

3D Environment Measurement Using Binocular Stereo and Motion Stereo by Mobile Robot with Omnidirectional Stereo Camera

Shinichi GOTO

Department of Mechanical Engineering
Shizuoka University
3-5-1 Johoku, Naka-ku, Hamamatsu-shi,
Shizuoka 432-8561, Japan
f0030021@ipc.shizuoka.ac.jp

Ryosuke Kawanishi Toru Kaneko
Department of Mechanical Engineering
Shizuoka University
3-5-1 Johoku, Naka-ku, Hamamatsu-shi,
Shizuoka 432-8561, Japan
{f5945015,tmtkane}@ipc.shizuoka.ac.jp

Atsushi Yamashita

Department of Precision Engineering
The University of Tokyo
7-3-1 Hongo, Bunkyo-ku,
Tokyo 113-8656, Japan
yamashita@ieee.org

Hajime Asama

Department of Precision Engineering
The University of Tokyo
7-3-1 Hongo, Bunkyo-ku,
Tokyo 113-8656, Japan
asama@robot.t.u-tokyo.ac.jp

Abstract

Map information is important for path planning and self-localization when mobile robots accomplish autonomous tasks. In unknown environments, they should generate environment maps by themselves. An omnidirectional camera is effective for environment measurement, because it has a wide field of view. There are binocular stereo and motion stereo in traditional methods for measurement by omnidirectional camera. However, each method has advantages and disadvantages. In this paper, we aim to improve measurement accuracy by integrating binocular stereo and motion stereo using two omnidirectional cameras installed on a mobile robot. In addition, stereo matching accuracy is improved by considering omnidirectional image distortion. Experimental results show the effectiveness of our proposed method.

1. Introduction

In recent years, it is expected to introduce autonomous mobile robots in various environments with robot technology development. There are disaster sites, the nuclear reactor inside, and so on, in these environments. They are often unknown environments. Therefore, mobile robots should measure the environments to generate the maps by themselves.

3D measurement using image data makes it possible to generate environment maps [1]. However, an image ac-



Figure 1. The omnidirectional camera which we use is shown in the left figure (a). The camera attaches a hyperboloid mirror in front of the camera lens. An omnidirectional image is shown in the right figure (b). The image has a 360-degree horizontal field of view.

quired by a conventional camera has a limited field of view. To solve this problem, cameras with a wide field of view have been developed.

Cameras with a wide field of view include an omnidirectional camera using a hyperboloid mirror [2, 3]. Taking account of installation on a mobile robot, an omnidirectional camera is suitable because it can get a surrounding view image at once. Gluckman [4] showed that an omnidirectional camera is effective for environment measurement. Therefore, we use omnidirectional cameras (Fig. 1(a)) for environmental measurement. The omnidirectional camera is equipped with a hyperboloid mirror. There are binocular stereo [5,6] and motion stereo [7] for measurement by using omnidirectional camera.

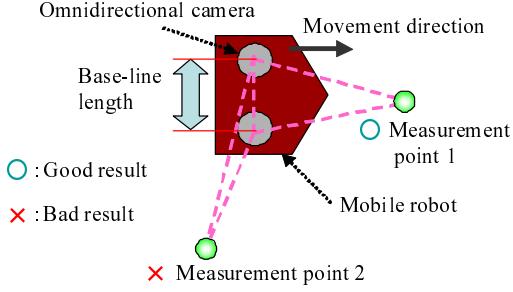


Figure 2. Advantages and disadvantages of binocular stereo measurement. Point 1 can be measured with high accuracy. However, it is difficult to measure point 2 with high accuracy.

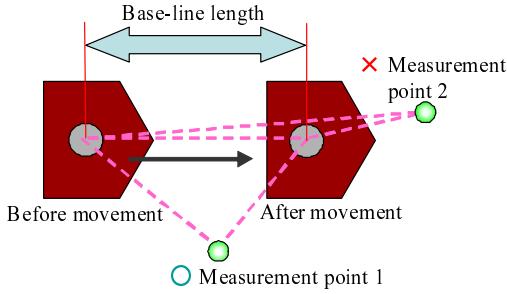


Figure 3. Advantages and disadvantages of motion stereo measurement. Point 1 can be measured with high accuracy. However, it is difficult to measure point 2 with high accuracy.

