Wearable Imaging System for Capturing Omnidirectional Movies from a First-person Perspective

Kazuaki Kondo*
Kyoto University

Yasuhiro Mukaigawa†
Yasushi Yagi‡
Osaka University

Abstract

We propose a novel wearable imaging system that can capture omnidirectional movies from the viewpoint of the camera wearer. The imaging system solves the problems of resolution uniformity and gaze matching that conventional approaches do not address. We combine cameras with curved mirrors that control the projection of the imaging system to produce uniform resolution. Use of the mirrors also enables the viewpoint to be moved closer to the eyes of the camera wearer, thus reducing gaze mismatching. The optics, including the curved mirror, have been designed to form an objective projection. The capability of the designed optics is evaluated with respect to resolution, aberration, and gaze matching. We have developed a prototype based on the designed optics for practical use. Capability of the prototype and effectiveness of first-person perspective omnidirectional movies were demonstrated through quantitative evaluations and presentation experiments to ordinary people, respectively.

Keywords: Omnidirectional Imaging System, Wearable Camera, First-person Perspective

1 Introduction

Movies and photographs that we see in daily life are generally captured by cameras placed at the edge of the target scene. This type of objective image recording is referred to as being from the third-party perspective, and enables viewers to easily understand the stories and constructions of the scene. On the other hand, a subjective view captured with a portable video camera or wearable camera is referred to as being from the "first-person perspective". This view allows viewers to experience realistic feelings. Furthermore, omnidirectional images can give viewers impressive and immersive sensations as if they were present in the target scene. A head-mounted omnidirectional camera is suitable for this objective. It can capture filming targets from just a person's viewpoint influenced dynamical motions of the person. Additionally it has high mobility, although filming with a vehicle type system is limited by the ground condition. For example, a head mounted system can easily capture a view of a person who walks up stairs.

In order to provide an even greater realistic sensation, two important issues need to be considered with respect to the omnidirectional images shown to viewers. The first is gaze matching between the camera wearer and his companion, which helps correct expressions of scenes and situations; for example, human interactions such as face-to-face conversations and eye contact. The other issue is resolution uniformity of the omnidirectional movies. Viewers can watch from any direction with these telepresence and omnidirectional display systems [Cruz-Neria et al. 1993; Hashimoto and Iwata 1999; Iwata 2004]. Therefore, uniform resolution images need to be provided independently of the direction of the viewers' gaze.

However conventional approaches that acquire omnidirectional movies do not satisfy the above requirements. These approaches are typically classified as omnidirectional cameras as typified by [Yamazawa et al. 1993] with either a convex mirror or multiple camera systems [Tanahashi et al. 2000; Ladybug]. The total resolution of the former group is not sufficient for practical use, whereas the latter group does not take into account resolution uniformity. Furthermore, neither approach addresses the problem of gaze matching, as the imaging system is not expected to be worn or to capture images from a first-person perspective. A configuration currently available allows the imaging system to be mounted on a person's head such as in [Yamazawa et al. 2002]. However this setup causes gaze mismatching between the camera wearer and his companion because the viewpoint of the imaging system is set higher than his eyes.

In this paper, we propose a novel wearable omnidirectional imaging system that meets the two requirements of resolution and gaze matching. It has been designed to provide uniform resolution for an omnidirectional panoramic view using curved mirrors. Gaze matching is achieved by moving the virtual viewpoint of the system closer to the eyes of the camera wearer. Evaluation of the prototype was done in two ways. One is quantitative evaluations of the above requirements about abilities as an omnidirectional capturing system. The other is a validation of whether first-person perspective omnidirectional movies can actually provide an immersive sensation, and whether it is really demanded. We presented video content created from movies captured by the prototype system to a number of ordinary people to obtain general and objective remarks.

2 Requirements for a wearable omnidirectional Imaging System

The requirements of an objective wearable omnidirectional imaging system are summarized in Table 1, where they are divided into optical and hardware requirements.

2.1 Optical requirements

FOV and total resolution
FOV and absolute resolution are configured based on ergonomics. We select a gaze stable view that is one of the natural views for humans. According to this view, FOV along the azimuth angle, the elevation angle, and the depression angle are $\pm 30 \sim 45 degrees$, $20 \sim 30 degrees$, and $25 \sim 40 degrees$, respectively. We adopt the latter two values as the FOV requirements in the vertical direction. Naturally, the horizontal FOV is $360 degrees$ for omnidirectional observation. The objective total resolution is determined with the above FOV. We assume that the width of the VGA ($640 pixels$) is

*e-mail: kondo@ccm.media.kyoto-u.ac.jp
†e-mail: mukaigaw@am.sanken.osaka-u.ac.jp
‡e-mail: yagi@am.sanken.osaka-u.ac.jp

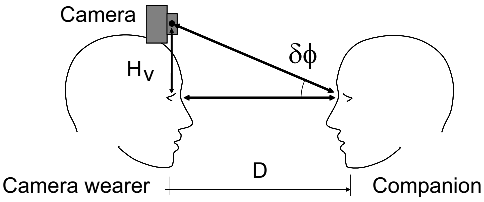


Figure 1: *Gaze mismatching caused by the gap between the camera's viewpoint and the eyes of the camera wearer. This gap should be minimized for gaze matching. The figure illustrates the case where a camera is mounted on the head.*

aligned to the horizontal FOV ($60 degrees$). This corresponds to a total resolution of about $3840 \times 590 pixels$.

