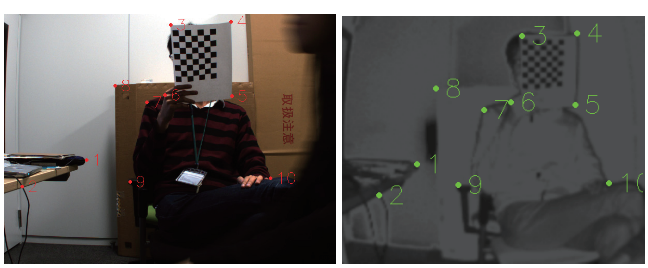


Figure 5. Correspondences between images from depth and color cameras..

C. Viewpoint transformation

Each frame of the depth camera is composed of 25000(≒176×144) points. This number of 3D points is not sufficient for view transformation because the resolution of the depth and color cameras as very different as shown in Figure 6.

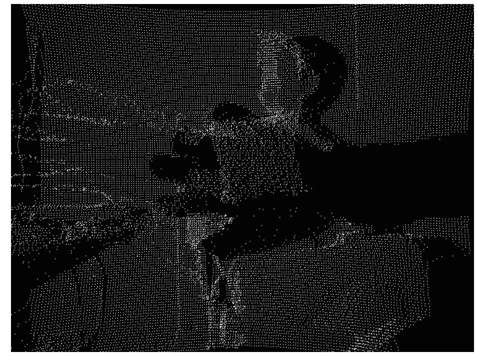


Figure 6. Example of insufficient 3D points.

In order to resolve this problem, we create the distance map by reducing the width and height of the depth map. In the image, the distance between each 3D point and the color camera is mapped. Therefore each pixel *(x, y)* in the distance map has the distance value *d(x, y)*.

The distance is obtained by projecting the 3D point to the image using the camera parameters. The result is normalized using the following equation:

$$I(x, y) = 255\ (d(x, y) - d_{min}) / (d_{max} - d_{min})$$

where d_{min} and d_{max} are respectively the distances between the color camera and the nearest/farthest 3D point from the viewpoint.

Since some occlusions exist in the depth map, we apply a median filter to fill the missing areas. However, this approach is not efficient when these areas are too big and should be completed by another in-painting method.

V. RESULT & DISCUSSION

We present results to check the availability of our proposed method using the following environment.

- CPU Core 2 Duo : 3.0 GHz
- Memory : 2 GB
- Resolution of color camera : 640 x 480
- Resolution of depth camera : 176 x 144

Figure 7 depicts the view transformation result. The depth image shown in the middle is transformed to match the upper image. The lower image is the result of our view transformation algorithm. The matching correctness of the result between the depth map and the color image is shown thanks to the red lines.

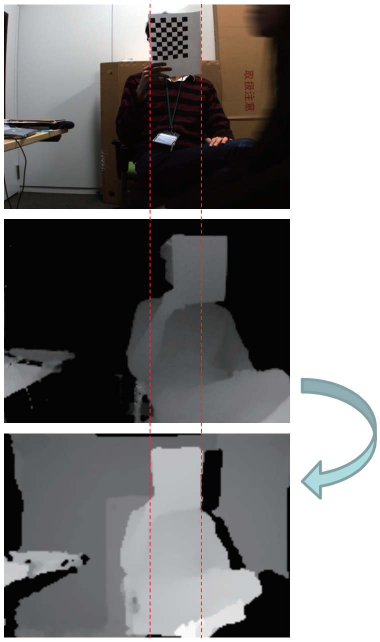


Figure 7. Result of viewpoint transfer of depth video. The top image is the input color image. The middle image is the original depth map. The bottom image is the transformed depth map.

Another result is presented in Figure 8. Some occlusion areas represented by black regions in the result image cannot be avoided using the median filter method. Black areas on the bottom and left parts are the consequences of the radial distortion correction applied by the hardware of the depth camera.

Figure 8. Result of viewpoint transfer of depth video.

As we described in section III A, the TOF depth camera can generate and transmit the data in real time. In that sense, our system needs computational cost only in the translation phase. The full process of our system runs at an average frame-rate of 10 frame / seconds. This can be improved by using a GPU framework.

Figure 9 shows the picture of the exhibition "InterBEE 2009 (International Broadcasting Equipment Exhibition)" held in November 2009. We have demonstrated our system that shows 3D video with an auto-stereoscopic display. Our system presented the 3D video of the captured real scene in real time and attracted from lots of participants because other demonstrations of 3D display in the exhibition required some glasses or displayed only CG objects.

Figure 9. System for real-time 3D capturing with depth and color for auto-stereoscopic display

VI. CONCLUSION

In this paper, we have presented a method to generate 3D content for auto-stereoscopic display from live captured real scenes. A depth camera associated to a color camera are used to fit the requirement of the 2D plus Depth algorithm. Since the positions of these cameras are slightly different, we apply fast image transformation to match the color and depth information.

In near future, we want to apply an extra algorithm to fill the holes generated from the occlusions like an in-painting method.

REFERENCES

[1] Neil A. Dodgson, "Autosteroscopic 3D Displays" , Computer, pp. 31-36, August, 2005.

[2] Thierry Oggier, et al., "An all-solid-state optical range camera for 3D real-time imaging with sub-centimeter depth resolution (SwissRanger)", Proc. SPIE, Vol. 5249, 534 (2004).

[3] Y. Taguchi, T. Koike, K. Takahashi, T. Naemura, "TransCAIP: A Live 3D TV System Using a Camera Array and an Integral Photography Display with Interactive Control of Viewing Parameters", IEEE Trans. on Visualization and Computer Graphics, vol..15, issue 5, pp.841-852, 2009.

[4] Christoph Fehn, "A 3D-TV approach using depth-image-based rendering ", The IASTED International Conference on Visualization, Imaging and Image Processing Vol. 3.

[5] J. Woetzel and R. Koch, "Real-time multi-stereo depth estimation on GPU with approximative discontinuity handling", CVPM 2004.

[6] R. Yang, M. Pollefeys, H. Yang and G. Welch, "A Unified Approach to Real-Time, Multi-Resolution, Multi-Baseline 2D View Synthesis and 3D Depth Estimation using Commodity Graphics Hardware",IJIG 2003.

[7] Vincent Nozick and Hideo Saito, "On-line Free-Viewpoint Video: From Single to Multiple View Rendering", International Journal of Automation and Computing, Volume 5, Issue 3, pp. 257-265, July 2008.

[8] PHILIPS 3D Solutions, "3D Interface Specifications, White Paper", http://www.philips.com/3dsolutions

[9] Christoph Fehn, "A 3D-TV system based on video plus depth information", Asilomar Conference on Signals, Systems, and Computers vol. 2 pp.1529- 1533 2003.

[10] Zhengyou Zhang, "A flexible new technique for camera calibration" IEEE Transactions on Pattern Analysis and Machine Intelligence, 22:1330-1334, 2000.

[11] David G. Lowe, "Object recognition from local scale-invariant features", Proceeding of IEEE International Conference on Computer Vision (ICCV), pp.1150-1157, Sep. 1999.