The measurement accuracy by binocular stereo depends on the baseline length. The longer the baseline length compared with distance to a measurement object is, the higher the accuracy is. However, the measurement accuracy of the method is limited because the baseline length cannot be made longer than the robot size. Therefore, the method can measure with high accuracy only near the robot (Fig. 2).

Motion stereo using a pair of images taken with a single camera at different observation points is equivalent to binocular stereo vision. This method can make the baseline length longer without restriction of the robot size. Therefore, objects far from the robot can be measured with higher accuracy than by binocular stereo. It is necessary to estimate camera motions (the relative relations of camera positions and orientations) in this method. The structure from motion (SFM) that is a kind of motion stereo estimates camera motions, and then measures objects [8].

However, the SFM is difficult to measure objects existing in baseline direction with high accuracy (Fig. 3). In addition, the scale is unknown by using SFM-based approaches.

Then, we aim to improve measurement accuracy by using two omnidirectional cameras installed on a mobile robot. In addition, stereo matching accuracy is improved by considering omnidirectional image distortion.

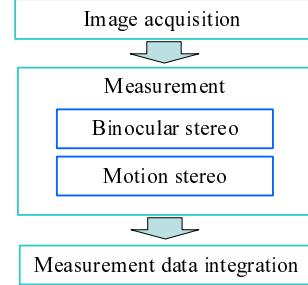


Figure 4. Procedure of our proposed method.

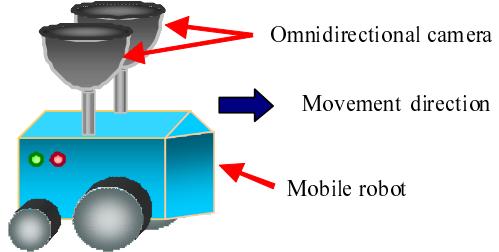


Figure 5. Camera configuration of our proposed method.

2. Outline

The process of our method is shown in Fig. 4. First, the mobile robot equipped with two omnidirectional cameras acquires an omnidirectional image sequence during its locomotion. Next, by using omnidirectional images, binocular stereo measurement and motion stereo measurement are executed. Finally, both measurement results are integrated to obtain high measurement accuracy.

Two omnidirectional cameras are installed parallel on the mobile robot. The baseline direction is decided perpendicular to the movement direction (Fig. 5).

3. Ray Vector Calculation

The coordinate system of our omnidirectional camera is shown in Fig. 6. The omnidirectional camera has a hyperboloid mirror in front of the lens of a conventional camera. A ray heading to image coordinates (u, v) from the camera lens is reflected on the hyperboloid mirror. In this paper, the reflected vector is called a ray vector. The extension lines of all ray vectors intersect at the focus of the hyperboloid mirror. The ray vector is calculated by the following equations.

$$\mathbf{r} = \begin{bmatrix} su \\ sv \\ sf - 2c \end{bmatrix}, \quad (1)$$

$$s = \frac{a^2 (f\sqrt{a^2 + b^2} + b\sqrt{u^2 + v^2 + f^2})}{a^2 f^2 - b^2 (u^2 + v^2)}, \quad (2)$$

where a , b and c are hyperboloid parameters and f is the focal length of the lens, respectively.

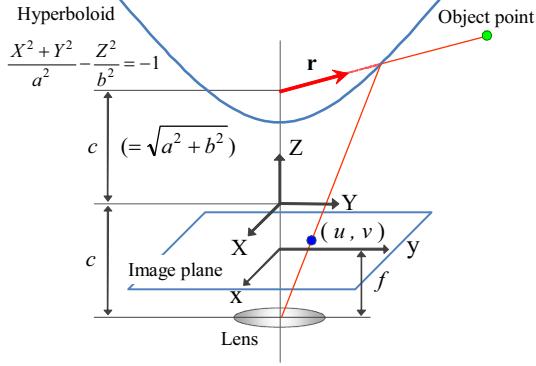


Figure 6. The coordinate system of the omnidirectional camera. A ray vector is defined as a unit vector which starts from the focus of a hyperboloid mirror.

