FREE VIEWPOINT IMAGE GENERATION SYNCHRONIZED WITH FREE LISNING-POINT AUDIO FOR 3-D REAL SPACE NAVICATION

Norishige Fukushima, Yendo Tomohiro, Toshiaki Fujii and Masayuki Tanimoto

Graduate School of Engineering Nagoya University

ABSTRACT

In this paper, we present a method that generates a free view/listening point image/audio by using a completely synchronized capturing system, named One Hundred Camera and Microphone System. This system can capture audio/image at 100 points with synchronized signal in a three-dimensional scene. This paper introduces an image interpolation method between nearest cameras to generate a free viewpoint image. The step of interpolating image is; calibrating all camera intrinsic / extrinsic parameter, rectifying each nearest stereo image pair, adjusting color balance among all images, stereo matching with occlusion detection, and generate middle image with interpolation along angle with disparity and occlusion map. The experimental results show that we can generate free view point image with free listening point audio. We have succeeded to render this data into 3-D cylindrical display.

Index Terms— large camera array, large microphone array, synchronized capturing system, free view/listening point movie

1. INTRODUCTION

Recently, members of researchers related to 3-dimensional television have been developing. Most of the researches are focused on three dimensional video rendering. Also we have researched the 3-D video field. However generating the input 3-D signal of photo-realistic images is difficult and audio signal is too. Thus we developed a system that can capture the multi-point synchronized audio and video signals by using one-hundred cameras and microphones.

It is too difficult to represent the complete 3-D scene with only raw data when using a huge array system [1]. 3-D reconstruction method of model based rendering is one of the ideas for generating 3-D data. But we request more realistic images. Thus we introduce a method of generating the realistic signal of 3DTV by Image Based Rendering [8].

The main contributions of this research are two points; first, we use ray-based rendering for grass-less 3D TV signal with image based rendering techniques, which include view interpolation / morphing [11], ray-space [5], light field rendering [6] and Lumigraph [7]. Secondly we combine the audio signal which can play free listening point audio. However, we focus on mainly 3-D video part. Free listening technologies are in [10] in more detail.

The key technology of 3-D video generation has 2part; color correction and image interpolation. Our experimental results show that we can render the free viewpoint image with free listening point audio, and then we have succeeded to render this data into 3-D cylindrical display [9](Figure4).

2. 100 CAMERAS AND MICROPHONES ARRAY

Figure1 shows the camera array which is arranged in a circular setup. Figure1 shows the microphone array. This system includes server computer, capturing node, and sync unit, and each object are connected with sync cable. The server and nodes communicate with TCP/IP. Each node can be connected with one camera and more or less four microphones. Figure 3 shows these units.

The image resolution of this camera is 1392x1040, 8 bit/pixel with Bayer matrix filter. We set the frame rate 29.4118 frames per second for audio visual synchronization. Sync generator synchronizes completely these microphones and cameras. The accuracy of synchronization is within 1 micro second.

Figure 1 Camera array

Figure 2 Microphone array

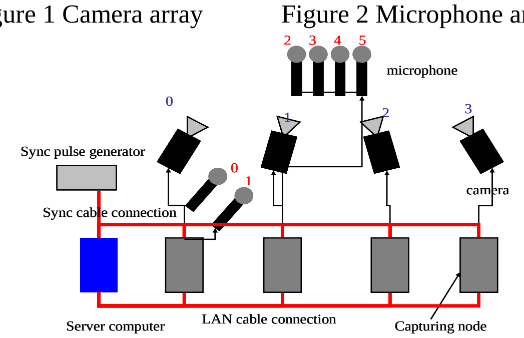

Figure 3 Image of this system

Figure 4 Cylindrical ray-based 3-D display

3. CAMERA ARRAY SETTING

We setup the camera in a circular arrangement accurately to display the ray-based cylindrical 3-D display [9]. This display needs many views on the cylindrical coordinate. However we use camera array, ray information is not sufficient. To generate free view-point image by means of ray-space method, it require view interpolation between nearest cameras.

