Pattern Anal Applic (2006) 9: 113–126
DOI 10.1007/s10044-005-0008-4

THEORETICAL ADVANCES

Sanae Shimizu · Kazuhiko Yamamoto · Caihua Wang
Yutaka Satoh · Hideki Tanahashi · Yoshinori Niwa

Moving object detection by mobile Stereo Omni-directional System (SOS) using spherical depth image

Received: 14 January 2005 / Accepted: 29 July 2005 / Published online: 8 November 2005

Abstract Moving object detection with a mobile image sensor is an important task for mobile surveillance systems running in real environments. In this paper, we propose a novel method to effectively solve this problem by using a Stereo Omni-directional System (SOS), which can obtain both color and depth images of the environment in real time with a complete spherical field of view. Taking advantage of the SOS that the frame-out problem never occurs, we develop a method to detect the regions of moving objects stably under arbitrary movement and pose change of the SOS, by using the spherical depth image sequence obtained by the SOS. The method first predicts the depth image for the current time from that obtained at the previous time and the ego-motion of the SOS, and then detects moving objects by comparing the predicted depth image with the actual one obtained at the current time.

Keywords Omni-directional stereo · Spherical image · Mobile surveillance system · Moving object detection · Ego-motion estimation

S. Shimizu (✉) · C. Wang · Y. Niwa
Research and Development Division, Softopia Japan,
4-1-7 Kagano, Ogaki, Gifu 503-8569, Japan
E-mail: shimizu@gifu-irtc.go.jp
Tel.: +81-584-77-1188
Fax: +81-584-77-1106

K. Yamamoto · S. Shimizu
Department of Information Science, Faculty of Engineering,
Gifu University, 1-1 Yanaido, Gifu 501-1193, Japan

Y. Satoh
Information Technology Research Institute,
National Institute of Advanced Industrial Science and Technology (AIST), Central2, 1-1-1 Umezono, Tsukuba, Ibaraki 305-8568, Japan

H. Tanahashi
Information System Division, IMIT-Gifu, 4-179-19 Sue,
Kakamigahara, Gifu 509-0108, Japan

1 Introduction

Detecting moving objects in a dynamic environment by using a mobile sensor is a basic and important task for many applications in computer vision and robotics. For example, an intelligent robot moving automatically in human environments usually needs to detect moving objects for its path planning or for finding human attention candidates for tracking and/or interaction. Similarly, mobile surveillance systems, which are attracting more and more attention recently, also involve the same task, because moving objects usually should be put under closest surveillance. This paper describes the method for moving object detection with a mobile image sensor in a real environment by using a new sensor.

Moving object detections can be classified into two groups: moving object detection with a fixed sensor and that with a mobile sensor. Compared with the former, where the background in the image is static, the image obtained by a mobile sensor contains changes over the entire image not only due to the actual motion of moving objects but also due to the change in the viewpoint of the sensor. The motion of the viewpoint causes changes in the background as well as the change from occlusions causing regions to appear and disappear. It is difficult to detect the change due to the motion of moving objects from these changes. Another difficulty is the frame-out problem, that is, the objects in attention may move out of the field of view (FOV) of the sensor due to the movements of the objects and/or the sensor. Therefore, moving object detection with a mobile sensor has been considered to be a hard problem until now.

In the past, both texture image-based approaches [1–8] and depth image-based approaches [9–18] have been explored to solve this problem. Generally, the texture-based methods suffer from appearance variations of the objects, the background and the occlusion when both the objects and sensor are moving at the same time. In contrast, since the depth image is the geometrical

information of the scene, the methods based on depth image can estimate the change of the view and occlusion due to the movement of the viewpoint and are more robust to the appearance of objects and the changes of the illumination. Particularly, depth image with a wide FOV is very effective for mobile surveillance systems because it is more stable regarding the frame-out problem caused by the movement of the sensor.

Along the second approach, several methods [9–12] have been proposed for moving object detection, by attempting to use depth image with a wide FOV. A representative approach of those methods is to use a panoramic laser range finder. Prassler et al. [9] describes a method which detects moving objects by calculating the difference between the current and the previous obstacle positions on a grid map. Since they estimate the robot position only by dead reckoning, the accumulated positional error of the robot may degrade the obstacle map, and therefore may result in detecting static obstacles as moving ones. Lu et al. [11] adopts a similar approach, whereas the range data is also used for ego-motion estimation to reduce the estimation error of dead reckoning. However, because those methods use line-scanning laser range finders, they can only detect objects which are intersected by the scanning plane at a fixed hight and may miss objects present only above or below the scanning plane.

Recently, Koyasu et al. [12] proposed a method for moving object detection using the omni-directional stereo system composed of a pair of omni-directional cameras, whose wide vertical FOV ensures more stability in moving object detection. The method first projects the panoramic depth image, which can be obtained by matching the stereo omni-directional images to the horizontal plane, and then generated the free space map by storing the nearest obstacles. The moving objects are detected by comparing the current observation with the map. The ego-motion of the sensor is estimated using the map to eliminate the error calculation in dead reckoning.

The purpose of the above methods is obstacle avoidance, so they only detect the moving objects closest to the sensor, and cannot detect moving objects behind some static obstacles. This may not be desirable for some applications such as mobile surveillance systems. Another problem is that although the above methods adopt sensors with a perfect horizontal FOV, they still have a limitation in the vertical FOV. This means that when the sensor is slanted, which often happens when it runs in a real environment, the changes in the visible FOV of the sensor may cause frame-out of the objects being observed, or yield a blind spot in the FOV of concern to the observer.

In this paper, we use the spherical depth image obtained by a mobile Stereo Omni-directional System (SOS) [21], which was developed by us, for moving object detection [24, 25]. The SOS has a complete spherical FOV and is able to capture high-resolution color and depth images of the entire surrounding environment simultaneously in real time. We propose a Motion compensatory inter-frame depth subtraction method to detect moving objects robustly with the mobile SOS, by using the depth image which is more to the changes of the illumination and the view due to the motion of the sensor and integrating the advantages of the complete spherical FOV, ability to maintain a stable FOV without blind spots during movement and pose changes of the SOS and ability to estimate ego-motion robustly to the occlusion by using the abundance of information by the spherical FOV.

Our method tries to detect the regions of all the moving objects even if they are partially occluded. We detect the regions rather than feature points of the moving object to achieve detection stability and reliability, because regions are generally considered to be more resistant to noise than feature points.

Our method adopts a similar approach as the frame difference method, except that we use the depth image instead of the texture. Since both the background and the regions of the moving objects will change when the SOS moves, the frame difference method cannot be applied directly. To deal with this problem, we first predict the depth image for the current time from the ego-motion of the SOS and the depth image obtained at the previous time, and then detect the moving objects by comparing the predicted depth image with the actual one obtained at the current time. The occlusion regions in the predicted depth image caused by the change of viewpoint are also estimated and distinguished from that of the moving objects. A similar method [19], which detects moving objects by predicting the texture image from ego-motion parameters and previous depth image, has been proposed, but the predicted texture comes under the influence of the change of the illumination and the error of ego-motion and depth map. We use the depth image directly which is more robust under these influences. Furthermore, we detect robustly despite the frame-out problem and an occlusion by using the spherical depth image obtained by the SOS.

2 Stereo Omni-directional System (SOS)

The SOS [21], provides both color and depth images of the surrounding environment in real time with a FOV of 360°×180° and high resolution. Fig. 1 and Table 1 show the prototype and the specification of the SOS, respectively.