4. Binocular Stereo Measurement

4.1. Epipolar Constraint

It is necessary to detect corresponding points between two images in stereo measurement. Then, in our method, we use template matching for searching corresponding points. We use epipolar constraint for template matching because the search range can be limited on the epipolar line. The epipolar line is obtained by projecting the ray vector calculated from a point in the image onto the other image (Fig. 7).

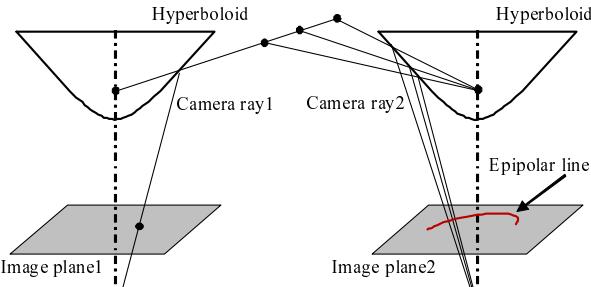


Figure 7. The epipolar line of the omnidirectional image. The epipolar line is obtained by projecting the ray vector.

4.2. Corresponding Point Acquisition

The omnidirectional image distortion and the projection distortion disturb corresponding point acquisition. The perspective projection conversion is used to solve this problem. The omnidirectional camera with a hyperboloid mirror can convert the omnidirectional image to the perspective projection image (Fig. 8). The perspective projection image is generated by projecting the omnidirectional image to the projection plane in the 3D space.

Calculation of the projection plane is important for decreasing various distortion. When the projection plane

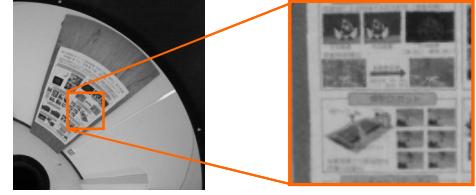


Figure 8. The image is obtained by perspective transformation. (a) Omnidirectional image. (b) Perspective image.

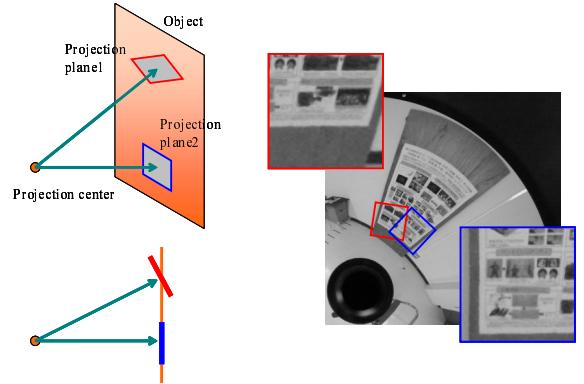


Figure 9. Projective distortion. The upper perspective image have projective distortion. The lower one does not have projective distortion.

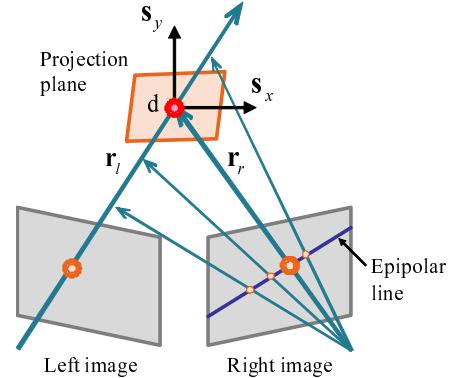


Figure 10. Detecting of corresponding point by comparing dissimilarity measures. The dissimilarity measure is calculated by SAD.

leans, the projection distortion arises to the perspective projection image (Fig. 9). The dissimilarity measure becomes small if the distortion is small when the feature quantity is compared. Therefore, we search for the center position and the gradient of the projection plane that minimizes dissimilarity measures as follows. Corresponding points are acquired from the perspective projection image calculated based on the center position and the gradient.

Let \mathbf{d} be the center position of the projection plane and \mathbf{s}_x and \mathbf{s}_y be the axes of the projection plane.

