Multiview Image Correction for Visually Equivalent Light Field 3D Display

Takasuke Nagai[1], Munekazu Date[1], Shinya Shimizu[1], Hideaki Kimata[1]

[1]NTT Media Intelligence Laboratories, Nippon Telegraph and Telephone Corporation, Japan
Keywords: light field display, parallelization, correction

ABSTRACT

The multiview-based light field displays assume that viewpoints of source images are strictly parallel and equally spaced. It is however difficult to arrange multiple cameras by actually satisfying such assumptions. In this paper, we propose a method to virtually parallelize multiple cameras and synthesize regularized light fields.

1 INTRODUCTION

The autostereoscopic display and the light field display can provide stereoscopic images without wearing a dedicated device such as an HMD. Figure 1 shows the visually equivalent light field 3D (VELF 3D) display proposed by Date *et al.* [1-6]. The display consists of a liquid crystal panel and a parallax barrier mounted in front of the backlight. By linearly blending the images of two adjacent viewpoints, the intermediate viewpoints are optically synthesized, and smooth motion parallax is realized with a small number of viewpoint images.

The VELF 3D display assumes that source multiview images are captured with the specific camera arrangement: all cameras are arranged in parallel and equally spaced on a straight line, as shown in the left of Figure 2. It is however difficult to satisfy such assumptions in real; the directions of the cameras are not parallel or cameras are not placed on a straight line at regular intervals as shown in the right of Figure 2. As the results, the quality of the stereoscopic image is degraded due to the deviation of the coordinates of the subject. Furthermore, irregular camera intervals cause view-dependent depth effects and unnatural motion parallax.

In this paper, we propose a light filed synthesis method for VELF 3D display by virtually rearranging multiple cameras so that they are aligned in parallel at equal intervals on a same straight line. We applied the proposed method to a five-camera system and confirmed that the synthesized light field could provide the natural 3D effects.

2 PROPOSED MULTIVIEW IMAGE CORRECTION

We propose a light filed synthesis method based on the multi-camera rectification and virtual camera shifting for VELF 3D display. The proposed method consists of three steps. First, we perform multi-camera calibration in order to obtain intrinsic and extrinsic camera parameters for each camera. Any method can be used; we adopted the Zhang's algorithm in the experiments [7].

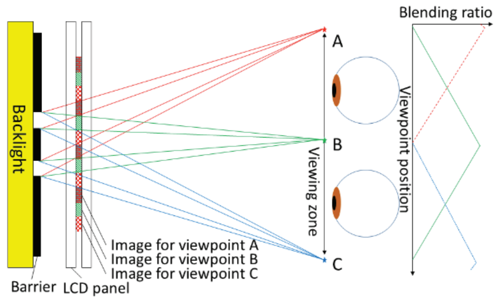

Fig. 1 VELF3D display system

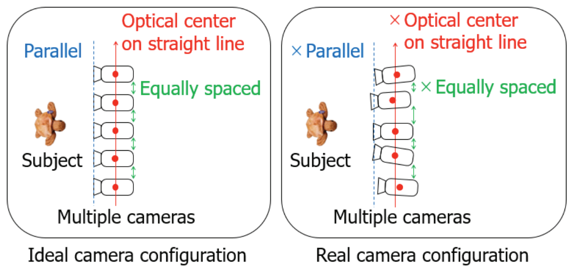

Fig. 2 Ideal and real camera configuration

The second step is to calculate the ideal intrinsic and extrinsic camera parameters, where all cameras are arranged in parallel and equally spaced on a straight line, because actual camera settings rarely satisfy such ideal conditions. We define the ideal intrinsic matrix A^{ideal} as the average of the calibrated intrinsic matrix of each camera.

Under the ideal camera configuration, the displacement vector between two adjacent cameras should be the same; the ideal camera positions can be expressed by two common vectors as follows:

$$C_i^{ideal} = C_1^{ideal} + (i-1)d^{ideal}$$

, where C_i^{ideal} and d^{ideal} expresses ideal camera position of i -th camera $(i = 1, 2, \dots, n)$ and a displacement vector between two adjacent cameras, respectively. Note that n is the number of cameras and the camera indexes are assigned consecutively according to physical layout.

Because an image at an ideal camera position is synthesized from captured image, inclusion of some distortions is inevitable; the amount of distortions depends on the distance between real and ideal camera positions. Therefore, we propose to obtain the ideal camera positions by solving the following minimization problem:

$$minimize \quad max\{D_1, D_2, \dots, D_n\}$$

$$D_i = \left\| C_i^{real} - C_i^{ideal} \right\|_2$$

, where C_i^{real} represents actual translation vector of i-th camera, which is obtained by the first step and $\|v\|_2$ expresses the L^2 norm of vector v. Since there might be no efficient solution on the above minimization problem, we propose to perform a simple search with the initial vectors, C_1' and d' obtained by solving the following minimization problem:

$$minimize \sum_{i=1}^{n} \left\| C_i^{real} - C_i' \right\|_2$$

$$C_i' = C_1' + (i-1)d'$$

In [8], it was proposed that the rotation matrix R' for a multi-camera system is calculated as follows:

$$R' = \begin{bmatrix} x' \\ y' \\ z' \end{bmatrix} = \begin{bmatrix} min \sum_{i=1}^{n} \|x - x_i\|_2 \\ min \sum_{i=1}^{n} \|y - y_i\|_2 \\ min \sum_{i=1}^{n} \|z - z_i\|_2 \end{bmatrix}$$

, where $[x_i, y_i, z_i]^T$ expresses actual rotation camera matrix of i-th camera. Note that the superscript T means the transpose operation.

