Binocular Spherical Stereo

Shigang Li

***Abstract*—A fish-eye camera has a wide field of view (FOV), and the realization of a binocular fish-eye stereo for sensing the surrounding 3-D information of the environment around a vehicle is useful for safe driving. However, since a fish-eye camera may have a wider-than-hemispherical FOV, the conventional stereo approach of obtaining a perspective image based on the pinhole camera model cannot directly be applied. However, using a spherical camera model and defining the disparity of a spherical stereo, the conventional binocular stereo problem is reformulated as a *binocular spherical stereo* problem. A binocular spherical stereo is a generalized paradigm that can cope with cameras having any FOV, including conventional cameras and fish-eye cameras. Moreover, by transforming the rectified spherical images to latitude–longitude representation, the feature point matching of the spherical stereo images can be sped up by using the processing used for perspective stereo images. The effectiveness of this approach is demonstrated by realizing a binocular spherical stereo using a pair of fish-eye cameras. Finally, the application of the proposed approach to vehicles in the future is considered.**

***Index Terms*—Driving assistance, fish-eye camera, stereo vision, 3-D information.**

I. Introduction

VISUAL sensing around a vehicle to avoid blind spots is important for safe driving. Nissan Motors has developed a visual sensing system that consists of four wide-view cameras mounted on the four sides of the vehicle. Fig. 1 shows the display of the visual sensing system. Since blind spots are almost completely eliminated, the driver can visually examine the entire area surrounding the vehicle, allowing safer and faster parking.

A similar system consisting of four fish-eye cameras has been proposed by Alps Electronics Company, as shown in Fig. 2. Since each of the fish-eye cameras has a maximum field of view (FOV) of 190°, the entire environment around the vehicle can be observed. Compared with systems that use normal cameras, which have a limited FOV, these two systems use a small number of cameras, which results in a simpler system configuration. Although a catadioptric omnidirectional camera consisting of a camera and a mirror can also observe a wide environment, this setup is not suitable for this task, because the camera lens is imaged at the center of the omnidirectional image [5], [6], [9], [17].

Manuscript received November 30, 2006; revised July 30, 2007, April 19, 2008, and May 7, 2008. First published November 7, 2008; current version published December 1, 2008. This work was supported in part by SpherEye Co. Ltd. The Associate Editor for this paper was S. Nedevschi.

This paper has supplementary downloadable material available at http://ieeexplore.ieee.org, provided by the authors.

The author was with the Faculty of Engineering, Iwate University, Morioka 020-8551, Japan. He is now with the Faculty of Engineering, Tottori University, Tottori 680-8550, Japan (e-mail: li@ele.tottori-u.ac.jp).

Digital Object Identifier 10.1109/TITS.2008.2006736

Fig. 1. Around view system of Nissan Motors.

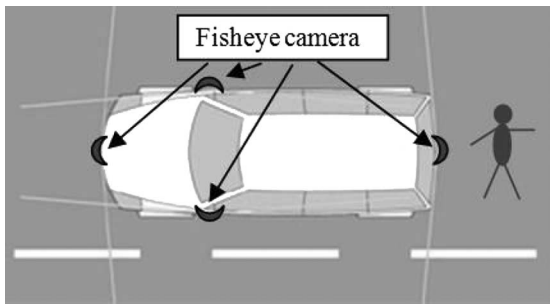


Fig. 2. Four-fish-eye-camera system of Alps Electronics Company.

Although the two aforementioned systems can provide visual information around a vehicle, it is difficult to detect the 3-D information of the environment. Although the motion stereo approach (which uses two images captured at different times at different places by the same camera) can be applied to 3-D environment analysis, this setup requires the environment to be static and the vehicle to move for a distance to obtain a baseline length for the stereo approach. Since a vehicle generally moves in a dynamic environment and safety should be considered in all situations, a natural choice is to employ the binocular stereo approach in which two cameras are used to observe the same environment at the same time. Not only can such a system provide the driver with visual information around the vehicle, but it can also detect obstacles around the vehicle. The latter information is very important when a driver wants to back up or park a vehicle.

The stereo method is a relatively mature concept in computer vision, and the stereo method has been investigated in a number of studies. However, most of these studies used normal cameras with a limited FOV, and the analyses were based on the pinhole camera model. This approach may result in blind spots for visual sensing around a vehicle. As shown in Fig. 3(a), the 3-D information can only be computed for the gray region, which

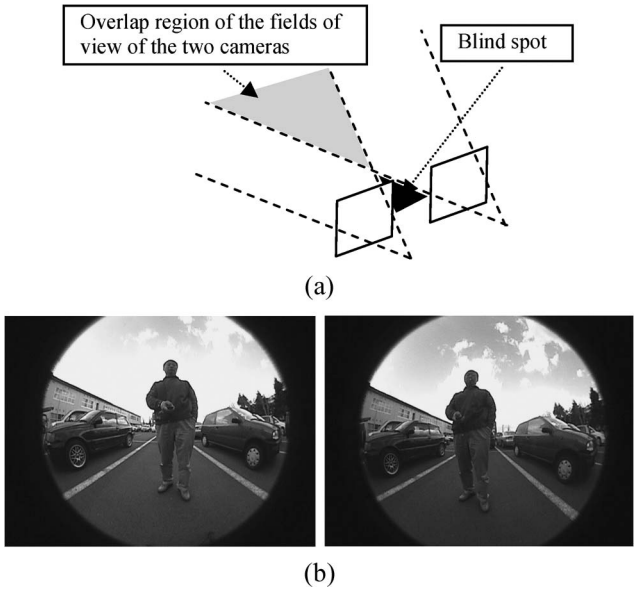


Fig. 3. (a) Using two normal cameras, the 3-D information can only be computed for the overlapped gray region, and a blind spot (shown in black) exists between these regions. (b) Pair of fish-eye images captured by the fish-eye stereo rig developed in this paper. Using a pair of stereo fish-eye images, the overlap region may be expanded to the entire area covered by the cameras.

is simultaneously observed by both cameras, and a blind spot (indicated in black) exists between the two cameras.

On the other hand, using a pair of stereo fish-eye images, as shown in Fig. 3(b), the overlap region of the FOVs of the two cameras may expand to the entire area covered by the cameras. In addition, the ability to compute the 3-D information of the overlap region of the observation fields of the two fish-eye cameras would be very helpful for safe driving.

However, a fish-eye camera may have a wider-than-hemispherical FOV. Therefore, the conventional pinhole camera model cannot be applied to the fish-eye camera.

In this paper, a binocular spherical stereo method based on a spherical camera model is proposed so that the approach used for conventional stereo for perspective images can be applied to cameras having any FOV. A spherical camera model is used to describe the fish-eye camera with a wider-than-hemispherical FOV. The conventional binocular stereo problem is reformulated by defining the disparity of the spherical stereo. Based on this definition, the binocular spherical stereo becomes a general model that can cope with any cameras having wide FOV, including conventional cameras. Moreover, by transforming the rectified spherical images to latitude–longitude representation, the feature point matching of spherical stereo images can be sped up by using the same processing as the conventional perspective stereo images. This paper is an extension of the research presented in [30].

The remainder of this paper is organized as follows: Related research is described in Section II. In Section III, the binocular spherical stereo method is introduced. A simulation for quantitative evaluation of errors is carried out in Section IV. The experimental results obtained using the two-fish-eye-camera stereo system are described in Section V. Finally, conclusions are presented in Section VI.

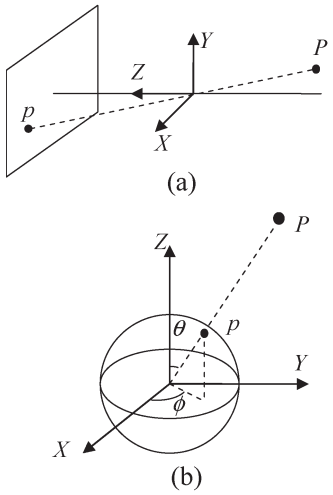


Fig. 4. (a) Perspective projection based on a pinhole camera model. (b) Spherical projection based on a spherical camera model.

II. Related Work

A. Pinhole Camera Model Versus Spherical Camera Model

In computer vision, the most popular camera model is the pinhole camera model, in which the environment is perspectively projected onto a planar image, as shown in Fig. 4(a) [1], [4], [7]. However, the pinhole camera model cannot cope with a wider FOV, compared with the hemispherical camera model, because the pinhole camera model requires an image of infinite size to represent a hemispherical FOV.

