博士論文 Doctoral Dissertation

Glassless Augmented Reality System Using High-Speed Marker Tracking

(高速マーカートラッキングによる裸眼拡張現実システム)

三河 祐梨

Glassless Augmented Reality System Using High-Speed Marker Tracking

Abstract

Augmented reality (AR) display is a novel technology that superimposes or produces digital three-dimensional (3D) images on real world, which may rewrite our experience of reality and facilitate our intuitive understanding of presented information. The latest research has mainly contributed to spatial augmented reality (SAR) using projection mapping and aerial displays to produce 3D images, where numerous optical techniques, tracking, and human visual mechanisms have been actively investigated. AR displays have been developed to be so matured to present projected planar images in wide area, rewriting the material appearance of object surface, and applied for advanced fabrication.

However, even the latest AR displays still have some difficulties to be solved for further immersive experience; among them, this thesis treats two main problems of wide presentation area and perceptual quality in stereoscopic vision. For the wide-area presentation, it is difficult to track the projected target object in wide area using conventional tracking markers whose design is complicated and therefore difficult to be tracked particularly with out-of-focus or low-resolution images. Also, the aerial image presentation in wide area is limited to only the usage in closed spaces like cinema as a huge screen is needed in the conventional methods.

For the visual perception in stereoscopic displays, the conventional displays using binocular parallax have not reproduced the real stereoscopic image faithfully as the eye information like variable focus and viewpoint is ignored. Recently, there is an emerging technology where the eye is continuously tracked and its information like position and gaze direction is reflected to the display presentation, which called gaze-contingent displays. However, it has lacked the speed in commercial device so far and the effect of gaze-contingent displays has not been examined precisely.

This thesis proposes several high-speed AR technologies related to the above difficulties. For the widely AR image presentation, two tracking methods are introduced: circumferential markers for ball tracking and multi-color LED markers in the case that the power usage is allowed. The concept of wide aerial image presentation system using distributed displays of binocular parallax is also proposed and two optics are validated for this concept. For high-speed gaze-contingent displays, the ocular parallax effect is examined using high-speed rendering system with a novel method of fast a stable eye's viewpoint tracking. Also, novel identification algorithm is proposed to distinguish the infrared light's reflection points in the corneal reflection method for eye tracking. Overall, this thesis uses stationary system to solve these problems as it enables high computational cost and high-speed processing necessary for immersive experience using SAR as well as enabling almost glassless viewing.

Careful simulations or actual device implementations, or both, have been conducted for all the above proposals. Evaluation experiments have shown the validity of the proposed method and examined stereoscopic perception while several demonstrations of useful applications, especially for sports, have been introduced. Each research in this thesis has some common points each other, such as the purpose of wide presentation, the usage of binocular parallax, tracking markers in wide area, and eye tracking. Thus, this thesis entirely contributes to solve several difficulties of latest AR research while being related to each other, which will lead immersive experience by glassless viewing in wide area.

Contents

Chapter 1

Introduction

1.1 Towards digitalized world with augmented reality

Augmented reality (AR) is one of the latest information technologies, which can superimpose digital information on the real world or reproduce three-dimensional (3D) images in mid-air. Since this is a new form of image presentation, AR has a potential to make a significant contribution to visual understandings. For example, 3D images floating in mid-air enable stereoscopic viewing not limited to text or flat images, thereby facilitating viewer to intuitive understanding. Furthermore, projecting a different texture onto objects represents another texture, in which a reality is likely to be redrawn and then the user experience can be changed. Also, a sticky projection onto an object with movies enables precise instructions in dynamic scenes like where to touch on the sports balls, as well as evoking a sense of wonder. Therefore, it can be said that AR will be a key technology for the future world development.

In recent years, AR can now do a variety of things in research fields, such as high-speed and wide-area projection [1, 2] and material representation [3, 4, 5], and AR in which the surface of object itself emits color freely [6, 7, 8]. AR is not only becoming quite technologically mature, but is also contributing to the diversification of social culture through digitalization. In fact, various AR products of glasses [9, 10] and stationary displays [11, 12] have been distributed, and their use has been increasing especially in industry, medical surgery, sports training, and entertainment.

However, AR cannot be freely presented under any circumstances even with state-of-the-art technologies. Among the numerous issues, this thesis focuses on two main difficulties: the wide presentation area of AR and the visual perception quality of stereoscopic vision.

1.2 Technical problems of the latest augmented reality

As shown in Fig. 1.1, AR systems can be broadly classified into two types: a stationary type, in which the display is placed in the environment, and a portable type, that is, the wearable displays on the head such as goggles or small portable devices like smartphones. The latter type is especially preferred in commercial use [9, 10], but in general, the smaller the computer, the lower the processing power and the rendering quality. Also, it is heavy and burdensome to wear due to the various components from sensors to optical elements mounted on the headset. On the other hand, the former stationary type does not have these problems, but it is difficult to extend the presentation range apart from the stationary device. In recent years, a system between the both stationary and portable types, where only the optical elements are mounted on glasses and the presentation device is placed in the environment to reduce the headset weight, has been proposed [15], but the presented image is limited to be small.

Projection mapping and aerial imagery are generally called spatial augmented reality (SAR) because their images pop up in space, and most stationary AR system presents these two. For wide-area projection mapping, the author proposed in her master's thesis a spatial scanning system

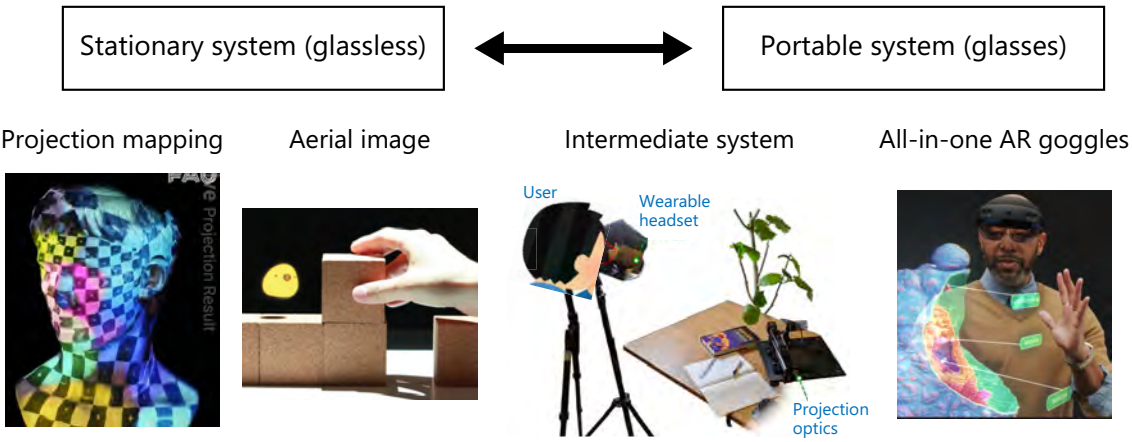

Fig. 1.1. Several types of displays for spatial augmented reality (SAR) with example researches: from left, seamless projection mapping [13], floating aerial images with interaction [14], an intermediate system of stationary projection optics and lightweight wearable devices [15] (edited image), and AR goggles that mounts every sensors and display optics [9].

using two-axis galvanometer mirrors called Saccade Mirror to project images in wide area [16, 2, 17]. On the other hand, projection mapping requires tracking of the target object to know where to project; the wider the area, the more blurring and the lower resolution occur in the camera image, which makes difficult to detect the object posture. Tracking markers are preferred to be used in the high-speed tracking in general, but the complicated design of conventional tracking markers such as AR markers is not suitable for such situations of blur and low resolution [18].

Aerial image displays can be broadly classified into binocular parallax and real image formation [19]. The latter method produces a faithful 3D image and has a variety of optics, such as light field displays [20], holographic displays [21], and volumetric displays [22]. However, in most of these systems, the farther away from the device, the lower the image resolution, and the narrower the viewing angle. One exception is the laser-plasma scanning way [23, 24], but it is dangerous for the user interaction in general situations. By contrast, the former method not only eliminates these problems but also simplifies setup by just presenting to both eyes an image with a slight shift in viewpoint position (binocular parallax image), and many 3D displays like head-mounted displays (HMDs) and 3D cinema use this method. However, the conventional parallax method requires a huge screen to be affixed to a wall to present a wide-area aerial image, limiting the locations and situations where it can be used, such as indoor cinemas and entertainment facilities. Such a system to present aerial images outdoor in wide field is expected to be developed.

1.3 Visual perception quality in stereoscopic vision

As mentioned in the previous section, many aerial image displays employ the stereoscopic vision using binocular parallax, which tricks the brain into seeing stereoscopic images rather than faithfully forming real images. However, this causes different problems in stereoscopic perception. As shown in Fig. 1.2 (a), in the stereoscopic vision using binocular parallax, the depth of focus (accommodation) is on the display surface while the eyes vergence angle directs the protruding position of the stereoscopic image. Therefore, those depth is different each other, causing a vergence-accommodation conflict. Also, the viewpoint position is slightly moved by eye movements like saccade but the small parallax, called ocular parallax, caused by this movement is ignored in most binocular parallax system. Therefore, as shown in Fig. 1.2 (b), the larger the disparity, the lower the quality of stereoscopic perception, such as depth of perception and binocular

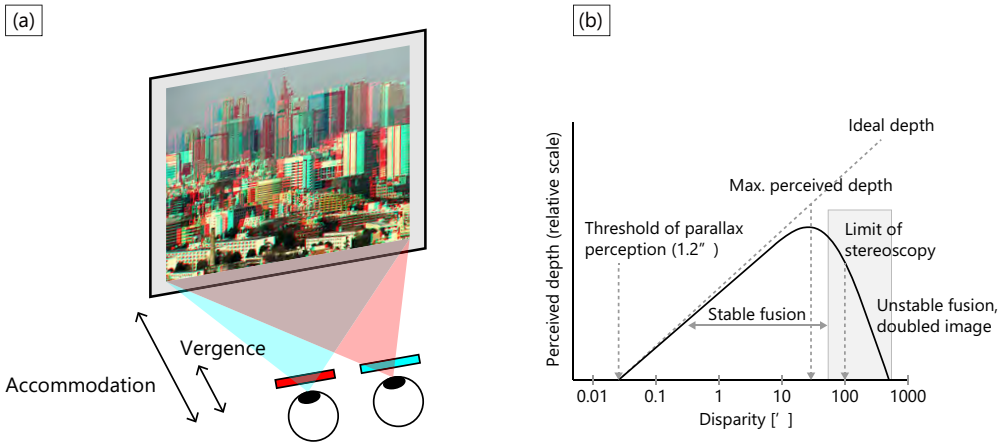

Fig. 1.2. Issues in stereoscopic vision using binocular parallax. (a) The example image of binocular parallax using anaglyph. (b) Relationship among disparity, perceived depth, and stereoscopic limits, cited by Fig. 10 in p.9 [25].

fusion vision.

From these facts above, it can be said that stereoscopic displays using binocular parallax are strongly related to eye information. Thus, research on gaze-contingent displays, where eye information obtained by eye tracking is reflected to the display, has been active in recent years. Gaze-contingent displays contribute to not only visual perception but also system performance in both virtual reality (VR) and AR. For example, varifocal displays adjust the depth of the formed image using a variable focal lens to reduce the vergence-accommodation conflict to improve stereoscopic vision [26, 27, 28]. Foveated displays can reduce computational cost using gaze information by rendering only the gazing point of the eye at high resolution and the peripheral field of view at low resolution [29].

This thesis focuses on ocular parallax since it is easy to be integrated into not only AR and but also VR systems, but its effect has not been examined in detail due to slow rendering speed in conventional commercial displays. Ocular parallax is, as described above, always produced by involuntarily eye rotation like saccade since the viewpoint of eye is apart from the rotational center of eye. Though this is a small parallax compared to binocular parallax and motion parallax, there have been already some reports in 20th century to imply the positive effect of ocular parallax on stereoscopic vision [30, 31, 32]. With the huge development of VR/AR systems, currently ocular parallax is easy to be integrated into a system by just attaching a small eye tracker to head-mounted displays (HMDs); actually, such systems have been already distributed in variety of commercial products [33, 34, 35]. However, the commercial products originally have a large system latency, which may have impaired the accurate verification of ocular parallax effect, as having reported in the previous research [36].

1.4 Purpose and position of this paper

The purpose of this thesis is immersive and glassless viewing of AR in wide space, which is the most natural way of experiencing 3D images floating or superimposing on the real world. Thereby, this thesis proposes several novel methods to solve the two main problem of current AR explained above: the wide presentation area of SAR (aerial displays and projection mapping) and the improvement of stereoscopic vision using binocular parallax. The overview of research introduced in this thesis is shown in Fig. 1.3 with specific chapters where each research is written. Every research in this thesis includes the development in the entire system to be fast and precise

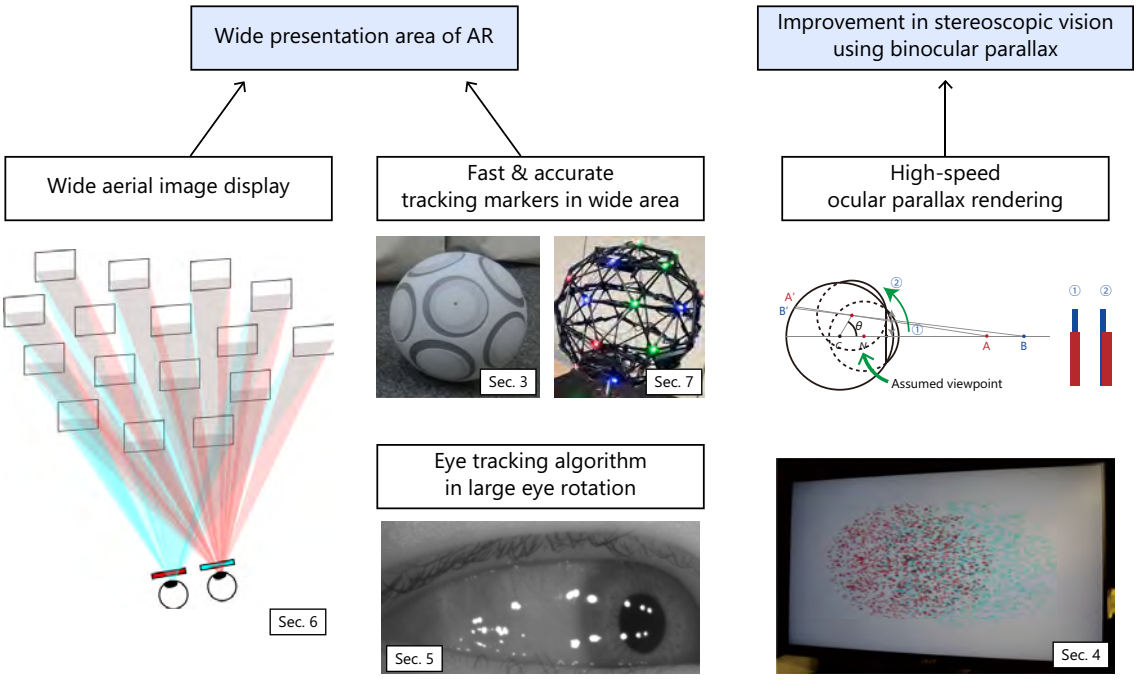

Fig. 1.3. The overview of this thesis.

measurement, recognition algorithm, and image presentation, as well as deep understanding of visual perception. By developing such a system, high computational cost is allowed to present richer images than current technologies can, which facilitates further development of rich AR and examination of visual perception. Furthermore, the novel point of stationary system is glassless viewing; users do not need to wear heavy and fuzzy equipment to experience immersive and rich AR presentation.

The author has been working on AR systems from her master's thesis to this doctoral thesis; the overview of her involvement in the specific AR issues is shown in Fig. 1.4.

As for the tracking markers in wide area, there are three proposals from this thesis [37, 38, 39] as well as a novel identification algorithm for corneal reflection method [40]. The wide-area projection mapping system [2] was proposed by her master thesis while the wide-area aerial image display is in this thesis. The investigation in stereoscopic visual perception is firstly taken in this thesis, but it also has a common point with these wide presentation optics in originally assembled high-speed system. The author's master thesis has also tackled with precise fabrication method using photochromic chemical phenomenon, which changes the color of object surface itself, and can emit color naturally and can be utilized for the digitalized daily fashion [41]. Those topics are challenging that has not yet been studied in AR research so far as well as difficulties in assembling high-speed systems, and this thesis provides a quantum leap forward for these topics.

1.5 Paper constitution

The structure of this paper, which was also shown in Table 1.1 with a novelty of each chapter, is described in detail in this section.

Chapter 1 describes the prospects of AR system and the technical components required for faithful AR presentation, as well as the need for thought in its application. The novelty of each described research, the relationship between them, and the contributions to AR research including the author's master thesis are presented.

Chapter 2 describes the elemental technologies related to each research presented in this paper,

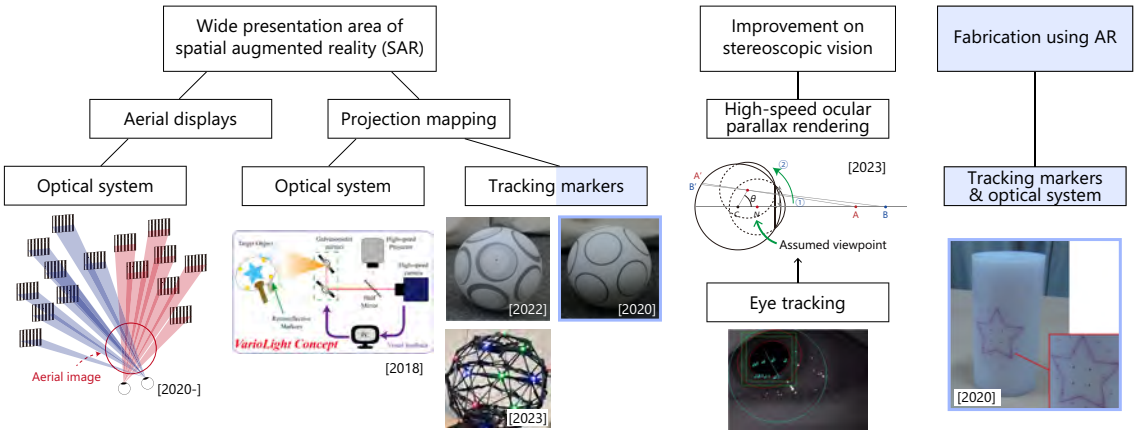

Fig. 1.4. The position of this and author's master thesis among the academic field of augmented reality. The brakets indicate the published year, and the principal author is always the author of this thesis though eliminated within this image. The work in the author's master thesis is indicated by blue area and lines.

Chapter	Topic	Novelty
Chapter 3	Wide projection mapping system & Circumferential markers	system, marker design, algorithm, application
Chapter 4	Investigation of the effect of high-speed ocular parallax rendering	system, algorithm, visual perception
Chapter 5	Model-based identification of corneal reflections	system, marker design, algorithm
Chapter 6	Aerial imaging display in wide area	system, application
Chapter 7	Multi-color LED marker for tracking in wide space	marker design, algorithm

Table 1.1. Overview of each section in this thesis.

ranging from tracking markers to SAR systems (projection mapping, aerial display, etc), as well as human visual perception and practical applications, primarily in sports.

Chapter 3 describes the design method and tracking algorithm of coded circumferential markers, called biased circumferential marker (BCM) for fast tracking of absolute posture and projection on the target sphere, as well as their quantitative evaluation. The uniform circumferential marker (UCM), which has no code on the marker, and the wide-area optical axis control system are the subject of my master's thesis [17]; however, their explanation is also included in this thesis for a comparison, and a comparative evaluation between various markers is performed.

Chapter 4 describes a study investigating how stereoscopic vision and depth perception are changed by high-speed binocular parallax presentation according to the eye's principal point. A reasonable approximation of eye's principal point is introduced to achieve both high speed and high accuracy, and its validity is evaluated through simulations. Then, an original system to track eyes and presents binocular parallax images at high speed and accurately is created from scratch. Using this system, a user study experiment was conducted to examine the stereoscopic perception, and the results was evaluated by statistical analysis.

Chapter 5 describes a model-based identification of multiple corneal reflections (CRs). In the eye tracking method using CRs, the gaze detection accuracy increases with a large number of observed CRs [42], but some of them are easily kicked and lost according to eye rotation. A

newly proposed method identifies a large number of CRs based on simplified eye model [43], and a user study evaluated that it is estimated to be sufficient especially for the participants whose corneal shape is similar to the assumed eye model.

Chapter 6 presents wide-area aerial image display using distributed displays. This system enables outdoor usage like a park, station, tunnels, and stadium, since the display placement becomes flexible by distributing each display compared to the conventional method of a huge screen. This concept was validated in two ways; one is laser scanning system, which was built, and one-point aerial image was evaluated, and the other is multiple monitors of parallax barrier, which was validated in simulation.

Chapter 7 proposes a method of multi-color LED tracking markers for an object moving in a wide area. While the purpose of this research is like that of Chapter 3, this method can be applied for any objects including spheres by using sparse marker placement under the allowance of the power usage. The LED marker itself is strong against image blur and low resolution because of stable light emission as well as circumferential markers and enables tracking in wide range. This method was evaluated at various distance of a color camera, which showed sufficient performance.

Chapter 8 presents the conclusions of this thesis.

Chapter 2

Related Works

Note that some sections reused the author's master thesis [17] and previous published papers and proceedings during the doctoral period [44, 45, 38].

2.1 History of augmented reality

There have been many studies on augmented reality (AR) systems for a long time, dating back to the 1960s [17]. A head-mounted display (HMD) was developed by Ivan Sutherland, who is credited as the founder of virtual reality (VR) and AR. The device was equipped with a head-tracking device and an optical transmissive display, where two cathode-ray tube (CRT) binocular parallax images were superimposed on the real world via a half-mirror. The device was heavy and had to be suspended from the ceiling, making it difficult for users to move around in the real world while wearing the device.

HMD-based AR had made little progress by the 1990s because of the enormous computational costs involved in modeling, matching, and accurately presenting the three-dimensional (3D) computer graphics (CG). In the 1970s and 1980s, however, intuitive interaction and user-controlled processes were demonstrated and studied using presentations that merged CG objects onto real images. In 1992, a system was developed in an airplane manufacturing plant that provided work support by overlaying the manufacturing process on the real world using a transmissive HMD. Since then, measurement and presentation technologies for AR displays have been developed, including spatial recognition [46], recognition of uneven shapes [47], and the development of portable AR displays [48]. In the 2010s, a goggle-type device called HoloLens by Microsoft [9] was developed to recognize spatial information and present aerial information. It also utilizes Simultaneous Localization and Mapping (SLAM) technology to simultaneously estimate self-position and map the environment [49], enabling sharing among multiple people. In addition, in the 2020s, smartphone AR games have been developed for private use [50].

Since the 2010s, AR presentation technology using stationary devices instead of an individual device like glasses has also been studied, and has been widely used in theater, museum exhibitions, and stage performances [51, 52]. Table-top aerial image presentation [14, 53, 54] and projection mapping [55, 1] have been studied extensively, where with the development of optical technology for measurement and presentation, it is easy to satisfy the three types of AR consistency (spatial, temporal, and optical consistency) [56].

2.2 Object tracking in augmented reality

Augmented reality (AR) usually requires the target tracking to accurately superimpose the virtual image onto the target object. The marker-based methods are mostly used because of its accuracy, speed, and stability rather than markerless methods [61, 62]. This section describes several ways of marker-based methods as shown in Fig. 2.1.

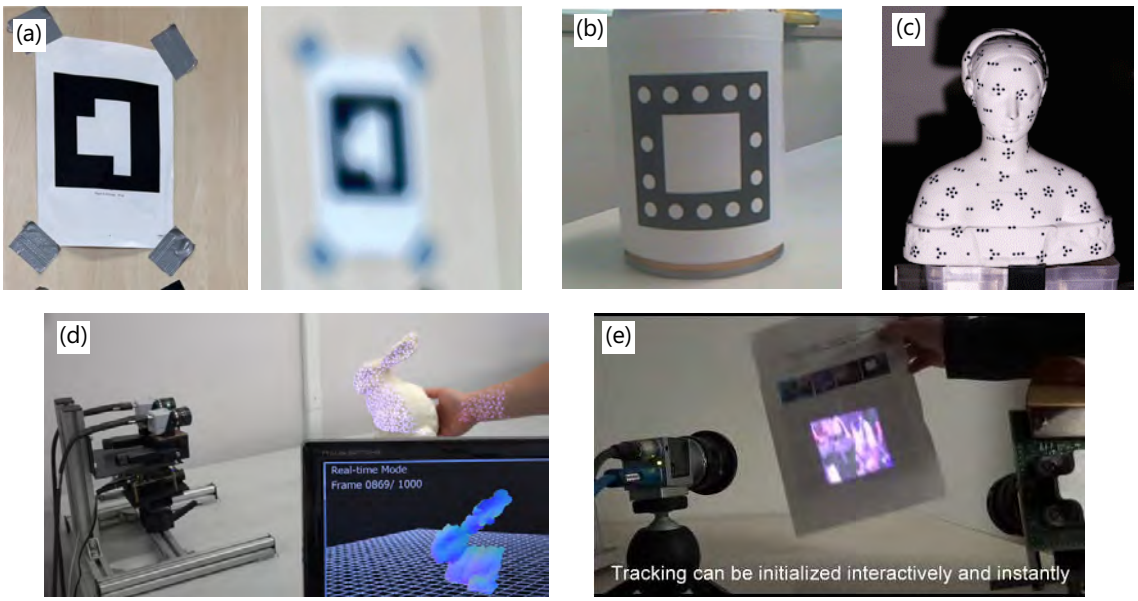

Fig. 2.1. Conventional methods of marker-based and markerless tracking. (a) An AR marker [18] and its blurred appearance. (b) AR markers on a cylinder [57]. (c) Extended dot cluster markers for the fast and accurate 3D object tracking [58]. (d) (e) Markerless tracking of a rigid object and a plane using pattern projection [59, 60].

2.2.1 Tracking in augmented reality

Tracking methods are mostly divided into marker-based and markerless methods. There are various markers such as AR markers and dot markers, and the latter, dot markers, are suitable for high-speed tracking because of its easy-to-detect shape. As for the markerless method, it is further divided into with or without lights projection; then, some of the light-projecting ways use a pattern projection while the others use uniform projection. For the high-speed and accurate tracking, the methods with light projection are preferable.

Note that most tracking methods uses the Perspective-n-Points problem [63] to solve the object posture assuming that the 3D coordinate of each marker or the object shape is already known. Furthermore, high-speed tracking uses the self-window method [64], which reduces the searched area for acceleration.

2.2.2 Marker-based tracking

Numerous markers have been developed to read the detailed information for various necessities in daily life [65, 66]. Typical examples are one-dimensional barcodes that are used to read product information and two-dimensional (2D) QR codes that are used to read local information with a handheld camera [67]. In particular, QR codes can store a large amount of character information, such as a URL, within a complicated 2D code surrounded by a quadrangular frame.

As the firstly proposed marker in the field of AR, AR marker [18], shown in Fig. 2.1 (a), has been popular for real-time tracking of the object posture and easy identification; however, the complicated shape within a rectangle leads low-latency such as 60 fps, which is not suitable for the high-speed tracking. Furthermore, in general, such a quadratic AR marker is restricted only to tracking of a flat board or a cylinder [57, 68]. AR marker is originally weak against image blurring when the camera moves out of the depth of field, resulting in out-of-focus, as shown in

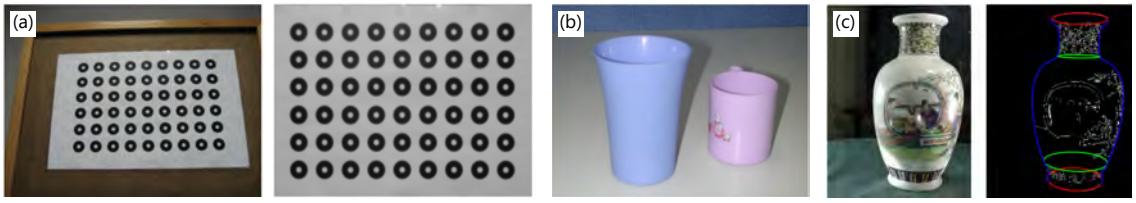

Fig. 2.2. Examples of calibration using circles. (a) Ring pattern for camera calibration [77]. (b) (c) Calibration methods from circles found in daily life, such as cups [78] and vases [79].

Fig. 2.1 (a). Moreover, this marker requires high-cost processing to detect a corner precisely, which exceeds the allowable presentation delay [69].

By contrast, high-speed marker-based tracking methods mostly use dot-based markers as the most detectable and easy to calculate markers; just calculating the center of gravity. However, it is difficult to obtain the unique geometric information of a 3D object from the dot arrangement alone. It usually necessitates manual initialization [70] despite of four-marker placement [2].

Focusing on that problem, a cluster marker consisting of a set of dots was proposed for expressing unique code, where automatic starting of the posture estimation was enabled. There have been two types of dot cluster markers; one is deformable dot cluster marker (DDCM) for non-rigid surfaces, such as paper and clothes [55] and the other is extended dot cluster marker (EDCM) for the rigid 3D object [58], as shown in Fig. 2.1 (b). However, these cluster markers are weak against low resolution and image blurring because small dots are gathered in a narrow area. There is a necessity to propose a novel marker to overcome such problems while enabling decoding, and this paper presents two novel methods.

2.2.3 Markerless high-speed tracking

Markerless methods have been preferred as its ideal style that the object posture can be known without pasting any markers. However, this way originally cannot track the rotationally symmetric objects such as a sphere and a cylinder on which this study focuses on, and it sometimes requires high resolution [71].

The majority of markerless methods projects patterns by light. A shape estimation method at 1,000 fps using pattern projection, known as the structured lighting method (Fig. 2.1 (c)), was proposed [59], but it exhibited the problem of coarse measurement due to discrete projected patterns. It also restricts the object shape to be uneven, which could not be applied to rotationally symmetric objects such as a sphere. Such problems have also appeared in the conventional commercial products of 3D sensing [72]. A high-speed tracking method on a flat object utilizing the projected texture was realized [60, 73, 74] (Fig. 2.1 (d)), but it also required high-resolution and unblurred images, and was still restricted to flat objects only. A radar-based tracking was previously reported [75], but because it uses time derivative, it cannot track the absolute posture and is not suitable for image presentation.

As for the method with uniform light projection, a tracking method based on contour point clouds was proposed [76], where the object shape was restricted to uneven ones and its speed is only about video frame rate because of its high computational cost.

As a conclusion, currently, marker-based methods are prior than markerless methods both in accuracy and speed and especially for a sphere, despite of hassle of attaching markers. This paper also considers various marker-based methods.

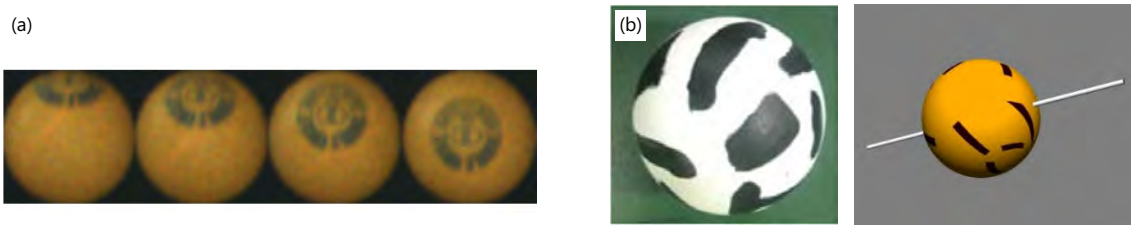

Fig. 2.3. Examples of sphere tracking. (a) The machine-learning-based method referring the original markers on the table tennis ball [98]. (b) Printed patterns on the surface of table tennis, which enables the rotation estimation with relatively large processing time [99, 100].

2.2.4 Image processing for circular shapes

This paper focuses on the circular shape for tracking markers, so this section describes several previous research that used circles for camera calibration and object tracking.

Camera calibration using circular shape has been considered robust against image blurring and low resolution, which is helpful for the similar consideration of circular-shaped markers of this study. In general, camera calibration commonly uses square-based patterns, but a method that uses multiple circles and rings has become popular for providing robust detection against blurring [77, 80, 81, 82], as shown in Fig. 2.2 (a). A circle has the feature whereby it is observed as an ellipse form in the perspective projection of the camera geometry [83], and the ellipse approximation is useful for enhancing the calibration accuracy. Moreover, focusing on the cross-ratio, which is an invariant component in perspective projection, calibration methods using concentric circles [83, 84, 85, 86], multiple circles [87, 88], and a circle and center point [89] have also been proposed. There have been methods that utilized concentric circles [90, 91] that can be found in daily life for calibration, such as cups [78] and vases [79] (Fig. 2.2 (b),(c)).

Conventional methods for ellipse detection that are related to our study are also described here. Because an ellipse has five parameters, it is difficult to improve its accuracy, particularly when the contours result in substantial image noise or when occlusion occurs. To deal with such problems, various methods such as iterative and linear solvers have been proposed [92, 93, 94, 95]. Applications of such ellipse detection methods have been proposed for pupils [96] and coins [97]. However, because these applications are singular or independent of one another, and sometimes include no prior information, the detection accuracy is likely to be low. Thus, we propose a [17] for using multiple known circles, which results in high recognition accuracy.

A circumferential marker is proposed previously by the corresponding author and introduced in my master thesis [17, 37]. It was used for the sphere tracking, where 12 circles were put over the entire surface of a sphere uniformly. Because of these aspects, it is called uniform circumferential markers, but its problem is that only the relative posture is obtained because of its uniformity; this paper investigates the absolute posture tracking using circumferential markers by embedding codes on them.

2.2.5 Tracking of spherical shape

In relation to the proposal of tracking markers for spheres in Chapter 3, this section will explain conventional methods of sphere tracking. A deep-learning-based method was proposed to estimate posture using the marks originally printed on the table tennis ball [98], as shown in Fig. 2.3 (a), which takes only 2.6 ms but is not adaptable for the dynamic situations including unexpected occlusion. A method estimating rotation speed and axis was also proposed by attaching a pattern to the table tennis ball, as shown in Fig. 2.3 (b), which determines the pixel-by-pixel brightness

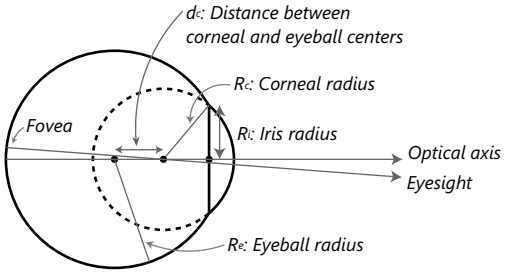

Fig. 2.4. Simplified eye model proposed by Le Grand [43]

difference between consecutive frames. But this previous study reported this process took 30-50 ms [99]. The similar research using a similar printed pattern also reported that the process from measurement to image presentation was within 1 s [100]. These methods are unsuitable for the high-speed and low-latency sphere projection mapping that is the goal of this thesis.

2.3 Eyeball tracking

2.3.1 Eye model

The eye is inherently complex in shape. Typical models include the Gullstrand No. 1 [101] and Gullstrand-Emsley [102]. The anterior surface of the cornea is nearly spherical in the center but flat in the periphery; therefore, the overall surface is considered aspheric [103, 104, 105]. Grand et al. replaced this complex model with a simple model consisting of only two spheres of the eyeball and cornea [43], as shown in Figure 2.4. Here, the overall corneal surface is a part of an ideal sphere, whose radius is the same as the curvature radius of the anterior corneal surface. As introduced in the following section, most model-based eye tracking algorithms assume this simple eye model.

Concerning the notable point of gaze estimation, eyesight is the line passing through the fovea and the eye's center, as shown in Figure 2.4. Eyesight is slightly different from the eye's optical axis passing through the center of the eye and corneal nodal point. This is because the central fovea of the retina, which has the highest resolution of acuity, is located slightly off the intersect point of the optical axis and retina. Ordinal gaze estimation requires calibration of such a deviation, and many methods have addressed this[106, 107].

2.3.2 Eye tracking

Bright/dark pupil method is a main method for eye tracking that irradiates infrared light to the eye utilizing retroreflectivity of pupil to extract pupil area, as shown in Fig. 2.5.

Model-based eye tracking methods mainly focus on two major features: corneal reflections (CRs) [40, 109] and an elliptic pupil [110, 111, 112, 113]. Note that this thesis focuses on CRs. Each feature has a different advantage. The methods using the ellipse approximation of a pupil cover a wide-range eye rotation but impair the accuracy [110]. By contrast, the methods based on CRs have a high estimation accuracy but are not strong against large eye rotation. Lai et al. combined these methods [114] using two IR rings around the neck of each stereo camera. For the front-facing eye, there are no missing CRs, and the eyesight can be estimated from both CRs and an elliptic pupil. However, for side-facing eyes, the estimation was conducted only for the elliptic pupil. This method has the advantage of easy setup and calibration, but in large eye rotations, only the elliptic method is used, and its drawbacks remain in terms of accuracy.

Various methods have been proposed for CR identification. An approach of switching infrared

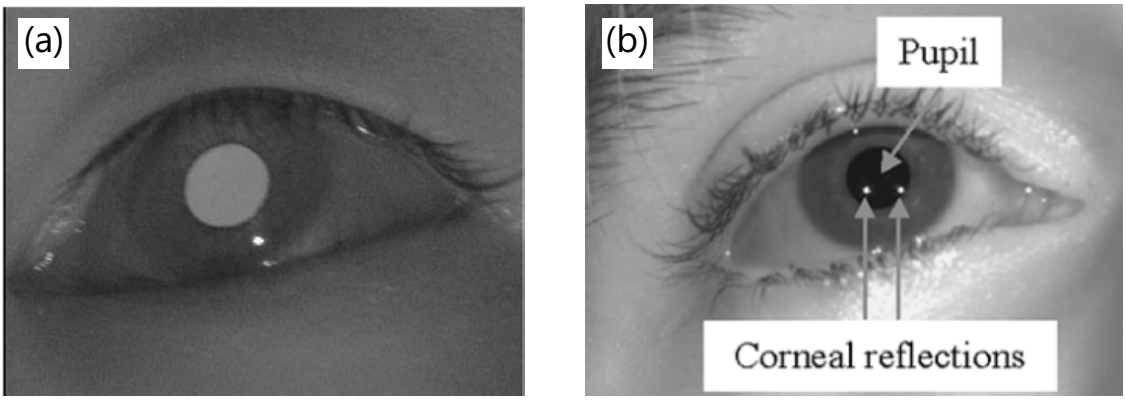

Fig. 2.5. Two-types of conventional eye tracking method methods; (a) bright pupil method [108] and (b) dark pupil method [40]. For the latter, the corneal reflections appear.

lights to make unique codes has been investigated [115, 116, 117]. However, it must synchronize the light switching and frame updating of the camera. Stoffregen et al. proposed a method for reaching 1000 fps using an event camera [118]. However, as a fundamental problem of the switching approach, it cannot use all light reflections per frame, and it impairs the accuracy in gaze estimation [42].

In methods using static infrared lights, the identification has a limitation in the eye rotational range and the number of CRs. For example, an approximated formula for the simple eye model can identify 9 CRs but is limited to only small angle of eye rotation ($\pm 10^{\circ}$) [119]. The pattern-matching approach [120] assumes that all true CRs appear near the pupil, limiting the eye rotational range. The homography matrix approach [121] does not consider the case in which true CRs are missed. The approach using CRs of corner-like shapes [122] realizes easy identification but uses few lights such that it does not correspond to wide eye rotation. Zihan et al. specialized in noise reduction in eye images [123]. However, the same problem persisted.

By contrast, with the development of methods based on convolutional neural network (CNN) in recent years [124, 125], Chugh et al. addressed the problem of kicked CRs [126] and achieved a high accuracy ($>$90%). However, using a CNN requires manually labeling each image [127]. This approach is robust against spurious reflections, but most of them are outside cornea and seem to be excluded using simple image processing.

This thesis targets to identify and fully utilize the multiple (¿10) CRs even if some of them are missing.

For PCCR-based eye tracking methods, the calibration of the entire system, including cameras and lights, is required. Previous research utilized a reference area such as a display monitor around which multiple infrared lights were attached [128, 129] or detected directly the area with high LED brightness [109]. By contrast, this study provides a method for calibrating a wearable eye-tracking device using a board to which attached both infrared lights and ChArUco board [130].

2.4 Spatial augmented reality

AR with high spatiality, such as aerial imaging display and projection mapping, is collectively called spatial augmented reality (SAR), and this section provides an overview of previous research on this topic.

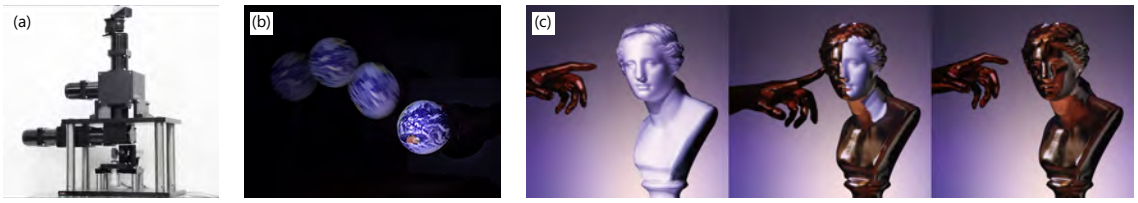

Fig. 2.6. Overview of high-speed dynamic projection mapping (DPM). (a) Saccade mirror for widely projection mapping [16]. (b) Projection on a widely moving sphere using saccade mirror [1]. (c) Material representation using dynamic projection mapping [5].

2.4.1 Dynamic projection mapping

Dynamic projection mapping has made a great progress in recent years. One key is the DynaFlash, a projector capable of projecting at 1000 fps with a minimum delay of 3 ms [131, 132, 133]. Using this, the projection mapping in response to object's dynamic motion is possible; without it, projection mapping is limited to shapes that do not change in appearance with the posture change, such as a sphere [1] and also limited to slow motion [76].

The wide-area projection mapping could be supported by the high-speed optical-axis control system [16] (Fig. 2.6 (a)), which is also used in this thesis. This technology was initially used for high-speed object tracking and videography but was then utilized for projection mapping to a table tennis ball by coaxing in a common 30 fps projector [134]. Furthermore, by introducing retroreflective material into the background, high-precision sphere tracking was also achieved [1] (Fig. 2.6 (b)).

By combining such a high-speed optical-axis control system with a high-speed projector, the author's master's thesis realized sticky dynamic projection mapping onto a moving plane such as a uchiwa, whose apparent shape changes due to posture change [2]. For the accurate posture tracking, retroreflective markers were used.

Research on material representation using projection mapping has also been developed, where the method of measuring its normal variation and projection mapping accordingly was proposed [5] (Fig. 2.6 (c)). As the other way for the material representation, projecting white light synchronously with multiple material display that rotates at high speed was also proposed [4, 3, 135]. Facial projection mapping [136] also used material representation like makeup was performed [137, 138].

Omnidirectional and/or shadowless projection mapping have also been proposed by projecting from multiple directions [139, 140, 13, 141].

2.4.2 Spatial Representation Using Spheres

A spherical display for SAR is described since this thesis treats various spherical displays. Unlike conventional flat displays, spherical displays have an advantage to be observed from all directions by multiple people. Therefore, its applications include the telexistence display [142, 143, 145] (Fig. 2.7 (a),(b)) and the widely floating display that is mounted on a drone [144] (Fig. 2.7 (c)). Projection mapping is adaptable for such a spherical display because no electrical equipment and power supply are required for the ball itself. This is effective especially for dynamic interactions such as catching balls [146].

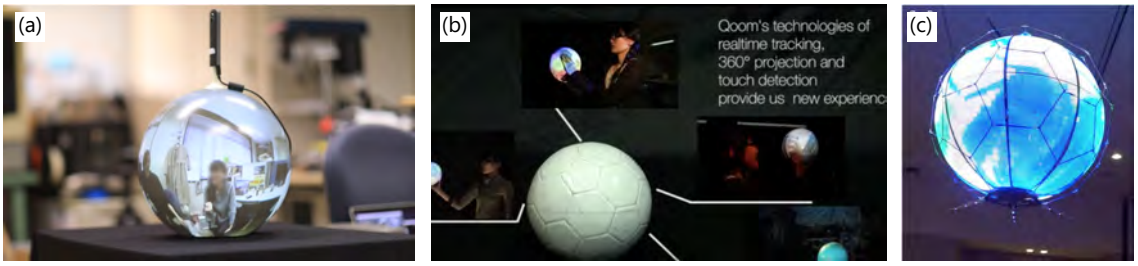

Fig. 2.7. Several examples for spherical displays; (a) A spherical display for telexistance [142]. (b) Projection-type spherical display with interactivity [143]. (c) Drone-mounted floating spherical display [144].

Fig. 2.8. Several examples for aerial displays; (a) Distant aerial image presentation using parallax barrier [147]. (b) Stereoscopic display using fly-eye lens [11]. (c) Dynamic parallax barrier named "Dynallax" [148].

2.4.3 Aerial imaging display

Aerial image displays using stationary devices can be broadly classified into two main types: those that faithfully produce 3D images using special optical elements and those that present binocular parallax on a flat display for stereoscopic viewing.

