A STUDY TO REALIZE A BOX-SHAPED 3D DISPLAY: A CALIBRATION METHOD TO ALIGN LENS ARRAY AND DISPLAY

Shunsuke Yoshida[1,2], *Roberto Lopez-Gulliver*[1,2], *Sumio Yano*[1,2], *and Naomi Inoue*[1,2]

[1] NICT Universal Media Research Center
[2] ATR Cognitive Information Science Laboratories
2-2-2 Hikaridai, Seika, Souraku, Kyoto, 619-0288, Japan
{*shun, gulliver, yano.s, na-inoue*}*@nict.go.jp*

ABSTRACT

We propose a handheld box-shaped 3D display as a communication tool for realizing natural human-object interactions in face-to-face communications. The display employs an integral photography technique for each surface. However, misalignment of the lens array and the LCD module occurs, and so adjustment is necessary. In this paper, we analyze the error factors caused by this misalignment and propose a convenient calibration method to minimize them. The method estimates the actual positions of each lens using one camera and several stripe pattern images. After brief manual alignment the method computes the calibration parameters. We verify the suitableness of the proposed method by applying the computed parameters into the image-rendering process.

***Index Terms*—** Three-dimensional displays, Calibration, Computer graphics, User interfaces, Virtual reality

1. INTRODUCTION

Figure 1 illustrates the goal of our handheld box-shaped 3D display [1], which is designed as a communication tool to allow users to directly and naturally share and interact with 3D images to enhance face-to-face collaborative tasks between multiple users. In this context, a glasses-free and viewpoint independent autostereoscopic 3D display are desired. Integral photography (IP) [2] is a technique to satisfy the above requirements; however, the displays are commonly assumed to be only viewed perpendicularly.

For a box-shaped display, the ability to display 3D images with a wide field of view is required because a maximum of three faces of the cube should be observed simultaneously [1]. To fulfill these requirements, we built a lens array with a 120° wide field of view for IP [3]. In the IP technique, each dot in each elemental image has an important role in the 3D image quality. However, alignment between an array of numerous lenses and minute elemental images is difficult. For easy calibration of a small 3D box-shaped display that consists of several faces, we propose a convenient method to adjust this misalignment using a camera and set stripe pattern images.

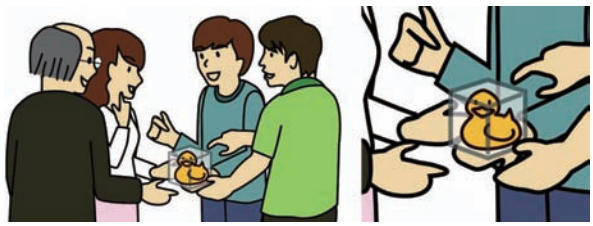


Fig. 1. Box-shaped 3D display

2. BUILDING A BOX-SHAPED 3D DISPLAY

2.1. Related Works

Although extensive research related to box-shaped displays has been reported, they only provide a single-viewpoint perspective 2D image [4]. Others require head-tracking devices and glasses for getting 3D clues [5]. Additionally, these systems are only for a single user.

For natural and collaborative work, our display employs IP as an autostereoscopic method. The original IP uses an array of convex lenses and a printed photograph of elemental images [2]. Recent IP research utilizes digital displays, and several studies have treated IP image quality. Arai [6] quantitatively analyzed the change in image quality caused by the misalignment of several device elements such as being hidden while captured and displaying the processes of elemental images. Lee [7] proposed a method to correct image distortion for lenticula-type autostereoscopic displays. Our proposed method conveniently corrects misalignment between the lens array and the elemental images of the IP technique. We employ a camera and several stripe pattern images to estimate error factors caused by misalignment.

2.2. Configuration of IP Display

For displaying the elemental images, we used a 3.5" color TFT LCD module whose display area is 72×52.56 mm and whose resolution is 640×RGB×480 dots. Horizontal and vertical pixel pitches are $(p_h, p_v) = (0.1125, 0.1095)$.