On the other hand, a spherical image has a full FOV. Spherical images have been used for object recognition [8], radiance mapping of the environment [3], environment mapping [2], and motion estimation [13], [16], [18]. In addition, the combination of a spherical projection with a spherical image has been used to estimate head pose from the surrounding marks for mapping the head motion to a virtual space [12] and to formulate the spherical stereo vision [15] so that the camera motion can be estimated and the structure of the environment can be recovered from spherical images.

Suppose that there is a sphere of radius f_s and a point P in space, as shown in Fig. 4(b). The intersection of the sphere surface with the line-intersecting point P and the center of the sphere is the projection of the point to the sphere and is called a *spherical projection* [15]. A spherical image is obtained by projecting all the visible points from the sphere. In other words, a spherical image is the surface of a sphere, with the focus point at the center of the sphere.

Let the coordinates (relative to the spherical image coordinate system) of a point P in space be $M_c = [X_c \quad Y_c \quad Z_c]^T$, and let the projection in the spherical image be $m = [f_s \sin\theta\cos\phi \quad f_s \sin\theta\sin\phi \quad f_s\cos\theta]^T$. The relation between the coordinates and the projection is

$$m = \lambda M_c \cong M_c \tag{1}$$

where $\lambda = f_s/\rho$, and $\rho = \sqrt{X_c^2 + Y_c^2 + Z_c^2}$.

Therefore, m is equal to M_c multiplied by a scale factor. Letting $f_s = 1$, the normalized spherical image coordinates are

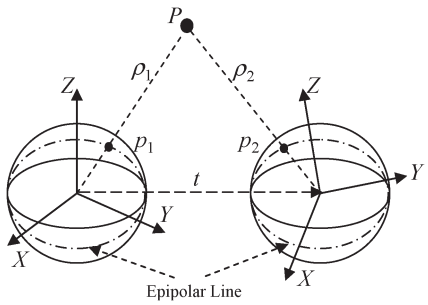


Fig. 5. Spherical stereo. A scene point is observed by a pair of spherical images.

then acquired. This corresponds to the projection onto a unit sphere.

Compared with the perspective projection of a planar image based on the pinhole camera model, spherical projections have three characteristics.

1) Every point in space can *equally* be represented in a spherical projection as long as it is visible from the spherical view.
2) The coordinate system of a spherical image is the *same* as the camera coordinate system.
3) Since every ray from the environment point to the spherical image is perpendicular to the spherical surface, *there is no definite optical axis* used as the conventional perspective projection.

B. Planar Stereo Versus Spherical Stereo

In [15], it is shown that, based on spherical projection, the conventional planar image stereo method (referred to hereinafter as the *planar stereo*) can be reformulated and applied to the spherical stereo images to measure the 3-D information of the environment points (referred to hereinafter as the *spherical stereo*), as shown in Fig. 5. The spherical stereo can be regarded as a unification of the planar and panoramic stereos. Using a full-view spherical image, the surrounding structure of the environment can directly be acquired.

However, although Li and Fukumori [15] gave the basic framework of the spherical stereo, unsolved problems for applying the spherical stereo in a dynamic environment remain.

1) In [15], a pair of spherical images is captured at different positions by moving the camera. This motion stereo approach cannot cope with a dynamic environment, as mentioned before. Considering a vehicle moving in a highly dynamic environment, it is necessary to use multiple cameras to capture multiple views at the same time, which is referred to as real-time stereo [10], [11], [20], [23]–[26], [32], for the acquisition of the 3-D structure of the environment. In this paper, a method of *binocular spherical stereo* that uses two spherical cameras is proposed.
2) To quickly compute the 3-D information, a pair of spherical cameras must be calibrated, and the captured spherical images must be rectified so that the correspondence of feature points can easily be carried out.

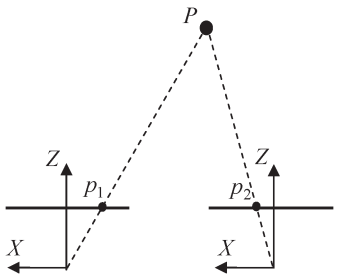


Fig. 6. Disparity definition for canonical stereo.

3) Instead of sparse 3-D information from only some corresponding feature points, as in [15], a dense 3-D representation is required for many tasks of an intelligent vehicle. On the other hand, the correlation-based stereo may be the most popular and practical approach to achieve this goal. Therefore, the question arises as to whether it is possible to realize a correlation-based binocular spherical stereo.

In this paper, a method that can overcome the preceding problems is proposed. The spherical stereo is defined for a stereo rig. After rectifying the pair of spherical images, they are transformed into a latitude–longitude representation so that the epipolar lines of spherical stereo are parallel to each other, and the conventional correlation method for stereo correspondence can directly be applied to acquire dense 3-D information of the environment with the same speed as the correlation-based planar stereo. Since the binocular spherical stereo is a general approach and is independent of the camera FOV, this method can be applied to fish-eye stereo cameras to acquire 3-D information around a vehicle.

III. Binocular Spherical Stereo

Fig. 6 shows a pair of rigidly attached pinhole cameras in a canonical configuration with baseline b. The disparity d of the planar stereo is defined as the difference in image coordinates, i.e., $x_r - x_l$, and the depth of the point z from the camera is computed as

$$z = b\frac{f_p}{d} \tag{2}$$

where f_p is the focal length of both pinhole cameras [7].

Two problems occur when this approach is applied to fish-eye cameras with a wider-than-hemispherical FOV.

1) The disparity is defined as the difference between the horizontal coordinates of two projections p_1 and p_2 of environment point P in the canonical stereo in Fig. 6. If environment point P significantly moves along the direction of the x-axis, the horizontal coordinates of p_1 and p_2 may become very large, which may cause a computational error in the disparity when a very small disparity is computed from two very large horizontal coordinates by a digital computer. The aforementioned problem will occur if the conventional pinhole camera model is applied to hemispherical-FOV fish-eye stereo images.

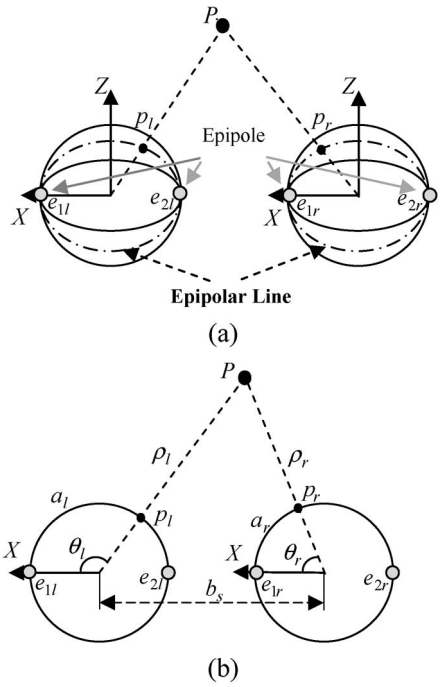


Fig. 7. (a) Definition of spherical stereo for a spherical stereo rig. (b) Definition of disparity for spherical stereo.

2) In the planar stereo approach, the *depth* (the z-coordinate value of the observed points at the camera coordinate system) is usually used to describe how far an environment point is located from the camera and is directly computed from (2). This is because the depth of an environment point imaged by a narrow-FOV camera is similar to the distance of the point from the camera. However, the *depth* may be significantly different from the distance of an environment point to the camera for the spherical stereo, because an environment point may be in any direction relative to a spherical camera and thus may have a great distance but a small *depth* value.

To cope with full-view spherical images, the disparity of spherical stereo must be defined, and the binocular stereo algorithm must be reformulated.

A. Disparity Definition of Spherical Stereo

Suppose that a pair of spherical cameras exists (which captures spherical images according to the spherical projection [15]). In addition, assume that each pair of axes of the camera coordinate systems is parallel and that the pair of x-axes is collinear, as shown in Fig. 7(a). The baseline of the spherical stereo is defined as the line connecting the two focuses, and the epipoles of each spherical image are the intersections of the baseline, with the spherical image located at the x-axis, and can be denoted by (± 1, 0, 0) in normalized spherical image coordinates. The epipolar lines are great circles, which are defined as circles that are the intersections of the sphere with a plane through the center of the sphere [34].

