ENVIRONMENT OBSERVATION BY STRUCTURE FROM MOTION WITH AN OMNI-DIRECTIONAL CAMERA

Tomoaki Harada, Atsushi Yamashita and Toru Kaneko

Department of Mechanical Engineering, Shizuoka University,
3-5-1 Johoku, Hamamatsu-shi, Shizuoka, 432-8561, Japan
Phone: +81-53-478-1604, Fax: +81-53-478-1604
Email: {f0430052, tayamas, tmtkane}@ipc.shizuoka.ac.jp

ABSTRACT

Measurement of a surrounding environment structure is important for mobile robots to make autonomous locomotion. In this paper, we propose a method for three-dimensional measurement of the environment using an omni-directional camera installed on a mobile robot. The method is based on structure from motion under the assumption that the environment is static. Along with the image sequence, it tracks feature points to get corresponding points for stereo matching. Using stereo pair images taken at selected observation points, the method estimates the relative relations of camera positions and orientations. With these relations and image coordinates of the feature points, three-dimensional coordinates of object points are calculated by triangulation. Integration of the individual measurement data realized by scale matching produces a whole map of the environment. Experimental results show the validity of the method.

1. INTRODUCTION

Map information is important for path planning and self-localization when mobile robots accomplish autonomous tasks. In an unknown environment, however, mobile robots should generate map information by themselves.

Three-dimensional measurement using image data makes it possible to generate map information, and image data is acquired ordinarily by using a conventional camera [1] [2] whose field of view is limited. To realize the measurement more efficiently, we have another way to use an omni-directional camera [3] which have a wide field of view. A stereo vision method using two omni-directional cameras equipped on a mobile robot is proposed [4]. The measurement accuracy by stereo vision depends on the baseline length, i.e., the longer the baseline length, the better the accuracy. Therefore, the measurement accuracy of the above method is limited because its baseline length cannot be longer than the robot size.

Structure from motion, using a stereo pair images each of which is taken with a single camera during robot locomotion, is equivalent to binocular stereo vision, but its baseline length can be longer without restriction of the robot size. Therefore, a method based on structure from motion can measure distant objects with higher accuracy than a method based on binocular stereo vision. A requirement to this method is to know a relative distance and orientation difference of the camera between two observation points.

The camera positions and orientations can be measured by using a dead reckoning function of the robot which is realize by counting the number of wheel rotation. However, dead reckoning gives erroneous data when wheels slip or run over uneven terrain, and then errors are accumulated during locomotion. Another method of estimating positions and orientations of the camera is to use image information. It makes it possible to measure the surrounding environment without any sensor except for one camera [2] [5].

As mentioned above, it is desirable for a mobile robot to employ an omni-directional camera with a wide view angle and a structure from motion strategy [6] which gives better three-dimensional measurement accuracy than binocular stereo vision does. Consequently, in this paper we propose a method of measuring the surrounding environment by a mobile robot equipped with an omni-directional camera. The method estimates relative differences of positions and orientations of the camera between two observation points by analyzing acquired image data. Integrating the all measurement data obtained for individual observation point pairs, the method generates a whole map of the environment by using the geometrical relation of the camera positions and orientations.

2. OUTLINE

Structure from motion is realized by stereo matching of images taken at different observation points during the robot locomotion. We use an omni-directional camera to get an image sequence along with robot locomotion. We employ a theory of weakly calibrated stereo vision, which enables us to perform stereo measurement without any other information than stereo images themselves. In this paper, we assume that the robot moves in a static environment.

In the proposed method, the first procedure is to find corresponding points among images in the image frame sequence. In the first frame image, we extract feature points which are so distinguishable to track in the sequence. Then we track these feature points by image processing technique, and then the tracked points in the other image of a stereo pair and the points in the first image are identified as corresponding points. By using these corresponding points, we estimate the relative relation in positions and

orientations of the camera at the two observation points. The geometrical relation and the image coordinates of the feature points give enable us to calculate three-dimensional coordinates with an ambiguity concerning to their scale.

