New View Video Synthesis via Real Time Depth Image Processing

Hideo SAITO, Yuhki TAKAYA, Songkran JARUSIRISAWAD, Yuko UEMATSU †, François de SORBIER † ‡

† : Keio University, {saito,takaya,songkran,yu-ko,fdesorbi}@hvrl.ics.keio.ac.jp
‡ : University Paris-Est, Marne-la-Vallée, fdesorbi@univ-mlv.fr

Abstract Depth video plays a significant role for generating 3D Video contents. In this paper, we first present a method for generating depth video in real-time from multiple video inputs. In the PGS framework, we employ plane sweeping algorithm using GPU for generating the depth video in real-time. Then we would also show the way of generating the depth video at the same camera parameters as the texture cameras in real-time, so that we can display the object scene with the auto-stereoscopic display in real-time.

1 Introduction

3D Video recently takes much attention as the next generation visual media for immersive communication and broadcasting. Even though variety of types of 3D videos are considered by a lot of researchers, the most important and significant aspect of the 3D Video is depth information of the captured scene, while the conventional video only captures 2D color images in variety resolution/formats.

In most of the situation where 3D video is shown to the audience, the depth video should be generated in real-time so that the 3D video can also be shown in real-time. Therefore generating the depth video in real-time is significant and important for a lot of applications.

In this paper, we first present a method for generating depth video in real-time from multiple video inputs. In this method, we propose to apply Projective Grid Space (PGS) for geometrical registration of the input cameras in 3D sense without estimating extrinsic parameters of the cameras. In the PGS framework, we employ plane sweeping algorithm for generating the depth video in real-time, since GPU computation is fit to the depth computation in the plane sweeping.

Such depth video is also very important for auto-stereoscopic display. Especially in the autos-stereoscopic display developed by Philips, one texture video along with depth video realizes autos-stereoscopic views. In this case, the camera parameters of both depth and texture cameras should completely co-incident. However, it is not possible to make all camera parameters of depth and texture cameras the same, as long as both videos are captured with different types of cameras.

In this paper, we would also show the way of generating the depth video at the same camera parameters as the texture cameras in real-time, so that we can display the object scene with the auto-stereoscopic display in real-time.

2 Depth Video Generation via Plane Sweeping in PGS Framework

2.1 Plane Sweep for Depth Generation

The plane-sweep algorithm [1] creates depth images and novel views of a scene from several input images. Considering a scene where the objects are exclusively Lambertian surfaces, the viewer should place the virtual camera somewhere around the real video cameras and define a near plane and a far plane such that every object of the scene lies between these two planes. Then, the space between near and far planes is divided into several parallel planes π_k. Here, we assume that a point lying on a plane whose projection on every input camera provides a similar color will potentially correspond to the surface of an object. Considering a visible object of the scene lying on one of these plane π_k at a point P, this point will be seen by every camera with the same color, i.e., the object color. Now consider another point P' lying on a plane but not on the surface of the visible object, this point will probably not be seen by the capturing cameras with the same color. Figure 1 illustrates this principal idea of the plane-sweep algorithm.

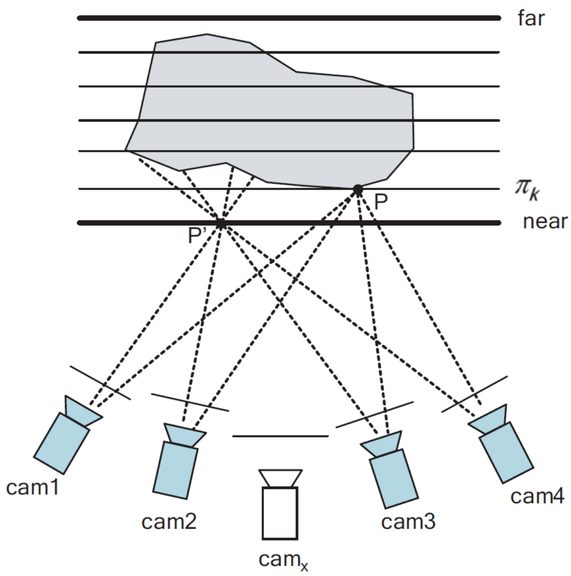
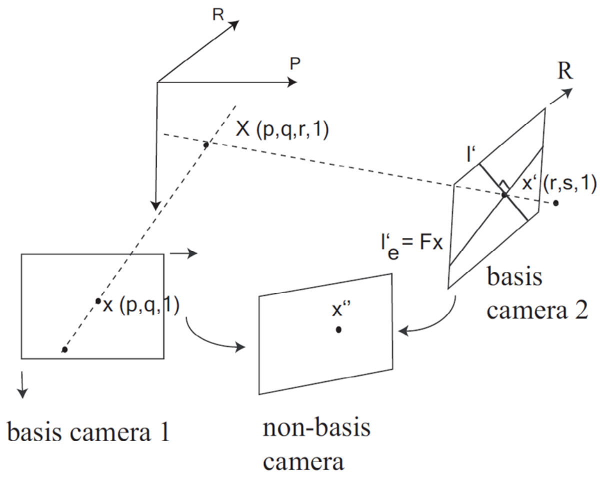