To define the disparity of spherical stereo, we focus on an epipolar plane, which is decided by an environment point and the two focuses, and depict it in Fig. 7(b). Let the angles of the projection of point P onto the spherical images with the x-axis be θ_l (with the polar angle relative to the x-axis) and θ_r, respectively. The disparity d_s of point P is defined as the difference in the lengths of arcs a_l and a_r subtended by θ_l and θ_r, respectively, i.e.,

$$d_s = a_l - a_r = f_s(\theta_l - \theta_r) \tag{3}$$

where f_s is the radius of the great circle, i.e., the focal length of the spherical image.

For a normalized spherical image, $f_s = 1$, and $d_s = \theta_l - \theta_r$, where d_s is the *normalized disparity* d_n of the spherical stereo, i.e.,

$$d_n = \theta_l - \theta_r. \tag{4}$$

Note that normalized disparity d_n corresponds to the *angle* $\angle P$ subtended by the rays from point P to the focuses of the pair of spherical images, independent of the intrinsic parameters of the spherical cameras. The aforementioned disparity is called the *spherical disparity* to distinguish it from the conventional disparity of planar stereo images.

B. Computing the Distance of Environment Points

Since the environment points of the surrounding environment are equally represented in spherical images, the point location in space is described by the *distance* in the spherical stereo, instead of the *depth*, as in the planar stereo, as previously mentioned. Suppose that the arcs subtended by angles θ_l and θ_r are a_l and a_r, respectively. In terms of the sine theorem, the distances of point P from the two spherical cameras are

$$\begin{aligned}
\rho_l &= b_s \frac{\sin(\theta_r)}{\sin(\theta_l - \theta_r)} \\
&= b_s \frac{\sin(a_r/f_s)}{\sin(a_l/f_s - a_r/f_s)} \\
&= b_s \frac{\sin(a_r/f_s)}{\sin(d_s/f_s)} \\
\rho_r &= b_s \frac{\sin(\pi - \theta_l)}{\sin(\theta_l - \theta_r)} \\
&= b_s \frac{\sin(\pi - a_l/f_s)}{\sin(a_l/f_s - a_r/f_s)} \\
&= b_s \frac{\sin(\pi - a_l/f_s)}{\sin(d_s/f_s)}
\end{aligned} \tag{5}$$

respectively, where θ_l and θ_r can simply be computed from the spherical image coordinates of the projection.

For a normalized spherical image, the preceding expression is simplified as

$$\begin{aligned}
\rho_l &= b_s \frac{\sin(\theta_r)}{\sin(d_n)} = b_s \frac{\sin(\theta_l - d_n)}{\sin(d_n)} \\
\rho_r &= b_s \frac{\sin(\pi - \theta_l)}{\sin(d_n)} = b_s \frac{\cos(\theta_r + d_n)}{\sin(d_n)}.
\end{aligned} \tag{6}$$

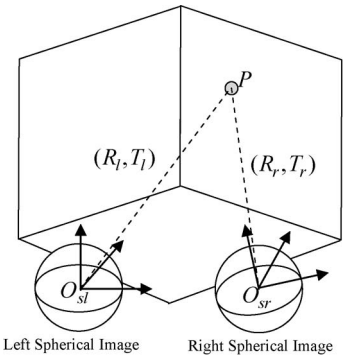


Fig. 8. Using a box pattern to calibrate a spherical stereo rig. The external parameters of each camera are used to rectify the input spherical images.

Thus, the spherical cameras can be rectified so that they are aligned, as shown in Fig. 7(a); the point correspondence of the environment points can be found from the spherical stereo image pairs, the disparity of the environment point can be computed as (3) or (4), and its distance from the spherical camera can be computed as (5) or (6).

C. Rectification of Spherical Stereo Images

Previously, the method of computing the 3-D information of the environment points is given under a spherical canonical stereo model. However, for a real stereo rig, the two cameras may not strictly obey the constraints. The captured spherical images must be rectified to acquire a pair of spherical images that obey the canonical stereo model. The algorithm of [28] can be revised and applied to the present case.

An environment point $P(X, Y, Z)$ in the global coordinate system attached to the calibration pattern is projected to spherical image $p(u, v, s)$ [(u, v, s) are the Cartesian coordinates on the spherical image] via external parameters R and T as follows:

$$p \cong [R \quad T]\widetilde{P} \tag{7}$$

where $\widetilde{P}$ is the homogeneous coordinate of P.

Suppose that the external parameters of the left and right spherical images are (R_l, T_l) and (R_r, T_r), respectively, as shown in Fig. 8. The coordinates of the origins of the two input spherical images $O_{\rm sl}$ and $O_{\rm sr}$ in the global coordinate system are computed as follows:

$$O_{\rm sl} = -R_l^{-1}T_l, \qquad O_{\rm sr} = -R_r^{-1}T_r. \tag{8}$$

Then

$$-R_l O_{\rm sl} = T_l, \qquad -R_r O_{\rm sr} = T_r. \tag{9}$$

The projections of environment point P onto the two spherical images $p_{\rm sl}$ and $p_{\rm sr}$ are represented as follows:

$$p_{\rm sl} \cong [R_l \quad -R_l O_{\rm sl}]\widetilde{P}, \qquad p_{\rm sr} \cong [R_r \quad -R_r O_{\rm sr}]\widetilde{P}. \tag{10}$$

To rectify the input spherical images, the two spherical images must be rotated so that the axes of the two coordinate systems of the rectified spherical images are parallel to each other, and one axis passes through the centers of both spherical images.

Suppose that, after the rectification, the projections of environment point P onto the two spherical images can be represented as follows in terms of (10):

$$p_{\rm sl} \cong [R \quad -RO_{\rm sl}]\widetilde{P}, \qquad p_{\rm sr} \cong [R \quad -RO_{\rm sr}]\widetilde{P}. \tag{11}$$

Rotation matrix $R = [r_1 \quad r_2 \quad r_3]^T$ is determined in three steps.

1) Let the x-axis be parallel to the line joining the two centers of the rectified spherical images. Thus, we have

$$r_1 = \frac{(O_{\rm sl} - O_{\rm sr})}{\|O_{\rm sl} - O_{\rm sr}\|}. \tag{12}$$

2) Let the unit vector of the z-axis of one input spherical image be k, and r_2 is determined as follows:

$$r_2 = k \times r_1. \tag{13}$$

3) Finally, r_3 is determined in terms of the property of a rotation matrix, i.e.,

$$r_3 = r_1 \times r_2. \tag{14}$$

Next, the brightness of the pixels must be mapped from the input spherical images to the rectified images. An explanation is given for the left spherical image.

Suppose that an environment point P is projected onto the input left spherical image as $p_{\rm slo}$ and the rectified left spherical image as $p_{s\,\ln}$. Then

$$p_{\rm slo} \cong [R_l \quad -R_l O_{\rm sl}]\widetilde{P}, \qquad p_{s\,\ln} \cong [R \quad -RO_{\rm sl}]\widetilde{P} \tag{15}$$

are obtained, and

$$P = O_{\rm sl} + \lambda_o R_l^{-1} p_{\rm slo}, \qquad P = O_{\rm sl} + \lambda_n R^{-1} p_{s\,\ln} \tag{16}$$

where λ_o and λ_n are the scale factors.

Next, we have

$$p_{s\,\ln} = \lambda R R_l^{-1} p_{\rm slo} \tag{17}$$

where $\lambda = \lambda_o/\lambda_n$.

Since the input spherical image and the rectified image have the same image sizes, $\lambda = 1$.

In practice, a spherical image is an intermediate representation of wide-FOV images. The aforementioned external parameters of the rectified spherical images can be acquired from the camera calibration process. An example of the calibration of a fish-eye camera stereo rig is given in Section V.

D. Correlation-Based Correspondence of Spherical Stereo

In contrast, the point correspondence searching of the planar stereo is limited on the parallel straight lines after the rectification of planar stereo images, and the point correspondence searching must be carried out along great circles in spherical stereo images, as previously described. Since searching along

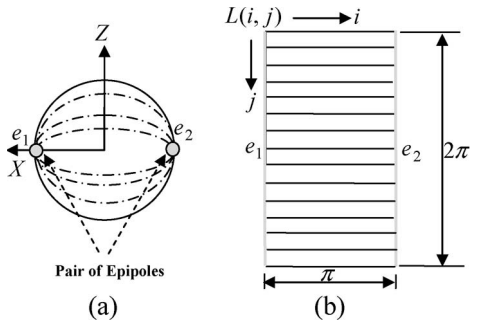


Fig. 9. (a) Pair of epipoles on the spherical image. (b) Extend the spherical image to obtain parallel epipolar lines by regarding the epipoles as the two poles and performing latitude–longitude sampling.

conic curves has a higher computation cost, compared with that along straight lines, this process must be sped up to apply this approach to real-time tasks of a vehicle. Next, an approach to achieve this goal is described.