To generate a whole map, we integrate the measurement data by matching their three-dimensional coordinates obtained for adjacent observation point pair. Figure 1 shows how to combine adjacent data. First, we estimate rotation matrix $\mathbf{R}_1$ and translation vector $\mathbf{t}_1$ by using two images taken at observation points 1 and 2. Using these values, we obtain three-dimensional coordinates $\mathbf{p}_{1,i}$ of the point from which the corresponding feature points in the images are originated, where i is the number of corresponding point. Then, we estimate rotation matrix $\mathbf{R}_2$ and translation vector $\mathbf{t}_2$, and measure $\mathbf{p}_{2,i}$ by using two images taken at observation points 2 and 3. Therefore, the relation of positions and orientations among observation points 1,2,3 is represented by rotation matrices and translation vectors $\mathbf{R}_1$,$\mathbf{R}_2$,$\mathbf{t}_1$,$\mathbf{t}_2$. However, because the measurement only uses the information of corresponding points, the distance between observation points is indeterminate. In other words, the measurement results are in relative scale of $|\mathbf{t}_i|$. Therefore, when the distance between observation points 1 and 2 is different from that between observation points 2 and 3, $|\mathbf{t}_1|$ and $|\mathbf{t}_2|$ need to be matched in relative scale of these distances. Scale matching is realized by making three-dimensional coordinates $\mathbf{p}_{1,i}$ and $\mathbf{p}_{2,i}$ that belong to the same point in three-dimensional space be closest. By this processing, we can combine three-dimensional data obtained at adjacent observation point pairs. To perform this procedure, we get a whole map data integrated from the image sequence.

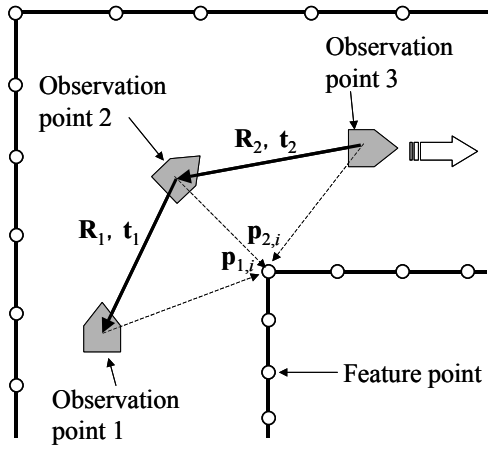

Fig. 1 Measurement Data Combination.

3. ALGORITHM

3.1 Corresponding Point Acquisition

For getting correspondent points between two images in the image frame sequence, we extract feature points in the first image and then track them along the sequence. In our method, we use KLT (Kanade-Lucas-Tomasi) tracker [7].

First, we extract g features in the first frame. Second, we track these feature points successively. Feature points that cannot be tracked because of less correspondence disappear, and remaining feature points that are tracked successfully to the image of the next observation point are regarded as corresponding points. The image of the next observation point is set to be the first image of the next measurement, and new feature points are extracted so that the number of the remaining feature points in the first image and the newly added feature points make the total number be constant g.

3.2 Ray Vector Calculation

We define a unit vector originating from the center of projection to an object point in three-dimensional space as a ray vector $\mathbf{r}=[x,y,z]^T$, where ^{T} stands for transposition of vector or matrix. An omni-directional camera we use has a hyperboloid mirror in front of a lens of a conventional camera [8]. Therefore, as shown in Fig. 2, ray vector $\mathbf{r}$ is directed from the focus of the hyperboloid mirror to the refection point of the ray on the mirror surface.

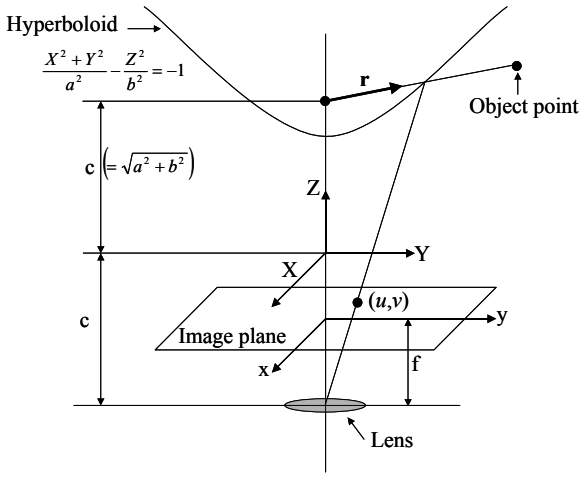

Fig. 2 Ray Vector.

