A line feature-based camera tracking method applicable to nuclear power plant environment

YAN Weida[1], ISHII Hirotake[2], SHIMODA Hiroshi[3], and IZUMI Masanori[4]

1. Graduate School of Energy Science, Kyoto University, Kyoto, 606-8501, Japan (yanweida@ei.energy.kyoto-u.ac.jp)
2. Graduate School of Energy Science, Kyoto University, Kyoto, 606-8501, Japan (hirotake@ei.energy.kyoto-u.ac.jp)
3. Graduate School of Energy Science, Kyoto University, Kyoto, 606-8501, Japan (shimoda@energy.kyoto-u.ac.jp)
4 .Fugen Decommissioning Engineering Center, Japan Atomic Energy Agency, Tsuruga, 914-8510, Japan (izumi.masanori@jaea.go.jp)

Abstract: Augmented reality, which can support the maintenance and decommissioning work of an NPP to improve efficiency and reduce human error, is expected to be pracitally used in an NPP. AR has indispensable tracking technology that estimates the 3D position and orientation of users in real time, but because of the complication of the NPP environment, it is difficult for its pracital use in the large space of an NPP. This study attempt to develop a tracking method for the pracital use in an NPP. Marker tracking is a legacy tracking method, but the preparation work necessary for that method is onerous. Therefore, this study developed and evaluated a natural feature-based camera tracking method that demands less preparation and which is applicable in an NPP environment. This method registers natural features as landmarks. When tracking, the natural features existing in the NPP environment can be registered automatically as landmarks. It is therefore possible to expand the tracking area to cover a wide environment in theory. The evaluation result shows that the proposed tracking method has the possibility to support field work of some kinds in an NPP environment. It is possible to reduce the preparation work necessary for the marker tracking method.
Keyword: maintenance and decommissioning support; augmented reality; tracking; line feature

1 Introduction

Hundreds of nuclear power plants (NPPs) are in operation throughout the world today. Compared with thermal plants of similar scale, the maintenance and decommissioning work of an NPP is much more difficult because of their complicated structure and radioactivity. Augmented reality (AR) technology, which has the possibility to offer support for maintenance and decommissioning work of an NPP to improve efficiency and safety, superimposes computer-generated information upon a worker's view intuitively at the correct 3D position in the real world [1][2]. For application of AR in plant field work, a tracking method that measures the 3D position and orientation of a worker with high accuracy in real-time is necessary. Many tracking technologies are used for AR, but most of them are inappropriate in a large indoor environment of NPP, which is invariably crowded with metal apparatus [2]. In other words, only a method based on vision sensors is useful with rather high accuracy and low cost in NPP environments. A vision sensor based method uses a camera (or some cameras) to capture images of the environment. It then extracts some features that can be recognized from the images through image processing. Using the extracted features, the 3D position and orientation of the camera (and the user which is holding the camera) is calculable from the geometric relation between the features and their projection on the captured images. The methods include marker tracking [3][4] and natural feature tracking [5]–[9]. Marker tracking has higher accuracy and stability, but it is troublesome because it is necessary to allocate numerous markers in the environment and to measure their 3D positions in advance. Natural features, which are visually recognizable parts of the environment such as corner points or edges of equipment, are more convenient for use as landmarks to calculate the position and orientation of the workers because of their low preparation workload. Point features are most widely used as natural features in tracking [5][6], but many occlusions exist in NPP environments because NPP have many more pipes and other components than other industrial plants have. Therefore point features are easily occluded, and

Received date: March 25, 2014
(Revised date: May 17, 2014)

many pseudo-points are detected. Moreover, some real feature points are not stable when the illumination or view angle changes. In contrast to point features, line features such as pipe boundaries and cables are abundant. Furthermore, line features have more pixels on an image; they can therefore be detected more reliably than point features because even if a part of a line is occluded, it is still detectable as a line.

The objective of this study is to attempt to use a line-feature-based camera tracking method that entails less preparation in NPP environment and evaluate its feasibility. In this tracking method, the camera pose is estimated by solving the perspective 3 Line (P3L) problem, which calculates the position and orientation of a camera through three 3D-lines and their projections (2D-lines) on an image. An image series captured from an NPP is used to evaluate the feasibility of the proposed method in the NPP environment. The tracking result is compared with the true data obtained from the NPP environment to analysis the performance and tracking error of the method.

