POSTER:Construction of Panoramic Depth Image

Koji Nishio
Osaka Institute of Technology
1-79-1, Kitayama
Hirakata, Osaka
Japan(573-0196)
nishio@is.oit.ac.jp

Keiko Hiraoka
Osaka Institute of Technology
1-79-1, Kitayama
Hirakata, Osaka
Japan(573-0196)
hiraoka@tipl.is.oit.ac.jp

Chikara Yamashita
Osaka Institute of Technology
1-79-1, Kitayama
Hirakata, Osaka
Japan(573-0196)
yamashita@tipl.is.oit.ac.jp

Ken-ichi Kobori
Osaka Institute of Technology
1-79-1, Kitayama
Hirakata, Osaka
Japan(573-0196)
kobori@is.oit.ac.jp

ABSTRACT
We propose a method which constructs a panoramic depth image. We use a digital still camera and a panoramic lens. The lens has omnidirectional mirror that is pointed in the camera direction and the camera can shoot a horizontal panoramic image through it. We use two panoramic images to construct a panoramic depth image. Each image is photographed at different height and there is the parallax difference between these images. Our method seeks corresponding point between these images and calculates the depth value at each pixel. In addition, we had some experience to evaluate effectiveness of our method. The results shows that our method is useful for making three-dimensional contents form the real scenery.

Keywords
Computer graphic, image processing, panoramic image, depth image, three-dimensional reconstruction.

1. INTRODUCTION
In recent years, flat panel displays such as LCD and plasma display have reached mass. Particularly these few years, 3D televisions are launched on the market by various manufacturers and the demand for 3D contents is increasing.

The easy way to get 3D contents is to record the real scenery with a stereo camera. Another way is to make three-dimensional objects and render them using some computer graphic software. However, such software demands the specialized knowledge of computer graphics and the operation skill. In addition, it requires much production time to construct virtual scenery.Then, the techniques constructing the virtual sceneries from photographed images taken by a camera have gotten a lot of attention recently. With this technique[Jou94a][Tra89a][Tra92a], the virtual scenery can be built up easily because it doesn't require the specialized knowledge about computer graphics. However it requires a huge number of shots to build up detailed virtual sceneries by using a conventional camera. Because of this reason, the method by using the omnidirectional mirror has gotten a lot of attention recently. With this mirror, a number of shots can be reduced because of its wide field of view.

When using computer graphics, the virtual scenery can be viewed stereoscopically by using the depth information preserved at each pixel. However it is necessary to give the depth information to the photographed image in order to view them stereoscopically because the photographed image doesn't have the depth information. The image with the depth information can be generated by using a conventional stereo camera. However there is no established standard method to reconstruct the depth information from photographed images through the omnidirectional lens. In the present report, we propose a method which generates the image with depth by storing the depth information to the omnidirectional image. Our method makes it easy to generate the virtual scenery from photographed omnidirectional images. We had some experiment to evaluate our method. In the experiment, we use a conventional digital still camera SONY DSLR-A900 and an omnidirectional mirror EIZOH Wide70 as shown in figure 1. Figure 2 shows the image photographed by this system.

Figure 1. Omnidirectional mirror

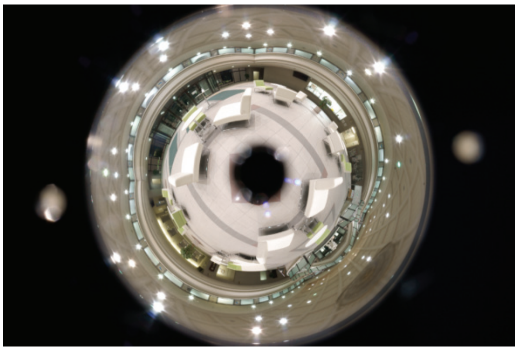


Figure 2. Omnidirectional image

2. OFFSET VERTICAL VISION

Our method uses two omnidirectional images and calculates depth information by applying stereo vision to these images. Stereo vision is a method which reconstructs the three-dimensional information from two images photographed at a different view point as shown in figure 3.