Although R' represents a good proxy of the rotation matrix of the camera array, the camera baseline is already optimized as d^{ideal}. We, therefore, propose to modify the ideal rotation matrix R^{ideal} as follows:

$$R^{ideal} = \begin{bmatrix} x^{ideal} \\ y^{ideal} \\ z^{ideal} \end{bmatrix} = \begin{bmatrix} \dfrac{d^{ideal}}{\|d^{ideal}\|} \\ z' \times x^{ideal} \\ x^{ideal} \times y^{ideal} \end{bmatrix}$$

, where $\|v\|$ expresses the norm of vector v. The ideal translation vector t_i^{ideal} can be defined as

$$t_i^{ideal} = -R^{ideal}\, C_i^{ideal}.$$

Finally, by using these ideal intrinsic and extrinsic camera parameters, the multiview images are synthesized from the captured images in order to virtually satisfy the ideal camera arrangement. We apply the simple homography-based image transformation. The homography matrix is estimated for each pair of real and ideal cameras as follows:

$$P_i^{ideal} = A^{ideal} \left[R^{ideal} \quad t_i^{ideal} \right]$$

$$m_{i,j}^{ideal} = P_i^{ideal}\, M_j$$

$$m_{i,j}^{ideal} = H_i\, m_{i,j}^{real}$$

, where P_i^{ideal}, M_j, and $m_{i,j}^{real}$ express the perspective transform matrix of the i-th ideal camera, the 3D world coordinate of the j-th reference object, and the 2D image coordinate of the j-th reference object in the i-th real camera's image plane $(j = 1, 2, \dots, m)$. Note that m is the number of the reference objects and we used corners of chessboard as the reference objects.

3 RESULTS

We applied the proposed image correction method to our five-camera system in order to assess the effectiveness. Figure 3 and 4 show the results with different camera arrangements. The first rows are original images captured by the cameras, and the second and third rows are the images corrected by [8] and the proposed method, respectively. In [8], the ideal rotation matrix is assumed to be R' and the ideal baseline distance between adjacent cameras is user defined configuration.

Although the cameras were mounted on the shared bar and the same baseline distances were applied with carefully checking the measure printed on the bar, the deviation of the coordinates of the subjects were visible in (a); the quality of the stereoscopic viewing with such images on the VELF 3D display had damaged. As it can be seen, such deviations were reduced in (b) and (c), and we confirmed that these images provided natural 3D effects including smooth motion parallax.

There were slight differences between (b) and (c), but it was difficult to identify the visual differences even in 2D and 3D. Table 1 and 2 show the squared error of the distance between ideal and real camera position of each camera of (b) and (c) in Figure 3 and 4. Compared to (b), the proposed method could reduce both total and maximum distances from the real camera. Although the differences are not visible due to the small baseline distances of cameras and the relatively shallow depth of scenes, it is obvious that less distance gives less distortion on the homography-based transformation.

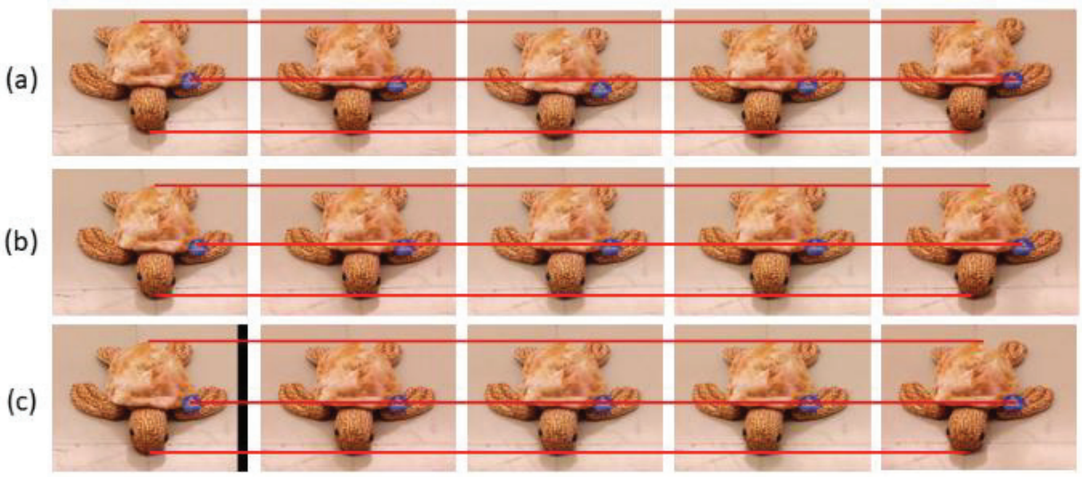

Fig. 3 Comparison of the results (a) original images (b) [8] method (c) proposed method