The idea is to transform the rectified spherical image, so that the conventional correlation-based approach for planar stereo images can directly be applied to the spherical stereo image. Since the epipolar lines in the spherical stereo are the great circles intersecting the epipoles, as shown in Fig. 9(a), the epipoles can be regarded as the two poles, and the spherical image is extended by latitude–longitude sampling. Thus, the great circles intersecting at the epipoles become parallel straight lines, and the conventional area-correlation-based correspondence method can be applied to compute the disparity of the spherical stereo images, as shown in Fig. 9(b).

Concretely, a point in the latitude–longitude sampling image $L(i,j), (i = 0, 1, \ldots, m;\ j = 0, 1, \ldots, n)$ can be mapped to the point in the rectified normalized spherical image $p(u, v, s)$ using the following expression:

$$\begin{aligned} u &= \cos\left(\frac{i}{m}\pi\right) \\ v &= \sin\left(\frac{i}{m}\pi\right)\cos\left(\frac{j}{n}2\pi\right) \\ s &= \sin\left(\frac{i}{m}\pi\right)\sin\left(\frac{j}{n}2\pi\right) \end{aligned} \tag{18}$$

where m and n correspond to the arc length between the two epipoles and the circumference of the great circle of the spherical image, respectively, and $(i/m)\pi$ and $(j/n)\pi$ can be regarded as the polar and azimuth angles, respectively, in the rectified spherical image if the two epipoles of Fig. 9(a) are defined as the two poles of the Earth, and (u, v, s) are the Cartesian coordinates computed from the polar and azimuth angles.

In terms of the internal and external camera parameters, the point in the rectified normalized spherical image $p(u, v, s)$ can be mapped onto a real captured digital image $I(x, y)$ [31].

IV. Simulation for Quantitative Evaluation

According to camera parameters, the direction of an environment point can be determined from its projection position in the image. The computation of the 3-D information of an environment point is equal to the distance of the environment point from the reference camera of a stereo rig. Theoretically, the distance from the environment points to the cameras can be computed in terms of (5) if the point matching between the two spherical images is given. In practice, there may be errors in the distance computation due to the errors in image digitization, camera parameter estimation, and point matching. Here, to quantitatively evaluate the computational error, a simulation experiment is performed. Considering the symmetrical properties of the spherical stereo image, the error analysis of the distance computation can be limited to half of the epipolar plane (see Fig. 7).

Performing a total differentiation of (5), the following inequality is obtained:

$$\begin{aligned} |\Delta\rho_l| \le & \left| -b_s \frac{\sin(a_r/f_s)\cos(a_r/f_s - a_l/f_s)}{\sin(a_r/f_s - a_l/f_s)\cdot f_s} \right| |\Delta a_l| \\ & + \left| -b_s \frac{\cos(a_r/f_s)}{\sin(a_r/f_s - a_l/f_s)\cdot f_s} \right. \\ & \left. + b_s \frac{\sin(a_r/f_s)\cos(a_r/f_s - a_l/f_s)}{\sin(a_r/f_s - a_l/f_s)\cdot \sin(a_r/f_s - a_l/f_s)\cdot f_s} \right| \\ & \times |\Delta a_r|. \end{aligned} \tag{19}$$

Since the errors in image digitization, camera parameter estimation, and point matching cause the changes in Δa_l and Δa_r, the value on the left-hand side of (19) gives a criterion of the error in the computed distance from the left camera.

Suppose that the two spherical cameras are separated by 20 cm (which is the same as the real fish-eye camera stereo rig used in the experiment presented in the next section) and that the environment points are arranged in a circle with a radius of 100 cm centered at the midpoint of the baseline of the two cameras, as shown in Fig. 10(a). The left camera is selected as the reference camera and marked as "+." Let the radius of the spherical image be $f_s = 480/\pi$ pixels, which corresponds to the focal length of an equidistant projection fish-eye camera covering a hemispherical FOV using a Video Graphics Array image, the size of which is the same as that of the following real-world experiment. While the environment point moves from the left side to the right side along the circle, the length of arc a_l between the left epipole and the projection of the environment point changes from 0 to 480 pixels [see Fig. 7(b)]. Assuming $\Delta a_r = \Delta a_l = 0.5$ pixels, the computation error $\Delta\rho_l$ in the distance of the environment points from the left camera is obtained, as shown in Fig. 10(b). As the projection position of the point approaches the epipoles, the difference between a_l and a_r becomes smaller. This results in a greater error in distance computation due to the errors in Δa_l and Δa_r. For the environment points on the baseline, except for the line segment joining the two cameras, the positions of their projections are just at the epipoles for both the left and right spherical images. Therefore, $a_r = a_l$, and the disparity becomes zero. In this case, the distance of such environment points cannot be computed in terms of (5). As shown in Fig. 10(b), the error in the estimated

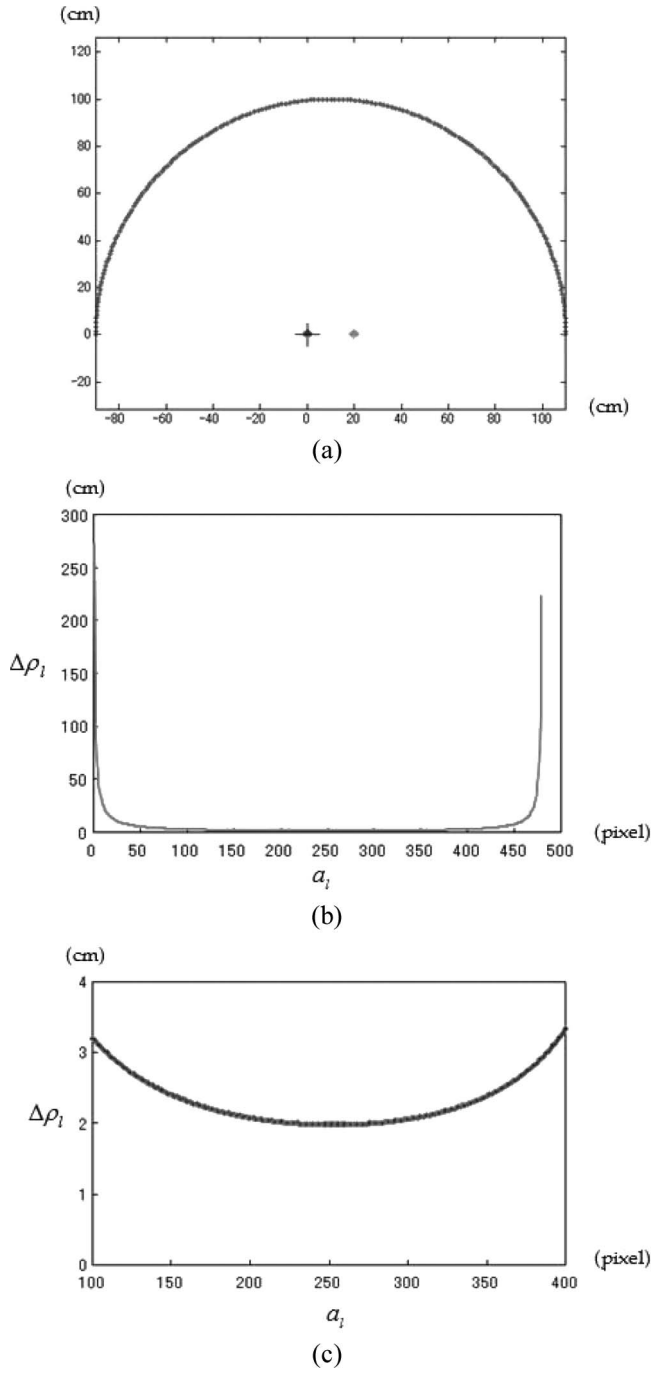


Fig. 10. (a) Environment points are arranged at the circle centered at the middle point of the baseline of the two cameras. (b) Errors in the estimated distance of the environment point by the binocular spherical stereo. (c) Enlarged part of (b) for the pixels in between 100 and 400.

3-D information abruptly becomes large as the projection of the points nears the epipoles.