2 Related works

Some applications use AR to support field work in an NPP. In an earlier study [10], dose-rate visualization was realized in an NPP environment, enabling field workers to avoid high-radiation areas. Another report of a study [11] described dismantling work supported by an AR system that refers the cutting parts and records the work progress. In another method [12], an AR system can simulate temporary placement and conveyance. Moreover, collision between the dismantled target and environment can be detected. Therefore, space verification is realized to support the dismantling planning. However, the NPP applications described above demand marker tracking, which limits the tracking area. In marker tracking, users must paste sufficient markers throughout the environment. Then they measure their 3D position in advance. Because NPPs are gigantic, the necessary number of markers is large, which increases the preparation workload for actual use in the whole NPP area. To resolve this problem, a natural-feature-based tracking method that entails less preparation work is anticipated. As described above in Section 1, point features are detected only unreliably because of the occlusion and the illumination change. Line features are better for use as a landmark for tracking in an NPP environment. Some related studies have examined tracking technology using line features. Bosse *et al.* [7] proposed a structure-from-motion (SFM) system using both point and line features extracted from omni-directional video sequences. Dailey *et al.* [8] used a trinocular stereo camera to estimate a map based on 3D line segments. Sola *et al.* [9] presented a monocular extended Kalman filter (EKF) SLAM using line landmarks. However, the methods described above entail some shortcomings if they are applied in an NPP environment. For example, only parallel lines or floor lines are used as in one method [7], but many line features in NPP have no parallel lines, or they are difficult to recognize as parallel lines. In another method described in the literature [8], line segments were used as landmarks, but in NPP, the line segments are not stable as landmark because the two endpoints of a line segment are difficult to recognize stably when the view angle or illumination changes. An infinite line is much more stable in such cases. In another study [9], the tracking camera is controlled by a robot to move along an accurately determined trajectory, which is impossible for workers when moving in an NPP environment.

In this study, infinite lines were used as landmarks. We estimate the camera pose by solving a P3L problem. We used an image series captured from an NPP environment to evaluate this method. No report in the relevant literature describes a similar study using line features for SLAM in an NPP environment.

3 NPP Environment and Tracking

In an NPP environment, it is invariably crowded with metal apparatus, which is the feature which affects the tracking most. Because of this feature, many existing tracking technology for AR is difficult to be applied. For example, the metal components limit the application of the tracking method based on the magnetic. As the described in [2], the tracking method with the highest cost performance is the vision sensor based method.

In the vision sensor based tracking, usually a video camera is used as a vision sensor. By capturing images of environment using camera, some features which are more distinguishable than other part of the image can be extracted from the images, such as

corners, edges, and planes. Then the relative position between the features and the camera can be estimated using a geometric method. The vision sensor based tracking includes two kinds: marker tracking and markerless tracking.

In the marker tracking, markers must be pasted in the tracking area, and their 3D position must be measured in advance. The tracking method is high accuracy and stability. However, in the NPP environment, the tracking area is always large, which requires that the marker size must large enough in the long distance tracking when the camera is far from the marker. And the complex environment which contains many pipe and other small components makes it difficult to paste large markers. Another problem is that the marker number would be very large to cover the whole NPP environment for tracking, which means the preparation work to paste and measure the markers is also very heavy. And even the preparation can be finished, the pasted markers which are with not so small size would disturb the maintenance or decommission work easily. Therefore, the application of marker tracking is limited in the NPP environment.

The markerless tracking has lower accuracy and stability than the marker tracking, but it require very light preparation work, because the features for tracking exist in the environment, and users don't have to paste or measure them in advance. In the markerless tracking methods, point features, such as the corner points of actual objects in the environment, are most widely used as natural features in tracking because it is the simplest feature and always abundant in most environment, but many occlusions exist in an NPP environment. Therefore point features are occluded easily, and many pseudo-points are detected. For example, the famous point feature-based tracking method PTAM[6] is very difficult to be used in NPP, espically in the environment with lots of pipes because of the occlusion and fewer point features. Moreover, some real feature points are not stable when the illumination or view angle changes. Compared with point features, line features such as the boundary of pipes and cables are abundant. Furthermore, the line feature has more pixels on an image, so it can be detected more reliably than point features because even a part of a line is occluded, it is still detectable as a line.