The SOS is composed of twelve stereo units, each one of which is mounted on the surface of a dodecahedron. Each stereo unit consists of three cameras, among which the left and central cameras form a horizontal stereo pair while the top and central cameras do a vertical pair. In order to preserve the necessary precision of the depth image obtained by stereo matching, sufficient length of baseline between the cameras of the stereo pairs must be kept. However, longer baselines will result in a larger

Fig. 1 The SOS prototype

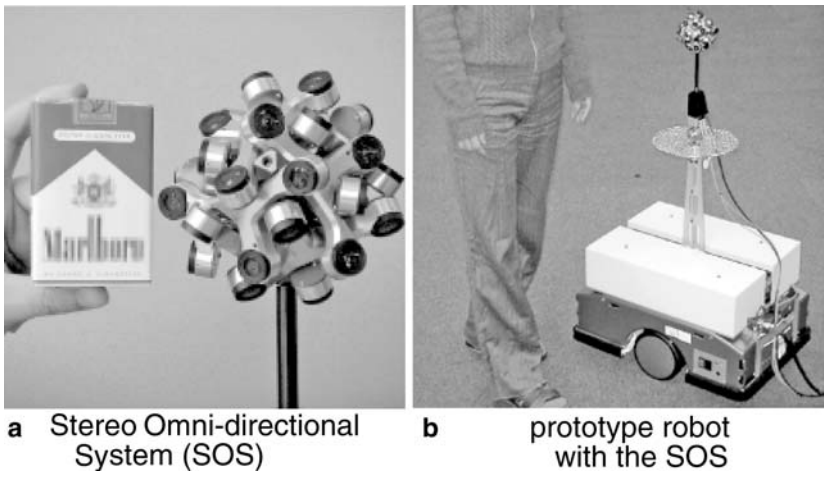


Table 1 The SOS specification

Shape	Dodecahedron
Image sensor	1/4' CMOS image sensor
Effective resolution	640×480
Focal length	1.9 mm
Field of view	101°×76°
Base line length	50 mm
Frame rate	15 fps
Diameter of camerahead	11.6 cm
Weight	615 g
Electric power consumption	9W (15V, 0.6A)

system. For our research, we selected a baseline length of 50 mm, with which the depth image can be obtained with satisfactory accuracy for the purposes of navigation, obstacle detection and moving object detection.

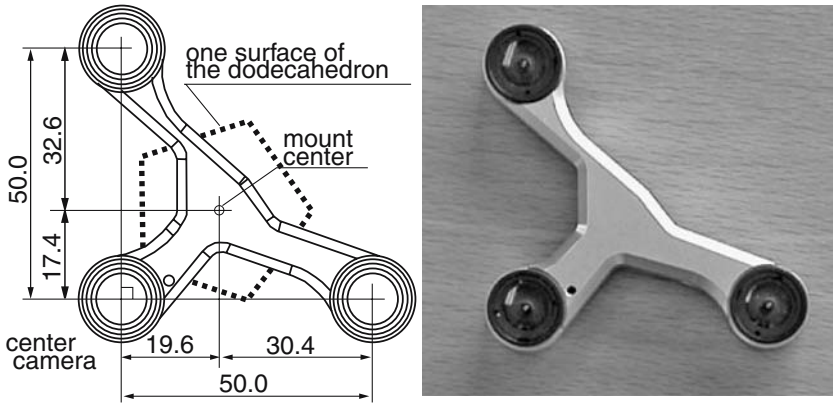
A total of twelve stereo units, each on of which is mounted on a face of the dodecahedron, cover a complete spherical FOV of the system with uniform high resolution. One may simply mount the stereo units to a dodecahedron which is large enough to contain one stereo unit in each face, however, the size of the overall system will become very large. In order to solve this problem, we first laid out three cameras in a stereo unit resembling a T-shaped arm (Fig. 2) and then mounted the stereo units to the dodecahedron whose faces are much smaller than the size of each stereo unit, nesting the arms so that the camera planes of neighboring stereo units cross with each other but such that their FOV are not obstructed by each other. As a result, we succeeded in reducing the size of the entire system (diameter: 11.6 cm; weight 615 g) while securing the baseline length (50.0 mm).

Here, three cameras of each stereo unit are in the same plane and their optical axes are parallel. The right and top cameras are laid out parallel to x and y axes of the image plane of the central camera, respectively. Therefore, simple horizontal and vertical epipolar constraints can be applied to reduce stereo matching cost. Furthermore, using two stereo pairs simultaneously has the benefits that we can obtain more accurate and reliable depth information from the matching results in two stereo pairs, and can get denser depth information using the complementary relation in two stereo matching results. Fig. 3 shows the images obtained by a 3-camera stereo unit.

We can integrate twelve color and depth images by projecting them to an expanded cylindrical image of a sphere. Figures 4 and 5 show examples of the spherical color and depth image obtained by the SOS, respectively.

Fig. 6 shows the system diagram of the SOS. Images obtained by the camera head are output over two fiber

Fig. 2 A stereo unit



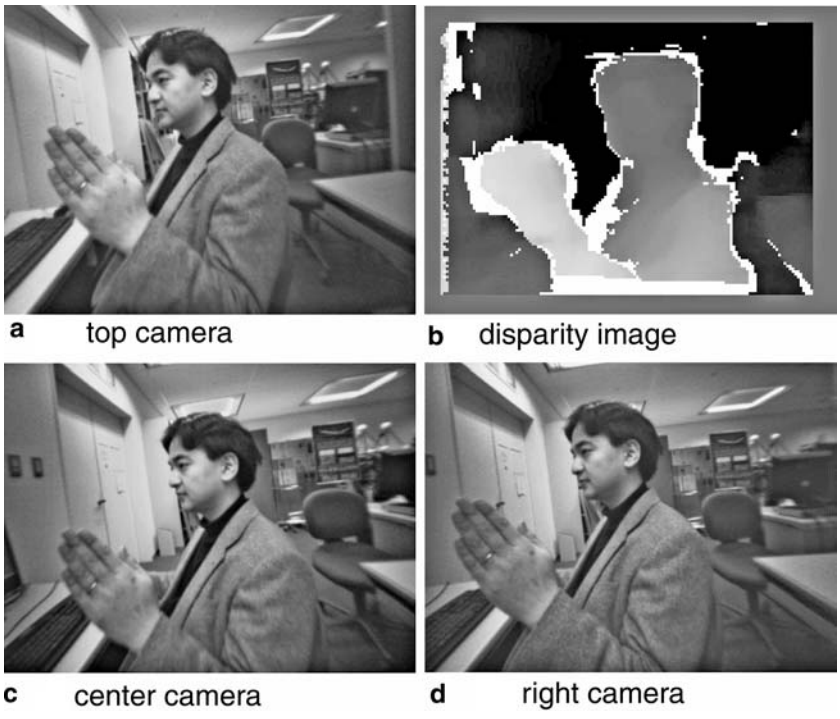


Fig. 3 An example of images obtained from a stereo unit

Fig. 4 Spherical color image

Fig. 5 Spherical depth image

cables, each transmitting data at 1.2 Gb/s. A memory unit and a control unit are implemented on a PCI board, so camera control and image capture can be operated by just one PC.

3 Motion compensatory inter-frame depth subtraction

3.1 Overview

Fig. 7 shows the overview of our method. First, we estimate the relative ego-motion parameters of the SOS between the previous time ($T=t$) and the current time ($T=t-\Delta t$) by using the spherical texture and depth image. Next, we generate a depth image for time t from the ego-motion parameters and depth image obtained at time $t-\Delta t$, when we assume that the scene is static. Finally, we compare the actual depth image obtained at the current time with the predicted depth image of the previous time, which is generated under a static condition, and detect the inconsistent regions as moving objects. Meanwhile, the occluded regions caused by the ego-motion of the SOS are estimated and used to reduce the influence of the occlusion. Here, we present pixels, points, surfaces, coordinates, axes and matrices with uppercase letters and values and vectors with lowercase letters.