A spherical image is an intermediate representation of a real image captured by image sensors, such as a fish-eye camera. Due to the errors in the estimation of the cameras' internal and external parameters, there may be distortion in the generated spherical images. Errors in the projection position on the spherical image and disparity estimation may also be generated. Fig. 11 shows the error $\Delta\rho_l$ in the computed distance of the environment points in Fig. 10(a), while Δa_l and Δa_r

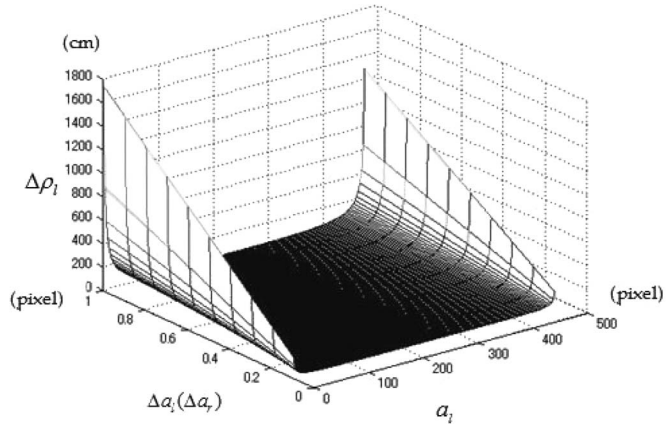


Fig. 11. Error in the computed distance of the environment in Fig. 10(a) while the error of the projection points simultaneously changes from 0.1 to 1.0 pixels.

simultaneously change from 0.1 to 1.0 pixels. $\Delta\rho_l$ increases with the increase in Δa_l and Δa_r. This implies that a more accurate distance can be computed if a higher resolution image is used.

The sensitivity of error $\Delta\rho_l$ in the distance computation for the binocular spherical stereo to the error in Δa_l and Δa_r is also dependent on the positions of the environment points. Fig. 12(a) shows a group of grid points on half of the epipolar plane. Fig. 12(b) shows the error $\Delta\rho_l$ in the computed distance at each point. In Fig. 12, the position of the reference camera of the computed distance is indicated by "+," and $|\Delta a_r| = |\Delta a_l| = 0.5$.

From the preceding error analysis on the distance computation, we can yield three conclusions.

1) In terms of (19), the error in the distance is inversely proportional to the radius f_s of the spherical images. Since a larger radius means a higher resolution of the spherical image, a higher resolution spherical image can improve the accuracy of the computed distance of the environment points.
2) In terms of (19) and the simulation result in Fig. 11, the error in the distance is proportional to the error in Δa_l and Δa_r. Therefore, to acquire accurate distance estimation, accurate projection position of the environment on the spherical image and correct stereo matching are required.
3) In the binocular spherical stereo, the accuracy of the computed distance abruptly decreases near the epipoles, as shown in Figs. 10 and 12. Therefore, there is a spot called a 3-D blind spot centered at the epipoles for which the computation error is too large to be used. The effective range of the spherical image points depends on concrete tasks. The higher the required accuracy of the computed 3-D information, the more narrow the effective range. Note that, if the conventional pinhole camera model is applied to hemispherical-FOV fish-eye images, there is a band for which the 3-D information cannot be acquired, because an infinitely large image is required to represent a hemispherical view using perspective projection.

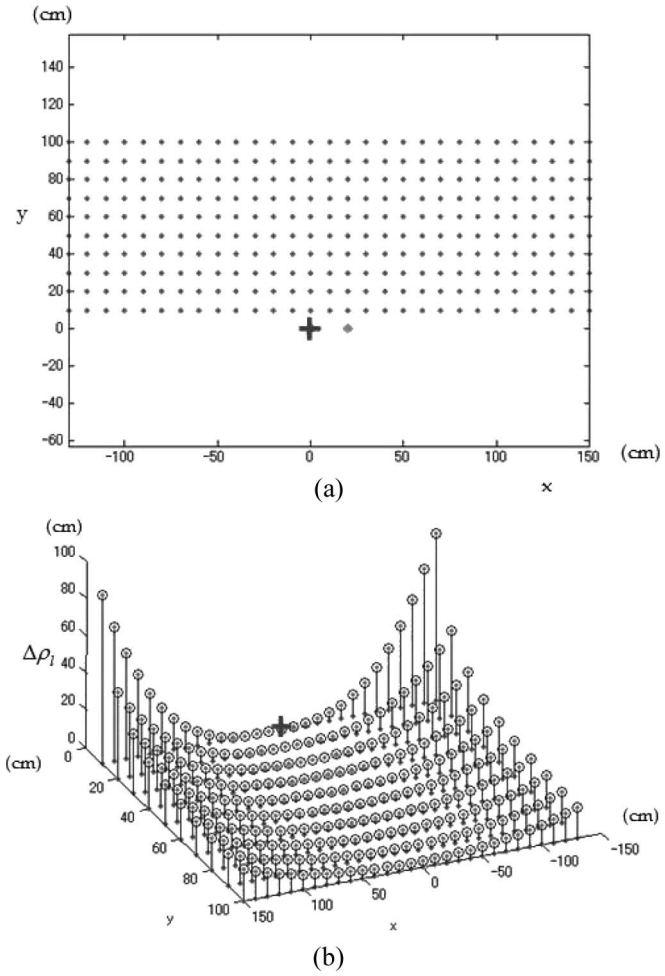


Fig. 12. (a) Group of grid points on half of the epipolar plane. (b) Error $\Delta\rho_l$ in the computed distance at each point.

V. Experiments by Stereo Fish-Eye Cameras

In the previous section, the definition of the disparity of the spherical stereo and the method of computing the 3-D information of the environment from stereo spherical images were presented. Note that this approach is a general approach, because an image with any FOV can be mapped to a spherical image. Therefore, this is an FOV-independent approach. In this section, this method is applied to the fish-eye cameras, which can be mounted at the rear of a vehicle for detecting obstacles, as described in Section I.

A. Calibration of the Spherical Stereo Rig

The remaining task for realizing a binocular spherical stereo is the calibration of a spherical stereo rig, which includes the internal parameters of each fish-eye camera and the external parameters between the cameras. Since a spherical camera may have a full view, it is necessary to use all of the surrounding information for the spherical camera calibration. Thus, methods based on the conventional pinhole camera model cannot directly be applied [19], [21], [22]. Here, a camera with any FOV can be calibrated by using the 3-D calibration pattern based on the spherical camera model. Since the internal parameters relate to a specific imaging device, the present method is explained using a fish-eye camera.

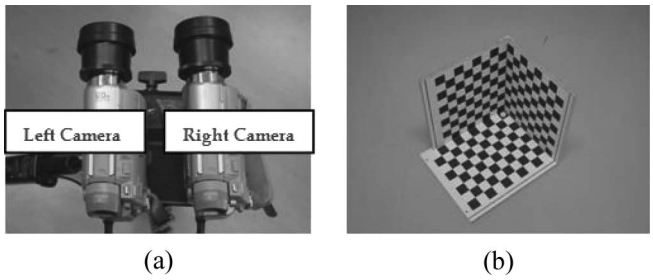


Fig. 13. (a) Spherical stereo rig consisting of two fish-eye cameras. (b) Half-box for camera calibration.

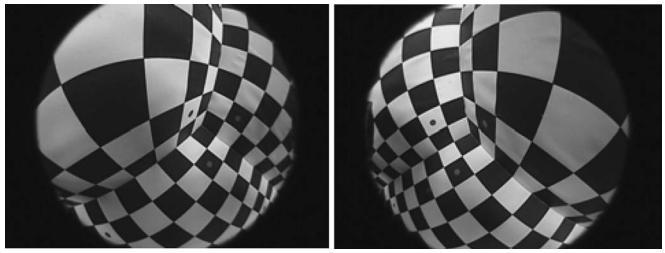


Fig. 14. Images captured by the two fish-eye cameras for the calibration of the spherical stereo rig.

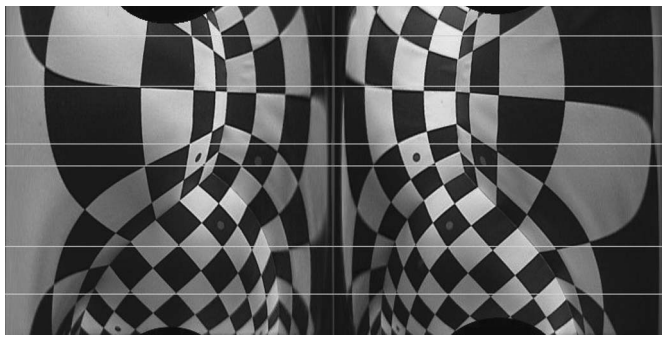


Fig. 15. Pair of latitude–longitude sampling images acquired by using the calibrated internal and external parameters of the spherical stereo rig. The gray horizontal epipolar lines that pass through the same feature points in the pair of images.