At first, the method calculates \mathbf{d} . As the first step, our

method acquires images taken with right and left cameras. Then, feature points are extracted from the left image. Ray vector \mathbf{r}_1 is calculated from these feature points. The epipolar line in the right image is calculated from \mathbf{r}_1 and the position relationship between the left and right cameras. \mathbf{d} is given by searching along the epipolar line of the \mathbf{r}_1 in the right image (Fig. 10).

In the next step, the ray vector \mathbf{r}_2 that passes each point on the epipolar line is calculated. The candidate point of \mathbf{d} is calculated as a cross point of \mathbf{r}_1 and each \mathbf{r}_2 .

In the third step, the projection plane is generated centering on each \mathbf{d} . An omnidirectional image is enlarged on the projection plane. And the projection plane is projected to the image plane.

After these steps, the right image and the left image are converted to the perspective projection image at each candidate point of \mathbf{d} . By comparing dissimilarity measures between these images, \mathbf{d} which minimizes dissimilarity measure is decided. The dissimilarity measure is calculated by SAD (Sum of Absolute Difference).

Then, \mathbf{s}_x and \mathbf{s}_y are rotated at the position of \mathbf{d} . Feature quantity is compared while rotating \mathbf{s}_x and \mathbf{s}_y . At last, corresponding points which are regarded as the same point in 3D space can be obtained.

4.3. 3D Measurement

Each ray vector is calculated from corresponding points of two images. Measurement points are obtained as the cross point of each ray vector.

5. Motion Stereo Measurement

5.1. Point Tracking

For getting correspondent points between images in the omnidirectional image sequence, the method extracts feature points in the initial image and then tracks them along the sequence. In our method, we use the Lucas Kanade tracker algorithm [9] with image pyramid representation [10]. These points are regarded as corresponding points between two images taken before and after the robot movement.

When the baseline direction is decided perpendicular to the movement direction, the other camera appears in the field of view. This might cause error in camera motion estimation. Therefore, in our method, the area where the other camera appears is excluded from each image.

5.2. Camera Motion Estimation

The ray vector is calculated from feature points in before and after images as follows.

$$\mathbf{r}_i = [x_i, y_i, z_i]^T, \mathbf{r}'_i = [x'_i, y'_i, z'_i]^T. \quad (3)$$

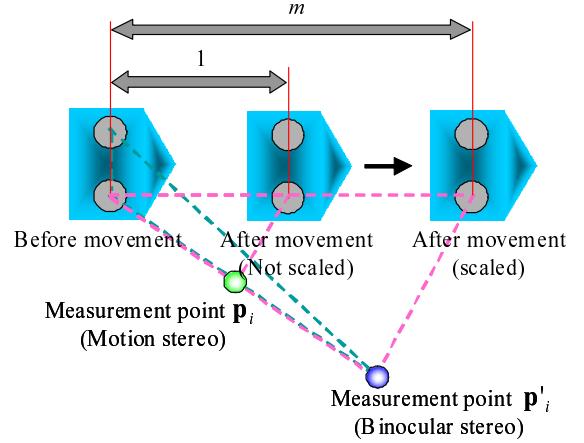


Figure 11. Scale matching.

Then, the essential matrix \mathbf{E} that describes camera positions and orientations is calculated. \mathbf{E} is expressed by Eq. (4), and Eq. (5) is obtained from Eq. (4).

$$\mathbf{r}'_i^T \mathbf{E} \mathbf{r}_i = 0, \quad (4)$$

$$\mathbf{u}_i^T \mathbf{e} = 0, \quad (5)$$

where \mathbf{u}_i and \mathbf{e} are expressed by the following equations (Eqs. (6) and (7)).

$$\mathbf{u}_i = [x_i x'_i, y_i x'_i, z_i x'_i, x_i y'_i, y_i y'_i, z_i y'_i, x_i z'_i, y_i z'_i, z_i z'_i]^T, \quad (6)$$

$$\mathbf{e} = [e_{11}, e_{12}, e_{13}, e_{21}, e_{22}, e_{23}, e_{31}, e_{32}, e_{33}]^T. \quad (7)$$

Then, \mathbf{E} is calculated by Eq. (8).