Uniformity of resolution
In order to obtain visual information of omnidirectional images uniformly, the density of the image resolution should be as uniform as possible. Since images have two-dimensional information, it is required that the resolution distribution be uniform in both the horizontal and vertical directions. To discuss density of resolution, a type of scene needs to be given. In this paper, we assume uniform resolution on a panoramic scene consisting of the azimuth angle and tangent of the elevation angle. This type of scene can be considered as a cylindrical surface with its axis aligned to the vertical axis. The panoramic projections generate wide, seamless panoramic images.

Gaze matching (degree of first-person perspective)
Gaze matching relates strongly to the vertical gap between the viewpoint of the imaging system and the eyes of the camera wearer. Generally, when the viewpoint of the imaging system does not coincide with the eyes of the camera wearer, the captured images cause discomfort to the viewers. This is caused by the different angle of gaze between the camera wearer and the imaging system, as shown in Fig. 1. This angle depends on the distance D and the gap of the two viewpoints. Let H_v be the size of the gap, then the angle $\delta\phi$ is calculated as $\delta\phi = arctan\frac{H_v}{D}$. Psychological research specifies that a $5.0 degree$ gaze mismatch is the maximum tolerable by viewers before discomfort sets in [Anstis et al. 1969]. Furthermore, we assume a $50cm$ limit on the distance that permits gaze matching when considering close situations such as face-to-face conversations and eye contact.According to these values and the above formulation, an admissible gap between viewpoints is about $4.4cm$, which we adopt as the requirement for gaze matching.

2.2 Hardware requirements

Hardware requirements are mainly for constructing a wearable, stand-alone system. The width and depth of view of the wearable omnidirectional imaging system must not be larger than that of a human, which is about $W60cm \times D40cm$. Wired power supplies and data transmission prevent stand-alone construction. Therefore, an integrated camera with a recording unit and battery is necessary for flexible movie recording. The maximum usage time should be considered for battery-powered equipment; $30 minutes$ seems to be enough for filming the sequences of a scene. In consideration of practical use, we have determined the maximum total weight including the optics and recording units based on the weight of a full-face type helmet.

3 How to design the optics

3.1 Design concept

We created an optical unit by combining a portable video camera and a curved mirror. The objective omnidirectional imaging system consists of four such optical units. The portable video cameras help to form a stand-alone system. The curved mirrors are able to not only expand the FOV but also control projection at the same time. The system can also create a virtual viewpoint for the camera, which is closer to the eyes of the camera wearer. We have designed the shape of the curved mirror so that the total optical properties meet the required objectives. Because total resolution depends mainly on the number of image sensors, a single camera is generally not sufficient. However, the size and weight increases as the number of cameras increases. Mindful of this trade-off, we have chosen to use four optical units. The construction based on the design described above is illustrated in Fig. 2(A). We address a single optical unit in the discussion that follows.

3.2 Algorithm for design of mirror

(1) Make an objective normal vector field
First, an objective normal vector field of the mirror is decided based on the objective FOV and camera parameters. The camera parameters are its focal length f and a matrix $\mathbf{P} = [\mathbf{R} \quad \mathbf{t}]$ that describes the camera coordinate system in terms of a standard one. Figure 2(B) illustrates the relationships of the camera, mirror, and the scene. We start at the objective projection determined by ideal correspondences between pixels on the image plane and rays running in the scene. In the case of uniform resolution of the panoramic scene, this is formulated as

$$\mathbf{V_s}(u,v) = \begin{bmatrix} tan(\frac{u}{U}(\theta_{max} - \theta_{min}) + \theta_{min}) \\ \frac{v}{V}(h_{max} - h_{min}) + h_{min} \\ 1 \end{bmatrix} \quad (1)$$

in the real-life coordinate system. (u, v) is the position of a pixel on the image plane whose size is $U \times V$. (θ_{min}, h_{min}) and (θ_{max}, h_{max}) form a rectangle of the objective FOV. h is the tangent of the elevation angle ϕ. On the other hand, (u, v) provides a projection of the camera itself $\mathbf{V_c}(u, v)$, which is formulated under a perspective projection as $\mathbf{V_c}(u, v) = [u \quad v \quad -f]^t$ in the camera coordinate system. The curved mirror relates $\mathbf{V_s}$ and $\mathbf{V_c}$ to its reflection ; a normal vector field on the mirror surface bisects the angle consisting of $\mathbf{V_s}$ and $\mathbf{V_c}$. Then we obtain the objective normal vector field $\mathbf{N_d}(u, v)$ of the curved mirror as

$$\mathbf{N_d} = \mathcal{N}\left[\frac{1}{2}\left(\mathcal{N}[\mathbf{V_s}] + \mathcal{N}[\mathbf{RV_c}]\right)\right]. \quad (2)$$

with a vector normalizing operator $\mathcal{N}[\mathbf{x}] = \frac{\mathbf{x}}{||\mathbf{x}||}$.

(2) Form a mirror shape
A mirror shape is formed so that its normal field is equal to $\mathbf{N_d}$. It is known that no mirror exists that perfectly satisfies the above condition in the majority of cases. Some approaches do solve this type of mirror designing problem [Swaminathan et al. 2004; Hicks 2005]. In this paper, we use the linear algorithm proposed by Swaminathan et al. [Swaminathan et al. 2004], which equalizes the shape of a curved mirror $\mathbf{S}(u, v)$ as the cross products of four-degree spline curves. $\mathbf{S}(u, v)$ is formulated as

$$\mathbf{S}(u,v) = \mathbf{RV_c}(u,v)\sum_{i,j} C_{ij} f_i(u) g_j(v) + \mathbf{t} \quad (3)$$

