Sky-ground Representation for Local Scene Description

Shigang Li

Faculty of Engineering Science
Iwate University
4-3-5, Ueda, Morioka, 020-0104, Japan
li@cis.iwate-u.ac.jp

ABSTRACT
This paper proposes a new representation, called a sky-ground representation, for describing scenes of environment at a local place. A sky-ground representation is a spherical image with full field of view, combined with a vertical reference which is determined by sensing the direction of gravity. The scene of environment at a local place is observed by a spherical image sensor. The acquired spherical image is divided into two parts, sky part and ground part, along the horizon according to their positions, above or below the horizon. The horizon is determined by sensing the gravity from an acceleration sensor.

Keywords
Spherical image, gravity reference, scene analysis

1. INTRODUCTION
The space in which humans live can coarsely be divided into the sky and the ground by ignoring some unevenness of terrains. While humans’ activities almost all happen in the ground, the sky rarely changes. This fact implies that the sky part is almost static compared with the highly dynamic ground part as shown in Fig.1. The boundary of the sky and the ground is the horizon which is the intersection of the sky and horizontal planes. If we cut the scene into two parts along the horizon, we can separate it as a highly dynamic scene and an almost static scene. The importance of the both of parts is changeable with tasks. The ground part is necessary for finding a route while the sky part is very useful for selecting landmarks to localize a place. We call this representation a sky-ground representation.

The same thing can be said for a mobile robot when we consider it as humans’ partner sharing space and communicating with us. Fig.2 shows a pair of hemispherical images acquired from our spherical image sensor mounted on a vehicle. For a mobile robot, obstacle-avoiding is the first step to have to do before completing other tasks. That is, the situation of the ground where there is a free space, static object or moving object, must be known. Many kinds of omnidirectional sensors are developed for this purpose. On the other hand, the situation of the sky part is also important if it wants to select some landmarks for its localization. The relation between the ground part and the sky part lets us know that a sky part stands on which ground.

0-7695-2128-2/04 $20.00 (C) 2004 IEEE
ground part) along the horizon which is determined by sensing the gravity from an inertial sensor.

The rest of this paper is organized as follows. In the next section, we describe related research. Section 3 presents our approach. Section 4 presents our preliminary experimental results. Finally, we conclude and present future work in the last section.

2. RELATED RESEARCH

Representation of environment is dependent not only on the task, but also on the used sensors. In order to sense the surrounding of environments, many kinds of omnidirectional image sensors are developed and omnidirectional vision has been widely researched [10][5][7]. However, most of these on mobile robots focus on the analysis of the ground part except some researches use objects standing on the ground for scene description[9]. There are rarely researches on sensing the whole surrounding scene and analyzing the visual cues for different tasks.

Visual cues play an important role in robots’ localization. Since an image sensor moves along with a mobile robot, the incoming scenes are different with the stored ones. The differences may stem from the change of environment for dynamic environments, but also (mainly) from an observation with a displacement and different orientation. If we have a common reference for storing scenes and matching with new scenes, it will enable image matching much easier. It is known that in biological systems the information provided by the vestibular system is fused at a very early stage with vision, playing a key role on the execution of visual movements such as gaze holding and tracking [1]. Using inertial sensor and vision cues for stabilization reflexes, dynamic vision, recovery of object shape and camera motion are reported [2][4][5][6].

Our method stems from a spherical image and is different from the above approaches. First, a spherical image with full field of view is used and it contains visual information which can cope with different tasks. Secondly, since we combine it with a common vertical reference of gravity sensed from an inertial sensor, it makes it easier to localize a robot based upon image matching of local scenes.

3. SKY-GROUND REPRESENTATION

Here we describe how to generate the sky-ground representation. Apart from the computer for data processing, our system consists of a spherical image sensor and an inertial sensor which sense the direction of gravity as a vertical reference as shown in Fig. 3.

3.1. Spherical image sensor

We use a fish-eye conversion lens, Olympus FCON-02, which provides max.185 degree field of view. Two samples of images captured by the fish-eye conversion lens are shown in Fig.2. The fish-eye conversion lens is based on an equidistance projection as shown in Fig.4. The projection equation is given as follows.

\[
\begin{bmatrix}
  r_i \\
  \theta_i
\end{bmatrix} = 
\begin{bmatrix}
  f & 0 \\
  0 & 1
\end{bmatrix} \begin{bmatrix}
  \phi \\
  0
\end{bmatrix}
\]

(1)

where \( r_i \) is the distance of the projection point \( p \) from the optical center and \( f \) is the focal length of the lens.

In order to have the optical center and the focal length, \( f \), we need to calibrate the fish-eye camera.

Figure 4 Equidistance projection of the fish-eye conversion lens used. (a) The projection of a point on a spherical image. (b) Its Corresponding position on the plane image via the fish-eye conversion lens.

In order to have the optical center and the focal length, \( f \), we need to calibrate the fish-eye camera.

Figure 3 The spherical image sensor combined with an inertial sensor.

Figure 5 A white-black stripe patterns for fish-eye camera calibration (on the top-left), an image of the white-black stripe pattern (on the top-right), and the found vanishing points (at the lower level).