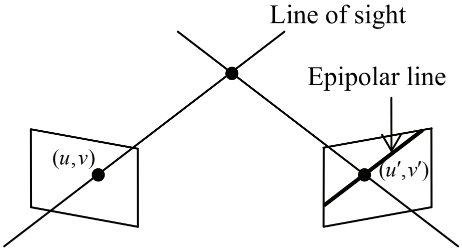


Figure 3. Stereo vision

To reconstruct the three-dimensional information on the principle of stereo vision, it is necessary to find the corresponding point between two images. These corresponding points are found with the constraint that spatial positions correlated to the pixel are on a line. The points corresponding to the pixel on an image plane are on the line of sight which projected to the other image plane. This line is called epipolar line. Then we fixed the camera to the tripod and take shots elevating the camera. This is called offset vertical vision and we define the image photographed at the upper view point as the upper image and the image photographed at the lower view point as the lower image. We can take the upper image and the lower image at a time with a single camera. In addition, there is vertical translation which is called vertical parallax between those images as shown figure 4.

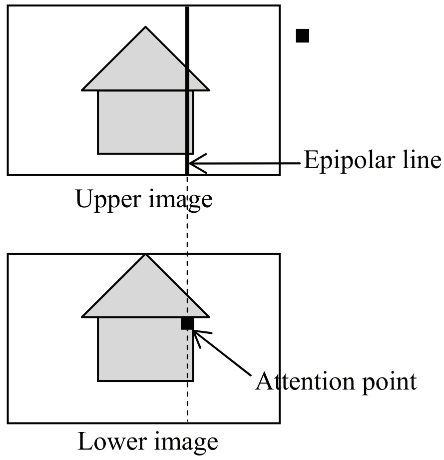


Figure 4. Offset vertical vision and its parallax

3. DEPTH CREATION

Our method generates the panoramic depth image from the omnidirectional image which photographed through the omnidirectional mirror. First, the upper image and the lower image are converted into panoramic images by a development projection. Finally, the depth information is calculated for each pixel from the parallax between developed images. However the process may find incorrect corresponding points by mistake. Then we set a limit to the search range for avoidance of mistake.

Searching for the Corresponding Point

Our method uses template matching to find corresponding points. Template matching searches for the point which is similar to the template image. The process generates the template image from the lower image and finds the corresponding point by searching for the similar region in the upper image.

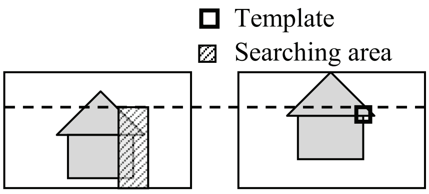
The corresponding point is found in each image photographed at the lower and the upper view point and the depth information for each pixel is calculated from the parallax of the corresponding point. We applied SSD(sum of squared differences) to calculating the similarity measure. This value is calculated by

$$SSD = \sum_{i=0}^{m} \sum_{j=0}^{n} \left(I(x_i, y_j) - T(x_i, y_j) \right)^2$$

I is the target image and T is the template image. (x_i,y_i) is the coordinate of an point in an image. m and n are size of target and template image. As SSD approaches to zero, the target image is similar to the template.

Limitation of the Searching Area

Our method uses the template matching with color information. If the process searches over whole the image, the possibility of miss search is increased at the regions where similar color is. Then we adopt a limitation of the searching area. We uses the epipolar lines, the property of the offset vertical vision and the standard deviation of color. First, as described in Chapter 2, the point in an image is on the epipolar line in the other image. Then the searching area can be limited to the epipolar line and the area around it. Second, objects in the lower image come out below the level of corresponding height in the higher image. Then the searching area can be limited to the area below the line that belongs to the top edge of the template. With these limitation, the searching area is reduced as shown in figure 5.



(a) Upper image (b) Lower image

Figure 5. Limitation of searching area

However, the process may find more than one matching result when there're not enough textures in the image. Then if the standard deviation of the template image is not enough high, the process doesn't proceed the template matching for the template because there're not enough textures.

Calculating the depth value

Our method calculates the depth value by triangulation. The camera position of the upper image is A and the camera position of the lower image is B. It is necessary to measure l, α, β, and r as shown figure 6.