3.2 Estimating ego-motion of the SOS

3.2.1 Slant estimation

The problem of estimating the slant of the SOS is to find the direction of the vertical axis (Z-axis) of the SOS in the world coordinate system. This problem is equivalent to the problem of finding the slant of the vertical-axis Z_w of the world coordinate system in the coordinate system

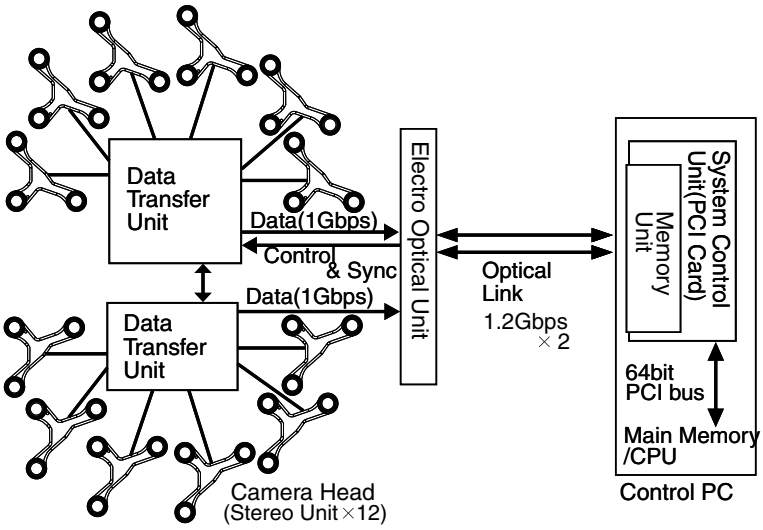


Fig. 6 System construction

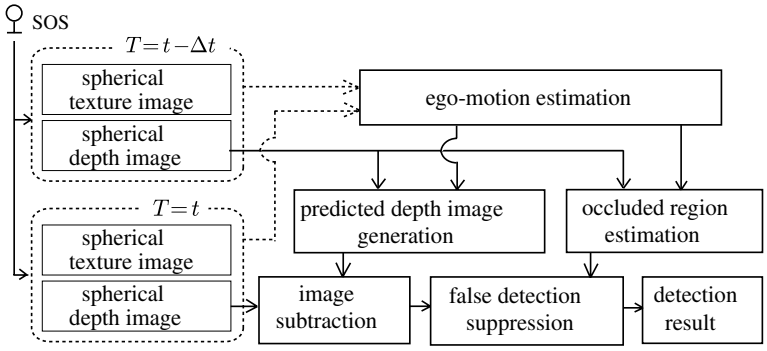


Fig. 7 Overview of Motion compensatory inter-frame depth subtraction.

of the SOS. As the direction of Z_w has the same direction as the vertical edges in the 3-dimensional (3D) space, we estimate the slant by finding the peak of the distribution of the vertical edge directions.

Suppose that a 3D vertical edge point P_w and its direction $\vec{e_w} = (X_{e_w} Y_{e_w} Z_{e_w})^\top$ in the world coordinate system are acquired as P_s and $\vec{e_s} = (X_{e_s} Y_{e_s} Z_{e_s})^\top$, respectively, in the coordinate system of the SOS. In the 3D space, since the vertical edge $P_s + \lambda \vec{e_s}$ and Z_w are parallel and they are coplanar, $P_s, \vec{e_s}$ and Z_w satisfy the relation $(Z_w \times P_s)^\top \vec{e_s} = (P_s \times \vec{e_s})^\top Z_w = 0$. As shown in Fig. 8, this relation shows Z_w is on a plane of $(XYZ)(P_s \times \vec{e_s}) = 0$. Obviously, all the homogeneous planes composed of the vertical edges pass through the intersection point $P_z = (x,y,1)$ of Z_w and $Z = 1$ plane. Generally, the vertical edges in the scene form a major edge group and they have a large distribution in the projecting plane. Therefore, the direction of Z_w can be estimate by detecting the peak $(x_{Pz}, y_{Pz}, 1)$ at p from the projecting plane. Then the yaw α and the pitch β are calculated from the peak $(x_{Pz}, y_{Pz}, 1)$ [22].

3.2.2 Estimating rotation and translation

We estimate the horizontal rotation ϕ and translation direction ω of the sensor after the slant recovery. In order to estimate these parameters, the approach [20] using a number of feature points for stable estimation in real time has been explored. However, the difficulties of feature selection and matching still exist. In order to estimate robustly without dealing with these problems, we use an edge histogram. An edge histogram is more robust to disturbance than feature points and discrete edges because an edge histogram is a kind of statistical feature. Moreover, the spherical images obtained by the SOS are global information of the entire surrounding environment and make estimation more robust to occlusions. First, we generate a panoramic edge image

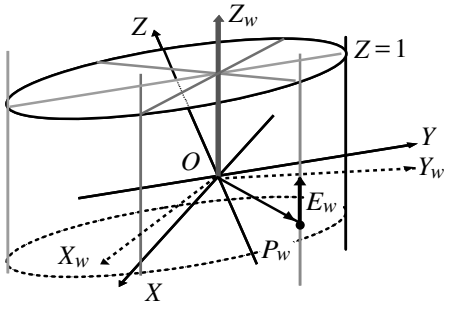


Fig. 8 Relation of vertical edges and slant of the SOS

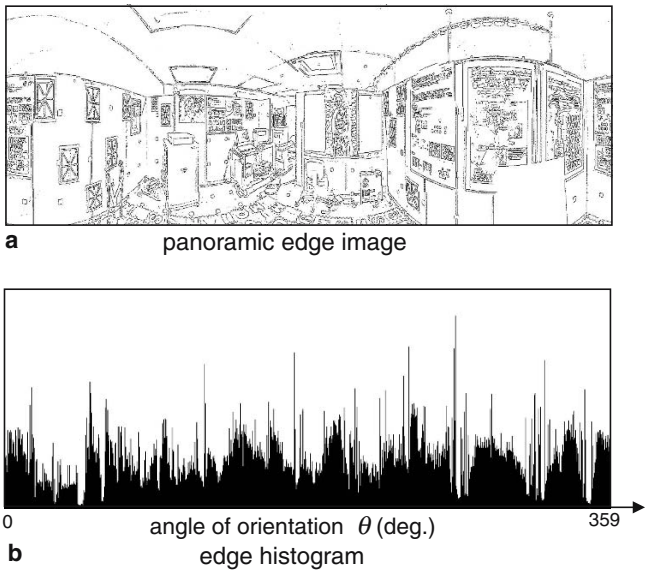


Fig. 9 An example of panoramic edge image and its edge histogram

from the spherical images with the slant recovered, as shown in Fig. 9a, and then generate an edge histogram by projecting each edge pixel vertically (Fig. 9b). As shown in Fig. 10, when the SOS moves in the direction ω from the position at time $t-\Delta t$, the edge histogram shift is small in direction $\omega \pm n\pi$, $n=0,1$, and large in direction $\omega \pm \frac{(2n+1)}{2}\pi, n=0,1$. The shift has the same characteristic as a sin curve, where the rotation of the SOS corresponds to the offset and translation direction of the SOS does to the zero phase. Then, the rotation and the translation direction are estimated by a sin curve fitting the shifts in the edge histograms, which can be found by matching the edge histograms at time $t-\Delta t$ and t [23].

3.2.3 Estimating movement distance

We estimate the altitude and translation distance using the spherical depth image. The altitude distance l_v is estimated by matching the horizontally projected depth histograms at time $t-\Delta t$ and time t.

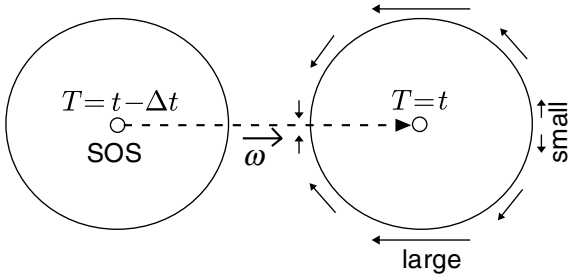


Fig. 10 Edge histogram shift by the SOS movement

Next, we estimate the translation distance l_h of the SOS in the translation direction ω. We generate the depth histogram by orthogonally projecting the spherical 3D points obtained at time $t-\Delta t$. Fig. 11a shows the depth histogram, we can see that parts of wall are presented as linear peaks. The distance l_h in the direction ω is estimated by the shifts of those linear peaks on the histograms at time t and $t-\Delta t$. We find the major surface groups by detecting the direction of the liner peaks in the depth histogram using the Hough transform, and calculate the normal vector $\vec{v_1}, \vec{v_2}$ to the major surface groups. Then, we find the nearest normal vector to the direction ω ($\vec{v_1}$ in Fig. 11a) and generate one-dimensional depth histograms at time t and $t-\Delta t$ (Fig. 11c and b). Finally, l_h is acquired by matching these one-dimensional histograms at time t and $t-\Delta t$.