In this paper, a pair of video cameras (Sony DCR-HC30) with fish-eye lenses (Olympus Fish-Eye Conversion Lens FCON-02) is used, as shown in Fig. 13(a). The length of the baseline of the spherical stereo rig is approximately 20 cm. Since the fish-eye lens has a maximum FOV of 185°, as shown in Fig. 13(b), a half-box is used to carry out the calibration of the rig.

The fish-eye camera is used to observe the box from the inside to acquire an image covered with the control points for camera calibration, as shown in Fig. 14. The size of the captured fish-eye images is 640 × 480 pixels.

The fish-eye lens used in this paper obeys equidistance projection. The expression for an equidistance projection of fish-eye lens is

$$\begin{bmatrix} r_i \\ \phi_i \end{bmatrix} = \begin{bmatrix} f_f & 0 \\ 0 & 1 \end{bmatrix} \begin{bmatrix} \theta \\ \phi \end{bmatrix} \tag{20}$$

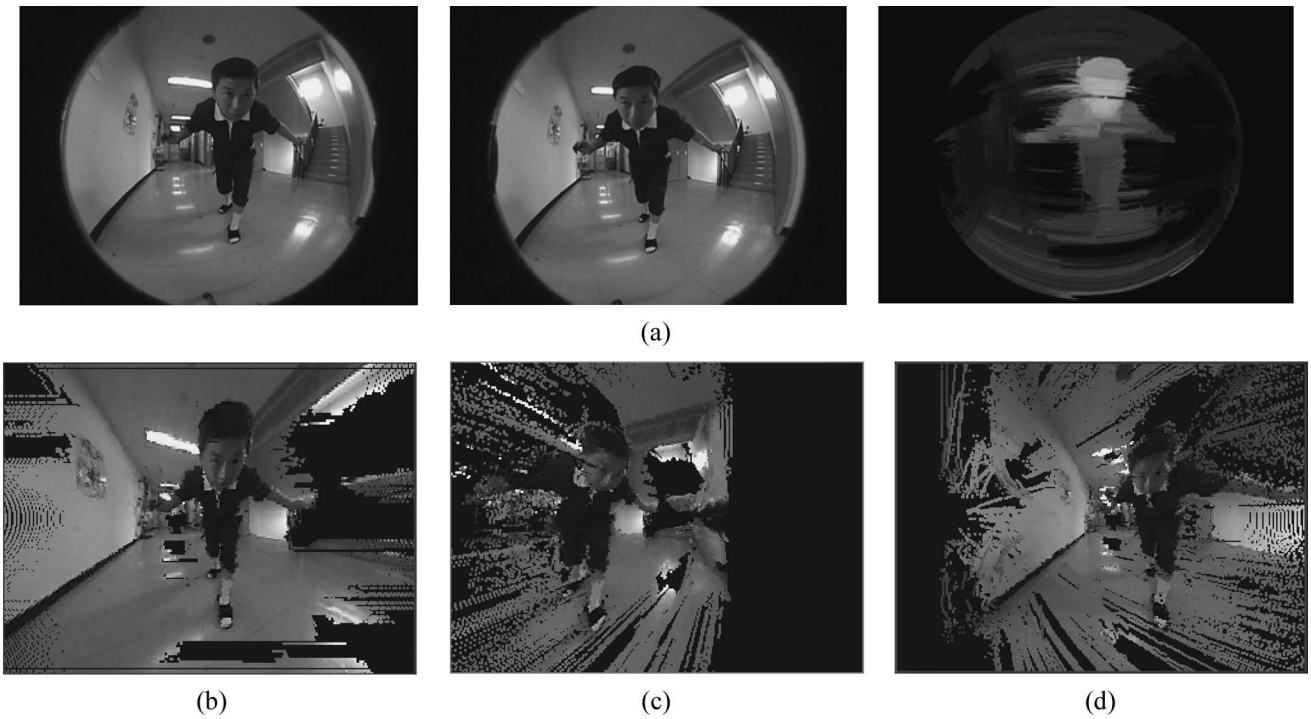


Fig. 16. (a) (Left) Pair of images captured by the binocular spherical stereo rig. (Right) Computed disparity map. (b)–(d) Perspective images projected from the computed 3-D distance at three different viewpoints.

where (r_i, ϕ_i) is the polar coordinate of the projection point in the fish-eye image, (θ, ϕ) are the polar angles of an environment point in the spherical coordinate system of the fish-eye camera, and f_f is the focal length of the fish-eye lens [31], [33]. Suppose that a global coordinate system is set on the calibration half-box. Let the coordinates of a control point on the box be M_w and M_c for the global and spherical image coordinate systems, respectively. Since the coordinate system of the global system can be aligned with the spherical coordinate system by rotation $R = [r_1 \quad r_2 \quad r_3]$ and translation $T = [t_x \quad t_y \quad t_z]^T$, we have

$$M_c = [R \quad T]\widetilde{M}_w. \tag{21}$$

Let the coordinates of the projection of the point on the spherical image be m. In terms of (1), we have

$$m \cong [R \quad T]\widetilde{M}_w. \tag{22}$$

Control point M_w can be mapped to a spherical image via (8) and then mapped onto the captured fish-eye image via (9). This implies that the projection position of a control point in the captured image of the fish-eye can be estimated by the internal and external camera parameters. After detecting the control points in the image, the internal parameters of a fish-eye camera and the external parameters between the fish-eye camera and the half-box can be computed by minimizing the distances for all control points as follows:

$$\sum_{i=1}^{n} d\left(p_i, \Gamma(M_{w_i})\right)^2 \tag{23}$$

where p_i is the detected position in the fish-eye image, and $\Gamma(M_{w_i})$ is the position in the fish-eye image mapped in terms of the camera parameters.

The optical center, focal length, and lens distortions, including the radial and decentering distortions, can then be calibrated. Detailed information of the fish-eye camera calibration can be found in [31].

B. Rectification of Spherical Stereo Images

Using the internal parameters of the fish-eye camera, the fish-eye image can be undistorted [31]. Using the external parameters of the two fish-eye cameras, the spherical image can be rectified to obtain the latitude–longitude image, as shown in Fig. 15, where the epipolar lines of the spherical stereo are parallel to each other, and the conventional correlation-based method for stereo correspondence can directly be applied to acquire the dense 3-D information of the environment.

C. Computation of the Dense Disparity Map

The computation of the dense disparity map for rectified planar stereo images has extensively been studied [10], [11], [32]. As shown in Fig. 15, the epipolar lines are parallel in the latitude–longitude images. Therefore, the conventional correlation-area method can be applied to acquire the dense spherical disparity map. In this paper, the open source code on computer vision OpenCV [35] is used to develop the proposed algorithm. The function used to compute stereo correspondence matching implemented based on [29] is used to compute the