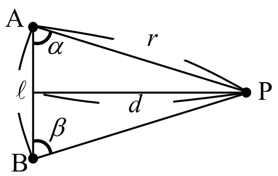
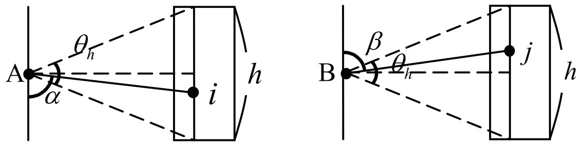


Figure 6. Triangulation

Where the vertical angle of view is θ_h and the corresponding points on the images are i and j, the height of the image is h. The α and β are calculated by

$$\alpha = \begin{cases} \frac{\pi}{2} + \tan^{-1}\left(\left(1 - \frac{2i}{h}\right) \tan\left(\frac{\theta_h}{2}\right) \right) & \left(0 \le i < \frac{h}{2}\right) \\ \frac{\pi}{2} - \tan^{-1}\left(\left(\frac{2i}{h} - 1\right) \tan\left(\frac{\theta_h}{2}\right) \right) & \left(\frac{h}{2} \le i < h\right) \end{cases}$$

$$\beta = \begin{cases} \frac{\pi}{2} - \tan^{-1}\left(\left(1 - \frac{2j}{h}\right) \tan\left(\frac{\theta_h}{2}\right) \right) & \left(0 \le j < \frac{h}{2}\right) \\ \frac{\pi}{2} + \tan^{-1}\left(\left(\frac{2j}{h} - 1\right) \tan\left(\frac{\theta_h}{2}\right) \right) & \left(\frac{h}{2} \le j < h\right) \end{cases}$$



(a) Upper image (b) Lower image

Figure 7. Relationship of the corresponding points and angles

r is calculated by

$$r = \frac{\ell \sin\beta}{\sin(\pi - (\alpha + \beta))}$$

The depth d as shown in figure 5 is calculated from the trigonometric ratio by

$$d = r \sin\alpha$$

Depth Value Correction

Some area is not searched as a result of the limitation of the searching area. Incorrect corresponding points may be found even if the searching area is limited. This reads to an incorrect depth value. Then we adopt a depth value correction. Unsearched area doesn't hold the depth value. First, these depth values are interpolated from neighboring pixels up and down as shown in figure8.

7	8	5
5	-	5
3	-	5
1	2	5

(a) Lack of depth

7	8	5
5	6	5
3	4	5
1	2	5

(b) Vertical interpolation

Figure 8. Depth value correction

Second, some depth is differ greatly from one of surrounding pixels. Then we apply median filter to every depth.

4. RESULT AND CONCLUSION

We had some experiments to evaluate the effectiveness of our method. In this experiment, depth image was generated from a pair of omnidirectinal images as shown in figure 9(a) and (b). The resolution of each image is 10,800 × 3,024. Figure 9(c) shows the result generated from these images. Where the blight area means near and the dark area means far. From this result, it is thought that our method can create the depth image almost successfully. However, some area results in black which means far despite near regions. In these areas, incorrect corresponding points are found by mistake. It is thought that this problem can be solved by setting appropriate aperture value to the camera and by using structured lights[Con03a].

5. REFERENCES

[Jou94a] Subbarao, M., Surya, G. Depth from defocus: A spatial domain approach. International Journal of Computer Vision 13, No.3 pp. 271-294, 1994.

[Tra89a] Hoff, William, Ahuja, Narendra. Surfaces from stereo: Integrating feature matching, disparity estimation, and contour detection. IEEE Transactions on Pattern Analysis and Machine Intelligence 11, No.2 pp.121-136, 1989.

[Tra92a] Adelson, Edward H., Wang, John Y.A. Single lens stereo with a plenoptic camera. IEEE Transactions on Pattern Analysis and Machine Intelligence 14, No.2, pp.99-106, 1992.

[Con03a] Scharstein, D., Szeliski, R. High-accuracy stereo depth maps using structured light. Proceedings of the IEEE Computer Society Conference on Computer Vision and Pattern Recognition 1, pp. I/195-I/202, 2003

(a) Upper image

(b) Lower image

(c) Created depth image

Figure 9