Fig. 1: Defining planes for plane-sweep.

During the new view creation process, every plane π_k is computed in a back to front order. Each point P of a plane π_k is projected onto the input images. A score and a representative color are the computed according to the matching of the colors found. A good score means every camera see a similar color. The computed scores and colors are projected onto the virtual camera cam_x. The pixel color in the virtual view will be updated only if the projected point P provides a better score than the current one. Then the next plane $\pi_k +1$ is computed. The final new view image is obtained once every plane has been computed. This method is detailed on [22].

2.2 Projective Grid Space (PGS)

Projective grid space (PGS) [2] is a 3D space defined by image coordinates of two arbitrarily selected cameras called basis camera 1 and basis camera 2. To distinguish this 3D space from the Euclidean one, we denote the coordinate system in PGS by P-Q-R axis. Figure 1 shows the definition of PGS. x and y axis in the image of basis camera 1 corresponds to the P and Q axis, while x axis of the basis camera 2 corresponds to the R axis in PGS.



Fig, 2: Definition of Projective Grid Space

2.3 Depth Image Synthesis via Plane-Sweep in PGS

For synthesizing the depth image and novel views from the virtual camera in PGS, we define the virtual camera and the plane for sweeping in the framework of PGS [3].

In our case where PGS is used, intrinsic parameters of any camera are unknown. Method for defining virtual camera in calibrated case is not applicable to our case. In our method, the position of the virtual camera is limited to only between two real reference cameras.

Any arbitrary near and far planes in PGS can be defined for doing plane-sweep. In our method we define the planes along the R axis (x image co-ordinate of basis camera 2) as shown in Figure 2. This approach makes the 3D near and far planes adjustment become easy since we can visualize them directly from the image of basis camera 2. This is impossible for the case of the normal plane-sweep algorithm in the Euclidean space in which full calibration is used. In that case, actual depth of a scene has to be measured so that near and far planes cover all volume of interest.

In our approach, basis camera 2 will not be used for color consistency testing during perform plane-sweep because every planes would be projected as a line in this image. So the basis camera 2 is needed only for weakly calibrated cameras to PGS, after that we can disable it to save CPU time.

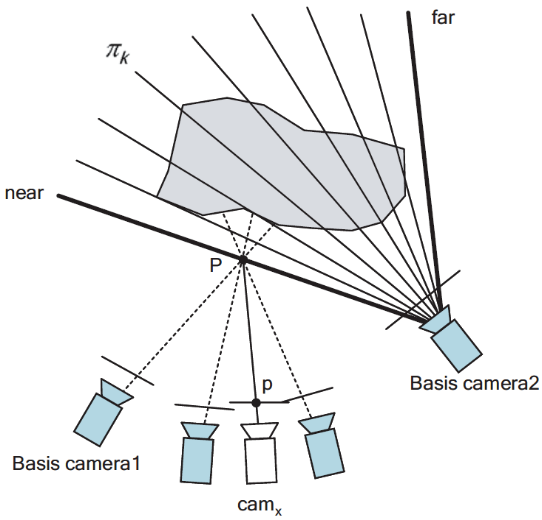


Fig. 3: Defining planes for plane-sweep in projective grid space.

2.4 Experimental Results

We tested our proposed method on PC Intel(R) Core(TM) 2 Duo 3.00GHz CPU with graphic card NVIDIA Quadro FX 570. Six Logitech web-cams with a resolution 320x240 are used to capture input videos. The camera setting is depicted by Figure 4, where camera 1 and camera 6 are selected as two basis cameras. We can see the planes for depth estimation via sweeping as the vertical lines in the basis camera 2 (cam6), which are shown in Figure 5(a). In Figure 5(c), the intermediate viewpoint image between cam1 and cam2 is shown.