$$\min_e \|\mathbf{B}\mathbf{e}\|^2, \quad (8)$$

$$\mathbf{B} = [\mathbf{u}_1, \mathbf{u}_2, \dots, \mathbf{u}_n]^T, \quad (9)$$

where n is the number of feature points. \mathbf{E} is expressed by Eq. (11) which is a product of a rotation matrix \mathbf{R} and a translation vector $\mathbf{t} = [t_x, t_y, t_z]^T$ [11].

$$\mathbf{E} = \mathbf{T}\mathbf{R}, \quad (10)$$

$$\mathbf{T} = \begin{pmatrix} 0 & -t_z & t_y \\ t_z & 0 & -t_x \\ -t_y & t_x & 0 \end{pmatrix}. \quad (11)$$

5.3. 3D Measurement

Each ray vector is calculated from corresponding points of before and after images. The measurement point is obtained as the cross point of each ray vector.

The motion stereo method cannot determine the distance $|\mathbf{t}|$ between two observation points because the method

which uses only images for the input does not get scale information. In our method, the 3D coordinates of the feature point measured by the binocular stereo method has scale information. Therefore, we use the measurement result by the binocular method for scale matching.

We measure the 3D coordinates of a point \mathbf{p}_i by the motion stereo method. The 3D coordinates of the same point \mathbf{p}'_i are measured by the binocular stereo method (Fig. 11). The scale m is calculated by Eq. (12).

$$m\mathbf{p}_i = \mathbf{p}'_i. \quad (12)$$

Other motion stereo measurement data are scaled by the scale factor m .

6. Measurement Data Integration

The measurement accuracy will be low if an object is close to the baseline direction or far from the camera. Therefore, the measurement data is a mixture of high and low accuracy. Then we remove the measurement data of low accuracy by taking the differentiation of measurement result \mathbf{p}_i by the image coordinates of the two feature points $[u_{m,i}, v_{m,i}]^T$ and $[u'_{m,i}, v'_{m,i}]^T$ as the evaluation of the measurement accuracy. Measurement results satisfying Eq. (17) are regarded as good results, where h is a threshold.

$$\mathbf{g}_i = [g_{x,i}, g_{y,i}, g_{z,i}]^T, \quad (13)$$

$$g_{x,i} = \left\| \frac{\partial \mathbf{p}_i}{\partial x_{1,i}} \frac{\partial x_{1,i}}{\partial u_{1,i}} + \frac{\partial \mathbf{p}_i}{\partial x_{1,i}} \frac{\partial x_{1,i}}{\partial v_{1,i}} + \frac{\partial \mathbf{p}_i}{\partial x_{2,i}} \frac{\partial x_{2,i}}{\partial u_{2,i}} + \frac{\partial \mathbf{p}_i}{\partial x_{2,i}} \frac{\partial x_{2,i}}{\partial v_{2,i}} \right\|, \quad (14)$$

$$g_{y,i} = \left\| \frac{\partial \mathbf{p}_i}{\partial y_{1,i}} \frac{\partial y_{1,i}}{\partial u_{1,i}} + \frac{\partial \mathbf{p}_i}{\partial y_{1,i}} \frac{\partial y_{1,i}}{\partial v_{1,i}} + \frac{\partial \mathbf{p}_i}{\partial y_{2,i}} \frac{\partial y_{2,i}}{\partial u_{2,i}} + \frac{\partial \mathbf{p}_i}{\partial y_{2,i}} \frac{\partial y_{2,i}}{\partial v_{2,i}} \right\|, \quad (15)$$

$$g_{z,i} = \left\| \frac{\partial \mathbf{p}_i}{\partial z_{1,i}} \frac{\partial z_{1,i}}{\partial u_{1,i}} + \frac{\partial \mathbf{p}_i}{\partial z_{1,i}} \frac{\partial z_{1,i}}{\partial v_{1,i}} + \frac{\partial \mathbf{p}_i}{\partial z_{2,i}} \frac{\partial z_{2,i}}{\partial u_{2,i}} + \frac{\partial \mathbf{p}_i}{\partial z_{2,i}} \frac{\partial z_{2,i}}{\partial v_{2,i}} \right\|, \quad (16)$$

$$\|\mathbf{g}_i\| < h. \quad (17)$$

Therefore, binocular stereo and motion stereo measurement results are integrated with high accuracy.