According to 3d-display-survey, the former, faithful production of 3D images, includes

- spatial scanning of two-dimensional images by a galvanometer scanner (spatial scanning method) [149, 22]
- the basic optical elements such as convex lenses and concave mirrors
- light ray reproduction methods such as using fly's eye lenses [150, 20]
- holographic displays [21]
- laser-plasma way [23, 24]
- imaging methods using a two-sided corner reflector array [151].

However, when considering wide-area aerial imagery, many of these methods are unsuitable due to the low resolution when presenting farther away from the device.

The latter binocular parallax displays can be further divided into glasses-type and glassless-type (naked-eye-type) displays. The glasses-type includes anaglyph and polarization systems for separating images for each eye on the glasses using filters, and shutter systems that synchronize with the frame update of the display. By contrast, the naked-eye type uses lenticular lenses, fly-eye lenses [11] (Fig. 2.8 (b)), and parallax barriers [152] into the display. Binocular parallax displays are originally suitable for distant presentation [147] (Fig. 2.8 (a)) and is currently used in 3D cinemas where 3D glasses are distributed to visitors [51].

Barriers and lenses used in the naked eye system have originally a limited spectroscopic point (hot spot), which is fixed with static barriers and lens systems. In recent years, dynamic parallax

barrier, called dynallax, has proposed to dynamically move the stereoscopic point by controlling the barrier portion using a liquid crystal panel [153, 148] (Fig. 2.8 (c)). Even so, crosstalk, which is a mixture of both parallax images and does not realize stereoscopy, is originally inevitable [152]. Therefore, research was also proposed to increase the number of viewpoints to four to make the hot spot a large area using a time-division display using a 240 fps monitor [154].

2.5 Stereoscopic vision and depth perception

A wide variety of cues are available for stereoscopic perception [155]. Non-visual cues are oculomotor ones, such as vergence and accommodation. Visual cues include binocular cues, such as retinal disparity of binocular parallax, and monocular cues, such as motion parallax [156, 157, 158], pictorial cues (occlusion, lighting, and size differences) [159], and ocular parallax where the eye's viewpoint continuously moves due to involuntary eye's rotation such as saccade, and therefore produces a parallax.

To improve stereoscopic perception, it is important to consider what stereoscopic image is shown and how to evaluate the perception. The conventional stereoscopic images include Gabor patch [160] and stripe bars [157], and among them, RDS, where binocular correspondence makes a 3D perception, is most frequently used [161]. Stereoscopic vision using binocular parallax images are typically presented either using glasses (anaglyph, polarization, and active shutter) or without glasses (lenticular, parallax barrier). The latter inherently has a problem of crosstalk [152], while the former can completely separate the lights. Our study takes the anaglyph approach as it is the easiest way: compared with other approaches, this approach only needs presenting red and cyan images and no special optical element.

There are several ways to examine the response of a perceived depth [155]: in addition to verbal numerical responses [160], using calipers [156], reaching out [162], and blind walking [163] (a method of triangulated blind walking is also discussed [164] since walking straight ahead is thought to be not versatile). Furthermore, there is another approach, using which we can estimate the time taken to reach the expected position [165]. Our study uses the verbal numerical response approach to avoid the effects of body sensory intervention on perception.

2.6 Gaze-contingent displays

Gaze-contingent displays are the displays that utilize eye information to improve stereoscopic viewing and to enhance VR/AR experience. For example, varifocal displays [27, 26] (Fig. 2.9 (b)) and multi-focal displays adjust the focus position. Foveated displays are famous for making high resolution in the fixated area and low resolution in the other peripheral field [29] (Fig. 2.9 (a)). Ocular parallax displays, on which this thesis also focuses, reproduces a parallax which is produced by involuntary eye rotation such as saccade [36, 166, 32] (Fig. 2.9 (c)).

Note that there have been few studies on the delay of gaze-contingent displays, that is, the delay from the eye measurement to visual presentation, while numerous studies about the latency in the interaction between users and visual information have been conducted [69, 167, 168, 169]. While it was shown that a latency of 50–70 ms is acceptable for the entire system [170], the other study showed that the effect of foveated display appears more in low-latency and fast system [171], which might be applied for ocular parallax, too. The discussion of eye tracking stability and acceptable delay, also discussed in this paper, was made for the pupil-swim problem [172].

2.7 Advanced sports with augmented reality

The explained SAR can be applied for the motion visualization of sports objects, including spheres, which convey quantitative motion information to athletes, which is expected to provide

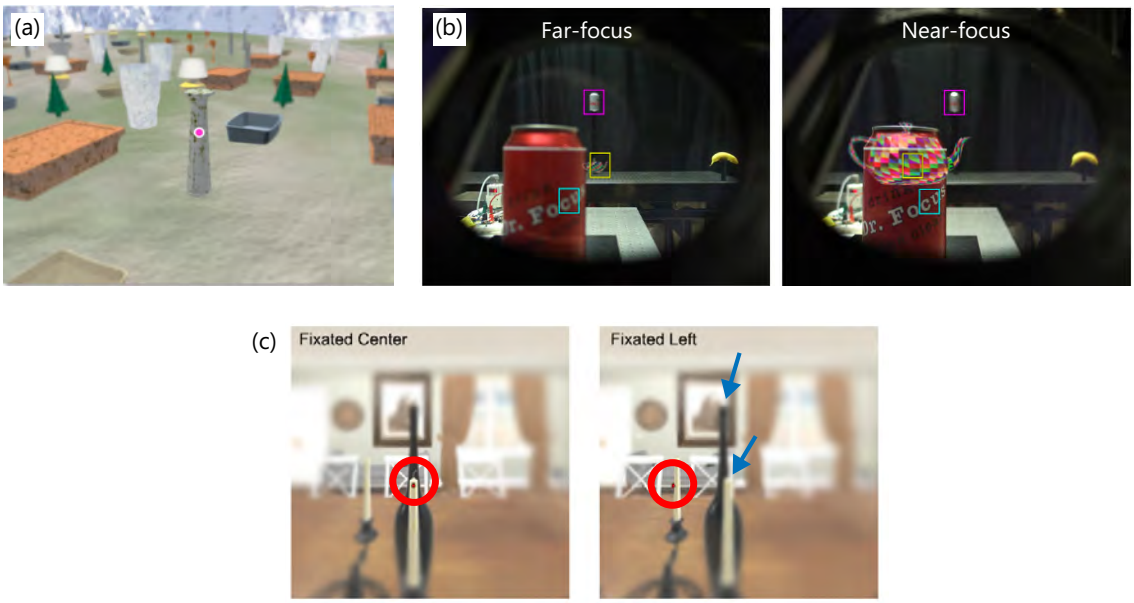

Fig. 2.9. Several examples for gaze-contingent displays; (a) Foveated display [29]. (b) Varifocal display [27]. (c) Ocular-parallax display [36].

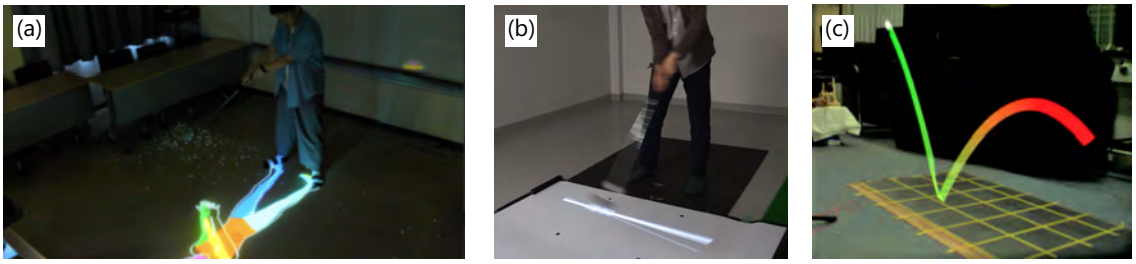

Fig. 2.10. Several examples for advanced sports using SAR like (a) (b) visualization of golf swing [173, 174] and (c) visualization of estimated ball orientation [175].

effective real-time feedback to athletes. Off-the-shelf devices that embed sensors in baseball balls and acquire their movement information remotely and asynchronously with tablets are available at the consumer level [176]. However, SAR can present such information on and around the target object in real time; for example, the predicted trajectory of a ball in AR glass [175] (Fig. 2.10 (c)) and golf swing orientation [173, 174] (Fig. 2.10 (a),(b)). The author's proposal about circumferential markers also enabled such SAR motion visualization to correspond to the absolute rotational posture of a sphere, which has previously been difficult owing to the lack of robustness of conventional tracking markers [37].

Many new e-sports where many people regardless of their physical ability have been investigated recently. An advanced ball sports using a drone-mounted spherical cage was proposed [177], where the cage moves autonomously in the wide field and players catch and throw it, and each player wearing AR glasses experiences AR around the cage. This thesis treats tracking method of a drone cage to realize such an e-sports.

Chapter 3

Uniform/Biased Circumferential Markers For Dynamic Sphere Projection Mapping

This section introduces uniform/biased circumference markers (UCMs/BCMs) as the suitable tracking markers for dynamic projection mapping onto a ball moving in a wide space, whose description reused the author's published paper [38]. Both types of circumferential markers are comprehensively discussed including the UCM proposed in my master's thesis [37, 17]. I would like to note that the first person is "we" because this is a joint work with co-authors.

3.1 Introduction

Spatial augmented reality (SAR) is an expanding research field that enables numerous people to feel the reality of a graphics presentation by projection mapping onto the object surface, thereby rewriting the texture of the real object as if by magic. SAR is generally based on real-time tracking of the target object. Markerless methods such as structured lighting [59, 60] have been attracting increasing attention in recent years, whereas marker-based tracking methods, represented by augmented reality (AR) markers [18], previously supported the development of AR.

However, emerging optical technology for SAR hardware that can realize sticking projection mapping corresponding to wide-area and dynamic movement, called Dynamic Projection Mapping (DPM) [131, 16, 2], requires a fast and highly precise marker-based tracking of three-dimensional (3D) objects with high robustness against blurring and occlusion. This technology is expected to have extensive applications in sports training and stage performances, such as juggling and dancing.

Markers that have been proposed for SAR are represented by AR markers [18] and dot-based markers [55, 58]. These markers have complicated shapes consisting of corners and irregularities in a narrow space to embed codes, which necessitates imaging at a high resolution with no blurring. In widely tracking, both low resolution and blurring inevitably occur when the object moves out of the focus field of the camera as it moves in the depth direction, which makes it difficult to measure the markers of complicated shapes accurately. Moreover, owing to their discrete placement, these markers are vulnerable to random large objects (such as bodies) and complicated occlusions (such as hands and fingers), which are unsuitable for dynamic scenes relating to human movement, such as the sports balls and juggling bats that are assumed in our study. Therefore, a novel marker design is required to solve such problems.

With reference to conventional research on camera calibration, numerous studies have attempted to deal with the problem of blurring and low resolution. These works focused on circles with smooth curves as blurring-resistant marker design, such as a grid pattern of circles

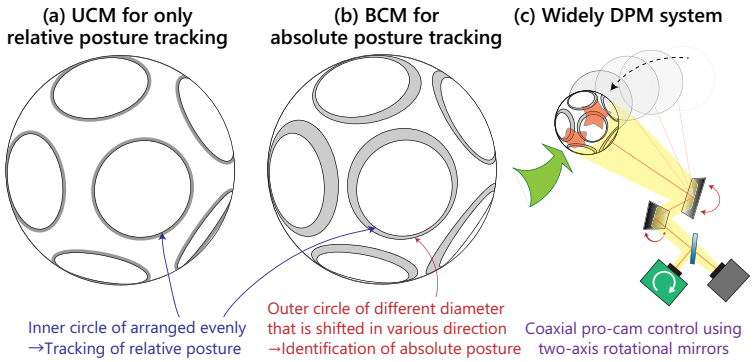

Fig. 3.1. Schematic of uniform/biased circumferential marker (UCM/BCM) and widely dynamic projection mapping (DPM) system. Cited by Fig. 1 in [38].

and rings [80, 81, 82, 77] and concentric circles [84, 83, 85, 86]. The circle shape can be maintained effectively even with blurring and low resolution, and it is possible to measure it with approximately the same degree of accuracy by image processing. Furthermore, in principle, a conical section can be observed as that of different parameters in the perspective projection of camera geometry; especially, a circle is observed as an ellipse. Such conic geometric parameters have been used for effective image-based recognition [83].

In this research, we consider the positive use of the circle-like shape for the design of the tracking marker. As a first step, we consider the marker design of a sphere because its cross-sections are always circles, and subsequently, the design can easily be conducted. A sphere is an object that is used extensively in dynamic situations such as sports balls and stage performances, and therefore, the development of a novel marker design for spheres may offer various applications in SAR.

First, we propose a uniform circumferential marker (UCM) for tracking the relative posture of the sphere, as illustrated in Fig. 3.1 (a). The marker is robust against blurring and low resolution, and it is also strong against random occlusion owing to a continuous line segment. The distance between the point group forming the circumference and the ellipse in the perspective projection is defined, and posture estimation is conducted by iterative optimization. This arrangement is evaluated as exhibiting almost the same high posture estimation accuracy as that of conventional dot markers [37]. However, it does not contain any code information, which results in only relative posture estimation.

Subsequently, focusing on the fact that one circumferential marker has two inner and outer circles, we propose a biased circumferential marker (BCM) that embeds unique code into itself by slightly shifting the outer circle from the inner circle, as illustrated in Fig. 3.1 (b). This design realizes easily identifiable code embedding while retaining the advantages of the circular shape described above. Moreover, a novel accurate and fast decoding method using the accurately estimated relative posture is developed. The proper marker design is discussed in detail in the following section.

We also propose a rough initialization for both markers by using an elliptic shape of the circumference that represents geometric information. We use a widely dynamic projection mapping (DPM) system by means of a coaxial optical system, as depicted in Fig. 3.1 (c). The DPM system can enhance the performance of the proposed UCM/BCM, and therefore, real-time visualization of rotation on the target surface of a randomly moving object can be realized for the first time.

A summary of the contributions of this study is presented in Table 3.1. The four contributing points are as follows: (1) dynamic and low-latency DPM, (2) wide-range motion in the depth direction, (3) interaction with humans, and (4) rotation visualization, which conventional methods have not achieved simultaneously.

Table 3.1. Summary of contributions of this study with section numbers of detailed explanations. The symbols have the following meanings: ◯: fully possible, △ moderately possible: ×: impossible.

Purpose / Motivation	Technical problem	Existing methods		Proposal	
		Dots	EDCM	UCM	BCM
(1) Dynamic and low-latency DPM	Low computational cost (high-speed)	◯	◯	◯	◯
(2) Wide-range motion in the depth direction	Out-of-focus (blur) / Low resolution	◯	×	◯	
(3) Interaction with humans	Automatic initialization / Overall occlusion	×	◯	◯	◯
	Partial occlusion	△	△	◯	◯
(4) Rotation visualization	Absolute posture estimation	×	◯	×	◯
	Estimation accuracy of the relative posture	◯	◯	△	△

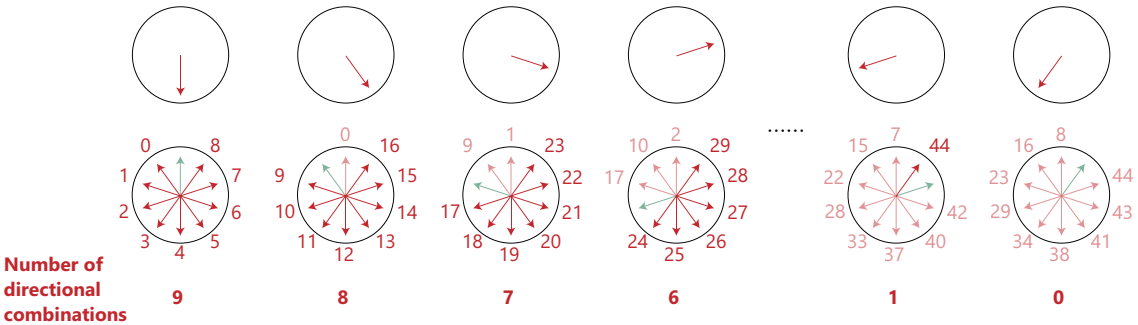

Fig. 3.2. List of directional combinations made by different outer circular axis directions of two markers; there are 45 directional combinations for 10 marker directions of two markers, indicated in the form of the opaque red direction. The semi-transparent green direction indicates a point symmetrical one, which is excluded for the combinations. The semi-transparent red direction indicates overlapping already shown before itself. Cited by Fig. 2 in [38].

3.2 Proposed markers

The purpose of this study is to design a tracking marker that enables tracking of the sphere posture with high speed and accuracy. To design a marker, we focus on the circle-like shape that is robust to blurring, which supports a wide range of movements in the depth direction. There is a characteristic whereby a circle is observed as an ellipse in the perspective projection of camera geometry, whose geometric characteristics, such as rotation and crush, indicate the posture cues leading to precise recognition. It can also be robust against random occlusion because it has many measurement points along the contour line and partial missing is permitted. Note that we assume that only one camera is used, because we think about to use the widely projection mapping system described in Fig. 3.1 (c) and Section 3.4, and the marker arrangement is already known.

We did think about putting multiple circles on the surface of the sphere as tracking markers for the sphere's posture because in principle, two different postures can be inferred from only one circle with a three-dimensional posture (leaning in the direction of roll or pitch), and thus, two or more circles whose arrangements are already known are required. Because there is a trade-off between the number and size of circles that can be placed on the spherical surface, the circles should be moderately large and should not be numerous in order to obtain a large number

of measurement points for accurate posture estimation of each circle markers. Moreover, the markers should be arranged isotropically for relative posture tracking of a sphere, whereby our initialization method is easy to conduct, as described in Section 3.3.2.

3.2.1 Overview of proposed marker design

We propose the UCM as a basic form of the sphere posture estimation, as depicted in Fig. 3.1 (a). We considered regular polyhedrons having 12 vertices to be preferable for the isotropic marker arrangement, and assuming inscription in the sphere, there are 12 circles with the same diameter, which are evenly distributed on the sphere surface. The circumferential shape is used for the marker, not the shape of the circle because it is easier to remove the outliers, which are caused by shielding or being located at the edge of the sphere, from the inner circle than the outer one by image processing thereby, the inner circle are used for sphere posture estimation. Each circle marker has the width as depicted by the gray area in Fig. 3.1 (a). The inner/outer parts of the circumference are called inner/outer circles, respectively.

However, the rotationally symmetric form of these markers only leads to a relative posture estimation of a sphere, which can only be used for the measurement of rotational speed and acceleration. For the absolute posture estimation of a sphere, it is necessary to embed unique code into each circumferential marker. As described in Section 2.2.2, the conventional marker consisting of complicated shapes exhibits the problem of being vulnerable to low resolution and blurring, which limits the movement range in the depth direction [55, 58]. Therefore, we consider the positive use of the smooth shape of the original circles that is resistant to blurring and low resolution [77, 37] to create the unique code, which has not been attempted to date. The original arrangement of the inner circles needs to be maintained as far as possible (for example, no breaking off of the line) for the stable estimation of the relative posture, and therefore, the coding expression should be carried out with the outer circles.

Here, we focus on a circle placed on a sphere. For such circles, a straight line passing through the center of the circle and perpendicular to the circle plane (called a circular axis in this paper) also passes through the center of the sphere. Note that outer/inner circular axis indicates the circular axis of the inner/outer circle, respectively. We thus consider changing the relationship of outer circular axes of adjacent markers to express the unique coding. Further, we propose a unique coding method of tilting the outer circular axis from the inner one slightly in various directions and combining the two-directional information of the adjacent outer circular axes. We consider various possibilities for expressing the coding without degrading the advantages of UCM, and simply tilting the outer circular axis is the best solution. This marker design is called the biased circumferential marker (BCM), and it is depicted in Fig. 3.1 (b). In the following section, we discuss the design of the unique coding by the outer circle shifting and recognition method.

3.2.2 Coding Design for Estimating Absolute Posture

The direction in which the outer circular axis is tilted relative to the inner one, which is called a marker direction here, indicates the direction of the marker itself. The direction of a single marker alone is only a relative value but the combination of the directions of neighboring markers gives an absolute value, which is referred to as directional combinations. Therefore, we create unique coding effectively with the directional combination.

In our recognition algorithm described later, to provide an accurate specification, the estimation of the absolute posture is conducted after that of the relative one, and each marker direction is specified according to the accurately estimated 3D relative posture of the inner circles on the sphere. Therefore, the variations of marker direction should be defined isotropically, based on the positions of the inner circles.

We need to consider the necessary number of variations of directional combinations for the

unique specifications among these. Based on the marker arrangement, which is designed according to the regular icosahedron inscribed in the sphere, each marker has five adjacent markers in equal directions, and there are 30 pairs of adjacent markers, that is equal to the number of sides of an icosahedron, to be distinguished. Subsequently, it can be assumed that one outer circular axis should have $5N$ ($N \in \mathbb{N}$) candidates for marker directions, and the number of directional combinations should exceed 30 pairs of adjacent markers.

The number of directional combinations of neighboring markers excluding the point-symmetrical ones, which cannot specify correspondence, is determined as follows:

$$(5N-1)+(5N-2)+\cdots+1=\frac{5N(5N-1)}{2}. \tag{3.1}$$

For example, in the case of $N = 1$, the number of all directional combinations without a point-symmetrical shape is 10, which is not sufficient for the identification of 30 pairs of adjacent markers. In the case of $N = 2$, the number is 45. According to the result of the brute-force search of the directional arrangement of the outer circular axes, the $N = 2$ case is sufficient to distinguish all adjacent marker combinations. It should be noted that the case of $N = 2$ indicates that there are $5N = 10$ numbers of isotropic marker directions. A schematic of all 45 directional combinations of the two markers in the $N = 2$ case is presented in Fig. 3.2.

In the actual design, as described in condition 1) of Section 3.2.4, the thick parts of the circumference should not interfere with one another as far as possible, while the outer circle diameter should be as large as possible for certain recognition in the limited space on the sphere surface. Therefore, the directional combinations, where the thickest parts of the markers face each other, should be excluded. The brute-force search shows that the marker could be designed by excluding at most three combination numbers (0, 8, and 16) when the marker direction is defined as the widest part between the outer and inner circles, and therefore, an optimized design was adopted based on this brute-force search.

3.2.3 Design of UCM

In this and the next section, we will consider specific design values for each marker.

For the UCM, as the tracking accuracy is evaluated to be almost the same as that of the dot markers in our previous research [37], the inner radius r_{in} and circumference breadth d with reference to the sphere radius r_{sphere} should use the same values:

$$r_{in}/r_{sphere} = 0.40 \tag{3.2}$$

$$d/r_{sphere} = 0.03 \tag{3.3}$$

3.2.4 Design of BCM

Because BCM is newly proposed in this paper, its proper design should be discussed well. Note that, in this paper, the BCM expresses the marker direction using the thickest part of the circumference (in practical terms, the marker direction can be defined as either the thinnest or the thickest part).

The design of the inner circles should be the same as that of the UCM to maintain the same tracking accuracy of the relative posture of the sphere. On the other hand, the space in which the outer circles can be placed is limited, and it is necessary to make effective use of the limited space for the outer circle design.

Design strategy

In arranging multiple outer circles in a limited space, the most important thing is to make the arrangement easy for image processing. While the markers themselves must not be interrupted the

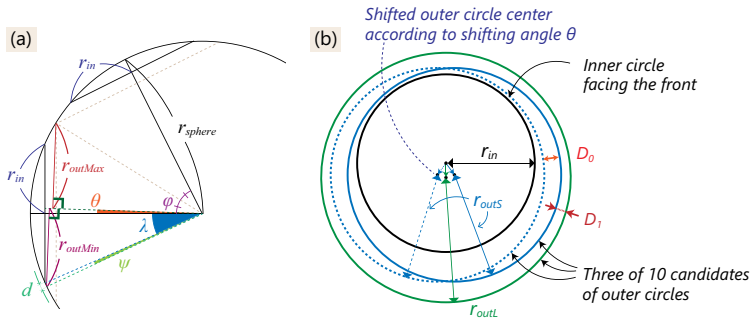

Fig. 3.3. Schematic of designing outer circles of markers. (a) Appearance of outer circular axis shifted from inner one. (b) Appearance of three of ten candidates of outer circles of one marker that alternates small and large radii in parallel projection. Each circle center is shifted according to the tilting angle of the circular axis θ. Cited by Fig. 3 in [38].

marker line, and they must not cross each other, the arrangement needs to have high identification accuracy. To satisfy these requirements, we noted the following four points to be observed in marker design. Note that the following explanation is given according to Fig. 3.3 (each parameter is explained in Table 3.2).

1. The thick parts of adjacent markers do not face one another (this has already been completed by coding design described in Section 3.2.2).
2. There is sufficient space between the adjacent outer circles of different markers.
3. The narrowest part of the inner and outer circles of one marker is not too narrow.
4. There is sufficient space between the adjacent candidates of the outer circles of one marker (e.g. D_0, D_1 in Fig. 3.3 (b) should have adequate values).

Notably, as long as the arrangement of these outer circle groups is symmetrical, they do not necessarily all have the same radius. This is because this strategy does not break the assumption that the outer circle is symmetric with respect to the inner circle. This strategy may have the potential to contribute to the accuracy of marker identification because each marker shows different shape characteristics in the image.

Here, we have the 10 candidate outer circles per marker. To be symmetrical, there should be either two or five variations. Since a large number of variations does not help with identification, two is thought to be reasonable. Therefore, there are two variations for r_{out}: r_{outS} and r_{outL}.

Detailed calculation process

According to the design strategy described above, the parameters of θ and $r_{out\{S,L\}}$ are determined in this section. As explained below, both parameters are dependent on each other, so we need to consider both at the same time.

When considering $r_{out\{S,L\}}$, we need to take into account its upper and lower limits, r_{outMax} and r_{outMin}, so that the markers themselves are not interrupted and they are not crossed each other, as mentioned in conditions 2) and 3). From Fig. 3.3 (a), the maximum value r_{outMax} is determined so as to prevent adjacent markers from colliding, and the minimum value r_{outMin} is determined by the minimum width of the circumference, which is assumed to be the circumferential width d in UCM. These are expressed as following equations:

$$r_{outMax} = r_{sphere} \sin(\varphi - \theta). \tag{3.4}$$

$$r_{outMin} = r_{sphere} \sin(\lambda + \theta + \psi) \tag{3.5}$$

Then, as described in condition 4), it is necessary to take as much space as possible between outer circles that are candidates for a single marker. That is, D_0 and D_1, depicted in Fig. 3.3, to

Table 3.2. Description of parameters used in Fig. 3.3.

Constant values	
r_{sphere}	The radius of the sphere
r_{in}	The radius of the inner circle
d	The breadth of the UCM / the minimum breadth of the BCM
Designed values	
θ	Tilting angle of the outer circular axis
$r_{out\{S,L\}}$	The small and large radii of the outer circle
$r_{out\{Min,Max\}}$	The minimum / maximum value of r_{out}
Other parameters used for calculation	
φ	The half of the angle joining the two inner circular axes of one pair of adjacent markers
λ	The angle forming the arc of an inner circle from the circular axis: $\lambda = \arcsin(r_{in}/r_{sphere})$
ψ	The angle that indicates the arc of the minimum breadth: $\psi = d/r_{sphere}$
D_0	Difference between adjacent outer circles with r_{outS}
D_1	Difference between adjacent outer circles with r_{outS} and r_{outL} each

have sufficient values, which are expressed as following:

$$D_0 = \frac{1}{2} r_{sphere} \sin\left(\frac{\pi}{10}\right) \sin\theta \cos(\lambda + \theta + \psi) \tag{3.6}$$

$$D_1 = r_{outL} - r_{outS} - \sin\theta \sqrt{d_L^2 - 2 d_L d_S \cos\left(\frac{\pi}{10}\right) + d_S^2} \tag{3.7}$$

It should be noted that $d_{\{S,L\}} = \sqrt{r_{sphere}^2 - r_{out\{S,L\}}^2}$.

All of mentioned parameters $r_{outMin}, r_{outMax}, D_0, D_1$ varies according to θ are shown in Fig. 3.4. Note that, in the formula of D_1, r_{outS} and r_{outL} are tentatively assumed to be r_{outMin} and r_{outMax}, respcetively. From this figure, the range of $r_{outMax} - r_{outMin}$ becomes narrower as θ increases, and there is a trade-off between D_0 and D_1. Not only the difference between r_{outL} and r_{outS} but also each value of D_0 and D_1 must be as large as possible. Based on these considerations and from the actual observed images of markers, each parameter is determined as follows.

$$\theta = 2.086^\circ \tag{3.8}$$

$$r_{outS}/r_{sphere} = 0.46 \tag{3.9}$$

$$r_{outL}/r_{sphere} = 0.49 \tag{3.10}$$

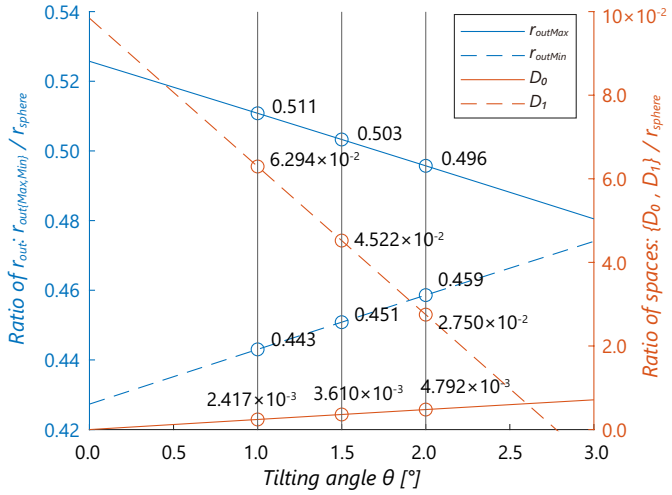

Fig. 3.4. The graph expressing the ratio of $r_{outMin}, r_{outMax}, D_0$, and$D_1$ to the r_{sphere}, which varies with θ. Note that in Eq. (3.7) for D_1, r_{outS} and r_{outL} are tentatively set to r_{outMin} and r_{outMax}, respectively. Cited by Fig. 4 in [38].

3.3 High-speed posture tracking algorithm

This section describes a 500 fps tracking algorithm for the UCM/BCM. Fig. 3.5 presents the processing flow. As processing for each frame, the determination of the relative posture using the inner circle is performed for both the UCM and BCM. For the BCM only, when the absolute posture is not known and after certain conditions are met, its estimation is conducted.

3.3.1 Image Processing

First, the image processing that is commonly performed for the early stages of each process is explained. The flow is depicted in Fig. 3.6. For the gray image obtained from the camera (a), binarization with an appropriate fixed threshold value and 1/2 resizing are simultaneously conducted for acceleration, as indicated in (b). Subsequently, contour extraction is performed on a binarized and resized image, as illustrated in (c). For each contour, the center of gravity is calculated. Simultaneously, the marker ID is assigned properly for each contour group according to the information of the previous frame in the case of frame-by-frame tracking. In the initial frame, the assignment algorithm described in Section 3.3.2 is applied.

Thereafter, as indicated in (d), the inner contours are extracted. In the case of the basic frame-by-frame tracking, this is achieved by using the elliptic geometry of the proper marker ID assigned using the information of the previous frame. Otherwise, in the case of the initial frame, no information is available regarding the proper marker ID for each contour, and therefore, the extraction is tentatively conducted using the center of gravity. Once the rough initialization described in Section 3.3.2 is completed, each contour is assigned with the proper marker ID.

Simultaneously, particularly for the BCM for which absolute posture estimation has not yet been conducted, outlier exclusion and outer point extraction are conducted. Outlier exclusion is necessary because certain parts of the BCM have a wide breadth owing to partial occlusion and being located at the edge of the sphere, unlike the UCM, and appear as a large number of outliers at the sphere edge and occlusion. The exclusion is conducted by using the search for pixel values in the tangential direction of the assigned ellipse geometry. The process for the outer point

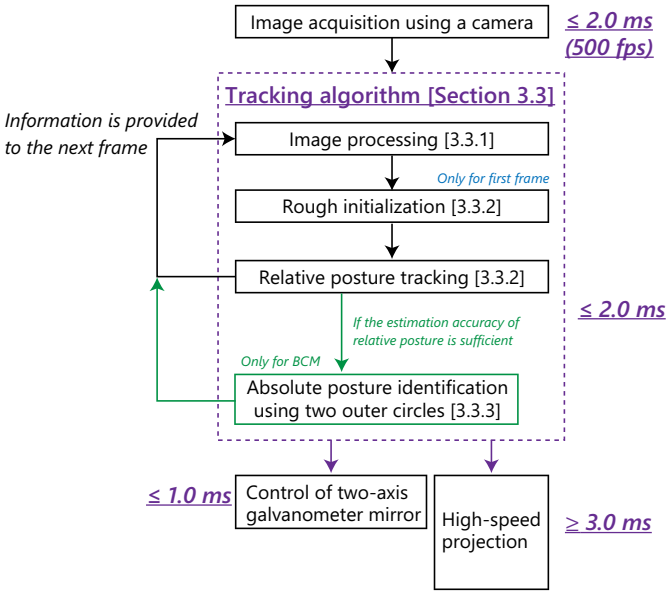

Fig. 3.5. Schematic of tracking flow using high-speed camera, projector, two-axis galvanometer mirror, and image processing. This figure is made referring to Fig. 5 in [38].

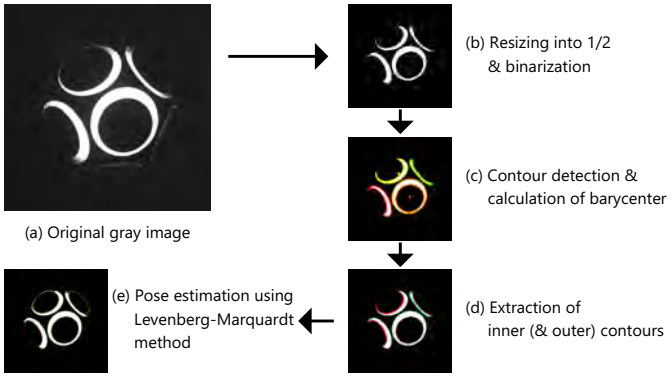

Fig. 3.6. Flow of image processing during relative posture tracking. Cited by Fig. 6 in [38].

extraction is the same as that for the inner points.

3.3.2 Rough Initialization of Relative Posture

Because there is no posture information in the initial frame, the estimation of the initial posture should be conducted by assigning the proper marker ID to each contour. Note that its estimation accuracy is not necessarily high because the iterative optimization in the process of estimating the relative posture converges well [37]. Given that all the inner circles in the circumferential marker have the same diameter, as described in Section 3.2.4, and are likely to have less outliers than the outer ones, we consider the estimation of the initial posture from the inner point groups.

To design the proper estimation process, it is assumed that the more points in the observed contour point group, the more it is likely to be facing forward to the camera. Also, the ellipse fitting is assumed to be performed well for such contour point groups. Based on these assumptions, the estimation process consists of following four points.

1. The two contour point groups with the largest number of points are extracted.

2. Ellipse fitting is performed for each contour point group.
3. The rotation matrix R is roughly estimated from the parameters of the two ellipses.
4. The translation parameter $\boldsymbol{t}$ is roughly estimated using the rotation matrix R and the parameters of the two ellipses.

Two contour point groups are used in this method because at least two ellipses are necessary for the unique posture estimation. In this case, since the ellipse fitting is performed to each contour point group separately, the posture estimation cannot be determined exactly and uniquely. But assuming the accuracy of the ellipse fitting to such a contour point group with a large number of points, it is thought to be enable to make generally correct estimation of the posture.

We explain this process in detail according to Fig. 3.7, and the used parameters are described in Table 3.3. Examples of following processes are shown in Appendix.

Selection of the two contour point groups

First, we select the top two inner point groups with the highest numbers that are tentatively extracted using the center of gravity, as described in Section 3.3.1.

Assuming that an ellipse fits effectively to such a large number of point groups, ellipse approximation is conducted for each point group based on a linear equation [92].

Then, two inner point groups are assigned to the IDs of certain adjacent marker pairs. Since these contour point groups are assumed to be facing forward, it is appropriate to assign the IDs of the two adjacent front-facing markers to them.

Rough estimation of rotation matrix

A rotation matrix R is roughly calculated where the coordinates of two adjacent markers facing front are rotated so that the slope of a straight line consisting of the two marker centers matches that of the estimated ellipses of the two contour point groups. When the difference of the two slope angles is α, R is expressed as follows:

$$R = R_z(\alpha) = \begin{pmatrix} \cos\alpha & -\sin\alpha & 0 \\ \sin\alpha & \cos\alpha & 0 \\ 0 & 0 & 1 \end{pmatrix} \tag{3.11}$$

Rough estimation of translation parameter

In this section, the rough construction of the translation parameter $\boldsymbol{t}$ is described. Note that $\boldsymbol{t}$ is equal to the center coordinate of the sphere in the camera coordinates.

This algorithm uses a lot of simple average calculations. Even with the contour point groups with a large number of points, as long as the points are arranged on a sphere, some missing points may occur. Since ellipse fitting is not robust to such missing point groups [92], highly accurate estimation of the initial posture is impossible in principle. In such a situation, weighting the estimated values according to the number of point groups will cause a lack of robustness. Calculating with a simple average value will give a stable but rough estimation, but this is acceptable because the relative posture tracking converges well, as mentioned earlier.

The estimation process is as follows; we first estimate t_z, the z-coordinate of the translation vector $\boldsymbol{t}$, and then estimate t_x, t_y, the x, y-coordinates, using t_z and the other parameters. This is because, as shown in Fig. 3.7, the z-coordinate of the ellipse center that applies to each contour point group can be easily obtained using the similarity relation of the triangles and can be used for the estimation of t_z.

Generally, the image point $\boldsymbol{p} = (x, y)$ can be transferred to the coordinate P on the image plane $z = 1$ in the camera coordinates using the camera focal parameters f_x and f_y and the principal

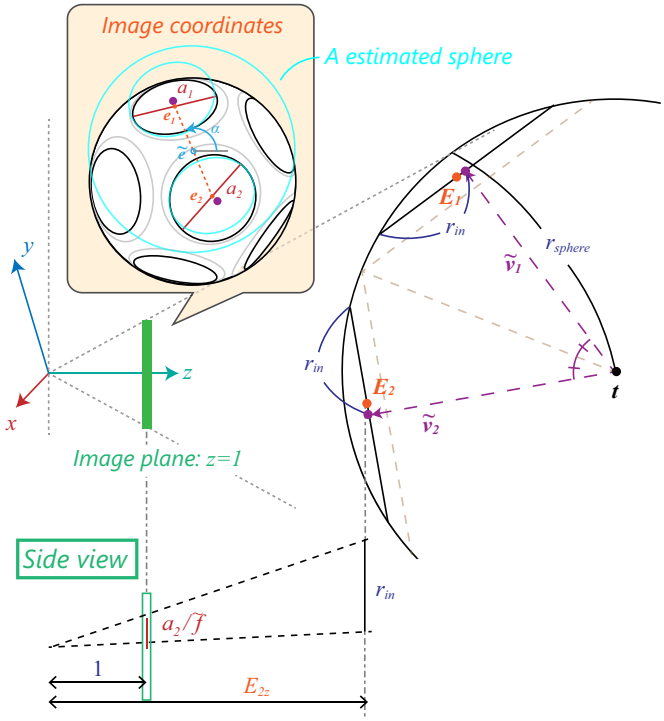

Fig. 3.7. Appearance of rough initialization at initial frame in the camera coordinate system. Cyan illustration in the orange balloon is related to model sphere transitioned by the roughly-estimated parameters R and $\boldsymbol{t}$. Cited by Fig. 7 in [38].

point parameters c_x and c_y as follows:

$$P = \left(\frac{x - c_x}{f_x}, \frac{y - c_y}{f_y}, 1 \right) \tag{3.12}$$

Similarly, a certain length observed in the image l can be also transferred to that on the image plane L. For simplicity, we assume the focal parameter as the average one, $\tilde{f} = (f_x + f_y)/2$, and it can be expressed as follows:

$$L \simeq \frac{l}{\tilde{f}} \tag{3.13}$$

Using this equation, the depth value of the ellipse center E_{iz} $(i = 1, 2)$ is approximated according to the similarity relation of the triangles, as shown in the bottom of Fig. 3.7. Assuming that the major axis of the ellipse a_i and the radius of the inner circle r_{in} are almost coincident, the calculation is as follows:

$$E_{iz} \simeq \frac{\tilde{f} r_{in}}{a_i}. \tag{3.14}$$

Thereafter, the z-coordinate of the translational vector t_{iz} estimated from each ellipse can be calculated as following:

$$t_{iz} = E_{iz} - \tilde{v}_{iz} \tag{3.15}$$

Note that $\boldsymbol{v}_i$ indicates the vector expressing the inner circular axis of each marker in the sphere model coordinates; in other words, the direction from the spherical center towards the inner circular center, as illustrated in Fig. 3.7. Its rotated form using the rotation matrix R is $\tilde{\boldsymbol{v}}_i = R\boldsymbol{v}_i$.

Table 3.3. Description of parameters used in Fig. 3.7. Note that $i = 1, 2$ indicates each of two ellipses, respectively.

a_i	The length of the main axis of the ellipse
$\boldsymbol{v}_i$	The directional vector of the circular axis
$\tilde{\boldsymbol{v}}_i$	$\boldsymbol{v}_i$ transferred to the camera coordinate using the rotation matrix R
$\boldsymbol{e}_i$	The 3D coordinate of the ellipse center
$c_{\{x,y\}}$	The camera principal point parameters in X and Y directions.
$f_{\{x,y\}}$	The camera focal parameters in X and Y directions.
$\tilde{f}$	The average value of both of camera focal parameters: $\tilde{f} = (f_x + f_y)/2$
$\boldsymbol{C_i}$	The image coordinate of the ellipse center
$\boldsymbol{C}$	The image coordinate of the estimated sphere center

Here, the center of the 3D inner circle of the marker is approximated to the value of ellipse center $\boldsymbol{E}_i$; they are not exactly the same [77], but it is acceptable in this rough estimation.

Therefore, the estimated sphere depth value $\tilde{t}_z$ is calculated as the average of two depth values t_{iz}:

$$\tilde{t}_z = \frac{t_{1z} + t_{2z}}{2} \tag{3.16}$$

The x, y-coordinates are roughly calculated using the image points of each ellipse center $\boldsymbol{e}_i$ and the estimated depth value $\tilde{t}_z$. Assuming that the sphere center exists approximately between two ellipses with large contour point groups, $\boldsymbol{t}$ is approximated as follows:

$$\boldsymbol{t} \simeq \tilde{t}_z \left(\frac{\tilde{e}_x}{f_x}, \frac{\tilde{e}_y}{f_y}, 1 \right) \tag{3.17}$$
$$\text{where } \tilde{\boldsymbol{e}} = \frac{\boldsymbol{e}_1 + \boldsymbol{e}_2}{2}$$

3.3.3 Tracking of Relative Posture

The tracking of the relative posture using the inner circles is the same as the previous method of circumferential markers [37]. Using the contours assigned with a proper marker ID, the posture is iteratively optimized using the Levenberg—Marquardt method that minimizes the ellipse–contour distance. During the optimization, according to each known posture change, an ellipse is calculated for each known inner circle by perspective projection. The estimation results are presented in Fig. 3.6 (e). Even after detecting the absolute posture described in Section 3.3.4, this method enables continued tracking thereof. The information of the estimated posture as well as the center of gravity and marker ID of each contour is retained for the next frame tracking process.

In this optimization, all errors are treated equally, and no predictions are included, which indicates that a simple error minimization is performed based on the results from the frame at a slightly earlier time and the data from the current frame. This is because this research assumes that the object to be projected moves in an unpredictable manner, i.e., it suddenly stops or changes its speed or direction, and thus, any prediction that is based on a physical model is not used in our method.

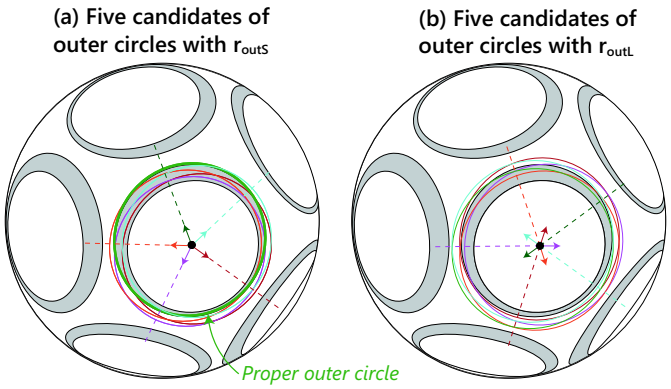

Fig. 3.8. As candidates for the outer circle with respect to the accurately measured outer circle, (a) a small circle r_{outS} and (b) a large circle r_{outL} are arranged in five different directions, respectively. The thick green line in (a) fits the actual outer measurement points best. Cited by Fig. 8 in [38].