Table 1: *Requirements for an objective imaging system, and achievements of the typical conventional omnidirectional imaging systems. Ladybug2 and HyperOmni-Vision with $a = b = 20mm$ are delegates of multiple camera systems and that of omnidirectional cameras with a convex mirror, respectively.*

terms	requirements	Ladybug2	HyperOmni-Vision	our system
Horizontal FOV	$360deg.$	$360deg.$	$360deg.$	$360deg.$
Vertical FOV	$-30 \sim +25deg.$	$-36.9 \sim +36.9deg.$	$-90 \sim +25.0deg.$	$-31.8 \sim +26.6deg.$
Resolution(horizontal)	$3840pixels$	$3580pixels$	$0 \sim 3220pixels$	$3960 \sim 5024pixels$
Resolution(vertical)	$590pixels$	$768pixels$	$520pixels$	$584 \sim 682pixels$
Resolution(distribution)	uniform	did not consider	did not consider	uniform
gap of viewpoints	$4.4cm$	over $18cm$	over $15cm$	$5.0cm$
Size of width	$60cm$	$10cm$	$10cm$	$40cm$
Size of depth	$40cm$	$11cm$	$10cm$	$40cm$
Additional equipments	no(stand-alone)	power and recorder	power and recorder	no(stand-alone)
Max. driving time	over $30minutes$	with power line	with power line	$85min.$
Weight	$2.0kg$	$1.2kg$(no recorder)	$0.5kg$(no recorder)	$2.4kg$

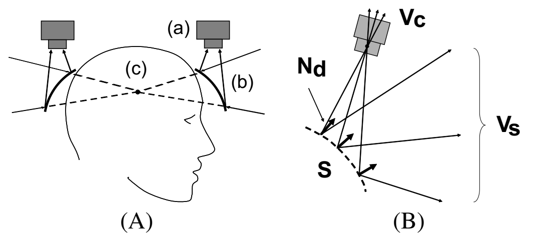


Figure 2: *(A) Basic construction. (a) Camera, (b) Curved mirror, and (c) Virtual viewpoint. Each camera around the head of the camera wearer is aligned to face downwards. The curved mirror supports projection and positioning of a virtual viewpoint. (B) Relationship between the camera, an objective projection, and normal vector field of the curved mirror.*

where C_{ij} and $f_i(u), g_j(v)$ are control points on the spline curves and four-degree spline bases, respectively. We obtain an optimal shape by solving the linear equations about C_{ij} that are stacks of

$$\frac{\partial \mathbf{S}}{\partial u} \cdot \mathbf{N_d} = \frac{\partial \mathbf{S}}{\partial v} \cdot \mathbf{N_d} = 0 \quad (4)$$

because the optimal shape should be perpendicular to the desired normal vector field $\mathbf{N_d}$. Although this algorithm certainly minimizes errors on the normal vector field, it tends to form a bumpy surface. So we apply a smoothing procedure to the shape formed by this algorithm. The smoothing approximates the shape with another smooth surface described by a small number of control points.

(3) Check projections
As described in the previous step, the designed mirror approximates the desired projection. Thus, we need to check the degree of the approximation. It is evaluated by

- sufficiency : how much does it cover the required FOV.
- redundancy : how much does it cover outside the FOV.
- uniformity : how uniformly does it distribute the image.

If the approximation is sufficient, the design advances to the next step. If not, we adjust the camera parameters to reduce projection errors and return to the first step.

(4) Check aberrations
Aberrations in the designed optics are checked to ensure well-focused images. These aberrations have not been considered in most of the conventional mirror design approaches. Generally, focusing and projection both relate to the shapes of the mirrors used in catadioptric imaging systems. Since a shape that produces a desired projection may not generate good focus, checking aberrations of the designed optics is necessary. Two important issues regarding aberrations in catadioptric imaging systems need to be considered [Swaminathan 2007]. The first is whether images are focused throughout the image plane. The curvature of mirrors tends to generate a distorted focus plane. This prevents objects from being recorded as sharp images on a normal flat image plane. The other issue is whether images are focused with respect to objects at any distance, thus corresponding to a wide depth of field. We construct aberration maps throughout the image plane at various depths to check whether the designed optics meet these two focusing requirements.

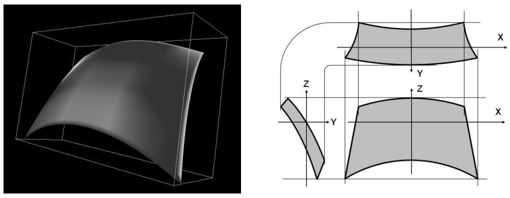
The amount of aberration can be estimated with a spot diagram, which is a spread image of a target object on the image plane. Rays departing from a target object are reflected by the curved mirror, and then refracted by the lens unit to concentrate on the image plane. Although they should ideally be focused onto a single point image, the curvatures of the mirror and lens unit prevent this. We can construct spot diagrams by tracing the rays that go through an aperture of the lens unit. RMS diameters of the spot diagrams correspond to the amount of aberration. In the case of cameras with digital photo sensors, the aberration needs to be less than a diagonal of the photo sensor to obtain focused images. If the aberrations appear to prevent image focusing, we adjust the camera parameters to reduce the aberration and return to the first step. The design process continues until the aberrations are acceptably small.

(5) Adjust camera parameters
Under the optical configuration assumed in this paper, using a wide FOV camera determined by the focal length f, reduces the projection error. Enlarging the size of the curved mirror determined by $\mathbf{t}$ reduces aberration. Decreasing the vertical angle of the camera determined by $\mathbf{R}$ reduces both of these. Note that the camera parameters cannot be freely changed because of the total size and self-occlusion of the camera.