7. Experiments

We used two omnidirectional cameras installed on the mobile robot (Fig. 12). The baseline length is 360 mm. The image size is 1920×1080 pixels.



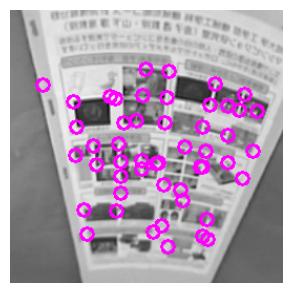
Figure 12. The mobile robot equipped with two omnidirectional cameras.



Figure 13. Feature points extracted by KLT algorithm.



(a)



(b)

Figure 14. Corresponding points obtained by template matching. (a) With perspective transformation. (b) With panoramic extension.

7.1. Experiment for Matching Comparison

We compared the binocular stereo matching with perspective transformation and matching with panoramic expansion. The distance between the omnidirectional camera and the measurement object is 1m. We used 50 feature points extracted by the Lucas Kanade tracker algorithm (Fig. 13). Figure 14 shows corresponding points obtained by each method. Mismatching points are 11 points by using a panoramic image. Mismatching points are 4 points by the proposed method. Figure 15 shows a histogram of matching score. The vertical axis is the number of corresponding points and the horizontal axis is the similarity. From those results, the proposed method using perspective transformation is possible to obtain corresponding points with high accuracy.

7.2. Measurement Experiment

In the experiment, we measured an indoor environment (Fig. 16). We acquired omnidirectional images (Fig. 17) by using the mobile robot. The top view of the measurement data by binocular stereo is shown in Fig. 18. The red marks show the measurement data. The blue marks show the camera position. The measurement data has error because of including error by camera calibration. The top view of the measurement data by motion stereo is shown Fig. 19. The yellow marks show the measurement data. The green marks

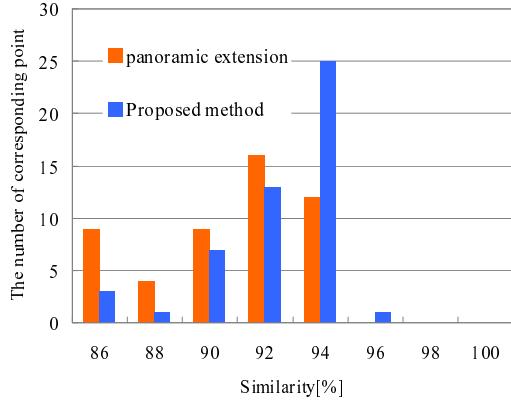


Figure 15. The histogram of matching score. The vertical axis is the number of corresponding points. The horizontal axis is the similarity.



Figure 16. Experimental environment.

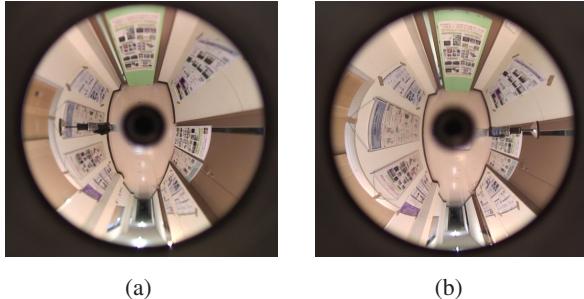


Figure 17. Omnidirectional stereo image pair are acquired by using mobile robot. (a) Left image. (b) Right image.

show the camera position. The measurement data with scale matching is shown Fig. 20. The top view of the measurement results integrated with binocular stereo data and motion stereo data is shown in Fig. 21. The bird's eye view of the measurement results is shown in Fig. 22. Table 1 shows the standard deviation from the least squares plane 1 and 2 (Fig. 21). These results show that the proposed method can measure the environment with high accuracy. Table 2 shows the processing time of each method. The processing time of binocular stereo takes 80.78sec to measure the area of $3m \times 4.5m \times 2m$. The processing time of motion stereo takes 62.23sec to measure the area of $0.5m \times 2.5m \times 1.5m$. Then, 90 images are used by motion stereo measurement.