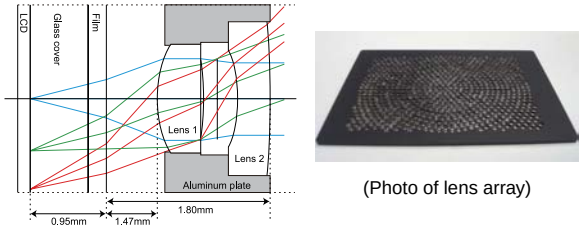


Fig. 2. Prototype of our lens array

Figure 2 shows our proposed lens array for realizing a 120° wide field of view [3]. The lens consists of a two-lens unit: a bi-convex lens (Lens 1) and a convex-concave lens (Lens 2). The lenses of 737 two-lens units are arranged on an aluminum plate in a circular layout. Such lens parameters as effective focal length are optimized for image quality. Effective focal length L_f was specified as 2.273 mm, after considering the thickness of the aluminum plates placed so that each lens focal point lays exactly on the LCD. The diameter of each elemental image is 20 pixels, and the lens diameter is also designed to cover this area. The separations of lenses are fixed to be 2.25 mm using horizontal pixel pitch p_h as the basis for one pixel.

3. CORRECTION OF MISALIGNMENT

3.1. Error Factors

In the IP technique, alignment accuracy between the lens array and the LCD is required in the submillimeter order. However, several alignment errors will occur while arranging them manually, and accurate manual adjustments of our display having six faces are troublesome. Also, adding customized microscopic-precise movement mechanisms is not possible. Hence, we provide a reasonable calibration method to avoid artifact effect in the image caused by misalignment errors.

We assume here that only three error factors occur during the alignment procedure, as shown in Figure 3. Let C_w be the ideal lens coordinate system and C_e be the actual lens coordinate system involving translational error factors (e_x, e_y) and rotational error factor e_θ, whose transformation $\mathcal{F}$ is expressed as follows:

$$\begin{aligned} C_e &= \mathcal{F}^{(e_x,e_y,e_\theta)}(C_w) \\ [x_e\ y_e\ z_e]^T &= R(e_\theta)[x_w\ y_w\ z_w]^T + [e_x\ e_y\ 0]^T. \end{aligned} \quad (1)$$

Here, $R(e_\theta)$ is a 3×3 rotation matrix along the Z-axis. For convenience we call the ideal lens coordinate system the world coordinate system.

3.2. Proposed Calibration Method

For convenience of explanation, assume a pinhole array instead of the lens array and a pinhole camera as a measurement

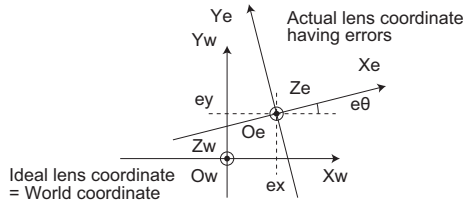


Fig. 3. Relationship of coordinates and error factors

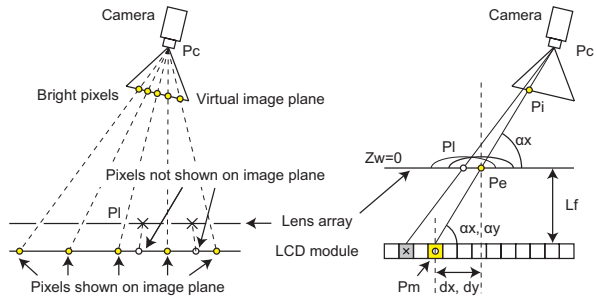


Fig. 4. Pixels observed by camera

device. Let the camera's focal point be $P_w^c = (x_w^c, y_w^c, z_w^c)$ and the pinhole's position (the lens's focal point in practice) be P_w^l as illustrated in Figure 4. Here, a subscript denotes the coordinate system, i.e. w is in the world coordinate system, and a superscript the point's name. In this case, only emitting LCD's pixels on the extended lines of $\overline{P_w^c P_w^l}$ should be brightly recorded on the corresponding location of a camera's image plane. In practice, P_w^l involves positioning errors as described in Section 3.1. The actual lens position might be P_w^e. As a result, the pixels, which should be observed by the camera through P_w^l, would not be recorded on the image plane; instead the pixels observed through P_w^e would be recorded.