3.3 Omni-directional depth image prediction

After the ego-motion of the SOS is estimated, we can generate a predicted spherical depth image for the current viewpoint from that observed at the previous viewpoint, if we assume that the scene is static. In our research, the predicated spherical depth image is generated for the ordinary pose of the SOS where its slant is recovered, because the detection results represented in an ordinary pose will be convenient for the follow-up processing. This is possible because the SOS has a complete spherical FOV and its visible range of FOV never changes at all in any slant.

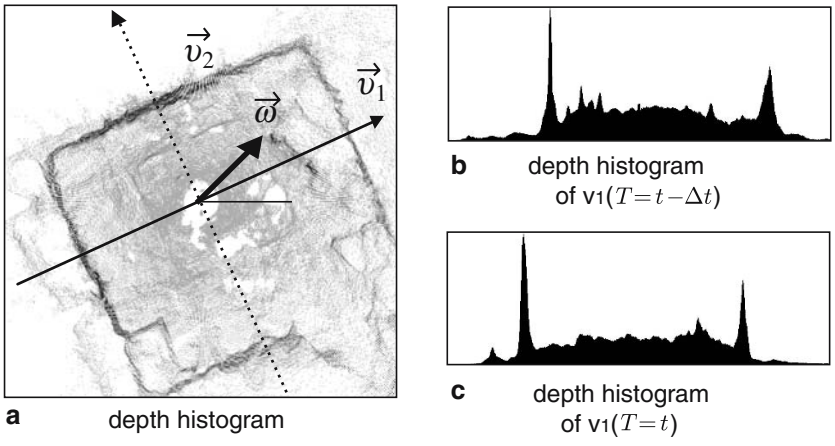


Fig. 11 An example of depth histograms

Let α and β be the yaw and pitch angles of the slant of the SOS respectively, the rotation matrix involving the slant of the SOS can be represented as follows.

$$R_{\alpha,\beta} = \begin{pmatrix} \cos\beta & 0 & \sin\beta \\ 0 & 1 & 0 \\ -\sin\beta & 0 & \cos\beta \end{pmatrix} \begin{pmatrix} 1 & 0 & 0 \\ 0 & \cos\alpha & \sin\alpha \\ 0 & -\sin\alpha & \cos\alpha \end{pmatrix} \quad (1)$$

Next, we transform each 3D point (x, y, z) in the depth image obtained at the previous viewpoint into the 3D point (x_p, y_p, z_p) in the current viewpoint, using the slant parameters (α, β), orientation parameter ϕ, and translation parameters (ω, l_h, l_v). The transformation can be represented as follows.

$$\begin{pmatrix} x_p \\ y_p \\ z_p \end{pmatrix} = \begin{pmatrix} \cos\phi & -\sin\phi & 0 \\ \sin\phi & \cos\phi & 0 \\ 0 & 0 & 1 \end{pmatrix} R_{\alpha,\beta} \begin{pmatrix} x \\ y \\ z \end{pmatrix} - \begin{pmatrix} l_h\cos\omega \\ l_h\sin\omega \\ l_v \end{pmatrix} \quad (2)$$

The predicted spherical depth image $d_p(\theta,\gamma)$ for the current viewpoint then is generated by mapping the transformed 3D points (x_p,y_p,z_p) to a spherical coordinate system. Here we use the cylindrical expansion of the sphere, and represent it by the orientation angle θ $(0<\theta<2\pi)$ and elevation angle γ $(-\pi/2<\gamma<\pi/2)$. The pixel $d_p(\theta,\gamma)$ in the spherical predicated depth image stands for the distance of the scene from the sensor in the view angle of (θ,γ).

3.4 Occluded region estimation

We estimate the occluded regions that have occurred due to the motion of the sensor. The occlusion is related to the sensor motion and the position of the objects in the environment in relation to the sensor. As shown in Fig. 12, when the SOS moves from the position at time $t-\Delta t$ to the position at time t, each pixel $P=(\theta,\gamma)$ on the spherical image at the position at time $t-\Delta t$ moves to pixel P' at the position at time t along vector $\vec{v}$, which is the direction of the intersection line of the tangent plane S_1 at pixel P on the spherical surface and the plane S_2 including pixel P and the translation direction vector $\vec{\omega} = (\omega, atan(l_v/l_h))$. As shown in Fig. 13, $\vec{v}$ can be represented as $\vec{v} = \vec{n_1} \times (\vec{n_1} \times \vec{\omega})$, where $\vec{n_1}$ is the normal vector to the surface S_1.

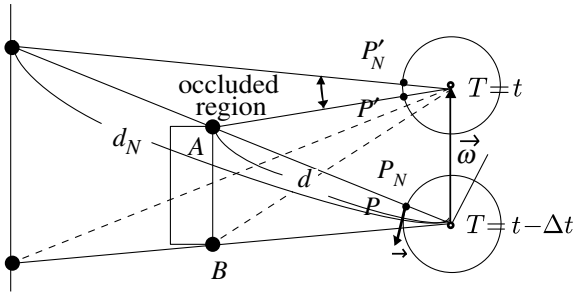


Fig. 12 Calculation of occluded regions

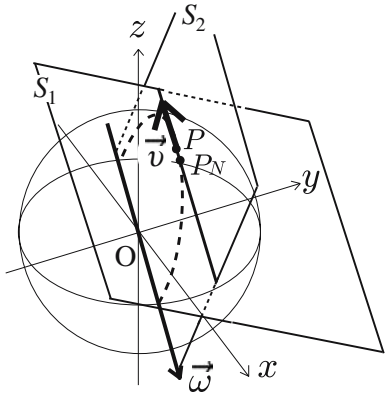


Fig. 13 Occlusion direction

Let d be the depth of pixel P and d_N be that of P_N which is adjacent to pixel P in direction $-\vec{v}$. The occluded region is estimated as the region between pixel P' and P'_N at the jump edge which satisfy $d_N - d > t_d$ (A in Fig. 12). t_d is a threshold to the difference of the depth, and is set to $t_d = 15$ cm in the experiment.

On the other hand, although a jump edge which satisfies $d_N - d < -t_d$ exists at position B in Fig. 12, such a jump edge doesn't need to be considered because it is overlapped by the front object in the predicted depth image.

3.5 Extracting Moving objects

In order to detect moving objects, the spherical depth image actually obtained at the current viewpoint is first transformed into the ordinary pose where the slant of the SOS has been recovered. Let α' and β' be the slant parameters of the SOS, the 3D points (x',y',z') obtained at the current viewpoint can be transformed to the coordinate system in the ordinary pose without slant as follows.

$$\begin{pmatrix} x_s \\ y_s \\ z_s \end{pmatrix} = R_{\alpha',\beta'} \begin{pmatrix} x' \\ y' \\ z' \end{pmatrix} \quad (3)$$

where $R_{\alpha',\beta'}$ in formula (3) is defined similarly as that in formula (1). Note that the pure rotation does not cause occlusion, so the occlusion region estimation procedure in Sect. 3.4 is not necessary here.

Similar to what was described in Sect. 3.3, the spherical depth image with the slant recovered can be generated by mapping 3D points (x_s,y_s,z_s) to a spherical coordinate system (θ,γ) as $d_s(\theta,\gamma)$. Then, we subtract the predicated image $d_p(\theta,\gamma)$ from the actual one $d_s(\theta,\gamma)$ of the current viewpoint, and get a differential image $\delta(\theta,\gamma) = d_s(\theta,\gamma) - d_p(\theta,\gamma)$. In the differential image $\delta(\theta,\gamma)$, the regions of the moving objects approaching to the

SOS satisfy $\delta(\theta,\gamma) > t_\delta$ and that receding from the SOS do $\delta(\theta,\gamma) < -t_\delta$.

