MULTI-LAYERED CONTENTS GENERATION FROM REAL WORLD SCENE BY THREE-DIMENSIONAL MEASUREMENT

M. K. Kim[1], Y. Nakajima[1, 2], T. Takeshita[1], S. Onogi[2], M. Mitsuishi[1] and Y. Matsumoto[1, 2]
[1]*School of Engineering, the University of Tokyo, Tokyo, Japan*
[2]*Intelligent Modeling Laboratory, the University of Tokyo, Tokyo, Japan*

Keywords: Layer Content, Three-dimensional measurement, Depth from Focus, Spatio-Temporal image analysis.

Abstract: In this paper, we propose a method to create automatically multi-layered contents from real world scene based on Depth from Focus and Spatio-Temporal Image Analysis. Since the contents are generated by layer representation directly from real world, the change of point of view is able to freely and it reduces the labor and cost of creating three-dimensional (3-D) contents using Computer Graphics. To extraction layer in the real images, Depth from Focus is used in case of stationary objects and Spatio-Temporal Image Analysis is used in case of moving objects. We selected above two methods, because of stability of system. Depth from Focus method doesn't need to search correspondence point and Spatio-Temporal Image Analysis has also simple computing algorithm relatively. We performed an experiment to extract layer contents from stationary and moving object automatically and the feasibility of the method was confirmed.

1 INTRODUCTION

Three-dimensional (3-D) contents are required in various field of Virtual Reality. It is various from cultural asset to cityscape. In case of cultural asset, its digitalization is required quite complex and precise description. The cultural asset is digitalized by using 3-D range measurement information of object and Computer Graphics for delicate contents. In this case, contents generation is also necessary to measurement apparatus such as helicopter or Global Positioning System. Hence, it needs huge labor and cost.

On the other hand, 3-D content digitalization of cityscape (e.g. building, car) is not necessary delicateness such as cultural asset. For instance, when you go sightseeing or visit virtual space, used all contents should not be elaborate like the cultural asset contents.

In the latter case, if real image is used to create 3-D contents directly, it is possible to generate simply more real 3-D contents by less effort and cost.

However, it has been a little problem. When we see the street through the Street View, we often don't go where we want to see. Since photograph data is expressed discretely.

Therefore it has limitation of viewpoint. Google Street View is one example of latter case.

To solve this problem, we employ a solution for 3-D scene description. The concept is 3-D representation with 2-D real image layer. In this method, 3-D space and contents is represented by using arrangement of 2-D layers and concerning psychology factor. This 3-D representation method is proposed and explained their effects by Ogi (Ogi07). There are the following two advantages. First, data volume is small, because one layer can express multi-point of view. Second, we can get continuous point of view that doesn't get before.

Our team with Ogi have been developing Dome Display Layer Representation System based on Figure 1. In this system, 2-D contents are presented in dome display and psychology factor (Seno08) is considered at the same time. So, high presence is obtained. In addition, data size is small because of layer representation. And this system is preformed three steps. First, 2-D layer contents are made and segmented from the real world image. Next, Layers are integrated according to the relation of the point of view and the relation of 3-D position of layer. Finally, these 2-D layer contents are presented on dome display in consideration of the distortion.

Kim M., Nakajima Y., Takeshita T., Onogi S., Mitsuishi M. and Matsumoto Y. (2009).
MULTI-LAYERED CONTENTS GENERATION FROM REAL WORLD SCENE BY THREE-DIMENSIONAL MEASUREMENT.
In *Proceedings of the Fourth International Conference on Computer Vision Theory and Applications*, pages 109-112
DOI: 10.5220/0001820501090112

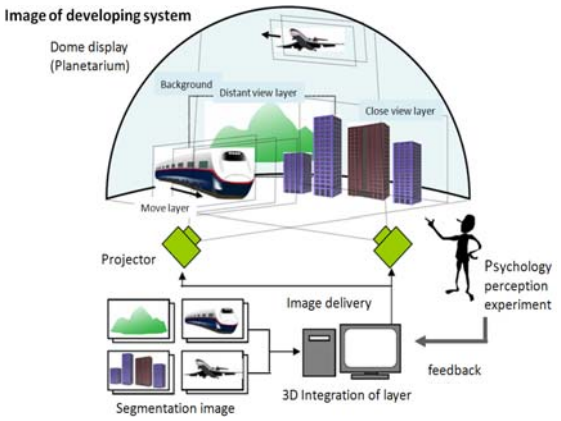

Figure 1: Dome Display Layer Representation System.

In this project, we have been developing automatic extraction method of layer contents with real world image. Layers are made from real world by 3-D measurement. 3-D measurement is necessary to extract layer from 2-D image. By using this method we can record cityscapes or landscapes and change point of view everywhere we want. Actually, the 3-D measurement performs discretely but we can go to midpoint between measurement points and get the scene naturally.

In this paper, two methods are proposed to layer extraction.