4 Prototype system

4.1 Optical construction

We designed an optical system for a prototype.The objective projection given for the mirror design is a uniform projection that covers a $110degree \times 60degree$ FOV. The reason that the horizontal FOV is wider than $90degrees = 360degrees/4$ is to create an overlap-



(A) (B)

Figure 3: *Designed mirror. (A) Overview (B) Orthographic views*

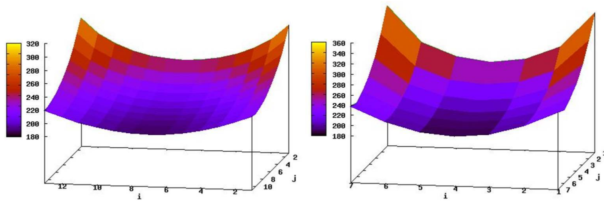
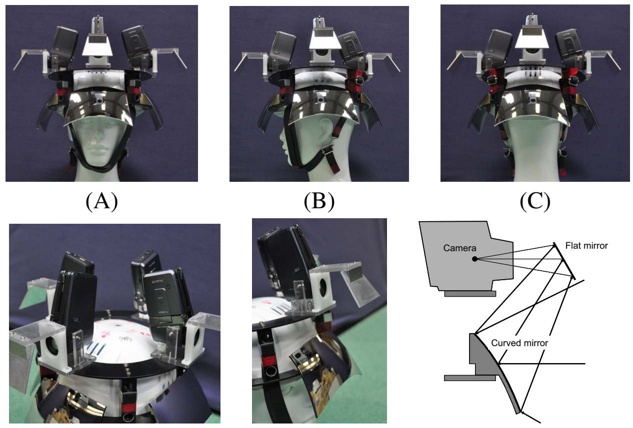


Figure 4: *Spline control points* $\mathbf{C_{ij}}$ *(Left) 13 × 11 control points (Right) 7 × 7 control points for a smooth surface*



(A) (B) (C)

(D) (E) (F)

Figure 5: *Prototype system. (A-E) Snapshots of the prototype. (F) Construction of the optical unit and its light paths.*

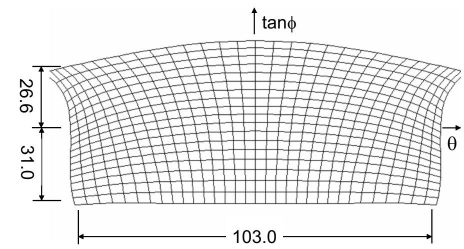


Figure 6: *Projection of equal sized squares from the image plane onto the panoramic scene. The horizontal and vertical axes represent the azimuth angle θ and tangent of the elevation angle ϕ, respectively. The effective FOV that corresponds to a maximum rectangle on the panoramic scene covered by the FOV is* $103 degrees \times -31.0 degrees \sim +26.6 degrees$.

ping area between the images captured by two adjacent cameras. This is useful when combining four image sequences to generate panoramic images. The number of control points is 13 × 11 for the mirror creation and 7 × 7 for the smoothing. The size of the curved mirror becomes about $158mm \times 100mm \times 25mm$. Fig. 3 and Fig. 4 show overviews of the designed mirror and its control points, respectively.

We have constructed an optical unit with a camera, the curved mirror, and an additional flat mirror, which aids compactness. Since flat mirrors do not affect optical properties such as focusing, we can safely ignore their influence. The total system consisting of four optical units is shown in Fig. 5, where each optical unit captures either front, right, left, or back scenes of the camera wearer. We have named the prototype system the FIrst-Person Perspective Omnidirectional imaging system : FIPPO. The performance of the FIPPO system is shown in the rightmost column of Table 1. Figure 6 shows a projection produced by the optical unit with a grid pattern on the image plane. Each optical unit covers $\pm 51.5 degrees$ in a horizontal direction and $-31.8 \sim +26.6 degrees$ in a vertical direction on the panoramic scene. When the curved mirror on front lies on a line that connects eyes of a camera wearer and that of his companion, they, of course, can not make eye contacts. Therefore the optical units are placed as close to the eyes of the camera wearer as possible without restricting his or her view. The viewpoint of the frontal optic unit is approximately $5.0cm$ higher than the eyes of the camera wearer.

Examples of captured images are shown in Fig. 12(A)-(D). These captured images have geometric warps (distortions) and chromatic differences. Image processing for solving these problems are described in Section 6.2.

4.2 Hardware construction

The cameras used in the FIPPO system are portable digital video cameras, SANYO Xacti DMX-HD2, that can capture $1280 \times 720 pixel$ images at a video rate of 30fps and 16 bit stereo sound. Captured video and audio medias are simultaneously compressed into an mpeg4 video file. We do not need an additional recording unit, thus reducing the size of the stand-alone system. The lens unit is replaced by a Fujinon DF9HA-1B with a focal length $f = 9.0mm$. The camera can be used continuously for $85 minutes$ without recharging the battery. The size of the FIPPO system is $40cm$ wide × $40cm$ deep and the total weight is about $2.4kg$. One effective approach to reduce the weight is to construct a capturing unit (currently portable video camera) with it separated to an optical part including lenses and an optical device, and a recording part consisting of other items. A recording part can be worn on another position e.g. his waist, which substantially reduces the weight of a head-mounted equipment. Additionally we can trim the weight of the metal jigs or replace them with resinic ones to reduce the total weight. Thus, being slightly heavy is not a serious problem. Unfortunately, the prototype system is not suitable for real time applications. With replacing the portable video cameras used in the prototype by online controlled cameras, it can be used for interactive applications.