3.3.4 Absolute Posture Estimation Using Outer Circles

Soon after the estimation accuracy of the relative posture is assumed to be sufficient, the process of determining the absolute posture is carried out. A schematic of the determination of a proper outer circle is presented in Fig. 3.8. As described in Section 3.2.2, we defined 10 as the appropriate number of candidates of the outer circles for each marker, which vary in terms of the diameter and the tilting direction of the circular axis and whose parameters in the camera coordinates are calculated using the estimated relative posture R and $\boldsymbol{t}$. Subsequently, the ellipse is calculated for the parameters of each candidate of outer circle in the same manner as the process conducted in the iterative optimization, as described in Section 3.3.3. Thereafter, the error between the outer point group and an ellipse is calculated for each of the 10 ellipse candidates. Those with the lowest error are adopted as the proper outer circle. Finally, according to the two marker directions of the two adjacent markers, the directional combination ID is estimated, and therefore, the absolute posture can be determined.

As noted in our previous study [37], the ellipse–contour distance (error) is calculated using all of the point groups, and therefore, the error differs significantly depending on true or false marker identification, which is effective for artifacts. The calculation of the ellipse–contour distance is conducted only 20 times (10 candidates for two outer point groups), whereas it is conducted more times in iterative optimization for relative posture tracking. Thus, the calculation cost is thought to be relatively low while retaining high estimation accuracy.

Note that in actual execution, for the purpose of acceleration, a point group with a large number is thinned out. Also, the marker design that is intended for less misidentification, as described in Section 3.2.4, also contributes to the high estimation accuracy.

3.4 Projection mapping system for widely moving object

The proposed UCM/BCM can be used for a conventional projector-camera system with a wide angle; however, we propose an optical system that projects onto a widely moving and rotating object, which enhances the performance. An overview of the system is presented in Fig. 3.9. Multiple optical axes of a high-speed camera, a high-speed projector [131], and infrared light are coaxialized using a beam splitter and the optical directions are rapidly scanned with two-axis galvanometer mirrors. Owing to a high-speed coaxial optical control system [16], high-speed

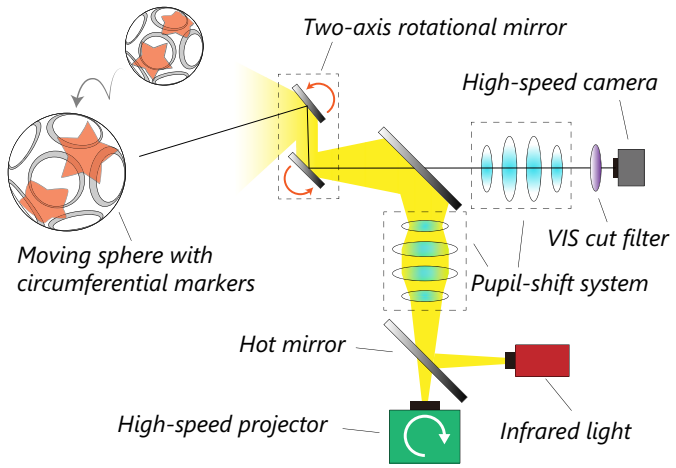

Fig. 3.9. Optical system for projecting onto widely moving and rotating object. Cited by Fig. 9 in [38].

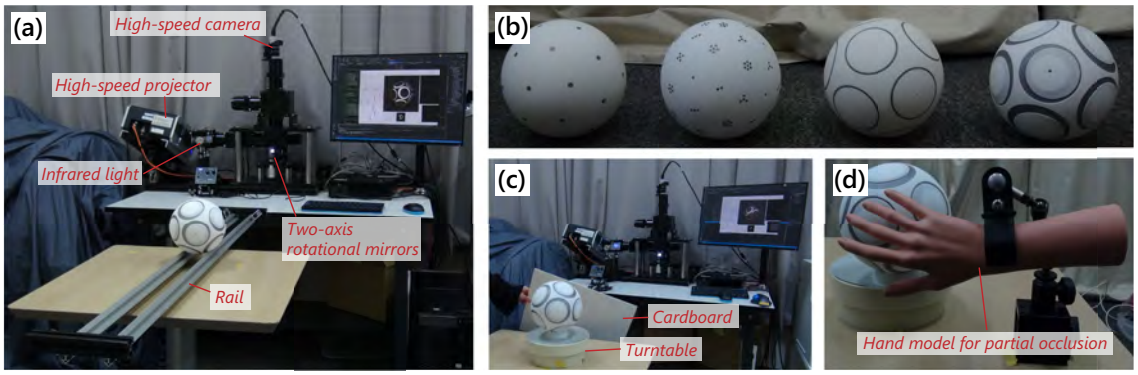

Fig. 3.10. (a) Appearance of overall evaluation experiment including rail for the movement in depth direction. (b) Spheres attached to dot markers, EDCM, and proposed UCM and BCM. (c) Appearance of overall evaluation experiment including turntable for rotational movement and a cardboard. (d) Appearance of hand model partially shielding the sphere surface. Cited by Fig. 10 in [38].

target shooting and projection are realized. It can also maintain high-resolution shooting against objects moving in a wide area using the camera with a narrow angle of view; therefore, it can enhance the tracking performance of the object translation compared to a conventional wide-angle projector camera system.

Because a more telephoto lens can be used in the coaxial optical system, the field depth is narrower than that of a conventional wide-angle system. However, the proposed circumferential marker compensates for this problem. Moreover, as shown in Fig. 3.5, owing to high-speed devices, the projection latency for per frame is approximately 7 ms from image acquisition to projection at 500 fps, which is almost the same value as the allowable delay [69]. In the following section, the highest tracking performance is evaluated using this system.

Using the proposed projection mapping system and tracking markers for enhancing the posture estimation performance, it will be possible to present SAR in conjunction with motion dynamics. One effective application is the real-time visualization of rotation information for athletic practices. For example, with the measurement of the high temporal resolution, the rotation speed can be calculated by thinning out, such as an average of 10 frames (20 ms), which is thought to be more accurate than that of a low frame rate. The BCM, which obtains the absolute posture, offers

more effective applications. The rotation information during complete occlusion can be interpolated from the information before and after. By using multiple cameras that measure from different directions, the integration of the absolute posture can also provide accurate measurements, regardless of occlusion. It is notable that our system can execute not only high-speed measurement, but also DPM of such information onto the target surface, owing to the high-speed optical system. With such projection mapping, athletes can obtain real-time motion feedback while concentrating on play. By using multiple projectors, it is possible to project such information on a sphere omnidirectionally. Of course, such projective expressions can also be applied in entertainment, such as juggling.

3.5 Evaluation

The experimental setup illustrated in Fig. 3.10 (a) consisted of a high-speed camera (Photron IDP Express R2000; 500 fps, 512×512 px, monochrome) with a telescope lens (Computar M2514-MP2, focal length: 25 mm), Novanta M2 galvanometer mirrors (scanning angle: ±30 deg), a DynaFlash high-speed projector (1,000 fps, minimum latency: 3 ms, 1024×768 px, monochrome), and infrared light source (IMAC IBF-LXS30AIR-850; wavelength: 850 nm). The camera, projector, and infrared light were placed coaxially using a beam splitter and hot mirror. The computer was a DELL PRECISION 7920 Tower (Windows 10 Pro 64 bit, Intel Xeon Gold 5118, NVIDIA Quadro P2000, 32 GB memory).

The camera lens was calibrated in advance. The visible light of a projector with a large light intensity that was uniform over the entire angle of view rather than infrared light was used for the evaluation. Infrared light was used for the demonstrations described in Section 3.6. Note that, because the area of the retroreflective material differs depending on the marker design, the intensity of the projected light was adjusted to suit each other, for accurate binarization.

Spheres with a diameter of 150 mm were attached to dot markers, EDCMs [58], and the proposed UCM/BCMs, respectively, as indicated in Fig. 3.10 (b). The sphere with dot markers had 32 markers that were evenly distributed. That with EDCMs had 59 markers, the design of which was calculated by a brute-force search. Note that the spheres with dot markers and UCMs were the same as those used in our previous research [37]. Original size images were used to track dot markers and EDCMs.

3.5.1 Evaluation of Tracking

The frame-by-frame tracking was compared for four markers: the dot marker, EDCM, UCM, and BCM. For the EDCM and BCM, only the process after the absolute posture estimation was used for evaluation. The evaluation was performed from three points of view: the basic tracking performance, robustness against occlusion, and translational movement in the depth and horizontal directions. The numerical values of the evaluation were calculated using three or more sequential data for each marker and each condition.

Evaluation of Basic Tracking Performance

As shown in Fig. 3.10 (c), the tracking performance was evaluated for the sphere rotating at a constant velocity of approximately 20 $^\circ/s$ while fixed on the center of the turntable. The center of the turntable was located approximately 122 cm from the front of the device. The data obtained for 5,000 ms at 500 fps; that is, 2,500 frames, were used for evaluation. The components used for the evaluation were the estimated rotational angle and the X, Y translational values. Because the rotational angle increases linearly owing to rotation with a constant velocity, the RMSE of the rotational angle was calculated using the error of each frame against a straight line that fit those data using the least squares method. The RMSEs of the estimated translational values X, Y were

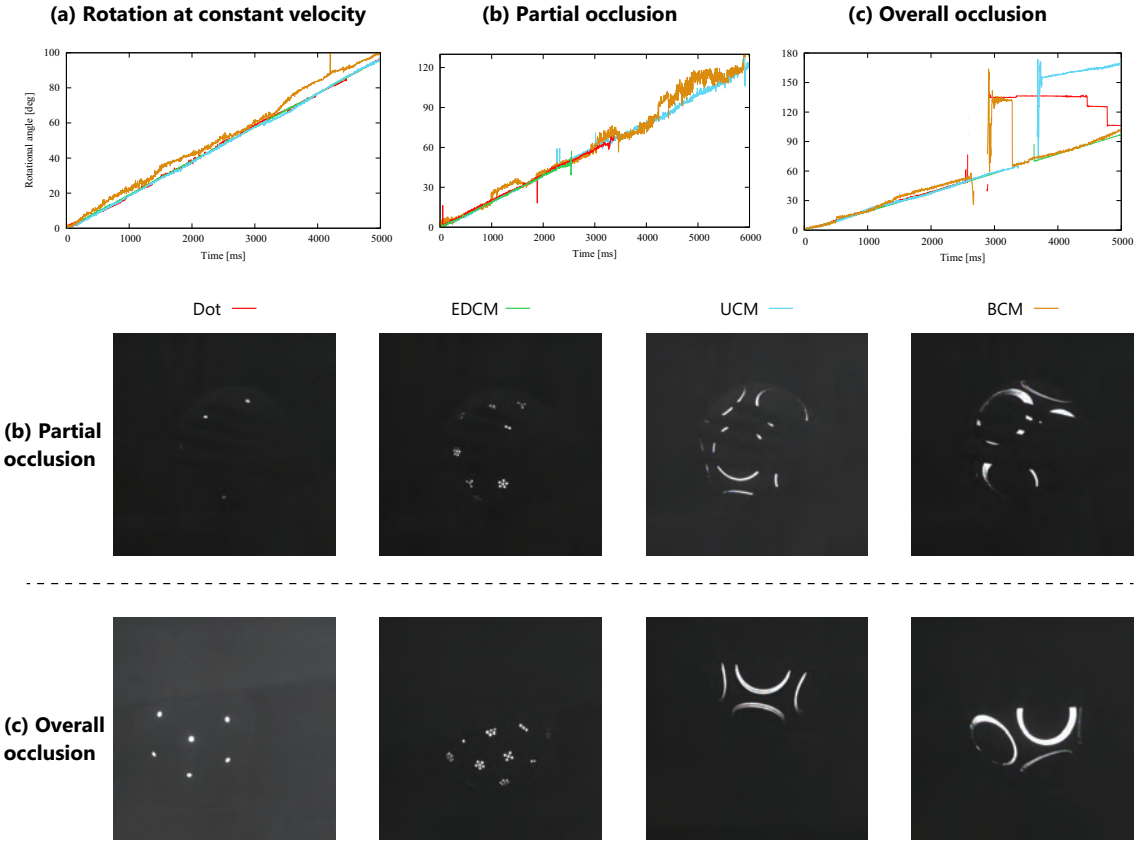

Fig. 3.11. Evaluation of (a) tracking performance and the robustness of (b) partial and (c) overall occlusion for a sphere rotating at a constant velocity. Cited by Fig. 11 in [38].

Table 3.4. RMSE of Estimated Rotational Angle and X/Y Translation During Constant Rotational Movement.

RMSE	Dot	EDCM	UCM	BCM
Angle [deg]	1.4703	1.069	1.501	2.090
X [mm]	0.4326	0.1833	0.4042	0.8956
Y [mm]	0.4022	0.2028	0.4546	0.8264

also calculated based on a value of 0 because the two-axis optical control system always tracks the object so that its center is always at the center of the image.

These results are presented in Table 3.4 and Fig. 3.11 (a). The UCM and BCM exhibited highly accurate estimations, with RMSEs of less than 2.1 ° in the rotational angle and 1.0 mm in X/Y. The EDCM exhibited the highest accuracy, while the dot markers had almost the same accuracy as that of the UCM.

From these results, it can be concluded that the EDCMs provided the most accurate performance. Although the EDCM has a more complicated shape than the dot marker, as markers located at the sphere edge are excluded whereby any of the remaining dots are lost according to the sphere rotation, it maintains highly accurate tracking. The dot markers located at the edge of the sphere have a chipped shape, and therefore, the estimation accuracy is likely to reduce compared with that of the EDCM. Based on the estimation value, the estimation accuracy of the UCM is assumed to be almost the same as that of the dot markers with a small number. The

Table 3.5. ES Rate [%], RMSE of Angle [deg], and Translation X/Y [mm] Only for Estimated Frame in Evaluation of Partial Occlusion.

	Dot	EDCM	UCM	BCM
ES rate [%]	89.24	46.17	100.0	100.0
RMSE of angle [deg]	12.98	1.648	2.163	3.877
RMSE of X [mm]	5.217	0.5897	1.002	1.967
RMSE of Y [mm]	4.685	0.3481	1.166	2.117

relatively large estimation error of the proposed BCM is assumed to be caused because the piece of the marker located at the end of the sphere was large and the outliers could not be completely removed. Moreover, the estimation accuracy of both UCM and BCM may be affected by the artifacts associated with the unevenness of the pasted curve marker, which occurs especially for the marker located at the edge of the sphere.

Evaluation of Partial/Overall Occlusion Robustness

In this section, we examine the robustness against partial/overall occlusion for each marker.

The evaluation of the partial occlusion was conducted using the model of the human hand and arm, as shown in Fig. 3.10 (d). By placing the sphere at the edge of the turntable, whose setting is the same as the previous section, the partially occluded area increased gradually according to the turntable rotation. The data of the tracking for 6,000 ms (3,000 frames at 500 fps) until the majority of the sphere was shielded were used for the evaluation of the RMSE of the angle and estimation success rate (denoted as the ES rate). Note that the initial posture of the sphere was changed manually for each evaluation. The evaluation of the overall occlusion was executed by shielding with a large piece of cardboard, shown in Fig. 3.10 (c), for approximately 1,000 ms during 5,000 ms of tracking of the sphere that was at the center of the turntable and rotated at the same velocity.

The results of the partial occlusion are presented in Table 3.5 and Fig. 3.11 (b). According to the ES rate, the dot marker and EDCM could not continue the posture estimation with the increasing occlusion area, whereas the UCM and BCM could complete the estimation. The RMSE of the angle and X/Y translation, displayed in Table 3.5, was calculated only for those for which the estimation was conducted. Those of the dot markers had large values because each dot ID was misrecognized according to the increasing occlusion area, while those of the EDCMs did not have large values for the reason indicated in Section 3.5.1. In comparison, those of the UCM and BCM increased significantly, which implies instability during the partial occlusion of complex shapes.

The results of the overall occlusion are depicted in Fig. 3.11 (c). The EDCM and proposed BCM achieved absolute posture estimation following overall occlusion, whereas the dot marker and UCM did not. However, only the UCM achieved relative posture estimation owing to the geometry of its ellipses, whereas the dot markers did not. According to the graph, the UCM and BCM could restart the tracking using the rough initialization algorithm introduced in Section 3.3.2. However, during the restarting process, they exhibited a larger estimation error and the convergence to the precise posture appeared to be later than that of the EDCM, which may have been caused by the difficulty of rough initialization during the partial occlusion. This trend was similarly observed in the evaluation of the ES rate described in Section 3.5.3.

Evaluation of Translational Movement in X/Y or Depth Direction

For this evaluation, the sphere was placed on the rail (125 cm length) and rolling motion started from its center to the end, and subsequently the other end. To evaluate the movement in the depth direction, the rail was placed vertically, and its end point was at a distance of approximately 60 cm

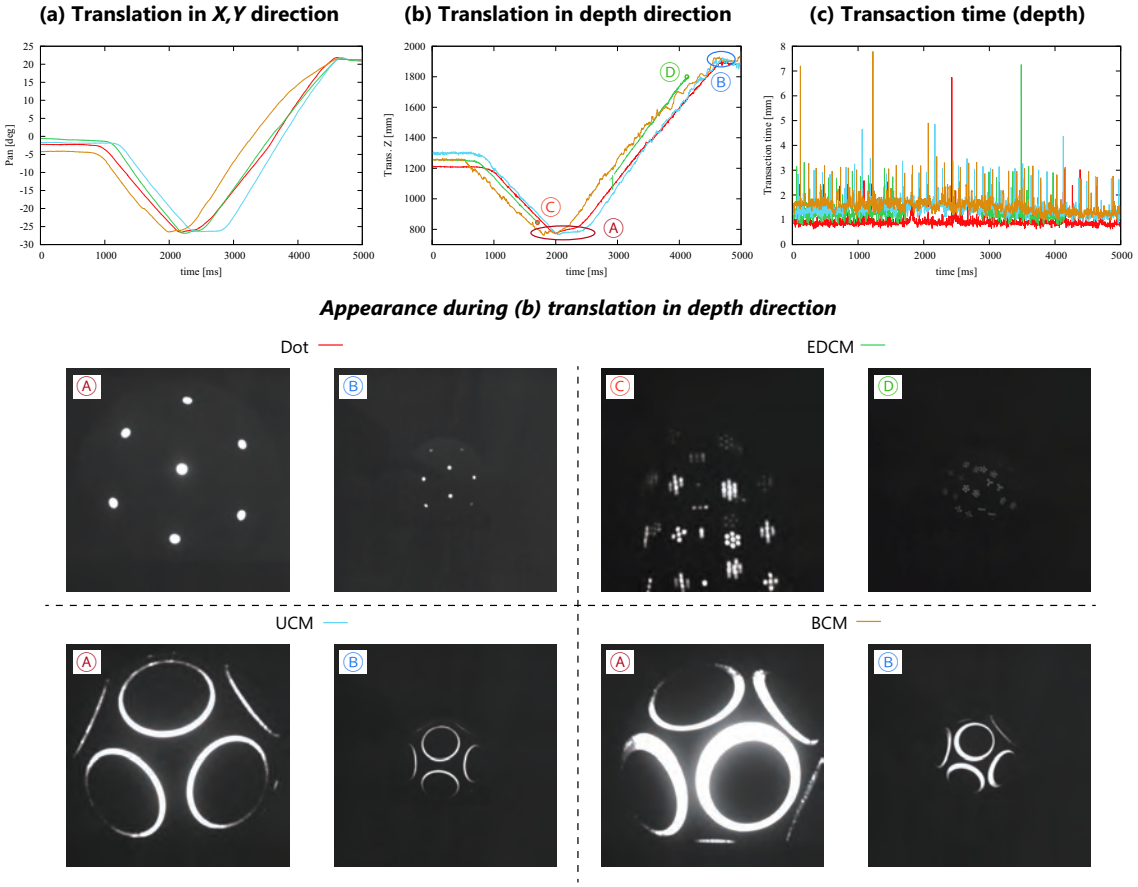

Fig. 3.12. Evaluation of translational motion in (a) horizontal and (b) depth directions with respect to device, and (c) calculation time. Cited by Fig. 12 in [38].

from the front of the device, as illustrated in Fig. 3.10 (a). To evaluate the movement in the lateral direction, the rail was placed horizontally, and its center line was at a distance of approximately 109 cm from the device.

The tracking performance of the horizontal movement was evaluated using the pan real angle of the galvanometer mirror, as depicted in Fig. 3.12 (a), implying that all of the markers could be tracked stably in the lateral direction.

The tracking performance of the movement in the depth direction was evaluated using the estimated translation value z, as illustrated in Fig. 3.12 (b). Only the EDCM could not achieve wide-range movement in the depth direction; its movement range was evaluated as approximately 840 to 1,810 mm. The other markers achieved wide-range movement in the range of approximately 770 to 1,900 mm. This result implies that the complicated shape of the EDCM is unsuitable for wide motion in the depth direction, which causes blurring, whereas the proposed UCM and BCM consisting of smooth curves are robust against blurring, and the BCM is compatible with code expression.

3.5.2 Evaluation of Calculation Time

The calculation time of each marker during tracking was evaluated using the data of the movement in the depth direction, where the sphere size changed continuously. The results are presented in Table 3.6 and Fig. 3.12 (c). Apart from certain frames with heavy processing owing to processor collision, in almost all frames, the transaction time was less than 2.0 ms for all of the markers.

Table 3.6. Calculation Time Per Process and Initialization [ms]

	Dot	Cluster	UCM	BCM	Initialization for BCM
Ave.	0.8942	1.247	1.4163	1.479	0.1451
Std.	0.2128	0.3172	0.3238	0.3341	3.269E-02

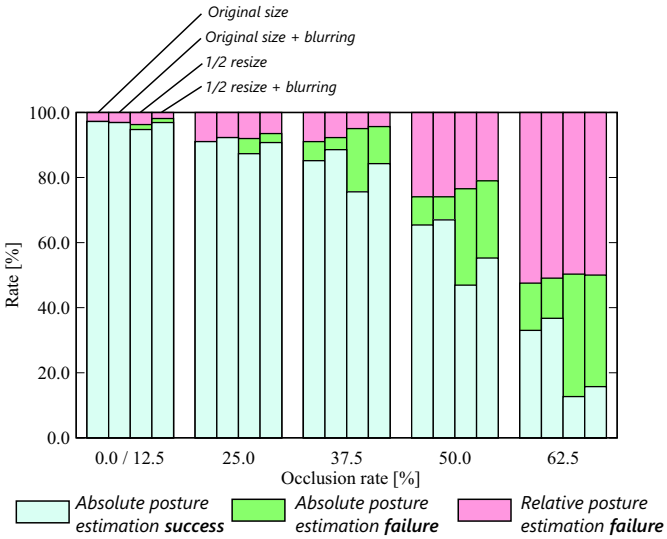

Fig. 3.13. Results of absolute posture estimation for 324 posture variations in CG simulation for images of original size and 1/2 resized with and without blurring. There were three types of estimation results: success (blue bar) and failure (green bar) cases of the absolute posture estimation, which was conducted after accurate relative posture estimation, and the case where even the relative posture estimation could not be completed (red bar). Cited by Fig. 13 in [38].

Therefore, it can be confirmed that the proposed UCM and BCM are sufficient for high-speed processing at 500 fps in addition to the other markers. In particular, the calculation times of the UCM and BCM were almost the same because both processes were the same during the tracking.

The calculation time of the absolute posture estimation of the proposed BCM was evaluated using 18 initialization data points, appearing in Table 3.6. It can be observed that sufficiently fast processing was achieved even when combined with the tracking process per frame.

3.5.3 Evaluation of Absolute Posture Estimation

We evaluated the success rate of the proposed estimation algorithm for the absolute posture for the BCM with various rotation types using a CG simulation. The target sphere was fixed at 1,300 mm away in the camera coordinates and rotated at a 10° pitch on the X and Y axes. The frame-by-frame tracking of each rotational posture was performed for 200 iterations using the estimation algorithm, which ended halfway when the absolute posture was estimated. The result of the posture estimation was expressed as the ratio against the 324 total variations in the rotation. The comparison was performed at two resolution types: original size (512×512 px) and 1/2 resized (used in the actual process), as well as with/without blurring (15 px at the original size), while changing the occlusion ratio of the overall image from the left side (0.0% to 62.5% at a pitch of 12.5%). The sphere was positioned at the center of the image and its diameter in the original image was 288 px. Therefore, the occlusion at 0.0% and 12.5% provided the same estimation

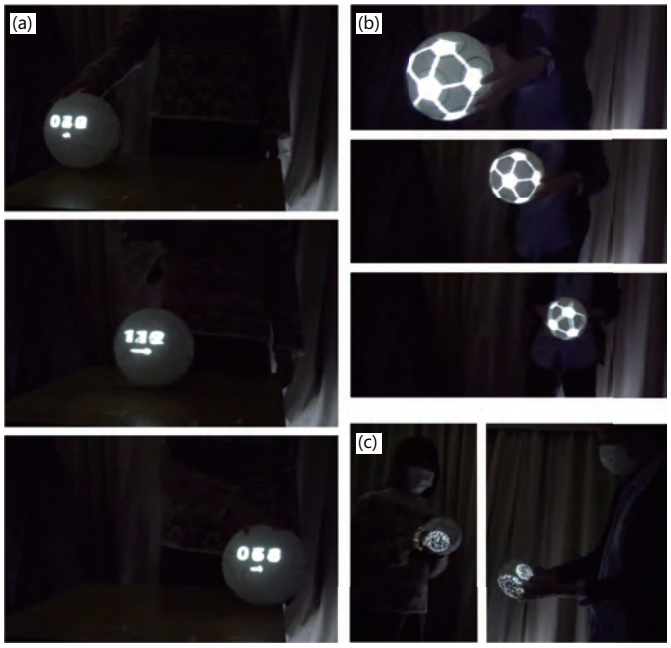

Fig. 3.14. Appearance of demonstration. (a) The posture of a sphere attached to the UCM is tracked, and the rotational velocity is presented as an arrow and rpm value. (b) Tracking and projection in a wide range using the UCM. (c) Spherical display presenting an earth map using BCM that is robust against the partial/overall occlusion. Cited by Fig. 14 in [38].

results.

The evaluation results are presented in Fig. 3.13. There were three types of estimation results: the success (blue bar) and failure (green bar) cases of the absolute posture estimation, and the case in which even the estimation of the relative posture could not be achieved; that is, the error of its estimation could not be lower than the threshold (red bar).

For any condition of the processed image, an ES rate of 91.05% or higher was maintained up to 25% occlusion (approximately 5.5% occlusion for the sphere area). It is clear from the graph that the ES rates of both the absolute and relative posture (blue and green bars) decreased according to the larger occlusion rate. The former was approximately 65.43% for the original sized image without blurring and 46.91% for the resized one at 50% occlusion; that is, half of the sphere was shielded. At the highest occlusion rate of 62.5%, almost all of the estimations failed (33.02% for the original sized image without blurring and 12.65% for the resized images). It is notable that in most cases the ES rate tended to increase with the blurring. This is thought to be because the artifacts were mitigated by blurring. Moreover, this result indicates that the blurring did not affect the reduction in the ES rate.

There were cases when even the estimation of the relative posture failed, as indicated by the red bars, because the adjacent BCMs with relatively thick parts facing one another were sometimes mistakenly recognized as the same marker. As described in Section 3.2.4, a greater distance between adjacent BCMs results in more limited spaces in which the outer circles can be placed and more cases of decoding failure (green bars). It was necessary to adjust the parameters and repeatedly observe the results in this simulation to identify the optimum marker arrangement.

3.6 Demonstration

The appearance of the demonstration is depicted in Fig. 3.14.

(a) The rotational velocity is presented on the sphere surface attached to the UCM as a result consisting of an arrow and rpm value. The recording of the rotational velocity is performed at a high frame rate owing to the high-speed optical system, which enhances the accuracy of the motion estimation. Furthermore, with the use of the BCM, the recording of the rotation during complete occlusion can be interpolated using that before and after the occlusion.

(b) The appearance of the tracking and projection onto a sphere with a diameter of 200 mm, which is larger than that used in the evaluation (150 mm), in the wide range using the UCM is shown (of course, the BCM could also realize this demonstration). As explained in Section 3.5.1, this method is robust against partial complex occlusion by the hands and arms, and is suitable for dynamic interaction with people grasping the sphere. Moreover, a wide range of motion is guaranteed in both the lateral and depth directions, as discussed in Section 3.5.1.

(c) It is also possible to use spherical displays for purposes such as presenting earth maps using BCM. Owing to the robustness against dynamics, the sphere can easily be handed over from one to another. Moreover, the consistent projection mapping corresponding to the absolute posture is realized even after overall occlusion, which simplifies the interaction with humans. With the proposed optical system, the image projection can be viewed at a high resolution.

Although this was not actually implemented in this study, owing to BCM, consistent projection mapping from all directions by many projectors can also be realized. By integrating measurements with multiple cameras, highly accurate rotation from a distance is also possible.

3.7 Discussion and conclusions

In this research, focusing on the circle shape that is resistant to blurring and occlusion, we have proposed the UCM and BCM for sticky projection mapping even in dynamic situations. The evaluation experiment confirmed a high tracking performance with an accuracy of less than 1.0 mm and 2.1 ° (Section 3.5.1), high speed (500 fps) (Section 3.5.2), and robustness against partial/overall occlusion and blurring (Sections 3.5.1 and 3.5.1). Moreover, using a simulation, the BCM was confirmed to exhibit a sufficient success rate for the absolute posture estimation even with occlusion (Section 3.5.3). With the wide projection mapping system (Section 3.4), projection mapping onto a sphere in dynamic situations, including wide-range motion and interaction with humans, and real-time rotation visualization onto the object surface can be realized for the first time (Section 3.6). According to the evaluated tracking accuracy, the UCM or BCM should be selected depending on the purpose: the UCM should be used when only relative attitude tracking is sufficient and the BCM should be used when considering omnidirectional measurement/projection from multiple cameras/projectors.

There are two notable points that we have proposed for appropriate circumferential marker design. The first is an isotropic arrangement that enables relative posture estimation. The other is that each marker has geometric information as an elliptical shape. Using this arrangement, even a small number of markers can be used as a clue to determine the relative posture roughly (Section 3.3.2), and therefore, the tracking of the relative posture can easily be achieved (Section 3.3.3). For the coding design, the accurately estimated relative posture is used as a basis for multiple outer circle candidates, which leads to high accuracy of the absolute posture estimation (Section 3.3.4).

Several notions should be considered in future works. In particular, the BCM has a large marker area made of retroreflective material; that is, the part where the light is returned in the arrival direction, which may lead to a dark projected texture being seen by the surrounding observers.

Moreover, it is difficult to attach the UCM/BCM to the sphere. In this study, a retroreflective material was attached to a 3D-printed sphere with a groove of the maximum breadth of each marker. In particular, a thin C-shaped ring was applied over the BCM to form a biased ring shape. Because many retroreflectors are sold only in the form of flat sheets, it is difficult to attach these

to the curved surface of a sphere, which may result in distorted paste and they may appear as artifacts in an image, which leads to an increase in the estimation error for both UCM and BCM. However, the spheres resemble a design such as a soccer ball or volleyball, which is thought to be useful for usual sewing work on a ball made of cloth.

Especially for BCM, it is considered that the thick piece of the marker caused owing to its location at the edge of the sphere or because of some occlusion may become a large outlier, and it significantly contributes to the estimation accuracy. Algorithms that efficiently remove such outliers should be reconsidered in the future.

The novel proposed concept of the conic-shaped marker design in an isotropic arrangement can easily be extended to other rotationally symmetrical objects with a conical shape of the cross-section, such as cylinders and cones. In the future, we should consider how to design such a conic-shaped marker for other shapes, which may be easy to implement for rotationally symmetrical objects such as rugby balls and bats.

Furthermore, an effective compensation algorithm that lacks rotational information owing to occlusion and the design of its visualization presentation, which has been expressed as only an arrow and rpm value in our study, should be considered.

Chapter 4

High-speed Gaze-contingent Display With Ocular Parallax Rendering

4.1 Introduction

Existing virtual/augmented reality (VR/AR) technologies rely heavily on three-dimensional (3D) images. Most devices use a binocular parallax image as the easiest way to present them; compared to technologies such as spatial light modulators and fly-eye lenses that produce 3D images in the air, binocular parallax can be executed by simply presenting images from different viewpoints for each eye. Furthermore, it can be easily set up in any flat displays, such as head-mounted displays (HMDs) and normal monitors.

By contrast, stereoscopic presentation using binocular parallax images cannot realize an ideal depth perception with large disparity. Figure 4.2 (b) shows a graph of the relationship between the disparity and perceived depth, as well as the fusible range [25]. There is a limit of disparity that allows binocular fusion, whose value differs from per user and under different conditions [178]. The perceived depth is not along the ideal line that is proportional to the disparity; it decreases significantly as it approaches the fusible limit [179]. This causes significant loss of reality in VR/AR and should be improved for immersive experience.

Numerous efforts have been made to understand and improve binocular fusion and depth perception. The common focus of these efforts is getting closer to the conditions of real-world stereoscopic viewing. There are various depth clues in visual perception [155]. Among them, motion parallax has been considered to have a major effect on depth perception, and because it is easy to conduct, many studies have been published on it [156, 157, 158]. However, in most situations, we consider objects without moving heads, whose depth can be naturally perceived. For such a static situation, that is, binocular parallax, oculomotor cues (accommodation and conflict) and pictorial cues (occlusion, size, light reflection and shadows) are important.

This study focuses on the ocular parallax assuming that it may provide an positive effect on the convergence work for depth perception, as suggested by several previous studies [180, 30, 31, 181, 182, 183, 184]. A faithful eye model is shown in Figure 4.1 (a). Conventional VR/AR systems have assumed the center of the eyeball C as a viewpoint position, but actually, it exists at the first nodal point of the cornea N_1, which is a bit closer to the corneal surface than C. As the eye constantly moves due to saccades, the viewpoint of N_1 always shifts slightly and repeatedly, and then ocular parallax is produced. Although its disparity is small, the faithful reproduction is considered to have a positive impact on the humans perception in VR/AR. However, there are only a handful of cases where the ocular parallax rendering has been applied to VR/AR [32, 36, 166].

Eye tracking is necessary to reproduce the ocular parallax in VR/AR systems. Owing to the

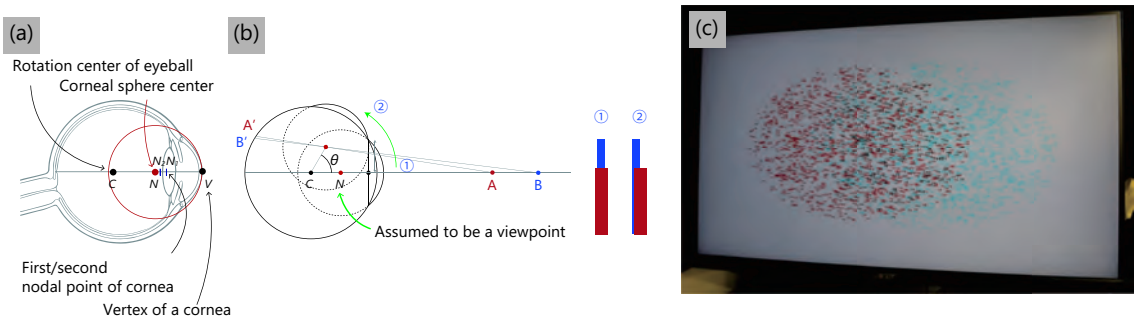

Fig. 4.1. (a) A complex eye model including first/second nodal points N_1, N_2, a corneal sphere which approximates the corneal anterior surface (red circle), and its center N. Although slightly shifted toward the retina from the actual viewpoint N_1, N is assumed as a viewpoint and tracked in this study from the algorithmic efficiency perspective. (b) A scene where an apparent gap appears due to ocular parallax. (c) The appearance of anaglyphic random dot stereogram (RDS) presented corresponding to ocular parallax. Small movements due to saccade are observed from the side (note: this image is superimposition of two frames)

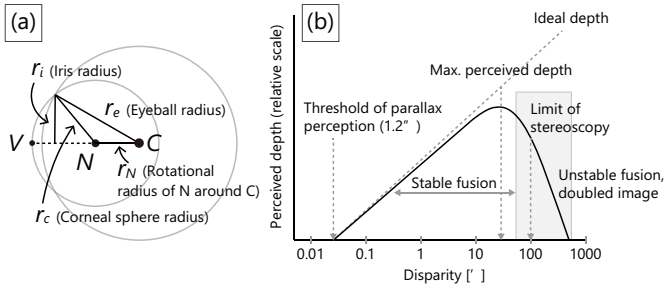

Fig. 4.2. (a) A schematic of a simple eye model [43]. (b) Relationship among disparity, perceived depth, and stereoscopic limits, cited by Figure 10 in p.9 [25] and reprinted by Fig. 1.2 (b).

advances made in eye tracking over the last two decades, many efficient eye trackers along with HMDs are commercially available [185, 33, 34, 35, 186]. In a previous study, the ocular parallax effect in such a state-of-art commercial HMD was validated [36]. A user study confirmed an improved experience, but only in a limited situation where two objects are in a back-and-forth relationship and their gap easily appears by ocular parallax (Figure 4.1 (b)). However, no difference was observed in the answering task of perceived depth by hand reaching. One possible reason for this is the over 20 ms latency of the used commercial device, as also discussed in that previous study.

Recent VR/AR systems require extremely small latency for better experience [69], and many studies have realized high-speed VR/AR system using high-speed display [131] and image processing [1, 55, 58, 5, 167, 38]. Assembling such a high-speed device is necessary for ocular parallax because it becomes closer to the conditions for real-world stereoscopic viewing and has a possibility to enrich VR/AR experience. Then, ocular parallax effect should be investigated in such a VR/AR system.

This study assembles a fast and accurate stereoscopic viewing system that reproduces ocular parallax faithfully to investigate its effect on the depth perception. The necessary specifications is quantitatively estimated through simulation. An algorithm of allowable approximation of a tracked nodal point, as shown in Figure 4.1 (a), is proposed to enable both accuracy and speed. Using this system, two user studies on binocular fusion and perceived depth were conducted;

Table 4.1. Summary of this paper and comparison to a previous study [36]. Note that 2AFC means two-alternative force choice test. ◯ and × mean whether a difference was observed, respectively.

	Konrad et al., [36]	Ours
Camera FPS	200 Hz	1000 Hz
Display FPS	90 Hz	360 Hz
Latency	$>$ 20 ms	Ave. 4.072 ms
Camera Type	Monaural	Stereo
Rendered scene	Clear gaps	RDS
Depth perception	Hand-reaching ×	Verbal answer ◯
Binocular fusion	–	Single response ◯
Reality	2AFC ◯	–

results show improvement in both.

A summary of the comparison between this study and the most relevant previous study by Konrad et al. [36] is shown in Table 4.1. The scene assumed in this study is not limited to the specific case of the obvious gap (Figure 4.1 (b)), but is a generic scene by just presenting a random dot stereogram (RDS) (Figure 4.1 (c)). Furthermore, our evaluated latency is significantly smaller than the previous one.

The following sections describe our elaborate strategies to realize such a system, followed with an explanation of how to assemble it and a quantitative evaluation on it using a simulation. The user study section confirms the effectiveness of our proposal with statistical analysis, followed by discussions based on these results. Note that our focus is a detailed design strategy on such a high-speed device and investigating its effect on human perception. Thus, we leave detailed comparisons with previous studies for future work.

4.2 Strategy of high-speed ocular parallax rendering

In this section, we describe the strategy of high-speed rendering corresponding to ocular parallax. As shown in Figure 4.1 (a), the optical principal point of the eye (viewpoint) is in principle considered the first nodal point N_1 of the cornea, not the eyeball center C. N_1 always involuntarily rotates around the C due to saccade, and ideally its position should be always tracked and the display rendering should correspond to its position to realize faithful ocular parallax. However, directly measuring N_1 has problems in terms of algorithmic speed and accuracy.

The used eye tracking method is the most popular and standard one, called pupil-center-corneal-reflection (PCCR) [40]. It assumes a simple eye model consisting of only two spheres [43], as shown in Figure 4.2 (a). It first finds the corneal sphere center N, then performs an elliptical approximation of the pupil, and finally estimates the line of sight. That is, once the gaze is obtained, it is possible to shift the viewpoint from the eyeball center C to the corneal node N_1 using the known length between them, which was actually conducted in the previous study [36, 166].

However, there are difficulties with the accurate elliptical approximation of the pupil, for example, as many as five parameters of an ellipse [111], difficulties of binarization with a constant threshold, and various noises caused by eyelash and eyelid. Although numerous efforts have been made to advance the available high-speed [113, 186] and precise methods [110], balancing both accuracy and speed is still challenging. Meanwhile, the corneal sphere center N can be linearly solved by using a stereo camera and its tracking is relatively fast and accurate. Since these meet the desired specifications of this study, it is desirable to track N, not N_1.

In the aforementioned faithful eye model, the distance from the vertex of the anterior corneal

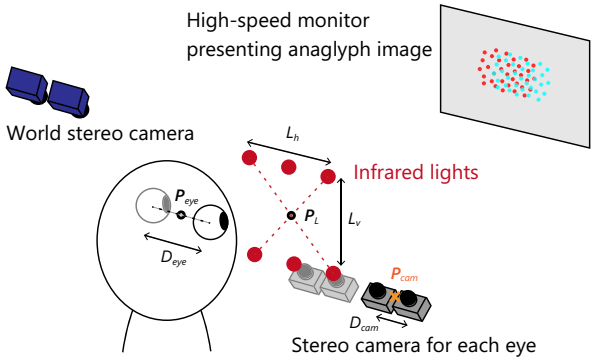

Fig. 4.3. Overall system diagram.

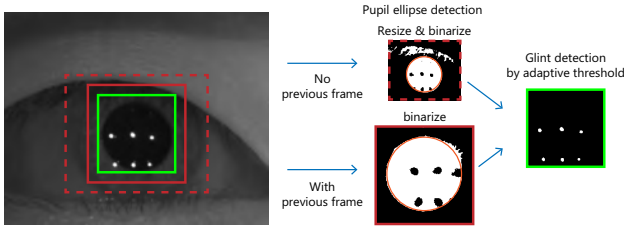

Fig. 4.4. Flow of high-speed image processing.

surface V to each of real and assumed nodal point of a cornea, N_1 and N, is quantified in Table 4.2 [36]. It shows that N is slightly displaced toward the retina from N_1. Therefore, the rotational radius of N around the eyeball center C is a bit smaller than that of N_1. As mentioned previously, although the tracking of N itself is advantageous to the system, there is concern about a poor experience due to this difference. This is discussed quantitatively in Section 4.4.

Table 4.2. Parameter values depending on each type of nodal points.

Nodal point types	Distance from the vertex of cornea V	Rotational radius of N around C
Real, relax: N_{1r}	$VN_{1r} = 7.0$	$r_{N_{1r}} = 6.2$
Real, acc.: N_{1a}	$VN_{1a} = 6.5$	$r_{N_{1a}} = 6.7$
Assumed: N	$VN = 7.8$	$r_N = 5.4$

4.3 System description

We describe the novel methodical way of system design to accelerate the process from eye's nodal point tracking to presenting the parallax images corresponding to its position.

The overview of system configuration is shown in Figure 4.3. The refresh rate of eye cameras and displays are assumed to be fast. There are as many as six lights to improve accuracy [42], and they are arranged in two rows for ease of handling.

In the following section, the detailed way of image processing to enable both accuracy and speed, proper rendering, and the way to calibrate the entire complex system are described. We also discuss how to estimate the overall system delay.

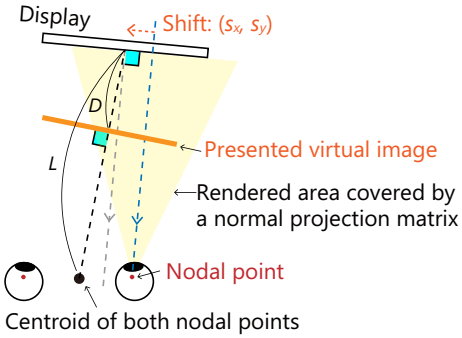

Fig. 4.5. Schematic of binocular parallax image rendering corresponding to the corneal sphere center N.

4.3.1 High-speed image processing for eye nodal point tracking

An overview of the fast image processing for tracking of a corneal sphere center N is shown in Figure 4.4. The used camera for each eye is a stereo because it enables linear calculation and does not necessitate the pre-calibration of any eye parameters (e.g., eyeball diameter, corneal curvature, and corneal refractive index) [40]. Each eye camera can capture the whole eye within the full angle of view.

The pupil area, which is black by the dark pupil method using infrared lights, is extracted by binarization with fixed threshold, and fast ellipse fitting through a least-squares method [92] is performed on the contour point clouds to obtain a rough idea of the area of corneal reflections. Since our study does not require eyesight measurement, no correction and noise elimination on the contour points are necessary; this strategy helps the acceleration. Self-windowing is used for the acceleration [64]; region-of-interest (ROI) is set centered at the ellipse center in the previous frame if available (red solid line); otherwise, its size and position are manually adjusted beforehand (red dashed line) and the ROI image is resized during binarization.

Then, corneal reflections (CRs) are detected using adaptive threshold, where self-windowing is also used and ROI is set centered at the center of the pupil ellipse (green solid line). If the number of CRs is assumed to be same as the number of infrared lights (six), IDs are assigned for each in a clockwise direction. Regarding the once-detected points, they can be assigned the same IDs in the next frame by searching the closest point.