If the relationship between the world, the camera, and the image plane coordinate system is known, position P_w^e can be determined. Let $\mathcal{G}$ be the transformation between the world and image plane coordinate systems. Note that $\mathcal{G}$ contains perspective projection. When a certain pixel located at LCD coordinate $P_l^m = (x_l^m, y_l^m)$ lights up and it is detected at P_i^i on the image plane coordinate as a light spot, we can assume a pinhole on an extension of line $\overline{P_w^c P_w^i}$. Here, P_w^i equals $\mathcal{G}^{-1}(P_i^i)$. That is, the observed actual lens position affected by error factors $P_w^e = (x_w^e, y_w^e, z_w^e)$ can be computed as the intersection point of the line and the plane of $Z_w = 0$.

After obtaining n pairs of the ideal lens's position points as ${}_iP_w^l$, where i denotes the ith pair and the observed lens's position as ${}_iP_w^e$ from the captured image by camera, we can then compute the error factors (e_x, e_y, e_θ) using the following procedure. ${}_iP_w^l$ can be transformed to ${}_iP_e^l$ based on Equation 1 by choosing arbitrary values as (e_x, e_y, e_θ). Here ${}_i\epsilon$ is the residual of observed positions ${}_iP_w^e$ and the ideal positions translated by values ${}_iP_e^l$, and a residual sum of squares S is

expressed as follows:

$$S = \sum_{i=1}^{n} {}_i\epsilon^2 = \sum_{i=1}^{n} ({}_iP_w^e - {}_iP_e^l)^2. \quad (2)$$

Here, error factors (e_x, e_y, e_θ) to minimize S could be selected as the best values to express the alignment errors between the lens array and the LCD module. These values are calculated by solving the following equations:

$$\partial S/\partial e_x = \partial S/\partial e_y = \partial S/\partial e_\theta = 0. \quad (3)$$

As described above, we can calculate all lens positions ${}_iP_w^e$ by individually controlling the light emitted for each pixel of the LCD and searching for the corresponding light spot on the image plane. Let $\mathcal{H}$ be the transformation from the LCD coordinate system to the world coordinate system. Then, d_x and d_y as shown in Figure 4 expressed as the following:

$$d_x = x_w^e - \mathcal{H}(x_l^m), \;\; d_y = y_w^e - \mathcal{H}(y_l^m). \quad (4)$$

If α is the angle of the elevation of $P_w^c - P_w^e$, then it can be expressed in terms of effective focal length L_f of the lens as follows:

$$\begin{array}{rclcl} \tan\alpha_x & = & L_f/d_x & = & z_w^c/(x_w^c - x_w^e) \\ \tan\alpha_y & = & L_f/d_y & = & z_w^c/(y_w^c - y_w^e). \end{array} \quad (5)$$

Finally, by solving these equations for x_w^e and y_w^e, we can accordingly obtain the ith lens observed position ${}_iP_w^e = (x_w^e, y_w^e, 0)$. The optimal error factors are computed from Equation 3 with these observed positions.

4. EXPERIMENT

4.1. Experimental Configurations

Relationships between all the coordinates systems are summarized in Figure 5. $\mathcal{F}$ was defined by Equation 1, and $\mathcal{H}$ can be expressed as:

$$C_w = \mathcal{H}(C_l) = [(x_l - D_h/2)p_h \; (D_v/2 - y_l)p_v]^T, \quad (6)$$

where (D_h, D_v) is the LCD's resolution (640, 480) and the origin of C_l is in the LCD's upper-left corner. The origin of C_w is in the center of the LCD.