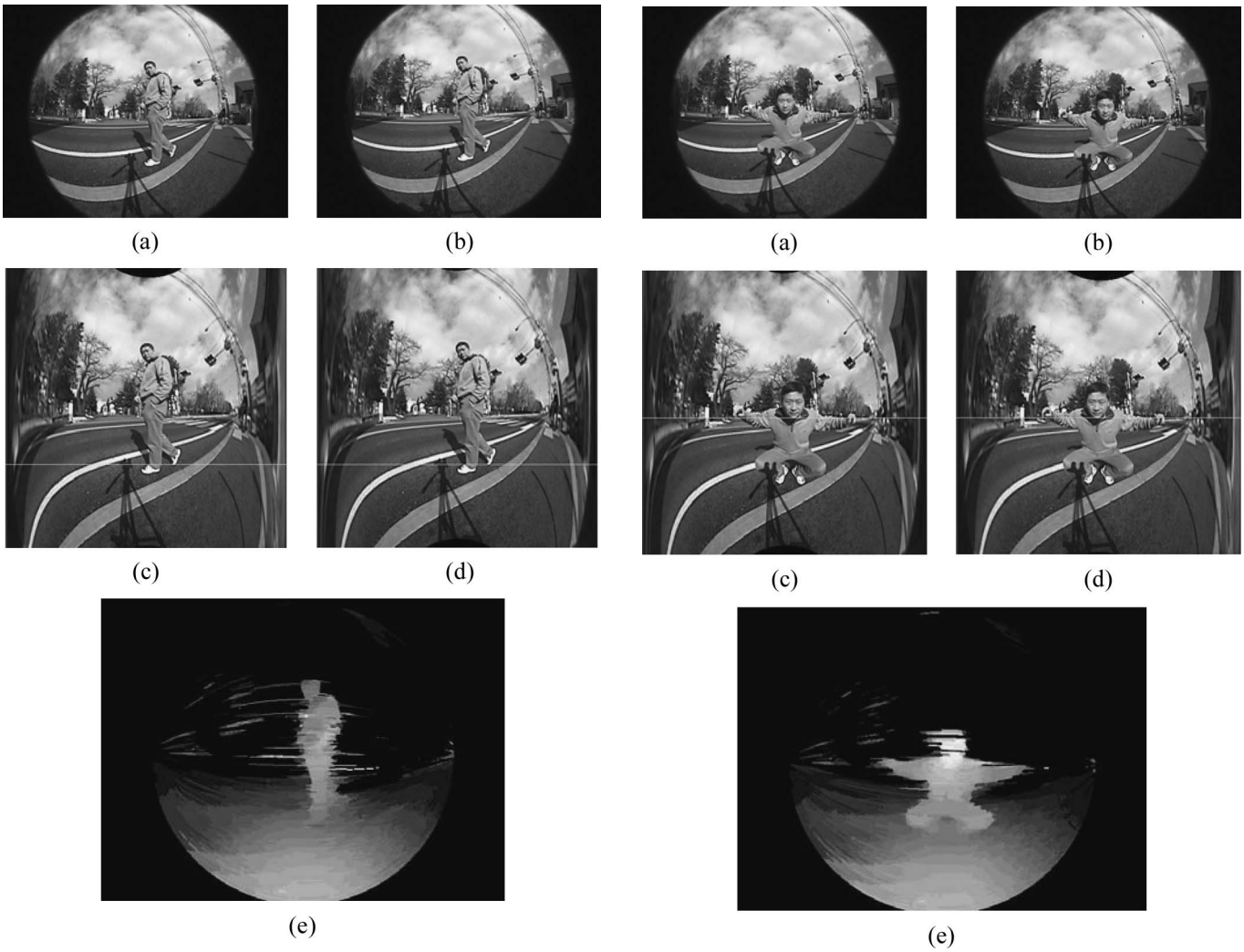


Fig. 17. (a) and (b) Two fish-eye stereo images. (c) and (d) Corresponding latitude–longitude images. (e) Computed disparity image. Brighter pixels correspond to greater spherical disparities, i.e., points closer to the scene.

Fig. 18. (a) and (b) Two fish-eye stereo images. (c) and (d) Corresponding latitude–longitude images. (e) Computed disparity image. Brighter pixels correspond to greater spherical disparities, i.e., points closer to the scene.

dense disparity value. Furthermore, the distance of environment points can be computed using their spherical disparities if necessary.

D. Test for Static Stereo Fish-Eye Images

First, a remote controller of the video camera of the stereo rig is used [shown in Fig. 16(a)] to capture a pair of fish-eye images. (The person is static while the images are captured.) Experiments are conducted for indoor and outdoor environments. Fig. 16 shows the pair of stereo images in an indoor environment. A pair of fish-eye stereo images and the computed dense spherical disparity map are shown in Fig. 16(a) (right). The brighter pixels in the spherical disparity map correspond to the closer distance of the observed environment point. By computing the distance of the space point in terms of (5), integrating with the color cues, and then using the camera parameters, the generated perspective 3-D map samples are shown in Fig. 16(b)–(d). Fig. 16(b) shows the perspective image with the viewpoint at the camera focus. Fig. 16(c) shows the perspective image with the viewpoint on the left of the camera. Fig. 16(d) shows the perspective image on the right of the camera. The black background points correspond to points of unknown distance, for which the correspondence or the occluded regions caused by front objects could not be computed.

More experimental results of outdoor environments are shown in Figs. 17 and 18, where panels (a) and (b) are two fish-eye stereo images, panels (c) and (d) are the corresponding latitude–longitude images, and panel (e) is the computed disparity image. The brighter pixel corresponds to the greater spherical disparity, i.e., the shorter distance between the environment points. As shown by the results, a stooping person can clearly be distinguished from the background, which shows the effectiveness of this approach for an intelligent vehicle.

E. Test for the Image Sequences of Two Fish-Eye Video Cameras

The algorithm is also tested using images directly captured from the two video cameras of the fish-eye camera stereo rig shown in Fig. 13. The two video cameras are linked with a video capture card (IO-DATA CB1394L) via IEEE 1394 cables.

TABLE I
COMPARISON BETWEEN THE CONVENTIONAL PLANAR STEREO AND THE PROPOSED SPHERICAL STEREO

	Binocular Planar Stereo	Binocular Spherical Stereo
Camera Model	Pin-hole camera model	Spherical camera model
FOV	Less than a hemisphere	Any wide FOV image
Definition of Disparity	Difference of coordinates in planar image	Angle subtended by the two cameras
Computed 3D Information	Depth	Distance
Range of Estimated 3D Information	Less than hemispherical FOV	Entire view, except for the 3D blind spot
Application to Fisheye Stereo Cameras with a Wider-than-Hemisphere FOV	The boundary region of the FOV cannot be described by a pin-hole camera model, which results in an inability to acquire the 3D information of the boundary region.	The entire view can be described by a spherical camera model. Thus, the 3D information of the entire view, except for the 3D blind spot centered at the epipoles, can be acquired.

The captured image size is 640 × 480 pixels. A lookup table from an input fish-eye image to a latitude–longitude sampling image is used to speed up the processing. The Sony PC VGN-K704 with a 2.8-GHz Intel Pentium 4 central processing unit has a processing speed of approximately 790 ms/frame. The disparity computation of a pair of latitude–longitude sampling images requires approximately 32 ms/frame, and the remainder is required for image capture, generation of latitude–longitude sampling image, and display of the relative results.

Here, two experimental examples are presented: 1) a person first moving backward and forward (see the video segment *moveBackForh.avi*) and 2) a person moving from left to right (see the video segment *moveLeftRight.avi*). This will be available at http://ieeexplore.ieee.org. The person is clearly detected in the output disparity image in the effective region previously mentioned. Although this preliminary experiment is conducted in an indoor environment, theoretically, this approach should also be able to cope with outdoor dynamic environments to realize an intelligent vehicle.

VI. CONCLUSION

Using two fish-eye cameras to sense the 3-D information around vehicles is useful for safe driving. However, the conventional pinhole camera model cannot be applied to fish-eye image processing, because fish-eye cameras may have a wider-than-hemispherical FOV. In this paper, a binocular spherical stereo approach is proposed to cope with this problem. Two conclusions are given here.

1) In theory, the main contribution of this paper is that the disparity of a spherical stereo is defined and that the binocular stereo problem is reformulated. Based on this definition, the binocular spherical stereo is a general model that can cope with cameras of any FOV, including fish-eye cameras and conventional cameras. By spherical stereo, the 3-D information of the spherical image points, except for the epipoles, can be computed. Even considering the computing error, the 3-D information of the spherical image points, except for the 3-D blind spot centered at the epipoles, can still be acquired. This means that, if two fish-eye stereo cameras are horizontally placed, the 3-D information of any high objects can be computed, even if the corresponding environment points appear at the boundary of the hemispherical FOV, as long as the points are not near the epipoles. Similarly, if the two fish-eye cameras are vertically placed, the 3-D information of any wide objects can be computed, even if the corresponding environment points appear at the boundary of the hemispherical FOV, as long as the points are not near the epipoles. However, if the conventional pinhole camera model is applied to such a fish-eye stereo system, there is a band near the boundary of the hemispherical view for which the 3-D information cannot be acquired, because an infinitely large image is needed for representing a hemispherical view using perspective projection. In Table I, a comparison is made between the conventional planar stereo and the proposed spherical stereo. Although false correspondences may appear along the horizontal direction when two spherical stereo cameras are horizontally placed, just like the conventional binocular stereo system, as shown by the experimental results, the results can further be improved by developing a trinocular spherical stereo system [23]–[27], [36].
2) In practice, the details of applying the binocular spherical stereo to a pair of fish-eye cameras with a hemispherical FOV are given. Using the conventional-pinhole-camera-model planar stereo, the entire boundary area of the FOV cannot effectively be computed, because an infinitely large image is needed to represent the boundary area. On the other hand, using the proposed binocular spherical stereo, only the 3-D information of the partial area near the two epipoles cannot accurately be estimated, as previously mentioned. The proposed method is effective for an intelligent vehicle to detect and monitor the surrounding environment. If hardware is used to generate the latitude–longitude sampling image from an input fish-eye image, there should be no problem in acquiring a processing rate of greater than dozens of frames per second.