Using these detected points, a corneal sphere center N is calculated as a general way of PCCR [40] (because of complexity, the detailed explanations is omitted in our paper). Hence, all positions of infrared lights in relation to the cameras should be known beforehand, whose calibration way is explained in Section 4.3.3.

4.3.2 Rendering

Here, we describe binocular parallax image rendering based on the detected nodal point. Because the previous studies used a HMD, it was sufficient to just shifting the viewpoint position from the eyeball center C to the corneal nodal point N_1 using the estimated gaze [36, 166]. Meanwhile, our study takes a different approach since the whole system is newly designed.

Virtual object position

Figure 4.5 shows an example of when that line length is L and the virtual image protrudes at a D distance from the display. The position of virtual image is assumed to be on the line passing through the display center and centroid of both nodal points, and its plane is orthogonal to that line

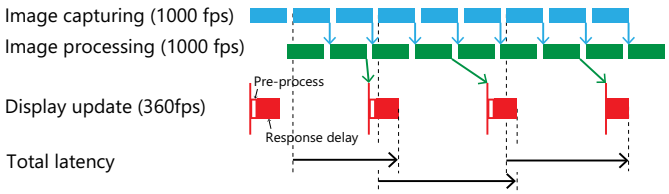

Fig. 4.6. System time flow.

(black dashed line). The display pose from the nodal point can be obtained using the relationship between camera and display, which can be obtained through calibration (explained in the next section).

Projection matrix calculation

A necessary adjustment is described so that the virtual image rendered on the display can be always properly observed by the user.

The 3D rendering normally uses a model, view, and projection matrix, called MVP (model-view-projection) matrix, to convert each vertex to be rendered to the viewpoint-origin coordinate system. Among several projection matrices, we use perspective projection matrix to facilitate realistic viewing. The most problematic point is that the target drawn plane (a display surface) is not in front of the viewpoint, while the projection from a viewpoint covers only the front area of the viewpoint, as shown in a yellow area. Therefore, a shifting matrix should be introduced to shift the rendered virtual image to the center of display; after multiplying MVP matrix to each vertex, the shifting matrix is multiplied to it.

The shifting matrix can be obtained as follows. Using a line passing through the nodal point and having the display normal as a vector (blue dashed line), the positional deviation in the display plane coordinate to display center is first calculated (an orange dashed arrow). Then, those values are converted into spatial units, then each of x and y coordinates becomes s_x and s_y. The shifting matrix firstly consists of 4x4 identity matrix; then, its x and y translational values, the elements of $(1, 4)$ and $(2, 4)$, are replaced to s_x and s_y, respectively.

4.3.3 System calibration

It is difficult to locate the infrared lights from the eye camera because they do not appear within its narrow angle of view. It is also difficult to make the correspondence between the eye camera and a display because they do not confront. Therefore, as shown in Figure 4.3, a wide-angle world camera is introduced to calibrate the entire system, which is stereo to improve the accuracy in depth direction.

To determine the position of infrared lights, not directly detecting the bright light area [109], but fixing them on a flat board and pasting an ArUco pattern sheet [187] would work well. The world camera detects the posture of display by presenting ArUco pattern on it. For confronting camera calibration between world and eye cameras, a board where ArUco pattern is printed on both sides is used. Note that the thickness of the board should be taken into consideration.

4.3.4 Overall time flow

The system time flow is shown in Figure 4.6. Image processing and display rendering are separated in threads for the acceleration. When the update frequency of image processing and that of the display are different, the delay from the end of image processing to the beginning of displaying varies significantly depending on the timing, which should be considered when estimating the total system latency.

4.4 Estimated effect in ocular parallax rendering

This section quantitatively discusses the ocular parallax effects based on the eye and display relationship, and the approximation of tracked nodal point N is verified. The necessary stability of the eye tracking is also estimated, which is referred when assembling the actual system.

4.4.1 General effect of ocular parallax rendering

To faithfully present stereoscopic images corresponding to the moving eye principal point, the parallax images are required to be displaced accordingly, as shown in Fig. 4.7 (a). This is the general effect of ocular parallax and is the most significant difference from conventional parallax image presentation, where the viewpoint is assumed to be static at the eyeball center C.

For the virtual image point P to be always presented at the intended position, the drawn point on the display Q must be the intersection between a display plane and the straight line passing through the user's viewpoint position N and virtual point P. Due to the involuntary eye rotation, viewpoint position N rotates θ around the eye center C and shifts to N'; then, the point Q on the display should be shifted to Q'. However, the conventional binocular disparity image does not correspond to this and only stays at Q. Whether this difference is perceived by the human eye should be discussed in terms of a human visual motion threshold; it was shown to be less than $20''$ at the 0 degree of retinal eccentricity and over $50''$ at the 10 degrees in the experiment for two subjects [188].

Let this difference be $\varphi = \angle QN'Q'$, which apparently has different values depending on the values of L (the distance from the display center to the viewpoint), D (that to the virtual image point P), and θ. The analysis on φ was performed by MATLAB 2021α.

Condition

Assuming that L is fixed to $L = 2.5$ m (same value as the actual experiment), the value of φ can be obtained for each D. Let D be signed, $D > 0$ if P is on the user's side from the display and $D < 0$ otherwise, and satisfy $-2.5 \leq D \leq 2.5$ m. The eye rotation angle θ is assumed to be caused by a saccade and three types of involuntary eye movement during fixation (i.e., tremor, drift, and flick). Saccades vary in size [189, 190], but here 6° is assumed, which is almost the same size as the ellipse presented in the user study. Meanwhile, the three involuntary eye movements are assumed to be $3'', 6'$ and $20'$, respectively. The rotational radius of N is 5.408 mm as taken from Table 4.2. The value of motion threshold takes two standards: $20''$ and $50''$.

The disparity α is also calculated for each D assuming the interocular distance be $D_{eye} = 63.5$ mm. Furthermore, as a verification on L over a wide range of $0 \leq L \leq 2.5$ m, the threshold disparities $\tilde{\alpha}$ where the ocular parallax would be perceived in the saccade of 4–6° and in two standards of motion thresholds were calculated.

Results

Figure 4.8 (a) shows the results of this simulation of fixated $L = 2.5$ m. All three types of involuntary eye movements are only noticed at very close to $D = 2.5$ m, where the disparity surpasses the two standards of motion threshold and is not general stereopsis. In contrast, in the 6° saccade, it can be considered that the ocular parallax would be perceived from $D = 0.7$ and 1.2 m for each standard of $20''$ and $50''$ motion thresholds, respectively, and their disparities are $\alpha \leq 1.5^\circ$, which is within the stable fusible range in, as shown in Figure 4.2 (b). In the negative disparities, only the 6° saccade is slightly over the motion threshold of $20''$ at $D \leq -1.5$, where α is close to be 0°. Therefore, ocular parallax effect is considered to appear due to the saccades from the general fusible range.

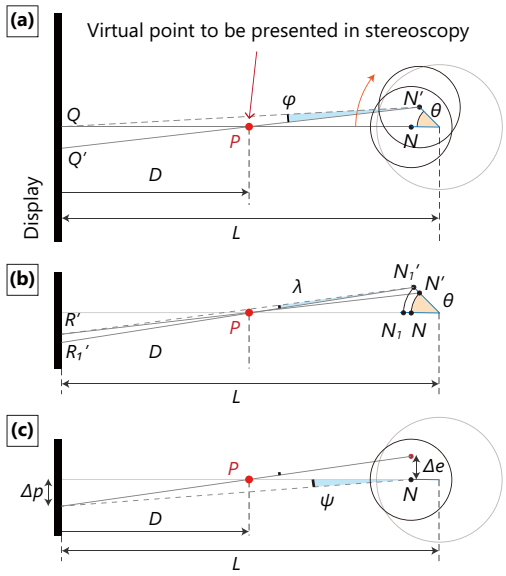

Fig. 4.7. Various ocular parallax effects considered in our study. (a) General effect. (b) Effect from the difference between ideal and tracked nodal points, N_1 and N. (c) Display shift Δp corresponding to the eye tracking error Δe.

The results of verification over the wide range of L are shown in the lower part of Figure 4.8 (c). The disparity where the stereoscopy would be perceived without discomfort in $L > 0.5$ m is calculated to be $0.5° \leq \tilde{\alpha} \leq 1.3°$ and $1.5° \leq \tilde{\alpha} \leq 2.3°$ for the lower and upper standard of motion threshold respectively. These range would be considered within the stable fusible range shown in Figure 4.7 (b).

4.4.2 Difference between ideal and assumed nodal points

As mentioned in Section 4.2, since the tracked corneal sphere center N and the corneal first nodal point (= actual viewpoint position) N_1 are slightly different, the effect caused by such a difference should be investigated.

A schematic diagram of the resulting misalignment is shown in Figure 4.7 (b). When the eye rotates θ, R'_1 should be presented on the display based on the position of N'_1 to precisely present the virtual point P, but in our study, R' determined by the position of N is actually rendered. Similarly, its difference is expressed as $\lambda = \angle R'_1 N'_1 R'$ and whether or not it would be perceived can also be judged by the motion threshold. Similarly, λ has different values depending on different L, D, and θ.

The result of the λ study for both relaxed and accommodated N_1 under the same conditions as before is shown in Figure 4.8 (b). Note that the rotational radius of each nodal point is referred from Table 4.2. It was shown that, for both types of N_1, λ due to three types of involuntary eye movements during fixation is similarly not perceived over a wide D range. In contrast, λ due to saccade is assumed to be perceived around $D = 1.5$ and 2.0 m for each of lower and upper standard of motion thresholds respectively, but not for $D \leq 0$.

A similar verification on the wide range of L is performed for accommodated N_1, which is plotted in the upper of Figure 4.8 (c). At the lower standard of motion threshold ($20''$) and in the condition of $L > 0.5$ m, the effect of difference between N_1 and N would appear from $\tilde{\alpha} = 2.5°$; however, since this standard is of $0°$ retinal eccentricity, it is a small value with high sensitivity to change and can be applied for only the case of fixating. In contrast, at the upper standard ($50''$), in the peripheral field in the other words, it appeared from $\tilde{\alpha} = 6°$, which is considered

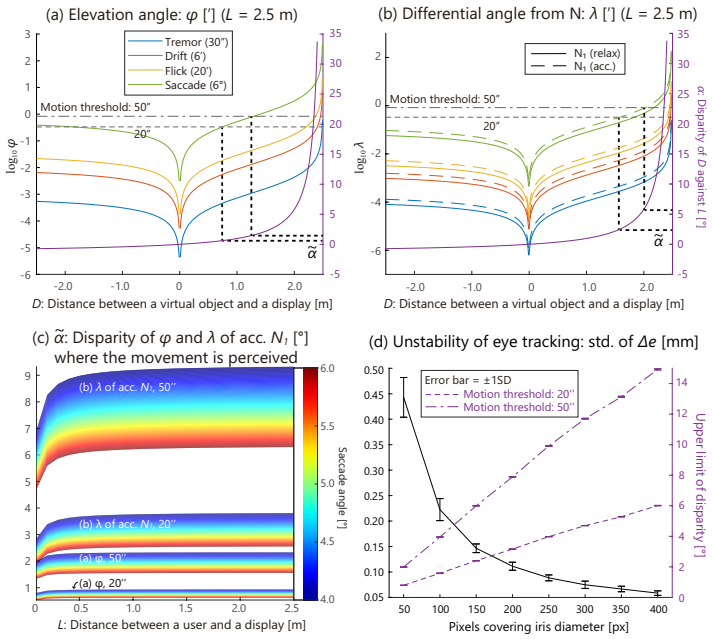

Fig. 4.8. Simulation results estimating the ocular parallax effect. (a) Magnitude of φ when the parallax image is not corrected according to the ocular parallax (Section 4.4.1). (b) That of λ due to the difference of N and N_1 (Section 4.4.2). (c) The estimated disparity where the ocular parallax would be perceived depending on a saccade angle for φ and λ of when N_1 is accommodated. (d) The instability of eye tracking, expressed as std. of Δe (left axis), and the upper limit of disparity where the eye tracking instability is not perceived (right axis) depending on the pixel resolution on the eye (Section 4.4.3).

relatively large disparity and not general stereopsis. From these results, it can be considered that this difference would appear only when fixating from the disparity slightly over the stable fusible range, which may be ignored in the normal stereopsis.

4.4.3 Effect of the stability of eye tracking

The accuracy and stability of the eye measurement are important in the system and should be examined. While the former totally depends on the accurate calibration, the latter depends on the measurement process. In this section, we focus on the latter.

The eye tracking used in this study, PCCR [40], assumes a simple eye model consisting of two spheres [43] (Figure 4.2 (a)). However, the actual cornea is considered to be aspherical in shape [103, 105, 104]; strictly, an elliptical shape whose horizontal axis is slightly longer than the vertical one [191]. While the assumption of the simple eye model makes the calculation easier, the shape difference may lead to unstable measurements; for example, the solution may fall into multiple local optimal solutions near the global optimal solution.

In a display system that reproduces ocular parallax, the eye-tracking results are reflected in the quality of the parallax image on the display accordingly. If the instability is large enough to be perceived by the user, it can be considered that the ocular parallax has not been reproduced accurately. Figure 4.7 (c) shows a rough understanding of the effect of eye-tracking error Δe to ocular parallax rendering. If Δe happens, a shift of Δp also occurs on the display, whose value is determined using L and D. The angular difference amount ψ expresses such a difference, and whether it is perceived or not is an index for stability and can be similarly judged by the motion threshold. Subsequently, the upper limit of the disparity where the image can be perceived without

Fig. 4.9. Experimental apparatus. (a) Appearance of the entire system. (b) shows it from the user's position. (c) Display presenting RDS with random dots backgrounds (Figure 4.1 (c) is that with white backgrounds). (d) World stereo camera.

discomfort is similarly calculated.

To improve stability, one way is to use multiple infrared point light sources, which can inherently enhance the stability [42]. The other is increasing the image resolution, whose effect is evaluated by following simulations in MATLAB 2021α.

Condition

The image resolution was considered as the number of pixels defined to be 50,100,150,...,400 px that covers the iris diameter, $2r_i = 12.0$ mm. This simulation assumed such a system, as shown in Fig. 4.3. Let the midpoint of the stereo camera be P_{cam}, and both were placed D_{cam} apart in the x axis direction. Let L_w and L_h be the width and height of the group of infrared light sources, P_L be the center of gravity of the lights. Let P_{eye} be the midpoint of the both eyeball center, which were separated by a interocular distance of D_{eye} in the x-axis direction as well. The eyeball center $\boldsymbol{C}$ was defined as the position of $VC = VN + r_N = 11.408$ mm apart from the vertex of anterior corneal surface V. The value of each parameter was defined in Table 4.3. The cornea shape follows an ellipsoid equation [103, 104, 105];

$$x^2 + y^2 + (1+q)z^2 = r_c^2, \tag{4.1}$$

where $q = -0.33$ and $r_c = 12$ mm [112].

In the simulation, the eyeball rotated at θ around the Y-axis of its center C in a range of $-10 \leq \theta \leq 10°$ with a step of $1°$. Then, the reflected position of each infrared light on the corneal surface was calculated in each eye posture. To estimate the stability, each position was randomly displaced at 1 px of each image resolution in random the direction, and the estimation of the corneal center in that condition was performed. This procedure was repeated 100 times, and the standard deviation of the estimation results for all trials was then measured as an index of stability (std. of Δe). This was calculated for each rotation angle, and its statistics over all the rotational angles were obtained.

The upper limit of disparity in each image resolution was similarly calculated using an interocular distance D_{eye}, a lower standard of motion threshold ($20''$), and the resulting values of std. of Δe, with varying L within $0 \leq L \leq 2.5$ m. The statistics of the upper limit of disparity was similarly taken over all the range of L for each image resolution.

Results

The result is shown in Figure 4.8 (d). As shown in the left axis, the higher the image resolution, the higher the stability, and it appears to be an inversely proportional relationship. The upper limit

Table 4.3. Parameters of the system assumed in the simulation. [mm]

Camera		Light		Eye	
$\boldsymbol{P}_{cam}$	$[-100, 0, 0]$	$\boldsymbol{P}_L$	[0, 50, 300]	$\boldsymbol{P}_{eye}$	[0, 50, 500]
D_{cam}	67	L_w	180	D_{eye}	63.5
		L_h	200		

of the disparity where the stereoscopy can be presented without discomfort is shown in the right axis; the higher the image resolution, the higher disparity limit was obtained.

If the upper limit of disparity is 2° (a realistic value), the image resolution of the cornea is required to be about 100 px. In this study, we investigate larger disparities, up to 5.5°, which requires a resolution of about 250 px. Although this is a strict value calculated using the lower motion threshold, this will be an index for the worst case; if the resolution is lower than this, there is a possibility that the user experiences instability in the stereoscopic presentation, and faithful ocular parallax may not be realized.

From these results, it can be concluded that higher resolution is preferable. While a latest high-speed camera has about 1M pixels [192], those used in this study have about 200K pixels. It can be considered that our camera satisfies the needed conditions for the faithful reproduction of ocular parallax.

4.5 Evaluation Experiment

4.5.1 Apparatus

The experimental setup is shown in Figure 4.9.

For eye tracking cameras, two Ximea MQ013RG-ON and two Basler acA800-510um, were used. Both were monochrome, and their ROI was cut to 512×410 px to have an acquisition rate of 1000 fps. Anaglyph was used to present binocular image, and a red dichroic color filter (Edmund Optics) and a cyan frame of commercial 3D glasses (Amazon) were used as filters. Both filters were selected from among a number of filters in the points that (1) completely separated the red and cyan random dots through these filters and (2) had high transmittance of infrared light. The Ximea cameras (infrared wavelength enhanced) corresponded to the right eye with the red filter where the infrared light intensity tends to drop, while the Basler cameras to the left with a cyan filter.

The six infrared lights were provided by Intelligent LED Solution (ILH-IW01-85SL-SC221-WIR200, 850 nm). They were arranged in a horizontal row, as shown in Figure 4.9 (b), and its height and width, L_h and L_w, were 18 and 20 cm, respectively.

The display was an LCD gaming monitor (Acer XV252QFbmiiprx, 1920×1080 px, 360 Hz standard, 390 Hz overclocked; latency of GTG (gray to gray): 1 ms standard, 0.5 ms overdrive at minimum). It was overclocked and overdriven for maximum performance. As a computer, GALLERIA ZA9C-R38 (Windows 10 Pro, CPU: Intel Core-i9 12900K, GPU: GeForce RTX 3080, RAM 64GB, SSD 2TB) was used. The world stereo cameras were from the Imaging Source Europe GmbH (DFK38UX253 with 4096×3000 at 30 fps and DFK38UX541 with 4504×4504 at 18 fps) (Figure 4.9 (d)). The entire system was covered by a black curtain.

The distance from the user's eye position to the display surface L was approximately 2.5 m, which was determined so that the pixel resolution of the display does not exceed the motion threshold. To fix the head position and set almost the same eye position for any user, a chin rest with an extendable height, an acrylic plate to press the forehead against, an occipital fixation bar, and a peephole were provided, as shown in Figure 4.9 (b). The peephole is particularly useful to set the head uniformly for a different head height and interocular distance.

Table 4.4. Statistics of display frame rate and response delay.

	Ave.	Std.
Frame rate [Hz]	359.827	109.232
Response delay [ms]	1.554	0.440

4.5.2 Evaluation of the system performance

Display frame rate and latency

To confirm the actual values, the frame rate and response delay of the display were measured using a high-speed camera: Photron FASTCAM SA-X2 (1024×672 px, color, 20k fps). In both measurements, a total of 20 red and cyan horizontal bars were presented to be equally spaced and alternately in the vertical direction on the white screen.

For the frame rate measurement, the process to render those bars and whiting out in the next frame was repeated. The video was recorded in approx. 2 s, from which the refresh rate was measured.

The response delay was measured as follows. After sending the command to render the borders, the microcontroller immediately sends a command to turn on the LED. Then, the appearance of a changing monitor and a lightening LED is recorded. Assuming that the commanding time is instant, the time from when the LED lights up to when the display completely rendered was regarded as the response delay. This procedure was repeated for 12 times within the 2 s recording window, and the statistics was calculated.

The results are shown in Table 4.4. The raw data of frame rate show that most of the frames were 410 fps, but there were some frame drops, resulting in an average of 360 fps with a large variance. It is considered that overclocking did not work well. The response delay was larger than that for GTG (0.5 ms at minimum) and therefore could be regarded as an appropriate value.

Processing time

The image-processing time for eye tracking, including the image acquisition, was measured using a video of 7000 frames of certain subject who participated in the following two experiments. The process time for the preparation for display rendering, described in Section 4.3.2, was also measured using 5000 frames data.

The statistical results are shown in Table 4.5. The total time of image processing is approximately 1 ms, which is sufficient to achieve 1000 fps, although the image-acquisition time appears to be relatively large. From the appearance of system operation, this image-acquisition time seems to include waiting time for the next frame. Therefore, although four images are being processed simultaneously per frame, there still appears to be a margin of time for computation. The time for rendering preparation is almost close to 0. These results can be attributed to a high-performance CPU.

Estimated total system latency

Using the parameters from Tables 4.4 and 4.5, the total system latency is roughly estimated to be 4.072 ± 0.2886 ms. Because the timing of both threads of image processing and display rendering is different, as explained in Section 4.3.4, this calculation is conducted by averaging the delay times for over 1000 display frames. It is notable that the variance is not too large. Note that this result is only limited to our study, where the anaglyph color (red and cyan) is presented on the white window, and not applicable for all the presentation.

Table 4.5. Processing time per frame. Ave. (Std.) [ms]

Image proc. all	0.9987 ($\pm$0.1818)
- Image acq.	0.4506 ($\pm$0.1445)
Rendering prep.	0.0181 ($\pm$0.0056)

4.5.3 User study on binocular fusion

Ocular parallax effects in binocular fusion were investigated. By presenting an RDS with random dots background (Figure 4.9 (c)), pictorial cues were completely excluded, and only the convergence work for the depth perception could be examined.

Stimuli

A uniform-density RDS was presented on the display in the anaglyph method. An ellipse would float with stereoscopic fusion, and its field-of-view (FOV) from the user's viewpoint was always 7.0 and 6.0° in the x, y directions, respectively, which covers the 6° saccade. The size of the ellipse is set to be always the same regardless of the protruding distance, which disables the pictorial size cues. There were 1800 points within the RDS ellipse, and the random dots on the background, indicating that the outside of the ellipse area had the same density. Thus, from the other side rather than viewpoint, a two-color random-dot ellipse shifted left and right, respectively. Further, a background consisting of black dots were observed, as shown in Figure 4.9 (c). RDS drawing was regenerated for every trial.

Condition

There were nine disparities: -0.75, -0.5, -0.25, 0.5, 1.5, 2.5, 3.5, 4.5, and 5.5°. Combined with and without (w/o) ocular parallax rendering, 18 different experimental conditions were used. Assuming that $L = 2.5$ m and $D_{eye} = 63.5$ mm, the protruding distance D for each disparity could be calculated to be in the range of -2.66 $\leq D \leq$ 1.98 m, which were thought to be appropriate values.

There were five trials for each condition, and therefore $5 \times 18 = 90$ trials in total. Regarding the trials order, to ensure that the order was not biased, five sets where 18 conditions were shuffled were prepared and connected.

Subjects

Twelve subjects (aged 21–35 years), having good vision at 2.5 m distance with the naked eyes or with contact lenses, were recruited by snowball sampling. All were Japanese, except one mixed-race Thai-Japanese woman and one Indian woman, and half were male and half female. All had black or brown eyes.

Procedure

The subject was on a seat, placed his/her chin on the chin rest, and pushed his/her forehead against the acrylic plate. The experimenter (= the corresponding author) adjusted the height of the chin rest and the eye camera's ROI so that his/her eyes were within peephole and observed in the center of the angle of view of each eye camera and then fixed the back of the head with a bar. After adjusting several eye-tracking parameters (ROI on the initial frame, shown as a dashed red line in Figure 4.4, binarization threshold for ellipse detection, etc.), the eye tracking was tested. Further, the corneal sphere center N at a certain frame was obtained, which was used as the viewpoint to generate the RDS in the condition of without the ocular parallax.

The process per each trial is as follows. First, infrared lights were turned on, and the experi-

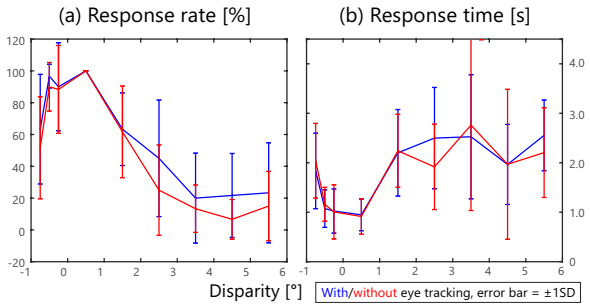

Fig. 4.10. (a) Statistical results of the (a) response rate and (b) response time of when the subjects answered. Both of them were calculated across all subjects data.

menter checked whether the eye tracking is actually working. Then, the RDS was generated and presented for up to 5 s, and the user immediately pressed a specified key when he/she perceived the ellipse image floating, as instructed beforehand. If the key was pressed, the presentation continued for additional 2 s and then disappeared. If not, it disappeared immediately after 5 s of presentation. Then, the screen blacked out, the infrared lights down, and a 3-s break followed.

Basically, there were no breaks other than a minor 3-s break, but the subjects could request long breaks. A few subjects actually took them, each lasting a few minutes. After these breaks, the series of initial adjustments described above were made again before resuming.

A total of 10 practice trials were conducted prior to the main experiments. Subjects who did not fully understand the experiment and requested more practice were given another 10 practice trials.

Since most eye-tracking errors happen during blinking, the screen was made to white out and turn off the RDS when blinking. It was implemented, regardless of the rendering corresponding to the ocular parallax. The downside was that the screen would continue to white out when eye tracking continued to fail, making it difficult to respond. Subjects in such cases were instructed to report immediately. After the report, the eye tracking was checked, and the same presentation was made again.

Analysis

For each subject and each condition of disparities and w/o ocular parallax rendering, the response rate (the proportion of trials that subjects could perceive the floating ellipse stimulus) among five trials was calculated. Then, the statistics for each condition were calculated across all subjects. The same process was also done for the time of only when responded.

Generalized linear mixed model (GLMM) [193] was used for the statistical analysis for both response rate and time, where the binomial and normal distribution with logit link were assumed, respectively. Both of them were performed across all the subject's data by taking the disparity and w/o ocular parallax rendering as two factors and the subject as a random effect. The analysis for response time used only the data of when responded; hence, the number of data for each condition differs.

If the analysis on GLMM has a significant value, post-hoc tests were conducted as pairwise t-tests between w/o ocular parallax at each disparity, with Bonferroni correction applied to the p-values.

Results and discussion

Figures 4.10 (a) and (b) show the statistics for response rate and time across all the subject's data, respectively. The individual data are shown in Appendix A. The response rate gradually decreased as the absolute disparity increased. However, the response rate was higher with our ocular parallax

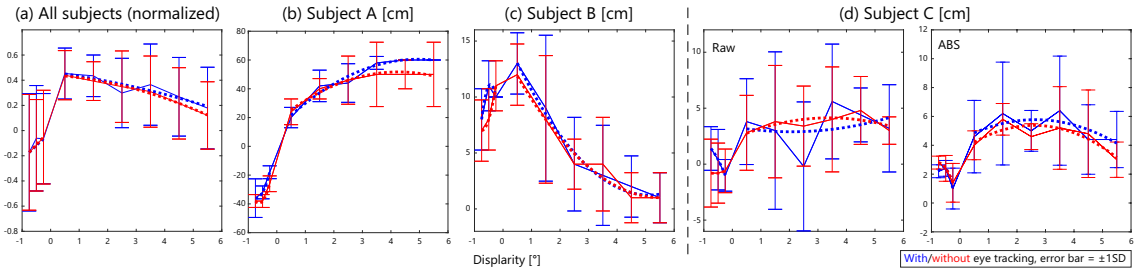

Fig. 4.11. Statistical results of the answered perceived depth depending on w/o ocular parallax rendering. (a) shows all subjects using each normalized data, while (b)-(d) show certain individual subjects. (d) shows both raw and absolute (ABS) graphs, while the others only show raw graphs. Quadratic curve fitting for each data of positive and negative disparities is shown as bold dashed lines.

Table 4.6. Summary of post-hoc t-tests results between w/o ocular parallax at each disparity, with Bonferroni correction applied to the p-values. Those p-values with a significant difference are noted with * mark.

Disparities [°]	-0.75	-0.5	-0.25			
p	0.1089	0.1590	0.5681			
Disparities [°]	0.5	1.5	2.5	3.5	4.5	5.5
p	NaN	0.8207	0.0093*	0.2888	0.0056*	0.1328

rendering in those conditions where disparity was equal to or greater than 2.5 °. This exceeds the estimated disparity limit of $L = 2.5$ m, where the effect of ocular parallax appears, as shown in Figure 4.8 (c). Although $0.5°$ was also included at minimum in the simulation of Section 4.4.1, this statistical result showed no statistical difference.

The GLMM analysis for each response rate and time focusing on the factor of w/o ocular parallax rendering was $\chi^2(8) = 21.02, p = 1.255E-2$ and $F(1, 587) = 1.7185, p = 0.2938$, respectively, where only the former showed statistically significant difference. In contrast, the interaction of the response rate was $\chi^2(8) = 5.359, p = 0.7186$ without statistically significant difference. Post-hoc t-tests were conducted only for the response rate, and its results are shown in Table 4.6. Only the disparities of 2.5° and 4.5° have a significant difference, where the ocular parallax rendering has a significant improvement over the conventional rendering.

The individual data show that some subjects showed a large and statistically significant difference in response rate, while others did not. It can be estimated that there were individual differences in the effect of convergence on binocular fusion. Originally, some subjects had difficulty in binocular fusion especially in larger disparities than $1.5°$. Since the effect of ocular parallax seems to appear from 2.5°, these subjects should not show any difference.

Therefore, high-speed ocular parallax rendering seems to have a significant effect on facilitating the binocular fusion even with the approximation of tracked nodal point N, but does not hasten it. This is believed to be effective for cases where convergence originally works well for binocular fusion. Note that the remaining results for the response rate and time are shown in Appendix C.

4.5.4 User study on perceived depth

An experiment on the verbal answer of perceived depth was conducted to examine the effect of ocular parallax on it.

Stimuli & Condition

The experimental conditions of the disparity and w/o eye tracking, RDS stimuli, experimental apparatus were the same as those in Section 4.5.3. However, the RDS background was white, as shown in Figure 4.1 (c), which reduced the difficulty of binocular fusion for any subjects.

Subjects

Subjects were recruited as the same way as in Section 4.5.3, and there were 11 black-eyed Japanese, aged 22–25 years, including two females. One of them was excluded due to an error in the experimenter's filling in the answered depth values.

Procedure

Several procedures, such as advance adjustment, large breaks, and practice before the main experiment, were the same as those in Section 4.5.3.

After the experimenter confirmed eye tracking, the RDS image of an ellipse was presented for 5 s, and the subjects were asked to answer verbally the numerical value of its perceived depth from the display in cm, as well as the direction against the display (front/back). Subjects were instructed to answer 0 value if no depth was perceived or a double image was observed. The presentation continued for 5 s even if the response was finished. After the screen was blacked out, the infrared light were turned off. After completing the response, the answered value was recorded with direction: the front as positive and the back as negative. Then, followed by a 3-second break.

The reason for using a verbal answer way, despite that a previous study used a hand-reaching way [36], is because the display is far from the user to generate faithful ocular parallax and it is difficult to reach out to the virtual image. Subjects were instructed to perceive the depth solely from the relationship between the display and the RDS as much as possible to avoid the usage of external cues for depth perception existing even in the dark room, such as a table edge.

Analysis

For each subject and each condition, the statistics was taken for the numerical values of the perceived depth. To take statistics across all the subjects, the values normalized by the maximum absolute value for each subject were used. For the data of each positive and negative disparity, quadratic curve fitting was performed to determine the overall trend. The analysis was conducted for raw data and absolute values.

GLMM was conducted for both answered depth values and its signature (direction) with the same condition as in Section 4.5.3, where normal and binomial distributions with logit link were used, respectively. The latter was conducted examine ocular parallax effect on the confused direction, and only non-zero data were used.

Results

Figure 4.11 (a) is a graph of statistics for all subjects. From the quadratic curve fitting, shown as dashed lines, the overall trend was assumed that the ocular parallax slightly improved the perceived depth in the disparity of 2.5° or greater. However, the result of GLMM showed no difference ($F(1, 882) = 6.873E - 05, p = 0.9934$). GLMM to examine the effect of ocular parallax rendering on the answered direction also showed no difference ($\chi^2(8) = 4.718, p = 0.8582$).

Certain subject data are shown in Figure 4.11 (b)-(d). Any answered depth value was much smaller than the actual depth, which is possibly due to a large deviation from accommodated cues and unstable binocular fusion, as well as the presentation of RDS with almost no pictorial cues. Furthermore, some subjects made a mistake in directions. Figures 4.11 (c) and (d) are such examples of the confused directions. For (d) subject C, the absolute values of the answered depth seems to be improved with an ocular parallax, but the raw data do not. For (c) subject B, all the

data were answered to be positive. Meanwhile, (b) subject A is an example of answering relatively large values (max. 80 cm) with correct directions. Any of them had greater results with ocular parallax rendering in the range of large values of either positive or negative disparity.

The remaining graphs are shown in Appendix B. The results of analysis not written here are shown in Appendix C.

Discussion

The effect of ocular parallax rendering on perceived depth seems to exist slightly in the graph, but no difference appeared in the GLMM analysis. Similarly, the answered direction also did not show a statistically significant difference. From the individual data, only some subjects showed a slight difference in the graph, especially in the absolute graph. Therefore, similarly as the effect on binocular fusion (Section 4.5.3), the appeared effect seems to differ from person to person. The cause is considered to be the same as before: individual differences on the convergence work on the perceived depth. From these results, it can be concluded that our high-speed devices is slightly effective for perceived depth. In contrast, no effect in the recognition of direction was observed. It is considered that the convergence is not effective for direction recognition; the blur and accommodation may be preferable rather than it [194].

It should be noted that, although the experimental room was covered with a black curtain, it was not completely dark due to the bright light from the white background of the display; these were in fact poor experimental conditions and should be improved in the future study.

4.6 Discussion and Conclusion

We assembled a high-speed and low-latency stereoscopic viewing device to reproduce the ocular parallax faithfully and investigate its effect on depth perception. The system was carefully designed to be fast, stable, and accurate using simulations, and its performance was evaluated to be sufficient. A user study using this system was conducted on the binocular fusion and perceived depth while presenting versatile scene with RDS, where improvements were observed with ocular parallax rendering. Therefore, our high-speed device contributed to presenting the positive impact of ocular parallax even with an approximation on the tracked nodal point N.

Our novel system can be used in future studies to investigate ocular parallax effects from multiple perspectives, for example, the allowable system delay where the ocular parallax effect still appears, and its effect on the vertical retinal image error. These studies will facilitate more immersive VR/AR experience. From the application standpoint, the proposed system cannot be commercialized anytime soon; its size should be reduced significantly so that it becomes as small as a wearable device such as HMDs.

Chapter 5

Model-based Identification of Multiple Corneal Reflections For Eye Tracking

5.1 Introduction

Eye tracking technology has played an important role in multiple situations such as psychophysical experiments, human-computer interaction, virtual reality, and many other fields. The most popular eye tracking method is called pupil-centered corneal reflection (PCCR) [40], which uses a set of infrared point light sources and accurately estimates the gaze using the corneal reflections (CRs). As having already used in various products [35, 33, 185], this method is popular for its high accuracy, especially with increased number of CRs [42]. However, identifying multiple CRs is difficult especially when the eye rotates largely and some CRs are easily kicked and lost. Therefore, an identification algorithm of CRs corresponding to wide eye rotation must be designed.

Some previous studies proposed the identification method of multiple CRs. One study performed identification of 9 CRs using approximated relationship between pupil-elliptic shape and group of CRs [119] based on LeGrand's simple eye model [43]. However, due to such an approximated formula, the allowable rotational angle is limited to within $\pm 10^\circ$. The LED-switching method [113, 115, 118] can use only limited number of CRs and therefore inevitably drops the accuracy. Furthermore, originally, PCCR assumes a simple eye model consisting of two spheres, but the actual eye has of course a complicated shape; therefore, changing available CRs frame by frame may lead to unstable measurements. The machine-learning-based method realizes high accuracy within $\pm 15^\circ$, but it requires manual labeling of a large number of images [126].

This study proposes an identification method of multiple CRs based on a simple eye model but without any approximation unlike Li et al [119]. This method is designed to identify even in a situation that some CRs are kicked and missed; the valid candidates are enumerated and the most valid one is efficiently extracted. The cost function is designed to have small calculation lost, which effectively utilizes the assumption of PCCR and realizes fast identification within one frame.

By the evaluation experiment, our method was validated to work within $\pm 20^\circ$ eye's rotational angle and have high accuracy even with lacking observed CRs, but there was a large standard deviation, that means, a large individual difference. Furthermore, our method was evaluated to be high speed as well as 1000 fps.

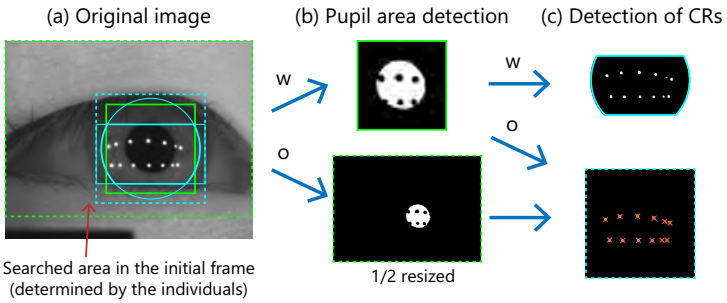

Fig. 5.1. Flow of image processing per frame.

5.2 Prior image processing

This section describes the image preprocessing conducted frame-by-frame to exclude the spurious reflections to the greatest extent possible. A schematic is shown in Figure 5.1.

5.2.1 Pupil detection

First, the pupil area is detected using the dark pupil method to roughly obtain the area of CRs on the corneal surface and to calculate the exact pupil center position used for the eyesight estimation. After binarization using a fixed threshold, the contours along the pupil area are extracted, and their distortion is corrected using the camera parameters which are previously calibrated. Then, ellipse fitting is conducted on the contour point groups using a linear least-squares method [92] by applying RANSAC [195]. As shown in Figure 5.1 (b), some CRs created a hole in some parts of the contour and lead to a large error when an ellipse is fitted to all contour points. Such contours are excluded using RANSAC, which determines only valid points that follow the desired ellipse. This method is also useful when the eyelid and eyelash cover the pupil and the pupil is detected as nearly a semi-circle and when an incorrect area is mixed because of a black eyelash. The center of the ellipse is assumed to be the pupil center on the image and used for the PCCR method [40]. To accelerate the process, the pixel-by-pixel search for precise pupil detection [110] is omitted here.

The search area is restricted by self-windowing [64] for efficiency and acceleration. If the tracking information of the previous frame is available, the search is performed within a limited window centered on the pupil position of the previous frame, as indicated by the green line in Figure 5.1 (a). If no previous tracking information is available (for example, an initial frame), the entire image is simultaneously binarized and resized. The same process of pupil area detection and ellipse fitting is conducted. If an apparently different area is in the entire image (for example, the black plate of the head holder shown in Figure 5.1 (a)), avoiding such an area by limiting the initial ROI (green dashed line) is effective.

5.2.2 Detection of CRs on corneal surface

CRs on the corneal surface are detected using adaptive thresholding from the original image within an ROI centered on the pupil ellipse. Using a limited ROI has the advantage of effectively excluding noisy reflections and acceleration [64]. Typical noisy reflections are extended CRs, glints around the eyelid, and secondary or tertiary Purkinje-Sanson images, as shown in Figure 4.1 (a). In most cases, such incorrect reflections can be properly excluded in the iterative search; however, misidentification sometimes occurs because the assumed simple eye model is essentially different from the actual model. To avoid such cases, excluding such noise using image processing is

necessary.

Because the noise around the CRs has some typical patterns, the ROI should be designed appropriately. Here, the intersection area of the rectangle and ellipse, each of which has an important contribution in properly detecting CRs, is used as an ROI, as shown by the cyan line in Figure 5.1 (a); this is called advanced self-windowing.

The area of an ellipse is centered at the pupil image center and has a multiplied axis with the length of that of the pupil, which nearly covers the overall corneal area and contributes to excluding the extended CRs at the base of the cornea. For extended CRs, the points cannot be determined precisely, which leads to a worse estimation. Therefore, excluding them is necessary. Because the change in pupil size over time is small, having an appropriately fixed value is acceptable for the axes-multiplication ratio. However, because the size of the pupil compared to that of the cornea varies by person, the ratio should be determined individually.

The rectangular area primarily contributed to excluding the upper and lower noisy reflections. Most of these reflections are glints of tears that accumulate at the boundary between the eyelid and eye, and secondary or tertiary Purkinje-Sanson images. Because most of these noisy reflections appear above or below the true CR area, careful design of the upper and bottom bounds is important. With high-speed tracking, the CRs in the current frame are assumed to be located at almost the same position as those in the previous frame; therefore, the upper and bottom bounds of the rectangular ROI are dynamically determined according to the CRs identified in the previous frame; both bounds are offset by around 20 px from the maximum and minimum y coordinates of all the true CRs. This strategy is helpful for tracking the area of true CRs and excluding the noise. If there is only one row of true CRs, whether the row is upper or bottom is recognized based on the assigned IDs, and the offset of the bound with no observed CRs will be extended more (around 100 px). Moreover, for the initial frame, because lacks tracking information, a fixed-sized square ROI is used.

In actual image processing, first CRs are detected within the rectangle area, then those that are not within the ellipse are removed.

5.3 Identification of multiple corneal reflections

This section explains the algorithm used to identify each CR. This is the most important component of the proposed approach, and it consists of grouping the observed CRs, an efficient enumeration algorithm, and a simple cost function. The overall purpose is to reduce the computational load for high-speed tracking without impairing the accuracy. Here, "CR is true" means that this CR is the correct reflection of a certain infrared light.

5.3.1 CR Grouping

The light sources are placed in two rows so that the CRs observed are easily grouped to effectively reduce the number of enumerations. In the case of circular placement, it is difficult to determine the reference point of observed CRs, especially when some CRs are kicked. By contrast, in the case of two-rows placement, the observed CRs are easily divided into two groups and the reference point, such as the right or left point, can be determined.

The first step is to determine whether the CRs are in one or two columns. The number of observed rows is two for most cases but can be one when the eye is directed upwards or downwards. Because of the image preprocessing, most CRs are assumed to be true; therefore, fitting a straight line for all CRs and checking whether this line is a good fit is sufficient for determining the number of rows of the observed CRs. The threshold for a good fit is determined appropriately beforehand.

Next, if the CRs are in two rows, grouping is conducted. This is a simple step: after averaging all CRs, grouping is conducted according to whether the Y coordinate of each CR is larger or

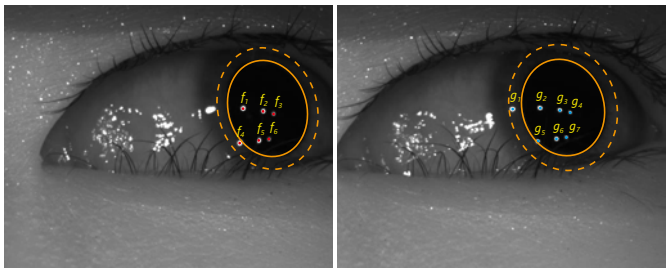

Fig. 5.2. Pupil detection using dark pupil method and corneal reflection detection.

smaller than that of the average position. This calculation assumes that the CRs of each row are nearly in a horizontal straight line. This could yield errors when the CRs are observed as an arc, which easily occurs depending on the distance between the light and eye. In such cases, other methods such as clustering using circle fitting are recommended.

5.3.2 Finding the matched CRs in each image

Next, the CRs matched between each of the stereo cameras are determined. This is another effort to reduce the search cost of identification in the next section. Rather than identifying the CRs in each image, identifying the matched CRs in both images can significantly reduce the search cost. Simultaneously, a noisy reflection observed in one camera but not in another can be eliminated. Moreover, some true CRs observed in one camera, but not in another, are also excluded; however, the advantages of reducing the cost can counterbalance this drawback.

Using Figure 5.2 as an example, which shows the eye images captured by stereo cameras, the matching between each CR is evidently $f_i = g_{i+1}, i = 1, 2, ..., 6$. This procedure is generalized and explained in the following subsection.

Procedure

Matching CRs between two images are simply constructed by focusing on the similarity between two images.

1. As a preprocess, each CR is laterally sorted for each row. Going forward, the procedure is explained as performed for each row.
2. All match candidates are enumerated assuming that there are k matched points. The following section describes the detailed procedure.
3. An appropriate correspondence between CRs is taken, and the cost function that focuses on the translational vector between two images is calculated.
4. Step (3) is performed for all candidates. Take the candidate with the smallest error, and if its error is less than the threshold determined in advance, it is assumed to be a correct match; otherwise, repeat (2)-(4) using $k \leftarrow k - 1, k \geq 2$.