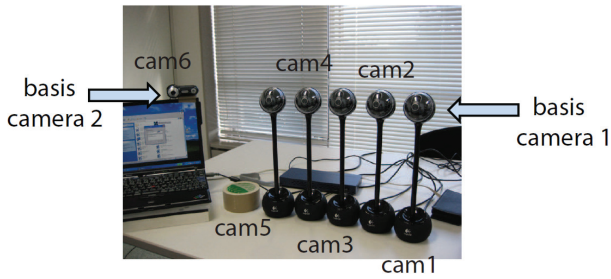
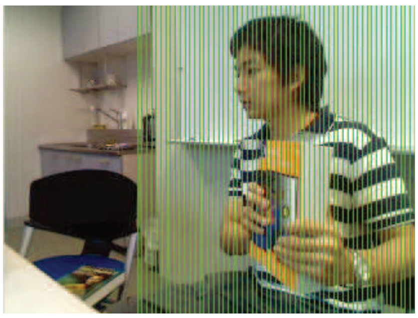


Fig. 4: Camera setup in the experiment.



(a) basis camera 2

(b)cam1, cam2

(c) Intermediate viewpoint between cam1 and cam2

Fig.5: Intermediate viewpoint image synthesized between cam 1 and cam2.

3 Depth Video Generation for Auto-stereoscopic Display

Using the method presented in the previous section, we can get both color video and depth video at the same virtual viewpoint. Such color video along with the depth video is important in particularly using auto-stereoscopic display. Therefore, the method presented in the previous section is the one of the possible method for generating 3D contents for such auto-stereoscopic 3D video presentation. However, the depth and color images generated at the virtual viewpoint using the plane sweeping method is not completely accurate and correct, so the displayed video is not always good for the 3D display presentation.

Recently, various depth capturing cameras are available. Especially, TOF (time of flight) based cameras [4] can capture accurate depth videos for almost all scenes within the pre-defined depth range, while depth cameras based on multiple viewpoint stereo cannot provide stable depth information in less textured area. Therefore, using the depth video captured with such TOF based depth camera can be considered as a simple and effective way to capture the 3D contents for displaying in auto-stereoscopic display. Unfortunately, the depth camera does not provide sufficient resolution color video to display in the auto-stereoscopic display, so capturing the color video should be performed with the high-resolution color cameras independently from the depth video capturing, , as shown in Figure 6.

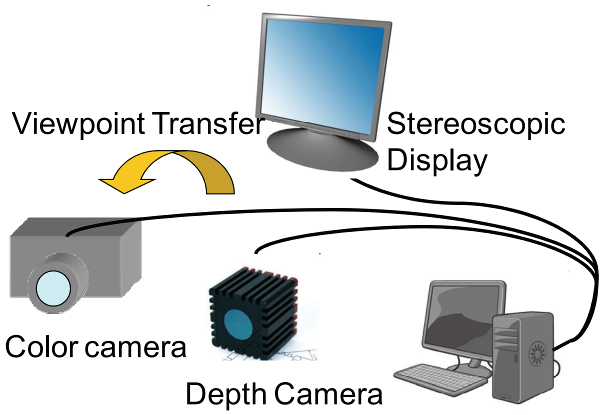


Fig. 6: Depth Camera with Color camera for video capturing to display onto auto-stereoscopic display.

(a) Color camera image (b) Depth camera image

Fig.7: Color Camera image and depth camera image. Each viewpoint is different from each other, so the captured position of the hand is not the same.

In this case, the viewpoint of color and depth videos in not the same as shown in Figure 7, but they should be the same for showing correct 3D contents in the auto-stereoscopic display. We can apply view synthesis technology for making them at the same viewpoint. We can change the viewpoint of either the color video or depth video, but we generate the new depth video at the same viewpoint of color camera so that we can avoid hall area for the color video which is more significant visual information than the depth video. Figure 8 shows an example of viewpoint transfer for depth video.