As the above detection results contain the occluded regions, we remove the occluded regions estimated in the Sect. 3.4 from the detection results. Moreover, we detect only the regions of moving objects by using an area filter, which measures each area of the regions and removes the regions whose areas are less than a threshold. It is difficult to distinguish the regions of moving objects and noise by a constant threshold because when the motion of moving objects is small, the overlapped regions of moving objects in the predicted and the current images are large and the region in the difference image is small, and the areas of objects in the image changes based on the distance from the SOS to the objects. In order to support in these cases, we define the threshold S_{th} of the area filter as follows.

$$S_{\text{th}} = \frac{2}{\text{res}^2}\arctan\left(\frac{l_m}{2d_m}\right)\left(\arctan\left(\frac{h - h_{\text{sos}}}{d_m}\right) + \arctan\left(\frac{h_{\text{sos}}}{d_m}\right)\right) \tag{4}$$

where l_m $(0 < l_m < w)$ is the minimum moving distance of moving objects, d_m is the distance from the SOS to the objects, res is the resolution to the orientation and elevation angle, and w and h are the width and the height of moving objects. h_{sos}, which is the height of the SOS from the floor at time t, was estimated by the horizontally projected depth histogram (Sect. 3.2.3) at time t. Here, we assume that a moving object is a person and moves on the floor. In the experiments in the next section, we set $l_m = 30$ cm, $d_m < 300$ cm, res $= 2.0$, $w = 30$ cm and $h = 160$ cm in order to detect the entire regions of moving objects (persons).

4 Experimental results

We tested our method in a real environment. Fig. 14a and b show the spherical texture and depth images obtained by the SOS with unknown slant at time $t-\Delta t$ and t, respectively. Three moving objects (persons) appeared in the scene, and are denoted as A, B, and C sequentially from the left of Fig. 14a. The positions and movements of the SOS and the moving objects are shown in Fig. 15, where A is moving along the same line as the SOS is moving, and B and C are moving parallel and aslant to the path of the SOS, respectively. The ego-motion parameters of the SOS obtained in Sect. 3.2 are $\alpha = 0.0$ rad, $\beta = 0.0$ rad, $\alpha' = 0.15$ rad, $\beta' = 0.19$ rad, $l_h = 29$ cm, $\omega = 1.73$ rad, $\phi = 0.54$ rad, $l_v = 14$ cm and $h_{\text{sos}} = 136$ cm. Fig. 14c and d are the depth images that are recovered from Fig. 14a and b respectively using the estimated slants of the SOS.

We can see that the slant can be estimated and recovered correctly and recovered image does not have the loss of the blind spot because the SOS has a spherical FOV.

Fig. 14e shows the predicted depth image for time t using the ego-motion parameters and the depth image of time $t-\Delta t$. Fig. 14g shows the subtraction results of the depth images in Fig. 14e and d, where the positive values are shown in white and the negative ones in gray. In this figure, the gray regions show the candidates of the objects that disappeared from time $t-\Delta t$ to time t, and white regions show the candidates of the objects that appeared. The regions of occlusions due to the movement of the sensor were estimated using the method described in Sect. 3.4 (Fig. 14f), and were used to suppress the false detections in Fig. 14g. After the area filter is used after this process for noise reduction, the final detection result is obtained and shown in Fig. 14h. We can see that the regions of the persons were detected suitably despite the sensor motion including slant. In particular, our method successfully detected person A moving in the same line as the SOS, which is a difficult case for optical flow-based methods.

Fig. 16 plots the depth values of one horizontal line in the central part of Fig. 14b, d, e. As in this graph, the predicted and the actual obtained depth values are consistent in the regions where no moving object exists. On the other hand, there are large differences in the regions where moving objects exist. This graph shows that there is enough information to detect moving objects.

Fig. 17 shows another experimental result in the case when a person is partially occluded by an obstacle in front of him, as shown in Fig. 18. The images obtained by the SOS with unknown slant are shown in Fig. 17a and b. Ego-motion parameters of the SOS were estimated as $\alpha = -0.15$ rad, $\beta = 0.09$ rad, $\alpha' = -0.05$ rad, $\beta' = -0.14$ rad, $l_h = 13$ cm, $\omega = 1.93$ rad, $\phi = -0.41$ rad, $l_v = 4$ cm and $h_{\text{sos}} = 122$ cm.

In this experiment, the person E in the middle of Fig. 17a was walking to the occluded area of a static obstacle. Fig. 17c shows the detection result, from which we can see that person E was detected even though the object was partially occluded.

Next, we carried out the experiment in the case when the area of the occluded regions is large. The position and movement of the SOS, a person and a static obstacle as shown in Fig. 20. The images obtained by the SOS are shown in Fig. 19a and b. Ego-motion parameters of the SOS were estimated as $\alpha = 0.0$ rad, $\beta = 0.0$ rad, $\alpha' = 0.0$ rad, $\beta' = 0.0$ rad, $l_h = 42$ cm, $\omega = 0.49$ rad, $\phi = -0.07$ rad, $l_v = 0$ cm and $h_{sos} = 150$ cm. In this experiment, since the position of the static obstacle is near the SOS and the SOS moves in the direction of the jump edge on the static obstacle in the depth image at time $t-\Delta t$, there large occluded regions near the static obstacle as shown in Fig. 19e. Although the area of the occluded regions is large, ego-motion parameters can be estimated stably by using a large amount of information obtained by the SOS. Although the area of the occluded regions is larger than the region of the moving object, the moving object can be detected by distinguishing the occluded regions.