Cost function

This subsection introduces the cost function. The translational vector $T_i = (x_i, y_i)$ where $i = 1 \sim k$, which translates one point to another in image coordinates, is considered. The average value is then calculated as $\bar{T} = \sum_{i=1\sim k} T_i/k$. For each matched point, the cost function is calculated as follows. The differential length between the translational and averaged vectors is calculated as $\Delta e_i = |T_i - \bar{T}|$ for $i = 1 \sim k$. Then, their sum, $\sum_{i=1\sim k} \Delta e_i$, is assumed to be the total error.

This procedure is possible if the stereo cameras are placed near each other and have approximately the same tilt angle. In such a situation, the two images from the stereo camera from slightly different perspectives are almost like each other. If the tilt angles of both cameras are significantly

Table 5.1. Parameters describing Figure 5.2.

i = Camera No.	N_i	N_{ui}	N_{bi}
1	6	3 $\{f_0, f_1, f_2\}$	3 $\{f_3, f_4, f_5\}$
2	7	4 $\{g_0, g_1, g_2, g_3\}$	3 $\{g_4, g_5, g_6\}$

different, additional parameters (such as an affine transformation), can be necessary to correct the rotational difference.

If there is only one point, the transition parameters are uniquely determined, the error is zero for any combination, and the optimal matching cannot be determined. Therefore, k should exceed 1, and a larger number of CRs yielded reliable results. Such a small lower limit is helpful in cases where the eye rotates significantly and number of observed CRs is small. If affine and homographic transformations are used (as they are uniquely determined by three and four points, respectively), the minimum required number of points is four and five, respectively. In such cases, many CRs are observed and are unsuitable for large eye rotations.

Enumeration of match candidates

This section describes the method of enumerating candidates for matching CRs between each image in stereo cameras. The basic idea is to select points of the same number from point clouds in each image.

The case with a single row of point clouds is the simplest, and the number of enumerations is small, because the points in each image are sorted horizontally. If the number of detected points in each camera is N_1 and N_2, and the number of selected points is k, the range of k is expressed as $k = \min(N_1, N_2), ..., 2$. Subsequently, the total number of combinations C_1 with k selected points is calculated as follows:

$$C_1(k) =_{N_1} C_k \cdot_{N_2} C_k. \tag{5.1}$$

If the point clouds are divided into two row groups, the enumeration of combinations becomes slightly more complicated. Similarly, let N_1, N_2 and k to have the same definitions. Furthermore, let k_u and k_b be the number of selected points from each row that satisfy the formula $k_u + k_b = k$, and let $N_{u1}, N_{b1}, N_{u2}, N_{b2}$ be the number of points in the upper and lower rows of each camera that satisfy $N_{ui} + N_{bi} = N_i$ for $i = 1, 2$.

The total number of combinations, C_2, selecting k_u, k_b points from each of the upper and lower points in each camera, is calculated as follows:

$$C_2(k_u, k_b) =_{N_{u1}} C_{k_u} \cdot_{N_{b1}} C_{k_b} \cdot_{N_{u2}} C_{k_u} \cdot_{N_{b2}} C_{k_b}, \tag{5.2}$$

where

$$k_u + k_b = k,\ 0 \leq k_i \leq \min(N_{i1}, N_{i2})\ for\ i = u, b. \tag{5.3}$$

Because Eq. (5.2) excludes some cases that mix the upper and bottom rows, the enumeration number is naturally reduced to

$$\sum_{k_u, k_b} C_2(k_u, k_b) < C_1(k). \tag{5.4}$$

Example

Using the picture in Figure 5.2 as an example, each parameter of N_i, N_{ui}, N_{bi} for $i = 1, 2$ is organized as follows.

The maximum number selected is $\min(N_1, N_2) = 6$. Thus, the candidates for the selected number k are $k = 6, 5, 4, 3, 2$. The range of k_j is $0 \leq k_j \leq 3$ for $j = u, b$. Therefore, the combinations of (k_u, k_b) that satisfy these conditions are as follows.

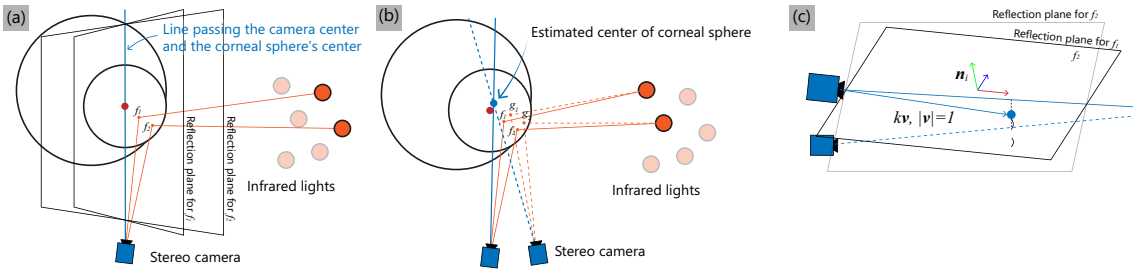

Fig. 5.3. Flow of identifying CRs. (a) The flow of PCCR method to estimate a corneal sphere center when the correspondence between each CR ad infrared light is known. (b) Our identification method leveraging that if each CR corresponds to an incorrect infrared light, the estimated corneal sphere center differs from the true one. (c) When the correspondence has a mistake, the estimated corneal sphere as an intersection point of lines is not on the plane. Such a feature is used for identification.

Table 5.2. Possible combinations of k_u and k_b under a certain k.

k	(k_u, k_b)
6	$(3,3)$
5	$(3,2),(2,3)$
4	$(3,1),(2,2),(1,3)$

Table 5.3. Example candidates of selecting points on each row; $k_u = 3$ and $k_b = 2$ from the upper and bottom rows.

Cam 1	Upper	$\{f_0, f_1, f_2\}$
	Bottom	$\{f_3, f_4\}, \{f_3, f_5\}, \{f_4, f_5\}$
Cam 2	Upper	$\{g_0, g_1, g_2\}, \{g_0, g_1, g_3\}, \{g_0, g_2, g_3\}, \{g_1, g_2, g_3\}$
	Bottom	$\{g_4, g_5\}, \{g_4, g_6\}, \{g_5, g_6\}$

For $k = 5$ and $(k_u, k_b) = (3, 2)$, the matching candidates are enumerated as follows. It can be concluded that there are $1 \times 3 \times 4 \times 3 = 36$ candidates in total.

5.3.3 Identification of CRs

This subsection explains a novel method for identifying detected CRs, which is designed based on the PCCR method [40]. As mentioned earlier, each reflection is assumed to be on the ideal spherical surface of the cornea; thus, a simple reflection equation is assumed. The key point is the center of the corneal sphere, and the novel points in our method are the design of such a simple cost function.

Basic idea of PCCR

Figure 5.3 (a) illustrates the PCCR method. When an infrared light corresponds to a certain corneal reflection point, a plane can be formed that passes through three points: the camera origin, corneal reflection point (CR), and infrared light. When there are two or more infrared lights (and, therefore, two or more CRs), there are naturally two or more planes. Then, the intersections of all planes become a straight line passing through the camera center and corneal sphere center. This is because, assuming a corneal sphere, the normal vector at the CR is on the same plane, and it also passes the center of the ideal corneal sphere. When there are two or more cameras, the intersection of the straight lines obtained for each camera creates an intersection point, becoming the center of the cornea.

Because the actual cornea differs slightly from an ideal sphere, even in the central area of the cornea, the intersection should be calculated by an approximation, such as using the least-squares method.

Identification Procedure

This subsection of the study proposes an identification algorithm that applies this procedure.

Figure 5.3 (b) shows an example in which infrared lights are assigned to CRs differently. The calculated planes and intersection lines differ from the correct one, as shown in Figure 5.3 (a); thus, the estimated center of the corneal sphere as an intersection point is also slightly different from the correct one. These two lines sometimes have a torsional relationship, as shown in Figure 5.3 (c). In such cases, the central position can be calculated using the least-squares method as the midpoint of the shortest distance between the two lines.

If the identification is correct, the estimated position of the center of the corneal sphere is exactly on each plane. However, with misidentification, the center of the corneal sphere does not exist on each plane. Focusing on this, we considered a cost function to determine the appropriate identification.

The identification procedure is described as follows. The definitions of k, k_u and k_b are the same as those in Sec 5.3.2, and the newly introduced variables N_u and N_b are the numbers of matched points in each upper and bottom row, respectively.

1. The detected points already have a match between two cameras and already belong to each row to reduce the computational cost of iterative search. Furthermore, the points of each row are laterally sorted.
2. All candidates are enumerated under the condition $k_u + k_b = k, 0 \leq k_j \leq N_j$ for $j = u, b$ (the following section provides a detailed explanation), and the error is calculated for each candidate.
3. If any of these candidates has the smallest error and is below the threshold determined in advance, it is considered to be the correct identification. (End)
4. If not found, Steps (2)-(4) are repeated with $k \leftarrow k - 1, k \geq 2$.

Generally, two or more points are required to estimate the corneal sphere's center; therefore, the lower limit of k is determined to be 2 and the matching algorithm.

Cost function

The design of the cost function is discussed here. Directly calculating the distance between a point and plane is the most straightforward method but has a high computational cost. Therefore, it is unsuitable for examining all candidates.

To simplify this procedure, this study focused on that, when the identification is correct, the following two observations are equivalent: (1) both the corneal sphere center and each camera origin are on the same each plane (2) the vector from the camera origin to the corneal sphere center and normal vector of each plane are orthogonal. By translating the estimated corneal center into the coordinates of the camera origin, the inner product of the normalized corneal coordinate and normal vector of each plane indicated whether it is on the plane. The sum of the absolute values of these inner products can be used as a simple error function. If the corneal sphere center position is correctly estimated, the calculation can be zero. This process is simple, fast, and contributes to reducing the computational cost.

Enumeration of identified candidates

A detailed method for enumerating candidates is described here. The strategy is similar to the CR matching described in Section 5.3.2.

Because most incorrect reflections are excluded, it is assumed that IDs can be assigned in ascending order from left to right. This assumption did not consider cases in which some noisy

reflections are sandwiched between true CRs or in which some of the true CRs are occluded by eyelashes. However, if such noise yields a large error, this enumeration method has more advantages in terms of simplicity.

Let the number of IR lights per column be N_l (e.g., if the total number of light sources is 14, $N_l = 7$), and let the other variables, N_u, N_b, k, k_u, k_b have the same definitions as above.

Let $j = u, b$ and the number of combinations to select k_j points from a certain row by calculated as ${}_{N_j}C_{k_j}$. Then, IDs are assigned to all k_j points, and the number of ID combinations is calculated as $N_l - k_j + 1$ by shifting the possible assigned ID. Then, the total number of candidates is computed as follows:

$$\sum_{k_u,k_b} \prod_{j=u,b} {}_{N_j}C_{k_j}(N_l - k_j + 1). \tag{5.5}$$

If only one row consists of CRs, let N be the number of all matched points between two images, and the number of enumerations is calculated as follows:

$$2{}_{N}C_{k}(N_l - k + 1). \tag{5.6}$$

Because the identification search is conducted for both rows of infrared light sources, the total number is doubled.

Example

Using the picture in Figure 5.2 as an example, let $N_l = 7, k_u = 3$ and $k_b = 1$. As shown in Table 5.1, both matched number of each row, N_u and N_b, are calculated as 3. Then, the number of identification candidates for each row is calculated as follows:

$${}_{N_u}C_{k_u}(N_l - k_u + 1) = {}_3C_3 \cdot (7 - 3 + 1) = 5 \tag{5.7}$$

$${}_{N_b}C_{k_b}(N_l - k_b + 1) = {}_3C_1 \cdot (7 - 1 + 1) = 21. \tag{5.8}$$

Therefore, the number of the ways is $5 \times 21 = 105$. The detailed candidates are enumerated as follows:

Upper $\{f_1,f_2,f_3\} \leftarrow \{1,2,3\},\{2,3,4\},\{3,4,5\},\{4,5,6\},\{5,6,7\}$
Bottom $\{f_4\},\{f_5\},\{f_6\} \leftarrow \{1\},\{2\},\{3\},\{4\},\{5\},\{6\},\{7\}$.

5.3.4 Continuous tracking and identification of new CRs

As mentioned in previous sections, our system with high-speed cameras enabled us to assume that the appearance in the current and previous frames does not differ significantly. Based on this, a point detected near a certain CR in a previous frame can be assigned the same ID.

However, as the eye rotates, new CRs may appear. They can then be identified using a similar procedure. First, the number of rows of all detected CRs is calculated, and the row to which new CRs belong is recognized. Then, for each new CR, the following procedure is conducted: it is assigned one ID among those unused in that row, and the cost function introduced in Section 5.3.3 is calculated. If the smallest error is below the threshold determined beforehand, the ID of that error is properly assigned to the new CR.

We introduce an example using Figure 5.2. If the CRs of $\{f_1, f_2\}$ are assigned with 3,4 IDs respectively, and the CR of $\{f_3\}$ remained unassigned, the candidates of the ID of $\{f_3\}$ are enumerated as 1,2,5,6, and 7. For each candidate, the cost function introduced in Section 5.3.3 is calculated, and those with the smallest error under the threshold are newly assigned a CR of $\{f_3\}$. In this case, the appropriate ID is 5.

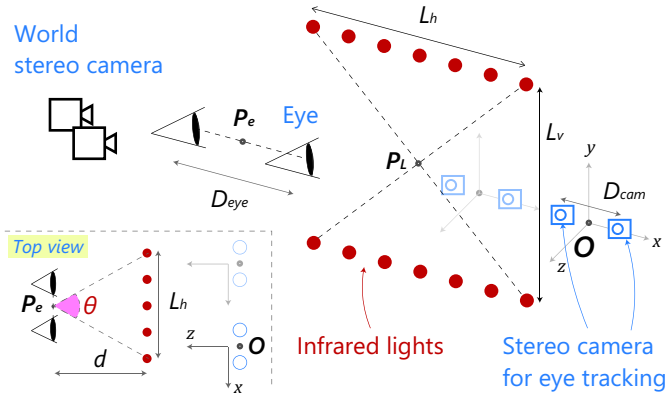

Fig. 5.4. System overview. The described parameters are used in the simulation.

5.4 System overview

Figure 5.4 shows a schematic of the system used in this study. Stereo cameras are prepared for each eye to obtain a high-resolution eye image. Four cameras are used in this study. This setup may seem complicated. Because the camera images did not include areas unrelated to the eyes, the setup enhanced both accuracy and speed. Furthermore, this type of arrangement has been used in several wearable products [185, 33, 35]. From this perspective, these advantages are considered valuable.

Multiple lights are placed between the user's eye and cameras. In this arrangement, the infrared light sources do not appear in view of the cameras, which makes calibration difficult. Here, a world stereo camera, located near the user's position and in front of both the light sources and eye cameras, is prepared. The calibration is conducted using a calibration board with checker patterns printed on both sides. Although the world camera has a wide view angle, the eye camera has a narrow view. Therefore, the pattern size of each side has a different suitable size for each camera.

Recently, with the availability of 3D printing and laser machining, infrared lights and cameras can be installed exactly as designed. However, the camera's principal point position may be located not exactly at the center of the lens, depending on the accuracy of the lens and other factors. The introduction of a stereo-world camera is believed to enable precise calibration.

5.5 Evaluation of actual eyes

5.5.1 Apparatus

Figure 5.5 shows the actual experiment setup. For the infrared-point light source, ILH-IW01-85SL-SC221-WIR200 from Intelligent LED Solutions (the built-in infrared light was OSLON(R) SFH 5716AS, 850 nm, radiant flux 1270 mW) was used. The stereo cameras for eye tracking were two Ximea MQ013RG-ON (monochrome, infrared light enhanced, 1280 $\times$ 1024 px, and 210 fps in full resolution) for the right eye and two Basler acA800-510um (monochrome, 800 $\times$ 600 px, and 511 fps in full resolution) for the left eye. Both cameras were set with an ROI of 512 $\times$ 410 px to achieve a frame rate of 1000 fps. All the cameras are attached with a filter passing only infrared light (Edmund Optics, TECHSPEC® High Performance Mounted Machine Vision Filters). The computing machine was a GALLERIA ZA9C-R38 (Windows 10 Pro, CPU: Intel Core-i9 12900K, 64 GB of RAM, 2-TB SSD). For the world stereo camera, color cameras of DFK38UX253 (4096$\times$3000 at 30 fps) and DFK38UX541 (4504$\times$4504 at 18 fps) from The

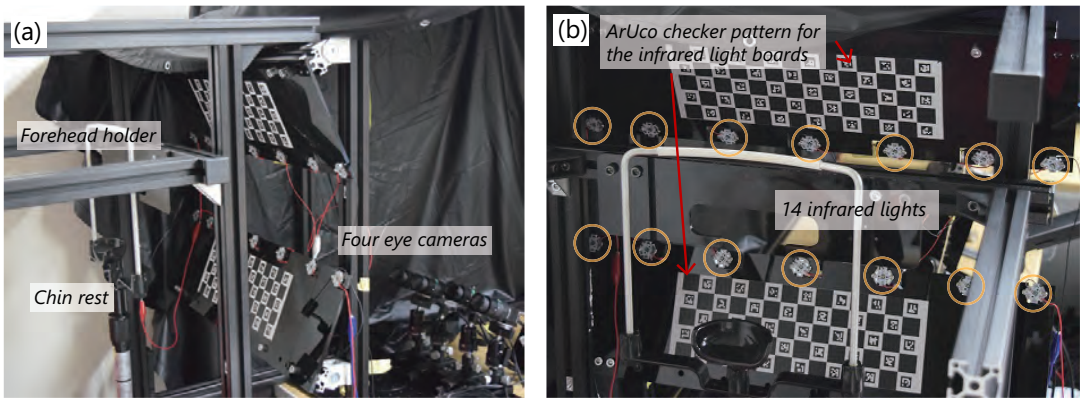

Fig. 5.5. Experiment apparatus from (a) left and (b) front sides.

Table 5.4. Image processing time per frame.

Time per frame [ms]	Ave.	Std.
All	0.4385	0.1112
Initial Ellipse Detection	0.2564	0.0936
Continuous Ellipse Detection	0.1148	0.0443
CR Detection	0.0576	0.0226
Gaze estimation	0.0252	0.0334
Matching	0.1411	0.1058
Initial identification	0.0934	0.1159
Identification for new CRs	0.0063	0.0020

Imaging Source Europe GmbH were used.

14 infrared point light sources were prepared, and its horizontal placement range was set to 54 cm. The distance from the user's eye to the IR lights d was approximately 20 cm from which θ, an indicator of the light placement range, was calculated to be $106.9°$.

Videos of eye rotation were recorded for 7 seconds (7,000 frames) from seven subjects: one Chinese male and six Japanese males. All subjects had black pupils and their ages ranged from 22 to 30. The subjects were instructed to place their chin on the chin rest and press their forehead against the forward forehead holder.

Three to five trials were conducted to capture video per person. Because videos of both eyes were recorded from both stereo cameras, four movies were recorded per trial. For each video, each user was instructed to rotate their eyes within a certain angle range, ranging from small to large rotations.

5.5.2 Processing time per frame

The results of each real-time processing time per frame, ranging from loading images to gaze estimation, was measured over a period of 7 seconds (7000 frames) per eye rotation video of one user. For the matching and identification process, because they are occasional processes, the average was taken only over the frames in which those processes were performed (not over all frames).

Table 5.4 shows the statistics for each process averaged over three trials. Although as many as four images were simultaneously processed within the same frame, the processing time was sufficiently small, less than 0.5 ms on average. It can be concluded that real-time high-speed eye tracking with correct CR identification was realized.

Table 5.5. Calculation time of 1,000 iterations for enumeration of matching and identification.

Ave. (Std.) [ms]	Matching	Identification
Single	0.4970 (0.1299)	6.3255 (1.5871)
Two parallel	0.3345 (0.2259)	8.3302 (3.0655)
Four parallel	0.5269 (0.2753)	4.9450 (2.7535)

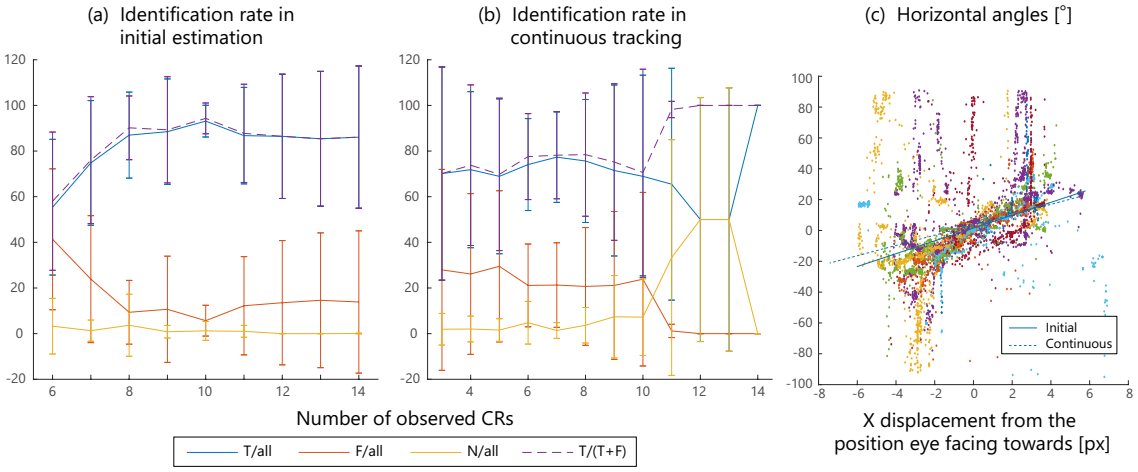

Fig. 5.6. Evaluation results on the estimation success. (a) Success rate of identifying CRs in the initial frame according to the number of observed CRs. (b) That in the continuous frames. (c) The estimated angle range of eye rotation where the estimation succeeds in the initial frame and the continuous tracking. The data points are the only initial points, thinned out by 5 points. The color of the data points differs by subject.

5.5.3 Identification performance

The success rate of CR identification on initial frames was investigated. For each subject, both left and right eye's movies of relatively large eye movements were taken. For a certain subject, the left eye's video could not be recorded properly; thus, two videos of his right eye were used instead.

For the identification test of initial frames, the tracking was always initialized, identification was conducted per frame, and the true or false of each frame was manually distinguished and recorded. By contrast, for those of continuous frames, only the frames where the number of CRs were changed were used for data. Statistics of identification success rate was taken for each subject and each number of observed CRs, and finally the overall statistics was taken across all the subject for each number of CRs.

Figure 5.6 (a) shows the statistical results for identification in initial frames. The success rate has a value of around 80% with 8 and larger number of observed CRs, and it drops with less than 8 observed CRs. While the success rate of solid blue line expresses those against all records, that of dashed purple line expresses those against recorded as true or false (none was excluded). Both graphs have a similar shape, and both have relatively large standard deviation. Therefore, there seems to be a large individual difference. As from the actual processing appearance of eye video, some subjects have high success, but others do not. It is thought to be due to the corneal shape difference from the assumed simple corneal shape.

Figure 5.6 (b) shows the statistical results for identification in continuous tracking. The success rate (dashed purple line expressing $T/(T+F)$) was also close to 100% with larger than 10

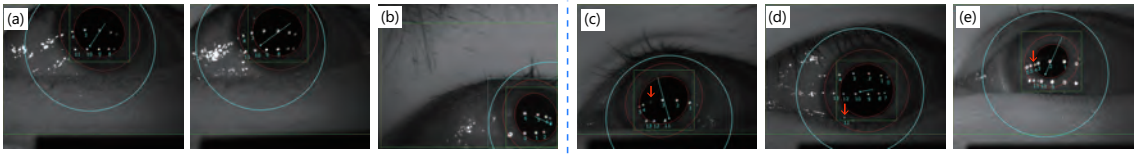

Fig. 5.7. Typical examples of when tracking does not succeed. (a) An example of the assigned IDs of the image on the right are shifted to left. (b) A misidentification example caused by occluded reflection, which is observed in the image on the left but not that on the right owing to issues such as an eyelash. (c) An example grouping failing to work: the CRs of two rows are incorrectly recognized as one row. (d) A glint on a tear accumulated on the eyelid and a Purkinje-Sanson image appeared; these were incorrectly assigned IDs of 11 and 5 in each image.

observed CRs, otherwise it stays around 70-80% with a large standard deviation. From the difference of solid blue and dashed purple lines, there were many cases where the identification could have no result (the orange line expressing N/all). The large individual difference is thought to be due to the same reason as before; the assumed simple eye model is not close to the eye's shape of some subjects. Furthermore, the lacking accuracy of small number of CRs also attributes to the low detection accuracy of CRs when eye directs to side. This is a fundamental problem for all PCCR-based eye tracking system.

Figure 5.6 (c) shows the estimated horizontal angle of eye rotation mapped according to C_x in both the initial and continuous frames, indicated by the blue solid and dashed lines, respectively. Note that these data were generated only from correctly identified CRs. The data points represent only the data from the initial frames, thinned out by 5 points. Both initial and continuous estimations were enabled within a horizontal eye rotation of $\pm 20^\circ$, which is a larger range than that in a previous study based on a CNN [126].

5.6 Discussion

Our method for the identification of multiple CRs achieved an accuracy of approx. 80% and nearly 100% on the initial and continuous frames respectively. However, both of them have a large standard deviation, which means that a large individual difference exists. Its reason is estimated that our identification method is effective only for those whose corneal shape matches to the assumed simple eye model, and ineffective for those does not.

The other reason to drop the accuracy should be noted here as referring to Fig. 5.7. Fig. 5.7 (a) and (b) are owing to the algorithm incorectness and inappropriate thresholds. Fig. 5.7 (a) shows the example of wrongly shifted assignment of IDs on the right image with compared to the correct assignment of the left image. Fig. 5.7 (b) shows that of incorrect recognition of number of CRs rows.

By contrast, as shown in Fig. 5.7 (c-e), some unexpected noisy reflections or missing reflections around the correct CRs decreased the accuracy, which can be excluded appropriately in the future. Fig. 5.7 (c) is an example of unexplained missing of a true CR, while (d) and (e) show the unexpected reflections at the border between cornea and lower eyelid and 2nd Purkinje image, respectively. These unexpected noises or lost cannot be corresponded by not only our method but also previous methods.

Since this study uses a computer with a state-of-the-art CPU, the processing time evaluated in Section 5.5.2 had so small value. This should be so because as many as four images need to be loaded at the same time; however, most of the current computer may have higher calculation time than ours, and thereby the tracking of CRs will have a limited performance.

5.7 Conclusion

This study proposes a method of model-based identification of multiple CRs. The identification algorithm assumed a simple eye model [43] but without further approximation [119], and was designed with reference to the most popular model-based eye tracking method, PCCR [40]. Since the fast detection is needed for involuntary and continuous eye's movement such as saccade, the candidates were efficiently enumerated by assuming the realistic situations, and cost function to distinguish the correct candidate was designed to be simple for the small computational cost.

The evaluation experiments for seven Asian people showed our method has a sufficient identification accuracy at average, but there existed a large standard deviation. This means that a large individual difference exists due to the individual corneal shape difference from the assumed simple eye model, which has been regarded as a standard model though. It is estimated that as the actual eye shape differs more from the simple eye model, the identification gets worse. However, from the actual eye images, some unexpected noisy or missing CRs were observed, which should have decreased the accuracy and can be excluded appropriately in the future. In summary, our method has yielded sufficient accuracy and speed, and then, it is applicable for fast tracking of widely rotating eyes as well as semi-automatic labeling for deep learning.

In the future, the identification algorithm should assume a more complicated eye model for high accuracy, which may lead to confront the problem of late speed. The individually different eye parameters such as corneal curvature radius and refractive index should be calibrated beforehand and used for the identification algorithm.

Chapter 6

Far-field Aerial Image Presentation Using Distributed Displays

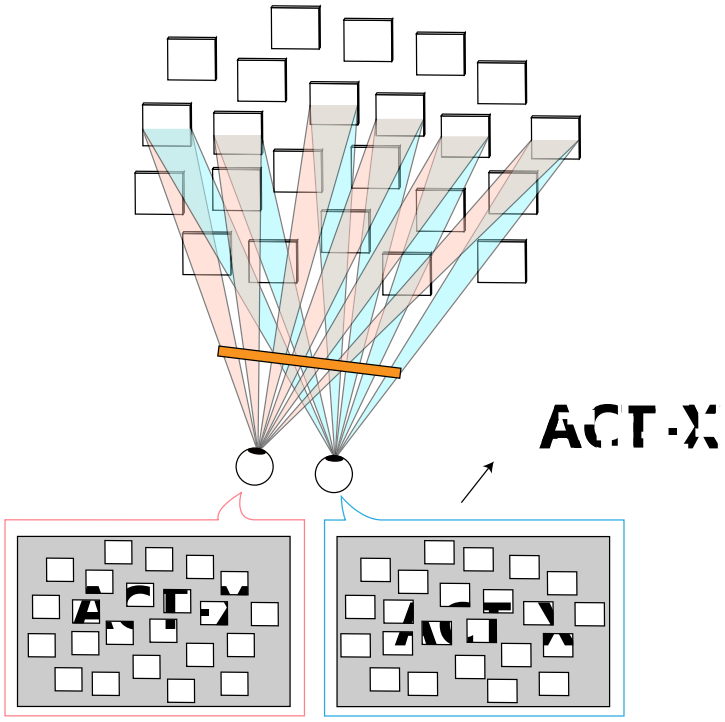

Fig. 6.1. Concept of distributed placement display. This figure shows the concept of a parallax display in particular.

6.1 Overview of proposal

The wide-area aerial image display, where the display is stationary and not wearable and the image can be observed with the naked eye or with lightweight glasses, is expected to be developed. Such a technology can be applied to alerting and guiding people in a wide-area space as well as to entertainment applications. In creating such aerial displays, binocular parallax displays are suitable fundamentally and for its easy setup, as mentioned in Section 2.4.3. By contrast, the conventional method requires the display to be huge and placed as covering the entire space [51], which is not suitable in some situations such as outdoors.

Therefore, this research proposes a display system in which multiple binocular parallax displays

are distributed, as shown in Fig 6.1. This concept eliminates the need to prepare a huge display or to spread small displays all over the wall. It also has the advantage that other equipment such as audio and tactile devices can be placed in the vacant space. By contrast, there is a drawback that only the small aerial image is presented or that missing parts may occur in the presented image; this paper assumes that humans can receive information correctly even if images are partially missing.

This section provides several studies towards such a concept of wide-area aerial presentation using distributed binocular parallax displays. Section 6.2 introduces a first study of separate rays, where a small aerial image is presented at the intersection of two rays and the convex lens is introduced to facilitate the focus to the aerial image. Section 6.3 conducts research to control such separate rays using a laser and two-axis galvanometer mirror. Section 6.4 realizes it using dynamic parallax barriers and conducts a simulation study to investigate the appropriate barrier parameters for the viewing by naked eyes.

These three studies are roughly separated in the former two using thin rays and last one using binocular parallax display, and each has different advantages and disadvantages. Those using separate rays is better in controlling the viewing zone especially with naked eyes than the binocular parallax way. By contrast, the binocular parallax way can produce dense images while the way of separate rays cannot. These concepts are investigated carefully, and the actual device was partially built.

It should be noted that there are several relations to the other sections. Chapter 4 is a study to improve stereoscopic perception and depth perception by applying eye tracking to conventional anaglyph-based binocular parallax displays. Chapter 5 is research providing new algorithm of identifying multiple corneal reflections during a large eye rotation, which can be applied for the wide-area tracking and presentation in such distributed displays, especially for the naked eye viewing.

6.2 Small aerial image presentation using two separate rays

(Note that this section is described with a reference to the author's conference proceedings paper [44], and the first person is "we" because this is joint research with co-authors.)

We propose a method of presenting aerial image at a distance using two separated rays, which is viewed by the left and right eye separately and presents binocular parallax images. Fig. 6.2 shows a schematic of the proposed optical system. In this system, each beam from the two distant light sources forms a real image of binocular parallax using convex and concave lenses, which can be seen by either the left or right eye (one at a time). The image of each beam is formed at the intersection point, where an aerial image can be seen naturally with a small vergence-accommodation conflict [196].

Although it is viewpoint-dependent for only one user and can form only a small image, it is sufficiently effective for several applications such as a small alert and advertisement in a wide space. In the following section, we explain our method to form an aerial image for a distant user corresponding to his/her various head positions and rotations. Evaluation experiments of the presentation of an aerial image for multiple head positions show the validity of the proposed method.

6.2.1 Proposed method

System overview

A schematic of the proposed distant aerial image method is shown in Fig. 6.2. A light ray forming a real image of the light source (display object) is generated by placing a collimating lens system

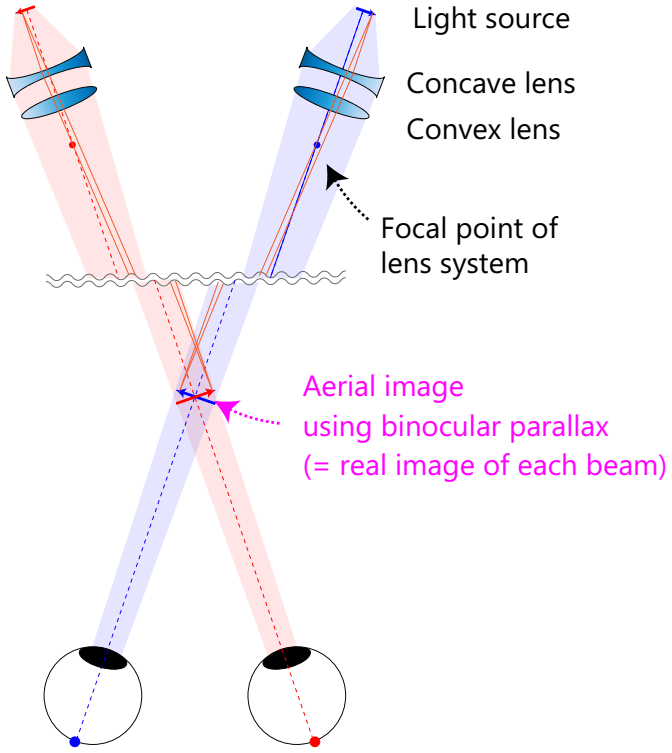

Fig. 6.2. System overview of distant aerial imaging display using binocular parallax. Cited by Fig. 1 in [44].

in front of the light source. By designing each ray direction to enter the pupil position of the left or right eye (one at a time), it is possible to present an aerial image at the rays' intersection position. The use of two rays overcomes the narrow aperture of the lens and enables the reliable presentation of an aerial image, regardless of the user's distance from the device and head's rotational angle.

The position of the aerial image is the rays' intersection, as mentioned above, and uniquely determined from the positions of the light sources and the position and rotation of the head. In this section, we formulate it that holds both in the bilaterally symmetrical case (the head is just in front of the device, as shown in Fig. 6.2) and the asymmetrical case (Fig. 6.3).

A schematic of our method in the asymmetric case with some parameters is shown in Fig. 6.3. We denote the coordinate of the user's intereye position as $P_e = (P_e x, P_e y)$, the interocular distance as d, and the rotational angle of the head as θ. The distance between the two light sources is D. The center of the two light sources is assumed to be an original point in the coordinate system. The coordinate of the presented image $P_a = (P_a x, P_a y)$ is expressed by

$$P_a = K(4P_{ex}P_{ey} - d\sin\theta(D + d\cos\theta), 4{P_{ey}}^2 - d^2\sin^2\theta) \tag{6.1}$$

The parameter K is expressed by

$$K = \frac{D}{4(DP_{ey} + d(P_{ey}\cos\theta - P_{ex}\sin\theta))} \tag{6.2}$$

In particular, when the line consisting of both eye positions is parallel to the x-axis, where the two light sources are located, the rotational angle of the head $\theta = 0$, and the position of the presented image P_a can be easily obtained by the similarity of the triangles,

$$P_a = \frac{D}{D + d}P_e \tag{6.3}$$

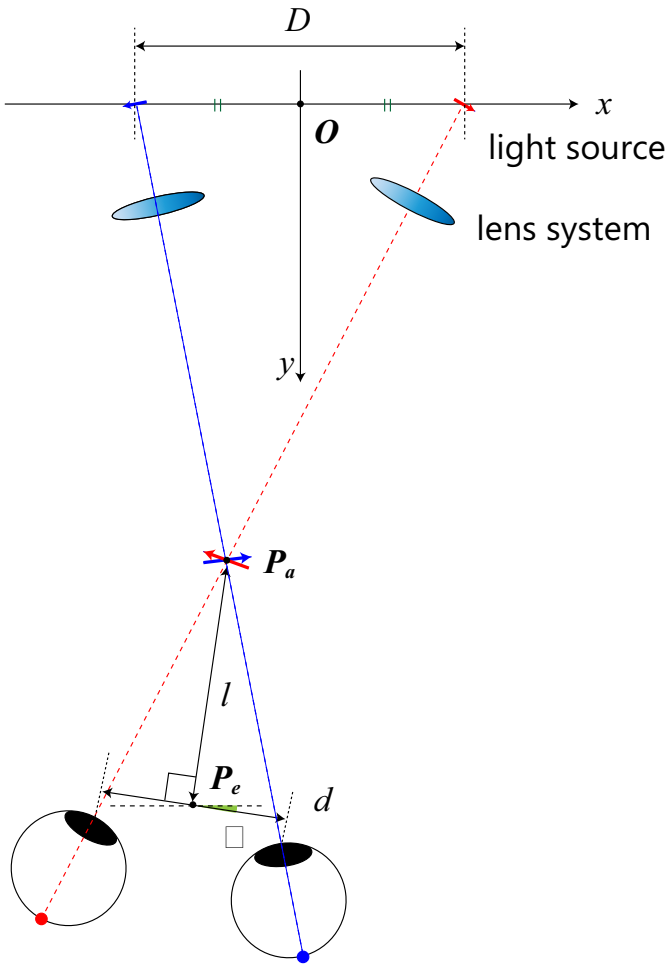

Fig. 6.3. Asymmetric case of the proposed system and relationship between the parameters that determine the position of the presented aerial image. Cited by Fig. 2 in [44].

When the aerial image is presented just in front of both eyes (the distance between the aerial image and each eye is the same), the position of the presented image P_a and its distance from the interocular position l are expressed by

$$P_a = P_e + l(\sin\theta, -\cos\theta) \tag{6.4}$$

and

$$l = \frac{d(P_{ey} + \sqrt{P_{ey}^2 + D(D\cos\theta + d)\cos\theta\sin^2\theta})}{2(D\cos\theta + d)\cos\theta} \tag{6.5}$$

respectively. Note that $P_e = (P_e x, P_e y)$ must satisfy

$$P_{ex}(l^2\cos^2\theta - \frac{d^2}{4}\sin^2\theta) + P_{ey}(l^2 + \frac{d^2}{4})\cos\theta\sin\theta = \frac{d^2 l}{4}\sin\theta \tag{6.6}$$

Multiple Lens System for Formation of a Real Image at Desired Size and Position

In addition to the formula for the aerial image's position, it is necessary to appropriately design the lens system so that each ray can form a real image with the same size at the same position P_a regardless the ray's length. In this study, we consider the use of a lens system in which concave

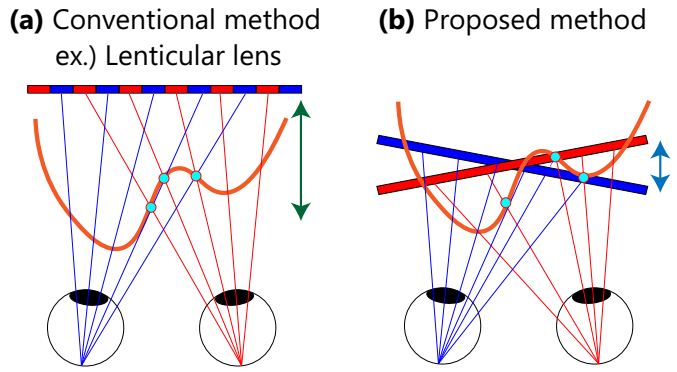

Fig. 6.4. The conventional method using stereoscopic vision for aerial images has a large vergence–accommodation conflict. (b) Our method has a small vergence–accommodation conflict. Cited by Fig. 3 in [44].

and convex lenses are arranged one by one in order from the light incident direction. This is the simplest system that can determine the positional relationship between the two lenses according to desired image magnification and image formation position (notably, a lower limit of the image magnification exists) [197]. According to the formula for the focus length of multiple lenses, the magnification of the real image to the object is smallest at the minimum distance between two lenses.

Consideration of the vergence–accommodation conflict

Our method uses stereoscopic vision to form aerial images, which has been used for many technologies, as described in Section 2. However, several of them suffer from a large vergence–accommodation conflict because of the fixed light-emitting surface position, as shown in Fig. 6.4 (a). In contrast, owing to the imaging lens, our method can place the light-emitting surface at almost the same position as that of the virtual object, as shown in Fig. 6.4 (b). Therefore, the aerial image can be viewed naturally with little discomfort. Regarding this content, we only discuss it and do not carry out evaluation experiments.

6.2.2 Evaluation experiment

Experiment Environment

Fig. 6.5 (a) and (b) shows the overall appearances of the experimental environment and optical system, respectively. A plano-convex lens (focal length: 50 mm) and plano-concave lens (focal length: -70 mm) with a diameter of 25 mm (SIGMAKOKI Co. Ltd.) were used. The minimum distance between two lenses was 25 mm owing to the carriers, which fixed the optical elements. A Nikon D3500 camera was used for the evaluation (camera lens: Nikon AF-S DX NIKKOR 18–55mm f/3.5-5.6G VR (zoom lens, 6000 $\times$ 4000 px)). A rail slider was used to imitate both eyes. A board indicating the focusing position of an aerial image was set statically. The interocular distance d was assumed to be 65 mm. An image with a size of approximately 1 mm with 9 pixels, was used as a printed image, as shown in Fig. 4 (c). The position of each lens was calculated using MATLAB 2020α to present an aerial image at the desired size and position.

Experiment Method

By this experiment, we confirmed that the proposed system can show an aerial image according to multiple head positions and angles.

First, we conducted an experiment on the system of using bilaterally symmetry ($\theta = 0°$) for two, near and far, cases ((a) 868.0 mm, (b) 950.0 mm). The distance between the aerial image

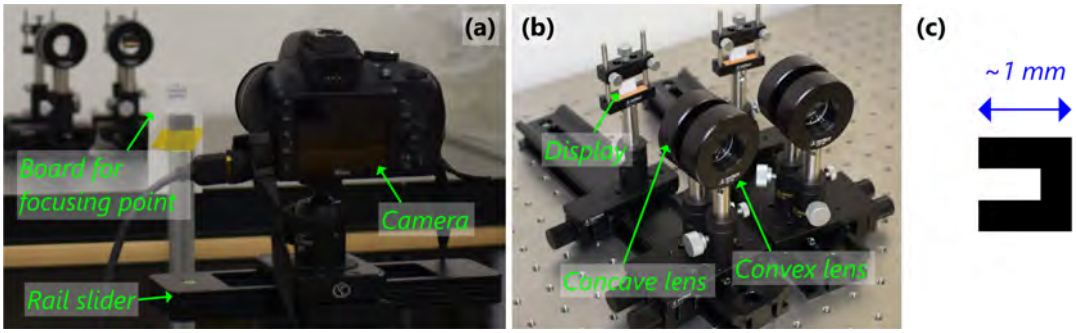

Fig. 6.5. Appearance of the evaluation experiment. (a) Entire system including an evaluation camera and board indicating the focusing position. (b) Optical system composed of lenses and tiny printed images. (c) Picture used in the tiny printed images. Cited by Fig. 4 in [44].

and intereye position l was 325 mm in both cases. When the distance between the two lenses had the minimum value (25 mm), the distances between the two light sources D were 108.8 and 150.0 mm, while the magnifications were 4.48 and 5.57, respectively. The board indicating the focusing position was set at the crossing position of the two rays. The lenses' positions were properly adjusted with cues of MATLAB 2020α calculation and visual feedback. The focal length of the zoom camera was 55 mm.

Second, we performed the same experiment in a bilaterally asymmetric case ((c)θ=-14.0° , l=61.1 mm, P_{ay}=471.9 mm, D=236.3 mm). The magnification was 3.33. The focal length of the zoom camera was 35 mm.

Results

The experimental results are shown in Fig. 6.6. When the focus of the camera at each eye's position was adjusted to the depth position where the board was located, the real image was observed at the same position above the arrow indicated by the board. As this is valid for all experimental cases, the two bilaterally symmetrical ones ((a) near and (b) distant) and asymmetric case (c), the proposed method was appropriate. The real image in (b), farthest in the three cases, has a lower brightness than those of the other cases. This could be attributed to the attenuation of light inversely proportional to the squared distance and low light density at the larger magnification in the distant imaging.

Discussion

Owing to the configuration of the optical system, the proposed system leads to two main problems, particularly noticeable at a distance. One of them is the limited size of a real image owing to vignetting by the aperture diameter, and the other is that the positional adjustment of the lens system requires a higher resolution of control. These problems can be overcome by selecting lenses with an appropriate aperture diameter and focal length according to the application.

Moreover, in a simple lens system composed of only two lenses, it is challenging to control two factors (imaging size and position) simultaneously because they are interdependent. However, it is possible to overcome this problem using a zoom lens.