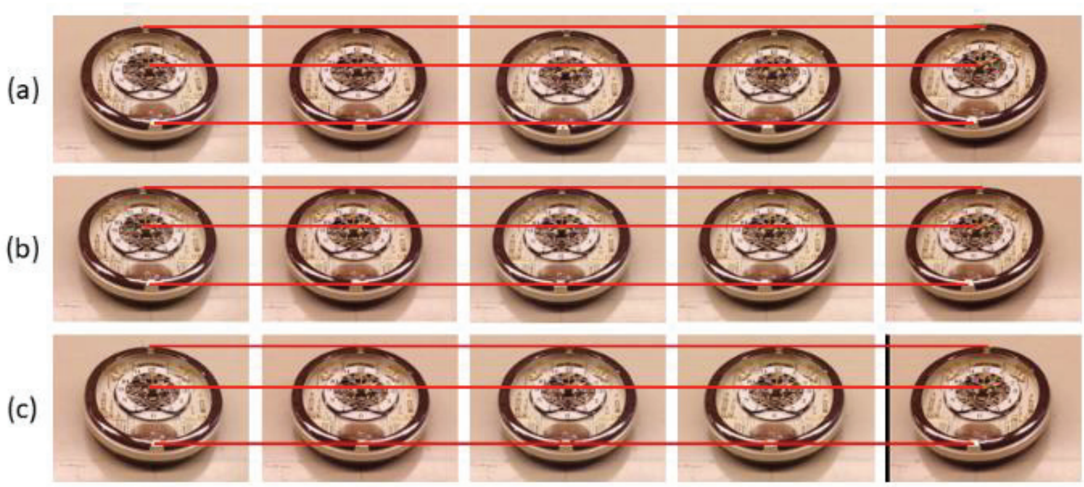

Fig.4 Comparison of the results (a) original images (b) [8] method (c) proposed method

Table 1 Squared error of distance between real and ideal camera position of Figure 3

	(b)	(c)
Camera1	1.28	0.57
Camera2	0.25	0.71
Camera3	4.87	3.80
Camera4	5.03	3.67
Camera5	2.27	0.35
Sum	13.7	9.10

Table 2 Squared error of distance between real and ideal camera position of Figure 4

	(b)	(c)
Camera1	3.60	1.06
Camera2	1.00	1.31
Camera3	1.24	1.44
Camera4	1.71	1.40
Camera5	1.07	0.47
Sum	8.61	5.68

4 CONCLUSIONS

We proposed the novel method to synthesize regularized light fields from miss-aligned multiple cameras system for the multiview-based light field display. The proposed method virtually parallelizes the multiple cameras and places them on a straight line at equally intervals. The experiments showed that the proposed method could reduce not only the maximum squared errors between the real and the ideal camera positions but also the total error. Note that it was difficult to confirm the subjective improvements in these experiments. It might be because the camera baselines were small and scene depth was relatively shallow. Our future works include conducting subjective viewings with different 3D scenes in order to assess subjective improvements.

REFERENCES

[1] M. Date et al., “Accommodation response for

visually equivalent light field 3D display", IEEE Industry Applications Society Annual Meeting, 2017-ILDC-0863 (2017).
[2] M. Date et al., "Flat Panel Visually-equivalent Light Field 3D Display", The Journal of The Institute of Image Electronics Engineers of Japan, 48(2), pp.264-272 (2019) [in Japanese].
[3] M. Date et al., "Full Parallax Visually Equivalent Light Field 3D Display Using RGB Stripe Liquid Crystal Display Panel", 3DSA DIGEST, LFII-4 (2018).
[4] M. Date at al., "Full Parallax Table Top 3D Display Using Visually Equivalent Light Field", Research Demo D31, IEEE VR (March 23-27, 2019, Osaka, Japan).
[5] M. Date et al., "Full Parallax Table Top 3D Display Using Visually Equivalent Light Field", SID Display Week 2019, 56-5, pp.791-794 (2019).
[6] M. Date et al., "Table-Top Photographic-Image 3D Display Using Visually Equivalent Light Field 3D", 3D-conf (July 4-5, 2019, Tsujidou, Japan) [in Japanese].
[7] Z. Zhang, "A Flexible New Technique for Camera Calibration", IEEE Transactions on Pattern Analysis and Machine Intelligence, 22(11), pp.1330-1334 (2000)
[8] J. Yang et al., "A Multi-View Image Rectification algorithm for matrix camera arrangement", Artificial Intelligence Research, 3(1), pp.18-29 (2014)