Ray vector $\mathbf{r}$ is calculated from image coordinates $[u,v]^T$ of the feature using Eq.(1), (2). In the equations, a, b and c are the hyperboloid parameters and f is the image distance (the distance between the center of projection, or the center of the lens, and the image plane) of camera.

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

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

3.3 Essential Matrix Calculation

Matrix $\mathbf{E}$ which satisfies Eq.(3) is called an essential matrix,

$$\mathbf{r}_i'^T\mathbf{E}\mathbf{r}_i=0 \quad (3)$$

where ray vectors $\mathbf{r}_i=[x_i,y_i,z_i]^T$, $\mathbf{r}_i'=[x'_i,y'_i,z'_i]^T$ are those of the corresponding point in two images, respectively.

Essential matrix contains information about relative position and orientation differences between two observation points.

Equation(3) is transformed into Eq.(4),

$$\mathbf{u}_i^T\mathbf{e}=0 \tag{4}$$

where

$\mathbf{u}_i=[x_ix'_i, y_ix'_i, z_ix'_i, x_iy'_i, y_iy'_i, z_iy'_i, x_iz'_i, y_iz'_i, z_iz'_i]^{\mathrm{T}}$

$\mathbf{e}=[e_{11}, e_{12}, e_{13}, e_{21}, e_{22}, e_{23}, e_{31}, e_{32}, e_{33}]^{\mathrm{T}}$

(e_{ab} is the row a and column b element of matrix $\mathbf{E}$).

Essential matrix $\mathbf{E}$ is obtained by solving simultaneous equations for more than 8 pair of corresponding ray vectors. This means that we solve Eq.(5), where n is the number of corresponding feature points.

$$\min_{\mathbf{E}}\sum_{i=1}^{n}\left(\mathbf{r}_i'^T\mathbf{E}\mathbf{r}_i\right)^2 \tag{5}$$

Equation (5) is transformed into Eq.(6).

$$\min_{\mathbf{E}}\left|\mathbf{U}\mathbf{e}\right|^2 \tag{6}$$

where $\mathbf{U}=[\mathbf{u}_1, \mathbf{u}_2, \cdots \mathbf{u}_n]^{\mathrm{T}}$.

$\mathbf{e}$ is given as the eigenvector of the smallest eigenvalue of $\mathrm{U^TU}$ and then essential matix $\mathbf{E}$ is obtained.

3.4 Outlier Rejection

All feature points tracked along the image sequence do not behave satisfactorily as corresponding points because of image noise etc. Feature points of mistracking should be rejected as outliers. To solve this problem, we employ a method of RANSAC [9].

In the procedure, we select randomly 8 feature points, which are of the minimum number of points for determining essential matrix $\mathbf{E}$, Let $\mathbf{E}_{\mathrm{rand}}$ be the essential matrix determined by using these feature points, and k be the number of feature points satisfying Eq.(7), where q is a given threshold..

$$\left|\mathbf{r}_i'^T\mathbf{E}_{\mathrm{rand}}\mathbf{r}_i\right|<q \tag{7}$$

We repeat this process of determining essential matrix $\mathbf{E}_{\mathrm{rand}}$ and number k for predetermined times. Then we choose the case with the maximum number of k, and we remove feature points that do not satisfy Eq.(7) as outliers. Finally, we calculate essential matrix $\mathbf{E}$ by Eq.(6) using the remaining feature points.

3.5 Estimating Position and Orientation

Essential matrix $\mathbf{E}$ is represented by rotation matrix $\mathbf{R}$ and translation vector $\mathbf{t}=[t_x,t_y,t_z]^{\mathrm{T}}$.

$$\mathbf{E}=\mathbf{R}\mathbf{T} \tag{8}$$

Here, $\mathbf{T}$ is a matrix given as follows.

$$\mathbf{T}=\begin{bmatrix}0 & -t_z & t_y\\ t_z & 0 & -t_x\\ -t_y & t_x & 0\end{bmatrix}.$$

The method cannot determine the distance $|\mathbf{t}|$ between two observation points because the measurement only uses images for input and does not get any scale information.