6.2.3 Conclusion

We realized a distant aerial image presentation using two rays, each of which shows a different binocular parallax image observed by the left and right eyes, respectively. Its performances at multiple head positions and angles were analyzed through evaluation experiments. However, a few

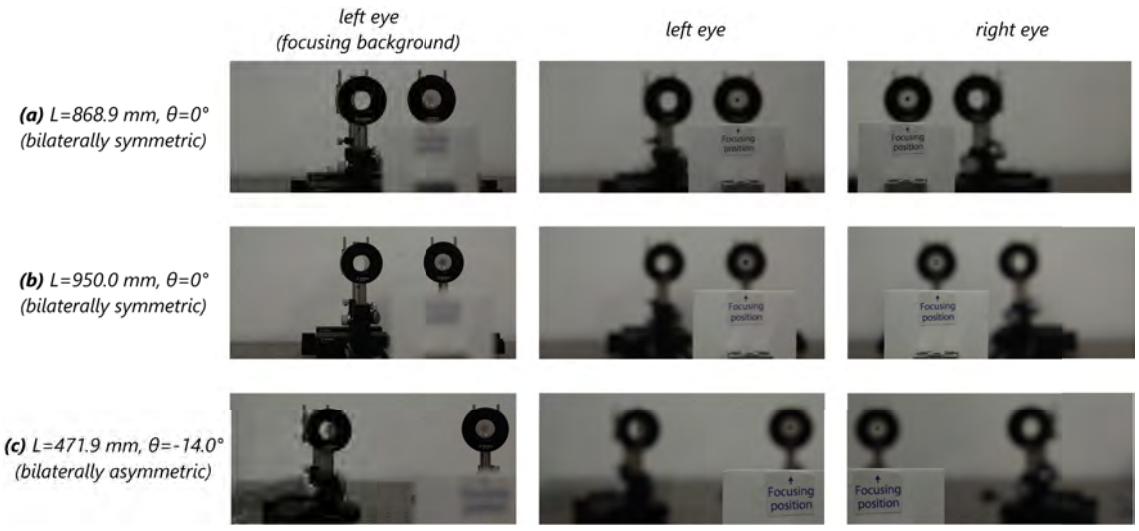

Fig. 6.6. Results of evaluation experiments of two near and far bilaterally symmetric cases (a), (b) and one bilaterally asymmetric case (c). Note that the focal length of the zoom camera was 55 mm for (a) and (b), and 35 mm for (c). Cited by Fig. 5 in [44].

problems still hinder the real-life application, as described in Section 5, which will be addressed in our future studies.

6.3 One-point distant aerial image presentation using laser scanning

(Note that this section is described with a reference to the author's conference proceedings paper [45], and the first person is "we" because this is joint research with co-authors.)

In this study, based on a previously developed system [44], we consider the method of presenting an aerial image to be observed at a farther distance using a laser light source with high directivity and low light attenuation. Then, we propose a system that uses a camera and a two-axis galvanometer mirror to present an aerial image of a single point in front of a distant user, as shown in Fig. 6.7. The laser beam is thin and has a narrow viewing angle at the convergence position that is distant from the device and near the user; therefore, it is necessary to measure the position of the pupil of the distant user and control the direction of the laser beam such that it always enters the pupil. For pupil measurement, we use the conventional method that coaxializes the camera and collimated infrared light [198]. A two-axis galvanometer mirror is used to control the laser beam direction. To build an entire calibrated system that can conduct laser scanning in 3D space, we propose a 3D calibration method and a fast-forwarding algorithm that is based on an established model [199], whereas most previous studies have focused on calibration on a plane for laser marking.

The performance of the proposed system is evaluated, and the appearance of an aerial image of a single point, depending on the incident vector to the camera, is confirmed. In this study, we do not verify whether automatic focus adjustment based on the focused beam occurs on one eye or whether proper beam focusing occurs at the focal point.

In the future, we will consider increasing the number of axes of the galvanometer or combining it with other optical elements such as concave mirrors to present patterns. Although the size of the aerial image is limited to that of the mirrors, many effective applications can be realized owing to the advantage of natural visibility at a distance from the device, even when only a small number

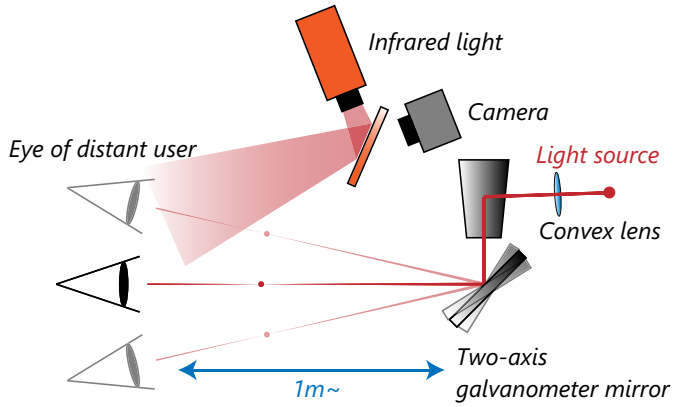

Fig. 6.7. Proposed basic system that realizes the presentation of a distant aerial image of a single point. Cited by Fig. 1 in [45].

of characters is present. The possible applications include alerting people in public spaces, such as train stations and crosswalks, and presenting a prompter to stage actors. This study will serve as a basis for these applications.

6.3.1 System

The proposed system for presenting an aerial image of a single point at a distance is shown in Fig. 6.7. The measurement component comprises a camera and an infrared light coaxialized by a beam splitter, and the presentation part consists of a two-axis galvanometer mirror and a straight light source, such as a laser beam. This system requires high-speed pupil tracking because the beam with a narrow viewing angle at the convergence position that is distant from the device and near the user must be irradiated into a narrow pupil area. Therefore, a camera with a high frame rate is used. Given the short camera exposure time, the bright pupil method is suitable as a tracking method, which is one of the pupil measurement methods based on the fact that the pupil of the human eye has a retroreflective component; therefore, the method in which the camera is coaxialized with infrared light with a beam splitter [198] is used. This involves a visual feedback system, where a camera senses the moving eye, and then, a two-axis galvanometer mirror controls the light beam such that the light beam is always incident on the eye.

In actual 3D tracking of the pupil, because only one pupil is involved, two or more cameras are required to perform measurement. However, in this study, we use only one camera because we assume that the pupil is substituted by the evaluation camera and tracked by multiple markers.

Calibration

For a precise control of the laser beam, the entire system should be calibrated appropriately. The model of the system used in this study, comprising a two-axis galvanometer mirror, laser source, and single camera, is shown in Fig. 6.8. The intrinsic parameters of the camera are calibrated using an existing method, and 3D reconstruction is assumed to be possible when appropriate cues of pose information of the target object are provided. It is assumed that the intrinsic model of the galvanometer mirror has 10 degrees of freedom, as established by Godineau et al. [199], enabling easy ray tracing by mapping the incident light orientation to the mirror plane coordinate system, even when the mirror thickness should be considered or the number of axes increases.

Herein, we propose the optimization of the relative orientations of the laser beam and the camera to the galvanometer mirror, which have four and six degrees of freedom, respectively, and are unknown because of manual alignment by humans, using 3D laser points in the camera coordinates. Although such a calibration for a two-dimensional plane for laser marking has been widely

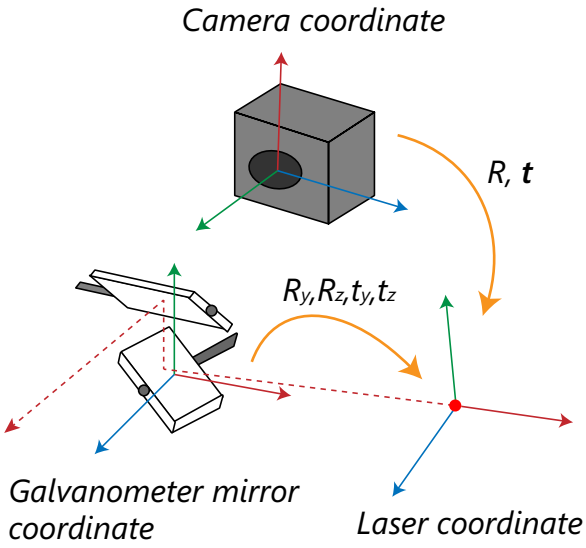

Fig. 6.8. Model of the system, comprising two-axis galvanometer mirror, laser source, and camera, used in this study. Cited by Fig. 2 in [45].

conducted previously, we target calibration in a 3D space to be scanned by a laser in this study. We assume that the parameters of the galvanometer mirror are factory defaults.

The specific procedure for the calibration of the galvanometer mirror system is as follows. Several ArUco markers [130] are placed on a black board and their positions are known; hence, the 3D orientation of the board in the camera coordinate system can be estimated from these markers. A visible laser beam is irradiated on the board; its position is detected via image processing, and its 3D position is estimated based on the board orientation. Subsequently, based on the 3D positions of the lasers and the corresponding actual control angles, the parameters are optimized iteratively using the Levenberg–Marquardt method.

Naturally, the higher the number of points of the laser, the better is the calibration accuracy. However, it is difficult to formulate the range of control angles of the galvanometer that can irradiate the laser over the entire area of the board when it has not yet been calibrated. In this study, we manually set the control angle that irradiates the laser at the center of the board and then control the light beam in a spiral pattern. Using AD conversion to obtain the control angle of the galvanometer mirror in parallel with image processing to detect the laser position, we can calculate roughly the control angle area of the galvanometer mirror that corresponds to the frame of the board.

Image-based Tracking

To estimate the orientation of the camera for evaluation, a board with retroreflective spherical markers at the four corners of the plane is attached to the camera. This allows easy marker detection and a precise posture estimation of the plane, thereby facilitating the estimation of the orientation of the lens plane. The monochrome camera that detects the markers is coaxialized with the infrared light by a beam splitter to enable high signal-to-noise ratio imaging, in which, only the areas with retroreflective markers appear bright. This process, which is shown in Fig. 6.9, enables a highly accurate and precise position detection of the marker locations. If the marker position in the previous frame is known, then the self-window method [64] can be used to accelerate the process. As the marker positions are known, the 3D orientation of the camera can be estimated by solving the perspective-n-point problem [63].

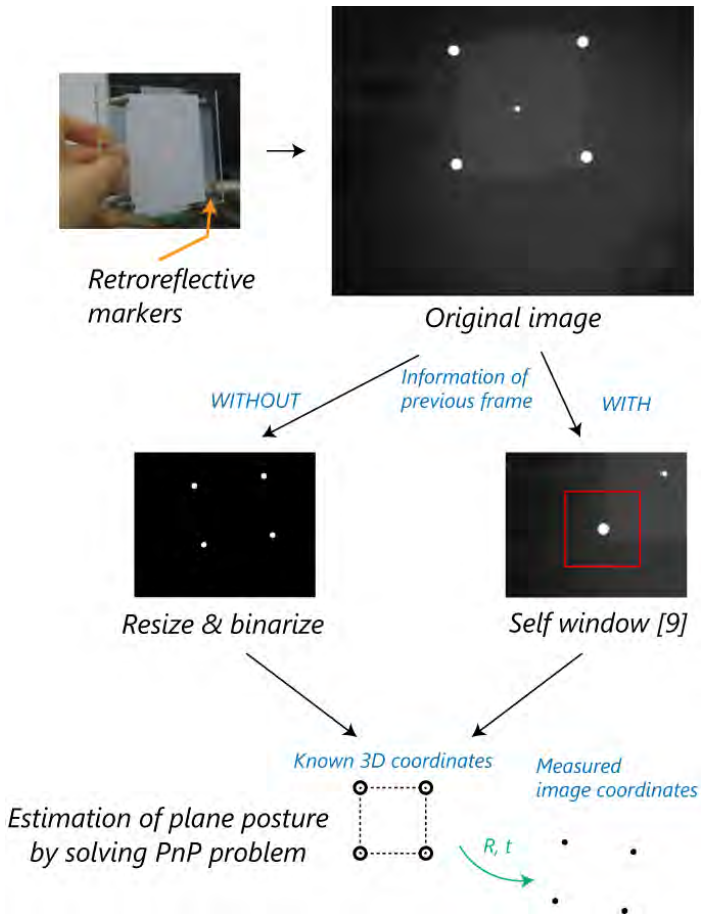

Fig. 6.9. Image process for board tracking. Cited by Fig. 3 in [45].

Control of Galvanometer Mirrors for Floating Image Presentation

To control the galvanometer mirror incident on the pupil position, a forwarding algorithm should be established. The model cannot be solved linearly because the two rotation axes are in a torsional relationship, and each mirror has a thickness in the direction perpendicular to the rotation axis [199], as shown in Fig. 6.8. Therefore, the following two forwarding models are considered in this study.

Formulation of the control angle by iterative optimization

The two variables of the X and Y mirror angles are optimized iteratively, and error minimization with the target 3D position is performed. In this study, we use the Levenberg–Marquardt method for the iterative optimization and formulate the pan and tilt angles to achieve the relevant positions.

Linear model without considering mirror thickness

When the mirror thickness is not considered, the rotation angle of the mirror corresponds linearly to the reflection angle of the ray and, hence, can be solved linearly. In this study, the rotation axis is regarded as the axis where the length of half the thickness of the mirror is offset from the original rotation axis in the direction to the mirror surface normal at a rotation angle of 0 °. This is because the simulation indicates that the error is the smallest in the range of the specified machine angle of the galvanometer mirror.

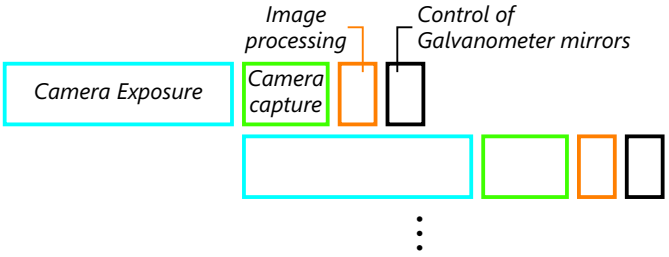

Fig. 6.10. Overall procedure. Cited by Fig. 4 in [45].

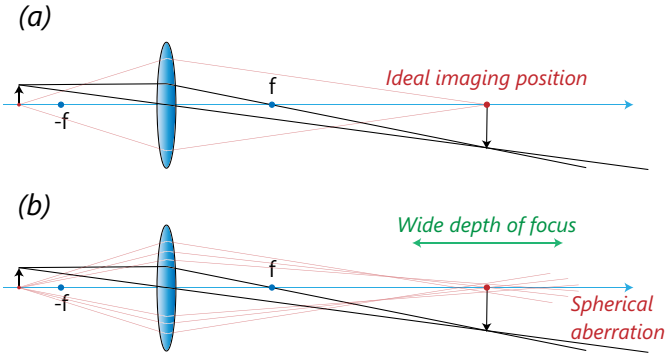

Fig. 6.11. Laser beam passing through a lens: (a) ideal ray path for a single convex lens and (b) ray path when spherical aberration is considered. Cited by Fig. 5 in [45].

Overall procedure

The temporal sequences of the tracking and control methods described thus far are shown in Fig. 6.10. Because the processing should be performed immediately after the camera captures the picture, it is assumed that parallel processing will occur; therefore, the update rate of tracking is the same as the frame rate of the camera. The latency depends on not only the exposure time of the camera, but also the time of the afterward processing for image retrieval from the camera, image processing, and control by the galvanometer mirror.

The time per process determines the allowable speed of pupil movement during actual operation. Assuming that the pupil diameter is e and the time required per processing is t, we can obtain the maximum value of the pupil movement speed, v_{max}, as follows:

$$v_{max} = e/2t. \tag{6.7}$$

6.3.2 Principle of Far-Field Aerial Image Presentation Using Laser Source

The aerial image formation in this study was performed using a convex lens, which is one of the most basic elements in optics. The image formation position and magnification of the real image were determined by the positional relationship between the lens and the real object, based on the lens formula, $1/a + 1/b = 1/f$. For discussion, we assume that a semiconductor laser is used. A semiconductor laser emits coherent straight light from a thin active layer (light-emitting layer) sandwiched between N- and P-type cladding layers; however, because the active layer is thin, i.e., in the nanometer order, it emits diffracted light that spreads out. Therefore, using a lens for

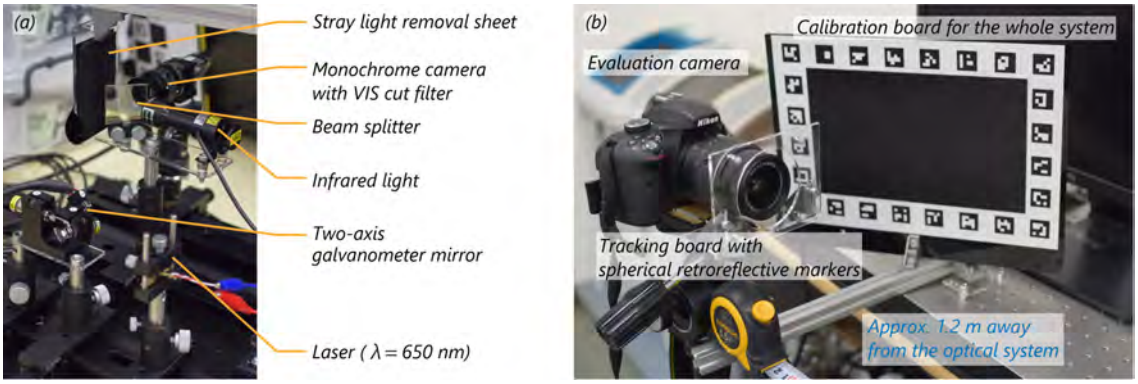

Fig. 6.12. Experimental environment. (a) The appearance of devices for measuring the posture of evaluation camera and presenting the aerial image. (b) Evaluation camera being tracked and calibration board for the whole system. Cited by Fig. 6 in [45].

focusing, while the collimated light is focused at the focal point of the lens, the broadened light of the semiconductor laser is focused at the position determined by the positional relationship with the lens. Ideally, based on the formula of the lens, the shorter the distance to the lens, the farther the light is focused, as shown in Fig. 6.11 (a).

Ideally, a converged laser beam from a microscopic perspective is a Gaussian beam with a beam waist of radius ω_0 instead of a point at the convergence position. Because wavelength λ and divergence angle θ are correlated as $\omega_0 = \lambda/\pi\theta$, it can be concluded that the smaller the aperture of the lens, the larger is the beam waist [200].

Focusing of laser beam assumed by optical system in this study

In the optical system used in this study, the laser passes through a galvanometer mirror, which serves as an aperture and reduces the numerical aperture, particularly when the flux is larger than the assumed beam diameter of the galvanometer mirror. In addition, as shown in Fig. 6.11 (b), the actual beam cannot converge by a spherical lens ideally because of spherical aberration. Off-axis aberrations, such as coma aberration and astigmatism, can be considered as well because of manual alignment errors. In addition to the beam waist of the Gaussian beam, various factors render it difficult to focus the beam on a single point.

Appearance of Floating Image From Tilted Eye

A far-field aerial image display combined with a concave mirror or multi-axis galvanometer mirror, which is expected to be developed based on this study, can be used in a manner whereby the laser is always incident on the pupil while air scanning is performed. In such cases, the direction of incidence to the pupil is not always perpendicular to the pupil surface. Because the actual shape of a pupil differs significantly from the spherical structure, off-axis aberrations such as astigmatism and coma aberration can be reduced. However, it is necessary to quantitatively evaluate whether the observed laser beam differs based on the incident direction.

6.3.3 Evaluation

Experiment environment

A photograph of the experimental environment is presented in Fig. 6.12. The camera used for tracking was a Ximea MQ013RG-ON (1280×1024 px, 200 fps, monochrome), the camera lens was a Computar M3514-MP (focal length 35 mm), and the camera was equipped with a visible light (VIS) cut filter. The galvanometer mirror was a Novanta 6210H 6 mm Silver (mechanical angle: $\pm 10°$, mirror thickness: 1.001 mm, assumed beam diameter: 6 mm), and the red laser mod-

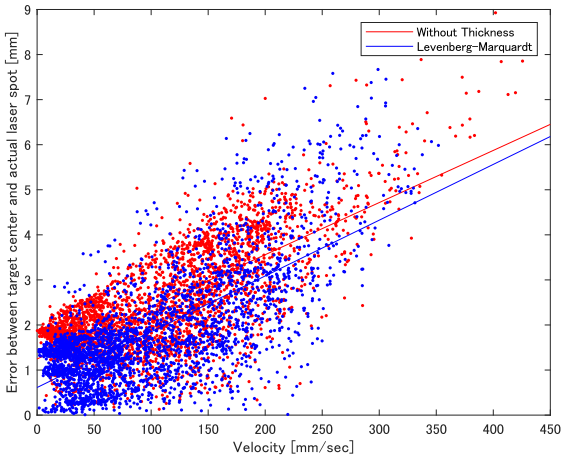

Fig. 6.13. Error between the moving speed of the board and the error of the laser irradiating position. Cited by Fig. 7 in [45].

ule was LM-101-A2 (Wen Tai Enterprise; average wavelength: 650 nm, output aperture: 3 mm, built-in spherical lens, focal length unknown), and the computer was a DELL Precision Tower 5820 (Intel Xeon Processor W-2123). The infrared light of a compact spotlight (IR3VP-8, Nissin Electronics) was diffused with a focusing lens (SH-F16) and coaxialized with the camera using a beam splitter (TS plate type B/S NIR 50R/50T, Edmund Optics). An anti-reflective material (Fine Shut Pole, KOYO Orient Japan Co., Ltd.) was used for stray light rejection.

A Nikon D3300 18-55 VR (6000×4000 px, open aperture) was used as the camera for evaluation. To estimate the pose of the camera, four spherical markers of retroreflective material (OptiTrack, 6.4 mm in diameter) were attached to the acrylic plate fixed on the outside of the camera lens in a 70 mm square, as shown in Fig. 6.12 (b). Similar to the performance evaluation experiment described in Section 6.3.3, the markers were placed on a flat board with a piece of white paper to verify the position of the laser beam. The entire system, comprising a galvanometer mirror, laser source, and camera, was calibrated using the method described in Section 6.3.1 with 500 laser irradiating points (100 points for each of the five different board postures), while the VIS cut filter was removed from the camera to verify the laser irradiating position. A black board with ArUco markers [130] lined up in the frame was used for the calibration.

Evaluation of control accuracy and speed of system

In this section, we evaluate the accuracy and speed of the proposed system, including the calibration among the galvanometer mirror, laser, and camera, and the forwarding algorithm for controlling the galvanometer mirror. These evaluations can be performed by freely moving the white board with the four spherical markers, controlling the laser irradiation to follow it, and evaluating the irradiation error. By removing the VIS cut filter of the monochrome camera, the four retroreflective material markers brightened by infrared light and the diffuse reflected light of the irradiated laser were simultaneously captured in the image. The board was allowed to move freely within the angle of view of the monochrome camera, and the laser beam followed it based on two different control methods for 15 s. The true value of the laser irradiation position was calculated based on the 3D orientation of the board recognized from the markers.

Fig. 6.13 shows data plots of the relationship between the board velocity and the laser position error for the two methods, the Levenberg–Marqruadt method, and the linear solution method; the graphs show the regression lines for each data. The parameters of the regression line and the mean z-coordinates of the moving board are shown in Table 6.1. The relationship between

Table 6.1. Parameters of the regression line shown in Fig. 6.13 and the mean value of the estimated Z coordinate of the moving board for each controlling method.

	Intercept	Slope	z [mm]
Without Thickness	1.2458	0.0116	944.1
Levenberg–Marquardt	0.6127	0.0124	935.3

Table 6.2. Time required for each process [ms].

	Ave.	Std.
Camera capture	0.4021	0.1812
Image process	0.4397	0.1986
Command calc		
Without Thickness	0.0232	0.0102
Levenberg-Marquardt	0.7353	0.3153
Vout	0.02758	0.005186
Total		
Without Thickness	0.9285	0.3664
Levenberg-Marquardt	1.5687	0.5844

the velocity of the board motion and the error was approximately proportional, and the linear solution without considering the mirror thickness had a larger error than that obtained using the Levenberg–Marqruadt method based on an accurate model.

The time required for each process for the two methods is listed in Table 6.2. The linear solution without considering the mirror thickness was less than 1 ms on average, whereas the control based on the Levenberg–Marqruadt method was approximately 1.57 ms per process on average owing to the high computational cost of iterative optimization.

Based on the processing time values, considering that the diameter of a human pupil is approximately 2˜6mm, the maximum allowable movement velocity of the pupil was approximately 152˜500 mm/s when the pupil was approximately 1 m away from the device (based on Eq. (6.7)), which is a reasonable value. However, as shown in Fig. 6.13, within these speeds, the control error was similar to the pupil diameter, which is thought to be due to the calibration accuracy. For continuous and precise observation of the laser beam, we should aim for sub-millimeter. Note that, the decrease in calibration accuracy is thought to be caused by the fact that the laser used in this project has a large and spreading spot diameter; especially when the board is tilted, the spot spreads out, preventing accurate image measurement.

Evaluation of camera directing differently

To evaluate the difference in the observed laser beam due to the difference in the incident vector of the laser to the pupil, we conducted an experiment using an evaluation camera in different orientations. The evaluation camera was focused at approximately 0, 20, 40, 60, 80, and 90 cm from the front of the device, and the laser beam was captured when the camera was facing the front of the device, and rotated several degrees in the roll, roll, tilt, and tilt directions, respectively. A ruler placed vertically at each position was used as the focusing cue. Each 3D orientation of the evaluation camera was recognized by the marker, and the control angles of the galvanometer mirror were calculated using the Levenberg–Marquardt method such that the laser beam was incident on the camera center. For each camera orientation, the incident vector of the laser beam in the camera coordinates was estimated using the calibrated model of the entire system.

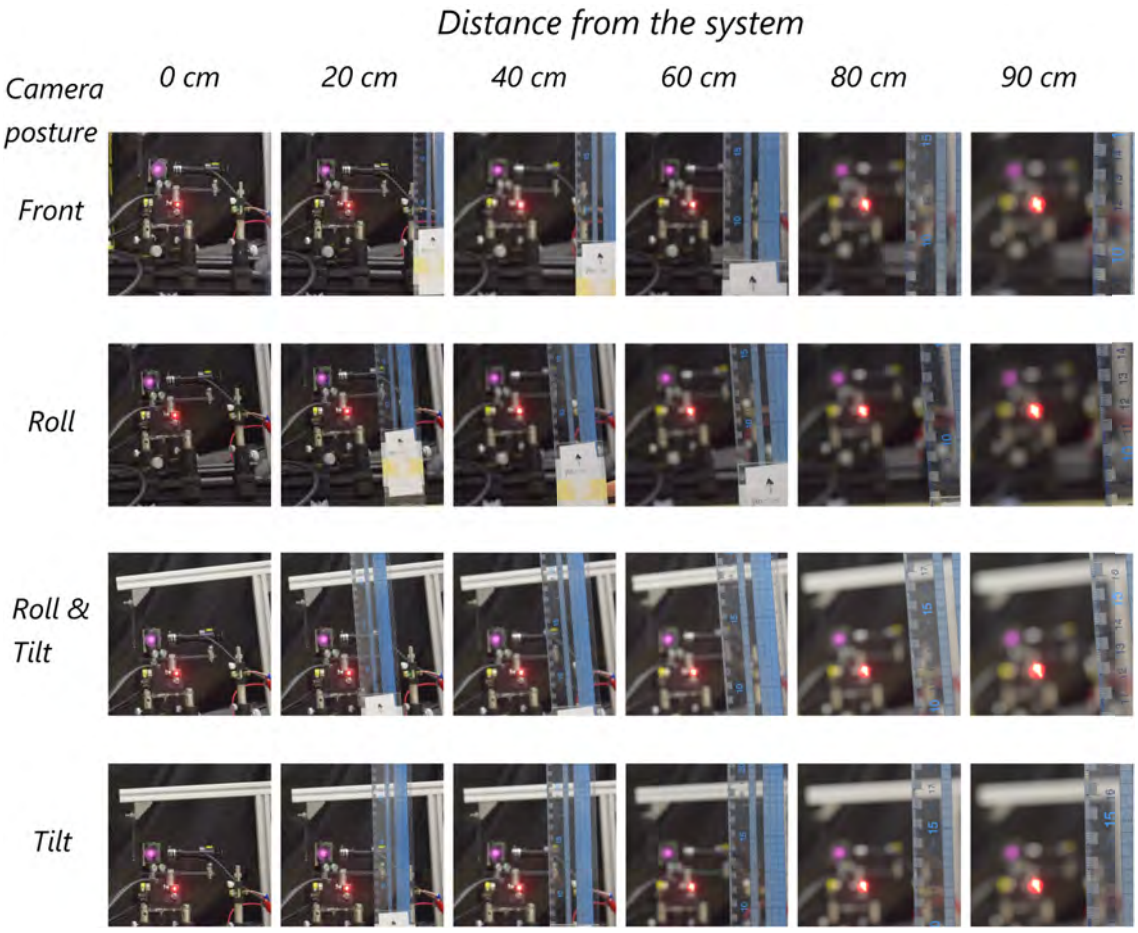

Fig. 6.14. Appearance of the laser beam from the evaluation camera at each orientation: facing the front of the device and rotating several degrees in the directions of roll, roll & tilt, and tilt. Cited by Fig. 8 in [45].

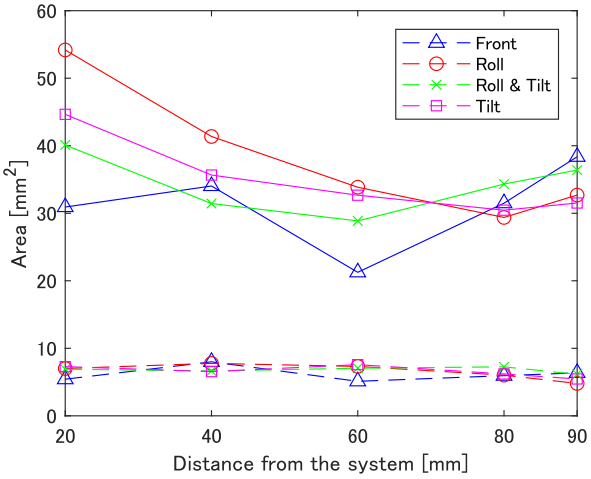

Fig. 6.15. A graph of the area of the white and red region of the laser beam in Fig. 6.14, expressed by dashed and solid lines, respectively. Cited by Fig. 9 in [45].

Table 6.3. Polar coordinates of the incidence vector of the laser in the camera coordinate at each camera orientation [°].

	Front	Roll	Roll& Tilt	Tilt
θ	2.852	8.911	7.013	6.850
ϕ	-56.44	11.07	-136.5	42.00

The results of the laser light observations are shown in Fig. 6.14. Also, a graph of the area of the white region, which is considered to be the convergence of the laser beam, and the red region, which is considered to be the blur of the laser beam, obtained by image processing according to the nearby ruler is shown in Fig 6.15. The estimated incident vector in polar coordinates for each camera orientation is listed in Table 6.3. It was confirmed that regardless of the rotational posture of the evaluation camera, the shape of the laser beam is generally the same. In particular, the white convergence area shows almost no change depending on the depth position and the camera orientation. However, even the white convergence area has about 7.5 mm^2, or a diameter of 3.1 mm at any depth position, which is too large for fine rendering. This was assumed to be caused by the laser beam that is not fully converged, as described in Section 6.3.2.

6.3.4 Conclusion

In this study, to realize the presentation of a far-field aerial image using a laser, we proposed the design of an optical system presenting a single point and conducted a basic study assuming that the laser beam is always incident on the camera. A model comprising a galvanometer mirror, laser source, and camera was established. Algorithms for calibration and forwarding control, as well as a formulation of the permissible movement speed of the pupil, were devised. Evaluation experiments indicated that the control speed of the system was reasonable, whereas the control accuracy based on calibration can be improved. We verified the visualization of the laser beam when the laser beam was controlled to be always incident on the center of the lens of the camera. We found that the laser beam was visualized with the same shape and brightness in any rotational posture of the evaluation camera, where its size was approximately the intended beam diameter of the mirror.

In the future, the size of the laser beam should be focused clearly, the calibration accuracy should be improved, and an optical system that presents images by laser scanning must be developed. Furthermore, the system should be evaluated by actual human eyes, and a presentation system for both eyes should be established.

6.4 Distant aerial image presentation using distributed multiple binocular parallax display

(Note that this section is described with a reference to the author's domestic conference proceedings paper [201], and the first person is "we" because this is joint research with co-authors.)

This research proposes a far-field aerial image presentation system based on a distributed placement of parallax barrier displays which presents a binocular parallax image, as shown in Fig 6.16 (b). By replacing the conventional large single display that covers an entire wall with a decentralized arrangement of multiple displays of general sizes or small displays, the system reduces installation costs and enables aerial image presentation in various situations beyond the traditional entertainment use. Compared to the usage of thin rays such as lasers (Section 6.3), there is an advantage that the image of each distributed display is dense and enables easy setup.

There is one crucial problem with this proposal; the parallax barrier enables naked-eye viewing,

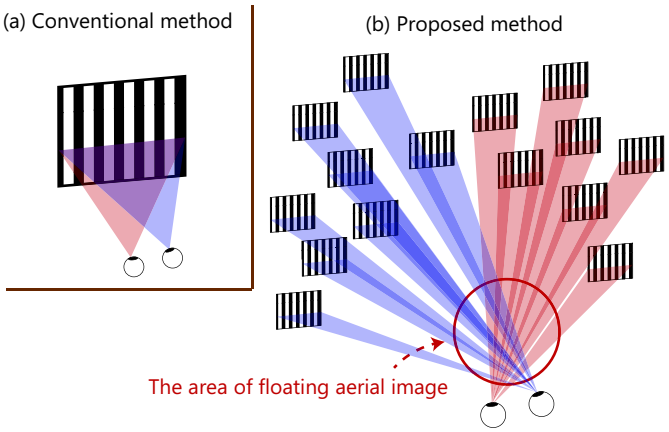

Fig. 6.16. Schematic illustration of (a) a conventional system and (b) a proposed system for far-field aerial image display. This figure is made referring to Fig. 1 in [201].

but the viewing area is narrow and limited. Then, this study spread such areas by using dynamic parallax barrier where the panel of liquid crystal display (LCD) is used for the barrier and the barrier parameter can be controlled electrically [153, 148]. Using such a dynamic barrier, the viewing area can be adjusted properly to the user's viewing position.

In this paper, we confirm the principle of barrier parameters according to several factors such as the distance between eyes and displays and distributed range, which was carefully examined through simulation.

6.4.1 Parallax barrier

Basic principle

In this section, we describe the basic principle of stereoscopic viewing with parallax barriers referring to Fig. 6.17 (a). Assuming that the distance between the eyes is e, the pixel pitch of the display is d, the number of display pixels used collectively for parallax images of each eye position is n, and the space between the display and the barrier is air, the distance t between barrier and display depending on the appropriate depth distance z_{opt} between user and barrier is obtained as follows [202].

$$t = \frac{ndz_{opt}}{e} \tag{6.8}$$

The calculation of the barrier pitch p is equal to the number of pixels times the pixel pitch of the display: almost twice nd, as shown in Fig. 6.17 (a), but noting the relationship of the triangle similarity regarding the distance of a particular pixel entering a certain eye, strictly it is obtained as follows.

$$p = \frac{2nd}{1 + nd/e} \tag{6.9}$$

From Eq. (6.8) and (6.9), the following relationship between the barrier pitch p and the user position z_{opt} is derived.

$$p = \frac{2e}{1 + z_{opt}/t} \tag{6.10}$$

Then, from Eq. (6.8) and (6.10), assuming that t is a static value, it can be said that both p, d are inversely proportional to z_{opt}. Note that α is the barrier duty ratio (ratio of barrier aperture to barrier pitch).

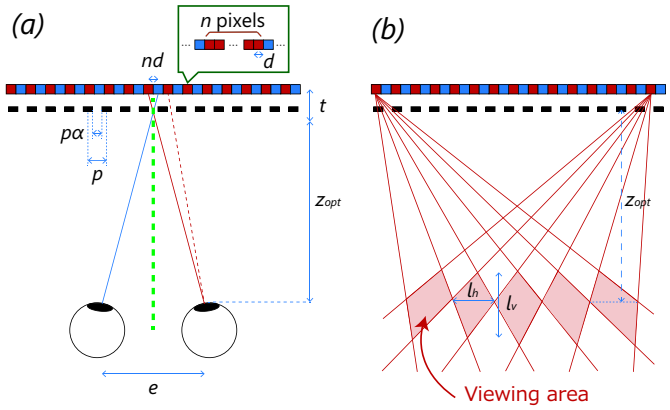

Fig. 6.17. Schematic diagram of stereoscopic viewing with parallax barriers. This figure is made referring to Fig. 2 in [201].

The region of each viewpoint created by the parallax barrier, i.e., the viewing area, is bounded by a group of viewpoint locations where both left and right images are observed in the same ratio. The viewing area consists of rectangular regions aligned parallel to the display plane and adjacent to each other at a depth position of $z = z_{opt}$, as shown in Fig. 6.17 (b). The horizontal and vertical lengths of the viewing area are l_h, l_v, respectively. Assuming that the width of the display is D (where $D = \tilde{D}$: the length is an even multiple of nd as the length of alternating pixels on the left and right), l_h and l_v are calculated as follows.

$$l_h = \frac{dz_{opt}}{t} = e \tag{6.11}$$

$$l_v \simeq 2z_{opt} \left| \frac{(\tilde{D} - 2nd)(e + nd)}{(\tilde{D} - 2nd)^2 - (e + nd)^2} \right| \tag{6.12}$$

In particular, l_h corresponds to the interpupillary distance e. Note that the viewing area that only left or right eye image can be observed is strictly a point, and the eye should be located at the center of the viewing area to clearly observe such images [152]; the other areas except for that point, the left and right images are mixed, which is called crosstalk.

It should be noted that naked-eye stereoscopic displays, including parallax barriers and lenticular lens, produce inverse stereopsis if each eye is located in the inverse viewing area. There are also methods that attempt to expand the viewing area [154], but they increase the number of horizontal divisions, resulting in a resolution reduction problem.

Configuration of dynamic parallax barrier

To address the aforementioned problem of the fixed viewable area of parallax barriers, many methods have been proposed to control images appropriately while tracking the position of the eyes or head [154]. In this study, we use dynamic parallax barriers in which the barrier parameters can be adjusted to appropriately shift the viewing area [148]. Dynamic parallax barrier uses the way that a LCD panel is taken out by disassembling a LED display and it is placed slightly forward from the other LCD [154, 153].

In particular, the shifting value of barrier in horizontal direction s to shift the viewing area horizontally is obtained as follows

$$s = \frac{t}{t + z_{opt}} h_x = \frac{nd}{nd + e} h_x \tag{6.13}$$

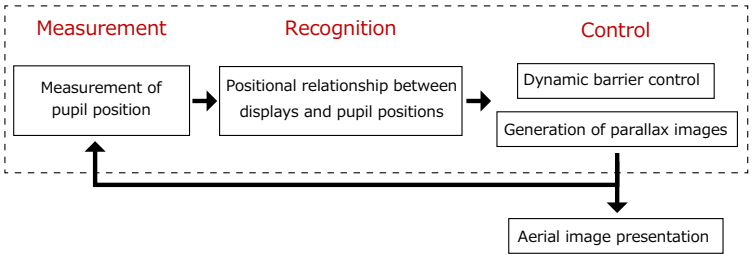

Fig. 6.18. Processing pipeline. This figure is made referring to Fig. 3 in [201].

6.4.2 Processing pipeline

As shown in Fig. 6.18, the proposed system performs a series of processes such as measuring the position of the user's pupil, recognizing the three-dimensional positional relationship between the displays and the pupil, and then presenting barriers and binocular parallax images. In this section, some of the processes are described in detail.

Pupil tracking

Since the presentation covers wide area, it is assumed that users of course move freely. To present dynamic aerial images to such a moving user, fast pupil tracking is required. For this purpose, the existing high-speed bright pupil method [198] is used.

The pupil diameter is approximately 2 ˜6 mm. The horizontal width of the visual field l_h is equal to the interpupillary distance e according to Eq. (6.11). Assuming that the ordinary length of e is 63˜65 mm, it can be said that the eye box size is sufficient for this system.

Dynamic control of parallax barriers

The horizontal shift of the barrier is controlled according to Eq. (6.13) so that the viewing area is appropriately shifted according to the measured eye position.

By contrast, this study does not control the depth of the viewing area. This is because, according to Eq. 6.11, the depth range of the viewing area, l_v, is considered to be wide enough if the presentation is distant (e.g. $z > 1.0$ m) and the displays contribute to stereoscopic presentation is moderately narrow (e.g. $D < 0.1$ m).

Parallax image generation

It is necessary to generate the parallax barrier image appropriately so that the image emerges in the intersecting area. This is done by considering the relationship between each pupil position of the user and display system, and then perspective projecting it onto each disparity display.

6.4.3 Consideration on viewing area

Fundamentally, the wider the scatter range of displays is, the larger the image can be, but the smaller the viewing area becomes. Furthermore, the barrier control is required to be accurate.

As described in the previous section, the general naked-eye binocular parallax images such as parallax barrier and lenticular lens always suffer from crosstalk, where the presented 3D image is a mixture of left and right images except for the proper viewing point. Barrier control with high accuracy is required to avoid the crosstalk, where the spatial resolution of LCD panel affects an upper limit of resolution.

In the following simulations, we consider only horizontal disparity shift. According to Eq. (6.13), the number of pixels used collectively n has a large influence on the raised problem; the

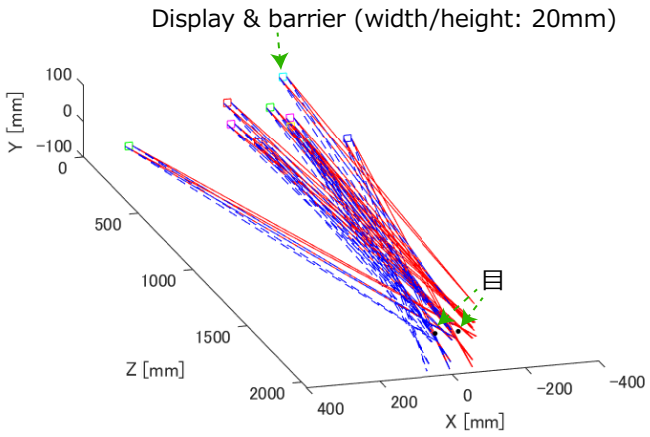

Fig. 6.19. Overall view of the system in the simulator. For readability, the number of displays in this figure is set to 10 while actually is 20. This figure is made referring to Fig. 4 in [201].

Table 6.4. Explanation and values of each parameter.

Variable	Explanation	Value
e	Interpupillary distance [mm]	63.0
z_{opt}	Optimal distance between user and display [mm]	2.00×10^3
D	Display width [mm]	20
d	Display pixel pitch [μm]	65.0
n	Number of pixels used collectively	5
t	Distance between display and barrier [mm]	10.32
p	Barrier pitch [μm]	646.7
α	Duty ratio of barrier	0.50

larger n, the smaller shift of viewing area s is enabled. However, supposing the high resolution of human vision [203], a large n may cause unintended vertical stripe patterns to be observed.

6.4.4 Evaluation experiment

In this section, we evaluate the resolution of horizontal shift of the viewing area according to the barrier shift value s and the scatter range of distributed displays through simulation. The overall system in the simulation is shown in Fig. 6.19. Twenty displays were randomly generated around the origin, and the center of the user's eyes was always at $(0, 0, 2000)$ [mm], facing the negative direction of Z axis. In addition, each display barrier was constructed with the parameters shown in Table 6.4. For simplicity, all display surfaces were assumed to be parallel to the XZ-plane and there is no user head tilt. MATLAB R2021α was used for the simulation.

6.4.5 Evaluation of horizontal shift of visual area due to barrier shift

The horizontal displacement of the viewing area h_x and the angular resolution of the display when the number of pixels used collectively n is varied between $[1, 10]$ and the barrier is moved by one pixel (d) was evaluated and its results are shown in Fig. 6.20. This graph indicates that there is a trade-off between the horizontal displacement resolution of the viewing area and the angular resolution of the display. Since the smaller the d value is, the higher the resolution of the display is. Since both parameters are related to the quality of the viewed image, it is difficult to determine

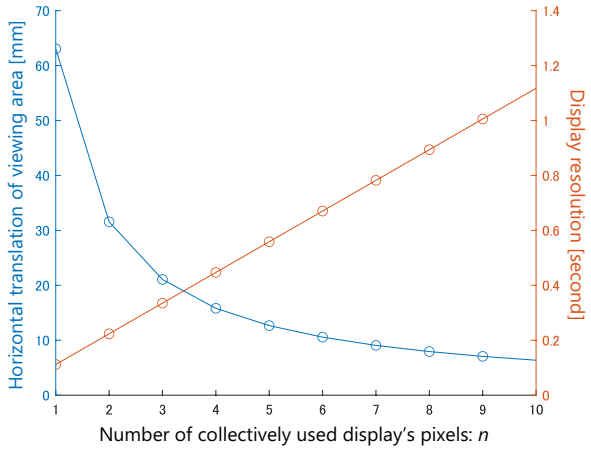

Fig. 6.20. Relationship between the number of pixels used collectively in a display n, the horizontal shift of the viewing area h_x, and the angular resolution of the display. This figure is made referring to Fig. 5 in [201].

the optimal n by this alone; a user study is considered necessary in the future.