The lens array and the LCD module are tightly bound using clamps. A brief manual alignment is performed using several markers in the underlying IP image. Each marker resembles a circle of 20 pixels in diameter. These markers are located in the center and at the four corners of the LCD, and their positions are identical as the positions of the specified lenses. The alignment between the lens array and the LCD module was adjusted so that each marker can be observed

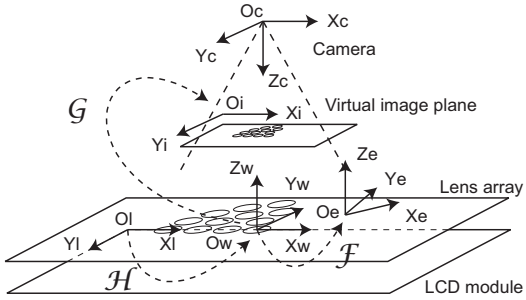


Fig. 5. Relationship of coordinates

from the specified lens. This configuration is fixed throughout the entire experimental procedure.

The camera used in this experiment was a Canon EOS Kiss Digital N. Its image plane size was 22.2×14.8 mm, and its resolution was 3456×2304 dots.

4.2. Identification of Camera Parameters

We obtained the remaining relationship $\mathcal{G}$ between the world and image plane coordinate systems using Tsai's calibration algorithm [8]. To identify reasonable camera parameters, the algorithm requires many pairs of corresponding feature points, which are points on a known flat plane and their recorded pixel coordinates on the image plane.

In our configuration, the lenses are arrayed on the flat plate whose alignment is known. When all LCD dots are lit and the lens array is captured by the camera, many blobs should be detected on the image plane. These blobs are the resultant images of the rays that passed through each lens. Therefore, the blob's centroids P^i are the actual lens positions P^e projected into the image plane coordinates. We can identify the correspondence of blobs and lenses using a coarse relationship of the lens array and camera that are roughly known due to configurations of the camera position and posture. Some other camera parameters such as radial lens distortion are also considered with this algorithm.

4.3. Simplified Sampling Using Strip Pattern Images

As described in Section 3.2, we can calculate the observed lens position by searching for the light spot on the image plane derived from certain emitting pixels of the LCD. However, image acquisitions and examinations for all pixels of the LCD are troublesome. Therefore, we took the following simplified procedure.

First, we chose a number s as the stripe pitch that is bigger than the lens pitch. Next, we prepared vertical and horizontal stripe pattern images with stripe separation s, the width of each stripe is one dot, and each image is shifted one line compared with the image having smaller ID number. For example, in the jth vertical stripe pattern image, the pixels, whose

residue numbers are j when dividing column number x by s, have luminosity. In this condition, two lines are never simultaneously displayed in a particular lens. We defined s as 25 due to lens pitch around 20 dots. Hence, the total number of captured images was 25×2.

The actual lens position involving error factors P_w^e and its projected position in image plane coordinate P_i^e are computed in the experiment in Section 4.2. Here, when investigating the changes of the luminosities of the arbitrary lens in all s images, we can select the ID number j of a certain image in which the lens is the brightest. This number denotes the position of the emitting pixel's column or row, and this position's pixel passed through P^e in practice. Finally, (d_x, d_y) can be obtained by comparing ID number j and the position of P^e in the LCD coordinates using relationship $\mathcal{H}$, because the appearing area of discrete lines called j^{th} columns or row numbers are limited, that would be near P_l^e within a $s/2$-radius vicinity.