In the procedure, the distance made by camera motion is set to be a unit, which is realized by changing the norm $\|\mathbf{E}\|$ so as to be $|\mathbf{t}|=1$. If $|\mathbf{t}|=1$, the Frobenius norm of $\mathbf{T}$ become $\|\mathbf{T}\|=\sqrt{2}$. Since multiplication of T by R does not change the norm in Eq.(8), we make the norm $\|\mathbf{E}\|$ as $\|\mathbf{E}'\|=\sqrt{2}$ as shown in Eq.(9).

$$\mathbf{E}'=\frac{\sqrt{2}}{\|\mathbf{E}\|}\mathbf{E} \tag{9}$$

In order to decompose matrix $\mathbf{E}$' into rotation matrix $\mathbf{R}$ and matrix $\mathbf{T}$, we perform singular value decomposition as shown in Eq.(10).

$$\mathbf{E}'=\mathbf{U}\Sigma\mathbf{V}^T \tag{10}$$

where

$$\Sigma=\mathrm{diag}(r,s,0).$$

Values r and s are the singular values of matrix E', each of which are close to 1. We change the values to be r=s=1 in order to make matrix $\mathbf{R}$ be a rotation matrix. By transforming Eq.(10) into Eq.(11), rotation matrix $\mathbf{R}$ and matrix $\mathbf{T}$ are obtained in the form of Eq.(12).

$$\mathbf{E}'=\mathbf{U}\mathbf{Y}\mathbf{V}^T\mathbf{V}\mathbf{Z}\mathbf{V}^T \tag{11}$$

$$\mathbf{R}=\mathbf{U}\mathbf{Y}\mathbf{V}^T,\quad \mathbf{T}=\mathbf{V}\mathbf{Z}\mathbf{V}^T \tag{12}$$

where

$$\mathbf{Y}=\begin{bmatrix}0 & \pm1 & 0\\ \mp1 & 0 & 0\\ 0 & 0 & \det\mathbf{U}\mathbf{V}^T\end{bmatrix},\quad \mathbf{Z}=\begin{bmatrix}0 & \pm1 & 0\\ \mp1 & 0 & 0\\ 0 & 0 & 0\end{bmatrix}.$$

Translation vector $\mathbf{t}$ is obtained from elements of matrix $\mathbf{T}$. There are multiple solutions for rotation matrix $\mathbf{R}$ and translation vector $\mathbf{t}$, and we choose those which make the direction of the feature point be the same as that of the ray vector.

3.6 Object Point Measurement

Three-dimensional coordinates of an object point which is projected as a feature point in the image are given based on triangulation with two cameras set in the geometrical relation given by rotation matrix $\mathbf{R}_m$ and translation vector $\mathbf{t}_m$, where m is the number of measurement.

Theoretically, the measurement result indicates that the coordinates of the object point are given at the cross-point of the corresponding two ray vectors, but in practice these ray vectors do not meet at one point because of image noise etc. Therefore, we choose the coordinates of the middle point on the line that connects the two ray vectors with the shortest distance. The resultant three-dimensional coordinates $\mathbf{p}_{m,i}$ of i–th feature point are obtained by Eq. (13).

$$\mathbf{p}_{m,i}=\frac{1}{2}\left\{\frac{A_{m,i}\mathbf{r}_{m,i}+B_{m,i}\mathbf{R}_m^T\mathbf{r}'_{m,i}}{\left(\mathbf{r}_{m,i}^T\mathbf{r}_{m,i}\right)\left(\mathbf{r}'^T_{m,i}\mathbf{r}'_{m,i}\right)-\left(\mathbf{r}_{m,i}^T\mathbf{R}_m^T\mathbf{r}'_{m,i}\right)^2}+\mathbf{t}_m\right\} \tag{13}$$

where

$$A_{m,i}=\left(\mathbf{r}_{m,i}^T\mathbf{t}_m\right)\left(\mathbf{r}'^T_{m,i}\mathbf{r}'_{m,i}\right)-\left(\mathbf{r}_{m,i}^T\mathbf{R}_m^T\mathbf{r}'_{m,i}\right)\left(\mathbf{r}'^T_{m,i}\mathbf{R}_m\mathbf{t}_m\right)$$

$$B_{m,i}=\left(\mathbf{r}_{m,i}^T\mathbf{t}_m\right)\left(\mathbf{r}'^T_{m,i}\mathbf{R}_m\mathbf{r}_{m,i}\right)-\left(\mathbf{r}_{m,i}^T\mathbf{r}_{m,i}\right)\left(\mathbf{r}'^T_{m,i}\mathbf{R}_m\mathbf{t}_m\right).$$