First, in case of extraction of stationary layer, we selected Depth from Focus method to extraction layer. DFF is depth measurement method by the degree of focus in different focus position image ((Nayar94) (Pentland87)). Mainly, using motion, stereo or focus method is used for 3-D measurement method from 2-D image. Motion and stereo utilize changes of point of view. So they need to search corresponding point. The search algorithm is quite complex. Meanwhile focus method doesn't need to search corresponding point and the algorithm is simple. Then we use focus information as extraction method. Second, extraction of moving object is performed by Spatio-Temporal Image Analysis method. This method is relatively more stable in terms of search less correspond point than other method. And it is also possible to generate 3-D layer contents from already existing video contents. But it is unsuitable to extract still object.

If consider two methods, both still object and moving object can be extracted at the same time. We proposed a method to generate multi-layer contents from real world scene by Depth from Focus and Spatio-Temporal image analysis and we extracted layer contents automatically.

2 METHOD

In this section, we introduce Depth from Focus and Spatio-Temporal Image Analysis methods that extract layer contents from real world scene.

2.1 Depth from Focus

In general, using motion and stereo camera method is widely used for depth measurement from camera to objects (Birch) (Krotkov). Motion and stereo methods basically utilize changes of point of view. These methods need to search and compute correspondence points. The search and computing algorithm is quite complex. Meanwhile Depth from focus method doesn't need to search corresponding point and the algorithm is simple.

Depth from focus method uses focus information. Focus condition depends on the geometry setting of the focal plane and object plane in a camera. Roughly speaking, we can get more blurred image, according to image plane is more distant from sensor plane. The geometry of imaging is shown in Figure 2.

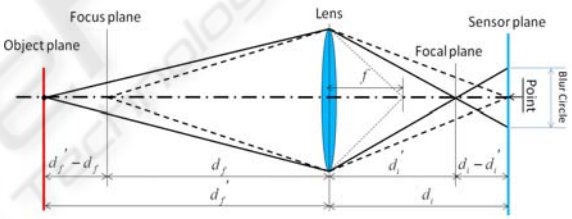

Figure 2: Geometry of imaging.

Depth estimation performed following steps. First, two or more images of different focused position in each image can be acquired by changing the sensor plane position or lens position at the identical viewpoint. Next, the focus measure is evaluated. In DFF, focus measure means degree of focus in each image pixel. In third step, best focus lens position is found. And then, the relation between camera setting and the focal plane is previously calculated by law of lens. Consequently, we can estimate depth information from this process.

2.2 Spatio-Temporal Image Analysis

An image hexahedron is obtained by arranging the video image sequence like Figure 3 (a). The cross section view of the hexahedron perpendicular to image plane is the image like Figure 3(b). The axis of the image is position-time. So the image is called Spatio-Temporal image. Spatio-Temporal image has

a feature at the position that color is not changed, the image becomes stripe pattern in the time direction. In this paper, the moving object is extracted using this feature. To detect the stripe pattern, we use frequency analysis by 2-D Fourier transformation.

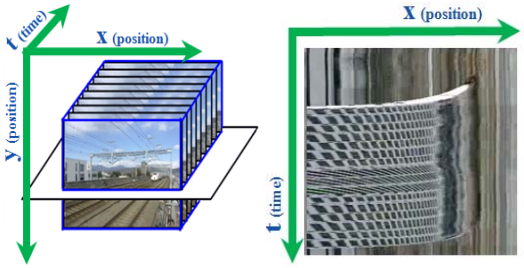

Figure 3: (a) image sequence (b) Spatio-Temporal image.

3 EXPERIMENTAL RESULTS

3.1 Layer Generation based on Depth from Focus

In this section, we computed the depth of object using DFF, and extracted layered object contents by dividing image layer according to depth information.

3.1.1 Depth Estimation with Apple and Orange

Depth distribution was estimated from real scene images shown in Figure 4. And the computed depth map of Figure 4 is shown in Figure 5 (all depth is expressed by color) by different focused images about 40 as Figure 4. For instance, an apple and an orange have different depth. Some areas are not computable (dark blue) because of low texture or CCD saturation.

Figure 4: (a) Defocused image. (b) Focused image.

3.1.2 Layer Segmentation Using Depth Information

Layer segmentation using depth information was performed. In the result, an apple and an orange were extracted from real scene as shown in Figure 6

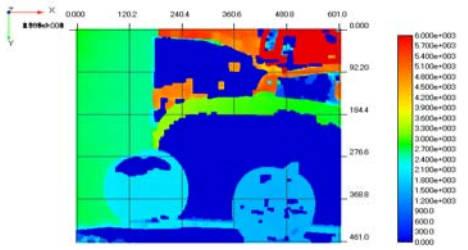

Figure 5: computed depth map.

And Figure 7 (a) shows the layers arrangement according to 3-D positions. And Figure 7 (b) is the same arrangement but different point of view.