5 Evaluations of image quality

5.1 Total resolution

We simulated the area of pixels included within the effective FOV of the FIPPO system as $360 degrees$ in a horizontal direction and $-31.8 \sim +26.6 degrees$ in a vertical one, excluding the overlapping areas. Pixels along the horizontal and the vertical axes depend on their positions in the scene. The FIPPO imaging system can obtain $3960 \sim 5024 pixels$ in the horizontal direction and

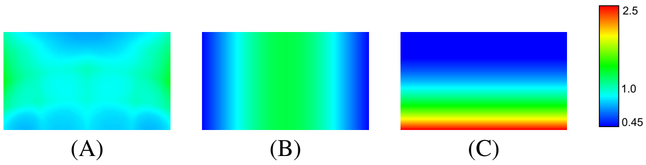


Figure 7: *Resolution maps normalized by their averages. Horizontal and vertical axes are the same as in Fig. 6. (A) FIPPO (var. = 0.01214) (B) Multiple camera (var. = 0.07670) (C) HyperOmni-Vision (var. = 0.42126)*

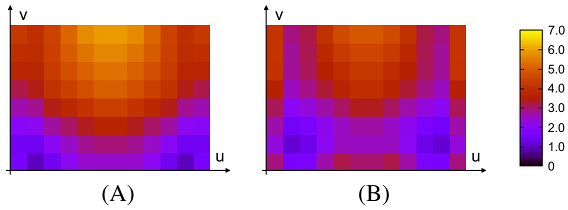


Figure 8: *RMS diameters of spot diagrams at discrete points on the $1280 \times 720 pixel$ image plane. (A) Objects at $50cm$ depth (B) Objects at $5m$ depth*

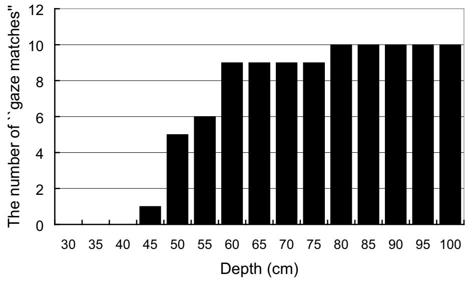


Figure 9: *Result of gaze matching voting. Horizontal and vertical axes represent the distance D and the number of "gaze matches" voting. The threshold seems to be in $50 - 55cm$.*

$584 \sim 682 pixels$ in the vertical direction. This satisfies the requirement of total resolution.

5.2 Resolution uniformity

In checking the resolution uniformity of the FIPPO, we calculated the distribution of the resolution. Because the relationship between the pixels and the rays in the scene are known, the spread of the scene corresponding to each pixel can be calculated. Resolution is then expressed by the inverse of the spread. Therefore, a panoramic scene consisting of the azimuth angle θ and tangent of the elevation angle ϕ produces resolution $R(\theta, h)$, calculated as

$$R(\theta, h) = \frac{\delta u(\theta, h) \delta v(\theta, h)}{\delta\theta\delta h} \tag{5}$$

where $h = tan\phi$, and $u(\theta, h), v(\theta, h)$ is the position of the pixel corresponding to ray (θ, h). This can be seen as the inverse of $\mathbf{V_s}(u, v)$. What we are aiming to evaluate is not the absolute value of the resolution but the distribution thereof. We therefore construct a resolution map normalized by the average value of the calculated resolution.

$$R_{normalized}(\theta, h) = \frac{R(\theta, h)}{1/S \sum_{\theta,h} R(\theta, h)} \tag{6}$$

where S defines an area of the effective FOV. Normalized resolution maps for a typical conventional omnidirectional imaging system are generated in the same way. We used a multiple camera system constructed from four pinhole cameras and an omnidirectional camera with a hyperboloidal mirror called the HyperOmni-Vision [Yamazawa et al. 1993]. For practical use, the parameters that determine the hyperboloidal mirror are $a = b = 20mm$. Figure 7 shows the maps together with their variances. The displayed ranges are $\pm 45 degrees$ along the horizontal axis and $-31.8 \sim +26.6 degrees$ along the vertical axis. This depicts a horizontal quarter of the effective FOV of the FIPPO system. Uniform color distribution corresponds to uniform resolution. The FIPPO system can produce sufficiently uniform resolution throughout the effective FOV in both horizontal and vertical directions. Although the multiple camera system produces uniform resolution in the vertical direction, resolution in the horizontal direction is not uniform. The omnidirectional camera with a hyperboloidal mirror has the reverse properties of the multiple camera system, and the degree of inconsistency is large. The results of the variances also indicate uniform resolution produced by the FIPPO system.

5.3 Aberrations

All optical units used in the FIPPO system have the same optical properties. Thus we discuss only one of these in this section. We constructed aberration maps on the image plane for scenes at two distances $50cm$ and $5m$, in order to check the focusing throughout the image and the wide range of depth. We can assume that focusing features change monotonically and are almost the same at distances over $5m$. Figure 8 shows aberration maps for $F_{number} = 8.0$. The size of the CCD image sensor used in the camera is $4.453\mu m$ square, and its diagonal is about $6.297\mu m$, which becomes the threshold to evaluate image focusing. At both distances, the diameters of the spot diagrams are less than the diagonal of an image sensor. This indicates that the FIPPO system can capture well-focused images at any distance over $50cm$. For both distances, the aberrations are larger in the upper half of the effective FOV than in the bottom half. This is because the degree of curvature of the mirror in the upper half is greater than that in the bottom half.

5.4 Gaze matching

The FIPPO system succeeds in reducing the gap related to gaze matching. We assume that the camera wearer and his companion stand face to face and gaze at each other as shown in Fig. 1. In this situation, the FIPPO system captures facial images of the companion at various distances. When the distance D is small, the facial images give viewers a "gaze does not match me" feeling. As the distance increases, this feeling is reduced until it finally changes to a "gaze matches" feeling. We displayed the facial images sequentially to ten subjects to obtain votes for "matches" and "does not match" for each distance. The results of the vote are shown in Fig. 9, and these indicate that the threshold for gaze matching is between $50cm$ and $55cm$. This satisfies the depth assumed in the optical requirements. This result confirms that the FIPPO system can appropriately express human interactions that are problematic for the conventional approaches. Swapping the front curved mirror with a curved one-way mirror with the same shape will enable the gaze mismatching distance to be eliminated.