4. OVERVIEW OF INTERPOLATION

To interpolate image between nearest cameras is as follow;

i. Compute the intrinsic/extrinsic parameters of cameras with camera calibration.
ii. Compute matrixes which rectify images in a parallel manner between nearest cameras.
iii. Compute color correction matrixes with normalized coefficient matching.
iv. Estimate the disparity and occlusion map on the nearest stereo camera with a state of the art method.
v. Interpolate inter image between cameras by using collinear approximation in this circle camera setting.
vi. Perform circle interpolation for direction parameter in the extrinsic camera parameter by using quaternion.

5. CALIBRATION AND RECTIFICATION

To get correspondences between cameras to interpolate middle image easily, we perform camera calibration and rectification. We use Zhang's camera calibration method [2].

Rectification means that changing intrinsic parameters in the nearest camera same and extrinsic parameters of direction is orthogonalized across the line between nearest focal point of camera.

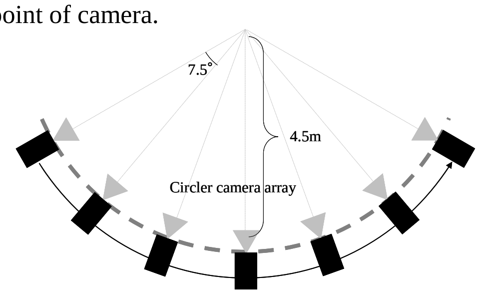

Figure 5 Camera array setup

First, we compute each left and right rotate matrix, R_L, R_R, which makes camera orientation orthogonal across the base line. Second, we set the virtual intrinsic parameter K_v, because all cameras have differential intrinsic parameter; K_L and K_R are left and right intrinsic parameter respectively. Finally, we can obtain the homograph matrix H_R, H_L, and can change image with these matrix.

$$H_R = A_R R_R A_v^{-1},\ H_L = A_L R_L A_v^{-1}$$

Following this step, corresponding point in each image are aligned in parallel x coordinate. This makes the search for correspondence easier. In addition, to interpolate middle camera between paralleled stereo cameras has been already researched by our Ray-Space interpolation method.

To simplify the method, we use linear approximation between camera to camera instead of correct circle coordinate in this paper. Figure 6 shows the cameras setting.

6. COLOR CORRECTION

For the assumption of Lambertian reflection rules, corresponded pixels must have same value. In ideal, we can find the line which is "$y=x$" in a scatter chart of correspondence pixels in the left and right image. However there are some differences of color intensity and temperature in each camera because of the variability characteristics. To relax these differences for states of the art stereo matching method and natural view synthesis, we perform color correction which is suitable for stereo matching.

The basic idea of color correction is to minimize the difference of correspondence pixel intensity among all cameras. Firstly, we perform the stereo matching with normalized-correlation based on a large window matching, and then we get corresponded sequences (you can see in figure 6). This method is robust for color or intensity difference while accuracy of matching in the area of object boundary is weak. Thus we must filer the missed pair of correspondence.

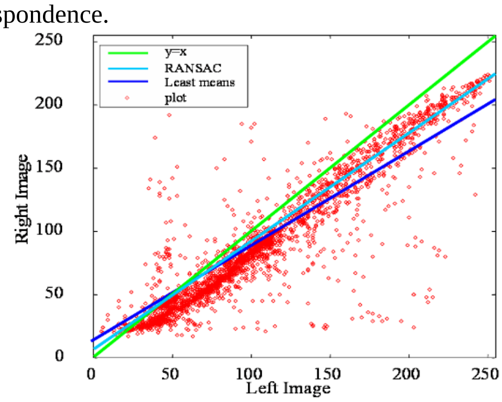

Figure 6 A scatter graph of left and right matches

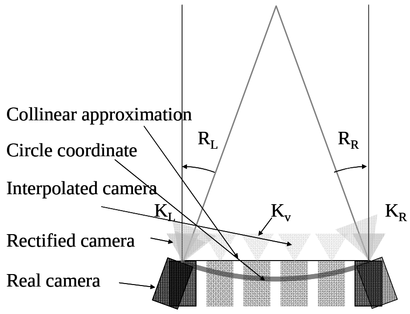

Figure 7 Ratification for virtual image interpolation


Figure 8 Images before color correction


Figure 9 Images after color correction

Second, we plot the correspondence pair, and find line. Because we assume that color difference between cameras has linear characteristics. In this step, we filter the miss-matched pair. We find the line by using RANSAC method, which is robust for outliers, and select the collect pair. Finally, we define the line, we multiply the only single image in the pair of stereo by this linear parameter to cancel difference or make this correspondence line "*y=x*".