3.7 Result Qualification

The accuracy of measurement is poorer as an object point lies closer to the baseline direction or it lies far from

the camera. Therefore, the measurement data is a mixture of high and low accuracy. Here, by taking the differentiation of measurement result $\mathbf{p}_{m,i}$ by the image coordinates of the two feature points $[u_{m,i},v_{m,i}]^{\mathrm{T}}$ and $[u'_{m,i},v'_{m,i}]^{\mathrm{T}}$ as the estimate of the measurement accuracy, we select measurement results satisfying Eq.(14), where h is a threshold.

$$\left|\frac{\partial \mathbf{p}_{m,i}}{\partial u_{m,i}}+\frac{\partial \mathbf{p}_{m,i}}{\partial v_{m,i}}+\frac{\partial \mathbf{p}_{m,i}}{\partial u'_{m,i}}+\frac{\partial \mathbf{p}_{m,i}}{\partial v'_{m,i}}\right|<h \qquad (14)$$

3.8 Result Combination

By the above procedure, we get individual measurement results and the geometrical relations of observation points by using pairs of stereo images that are selected in the image sequence.

In order to unite these measurement results, we should solve the problem of scale ambiguity among individual measurements. In the above, we assumed that the distance between any set of two observation points to be 1. Therefore, measurement results are mismatched when actual distances are different to one another.

Figure 3 shows this scale mismatching problem, where the broken lines indicate the current result and the solid lines indicate the previous result. The scale of current result should be matched with the scale of previous result. Here, let observation point c be the common point for the previous and current measurements and have the previous measurement result $\mathbf{P}_{m,i}=[x_{m,i},y_{m,i},z_{m,i}]^{\mathrm{T}}$ and the current measurement result $\mathbf{P}_{m+1,i}=[x_{m+1,i},y_{m+1,i},z_{m+1,i}]^{\mathrm{T}}$ for a common object point projected in the image as i-th feature point. Scale matching is realized by making the three-dimensional coordinates of the common object point be as close as possible.

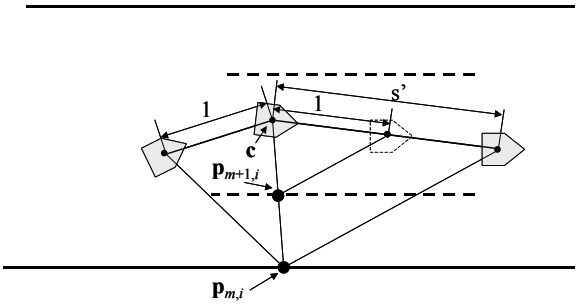

Fig. 3 Scale Mismatching.

Minimization of deviation of the two resultant coordinates of the common object point is more sensitive when the object point lies farther from the observation point. Therefore, it is appropriate to minimize the logarithmic distances between common object points rather than minimize the linear distances. Scale s' of the current measurement is obtained by Eq.(15).

$$\min_{s'}\sum_{i=1}^{n}\left|\log\left(\mathbf{p}_{m,i}-\mathbf{c}\right)-\log\left(s'\mathbf{p}_{m+1,i}-\mathbf{c}\right)\right|^2 \qquad (15)$$

4. EXPERIMENT

4.1 Omni-directional Camera

The omni-directional camera we used in the experiment is a combination of an HDV camera (HDR-HC1) and a hyperboloid mirror (SOIOS70-scope) shown in Fig. 4. A motion image sequence was acquired with the image size of 1920 × 1080pixels, and the image capture rate was 30frame/s.