6 Creation and Display of Video Content

6.1 Scene Capturing

In order to obtain material for video content, we filmed omnidirectional scenes from the viewpoint of an actor equipped with the FIPPO. The scenes are categorized into daily and not daily scenes. With a movie captured in a daily scene, viewers can easily compare their response to the movie to that provided by a real scene. Sports games were selected as not daily scenes. Generally, we watch sports games from outside the field or in movies captured from a third-party perspective. This is an easy way to understand the current situation of a game, but is not effective to provide a highly realistic feeling. Video capturing with the FIPPO mounted on the head of a sport player makes viewers able to feel his sense. Videos were captured from daily scenes, like the fall foliage, the pond, and the passage in a park, as well as basketball games from the viewpoint of players and referee. Attractive cuts in which people and interesting objects omnidirectionally appear were chosen to create the video content.

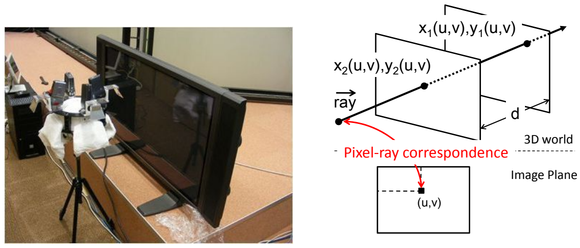
6.2 Image Processing

Correction of Image Warping
As mentioned before, images captured by the FIPPO have geometric warps that should be corrected for display. Since the curved mirrors used in the FIPPO mainly raise the warp, calibrations of the distorted projections produced by the entire optical system, including the curved mirrors, allow the images to be corrected. The calibrations were conducted with a particular scene construction illustrated in Fig. 10 in order to homologize image pixels and rays from the scene. The FIPPO placed at the front of a wide flat panel monitor captures coded patterns displayed on the monitor. The coded patterns give correspondences between each pixel on the image plane and each 2D position on the monitor. We used temporal horizontal and vertical gray code patterns. This correspondence targets a single plane at a particular depth. Measurements for planes at several depths are necessary for pixel-ray correspondence. When measurements are taken at two depths whose distances d are given, the pixel-ray correspondences can be formulated by

$$\overrightarrow{\mathbf{ray}}(u,v) = \begin{bmatrix} \frac{\overrightarrow{\mathbf{p1}} - \overrightarrow{\mathbf{p2}}}{d} \end{bmatrix} = \begin{bmatrix} \frac{x_1(u,v) - x_2(u,v)}{d} \\ y_1(u,v) - y_2(u,v) \end{bmatrix} \quad (7)$$

where $\overrightarrow{\mathbf{ray}}(u,v) = [r_x, r_y, r_z]^t$, and $\overrightarrow{\mathbf{p_i}} = [x_i, y_i]^t$ denote a ray in the world corresponding to a pixel on the image plane, and a 2D position on the monitor plane, respectively. Eq. (8) says that directions of rays are determined, but not their position. Fortunately, the FIPPO can be approximated as a single viewpoint optical system. Thus, we assume that all rays captured by the FIPPO concentrate on a center of the FIPPO. Once the calibration has been done, the image warp can be corrected by back projection of the input image onto a virtual screen in the world based on the acquired pixel-ray correspondences. We generated unwarped images with 2D texture mapping implemented in OpenGL. Figure 11 shows an example of the results in the case of back projection onto a virtual planar screen (a perspective projection). We can see that curves that come from straight lines in the original scene correctly appear as straight lines in the unwarped image. A result of a projection onto a virtual cylindrical screen (a panoramic projection) is also shown in Fig. 12.

Correction of Color Space
It is also necessary to correct chromatic differences that are mainly attributable to individual differences in the cameras. We solved



(A) (B)

Figure 10: *Calibration of the FIPPO's projection. (A) Construction for the calibration (B) How to determine pixel-ray correspondence*

(A) (B)

Figure 11: *A result of image unwarping (perspective). (A) Input image (B) Corrected image*

this problem by transforming color spaces. When the same object is captured by different cameras with the same illumination, the color of the captured images should match. Accordingly, transformations between the color spaces can be estimated, and once completed, chromatic correction can be applied for all captured images based thereon. We assumed affin transformations between the color spaces of two cameras. Although affin transformations may not approximate many different color spaces given by entirely-different imaging devices, a good approximation is expected for similar color spaces produced by the same model cameras used in the FIPPO. Under the affin transformation, two pixels on two images that should be the same color are related by

$$\begin{bmatrix} R_m \\ G_m \\ B_m \end{bmatrix} = \begin{bmatrix} p_{11} & p_{12} & p_{13} & p_{14} \\ p_{21} & p_{22} & p_{23} & p_{24} \\ p_{31} & p_{32} & p_{33} & p_{34} \end{bmatrix} \begin{bmatrix} R_n \\ G_n \\ B_n \\ 1 \end{bmatrix} \quad (8)$$

where $[R_k, G_k, B_k]^t$ and $p_{ij} (i = 1,2,3, j = 1,2,3,4)$ are RGB colors of the same object on the k-th camera and coefficients of the affin transformation, respectively. Since Eq. (9) forms three linear equations for one color correspondence, at least four color correspondences are necessary to determine twelve unknowns in p_{ij}. We used the Gretag Macbeth color checker to find twenty-four color correspondences, and then calculated the coefficients by the LMS algorithm. Figure 13 shows the results of chromatic correction. The images in the figure show a neighborhood of image mosaics. The vertical line at the horizontal center corresponds to the border between two contiguous images. Blue components that were relatively strong assume natural coloring after the correction.