6.4.6 Evaluation of the area of scatter and visual area

In this section, we examine the relationship between the scatter range of multiple displays and the viewing area. The horizontal (X-axis), vertical (Y-axis), and normal (Z-axis) scatter ranges of the display surface are defined as D_X, D_Y, and D_Z, respectively, and their range is defined as follows.

$$
\begin{aligned}
-d_x/2 \leq D_X \leq d_x/2, d_x &= \{200, 400, 600, 800, 1000\} \\
-50 \leq D_Y \leq 50& \\
0 \leq D_Z \leq d_z, d_z &= \{0, 100, 200, 300, 400, 500\}
\end{aligned}
$$

Note that the units are [mm].

Since D_Y is fixed to be a constant value, the evaluation was performed in $5 \times 6 = 30$ ways, where only D_X and D_Z vary. Displays were randomly generated within the scatter range, and the area of the viewing area was measured 15 times.

The result graph of the average area is shown in Fig. 6.21. Although there are fluctuations because of the randomized display position, the larger the scatter range, the smaller the visual area becomes, especially the value of D_x makes a large contribution. The maximum viewing area is 20713.9 mm^2 when $(d_x, d_z) = (200, 0)$, and the minimum one is 2963.14 mm^2 when $(d_x, d_z) = (1000, 500)$; the radii of each area is supposed to be 81.2 and 30.7 mm, respectively. Both radii are well above the pupil diameter, and the eye box size is considered to be large enough.

6.4.7 Conclusion

This study proposed a method of presenting aerial images using a distributed parallax barrier display and described its details. The performance of the system was briefly evaluated through simulation experiments. In the future, we will evaluate the allowable moving speed of a user based on the performance of the display, and then, we will select appropriate equipment and attempt to construct an actual system.

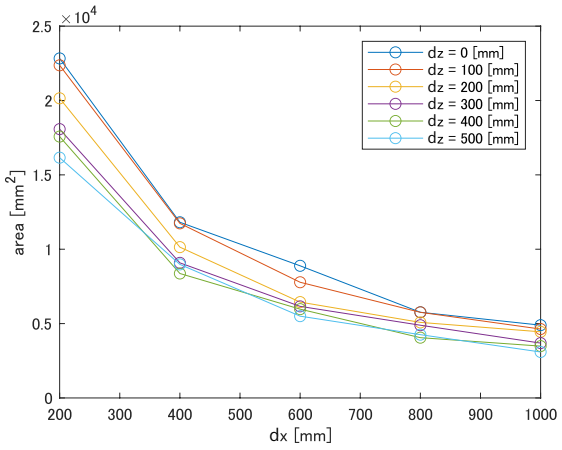

Fig. 6.21. Graph showing the relationship between the randomly generated range of distributed display and the viewing area. This figure is made referring to Fig. 6 in [201].

6.5 Summary

To realize aerial image display using distributed binocular parallax displays, two types of systems were proposed; one (Section 6.2 and 6.3) is a laser-scanning system and the other (Section 6.4) consists of multiple flat LCD monitors. Both has different advantages; the former has narrow viewing area while the latter is easy to setup. Since easy setup is the most preferable in actual situations, the latter way will be proceeded in the future. Since it is first to present binocular parallax images in the separated displays, further investigation based on user study is necessary about stereoscopic viewing and depth perception, as well as Chapter 4. Also, the allowable ratio of missing parts due to distributed displays should be examined.

Chapter 7

Multi-color LED Marker for Dynamic Target Tracking in Wide Area

The description of this section reused the author's paper [204] and the first person is "we" because this is joint research with co-authors.

7.1 Background and purpose

Fast and accurate tracking for widely moving objects is required to present augmented reality (AR) in dynamic situations like sports. In such a situation, putting tracking markers onto the target surface is suitable in terms of accuracy and speed, and their placement should be designed to be easily recognized even in low resolution.

LEDs are focused as markers in this research as they can consistently emit strong light in the same color range. It should be noted that their required power is small, which is not a problem especially in some cases, like combination with a drone. The visible colorful LEDs may limit AR image superimposition onto the target object, but this will be improved by the future usage of infrared (IR) lights of multiple wavelengths and an IR-color camera which detects different IR wavelengths [205].

Many previous methods using LED markers have taken blinking ways [206], which generally needs synchronization with the camera and is not suitable for wide tracking. Multi-color-coding way is considered in this research, which has been explored only in the printed colors and is weak against changing lighting conditions [207].

We propose multi-color LED markers for wide-area target tracking. LEDs have an advantage of stable color and luminance under any lighting condition compared to traditional color-printed markers. The marker shape is simple but can express unique IDs using a few color pixels. This marker also has an advantage in the image processing; only the bright areas of a marker can be extracted by limiting camera exposure time, which is also effective for eliminating motion blur in the fast movement. For the posture identification, the color placement on each LED is considered so that the combination of adjacent marker colors can express unique IDs. Note that each color is always static; IDs are recognized just from the relationship of adjacent marker colors.

As a first step, this research focuses on a spherical cage made of wire frame as a target object, which mounts a drone and autonomously moves around the wide field, used in a new e-sports game combined with AR [177], as shown in Fig. 7.1. This research discusses the proper arrangement of different colors and ID-recognition algorithm with a reference to previous research [58], and evaluates if our marker works in the wide range of distances and fast rotation.

Fig. 7.1. A teaser figure expressing new superhuman sports cited from Figure 1 in [177].

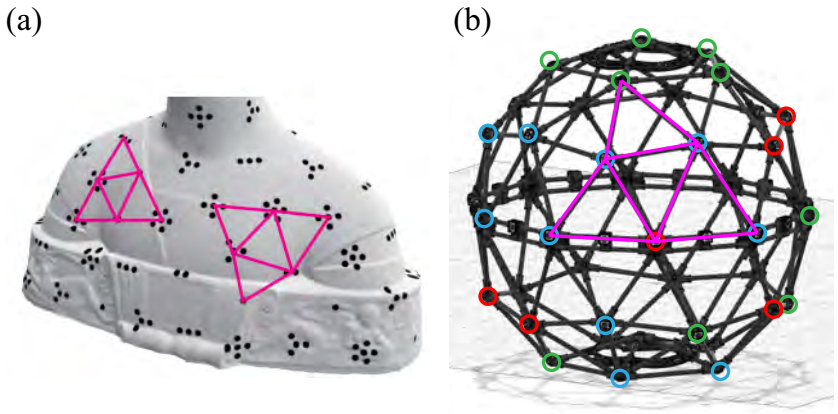

Fig. 7.2. (a) Extended dot cluster marker, cited by Figure 2 in [58]. (b) Our multi-color LED markers whose placement refers to (a).

7.2 Method

The placement of different colors and the ID-recognition algorithm refer to a previous study of extended dot cluster marker (EDCM) [58], as shown in Figure 7.2 (a). This previous marker consisted of different number of dots, and its unique ID could be recognized using adjacent four triangles providing six markers. We replace this dot cluster with a multi-color LED.

One different factor in our study compared to the previous one [58] is, that our adjacent markers can have the same color. There are two reasons for it. One reason is that if a part of the LED markers is occluded, the color does not greatly change and the tracking performance is not affected; the previous method may be susceptible to this, because if a part of a cluster marker is occluded by hands and fingers, the recognized number of dots within that marker might be reduced. The second reason is that, since some adjacent markers would have the same color, the number of different colors can be reduced, which leads to enhance the color-extraction accuracy.

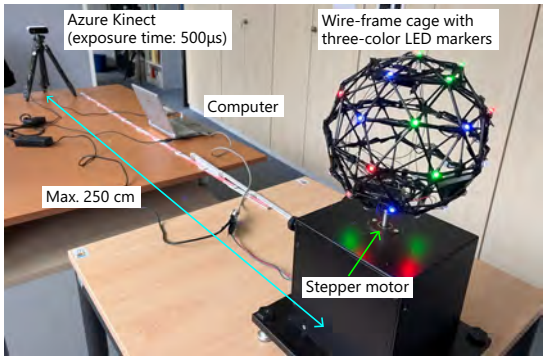

Fig. 7.3. Evaluation apparatus.

In this study, we target a cage made of triangle shapes for e-sports [177] as shown in Figure 7.2 (b). By brute-forth search to find unique arrangement of markers, it was revealed that our LED marker can work with only three different colors. This is owing to the sparse wire frame of target cage; if a large number of LEDs are put onto the target object, the necessary number of different colors would increase.

Note that, our method is also different from the previous one in the marker placement; the number of adjacent markers for one marker is not always six, sometimes five, while it was always six in the previous method. This allows to keep the usage of four adjacent triangles while loosing the restriction of marker placements.

As the other novel point, our methods require to set the camera's exposure time to be short to detect only the brightest points and eliminate motion blur even in the fast object movement.

7.3 Evaluation

7.3.1 Evaluation environment

The picture of the evaluation environment is shown in Figure 7.3.

The multi-color LEDs (WS2812B, BTF-LIGHTING), each color is programmable, were put onto a cage that was made of black triangle wire frames assuming the usage in the new e-sports [177]; the wire frame is sparsely placed. The three colors of LEDs were red, green, and blue. The cage diameter was approx. 23 cm. A stepper motor (Nema 17, OSM Technology Co.,Ltd.) with a controller (Stepper Brick DRV8811, Tinkerforge GmbH) was used for automatic rotation. The used notebook computer was Panasonic *Let's note* CF-SV (Intel® Core™ i7-8565U).

The used camera was Azure Kinect (Microsoft, 1280×720 px, 30 fps, color, field-of-view (FOV): W90°×H59°), whose exposure time was set to 500 μs. The cage was fixed on the stepper motor, and the distance between the camera and the center of cage varied at 0.5, 1.0, 1.5, 2.0, and 2.5 m. The height of the camera and the cage was always the same.

Since the FOV of the used camera is too wide, it is helpful to consider the conversion of distance depending on the camera's FOV. Let the cage radius be r (=11.5 cm) and the cage distance from the principal point of camera be d. By making the cage to occupy the same number of image pixels at any distance, the converted distance d' at different FOV ($= fov'$) from original distance d at original FOV ($= fov$) can be formulated as

$$d' = \frac{r}{\tan(fov' * \theta/2)} \tag{7.1}$$

$$\text{where } \theta = \frac{2\arctan(r/d)}{fov}. \tag{7.2}$$

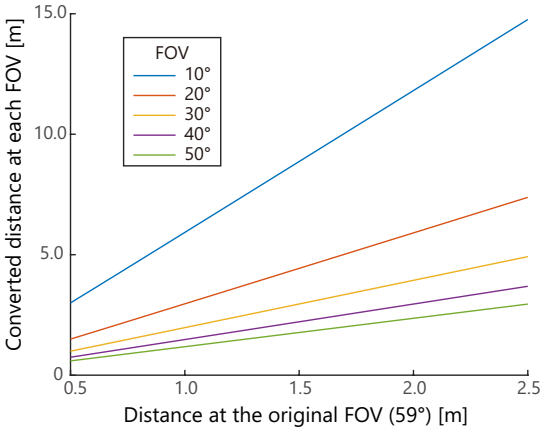

Fig. 7.4. The relationship expressing a distance d' at a certain FOV (fov') (field-of-view) converted from a distance d at an original FOV ($fov = 59°$).

Table 7.1. The success rate of ID recognition of multi-color LED markers at various distances

Distance [m]	0.5	1.0	1.5	2.0	2.5
Success rate [%]	100.0	100.0	99.0	99.0	83.0

The graph expressed by these formulas is shown in Figure 7.4. For example, assuming that the original FOV is $fov = 59°$, the vertical FOV of the used camera, the success rate at d =2.0 m in our system would be converted into (d', fov') = (3.937 m, 30°) and (11.81 m, 10°), which is said to be a sufficient range of distance.

7.3.2 Evaluation of tracking

The cage rotated at 9° steps all around, and 5 frames were captured by the camera for each rotation, resulting in 200 frames in total. The ID recognition was conducted for each frame, whose result was directly drawn onto the image and used for the manual judgement of tracking success by the corresponding author.

The evaluation result is shown in Table 7.1. The success rate was almost 100% at 0.5-2.0 m, and it dropped to 83.0 % at 2.5 m. The examples of captured and result-drawn images at 0.5 and 2.5 m are shown in Figure 7.5 (a) and (b). The cage diameter in the image at 0.5 m distance was 330 px, much smaller than the image height (720 px), which implies the possibility of tracking at further small distance in small FOV. That at 2.5 m was 60 px; the success rate over 80 % is substantial even with such a small image size.

The fast rotation at 300 RPM was also tested, and its appearances from both iPhone 13 Pro (30 fps, the estimated exposure time: around 30 ms) and Azure Kinect are shown in Figure 7.5 (c). The latter has no motion blur thanks to short exposure time while the former has it.

The processing time was calculated as 4.950 ms on average with standard deviation of 1.811 ms using a 1000-frame recorded video of the cage rotating at 22.5°/s. This means almost 200 fps tracking is possible, which is sufficient for normal AR superimposition.

7.4 Discussion and conclusion

It is notable that the bright colorful LEDs work as tracking markers and a camera of limited exposure time facilitates their easy detection and ID recognition. This method is also robust against the specular reflection since LEDs are further bright. Though, sunlight may obstruct them

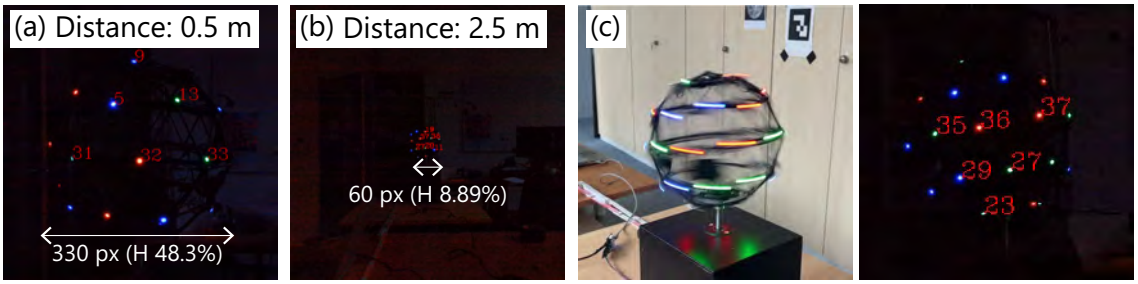

Fig. 7.5. (a,b) Example of captured images using Azure Kinect at 50 and 250 cm, respectively. (c) Captured images of fast rotation (300 RPM); left shows a picture taken by iPhone 13 Pro at 30 fps while right shows the image taken by Azure Kinect with short exposure time at 100 cm distance.

because of similar luminance, hence the indoor usage is preferable.

The cause of false detection at a distance is estimated due to mistriangulation, which is owing to missing LEDs caused by a lack of LED luminance or too short exposure time. However, this would be improved by using the camera of a narrow FOV, as discussed in Section 7.3.1 and 7.3.2.

Chapter 8

Conclusion

This thesis presented several novel proposals to solve two main problems of current AR research: the wide presentation area and the improvement in stereoscopic vision. Stationary system was used in each research of this thesis to present high-speed SAR (projection mapping and aerial image) which allows a high computational cost necessary for precise and fast measurement and recognition and leads immersive experience in the future. This thesis introduced several novel ideas in design of tracking marker and algorithm, high-speed and accurate image presentation devices, and the investigation of human visual perception. Furthermore, new and socially beneficial applications were explored, especially in the field of sports.

The conclusions of each chapter are described below.

In Chapter 3, circumferential markers for wide-area dynamic projection mapping on a sphere was proposed, which is new in that it focuses on the shape of a circle, as compared to conventional markers with complicated shapes. Many methods have used circular shapes for tracking and calibrations. Coins [97] and pupil [96] was tracked utilizing elliptic shape, which might impair accuracy due to as many as five parameters of ellipse. Circle and ring array patterns were proposed for blur-resistant calibration [80, 81, 82, 77], and the calibration way that utilizes circles found in daily life such as cups [78] and vases [79] was also proposed. The proposal of this thesis, by contrast, firstly uses multiple circular markers for high-speed and accurate tracking, which is called circumferential markers in the point of having narrow width. This thesis introduces another new approach to express the code in the circumferential markers by shifting the outer circle of the circumferential marker with respect to the inner circle, which enables absolute posture estimation. The several evaluations confirmed its performance and the demonstration of sports motion visualization and spherical displays showed its beneficial applications.

In Chapter 4, the ocular parallax was focused that it is generally produced according to the viewpoint shift by involuntary eye rotation such as saccade, which has been ignored in conventional VR/AR stereoscopic system. One previous study reproduced it using state-of-art commercial head-mounted display (HMD) [36], but it suffers from large delay from eye position measurement to stereoscopic image presentation and confirmed only the enhance in reality feelings. This thesis firstly assembled a system of high-speed eye tracking and presentation of binocular parallax images from scratch, where the novel approximation of viewpoint was introduced as confirmed to be valid using elaborate simulation. The user study of binocular fusion and depth perception was conducted; the former had a significant difference while the latter did not in the analysis using generalized linear mixed model (GLMM) [193], but the graph seemed to have a large individual difference. Note that ocular parallax is one of gaze-contingent displays, which have been popular as followed with the recent trend of eye trackers that can be easily integrated into HMD.

In Chapter 5, the fact was focused that the most popular eye tracking method, the pupil-centered corneal reflection (PCCR) [40], is not resistant to wide-angle rotation of the eye because the corneal reflections (CRs) are easily kicked and missed according to eye rotation. I proposed an algorithm for identification of each CR using the simplified eye model [43] while conventional

methods relied on deep-learning method [126] or the approximated formula [119]. Since the continuous tracking of CRs is also important for continuously rotating eyes, high-speed ($\tilde{1}000$ fps) continuous tracking and identification were performed. The evaluation on seven Asian people confirmed its sufficient operation with a moderate individual difference.

In Chapter 6, the concept of distributed displays was proposed for wide-area aerial image presentation. Conventional aerial displays have forced to use a huge screen to present an aerial image in wide area, which has limited the usage only in the closed space such as cinema. By contrast, this method enables flexible placement even in open space such as parks, station, pedestrian crossing, and tunnels. Two methods to actually conduct them were proposed: laser scanning and binocular parallax displays such as dynamic parallax barriers [153, 148]. The former, the laser scanning method, was tested on a real device to present an aerial image of a point while the latter was only verified by simulation.

In Chapter 7, the multi-color LED markers were proposed for the robust tracking for the widely moving object supposing that the power usage is allowed. This method can realize wide-area tracking as well as Chapter 3 having robustness against image blur and low resolution thanks to consistent light with the same color range. Furthermore, the target shape is not limited to spheres because of the flexible design of point-based LED markers. The detection and identification algorithm were proposed as similar way to extended dot cluster marker (EDCM) [58], and the evaluation at the various distance between the target object and the color camera confirmed its accuracy and validity as well as its tracking speed.

Thus, this paper covers a wide range of advanced AR research, from tracking techniques to 3D image presentation, and investigates stereoscopic perception and beneficial applications, with a particular technical focus is on high speed and wide area. I hope that this paper as a whole will make a major step forward in AR technology, and further developments both in technical and perceptual terms are expected in the future.

Acknowledgements

In completing my doctoral thesis, I would like to express my sincere gratitude to all the professors, laboratory members, colleagues, and family members who supported me during my three years in the doctoral program.

My supervisor, Prof. Hiroyuki Shinoda, taught me many essentials of research that were eye-opening to me. I felt at the first meeting that what he said was like a jewel box, and that feeling has persisted over the past three years. He taught me a wide range of essentials for research such as the way of thinking outside the box, the process of research that minimizes time, effort, and monetary costs, how to write effective arguments, and the techniques for creating breakthrough research that contributes to the society significantly. I also greatly respect the fact that, despite his outstanding talent and the many things he has accomplished, he has a humble attitude and does not hold this in high esteem. I also appreciated that he allowed me to freely express my opinions, both at the meeting and in personal communication, even though I was still a little green. I think that all of my opinions were not well understood, hasty and immature, but he always took them seriously, preached them enthusiastically, and conveyed his sincere desire to nurture me. I believe that his natural character is also an important factor that makes the laboratory harmonious, and it is a point that I hope will never change. He also helped us with many administrative procedures related to budget execution, scholarships, and salaries, which was helpful. I am also grateful for the generous budget and facilities of the laboratory. In particular, the machine tools made it possible for me to enjoy the pleasure of making things with my own hands many times over.

My associate advisor, Assoc. Prof. Yasutoshi Makino, always inspired me with his flexible and novel ideas. He was also close to his students and created a friendly atmosphere in the laboratory. Also, Dr. Masahiro Fujiwara, a lecturer at Nanzan University and previously Project Research Associate at my laboratory, was the closest faculty stuff to the students during the two and a half years of my doctoral student life. He took meticulous care of me by reviewing my thesis and budget applications, attending to my interview practice, and taught me many important research essentials. I would like to express my sincere gratitude to both.

The secretary of the laboratory, Yasuko Konagai, and the administrative staff of the Graduate School of Frontier Sciences took care of my budget, salary, studying abroad, and other administrative matters, which caused them a lot of trouble. All of these procedures were necessary for me to proceed with my research. I would like to thank you them all.

I would also like to thank the Ishikawa Group Laboratory, to which I belonged during my master's course. During my first year as a doctoral student, I belonged to the group as a technical assistant and could produce one of my most important papers. I would also like to express my sincere gratitude for renting some equipment, including those that I could not obtain due to the shortage of semiconductors. I had a good time with Project Prof. Masatoshi Ishikawa, who encouraged me and gave me a lot of motivation to do my research.

I would like to express my sincere gratitude to my colleagues, seniors, and juniors in the Shinoda Makino Laboratory, as well as to the assistant staff and my peers who entered the doctoral program with me from the master's program in the Ishikawa Group Laboratory. I enjoyed my research life in Hongo campus with the junior students and had a good break from my research. Outside the laboratory, there are seniors and peers with whom I have had flat discussions across the humanities and sciences, and I would like to express my sincere gratitude to them as well.

I would like to thank Prof. Gudrun Klinker of Forschungsgruppe Augmented Reality (FAR) group at the Technical University of Munich, where I have been studying during my third year of the doctoral program, for accepting my sudden request for acceptance and for always being kind to me during my stay. I would like to thank Mr. Christian Eichhorn, with whom I have worked closely on the LED tracking project and had many discussions. Even though he was busy, he was always attentive to our discussions, which made my first international project smooth and enjoyable. I would like to thank my colleagues Ms. Linda Rudolph and Mrs. Chloé Eghtebas. They always helped me with anything I did not understand for both the laboratory and the life in Germany. I was encouraged by the presence of these two career-minded strong women assuming that there are few women in information sciences. We had a good time during my mindless study abroad, chatting about other things in between research, going out to dinner, and so on. I would also like to express my sincere gratitude to Assoc. Prof. Yuta Itoh for introducing this laboratory to me.

I would also like to thank Dr. Taiki Fukiage of NTT Communication Science Laboratories, who gave me a lot of accurate advice for my research on ocular parallax. I look forward to working with you from the next year.

I would also like to thank the Japan Society for the Promotion of Science (JSPS), the Japan Science and Technology Agency (JST), the Toyota-Dwango Advanced Artificial Intelligence Scholarship, and Microsoft Research Asia for their budget and financial support during my doctoral program.

Finally, I would like to thank my family for their support over the past three years. Once again, I express my sincere thanks to all those who have supported and encouraged me.

Publication

Journals

1. Yuri Mikawa, Tomohiro Sueishi, Yoshihiro Watanabe, and Masatoshi Ishikawa: Dynamic Projection Mapping for Robust Sphere Posture Tracking Using Uniform/Biased Circumferential Markers, IEEE Transactions on Visualization and Computer Graphics, doi: 10.1109/TVCG.2021.3111085.
2. 三河祐梨，牧野泰才，篠田裕之: 深層学習を用いた押下動作映像からの硬さ推定, 日本バーチャルリアリティ学会論文誌, Vol.23, No. 4, pp.239-248 (2018) (Japanese)

Awards

1. 2022 年 学生奨励賞, 映像情報メディア学会 立体映像技術研究会 (3DIT)
2. SICE Young Author's Award, SICE Annual Conference 2021

International Conference

Invited oral presentation

1. Yuri Mikawa: Dynamic Projection Mapping for Robust Sphere Posture Tracking Using Uniform/Biased Circumferential Markers, ACM SIGGRAPH Asia 2022, Daegu, South Korea, Dec. 7, 2022. (Invited as a few top-quality paper among all the published TVCG papers in 2022)
2. Yuri Mikawa, Tomohiro Sueishi, Yoshihiro Watanabe, and Masatoshi Ishikawa: Dynamic Projection Mapping for Robust Sphere Posture Tracking Using Uniform/Biased Circumferential Markers, IEEE Conference on Virtual Reality and 3D User Interfaces (VR), Virtual (Christchurch, New Zealand), 16th, Mar. 2022.

Oral presentation

1. Yuri Mikawa, Masahiro Fujiwara, Takefumi Hiraki, Yasutoshi Makino, Hiroyuki Shinoda: Far-field Aerial Image Presentation of One Point by a Laser Source using Beam Scanning by Two-axis Galvanometer Mirror, in Proceedings of SICE Annual Conference 2021, pp. 137-143, Sep. 8, 2021, Online / Regular Paper [SICE Young Author's Award]
2. Yuri Mikawa, Masahiro Fujiwara, Hiroyuki Shinoda: Distant Aerial Image Presentation Using Two Separate Rays of Binocular Parallax Images, The 27th International Display Workshops (IDW '20) (Online, 2020.12.11) / Proceedings, 3D2-3.
3. Yuri Mikawa, Tomohiro Sueishi, Yoshihiro Watanabe, and Masatoshi Ishikawa: Projection mapping system to a widely dynamic sphere with circumferential markers, IEEE International Conference on Multimedia and Expo (ICME2020) (London, 2020.07.09), pp. 1-6.
4. Yuri Mikawa, Tomohiro Sueishi, Tomohiko Hayakawa, Masatoshi Ishikawa: Laser-based

drawing method for posture-free objects by photochromic active marking with high-speed coaxial gaze control, SPIE Photonics West 2020 (San Francisco, California, USA. 2020.02.05)/ Oral Session

Demonstration

1. Yuri Mikawa, Tomohiro Sueishi, Yoshihiro Watanabe and Masatoshi Ishikawa: Vario-Light: Hybrid Dynamic Projection Mapping Using High-speed Projector and Optical Axis Controller, SIGGRAPH Asia 2018 Emerging Technologies (p.17), Tokyo, Japan, 4-7 Dec. (2018)

Poster

1. Yuri Mikawa, Christian Eichhorn and Gudrun Klinker: Multi-color LED Marker for Dynamic Target Tracking in Wide Area, 2023 IEEE Conference on Virtual Reality and 3D User Interfaces (VR), Shanghai, China, 25-29 Mar. (2023) [to appear]
2. Yuri Mikawa, Masahiro Fujiwara, Yasutoshi Makino, Hiroyuki Shinoda: Fast and Low-Latency Ocular Parallax Rendering Improves Binocular Fusion in Stereoscopic Vision, 2023 IEEE Conference on Virtual Reality and 3D User Interfaces (VR), Shanghai, China, 25-29 Mar. (2023) [To appear]
3. Yuri Mikawa, Tomohiro Sueishi, Tomohiko Hayakawa and Masatoshi Ishikawa: Laser-based Photochromic Drawing Method for Rotating Objects with High-speed Visual Feedback, 2019 IEEE Conference on Virtual Reality and 3D User Interfaces (VR), Osaka, Japan, 23-27 Mar. (2019)

Domestic Conference

Oral presentation

1. 三河祐梨, 篠田裕之: 分散配置の視差ディスプレイによる広域空中像に向けた欠損視差映像による立体視の品質の調査, 映像情報メディア学会 立体メディア技術研究会（3DMT）, 2022 年 3 月 6–7 日 [to appear]
2. 三河祐梨, 藤原正浩, 牧野泰才, 篠田裕之: 高速・低遅延な Ocular parallax レンダリングによる立体知覚への効果の調査, 映像情報メディア学会 立体メディア技術研究会（3DMT）, 2022 年 3 月 6–7 日 [to appear]
3. 阿部拓実, 三河祐梨, 藤原正浩, 牧野泰才, 篠田裕之: 特定方向からの接触に対する超音波触覚提示に適した複数カメラによる指先センシング, 第 23 回計測自動制御学会システムインテグレーション部門講演会 (千葉, 2022 年 12 月 14-16 日)
4. 阿部拓実, 三河祐梨, 藤原正浩, 牧野泰才, 篠田裕之: 超音波触覚提示のための複数カメラによる指先 3 次元位置高精度センシングの検証, 計測自動制御学会 第 39 回センシングフォーラム (東京, 2022 年 9 月 21 日) ／ 1B3-4
5. 三河祐梨, 藤原正浩, 牧野泰才, 篠田裕之: 視点移動対応の分散型両眼視差ディスプレイによる遠方空中像提示の検討, 映像情報メディア学会 立体映像技術研究会 (3DIT), 2022 年 3 月 7 日 [学生奨励賞]
6. 佐竹空良, 三河祐梨, 藤原正浩, 牧野泰才, 篠田裕之: 空中超音波触覚による移動刺激の手掌部における二点弁別閾の調査, 第 22 回計測自動制御学会システムインテグレーション部門講演会論文集, 2F1-01, オンライン, 2021.12.16.
7. 三河祐梨, 末石智大, 石川正俊: 球体姿勢に対応した回転相殺テクスチャの高速投影の残像効果による一軸回転可視化法の提案, 第 26 回日本バーチャルリアリティ学会大会

(VRSJ2021) (オンライン, 2021.9.13) ／論文集, 2D2-5.
8. 三河祐梨, 鈴木大河, 藤原正浩, 牧野泰才, 篠田裕之: 分散配置の動的パララックスバリアによる遠方空中像ディスプレイの基礎検討, 第 26 回日本バーチャルリアリティ学会大会 (VRSJ2021) (オンライン, 2021.9.12) ／論文集, 1D3-1.
9. 三河祐梨, 末石智大, 渡辺義浩, 石川正俊: VarioLight2: 円周マーカを用いた球体への広域かつ遮蔽に頑健なダイナミックプロジェクションマッピング, 第 27 回画像センシングシンポジウム (SSII2021) (オンライン, 2021.6.9) ／講演論文集 IS1-25.
10. 神宮亜良太, 三河祐梨, 藤原正浩, 牧野泰才, 篠田裕之: 空中集束超音波を用いた唇部への非接触触覚提示の基礎的検討, 第 25 回日本バーチャルリアリティ学会大会 (VRSJ2020) (オンライン, 2020.9.17) ／論文集, 2D3-3.
11. 三河祐梨, 末石智大, 早川智彦, 石川正俊：フォトクロミズムと高速光軸制御による発色型ファブリケーションに向けた動的描画システム, 第 24 回日本バーチャルリアリティ学会大会 (VRSJ2019) (東京, 2019.9.11) ／論文集, 3C-03.
12. 三河祐梨, 末石智大, 渡辺義浩, 石川正俊: VarioLight：高速プロジェクタ及び光軸制御による非対称な移動物体への投影型拡張現実感システム, 第 23 回日本バーチャルリアリティ学会大会 (VRSJ2018) (仙台, 2018.9.19) ／論文集, 14D-4.
13. 三河祐梨, 末石智大, 石川正俊: 動的プロジェクションマッピングに向けた輪郭情報に基づく高速球体トラッキング, 第 22 回日本バーチャルリアリティ学会大会 (VRSJ2017) (徳島, 2017.9.27) ／論文集, 1E4-03.

Demonstration

1. 三河祐梨, 末石智大, 石川正俊: 球体姿勢に対応した回転相殺テクスチャの高速投影の残像効果による一軸回転可視化法の提案, 第 26 回日本バーチャルリアリティ学会大会 (VRSJ2021) 技術展示 (オンライン, 2021.9.13) ／ 2G-6

Patents

1. 末石智大, 三河祐梨, 宮地力, 石川正俊: ”情報処理システム、情報処理装置、情報処理方法及びプログラム,” 特開 2021-190024(P2021-190024A), 日本, 2020 年 6 月 4 日

Bibliography

[1] Tomohiro Sueishi, Hiromasa Oku, and Masatoshi Ishikawa. Lumipen 2: Dynamic projection mapping with mirror-based robust high-speed tracking against illumination changes. *Presence*, Vol. 25, No. 4, pp. 299–321, 2016.

[2] Yuri Mikawa, Tomohiro Sueishi, Yoshihiro Watanabe, and Masatoshi Ishikawa. VarioLight: Hybrid Dynamic Projection Mapping Using High-speed Projector and Optical Axis Controller. In *SIGGRAPH Asia 2018 Emerging Technologies*, p. 17. ACM, 2018.

[3] Leo Miyashita, Kota Ishihara, Yoshihiro Watanabe, and Masatoshi Ishikawa. Zoematrope: A system for physical material design. In *ACM SIGGRAPH 2016 Emerging Technologies*, pp. 1–1. 2016.

[4] Takatoshi Yoshida, Yoshihiro Watanabe, and Masatoshi Ishikawa. Phyxel: Realistic display of shape and appearance using physical objects with high-speed pixelated lighting. In *Proceedings of the 29th Annual Symposium on User Interface Software and Technology*, pp. 453–460, 2016.

[5] Leo Miyashita, Yoshihiro Watanabe, and Masatoshi Ishikawa. Midas projection: Markerless and modelless dynamic projection mapping for material representation. *ACM Transactions on Graphics (TOG)*, Vol. 37, No. 6, pp. 1–12, 2018.

[6] Daniel Saakes, Kevin Chiu, Tyler Hutchison, Biyeun M Buczyk, Naoya Koizumi, Masahiko Inami, and Ramesh Raskar. Slow display. In *ACM SIGGRAPH 2010 Emerging Technologies*, pp. 1–1. 2010.

[7] Tomoko Hashida, Yasuaki Kakehi, and Takeshi Naemura. Photochromic sculpture: volumetric color-forming pixels. In *ACM SIGGRAPH 2011 Emerging Technologies*, pp. 1–1. 2011.

[8] Yuhua Jin, Isabel Qamar, Michael Wessely, and Stefanie Mueller. Photo-chromeleon: Reprogrammable multi-color textures using photochromic dyes. In *ACM SIGGRAPH 2020 Emerging Technologies*, pp. 1–2. 2020.

[9] Microsoft. Microsoft HoloLens — Mixed Reality Technology for Business, 2022.

[10] Magic Leap Inc. Magic Leap One, 2023.

[11] SONY. Spatial Reality Display, 2021.

[12] Looking Glass Factory: The Hologram Company. Looking glass facotry, March 2022.

[13] Philipp Kurth, Markus Leuschner, Marc Stamminger, and Frank Bauer. Content-aware brightness solving and error mitigation in large-scale multi-projection mapping. *IEEE Transactions on Visualization and Computer Graphics*, 2022.

[14] Hanyuool Kim, Issei Takahashi, Hiroki Yamamoto, Satoshi Maekawa, and Takeshi Naemura. Mario: Mid-air augmented reality interaction with objects. *Entertainment Computing*, Vol. 5, No. 4, pp. 233–241, 2014.

[15] Yuta Itoh, Takumi Kaminokado, and Kaan Akşit. Beaming displays. *IEEE Transactions on Visualization and Computer Graphics*, Vol. 27, No. 5, pp. 2659–2668, 2021.

[16] Kohei Okumura, Keiko Yokoyama, Hiromasa Oku, and Masatoshi Ishikawa. 1 ms auto pan-tilt–video shooting technology for objects in motion based on saccade mirror with background subtraction. *Advanced Robotics*, Vol. 29, No. 7, pp. 457–468, 2015.

[17] Yuri Mikawa. Projection-based and color-forming augmented reality system using optical 3d tracking. Master's thesis, The University of Tokyo, Hongo, Bunkyo-ku, Tokyo, Japan,

3 2020.

[18] Hanhoon Park and Jong-Il Park. Invisible marker tracking for AR. In *Third IEEE and ACM International Symposium on Mixed and Augmented Reality*, pp. 272–273. IEEE, 2004.

[19] 宮崎大介. 三次元像の空中表示技術の現状と展開. 光学, Vol. 40, No. 12, pp. 608–615, 2011.

[20] Toru Iwane. Light field display and 3d image reconstruction. In *Three-Dimensional Imaging, Visualization, and Display 2016*, Vol. 9867, pp. 214–225. SPIE, 2016.

[21] Fahri Yaraş, Hoonjong Kang, and Levent Onural. State of the art in holographic displays: a survey. *Journal of display technology*, Vol. 6, No. 10, pp. 443–454, 2010.

[22] Gregg E Favalora, Joshua Napoli, Deirdre M Hall, Rick K Dorval, Michael Giovinco, Michael J Richmond, and Won S Chun. 100-million-voxel volumetric display. In *Cockpit Displays IX: Displays for Defense Applications*, Vol. 4712, pp. 300–312. SPIE, 2002.

[23] Hideo Saito, Hidei Kimura, Satoru Shimada, Takeshi Naemura, Jun Kayahara, Songkran Jarusirisawad, Vincent Nozick, Hiroyo Ishikawa, Toshiyuki Murakami, Jun Aoki, et al. Laser-plasma scanning 3d display for putting digital contents in free space. In *Stereoscopic Displays and Applications XIX*, Vol. 6803, pp. 93–102. SPIE, 2008.

[24] Yoichi Ochiai, Kota Kumagai, Takayuki Hoshi, Jun Rekimoto, Satoshi Hasegawa, and Yoshio Hayasaki. Fairy lights in femtoseconds: aerial and volumetric graphics rendered by focused femtosecond laser combined with computational holographic fields. *ACM Transactions on Graphics (TOG)*, Vol. 35, No. 2, pp. 1–14, 2016.

[25] Takaki Yasuhiro. *Spatial 3D Presentation and User Interface (Japanese)*. S&T Publishing Inc., Tokyo, Japan, 2019.

[26] Andrew T Duchowski, Donald H House, Jordan Gestring, Rui I Wang, Krzysztof Krejtz, Izabela Krejtz, Radosław Mantiuk, and Bartosz Bazyluk. Reducing visual discomfort of 3d stereoscopic displays with gaze-contingent depth-of-field. In *Proceedings of the acm symposium on applied perception*, pp. 39–46, 2014.

[27] David Dunn, Cary Tippets, Kent Torell, Petr Kellnhofer, Kaan Akşit, Piotr Didyk, Karol Myszkowski, David Luebke, and Henry Fuchs. Wide field of view varifocal near-eye display using see-through deformable membrane mirrors. *IEEE transactions on visualization and computer graphics*, Vol. 23, No. 4, pp. 1322–1331, 2017.

[28] Christoph Ebner, Shohei Mori, Peter Mohr, Yifan Peng, Dieter Schmalstieg, Gordon Wetzstein, and Denis Kalkofen. Video see-through mixed reality with focus cues. *IEEE Transactions on Visualization and Computer Graphics*, Vol. 28, No. 5, pp. 2256–2266, 2022.

[29] Brian Guenter, Mark Finch, Steven Drucker, Desney Tan, and John Snyder. Foveated 3d graphics. *ACM Transactions on Graphics (TOG)*, Vol. 31, No. 6, pp. 1–10, 2012.

[30] Hiroaki Kudo and Noboru Ohnishi. Study on the ocular parallax as a monocular depth cue induced by small eye movements during a gaze. In *Proceedings of the 20th Annual International Conference of the IEEE Engineering in Medicine and Biology Society. Vol. 20 Biomedical Engineering Towards the Year 2000 and Beyond (Cat. No. 98CH36286)*, Vol. 6, pp. 3180–3183. IEEE, 1998.

[31] Hiroaki Kudo, Masaya Saito, Tsuyoshi Yamamura, and Noboru Ohnishi. Measurement of the ability in monocular depth perception during gazing at near visual target-effect of the ocular parallax cue. In *Proceedings of 1999 IEEE International Conference on Systems, Man, and Cybernetics (SMC)*, Vol. 2, pp. 34–37. IEEE, 1999.

[32] Hiroaki Kudo and Noboru Ohnishi. Effect of the sight line shift when a head-mounted display is used. In *Proceedings of the 22nd Annual International Conference of the IEEE Engineering in Medicine and Biology Society (Cat. No. 00CH37143)*, Vol. 1, pp. 548–550. IEEE, 2000.

[33] Vive. Vive Pro Eye, 2021.

[34] Pupil Labs GmbH. Pupil Labs VR/AR, 2021.

[35] Pico Immersive Pte. Ltd. Pico neo 3, 2022.

[36] Robert Konrad, Anastasios Angelopoulos, and Gordon Wetzstein. Gaze-contingent ocular parallax rendering for virtual reality. *ACM Transactions on Graphics (TOG)*, Vol. 39, No. 2, pp. 1–12, 2020.

[37] Yuri Mikawa, Tomohiro Sueishi, Yoshihiro Watanabe, and Masatoshi Ishikawa. Projection mapping system to a widely dynamic sphere with circumferential markers. In *2020 IEEE International Conference on Multimedia and Expo (ICME)*, pp. 1–6. IEEE, 2020.

[38] Yuri Mikawa, Tomohiro Sueishi, Yoshihiro Watanabe, and Masatoshi Ishikawa. Dynamic projection mapping for robust sphere posture tracking using uniform/biased circumferential markers. *IEEE Transactions on Visualization & Computer Graphics*, Vol. 28, No. 12, pp. 4016–4031, 2022.

[39] Yuri Mikawa, Christian Eichhorn, and Gudrun Klinker. Multi-color led marker for dynamic target tracking in wide area. In *2023 IEEE Conference on Virtual Reality and 3D User Interfaces (VR)*. IEEE, 2023.

[40] Elias Daniel Guestrin and Moshe Eizenman. General theory of remote gaze estimation using the pupil center and corneal reflections. *IEEE Transactions on biomedical engineering*, Vol. 53, No. 6, pp. 1124–1133, 2006.

[41] Yuri Mikawa, Tomohiro Sueishi, Tomohiko Hayakawa, and Masatoshi Ishikawa. Laser-based drawing method for posture-free objects by photochromic active marking with high-speed coaxial gaze control. In *Laser 3D Manufacturing VII*, Vol. 11271, pp. 100–107. SPIE, 2020.

[42] Clara Mestre, Josselin Gautier, and Jaume Pujol. Robust eye tracking based on multiple corneal reflections for clinical applications. *Journal of biomedical optics*, Vol. 23, No. 3, p. 035001, 2018.

[43] Y. Le Grand. *Light, Color and Vision*. Wiley, New York, NY, USA, 1957.

[44] Yuri Mikawa, Masahiro Fujiwara, and Hiroyuki Shinoda. Distant aerial image presentation using two separate rays of binocular parallax images. In *The 27th International Display Workshops (IDW '20)*, pp. 3D2–3, 2020.

[45] Yuri Mikawa, Masahiro Fujiwara, Takefumi Hiraki, Yasutoshi Makino, and Hiroyuki Shinoda. Far-field aerial image presentation of one point by a laser source using beam scanning by two-axis galvanometer mirror. In *2021 60th Annual Conference of the Society of Instrument and Control Engineers of Japan (SICE)*, pp. 137–143. IEEE, 2021.

[46] George W Fitzmaurice. Situated information spaces and spatially aware palmtop computers. *Communications of the ACM*, Vol. 36, No. 7, pp. 38–50, 1993.

[47] David T Chen, Chris Tector, Andrew Brandt, Hong Chen, Ryutarou Ohbuchi, Mike Bajura, Henry Fuchs, et al. Case study: observing a volume rendered fetus within a pregnant patient. In *Proceedings of the conference on IEEE Visualization '94*, pp. 364–368, 1994.

[48] Jun Rekimoto and Katashi Nagao. The world through the computer: computer augmented interaction with real world objects. In *Proceedings of the 8th annual ACM symposium on User interface software and technology*, pp. 29–36. ACM, 1995.

[49] Takafumi Taketomi, Hideaki Uchiyama, and Sei Ikeda. Visual slam algorithms: A survey from 2010 to 2016. *IPSJ Transactions on Computer Vision and Applications*, Vol. 9, No. 1, pp. 1–11, 2017.

[50] Zornitza Yovcheva, Dimitrios Buhalis, and Christos Gatzidis. Smartphone augmented reality applications for tourism. *E-review of tourism research (ertr)*, Vol. 10, No. 2, pp. 63–66, 2012.

[51] Bernard Mendiburu. *3d TV and 3d cinema: tools and processes for creative stereoscopy*. Routledge, 2012.

[52] teamLab. teamLab, 2022.

[53] Hirotsugu Yamamoto and Shiro Suyama. Aerial 3d led display by use of retroreflective sheeting. In *Stereoscopic Displays and Applications XXIV*, Vol. 8648, pp. 210–217. SPIE, 2013.

[54] Masahiko Yasui, Yoshihiro Watanabe, and Masatoshi Ishikawa. Occlusion-robust 3d sensing using aerial imaging. In *2016 IEEE International Conference on Computational Photography (ICCP)*, pp. 1–10. IEEE, 2016.