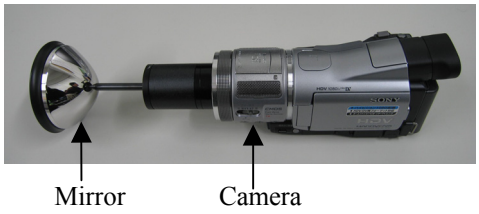

Fig. 4 Omni-directional Camera.

4.2 Experimental Result

In the experiment we measured a passageway including an L-shape corner. Figure 5 shows the input image in which extracted feature points are superimposed. The arrow indicates the direction of camera motion. Points with mark of ◇ are those rejected by RANSAC. Points with mark of △ are those judged as having poor accuracy. Points with mark of ○ are those used in combination of results as reliable points. The number of feature points extracted in the first frame was set to be g=200. The threshold for RANSAC was q=0.01. The threshold for judging whether the measurement has good accuracy was h=0.05. In the next course of measurement beginning with the image shown in Fig.5, the feature points with mark of ○ remained, the rest were discarded, and new feature points were extracted and added to make g be constant as 200.

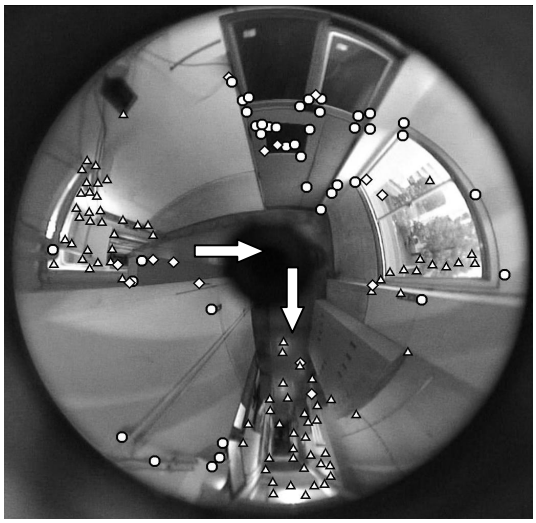

Fig. 5 Feature Points in Input Image.

Figures 6 and 7 show the top view of combined measurement result for the motion image sequence. Figure 6 shows the result without procedures of outlier rejection in Sec.3.4 nor result qualification in Sec. 3.7. Figure 7 shows the result with outlier rejection and result qualification procedures. The coordinate system is that of the last image in the sequence, and the scale is set so as to make the motion distance of the last measurement be a unit. Black points in the figures indicate the measurement results of the feature points. Gray points indicate the estimated camera position in each measurement. The arrow is the direction of camera motion. Figure 8 shows the same result of Fig. 7 in bird's-eye view.

In Fig.6, the measured positions of feature points spread over. Its scale matching failed, and the widths of passageway at the starting point and the endpoint differ much to each other. On the other hand, the measured feature points in Fig.7 are located in correct positions of the passageway. The two sides of the passageway are measured as parallel lines. This result shows that the proposed method is not affected by error accumulation which is unavoidable in dead reckoning.

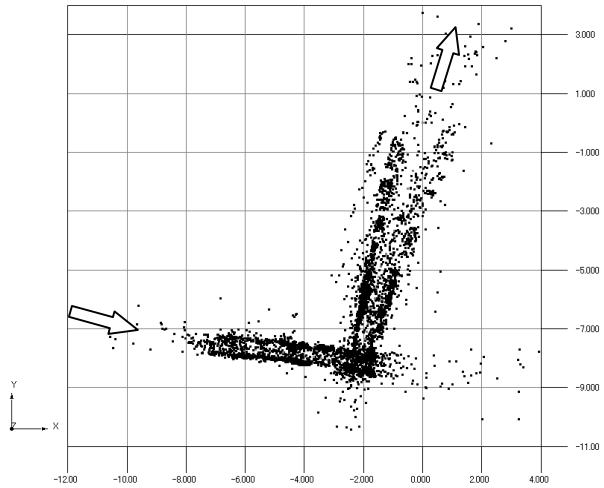

Fig. 6 Result without Outlier Rejection nor Result Qualification.