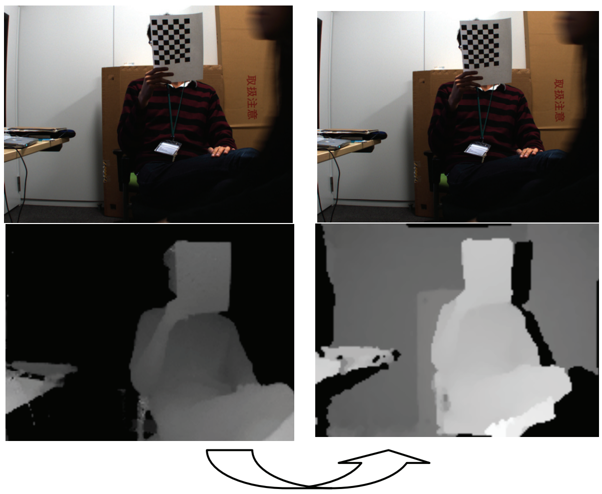


Fig.8: Example of viewpoint transfer of depth video. The depth image shown in the bottom left is transferred to bottom right, so that the viewpoint can be at the same position as the color image. (The color images on the top is the same, but shown for the comparison of the depth images.

Even though the transfer of the depth video to the color camera viewpoint is well known technique in computer vision area, there are some research issues to be considered: rendering speed, hall region in the transferred depth video by occlusion, and resolution difference between color camera and depth camera. The hall region cannot be avoided along the area of the object boundary as shown in the bottom right in Figure 8. This is a significant issue in such viewpoint transfer from captured depth camera. We will tackle with those issues to improving the quality of the 3D video contents in near future research.

In InterBEE 2009 (International Broadcasting Equipment Exhibition), held in November 2009, we demonstrated the prototype system that shows 3D video captured with both color cameras and depth cameras onto an auto-stereoscopic display. Figure 9 shows the picture of the exhibition.

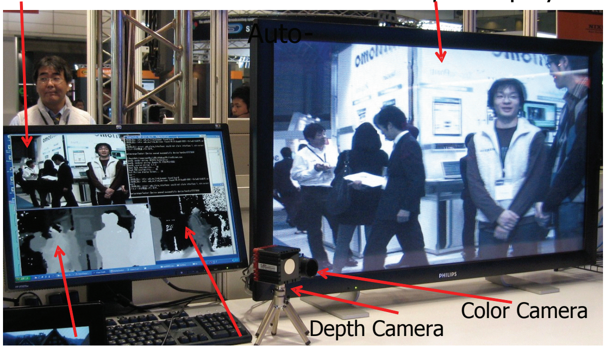


Fig9: System for real-time 3D capturing with a depth and a color for auto-stereoscopic display.

4 Conclusion

In this paper, we first present a method for generating depth video in real-time from multiple video inputs using a plane sweeping algorithm in the PGS framework. Then we also show the way of generating the depth video at the same camera parameters as the color cameras in real-time, so that we can display the object scene with the auto-stereoscopic display in real-time.

Acknowledgment This work is partially supported by NICT (National Institute of Information and Communications Technology).

References

[1] Vincent Nozick, Sylvain Michelin and Didier Arquès, Real-time Plane-sweep with local strategy. in the Journal of WSCG, The 14-th International Conference in Central Europe on Computer Graphics, Visualization and Computer Vision, 14(1-3):121–128, feb 2006.

[2] H. Saito, T. Kanade, Shape reconstruction in projective grid space from large number of images, in: Proceedings of the IEEE Computer Society Conference on Computer Vision and Pattern Recognition (CVPR'99), Vol. 2, 1999, pp. 49-54.

[3] Songkran Jarusirisawad, Hideo Saito, Vicent Nozick, Real-Time Free Viewpoint Video from Uncalibrated Cameras Using Plane-Sweep Algorithm, The 2009 IEEE International Workshop on 3-D Digital Imaging and Modeling (3DIM2009), pp. 1740-1747, Kyoto, Japan, 2009.

[4] 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)

Saito, Hideo: received Ph.D. degree in Electrical Engineering from Keio University, Japan, in 1992. He has been on the faculty of Department of Electrical Engineering, Keio University, since 1992, and now he is a Professor. In 1997 to 1999, he stayed in the Robotics Institute, Carnegie Mellon University as a visiting researcher. He is a senior member of IEEE, and IEICE, and a member of IPSJ, SICE, VRSJ, and ITE.

Takaya, Yuhki: received B.E. degree in Information and Computer Science from Keio University, Japan, in 2009. He is now a graduate student for master degree in Keio University.

Jarusirisawad, Songkran: received Ph.D. degree in Engineering from Keio University, Japan, in 2009. He is now working at Freewill Solutions, Thailand.

Uematsu, Yuko: received Ph.D. degree in Engineering from Keio University, Japan, in 2009. She is now a Research Associate in Keio University.

De Sorbier, François: received Ph.D. degree in Computer Science from University Paris-Est, Marne-la-Vallée, France, in 2008. He is now a JSPS Fellowship and joining in Keio University.