The experiments described in this paper are preliminary for applying the proposed approach to a real vehicle. Comparatively evaluating the distance measured by the proposed method with ground truthing using a laser scanner for an outdoor environment, finding obstacles near a vehicle based on computed 3-D information, and testing this approach on a vehicle moving in various environments will be performed in future studies. While a wide-FOV camera observes a wide range of the environment, the resolution of the objects in the image is lower than that when the objects are observed by narrow-FOV cameras with the same charge-coupled device. This results in poorer accuracy of 3-D information computation due to the decrease in image resolution of the environment objects, as

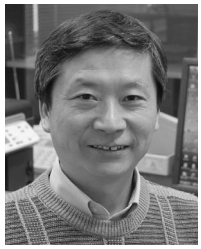
mentioned in the error analysis. However, the accuracy may be improved using higher resolution cameras. Considering the speedy development of digital devices, there may be no need to wait a long time for this approach to be applied to a vehicle.

ACKNOWLEDGMENT

The author would like to thank the anonymous reviewers and Associate Editor for their helpful comments and suggestions and Dr. H. Zhu and Y. Hai for his experiment of the error analysis.

REFERENCES

[1] J. Aloimonos, "Perspective projection," *Image Vis. Comput.*, vol. 8, no. 3, pp. 177–192, 1990.
[2] S. Coorg and S. Teller, "Spherical mosaics with quaternions and dense correlation," *Int. J. Comput. Vis.*, vol. 37, no. 3, pp. 259–273, Jun. 2000.
[3] P. E. Debevec, "Rendering synthetic objects into real scenes: Bridging traditional and image-based graphics with global illumination and high dynamic range photography," in *Proc. SIGGRAPH*, 1998, pp. 189–198.
[4] O. Faugeras, *Three-Dimensional Computer Vision*. Cambridge, MA: MIT Press, 1993.
[5] Y. Yagi, S. Kawato, and S. Tsuji, "Real-time omnidirectional image sensor (COPIS) for vision-guided navigation," *IEEE Trans. Robot. Autom.*, vol. 10, no. 1, pp. 11–22, Feb. 1994.
[6] J. Hong, X. Tan, B. Pinette, R. Weiss, and E. M. Risemani, "Image-based homing," in *Proc. Robot. Autom.*, 1991, pp. 620–625.
[7] B. K. P. Horn, *Robot Vision*. Cambridge, MA: MIT Press, 1986.
[8] K. Ikeuchi, "Recognition of 3-D objects using the extended Gaussian image," in *Proc. Int. Joint Conf. Artif. Intell.*, 1981, vol. 7, pp. 595–600.
[9] H. Ishiguro, M. Yamamoto, and S. Tsuji, "Omni-directional stereo," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 14, no. 2, pp. 257–262, Feb. 1992.
[10] K. Konolige, "Small vision systems: Hardware and implementation," in *Proc. 8th Int. Symp. Robot. Res.*, 1997, pp. 203–212.
[11] L. Mathies, "Stereo vision for planetary rovers: Stochastic modeling to near real-time implementation," *Int. J. Comput. Vis.*, vol. 8, no. 1, pp. 71–91, Jul. 1993.
[12] S. Li, F. Ishizawa, and N. Chiba, "Wandering in VR environments by estimating head pose using an omnicam," in *Proc. 10th Pacific Conf. Comput. Graphics Appl.*, 2002, pp. 318–324.
[13] S. Li, "Estimating head pose based upon sky-ground representation," in *Proc. IEEE/RSJ IROS*, 2005, pp. 1847–1856.
[14] S. Li, "Sky-ground representation for local scene description," in *Proc. ICPR*, 2004, pp. 252–255.
[15] S. Li and K. Fukumori, "Spherical stereo for the construction of immersive VR environment," in *Proc. IEEE VR Conf.*, 2005, pp. 217–222.
[16] A. Makadia, L. Sorgi, and K. Daniilidis, "Rotation estimation from spherical images," in *Proc. ICPR*, 2004, pp. 590–593.
[17] S. Nayar, "Catadioptric omnidirectional camera," in *Proc. Comput. Vis. Pattern Recog.*, 1997, pp. 482–488.
[18] R. C. Nelson, "Finding motion parameters from spherical flow fields," in *Proc. IEEE Workshop Vis. Motion*, 1987, pp. 145–150.
[19] R. Swaminathan and S. K. Nayar, "Nonmetric calibration of wide-angle lenses and polycameras," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 22, no. 10, pp. 1172–1178, Oct. 2000.
[20] R. Yang and M. Pollefeys, "Multi-resolution real-time stereo on commodity graphics hardware," in *Proc. CVPR*, 2003, pp. I-211–I-217.
[21] R. Y. Tsai, "A versatile camera calibration technique for high accuracy 3D machine vision," *Int. J. Robot. Autom.*, vol. 3, no. 4, pp. 323–344, 1987.
[22] Z. Zhang, "Flexible camera calibration by viewing a plane from orientations," in *Proc. ICCV*, 1999, pp. 666–673.
[23] N. Ayache and F. Lustman, "Fast and reliable passive trinocular stereovision," in *Proc. ICCV*, 1987, pp. 422–427.
[24] Y. Kitamura and M. Yachida, "Three dimensional data acquisition by trinocular vision," *Adv. Robot.*, vol. 4, no. 1, pp. 1407–1417, 1991.
[25] Y. Ohta, M. Watanabe, and K. Ikeda, "Improving depth map by right-angled trinocular stereo," in *Proc. 8th ICPR*, 1986, vol. I, pp. 519–521.
[26] M. Okutomi and T. Kanade, "A multiple-baseline stereo," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 14, no. 4, pp. 355–363, Apr. 1993.
[27] J. Mulligan, V. Isler, and K. Daniilidis, "Trinocular stereo: A new algorithm and its evaluation," *Int. J. Comput. Vis.*, vol. 47, no. 1, pp. 51–61, Apr. 2002.
[28] A. Fusiello, E. Trucco, and A. Verri, "A compact algorithm for rectification of stereo pairs," *Mach. Vis. Appl.*, vol. 12, no. 1, pp. 16–22, Jul. 2000.
[29] S. Birchfield and C. Tomasi, "Depth discontinuities by pixel-to-pixel stereo," Stanford Univ., Palo Alto, CA, Tech. Rep. STAN-CS-TR-96-1573, Jul. 1996.
[30] S. Li, "Real-time spherical stereo," in *Proc. ICPR*, 2006, pp. 1046–1049.
[31] S. Li, "Monitoring around a vehicle by a spherical image sensor," *IEEE Trans. Intell. Transp. Syst.*, vol. 7, no. 4, pp. 541–550, Dec. 2006.
[32] W. Mark and D. M. Gavrila, "Real-time dense stereo for intelligent vehicles," *IEEE Trans. Intell. Transp. Syst.*, vol. 7, no. 1, pp. 38–50, Mar. 2006.
[33] K. Miyamoto, "Fish eye lens," *J. Opt. Soc. Amer.*, vol. 54, no. 8, pp. 1060–1061, 1964.
[34] D. W. Henderson and D. Taimina, *Experiencing Geometry: Euclidean and Non-Euclidean With History*. Englewood Cliffs, NJ: Prentice–Hall, 2005.
[35] OpenCV beta 3.1, 2003. [Online]. Available: http://www.sourceforge.net/projects/opencvlibrary
[36] S. Li, "Trinocular spherical stereo," in *Proc. IROS*, 2006, pp. 4786–4791.



Shigang Li received the B.S. degree in electrical engineering from Tsinghua University, Beijing, China, in 1985 and the M.S. and Ph.D. degrees from Osaka University, Osaka, Japan, in 1990 and 1993, respectively.

After receiving the Ph.D. degree, he joined Osaka University, as a Research Associate. In 1995, he joined the Faculty of Information Sciences, Hiroshima City University, Hiroshima, Japan, as an Associate Professor. In 2001, he joined the Faculty of Engineering, Iwate University, Morioka, Japan. Since October 2007, he has been a Professor with the Faculty of Engineering, Tottori University, Tottori, Japan. His research interests include computer/robot vision, intelligent transportation systems, and mixed reality systems.