We perform above step from left camera to right camera, and right camera to left camera iteratively. As a result, we can reduce the difference among each camera.

7. IMAGE INTERPOLATION BETWEEN NEAREST CAEMRAS

To interpolate middle image in the local stereo pair, we estimate the disparity map and occlusion map on each left and right image. This paper does not deal this computation method with detail. Tanilab However many states of the art methods [3] [4] are researched for computing this maps.

Given these maps, we can interpolate the middle image following the manner of Lambertian reflection and half-occlusion. Figure 10 shows an example of interpolation. In the case of interpolating the middle virtual camera with this scene, we need the information of disparity and occlusion on the left and right image. Figure 10 shows the true disparity map and the occlusion map. In this scene, α shows the position of virtual camera between left and right camera .This parameter must satisfy $0 \le \alpha < 1$.

To generate the middle virtual image, we must compute the disparity map and the occlusion map on virtual image by using left and right these maps. When we generate the middle image which is posited in alpha, we can generate middle disparity map d_{mid} as theses equations. δ is disparity and $d_{left,right}$ is the disparity map at each images.

$$\delta_l = -\alpha\delta, \delta_r = (1-a)\delta$$

$$d_{mid}(x) = \arg\min_{\delta}\{(| d_{left}(x-\delta_l) - d_{right}(x-\delta_r)| + | d_{left}(x-\delta_l) - \delta | + d_{right}(x-\delta_l) - \delta |)/3\}$$

After computer disparity map on the virtual image, we interpolated the image of middle view.

$$I_\alpha(u) = \begin{cases} Occ_L(u+\delta_l, v) = 1, I_r(u+\delta_r). \\ Occ_R(u+\delta_r, v) = 1, I_l(u+\delta_l). \\ else, (1-\alpha)I_r(u+\delta_r) + \alpha I_l(u+\delta_l). \end{cases}$$

Occ is occlusion map and $I_\alpha(u)$ is interpolated image. Basically we interpolate middle image bi-linearly by using corresponded pair of pixels with position weighted blending. In the case of occlusion, if using pixel in the left of right image are masked by occluded, we interpolate middle image by copy by means of either image. Performing these process from that alpha is zero to one, enough view between left and right camera are generated.

8. INTRINSIC / EXTRINSIC CAMERA PARAMETER INTERPOLATION

To preserve the captured image with interpolation, we interpolate the camera parameter of virtual camera too. We interpolate the intrinsic parameter and orientation of extrinsic parameter, while the position of extrinsic parameter is already known. Camera intrinsic parameter *K* is descried as follow.

$$K = \begin{bmatrix} f_u & \gamma & u_0 \\ 0 & f_v & v_0 \\ 0 & 0 & 1 \end{bmatrix}$$

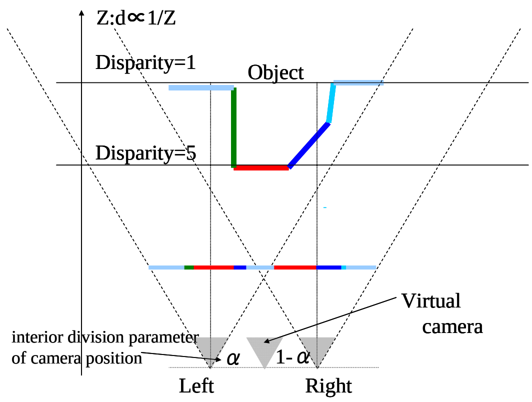

Figure 10 example of interpolation

If the parameter γ is nearly zero, this matrix is orthogonal matrix; the cameras of recent are more or less zero. Thus we can interpolate middle camera intrinsic parameter K_α with both left and right one bi-linear interpolation.

$$K_\alpha = \alpha K_{left} + (1-\alpha) K_{right}$$

α is the position of virtual camera between left and right. Orientation matrix of extrinsic parameters is interpolated by SLERP (spherical linear interpolation) method, in the context of quaternion interpolation. This matrix is not orthogonal, thus we cannot interpolate in the same manner of intrinsic parameter. If we try to interpolate the rotation matrix by splitting the three coordinate; pitch, roll and yaw, it becomes non-natural interpolation. Thus we transform rotation matrix *R* into quaternion *q*, and then we interpolate middle quaternion q_α with SLERP. The quaternion is representing by a real number and three imaginary numbers.