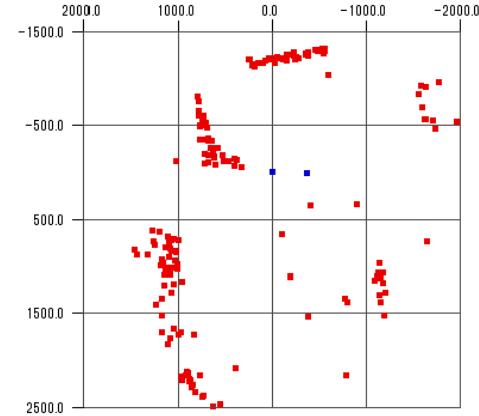


Figure 18. Measurement data by binocular stereo. The red marks show the measurement data. The blue marks show the camera position.

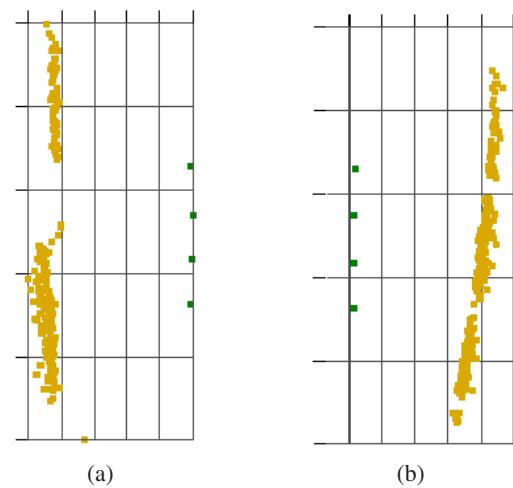


Figure 19. Measurement data by motion stereo. The yellow marks show the measurement data. The green marks show the camera position. (a) Left camera. (b) Right camera.

Table 1. Standard deviation from the least squares plane.

	Standard deviation
Plane1	9.09mm
Plane2	17.21mm

8. Conclusions

In this paper, we propose a three-dimensional measurement method using binocular stereo together with motion stereo by two omnidirectional cameras installed on a mobile robot. Experimental results showed the effectiveness of the proposed method.

As future works, the camera internal parameters should be estimated accurately to improve the measurement accuracy and the processing time. Many points should be used in scale matching to improve the measurement accuracy.

Table 2. Processing times.

Measurement method	Processing time
Binocular stereo	80.78sec
Motion stereo	62.23sec

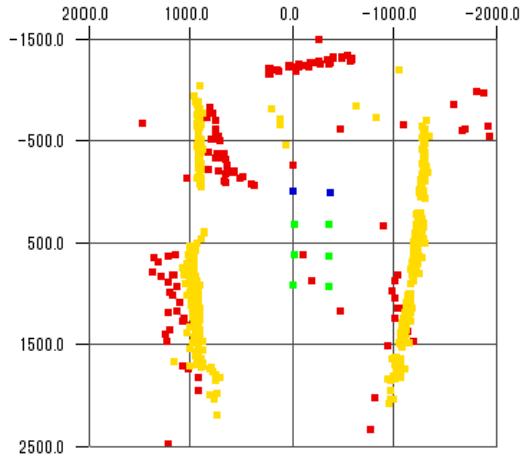


Figure 20. Measurement data with scale matching. The red marks and yellow marks show the measurement data by binocular stereo and motion stereo respectively. The blue marks and green marks show the camera position by binocular stereo and motion stereo respectively.

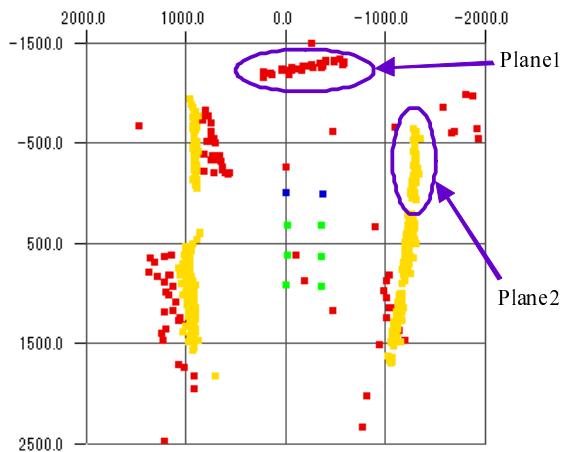


Figure 21. Top view of integrated measurement result. The red marks and yellow marks show the measurement data by binocular stereo and motion stereo respectively. The blue marks and green marks show the camera position by binocular stereo and motion stereo respectively.