4.4. Experimental Results

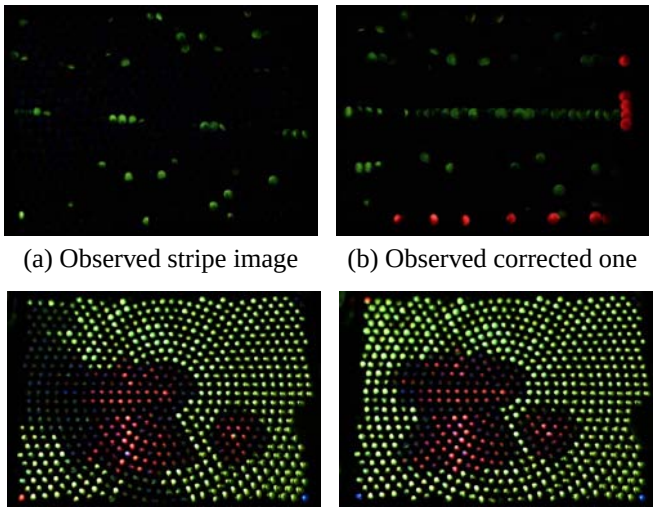
Due to the brief alignment described in Section 4.1, error factors are restricted to certain ranges of $(e_x, e_y) \leq \pm 2$ mm and $e_\theta \leq \pm 3°$. Hence, we numerically minimized Equation 2 in those ranges. As a result, the parameters of $(e_x, e_y) = (-0.64, 0.36)$ mm and $e_\theta = -1.03°$ were obtained using only green-channel stripe pattern images.

Figure 6 shows several calibration results by applying the obtained parameters. (a) is one observed stripe pattern image used in the calibration procedure. By rotation and translation with the obtained parameters on the image, as a result, we can see a horizontal straight line in a middle of the image (b). (c) and (d) are photos of the display shot at a certain viewpoint. Their elemental images were rendered using ideal lens positions P^l for (c) and computed actual lens positions P^e for (d). We can see a dark area caused by misalignment on the left side in (c), but (d) does not have such an area. These results mean the error factors have been removed from the image.

A translational error factor in a perpendicular direction of the lens array would also be considered in some circumstances. In this experiment, we ignored this effect because it is limited within the manufacturing precision of the lens array.

5. CONCLUSION

We proposed a calibration method to remove the error factors caused by misalignment between the lens array and the LCD module in the integral photography technique. The method estimates the actual positions of each lens using one camera and several stripe pattern images. After brief manual alignment and computation of the calibration parameters, we obtained better 3D images as a result.



(a) Observed stripe image (b) Observed corrected one

(c) Before calibration (d) After calibration

Fig. 6. Resultant calibration images

In the future, we will build a small five-face 3D display as a prototype for studying the interaction methods of our display. Additionally, we need to improve the 3D image quality. Especially, the trade-off of 3D image resolution and the number of viewable directions must be considered.

6. REFERENCES

[1] S. Yoshida et al., "A study of a box-shaped 3d display for multi-user environments (in japanese)," in *VRSJ 12th Conf.*, 2007, pp. 2C2–4.

[2] M. G. Lippmann, "Épreuves réversibles donnant la sensation du relief," *J. Phys*, vol. 7, pp. 821–825, 1908.

[3] R. Lopez-Gulliver et al., "Toward an interactive box-shaped 3d display," in *IEEE Sym. on 3D UI*, 2008.

[4] N. Kawakami et al., "Object-oriented displays," in *IEEE Systems Man and Cybernetics Conf.*, 1999, p. 493.

[5] H. Hua et al., "System and interface framework for scape as a collaborative infrastructure," *Presence*, vol. 13, no. 2, pp. 234–250, 2004.

[6] J. Arai et al., "Geometrical effects of positional errors in integral photography," *J. Opt. Soc. Am. A*, vol. 21, no. 6, pp. 951–958, 2004.

[7] Y. G. Lee and J. B. Ra, "Image distortion correction for lenticula misalignment in three-dimensional lenticular displays," *Opt. Eng.*, vol. 45, no. 1, pp. 017007, 2006.

[8] R. Y. Tsai, "An efficient and accurate camera calibration technique for 3d machine vision," in *IEEE Conf. on Computer Vision and Pattern Recognition*, 1986, pp. 364–374.