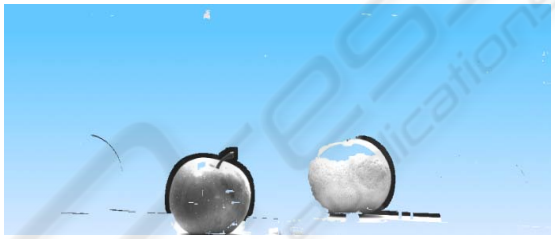

Figure 6: Extracted layer of orange and apple.

Figure 7: (a) Object layers arrangement according to three dimensional positions. (b) The arrangement virtually changed view point (by camera position).

3.2 Layer Generation based on Spatio-Temporal Image Analysis

In the experiment, the extraction of moving train in video image was conducted by shown as Figure 10. One of the spatio-temporal images is shown in Figure 7(a). And in Figure 7 (b) the background region judged by our method was expressed with blue. The final result is shown in Figure 8.

Figure 8: Train image.

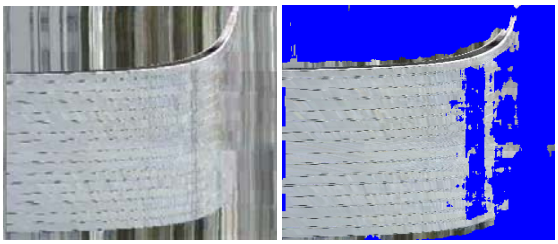

Figure 9: (a) Spatio-Temporal Image. (Blue is still region).

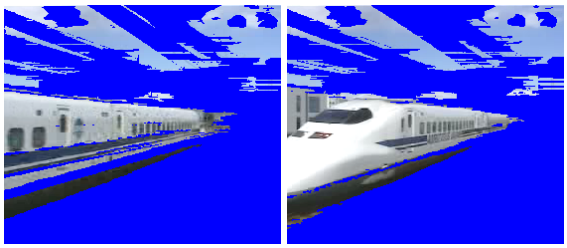

Figure 10: Extracted train.

3.3 Presenting Extracted Contents to Dome Display

We use a hemispherical dome display with 3 meter of diameter for presenting layered images in Intelligent Modeling Laboratory of the University of Tokyo. A design drawing and actual appearance of display are shown in Figure 13. The layered images were integrated into a projection image. Then the image was projected from the floor just under the screen. We performed direct fusion of real-image layer on the display by using depth information.

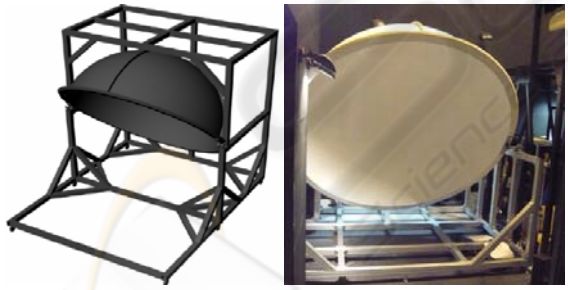

Figure 11: (a) Design drawing. (b) Appearance of dome display.

4 DISCUSSION

Two major problems occurred in implementing the system. First problem is concerning texture of object. If object has no texture or too low texture, it is difficult to detect the motion or degree of focusing. So position detection becomes unstable. This problem is also occurred when the object is too bright or dark. In this case, the object has no texture because of the saturation of image sensor. To cope with it, the region enclosed by one object layer.

The next problem is about boundaries of object. To calculate Fourier transform in Spatio-Temporal image or the degree of focusing of a point, information of surrounding area information is needed in the images. Therefore, in Figure 6 and Figure 10, the edge of object is larger than real object size. So to extract objects accurately, it is necessary to optimize Fourier transform window or filter size.

5 CONCLUSIONS

We proposed a method to generate multi-layer contents from real world scene by using Depth from Focus and Spatio-Temporal Image Analysis method. Although some problems mentioned above are remained, the feasibility of the method was confirmed.

REFERENCES

T. Ogi, M. Hayashi, 2007, *Dome Image Contents based on Layered Image Representation, ASIAGRAPH 2007 in Tokyo Proceedings*, Vol.1, No.2, pp.113-118, 2007.10.

Takeharu Seno, Masahiro Hayashi, Tetsuro Ogi, Takao Sato: *Virtual Depth Effects for Non-Stereoscopic Dome Images -The Estimation of the Depth Effects of the Dome Image by Psychophysics, ASIAGRAPH 2008 in Shanghai Proceedings*, Vol.2, No.1, pp.121-126 2008.6.

S. K. Nayar and Y. Nakagawa, 1994, *Shape from focus: An effective approach for rough surface, IEEE Transaction, On Pattern Analysis and Machine Intelligence*, 16(8):824-831.

A. Pentland, 1987, *A new sense for depth of field, IEEE Transaction, On Pattern Analysis and Machine Intelligence*, 9(4):523-531.

S. Birch_eld and C. Tomasi, 1999, *Multiway cut for stereo and motion with slanted surfaces, in ICCV*, pp.489-495.

E. Krotkov and R. Bajcsy, 1993, *Active vision for reliable ranging: cooperating focus, stereo, and vergence, International Journal of Computer Vision*, 11:187–203.