[55] Gaku Narita, Yoshihiro Watanabe, and Masatoshi Ishikawa. Dynamic projection mapping onto deforming non-rigid surface using deformable dot cluster marker. *IEEE transactions on visualization and computer graphics*, Vol. 23, No. 3, pp. 1235–1248, 2016.

[56] Dieter Schmalstieg and Tobias Hollerer. *Augmented reality: principles and practice*. Addison-Wesley Professional, 2016.

[57] Asahi Suzuki, Yoshitsugu Manabe, and Noriko Yata. Design of an ar marker for cylindrical surface. In *2013 IEEE International Symposium on Mixed and Augmented Reality (ISMAR)*, pp. 293–294. IEEE, 2013.

[58] Yoshihiro Watanabe, Toshiyuki Kato, et al. Extended dot cluster marker for high-speed 3d tracking in dynamic projection mapping. In *2017 IEEE International Symposium on Mixed and Augmented Reality (ISMAR)*, pp. 52–61. IEEE, 2017.

[59] Satoshi Tabata, Shohei Noguchi, Yoshihiro Watanabe, and Masatoshi Ishikawa. High-speed 3D sensing with three-view geometry using a segmented pattern. In *2015 IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS)*, pp. 3900–3907. IEEE, 2015.

[60] Shingo Kagami and Koichi Hashimoto. Animated stickies: Fast video projection mapping onto a markerless plane through a direct closed-loop alignment. *IEEE transactions on visualization and computer graphics*, Vol. 25, No. 11, pp. 3094–3104, 2019.

[61] James M Rehg and Takeo Kanade. Visual tracking of high dof articulated structures: an application to human hand tracking. In *European conference on computer vision*, pp. 35–46. Springer, 1994.

[62] Z. Cao, G. Hidalgo Martinez, T. Simon, S. Wei, and Y. A. Sheikh. Openpose: Realtime multi-person 2d pose estimation using part affinity fields. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 2019.

[63] Richard Hartley and Andrew Zisserman. *Multiple View Geometry in Computer Vision*. Cambridge university press, 2003.

[64] Idaku Ishii and Masatoshi Ishikawa. Self-windowing for high speed vision. In *Proceedings 1999 IEEE International Conference on Robotics and Automation*, Vol. 3, pp. 1916–1921. IEEE, 1999.

[65] Ankit Mohan, Grace Woo, Shinsaku Hiura, Quinn Smithwick, and Ramesh Raskar. Bokode: imperceptible visual tags for camera based interaction from a distance. In *ACM SIGGRAPH 2009 papers*, pp. 1–8. 2009.

[66] Homayoun Bagherinia and Roberto Manduchi. A theory of color barcodes. In *2011 IEEE International Conference on Computer Vision Workshops (ICCV Workshops)*, pp. 806–813. IEEE, 2011.

[67] Tan Jin Soon. Qr code. *Synthesis Journal*, Vol. 2008, pp. 59–78, 2008.

[68] 奥山瑞希, 的場やすし, 椎尾一郎. 円柱体の id・位置検出のための円筒マーカ. インタラクション, Vol. 2018, pp. 307–311, 2018.

[69] Albert Ng, Julian Lepinski, Daniel Wigdor, Steven Sanders, and Paul Dietz. Designing for low-latency direct-touch input. In *Proceedings of the 25th annual ACM symposium on User interface software and technology*, pp. 453–464, 2012.

[70] Hideaki Uchiyama and Hideo Saito. Random dot markers. In *2011 IEEE Virtual Reality Conference*, pp. 35–38. IEEE, 2011.

[71] Sora Hisaichi, Kiwamu Sumino, Kunihiro Ueda, Hidenori Kasebe, Tohru Yamashita, Takeshi Yuasa, Uwe Lippmann, Petra Aswendt, Roland Höfling, and Yoshihiro Watanabe. Depth-aware dynamic projection mapping using high-speed rgb and ir projectors. In *SIGGRAPH Asia 2021 Emerging Technologies*, pp. 1–2. 2021.

[72] Leonid Keselman, John Iselin Woodfill, Anders Grunnet-Jepsen, and Achintya Bhowmik.

Intel realsense stereoscopic depth cameras. In *Proceedings of the IEEE conference on computer vision and pattern recognition workshops*, pp. 1–10, 2017.

[73] Shingo Kagami and Koichi Hashimoto. Interactive stickies: Low-latency projection mapping for dynamic interaction with projected images on a movable surface. In *ACM SIGGRAPH 2020 Emerging Technologies*, pp. 1–2. 2020.

[74] Takefumi Hiraki, Shogo Fukushima, Hiroshi Watase, and Takeshi Naemura. Pixel-level visible light communication projector with interactive update of images and data. In *Proceedings of the International Display Workshops*, pp. 1192–1195, 2018.

[75] Robert J Leach, Stephanie E Forrester, AC Mears, and Jonathan R Roberts. How valid and accurate are measurements of golf impact parameters obtained using commercially available radar and stereoscopic optical launch monitors? *Measurement*, Vol. 112, pp. 125–136, 2017.

[76] Yuki Morikubo, Eugene San Lorenzo, Daiki Miyazaki, and Naoki Hashimoto. Tangible Projection Mapping: Dynamic Appearance Augmenting of Objects in Hands. In *SIGGRAPH Asia 2018 Emerging Technologies*, p. 14. ACM, 2018.

[77] Ankur Datta, Jun Sik Kim, and Takeo Kanade. Accurate camera calibration using iterative refinement of control points. In *2009 IEEE 12th International Conference on Computer Vision Workshops, ICCV Workshops*, pp. 1201–1208. IEEE, 2009.

[78] Yihong Wu, Haijiang Zhu, Zhanyi Hu, and Fuchao Wu. Camera calibration from the quasi-affine invariance of two parallel circles. In *European Conference on Computer Vision*, pp. 190–202. Springer, 2004.

[79] Carlo Colombo, Dario Comanducci, and Alberto Del Bimbo. Camera calibration with two arbitrary coaxial circles. In *European Conference on Computer Vision*, pp. 265–276. Springer, 2006.

[80] Janne Heikkila. Geometric camera calibration using circular control points. *IEEE Transactions on pattern analysis and machine intelligence*, Vol. 22, No. 10, pp. 1066–1077, 2000.

[81] Perng Fei Luo and Joyce Wu. Easy calibration technique for stereo vision using a circle grid. *Optical Engineering*, Vol. 47, No. 3, p. 033607, 2008.

[82] Tomohiro Sueishi and Masatoshi Ishikawa. Circle grid fractal pattern for calibration at different camera zoom levels. In *SIGGRAPH Asia 2015 Posters*, pp. 1–1. 2015.

[83] Vincent Fremont and Ryad Chellali. Direct camera calibration using two concentric circles from a single view. In *International Conference on Artificial Reality and Telexistence*, pp. 93–98. Citeseer, 2002.

[84] Jun Sik Kim, Ho Won Kim, and In So Kweon. A camera calibration method using concentric circles for vision applications. *ACCV2002, Melbourne, Australia*, 2002.

[85] Francisco Abad, Emilio Camahort, and Roberto Vivó. Camera calibration using two concentric circles. In *International Conference Image Analysis and Recognition*, pp. 688–696. Springer, 2004.

[86] Guang Jiang and Long Quan. Detection of concentric circles for camera calibration. In *Tenth IEEE International Conference on Computer Vision (ICCV'05) Volume 1*, Vol. 1, pp. 333–340. IEEE, 2005.

[87] Qian Chen, Haiyuan Wu, and Toshikazu Wada. Camera calibration with two arbitrary coplanar circles. In *European Conference on Computer Vision*, pp. 521–532. Springer, 2004.

[88] Yihong Wu, Xinju Li, Fuchao Wu, and Zhanyi Hu. Coplanar circles, quasi-affine invariance and calibration. *Image and Vision Computing*, Vol. 24, No. 4, pp. 319–326, 2006.

[89] Zijian Zhao and Ying Weng. Recovering euclidean structure from principal-axes paralleled conics: applications to camera calibration. *JOSA A*, Vol. 31, No. 6, pp. 1186–1193, 2014.

[90] Jun Sik Kim, Pierre Gurdjos, and In So Kweon. Geometric and algebraic constraints of projected concentric circles and their applications to camera calibration. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, Vol. 27, No. 4, pp. 637–642, 2005.

[91] B W Zhang, You Fu Li, and S Y Chen. Concentric-circle-based camera calibration. *IET image processing*, Vol. 6, No. 7, pp. 870–876, 2012.

[92] Andrew Fitzgibbon, Maurizio Pilu, and Robert B Fisher. Direct least square fitting of ellipses. *IEEE Transactions on pattern analysis and machine intelligence*, Vol. 21, No. 5, pp. 476–480, 1999.

[93] Fred L Bookstein. Fitting conic sections to scattered data. *Computer graphics and image processing*, Vol. 9, No. 1, pp. 56–71, 1979.

[94] Walter Gander, Gene H Golub, and Rolf Strebel. Least-squares fitting of circles and ellipses. *BIT Numerical Mathematics*, Vol. 34, No. 4, pp. 558–578, 1994.

[95] Gabriel Taubin. Estimation of planar curves, surfaces, and nonplanar space curves defined by implicit equations with applications to edge and range image segmentation. *IEEE Transactions on Pattern Analysis & Machine Intelligence*, No. 11, pp. 1115–1138, 1991.

[96] Lech Świrski, Andreas Bulling, and Neil Dodgson. Robust real-time pupil tracking in highly off-axis images. In *Proceedings of the Symposium on Eye Tracking Research and Applications*, pp. 173–176, 2012.

[97] Zhi Liu and Yanru Sun. A lightweight coin detection and classification algorithm for low quality images. In *2017 2nd International Conference on Automation, Mechanical Control and Computational Engineering (AMCCE 2017)*. Atlantis Press, 2017.

[98] Jonas Tebbe, Lukas Klamt, Yapeng Gao, and Andreas Zell. Spin detection in robotic table tennis. In *2020 IEEE International Conference on Robotics and Automation (ICRA)*, pp. 9694–9700. IEEE, 2020.

[99] Akira Nakashima, Yuki Ogawa, Chunfang Liu, and Yoshikazu Hayakawa. Robotic table tennis based on physical models of aerodynamics and rebounds. In *2011 IEEE International Conference on Robotics and Biomimetics*, pp. 2348–2354. IEEE, 2011.

[100] 阿部孝則, 宮地力, 吉田和人, 奥寛雅. 卓球サーブ練習のための回転情報リアルタイム提示システム. 映像情報メディア学会冬季大会講演予稿集 映像情報メディア学会 2015 年冬季大会講演予稿集, pp. 13B–2. 一般社団法人 映像情報メディア学会, 2015.

[101] Hermann von Helmholtz and James PC Southall. Helmholtz's treatise on physiological optics, vol. 1, trans. 1924.

[102] Arthur George Bennett and Ronald B Rabbetts. *Clinical visual optics*. Butterworth-Heinemann, 1989.

[103] Atsushi Nakazawa, Christian Nitschke, and Toyoaki Nishida. Registration of eye reflection and scene images using an aspherical eye model. *JOSA A*, Vol. 33, No. 11, pp. 2264–2276, 2016.

[104] Michel Guillon, Donald PM Lydon, and Christine Wilson. Corneal topography: a clinical model. *Ophthalmic and Physiological Optics*, Vol. 6, No. 1, pp. 47–56, 1986.

[105] Jinglu Ying, Mingguang Shi, and Bo Wang. Anterior corneal asphericity calculated by the tangential radius of curvature. *Journal of biomedical optics*, Vol. 17, No. 7, p. 075005, 2012.

[106] Brian John Mun Lui. *A point-of-gaze estimation system for studies of visual attention*. 2003.

[107] Thiago Santini, Wolfgang Fuhl, and Enkelejda Kasneci. Calibme: Fast and unsupervised eye tracker calibration for gaze-based pervasive human-computer interaction. In *Proceedings of the 2017 chi conference on human factors in computing systems*, pp. 2594–2605, 2017.

[108] Carlos Hitoshi Morimoto, Dave Koons, Arnon Amir, and Myron Flickner. Pupil detection and tracking using multiple light sources. *Image and vision computing*, Vol. 18, No. 4, pp. 331–335, 2000.

[109] Sheng-Wen Shih and Jin Liu. A novel approach to 3-d gaze tracking using stereo cameras. *IEEE Transactions on Systems, Man, and Cybernetics, Part B (Cybernetics)*, Vol. 34, No. 1,

pp. 234–245, 2004.

[110] Lech Swirski and Neil Dodgson. A fully-automatic, temporal approach to single camera, glint-free 3d eye model fitting. *Proc. PETMEI*, pp. 1–11, 2013.

[111] Kai Dierkes, Moritz Kassner, and Andreas Bulling. A novel approach to single camera, glint-free 3d eye model fitting including corneal refraction. In *Proceedings of the 2018 ACM Symposium on Eye Tracking Research & Applications*, pp. 1–9, 2018.

[112] Kai Dierkes, Moritz Kassner, and Andreas Bulling. A fast approach to refraction-aware eye-model fitting and gaze prediction. In *Proceedings of the 11th ACM Symposium on Eye Tracking Research & Applications*, pp. 1–9, 2019.

[113] Anastasios N. Angelopoulos, Julien N.P. Martel, Amit P. Kohli, Jörg Conradt, and Gordon Wetzstein. Event-based near-eye gaze tracking beyond 10,000 hz. *IEEE Transactions on Visualization and Computer Graphics*, Vol. 27, No. 5, pp. 2577–2586, 2021.

[114] Chih-Chuan Lai, Sheng-Wen Shih, and Yi-Ping Hung. Hybrid method for 3-d gaze tracking using glint and contour features. *IEEE Transactions on Circuits and Systems for Video Technology*, Vol. 25, No. 1, pp. 24–37, 2014.

[115] Atsushi Nakazawa and Christian Nitschke. Point of gaze estimation through corneal surface reflection in an active illumination environment. In *European Conference on Computer Vision*, pp. 159–172. Springer, 2012.

[116] Viktor Wase, Erik Ljungzell, Mark Ryan, Chiara Giordano, Rickard Lundahl, and Pravin Kumar Rana. Eye tracking system, March 17 2022. US Patent App. 17/459,442.

[117] Pravin Rana, Daniel Tornéus, and Jonas Andersson. Method and system for controlling illuminators, June 18 2020. US Patent App. 16/506,398.

[118] Timo Stoffregen, Hossein Daraei, Clare Robinson, and Alexander Fix. Event-based kilohertz eye tracking using coded differential lighting. In *Proceedings of the IEEE/CVF Winter Conference on Applications of Computer Vision*, pp. 2515–2523, 2022.

[119] Feng Li, Susan Kolakowski, and Jeff Pelz. Using structured illumination to enhance video-based eye tracking. In *2007 IEEE International Conference on Image Processing*, Vol. 1, pp. I–373. IEEE, 2007.

[120] Craig A Hennessey and Peter D Lawrence. Improving the accuracy and reliability of remote system-calibration-free eye-gaze tracking. *IEEE transactions on biomedical engineering*, Vol. 56, No. 7, pp. 1891–1900, 2009.

[121] Dan Witzner Hansen, Lars Roholm, and Iván García Ferreiros. Robust glint detection through homography normalization. In *Proceedings of the Symposium on Eye Tracking Research and Applications*, pp. 91–94, 2014.

[122] Antonio Peréz, M Luisa Córdoba, A Garcia, Rafael Méndez, ML Munoz, José Luis Pedraza, and F Sanchez. A precise eye-gaze detection and tracking system. 2003.

[123] Zihan Ding, Jiayi Luo, and Hongping Deng. Accelerated exhaustive eye glints localization method for infrared video oculography. In *Proceedings of the 33rd Annual ACM Symposium on Applied Computing*, pp. 620–627, 2018.

[124] Zhengyang Wu, Srivignesh Rajendran, Tarrence Van As, Vijay Badrinarayanan, and Andrew Rabinovich. Eyenet: A multi-task deep network for off-axis eye gaze estimation. In *2019 IEEE/CVF International Conference on Computer Vision Workshop (ICCVW)*, pp. 3683–3687. IEEE, 2019.

[125] Lijinliang Niu, Zhaopeng Gu, Juntao Ye, and Qiulei Dong. Real-time localization and matching of corneal reflections for eye gaze estimation via a lightweight network. In *The Ninth International Symposium of Chinese CHI*, pp. 33–40, 2021.

[126] Soumil Chugh, Braiden Brousseau, Jonathan Rose, and Moshe Eizenman. Detection and correspondence matching of corneal reflections for eye tracking using deep learning. In *2020 25th International Conference on Pattern Recognition (ICPR)*, pp. 2210–2217. IEEE, 2021.

[127] Joohwan Kim, Michael Stengel, Alexander Majercik, Shalini De Mello, David Dunn,

Samuli Laine, Morgan McGuire, and David Luebke. Nvgaze: An anatomically-informed dataset for low-latency, near-eye gaze estimation. In *Proceedings of the 2019 CHI conference on human factors in computing systems*, pp. 1–12, 2019.

[128] Jiannan Chi, Jiahui Liu, Feng Wang, Yingkai Chi, and Zeng-Guang Hou. 3-d gaze-estimation method using a multi-camera-multi-light-source system. *IEEE Transactions on Instrumentation and Measurement*, Vol. 69, No. 12, pp. 9695–9708, 2020.

[129] Elias Daniel Guestrin. *Remote, non-contact gaze estimation with minimal subject cooperation*. PhD thesis, University of Toronto, 2010.

[130] Francisco J Romero-Ramirez, Rafael Muñoz-Salinas, and Rafael Medina-Carnicer. Speeded up detection of squared fiducial markers. *Image and vision Computing*, Vol. 76, pp. 38–47, 2018.

[131] Yoshihiro Watanabe, Gaku Narita, Sho Tatsuno, Takeshi Yuasa, Kiwamu Sumino, and Masatoshi Ishikawa. High-speed 8-bit image projector at 1,000 fps with 3 ms delay. In *22nd International Display Workshops, IDW 2015*, pp. 1421–1422. International Display Workshops, 2015.

[132] Yoshihiro Watanabe and Masatoshi Ishikawa. High-speed and high-brightness color single-chip dlp projector using high-power led-based light sources. In *26th International Display Workshops, IDW 2019*, pp. 1350–1352. International Display Workshops, 2019.

[133] U Lippmann, P Aswendt, R Höfling, K Sumino, K Ueda, Y Ono, H Kasebe, T Yamashita, T Yuasa, and Y Watanabe. High-speed rgb+ ir projector based on coaxial optical design with two digital mirror devices. In *International Display Workshops*, pp. 636–639, 2021.

[134] Kohei Okumura, Hiromasa Oku, and Masatoshi Ishikawa. Lumipen: Projection-based mixed reality for dynamic objects. In *2012 IEEE International Conference on Multimedia and Expo*, pp. 699–704. IEEE, 2012.

[135] Ray Asahina, Takashi Nomoto, Takatoshi Yoshida, and Yoshihiro Watanabe. Realistic 3d swept-volume display with hidden-surface removal using physical materials. In *2021 IEEE Virtual Reality and 3D User Interfaces (VR)*, pp. 113–121. IEEE, 2021.

[136] Hao-Lun Peng and Yoshihiro Watanabe. High-speed dynamic projection mapping onto human arm with realistic skin deformation. *Applied Sciences*, Vol. 11, No. 9, p. 3753, 2021.

[137] Amit H Bermano, Markus Billeter, Daisuke Iwai, and Anselm Grundhöfer. Makeup lamps: Live augmentation of human faces via projection. In *Computer Graphics Forum*, Vol. 36, pp. 311–323. Wiley Online Library, 2017.

[138] PRTIMES. 高速プロジェクションマッピング技術と色補正技術を活用した実際の顔で体験できるメイクシミュレーションシステムを開発　～メイクシミュレーター「COLOR MACHINE」に応用～ (Japanese page), 2022.

[139] Takashi Nomoto, Wanlong Li, Hao-Lun Peng, and Yoshihiro Watanabe. Dynamic multi-projection mapping based on parallel intensity control. *IEEE Transactions on Visualization and Computer Graphics*, Vol. 28, No. 5, pp. 2125–2134, 2022.

[140] Christian Siegl, Matteo Colaianni, Lucas Thies, Justus Thies, Michael Zollhöfer, Shahram Izadi, Marc Stamminger, and Frank Bauer. Real-time pixel luminance optimization for dynamic multi-projection mapping. *ACM Transactions on Graphics (TOG)*, Vol. 34, No. 6, pp. 1–11, 2015.

[141] Kosuke Hiratani, Daisuke Iwai, Parinya Punpongsanon, and Kosuke Sato. Shadowless projector: Suppressing shadows in projection mapping with micro mirror array plate. In *2019 IEEE Conference on Virtual Reality and 3D User Interfaces (VR)*, pp. 1309–1310. IEEE, 2019.

[142] Zhengqing Li, Shio Miyafuji, Erwin Wu, Hideaki Kuzuoka, Naomi Yamashita, and Hideki Koike. OmniGlobe: An Interactive I/O System For Symmetric 360-Degree Video Communication. In *Proceedings of the 2019 on Designing Interactive Systems Conference*, pp. 1427–1438. ACM, 2019.

[143] Shio Miyafuji, Toshiki Sato, Zhengqing Li, and Hideki Koike. Qoom: An interactive omnidirectional ball display. In *Proceedings of the 30th annual acm symposium on user interface software and technology*, pp. 599–609, 2017.

[144] Wataru Yamada, Kazuhiro Yamada, Hiroyuki Manabe, and Daizo Ikeda. iSphere: Self-Luminous Spherical Drone Display. In *Proceedings of the 30th Annual ACM Symposium on User Interface Software and Technology*, pp. 635–643. ACM, 2017.

[145] Dylan Brodie Fafard, Qian Zhou, Chris Chamberlain, Georg Hagemann, Sidney Fels, and Ian Stavness. Design and implementation of a multi-person fish-tank virtual reality display. In *Proceedings of the 24th ACM Symposium on Virtual Reality Software and Technology*, p. 5. ACM, 2018.

[146] Shio Miyafuji, Masato Sugasaki, and Hideki Koike. Ballumiere: Real-time tracking and spherical projection for high-speed moving balls. In *Proceedings of the 2016 ACM International Conference on Interactive Surfaces and Spaces*, pp. 33–37. ACM, 2016.

[147] H Nishimura, T Abe, H Yamamoto, Y Hayasaki, and N Nishida. Development of 140-inch autostereoscopic display by use of full-color led panel. In *Light-Emitting Diodes: Research, Manufacturing, and Applications XI*, Vol. 6486, p. 64861B. International Society for Optics and Photonics, 2007.

[148] Tom Peterka, Robert L Kooima, Daniel J Sandin, Andrew Johnson, Jason Leigh, and Thomas A DeFanti. Advances in the dynallax solid-state dynamic parallax barrier autostereoscopic visualization display system. *IEEE transactions on visualization and computer graphics*, Vol. 14, No. 3, pp. 487–499, 2008.

[149] Daisuke Miyazaki, Noboru Hirano, Yuuki Maeda, Keisuke Ohno, and Satoshi Maekawa. Volumetric display using a roof mirror grid array. In *Stereoscopic Displays and Applications XXI*, Vol. 7524, pp. 209–217. SPIE, 2010.

[150] Jisoo Hong, Sung-Wook Min, and Byoungho Lee. Integral floating display systems for augmented reality. *Applied optics*, Vol. 51, No. 18, pp. 4201–4209, 2012.

[151] Daisuke Miyazaki, Noboru Hirano, Yuki Maeda, Siori Yamamoto, Takaaki Mukai, and Satoshi Maekawa. Floating volumetric image formation using a dihedral corner reflector array device. *Applied optics*, Vol. 52, No. 1, pp. A281–A289, 2013.

[152] Sung Kyu Kim, Ki Hyuk Yoon, Seon Kyu Yoon, and Heongkyu Ju. Parallax barrier engineering for image quality improvement in an autostereoscopic 3d display. *Optics express*, Vol. 23, No. 10, pp. 13230–13244, 2015.

[153] Tom Peterka, Robert L Kooima, Javier I Girado, Jinghua Ge, Daniel J Sandin, Andrew Johnson, Jason Leigh, Jurgen Schulze, and Thomas A DeFanti. Dynallax: Solid state dynamic parallax barrier autostereoscopic vr display. In *2007 IEEE Virtual Reality Conference*, pp. 155–162. IEEE, 2007.

[154] Hideki Kakeya, Ken Okada, and Hayato Takahashi. Time-division quadruplexing parallax barrier with subpixel-based slit control. *ITE Transactions on Media Technology and Applications*, Vol. 6, No. 3, pp. 237–246, 2018.

[155] Rebekka S Renner, Boris M Velichkovsky, and Jens R Helmert. The perception of egocentric distances in virtual environments-a review. *ACM Computing Surveys (CSUR)*, Vol. 46, No. 2, pp. 1–40, 2013.

[156] Hiroshi Ono and Hiroyasu Ujike. Motion parallax driven by head movements: Conditions for visual stability, perceived depth, and perceived concomitant motion. *Perception*, Vol. 34, No. 4, pp. 477–490, 2005.

[157] Hiroyasu Ujike and Hiroshi Ono. Depth thresholds of motion parallax as a function of head movement velocity. *Vision Research*, Vol. 41, No. 22, pp. 2835–2843, 2001.

[158] Satoko Ohtsuka, Hiroyasu Ujike, and Shinya Saida. Relative distance cues contribute to scaling depth from motion parallax. *Perception & psychophysics*, Vol. 64, No. 3, pp. 405–414, 2002.

[159] G Lee Zimmerman, Gordon E Legge, and Patrick Cavanagh. Pictorial depth cues: a new

slant. *JOSA A*, Vol. 12, No. 1, pp. 17–26, 1995.

[160] Patricia M Cisarik and Ronald S Harwerth. Stereoscopic depth magnitude estimation: effects of stimulus spatial frequency and eccentricity. *Behavioural brain research*, Vol. 160, No. 1, pp. 88–98, 2005.

[161] Andrew Glennerster, Brian J Rogers, and Mark F Bradshaw. Stereoscopic depth constancy depends on the subject's task. *Vision research*, Vol. 36, No. 21, pp. 3441–3456, 1996.

[162] Phillip E Napieralski, Bliss M Altenhoff, Jeffrey W Bertrand, Lindsay O Long, Sabarish V Babu, Christopher C Pagano, Justin Kern, and Timothy A Davis. Near-field distance perception in real and virtual environments using both verbal and action responses. *ACM Transactions on Applied Perception (TAP)*, Vol. 8, No. 3, pp. 1–19, 2011.

[163] Jeffrey Andre and Sheena Rogers. Using verbal and blind-walking distance estimates to investigate the two visual systems hypothesis. *Perception & Psychophysics*, Vol. 68, No. 3, pp. 353–361, 2006.

[164] Adam R Richardson and David Waller. Interaction with an immersive virtual environment corrects users' distance estimates. *Human Factors*, Vol. 49, No. 3, pp. 507–517, 2007.

[165] Timofey Y Grechkin, Tien Dat Nguyen, Jodie M Plumert, James F Cremer, and Joseph K Kearney. How does presentation method and measurement protocol affect distance estimation in real and virtual environments? *ACM Transactions on Applied Perception (TAP)*, Vol. 7, No. 4, pp. 1–18, 2010.

[166] Brooke Krajancich, Petr Kellnhofer, and Gordon Wetzstein. Optimizing depth perception in virtual and augmented reality through gaze-contingent stereo rendering. *ACM Transactions on Graphics (TOG)*, Vol. 39, No. 6, pp. 1–10, 2020.

[167] Himari Tochioka, Haruka Ikeda, Tomohiko Hayakawa, and Masatoshi Ishikawa. Effects of latency in visual feedback on human performance of path-steering tasks. In *25th ACM Symposium on Virtual Reality Software and Technology*, pp. 1–2, 2019.

[168] Andriy Pavlovych and Wolfgang Stuerzlinger. The tradeoff between spatial jitter and latency in pointing tasks. In *Proceedings of the 1st ACM SIGCHI symposium on Engineering interactive computing systems*, pp. 187–196, 2009.

[169] Jason D Moss, Jon Austin, James Salley, Julie Coats, Krysten Williams, and Eric R Muth. The effects of display delay on simulator sickness. *Displays*, Vol. 32, No. 4, pp. 159–168, 2011.

[170] Rachel Albert, Anjul Patney, David Luebke, and Joohwan Kim. Latency requirements for foveated rendering in virtual reality. *ACM Transactions on Applied Perception (TAP)*, Vol. 14, No. 4, pp. 1–13, 2017.

[171] Luke Hsiao, Brooke Krajancich, Philip Levis, Gordon Wetzstein, and Keith Winstein. Towards retina-quality vr video streaming: 15ms could save you 80% of your bandwidth. *ACM SIGCOMM Computer Communication Review*, Vol. 52, No. 1, pp. 10–19, 2022.

[172] Phillip Guan, Olivier Mercier, Michael Shvartsman, and Douglas Lanman. Perceptual requirements for eye-tracked distortion correction in vr. In *ACM SIGGRAPH 2022 Conference Proceedings*, pp. 1–8, 2022.

[173] Tomohiro Sueishi, Chikara Miyaji, Masataka Narumiya, Yuji Yamakawa, and Masatoshi Ishikawa. High-speed projection method of swing plane for golf training. In *Proceedings of the Augmented Humans International Conference*, pp. 1–3, 2020.

[174] Atsuki Ikeda, Yuka Tanaka, Dong-Hyun Hwang, Homare Kon, and Hideki Koike. Golf training system using sonification and virtual shadow. In *ACM SIGGRAPH 2019 Emerging Technologies*, pp. 1–2. 2019.

[175] Yuta Itoh, Jason Orlosky, Kiyoshi Kiyokawa, and Gudrun Klinker. Laplacian Vision: Augmenting Motion Prediction via Optical See-Through Head-Mounted Displays. In *Proceedings of the 7th Augmented Human International Conference 2016*, p. 16. ACM, 2016.

[176] Irena Bojanova. It enhances football at world cup 2014. *IT Professional*, Vol. 16, No. 4, pp. 12–17, 2014.

[177] Christian Eichhorn, Adnane Jadid, David A Plecher, Sandro Weber, Gudrun Klinker, and Yuta Itoh. Catching the drone-a tangible augmented reality game in superhuman sports. In *2020 IEEE International Symposium on Mixed and Augmented Reality Adjunct (ISMAR-Adjunct)*, pp. 24–29. IEEE, 2020.

[178] Whitman Richards and Martin G Kaye. Local versus global stereopsis: two mechanisms? *Vision Research*, Vol. 14, No. 12, pp. 1345–1347, 1974.

[179] Hwan S Lee and Allan C Dobbins. Perceiving surfaces in depth beyond the fusion limit of their elements. *Perception*, Vol. 35, No. 1, pp. 31–39, 2006.

[180] Geoffrey P Bingham. Optical flow from eye movement with head immobilized: "ocular occlusion" beyond the nose. *Vision Research*, Vol. 33, No. 5-6, pp. 777–789, 1993.

[181] Mark Nawrot. Eye movements provide the extra-retinal signal required for the perception of depth from motion parallax. *Vision research*, Vol. 43, No. 14, pp. 1553–1562, 2003.

[182] Mark Nawrot. Depth from motion parallax scales with eye movement gain. *Journal of Vision*, Vol. 3, No. 11, pp. 17–17, 2003.

[183] Mark Nawrot, Benita Nordenstrom, and Amy Olson. Disruption of eye movements by ethanol intoxication affects perception of depth from motion parallax. *Psychological Science*, Vol. 15, No. 12, pp. 858–865, 2004.

[184] Keith Stroyan and Mark Nawrot. Visual depth from motion parallax and eye pursuit. *Journal of mathematical biology*, Vol. 64, No. 7, pp. 1157–1188, 2012.

[185] Tobii Pro. Tobii pro glasses 3, March 2021.

[186] EyeLink. Eyeiink 1000 plus, 2022.

[187] G. Bradski. The OpenCV Library. *Dr. Dobb's Journal of Software Tools*, 2000.

[188] Suzanne P McKee and Ken Nakayama. The detection of motion in the peripheral visual field. *Vision research*, Vol. 24, No. 1, pp. 25–32, 1984.

[189] Han Collewijn, Casper J Erkelens, and Robert M Steinman. Binocular co-ordination of human horizontal saccadic eye movements. *The Journal of physiology*, Vol. 404, No. 1, pp. 157–182, 1988.

[190] Han Collewijn, Casper J Erkelens, and RM Steinman. Binocular co-ordination of human vertical saccadic eye movements. *The Journal of physiology*, Vol. 404, No. 1, pp. 183–197, 1988.

[191] Allen O Eghrari, S Amer Riazuddin, and John D Gottsch. Overview of the cornea: structure, function, and development. *Progress in molecular biology and translational science*, Vol. 134, pp. 7–23, 2015.

[192] PHOTRON LIMITED. Infinicam, 2022.

[193] Benjamin M Bolker, Mollie E Brooks, Connie J Clark, Shane W Geange, John R Poulsen, M Henry H Stevens, and Jada-Simone S White. Generalized linear mixed models: a practical guide for ecology and evolution. *Trends in ecology & evolution*, Vol. 24, No. 3, pp. 127–135, 2009.

[194] Marina Zannoli, Rachel A Albert, Abdullah Bulbul, Rahul Narain, James F O'Brien, and Martin Banks. Correct blur and accommodation information is a reliable cue to depth ordering. *Journal of Vision*, Vol. 14, No. 10, pp. 138–138, 2014.

[195] Konstantinos G Derpanis. Overview of the ransac algorithm. *Image Rochester NY*, Vol. 4, No. 1, pp. 2–3, 2010.

[196] Gregory Kramida. Resolving the vergence-accommodation conflict in head-mounted displays. *IEEE transactions on visualization and computer graphics*, Vol. 22, No. 7, pp. 1912–1931, 2015.

[197] Max Born and Emil Wolf. *Principles of optics: electromagnetic theory of propagation, interference and diffraction of light*. Elsevier, 2013.

[198] Tomohiro Sueishi, Arata Jingu, Shoji Yachida, Michiaki Inoue, Yuka Ogino, and Masatoshi Ishikawa. Dynamic iris authentication by high-speed gaze and focus control. In *2021 IEEE/SICE International Symposium on System Integration (SII)*, pp. 813–814. IEEE,

2021.

[199] Kevin Godineau, Sylvain Lavernhe, and Christophe Tournier. Calibration of galvanometric scan heads for additive manufacturing with machine assembly defects consideration. *Additive Manufacturing*, Vol. 26, pp. 250–257, 2019.

[200] Javier Alda. Laser and gaussian beam propagation and transformation. *Encyclopedia of optical engineering*, pp. 999–1013, 2003.

[201] 三河祐梨, 鈴木大河, 藤原正浩, 牧野泰才, 篠田裕之. 分散配置の動的パララックスバリアによる遠方空中像ディスプレイの基礎検討. 第 26 回日本バーチャルリアリティ学会大会論文集, pp. 1D3–1. 日本バーチャルリアリティ学会, 2021.

[202] Hirotsugu Yamamoto, Syuji Muguruma, Takeshi Sato, Kasai Ono, Yoshio Hayasaki, Yoshifumi Nagai, Yoshinori Shimizu, and Nobuo Nishida. Optimum parameters and viewing areas of stereoscopic full-color led display using parallax barrier. *IEICE transactions on electronics*, Vol. 83, No. 10, pp. 1632–1639, 2000.

[203] Michael F Deering. The limits of human vision. In *2nd International Immersive Projection Technology Workshop*, Vol. 2, p. 1, 1998.

[204] Yuri Mikawa, Masahiro Fujiwara, Yasutoshi Makino, and Hiroyuki Shinoda. Fast and low-latency ocular parallax rendering improves binocular fusion in stereoscopic vision. In *2023 IEEE Conference on Virtual Reality and 3D User Interfaces (VR)*. IEEE, 2023.

[205] Yasushi Nagamune. Image capturing device and image capturing method, U.S. Patent, No. 8836795B2, 2014. U.S. Patent No. 8836795B2.

[206] Daiki Tone, Daisuke Iwai, Shinsaku Hiura, and Kosuke Sato. FibAR: Embedding optical fibers in 3d printed objects for active markers in dynamic projection mapping. *IEEE transactions on visualization and computer graphics*, Vol. 26, No. 5, pp. 2030–2040, 2020.

[207] Volker Scholz, Timo Stich, Marcus Magnor, Michael Keckeisen, and Markus Wacker. Garment motion capture using color-coded patterns. In *ACM SIGGRAPH 2005 Sketches*, p. 38. 2005.

A

Optimal infrared light placement for accurate identification of multiple corneal reflections

This section describes the appropriate placement of infrared lights for tracking of widely rotating eye referred in Chapter 5. The placement is designed so that the possibility of misidentification would be reduced, which has been investigated by simulation.

Normally, with misidentification, the properly designed cost function is assumed to become larger than that of the correct identification. However, there are cases in which the cost function for incorrect identification cannot be as large. There are several factors for such a mistake, for example, system calibration, different corneal shapes between users, and inappropriate placements of light sources. Among these factors, considering the optimal placement of the light sources is necessary such that the error can obtain a large value in incorrect identification and a small value without error to enable clear identification. This issue was examined through simulation.

A simulation of the optimal placement of the light sources was performed using MATLAB R2021α. The results significantly vary depending on the position of the camera, lights, and user; examining all cases is impossible. Therefore, we strongly recommend case-by-case simulation of an actual system. Our study conducted a simulation with nearly the same placement as that used in the evaluation of an actual eye.

A.1 Conditions

To simulate a real eye as much as possible, the aspherical model was used for the corneal surface in this simulation. According to previous studies [103, 105], this is generally formulated as follows:

$$x^2 + y^2 + (1 + Q)z^2 = R_c^2, \tag{A.1}$$

where $Q = -0.33$. Figure 5.4 shows the assumed positioning of the user's eyes, the camera, and lights, and each parameter was determined as shown in Table A.1.

The stereo cameras were spaced D_{cam} apart in the x-axis direction, and their center was the origin of the coordinate system. The same number of lights was placed in two rows within the range of L_h along the x-axis. These two rows were spaced L_v apart in the y-axis direction, and the center of gravity of all light sources was P_L. The lights were aligned in a straight horizontal line unless otherwise indicated. The number of lights per row was N_l, and the total number of lights was N_L, which has a relationship of $N_L = 2 * N_l$.

Both eyes were spaced D_{eye} apart, called the interocular distance, in the x-axis direction, and the mid-position of both eyes was P_e. The angle from the user to the horizontal row of lights was

Table A.1. Parameters of the system simulation including the user's eyes, the camera, and light sources.

L_h [mm]	100	L_v [mm]	111
D_{cam} [mm]	67	R_c [mm]	7.8
D_{eye} [mm]	63.5	R_i [mm]	6.0
$\boldsymbol{P}_e$ [mm]	[$D_{eye}/2$, 50, 500]	N_l	7
$\boldsymbol{P}_L$ [mm]	[$D_{eye}/2$, 50, 300]	R_e	12

set to θ, as shown in the top view. The light placement range along the x-axis is defined by an angle of θ, and a relationship $\theta = \arctan(L_v/2d)$ exists, where $d = |P_{ez} - P_{Lz}|$.

For all simulations, the eye was assumed to rotate at several angles ϕ around the y-axis, passing through the center of the eyeball from -45 to 45 $^\circ$ with 5° steps, resulting in a total of 19 patterns.

For simplicity, the spacing between lights was assumed to be the same everywhere. The vertical distance between two rows of lights, L_v, is preferably narrow to have high accuracy in eyesight estimation; however, a placement that is too narrow for the obstacle user's view is unacceptable.

A world stereo camera was placed near the user's eye at the appropriate position. No significant rules dictate the placement here; placing the camera where both the light source and eye camera can be observed was sufficient.

A.2 Evaluation method

Next, we describe the evaluation method. It is assumed that there are no noisy reflections and only true CRs, but some of these were missed owing to the occlusion. For each rotational angle of the eyeball, all promising identification candidates were enumerated assuming that some points were missed: six points at maximum when $N_L \geq 8$ and otherwise $N_L - 2$. The value of the cost function was then calculated for each point.

Among the errors of all candidates for a certain rotational angle, let the maximum error of correct identification be Δe_T and the minimum error of incorrect identification be Δe_F. Subsequently, the evaluation value γ is defined as

$$\gamma = \Delta e_F - \Delta e_T. \tag{A.2}$$

If of the value of γ is large, the identification is clear. The ratio of γ to Δe_T was unused because the evaluation using this ratio was unsuitable for actual use; the threshold to distinguish true identification was a static value. The statistics of γ were evaluated for all ϕ that satisfy $\gamma \geq 0$.

Some cases have $\gamma < 0$, which are special and rare cases in actual situations, indicating that an incorrect identification would likely be selected. Whether minus γ appears was enumerated for each rotational angle of the eyeball ϕ, and the ratio against the total cases of rotational angle was calculated.

A.3 Evaluation results

A.3.1 Horizontal range of the placement

Here, the optimal range of the light's placement in the x-direction, θ, was investigated. The number of lights, N_L, used 10 and 14. The range of the light placement was set to $\theta = 40, 60, 80, ..., 160^\circ$.

Figure A.1 (a) shows the results. If $\theta < 120^\circ$, the γ of both $N_L = 10$ and 14 increases according to θ; otherwise, only those with $N_L = 14$ continue to increase according to θ. The ratio

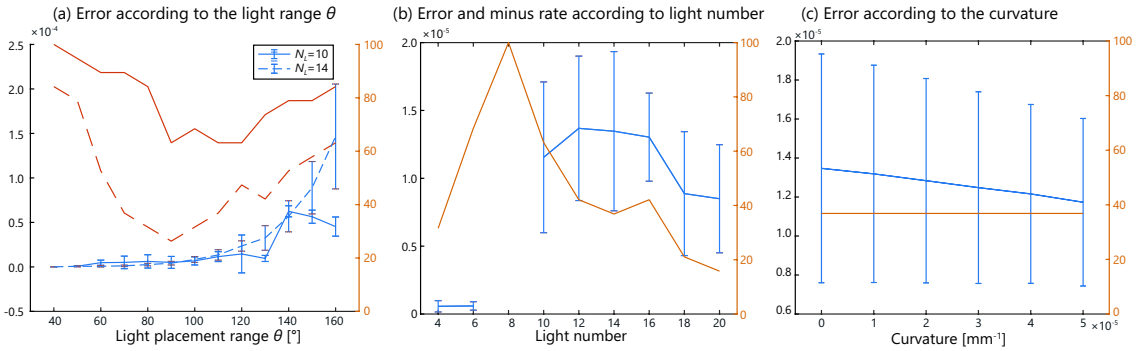

Fig. A.1. Results of the identification simulation for the evaluation value γ (left axis, blue lines) and the probability of $\gamma < 0$ [%] (right axis, orange lines) according to the following different conditions. *(a)* The range of the placement of lights θ of $N_L = 10$ or 14 indicated by solid and dashed lines, respectively. *(b)* The number of lights N_L when $\theta = 110[^\circ]$. *(c)* The curvature of the light placement when $N_L = 14$ and $\theta = 110[^\circ]$. The blue line indicates γ, and the orange line indicates the ratio of $\gamma < 0$.

of $\gamma < 0$ to $N_L = 10$ is much larger than to $N_L = 14$ for all θ. However, for $N_L = 14$, the minimum peak is observed at $\theta = 90^\circ$ at 30 %, which seems to be a somewhat large value.

In conclusion, as a useful value for the actual situation and a reasonably evaluated value, $\theta = 110^\circ$ was assumed to be proper, which was used in the following situations. However, the problem was that the ratio of $\gamma < 0$ was reasonably large, even at the minimum peak of $N_L = 14$.

A.3.2 Number of lights

With the light placement range of $\theta = 110^\circ$, the effect of the number of lights, $N_L = 4, 6, 8, ..., 20$ was investigated. Figure A.1 (b) shows the results. Note that the evaluation value γ for $N_L = 8$ does not appear because $\gamma < 0$ for all rotational angles.

The maximum γ was reached at approximately $N_L = 12$ and 14, and the rate of $\gamma < 0$ was likely to drop at a large N_L. Based on this, a moderate number of N_L, such as 12 or 14, was concluded to be sufficient for clearly identifying neighborhood lights while maintaining a relatively small probability of $\gamma < 0$.

A.3.3 Shape of the placement

Here, the effect of changing the shape of the lights is investigated. This was investigated because the typical placement involves straight lines and circular placements. Assuming realistic cases, the shape ranged from a straight line to an inward curve. Therefore, the curvature value was set to $0 - 5.0 \times 10^{-5}$ with 1.0×10^{-5} steps, resulting in six patterns. From the results of the previous sections, $\theta_{=}110^\circ$ and the number of lights $N_L = 14$ were used.

Figure A.1 (c) shows the results according to the curvature. No significant differences were observed among each condition; rather, γ decreased as more curves are added. Therefore, a straight line was considered the most preferable.

A.4 Conclusion

According to the simulation results, the horizontal range of the infrared light placement should have a large value, but as a reasonable value, this paper concluded that 110° is appropriate. The number of two-row lights should be 14 in total, and there is no need to have a curvature on each

row. Note that these values are not general and are applicable just for the specified system placement (the distance between cameras and a user, and that between cameras and infrared lights). Such simulation should be done for each system dimension you desire to use.