$$R \Rightarrow q = a + bi + cj + dk$$

And SLERP is as follow.

$$q_\alpha = q_r \frac{\sin(\theta(1-\alpha))}{\sin\theta} + q_l \frac{\sin(\theta_t)}{\sin\theta}$$

q_r is quaternion of left camera and q_l is one of right one. α is the interpolation position.
After interpolating middle quaternion, we back transform into the middle rotation matrix R_α, and compute the projective transformation matrix H_α.

$$H_\alpha = K_v R_\alpha K_\alpha^{-1}$$

Finally, we multiply the image by this homography matrix.

9. EXPERIMENTAL RESULT

In our experimental result, we use 17 cameras with figure 5 setup, and perform color correction and view interpolation. Figure 12 is the color correction result and view interpolation result between left and right cameras. Our result of free viewpoint image and audio can be downloaded at the site [10]. And Figure13 is the rendering result of 3-D display.

Figure 12 the result of generated view between nearest camera

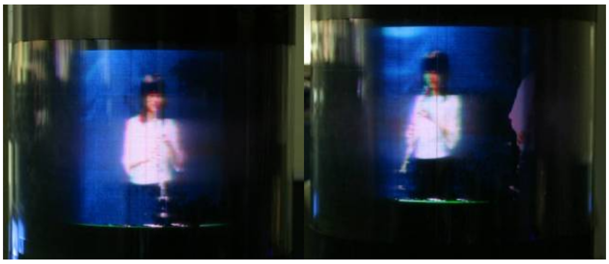

Figure 13 The result of rendering by using cylindrical 3D display

10. CONCLUSION

In this paper, we have presented the method of generating the free viewpoint images. Main contributions of this paper are two point; color correction and view interpolation with smooth. The natural camera parameter interpolation method combining middle view interpolation method with color correction make free view point image sequences smoothly. In the nearly feature work, we render these data into ray re-projective 3-D display with 3-D audio.

ACKNOWLEDGMENT

The authors would like to thank Kenta Niwa, a member of Takeda laboratory of Nagoya University for researching free listening audio part of this experiment. We are also grateful to Collins Computer Co.,Ltd. for designing and developing the system. This work is supported by Nagoya University 21st century COE program (Intelligent Media Integration).

11. REFERENCES

[1] B. Wilburn, *et al.* High Performance Imaging Using Large Camera Arrays. ACM SIGGRAPH 2005, July 2005.
[2] Z. Zhang, A Flexible New Technique for Camera Calibration, IEEE Trans. on PAMI, 22, pp.1330-1334, 2000.
[3] Y. Boykov, *et al*, Fast approximate energy minimization via graph cuts, IEEE TPAMI,23(11): pp. 1222-1239, 2001.
[4] J. Sun, N. Zheng, and H. Shum, Stereo matching using belief propagation, PAMI, 25(7):pp. 787-800, 2003.
[5] T. Fujii, *et al*, Ray space coding for 3D visual communication, Proc. Picture Coding Symp.'96, vol.2, pp. 447-451, Mar 1996.
[6] M. Levoy, P. Hanrahan, Light Field Rendering', Proc. ACM SIGGRAPH '96, pp. 31-42, 1996.
[7] S. J.Gortler, R. Grzeszczuk, R. Szeliski, M. F. Cohen, The Lumigraph, Proc. ACM SIGGRAPH '96, pp. 43-54, 1996.
[8] H. Y. Shum, *et al.* Survey of image-based representations and compression techniques, IEEE Trans.CSVT2003, pp. 1020–1037, 13, 11, 2003.
[9] T. Yendo, *et al*. Seelinder: The Cylindrical Lightfield Display, Proc. ACM SIGGRAPH 2005.
[10] http://www.sp.m.is.nagoya-u.ac.jp/~niwa/fvlp.html
[11] L. Zitnick, *et al*, High-quality video view interpolation using a layered representation ACM Trans. Graphics, vol. 23, no.3, pp. 598-606, 2004.