Acknowledgements

This work was in part supported by MEXT KAKENHI, Grant-in-Aid for Young Scientist (A), 22680017.

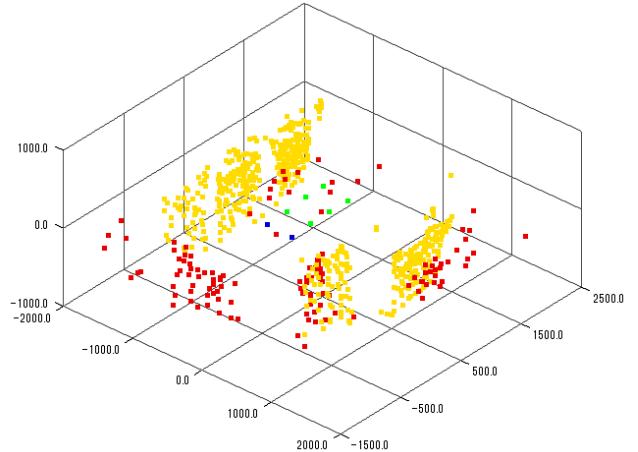


Figure 22. Bird's eye view of integrated measurement result. The red marks and yellow marks show the measurement data by binocular stereo and motion stereo respectively. The blue marks and green marks show the camera position by binocular stereo and motion stereo respectively.

References

- [1] A. J. Davison: "Real-Time Simultaneous Localisation and Mapping with a Single Camera", *Proceedings of 9th IEEE International Conference on Computer Vision*, Vol. 2, pp. 1403-1410, 2003.
- [2] C. Geyer and K. Daniilidis: "Omnidirectional Video", *The Visual Computer*, Vol.19, No.6, pp.405-416, 2003.
- [3] R. Bunschoten and B. Krose: "Robust Scene Reconstruction from an Omnidirectional Vision System", *IEEE Transactions on Robotics and Automation*, Vol.19, No.2, pp.351-357, 2003.
- [4] J. Gluckman and S. K. Nayar: "Ego-motion and Omnidirectional Cameras", *Proceedings of the 6th International Conference on Computer Vision*, pp. 999-1005, 1998.
- [5] H. Ishiguro, M. Yamamoto and S. Tsuji: "Omni-Directional Stereo", *IEEE Transactions on Pattern Analysis and Machine Intelligence*, Vol. 14, No. 2, pp. 257-262, 1992.
- [6] J. Takiguchi, M. Yoshida, A. Takeya, J. Eino and T. Hashizume: "High Precision Range Estimation from an Omnidirectional Stereo System", *Proceedings of the 2002 IEEE/RSJ International Conference on Intelligent Robots and Systems*, pp. 263-268, 2002.
- [7] R. Kawanishi, A. Yamashita and T. Kaneko: "Three-Dimensional Environment Model Construction from an Omnidirectional Image Sequence", *Journal of Robotics and Mechatronics*, Vol.21, No.5, pp.574-582, 2009.
- [8] P. Chang and M. Hebert: "Omni-directional Structure from Motion", *Proceedings of the 2000 IEEE Workshop on Omnidirectional Vision*, pp.127-133, 2000.

- [9] J. Shi and C. Tomasi: “Good Features to Track”, *Proceedings of the 1994 IEEE Computer Society Conference on Computer Vision and Pattern Recognition*, pp.593-600, 1994.
- [10] J. Y. Bouguet: “Pyramidal Implementation of the Lucas Kanade Feature Tracker Description of the Algorithm”, *OpenCV*, Intel Corporation, 2000.
- [11] R. Hartley and A. Zisserman: *Multiple View Geometry in Computer Vision, Second Edition*, Cambridge University Press, 2004.