6.3 Omnidirectional Theater

Presentation with only a usual flat panel monitor cannot effectively show omnidirectional movies. Omnidirectional movies should be displayed all around on viewers with a wide FOV. Researchers have proposed omnidirectional display systems for such a situation. These are categorized into personal use equipment [Hashimoto and

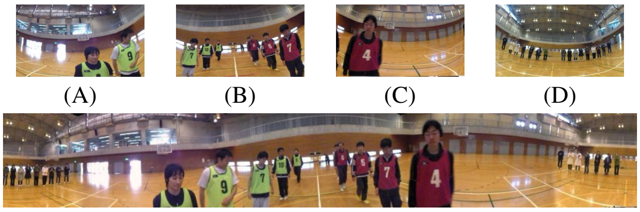


Figure 12: *A result of image unwarping (panoramic). (A)-(D) Input images for each direction, left, front, right, and back (E) Unwarped and mosaiced image.*

Figure 13: *Chromatic correction. (A) Color checkers captured by different cameras. (B) Mosaiced images without the correction. (C) That with the correction.*

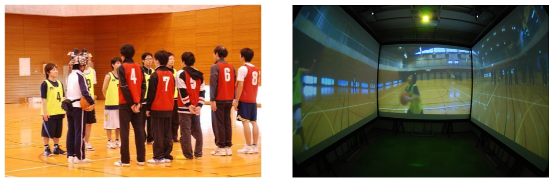


Figure 14: *Omnidirectional theater. (A) Example of captured scenes (basketball game). (B) Its display.*

Table 2: *Contents of the questionnaire.*

Questions about realistic sensation(five-grade scales).				
A. Did you got immersive feeling as you were in the scene ?				
1. Not at all	2. Not much	3. As usual	4. Fairly	5. Much
B. How much did you feel reality compared with a single front movie ?				
1. Not at all	2. Not much	3. As usual	4. Fairly	5. Much
C. How was the image quality ?				
1. Bad	2. Not good	3. Normal	4. Good	5. Great
Questions with free-form spaces.				
D. What scenes do you recommend to experience ?				
E. What additional features are necessary for the current system ?				

Iwata 1999; Nagahara et al. 2006] and dome or room type systems for multiple persons [Cruz-Neria et al. 1993]. We developed the latter type omnidirectional theater to emphasize that multiple viewers share the same feeling. The theater consists of four projectors and four $3m \times 2m$ flat screens standing like walls of a square room(Fig. 14). Although a cylindrical screen is actually a better shape to utilize the uniform resolution provided by the FIPPO, we considered that the flat screens are enough to validate the effectiveness of first-person perspective omnidirectional movies for producing immersive sensations. The omnidirectional theater also has a 3D sound system with eights speakers. In order to avoid the image display being block, the speakers are placed behind the screens especially manufactured so that sound can easily go through.

7 Application for Virtual Experience

Capabilities of the FIPPO itself have been evaluated in the previous section. Here we validate the effectiveness and demands of first-person perspective omnidirectional movies through a virtual experience appreciation.

7.1 Experimental Description

We presented the video content to ordinary people in the omnidirectional theater to evaluate an application for a virtual experience given by first-person perspective omnidirectional movies. At first, we spent a few minutes explaining the outline of our research including the objective, the capturing system (FIPPO), image processing, and the construction of the theater. We then presented the 3 minute video content consisting of the scenes addressed in section 6.1. for each group consisting of $3 \sim 5$ people. After that, we conducted questionnaires and evaluated our system based on its results. Questionnaire items are listed in Table 2. These are five-grade scale questions and questions with free-form answer spaces. The former is related to realistic sensation that the viewers felt. The latter is to obtain objective opinions about demands of first-person perspective omnidirectional movies and issues that should be improved. The above experiment had been executed in an outreach event held at the National Museum of Emerging Science and Innovation in Tokyo for three days. We got about 750 valid responses from more than $1,100$ viewers who experienced our system. Since the viewers were in a wide age range, and included groups such as couples, friends, and families, we can expected general and objective evaluations.

7.2 Results and Discussions

The results of the five-grade scale questions are summarized in Fig. 15. We can see that most subjects felt highly realistic sensations, which demonstrates the effectiveness of first-person perspective omnidirectional movies. Unfortunately, image quality got a low score. There seems to be mainly two reasons. One is optical construction of the FIPPO. Since a scene is reflected multiple times to be projected on image planes, light quantity decreases in each reflection, resulting in low quality images. The mirrors used in the prototype FIPPO are covered with low reflective material for low cost. Therefore, this problem can be solved by coating the mirrors with high reflective material to suppress light quantity decreasing. The other reason is less image contrast at the display stage. There are some causes such as output contrast of the projectors and inter reflections between the screens. One of improvement is to use gray colored and/or retroreflect screens.