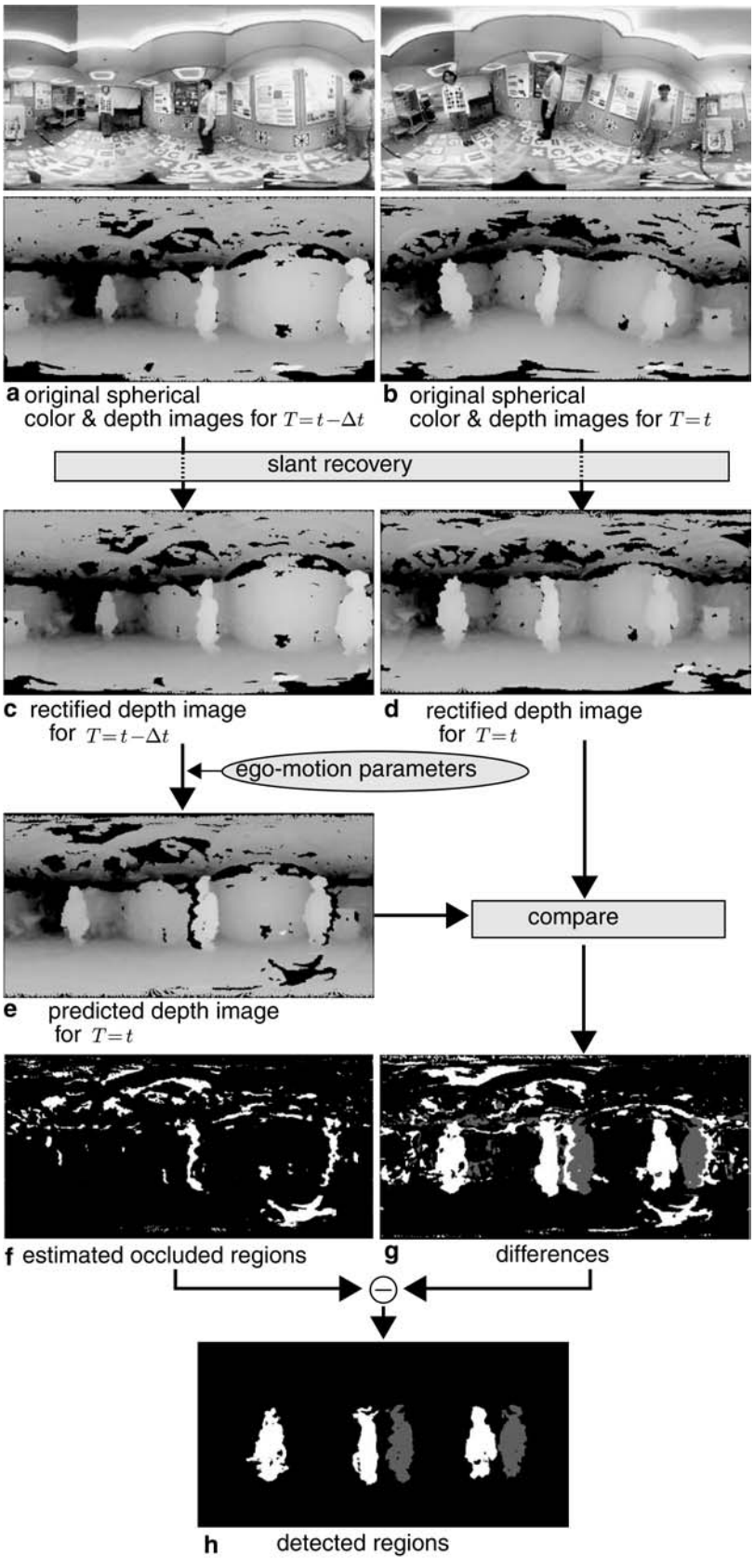


Fig. 14 Experimental results

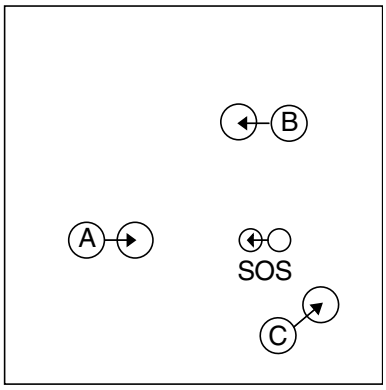


Fig. 15 Experimental condition

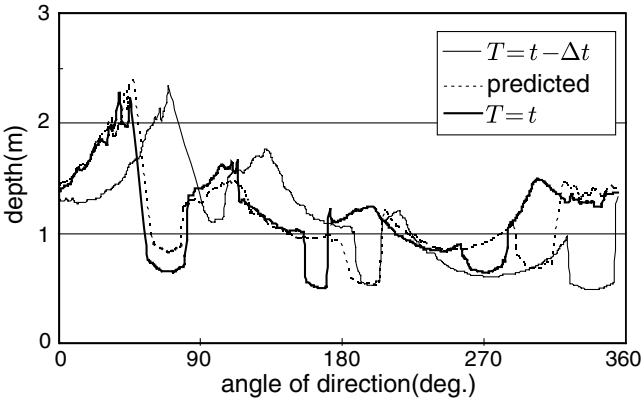


Fig. 16 Profiles of depth images for time $t-\Delta t$ and time t, and the predicted depth image

This method works well in the case when ego-motion parameter estimation is successful and the motion of moving objects is large enough to detect. First, since we estimate ego-motion by matching of edge histogram based on the edge amount in each direction, this method can not work well for two reasons. One of them is that the edge histograms of two frames will appear totally different due to the large motion of the sensor. Another is the area of the occluded regions which occur due to the motion of the sensor. The former is not a problem in the situation which we are dealing with, because we use global information of the spherical image obtained by the SOS. As for the latter problem, stable estimation is generally difficult when the area of occluded regions is large within the total FOV. However, with an occlusion of the same size, since the ratio of the occluded regions in the spherical FOV of the SOS is smaller than that in the FOV of an ordinary camera, the spherical FOV makes more robust estimation. The occluded regions on the images obtained by a unit (FOV:101°×76°) and the SOS which has the spherical FOV are shown as in Fig. 21 and

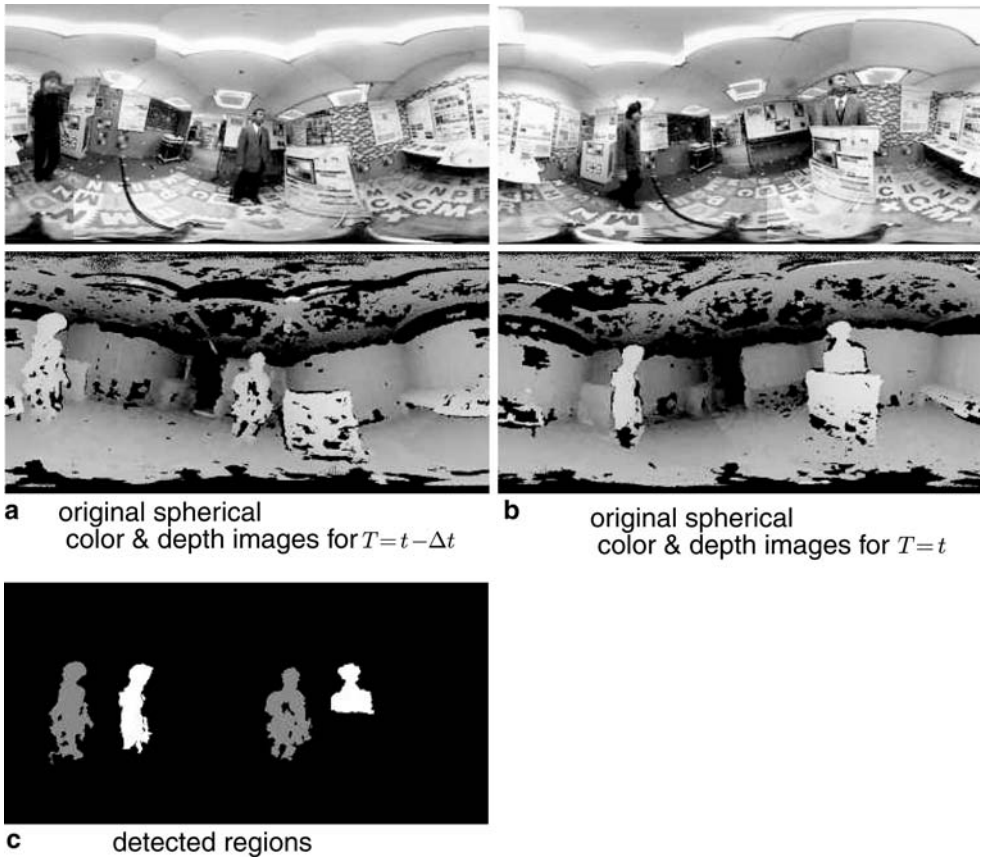


Fig. 17 Experimental results in the case when a person is partially occluded by an obstacle

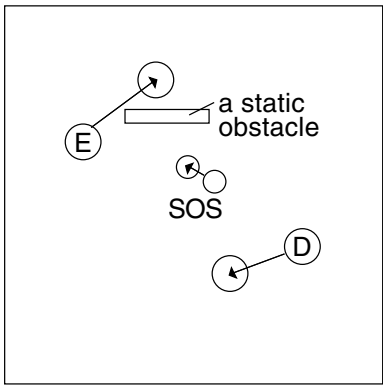


Fig. 18 Experimental condition in the case when a person is partially occluded by an obstacle

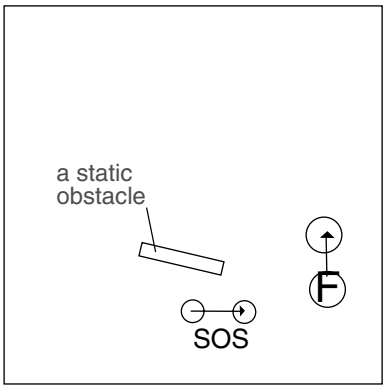


Fig. 20 Experimental condition in the case when the area of the occluded regions is large