Fig.5 shows an image of a white-black stripe pattern. The straight lines of the stripe pattern are parallel and intersect at infinite. The point of intersection is projected onto an image as vanishing points. Since the fish-eye conversion lens has about 185 degree field of view, the vanishing points may appear in
images. Suppose there are two groups of parallel straight lines. They result in two pairs of vanishing points. Thus, the optical center is found as the intersection of the two lines joining the two pairs of vanishing points, respectively. The focal length, \( f \), is determined from the range, \( r_p \), of a pair of vanishing points in term of the following equation.

\[
f = \frac{r_p}{\pi}
\]  \( (2) \)

We use two groups of parallel straight lines which are almost perpendicular to each other as shown in Fig. 6, find their vanishing points respectively, and determine the optical center and focal length as mentioned above. The circle corresponding with the hemispherical view determined from the above estimates is shown in Fig. 6, and its optical center is shown as '+'. Since the fish-eye conversion lens has a 185 degree field of view, there might be overlapping parts of the two opposite views. We further refine the camera parameters based upon image correlation. The detailed description of calibrating a fish-eye camera for acquisition of a spherical image is given in [3].

A sample of spherical image captured from our image sensor is shown in Fig. 10.

### 3.2. Estimating an inclination from inertial sensors

In order to divide a spherical image into a sky part and a ground part, we need to know the horizon. The horizon can be determined easily if we know the gravity vector at the local place. Suppose there are the world coordinate system, \( \text{Ow} \), with its X-axis along the gravity vector and that, \( \text{Os} \), of a spherical image with any orientation as shown in Fig. 7. The orientation of \( \text{Os} \) can be aligned with \( \text{Ow} \) by a rotation transformation. Since any rotation, \( R \), can be expressed as a roll (around Z-axis), \( R_z \), a pitch (around Y-axis), \( R_y \), and a yaw (around X-axis), \( R_x \), \( R = R_z R_y R_x \). We can rectify the coordinate system of a spherical image, \( \text{Os} \), to align its X-axis with the gravity vector if we know the roll and pitch angles.

Let the unit column vectors along \( X_w, Y_w, Z_w \) be \( i_w, j_w, k_w \), respectively. Let the unit column vectors along \( X_s, Y_s, Z_s \) be \( i_s, j_s, k_s \), respectively.

\[
\begin{bmatrix}
i_s & j_s & k_s
\end{bmatrix} = R_z R_y R_x \begin{bmatrix}
i_w & j_w & k_w
\end{bmatrix}
\]

If we know the pitch and roll angles,

\[
R_y^{-1} R_z^{-1} \begin{bmatrix}
i_s & j_s & k_s
\end{bmatrix} = R_x \begin{bmatrix}
i_w & j_w & k_w
\end{bmatrix}
\]

Thus, we can rectify a spherical image to align it with a vertical reference.

In general, it is difficult to determine the orientation of the spherical image sensor by image analysis even if given the correct initial values. It is known that motion estimation from image is noisy and the errors are accumulated. In this paper, we use an inertial sensor. An inertial sensor can measure the inclination caused by the gravity and we can obtain the roll and pitch angles. It means that we can rectify the spherical image sensor to align it with the gravity vector by using an inertial sensor if we know the relative orientation between the acceleration sensor and the spherical image sensor.

### 3.3. Calibrating the orientation of the spherical image sensor relative to the inertial sensor

![Figure 8 An inclination of the spherical image sensor causes a deviation of the vanishing points from the polarities](image)
We can calibrate the orientation of the spherical image sensor relative to the inertial sensor by observing a group of parallel vertical lines which exist a lot in man-made of indoor environment. A group of parallel vertical lines generate a pair of vanishing points. If there is not an inclination for the spherical image sensor, the pair of vanishing points appears at the opposite two polarities. An inclination of the spherical image sensor causes a deviation of the vanishing points from the polarities as shown in Fig.8.

In practice, since the extracted edge from natural environment is very noisy, we put the stripe patterns on a vertical plane and let the stripe parallel to vertical edges of environments. Thus we can obtain clear edges. By fitting the edges with conic curves by the same method as in Fig. 5 [3], we obtain the pair of vanishing points. By using the calibrated camera parameters, we can have the coordinates of the pair of vanishing points on a unit sphere, $V_n$ and $V_s$, corresponding with the north polar and south polar, respectively. Let the roll and pitch angles of the spherical image sensor relative to the vertical reference are $\beta$ and $\gamma$. In terms of (3), we have

$$R^{-1}(\beta)R^{-1}(\gamma)V_n = \begin{bmatrix} 1 & 0 & 0 \end{bmatrix}^T,$$

$$R^{-1}(\beta)R^{-1}(\gamma)V_s = \begin{bmatrix} -1 & 0 & 0 \end{bmatrix}^T$$

(4)

Since there are only tow unknowns, we use the least square method to solve the above equations.

4. EXPERIMENTS

We carried out experiments using our systems. Fig. 9 shows the results of rectifying the spherical images by compensating the inclination measured from the inertial sensor. The three images from the left to the right have rotations with roll, pitch and yaw, (9.668, 3.098, 0.066), (-0.286, 25.159, 5.669) and (13.975, 10.481, -15.798) (measured from the inertial sensor as degrees), respectively.

Fig.10 shows the images captured by our spherical image sensor which is mounted on a moving car. Based upon the horizon, we divided it into a sky part and a ground part. They can be further processed for different tasks.

5. CONCLUSIONS

We propose a new representation, called a sky-ground representation, for describing scenes of environment at a local place. A sky-ground representation is a spherical image with full field of view, combined with a vertical reference which is determined by sensing the direction of gravity. How to use this representation is our future work.

REFERENCES