The representative opinions written in the free-form space are listed in Table 3. Various answers were given, including many impossible or difficult scenes to experience. It denotes the needs and demands of virtual experiences provided by first-person-perspective omnidirectional movies. Video quaking that means image shakes and blurs caused by rapid ego-motions and limited spatial sense were pointed out as problems that should be solved, at the same time as the low image quality. Some subjects said that they felt nauseous or got dizzy. Because video quaking had been recognized in advance as being a bad influence, we carefully filmed scenes where the wearer's head was not swinging rapidly. However, the experimental results prove that this is not sufficient to suppress video quaking. In a way, our system can truly reproduce a first-person perspective, including head swing, but this is actually worse to be

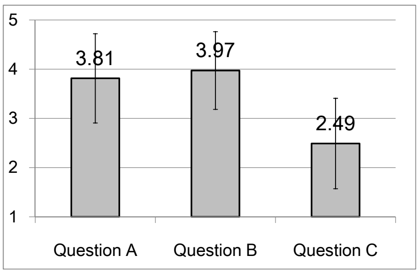


Figure 15: *Results of the five-grade scale questions.* 1 *and* 5 *denote lowest and highest scores, respectively.*

Table 3: *Representative answers written in the free-form spaces*

Answers replied to the question D	
a.	Daily life scenes
	- Crowded street, Bargain sale
	- Walking scenes in streets and forests
b.	Sports game scenes
	- Soccer, Baseball(hitter and pitcher), Tennis, Boxing
	- Ski, Surfing, Bungy jump, Wrestling, Sky diving
	- Scuba diving, American football, Canoe
c.	Vehicle scenes
	- Cycling, Roller coaster, Moter bike, Ferris wheel
d.	Experiences at polar fields
	- In the sea, sky, cosmo space
e.	Entertainments
	- Horror film, Haunted house, caving
	- Singer's view on a stage
f.	Animal
	- Dog's view, man eaten by a lion
Answers replied to the question E	
a.	Quality of the medias
	- Improveme image quality (resolution and contrast)
b.	Reproduction of spatial sense
	- Desire sense of correct depth
	- Recognize shapes of objects and scenes
c.	Video quaking
	- Little sick, nauseous, dizzy
	- Film without head swinging

displayed to static viewers. Recording a head state with a gyro sensor or ego-motion estimation algorithms that use horizontal cyclic property of omnidirectional movies will help with video stabilization. Fundamental approaches are needed to provide a spatial sense. Omnidirectional scenes must be spatially constructed, which requires special capturing equipment. A three dimensional display all around viewers is also a challenging issue.

8 Conclusion

In this paper, we proposed a novel wearable omnidirectional imaging system to capture images from the first-person perspective. The prototype system captures omnidirectional images with uniform resolution, using specially designed optics that include curved mirrors. As the viewpoint of the system is positioned close to the eyes of the camera wearer, gaze matching is achieved for near objects. Effectiveness and demands of first-person perspective omnidirectional movies had been validated through presentation experiments for ordinary people. The prototype FIPPO system is relatively large and heavy. A more compact system with less weight is needed for practical use. Improvement of image quality and producing spatial sense are also necessary for the current display system. We now plan to solve these issues to provide a more immersive visual and acoustic media. Additional experience-based demonstration that validates the system from its design aspect is also necessary.

Acknowledgment

This work is supported by the Special Coordination Funds for Promoting Science and Technology of Ministry of Education, Culture, Sports, Science and Technology. The authors wish to thank SANYO Electric Co. Ltd. for providing the specially modified portable video cameras.

References

ANSTIS, S. M., MAYHEW, J. W., AND MORLEY, T. 1969. The perception of where a face or television portrait is looking. *American Jounal of Psychology 82*, 4 (Dec.), 474–489.

CRUZ-NERIA, C., SANDIN, D. J., AND DEFANTI, T. A. 1993. Surround-screen projectorion-based virtual reality : The design and implementation of the cave. In *Proc. of SIGGRAPH1993*, 135–142.

ENOMOTO, S., IKEDA, Y., ISE, S., AND NAKAMURA, S. 2008. Three-dimensional sound field reproduction and recording system based on the boundary surface control principle. In *The 14th Int. Conf. on Auditory Display*, o_16.

GLUCKMAN, J., AND NAYAR, S. K. 1998. Egomotion and omnidirectional cameras. In *Proc. of Int. Conf. on Computer Vision*, 999–1005.

HASHIMOTO, W., AND IWATA, H. 1999. Ensphered vision: Spherical immersive display using convex mirror. *Trans. of the Virtual Reality Society of Japan 4*, 3, 479–486.

HICKS, R. A. 2005. Designing a mirror to realize a given projection. *Journal of Optical Society of America 22*, 2, 323–329.

IWATA, H. 2004. Full-surround image display technologies. *International Journal of Computer Vision, 58*, 3, 227–235.

NAGAHARA, H., YAGI, Y., AND YACHIDA, M. 2006. Super wide field of view head mounted display using catadioptrical optics. *Presence: Teleoperators & Virtual Environments 15*, 5, 588–598.

POINT GREY RESEARCH INC. *Ladybug2*. http://www.ptgrey.com/products/ladybug2/index.asp.

SWAMINATHAN, R., NAYER, S. K., AND GROSSBERG, M. D. 2004. Designing of mirrors for catadioptric systems that minimize image error. In *Proc. of IEEE Workshop on Omnidirectional Vision and Camera Networks*.

SWAMINATHAN, R. 2007. Focusing in catadioptric imaging system. In *Proc. of IEEE Workshop on Omnidirectional Vision and Camera Networks*, 1–7.

TANAHASHI, H., YAMAMOTO, K., WANG, C., AND NIWA, Y. 2000. Development of a stereo omni-directional imaging system (sos). In *Proc. of IEEE IECON-2000*, 289–294.

YAMAZAWA, K., YAGI, Y., AND YACHIDA, M. 1993. Omnidirectional imaging with hyperboloidal projection. In *Proc. of IEEE/RSJ Int. Conf. on Intelligent Robotics and Systems*, vol. 2, 1029–1034.

YAMAZAWA, K., TAKEMURA, H., AND YOKOYA, N. 2002. Telepresence system with an omnidirectional hd camera. In *Proc. of 5th Asian Conf. on Computer Vision*, vol. 2, 533–538.