Fig. 19 Experimental results in the case when the area of the occluded regions is large

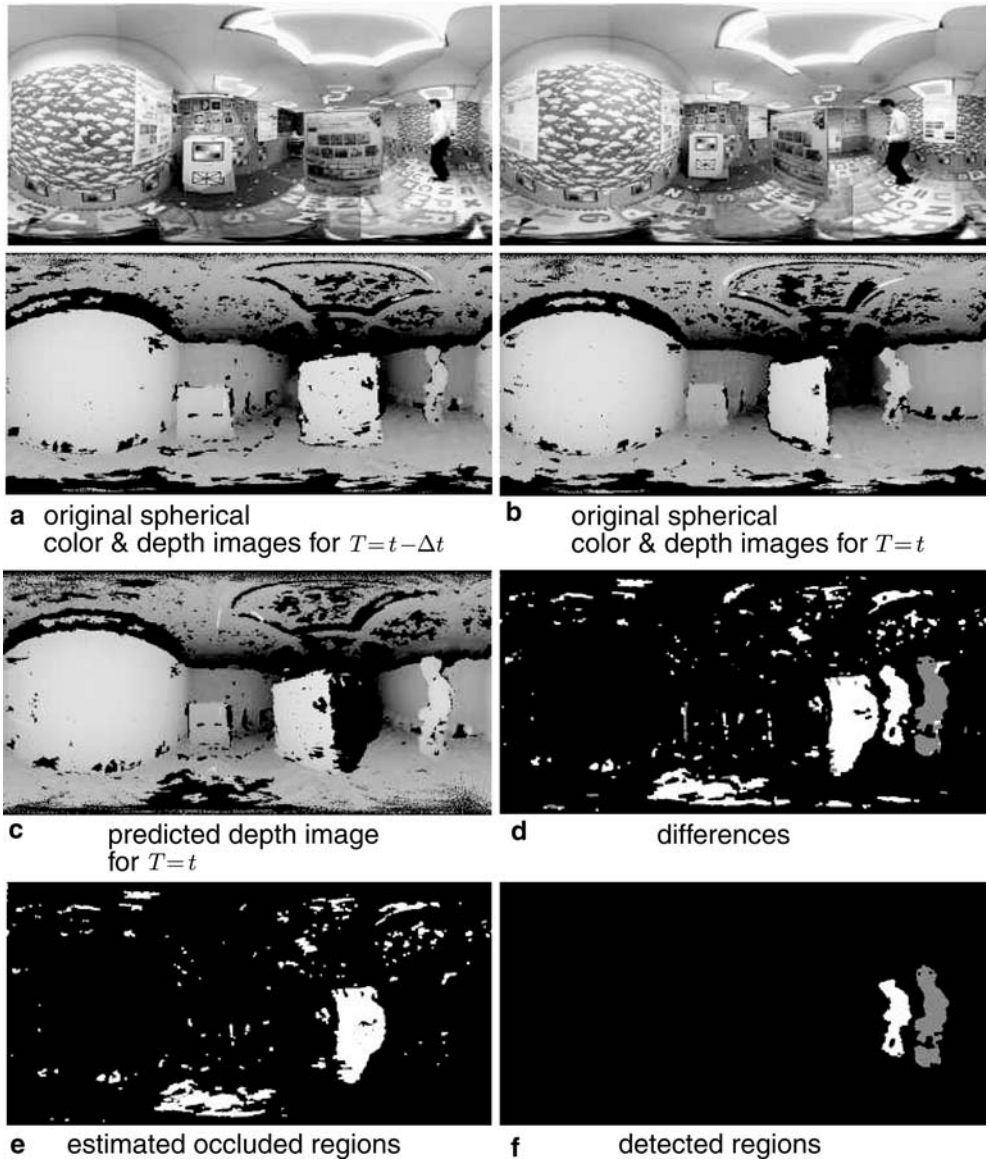


Fig. 19e. In the experimental situation, our method using the SOS can estimate stably, which is a difficult case for methods using an ordinary camera. The minimum moving distance l_m of moving objects is defined in Sect. 3.5. Here, we set $l_m = 30$ cm and detect moving object which moves more than 30 cm.

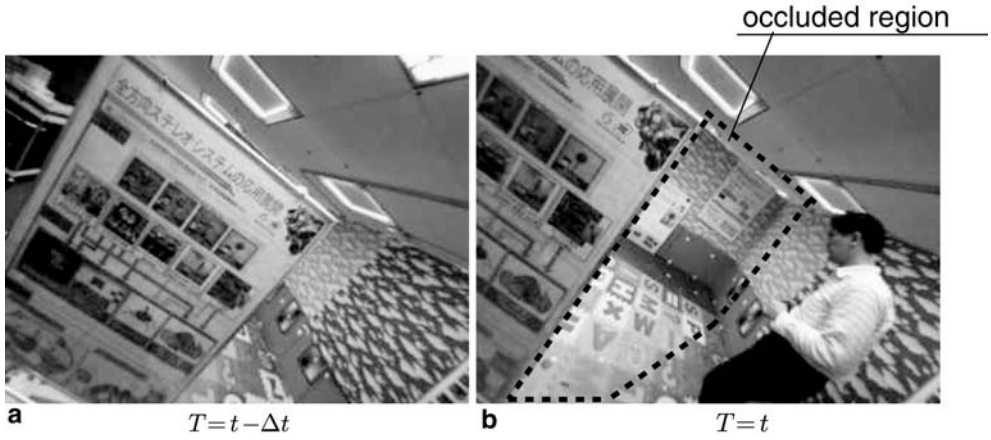


Fig. 21 The images obtained by one of twelve units of the SOS in the case when the area of the occluded regions is large

Detection capability of this method depends on the accuracy of depth information. The error of depth is theoretically proportional to the square of the depth values of the objects and also dependent on the stereo corresponding errors. Although it cannot be stated precisely, in practice, the margin of error tends to be approximately 6–7cm in such a case. This shows a human whose width usually is more than 30 cm can be detected, even if that human moves along a wall.

5 Conclusions

In this paper, we proposed a novel method called Motion compensatory inter-frame depth subtraction to solve the problem of moving object detection with a mobile sensor, by using the spherical depth images obtained by the SOS, whose complete spherical FOV allows us to obtain stable information of the surrounding environment regardless of its pose. Based on the characteristic of the SOS, our method could deal with arbitrary motion of the sensor by estimating its ego-motion parameters with 6 degrees of freedom. The moving object detection was carried out by first predicting a depth image for the current time from that obtained at the previous time and the ego-motion of the sensor and then comparing it with that actually obtained at the current time. Our method also estimated the occluded regions caused by the motion of the sensor to reduce their influences. The region based approach adopted in our method allowed us to detect moving objects more stably and reliably under noise and partial occlusions than using feature points. Experimental results in real environments showed the effectiveness of our method.

The quantitative evaluation and comparison with feature point based technique, tracking moving objects and analysis attributes of moving object regions are our future work.

6 Originarity and contributions

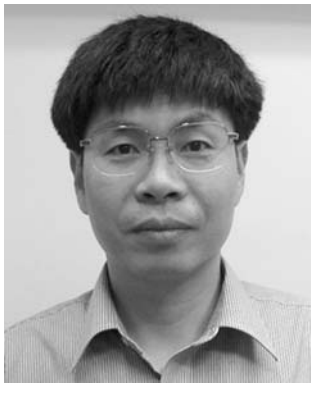
We explored a new approach of using depth images with a complete spherical field of view, which can be obtained by the SOS developed by us, for moving objects detection while the sensor is also moving. We dealt with the most general form of this problem by taking arbitrary motion of the sensor including its slant into account, taking advantage of the SOS that no frame-out problem exists. A novel method called Motion compensatory inter-frame depth subtraction was also proposed in this paper to detect the regions of moving objects stably and reliably even given noise and partial occlusions. The proposed method can be applied to mobile surveillance systems, where moving object detection with sensor movement is usually a critical task for detecting and tracking intruders, as well as for intelligent robots moving automatically in human environments for path planning or for finding human attention candidates for tracking and/or interaction.