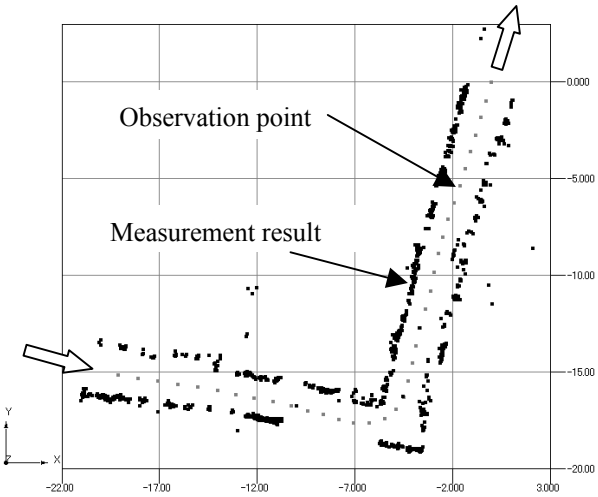

Fig. 7 Result with Outlier Rejection and Result Qualification.

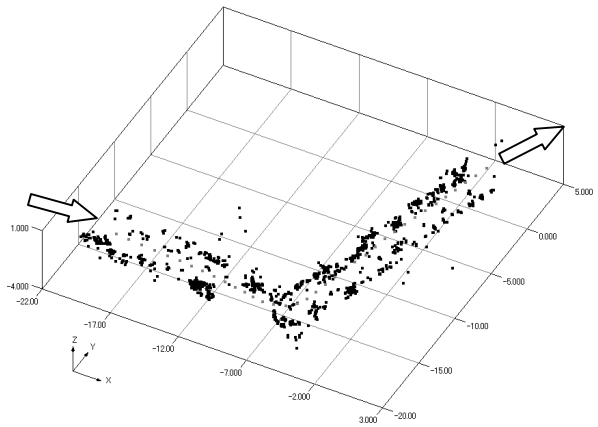

Fig. 8 Bird's-Eye View of Fig.7.

5. CONCLUSION

We proposed a method for three-dimensional measurement of environment based on structure from motion with images taken by an omni-directional camera. In order to make a whole map, the method integrates the measurement results obtained at multiple observation points along the image sequence. Advantages of the method are that the system configuration is simplified by using only image information taken by one camera, and that it can produce reliable data when a robot equipped with a camera moves unstably on rough surfaces where dead reckoning fails. Experimental results show the validity of the proposed method.

ACKNOWLEDGEMENT

This research was partly supported by CASIO Science Promotion Foundation, Japan.

REFERENCES

[1] K. Yamazaki, M. Tomono, T. Tsubouchi and S. Yuta: "3-D Object Modeling by a Camera Equipped on a Mobile Robot", Proceedings of the 2004 IEEE International Conference on Robotics and Automation, pp.1399-1405, 2004

[2] R. Hartley, R. Gupta and T. Chang: "Stereo from Uncalibrated Cameras", Proceedings of the 1992 IEEE Computer Society Conference on Computer Vision and Pattern Recognition, pp.761-764, 1992.

[3] Y Yagi: "Omnidirectional Sensing and Its Applications" IEICE Transactions on Information and Systems, Vol.E82-D, No.3, pp.568-579, 1999.

[4] H. Koyasu, J. Miura, and Y. Shirai: "Recognizing Moving Obstacles for Robot Navigation Using Real-time Omnidirectional Stereo Vision", Journal of

Robotics and Mechatronics, Vol. 14, No. 2,pp.147-156, 2002.
[5] G. Xu and Z. Zhang: “Epipolar Geometry in Stereo, Motion and Object Recognition”, Kluwer Academic Publishers, 1996.
[6] M. Oe, T. Sato and N. Yokoya: “Estimating Camera Position And Posture by Using Feature Landmark Database”, Proceedings of the 14th Scandinavian Conference on Image Analysis (SCIA2005), pp.171-181, 2005.
[7] 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.
[8] K. Yamazawa, Y. Yagi and M. Yachida: “Omnidirectional Image Sensor -HyperOmni Vision-”, Proceedings 3rd International Conference on Automation Technology, Vol.5, pp.125-132, 1994.
[9] M. A. Fischler and R. C. Bolles: “Random Sample Consensus: A Paradigm for Model Fitting with Applications to Image Analysis and Automated Cartography”, Communications of the ACM, Vol. 24, pp.381-395, 1981.