7 About the authors

Sanae Shimizu received her B.S. and M.S. degree in science from Nagoya University in 2000 and 2002. She is now in a Ph.D. candidate in engineering at Gifu University. She has been with Gifu Prefecture Research Institute of Manufactural Information Technology in 2002. She was a researcher of the HOIP project at Softopia Japan Foundation from 2002 to 2004. She is a researcher of the R&D Department of Softopia Japan Foundatin. Her interests are computer vision and image processing.

Kazuhiko Yamamoto received his B.E., M.E. and Ph.D. degree in engineering from Tokyo Denki University in 1969, 1971 and 1983 respectively. From 1971 to 1995, he was with the Electrotechnical Laboratory (ETL). He was the head of image understanding section of ETL from 1986 to 1995. From 1979 to 1980, he was a visiting researcher at the Computer Vision Laboratory, University of Maryland. Since 1995, he has been a professor in the Department of Information Science, Faculty of Engineering at Gifu University. His research interests are pattern recognition and artificial intelligece. Dr.Yamamoto is a fellow of IAPR and IEICE, and a member of IEEE Computer Society.

Caihua Wang received his B.S. in mathematics and M.E. in computer science from Renmin University of China, Beijing, China in 1983 and 1986 respectively, and his Ph.D. from Shizuoka University, Hamamatsu, Japan in 1996. He had done his post doctoral research at Electrotechnical Laboratory from 1996 to 1999 as a JST domestic research fellow. He was a senior researcher of the HOIP project at Softopia Japan Foundation from 1999 to 2004. He is currently working in Fuji Photo Film Co., Ltd. His research interests include computer vision and image processing. Dr. Wang is a member of the IPSJ.



Yutaka Satoh received his B.E. and M.E. degrees in Engineering from Tokyo University of Agriculture and Technology in 1996 and 1998 respectively, and the Ph.D. degree in Engineering from Hokkaido University in 2001. He worked as a senior researcher in the HOIP project at Softopia Japan Foundation from 2001 to 2004. He is currently working in National Institute of Advanced Industrial Science and Technology (AIST), Japan. His research interests include machine vision and pattern recognition. He is a member of JSPE and ITE.

Hideki Tanahashi received his B.E. and Ph.D. degrees in engineering from Gifu University in 1985 and 1998 respectively. He has been with Gifu Prefecture Research Institute of Industrial Technology since 1985. He was a senior researcher of the HOIP project at Softopia Japan Foundation from 1999 to 2002. He has been a senior researcher of Gifu Prefecture Research Institute of Manufactural Information Technology. His research interests include computer vision, especially 3D shape reconstruction from images. Dr. Tanahashi is a member of the IEICE.

Yoshinori Niwa received his B.E. from Nagoya University in 1974. He had been with the Metallurgy Research Institute of Gifu Prefecture since 1978. He was the director of the R&D Department of Softopia Japan Foundation and the director of the HOIP project. He is a Chief Information Officer of Softopia Japan Foundation. His research interests include image recognition and CAD/CAM. He is member of the IEEE, IPSJ and ITE of Japan.

References

1. Murray D, Basu A (1994) Motion tracking with an active camera. IEEE Trans Pattern Anal Mach Intell 16(5):449–459
2. Nair D, Aggarwal J (1994) Detecting unexpected moving obstacles that appear in the path of navigation robot. In: Proceedings of the IEEE international conference on image processing, pp 311–315
3. Odobez J, Bouthemy P (1994) Detection of multiple moving objects using multiscale MRF with camera motion compensation. In: Proceedings of the IEEE international conference on image processing Vol 2, pp 257–261
4. Iraniand M, Rousso B, Peleg S (1992) Detecting and tracking multiple moving objects using temporal integration. In: Proceedings of European conference on computer vision, pp 282–287
5. Araki S, Matsuoka T, Yokoya N, Takemura H (1999) Real-time tracking of multiple moving object contours in a moving Camera Image Sequence. Syst Comput Jpn 30(9):25–33
6. Frazier J, Nevatia R (1992) Detecting moving objects from moving platform. In: Proceedings of the IEEE international conference on robotics and automation, pp 1627–1633
7. Nagai A, Kuno Y, Shirai Y (1999) Detection of moving objects against a changing background. Syst Comput Jpn 30(11):107–116
8. Watanabe M, Takeda N, Onoguchi K (1996) A Moving object recognition method by optical flow anaysis. In: Proceedings of the international conference on pattern recognition, pp 528–533
9. Prassler E, Scholz J (2000) Tracking multiple moving objects for real-time robot navigation. Autonomous Robots 8(2):105–116
10. Lindstrom M, Eklundh J-O (2001) Detecting and tracking moving objects from a mobile platform using a laser range scanner. In: Proceedings of the international conference on intelligent robots and systems, pp 1364–1369
11. Lu F, Milios EE (1994) Robot pose estimation in unknown environments by matching 2d range scans. In: Proceedings of the international conference on pattern recognition, pp 935–938
12. Koyasu H, Miura J, Shirai Y (2002) Recognizing moving obstacles for robot navigation using real-time omnidirectional stereo vision. Robotics Mecatron 14(2):147–156
13. Okada K, Kagami S, Inaba M, Inoue H (2001) Walking human avoidance and detection from a mobile robot using 3d depth flow. In: Proceedings of the IEEE international conference on robotics and automation, pp 2307–2312
14. Beymer D, Konolige K (1999) Real-time tracking of multiple people using continuous detection. In: Proceedings of the IEEE international conference on computer vision
15. Davis L, Philomin V, Duraiswami R (2000) Tracking humans from a moving platform. In: proceedings of the international conference on pattern recognition Vol 4, pp 171–178
16. Fod A, Howard A, Mataric MJ (2002) A laser-based people tracker. In: Proceedings of the IEEE international conference on robotics and automation, pp 3024–3029
17. Kluge B, Kohler C, Prassler E (2001) Fast and robust tracking of multiple objects with a laser range finder. In: Proceedings of the IEEE international conference on robotics and automation, pp 1683–1688
18. Reid DB (1979) An algorithm for tracking multiple targets. IEEE Trans Automatic Control 24(6):843–854
19. Bjorkman M, Eklundh J-O (2000) A Real-time system for epipolar geometry and ego-motion estimation. In: Proceedings of the IEEE computer vision and pattern recognition Vol 2, pp 506–513
20. Zhang T, Tomasi C (1999) Fast, robust, and consistent camera motion estimation. In: Proceedings of the IEEE computer vision and pattern recognition Vol 1, pp 164–170
21. Tanahashi H, Shimada D, Yamamoto K, Niwa Y (2001) Acquisition of three-dimensional information in a real environment by using the Stereo Omni-directional System (SOS). In: Proceedings of the IEEE third international conference on 3-d digital imaging and modeling, pp 365–371
22. Wang C, Tanahashi H, Satoh Y, Hirayu H, Sato J, Niwa Y, Yamamoto K (2003) Slant estimation for active vision using edge directions in omni-directional images. In: Proceedings of the IEEE international conference on image processing, pp 841–844
23. Wang C, Tanahashi H, Satoh Y, Hirayu H, Niwa Y, Yamamoto K (2004) Location and pose estimation for active vision using panoramic edge histograms. Syst Comput Jpn 35(12):32–43
24. Shimizu S, Yamamoto K, Wang C, Satoh Y, Tanahashi H, Niwa Y (2004) Moving object detection using depth information obtained by mobile Stereo Omni-directional System (SOS). In: Proceedings of the asian conference on computer vision, pp 336–341
25. Shimizu S, Yamamoto K, Wang C, Satoh Y, Tanahashi H, Niwa Y (2004) Moving object detection with mobile Stereo Omni-directional System (SOS) based on motion compensatory inter-frame depth subtraction. In: Proceedings of the international conference on pattern recognition Vol 3, pp 248–251