In generally, the line feature based tracking has high cost performance in the NPP environment. Consider that its accuracy and stability is lower than the marker tracking, another choice is a hybrid tracking method which combines the marker tracking and the line feature based tracking. In the hybrid tracking method, the long distance tracking can be executed by the line feature based tracking, so the marker size can be smaller enough to past it on a small component for the short distance tracking with high accuracy and stability. Moreover, the marker can be pasted only in the area which requires high accuracy and stability. Compared with marker tracking, the preparation and the affection on the maintenance or decommission work is reduced a lot in the hybrid tracking method. However, only line feature based tracking is used in this study because its performance must be evaluated first. If its accuracy and stability is enough for practical use, the hybrid method which still requires some preparation is then not so necessary. In other words, the hybrid method can be developed in the future work if necessary.

4 Proposed method

4.1 Profile of the Method

The main flow of the method is depicted in Figure 1. Eight 3D lines of which the position and direction are measured in advance are registered into databases as initial landmarks. When calculating the camera pose, four 3D lines are sufficient. However, more lines are necessary to improve the initial accuracy. Using the initial landmarks, the method starts tracking. It registers new landmarks from the environment while the camera is moving. At the first image frame, the eight line landmarks are matched with their corresponding 2D-line features. After capturing a new image, 2D-lines are detected on the image. Then they are matched with the features of the prior image. If a line is matched with the feature corresponding to a landmark, then it can be used for calculating the position and orientation of camera in this frame. Otherwise it is a candidate line that can be registered as a new landmark through a triangulation method. Bundle adjustment [13] is used in the background to update the registered landmarks and thereby improve

the accuracy of their position and direction. The main steps of the method are the following.

1. Initialization. (section 4.2)
2. Detect 2D line features from an image using image processing. (section 4.3)
3. Match the detected 2D line features with 3D landmarks. To solve the P3L problem, correspondence between 3D-lines (landmarks) and 2D-lines (projection of landmarks) is necessary. In this study, detected line features are matched with those of the prior frame or key frames to ascertain which are 3D landmark projections. (section 4.4)
4. Estimate the camera pose using RANSAC-based method. (section 4.5)
5. Register new landmarks into a database. We register new landmarks from unknown environment. A triangulation method is applied to estimate the corresponding 3D-line of a 2D-line that is not matched with any landmark registered in the database. If the estimated 3D-line satisfies some constraints, meaning that it is sufficiently reliable, then it will be registered as a new landmark. (section 4.6)

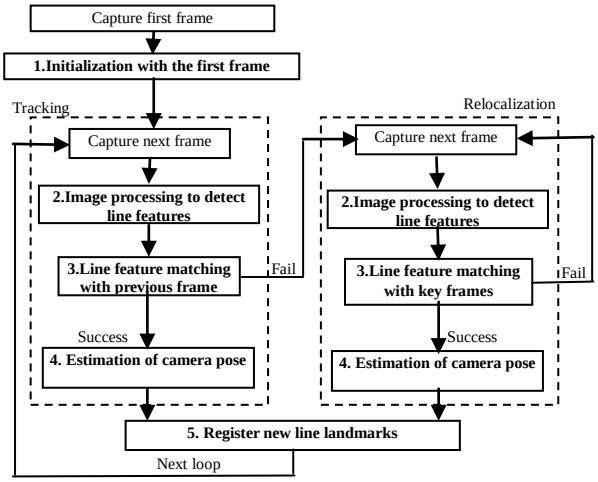


Fig.1 Flow chart of the proposed method.

4.2 Initialization

For this study, two rectangular markers, resembling those depicted in Figure 2, were used as initial landmarks. At the first frame, the corresponding 2D lines of the eight registered initial 3D landmarks must be extracted from the image. After binary processing of the image and detecting edge points (pixels with a large grayscale gradient) using Canny operator [14], a parallelogram is sought on the image. Then the extracted parallelograms are checked to ascertain whether they are the initial landmarks or not. As shown in Fig.3, we assume that the average grayscale of part A (black part) and part B (shadow part) are v_A and v_B, respectively. If $v_B - v_A > 50$, then four edges of the parallelogram are regarded as the corresponding lines of an initial marker.

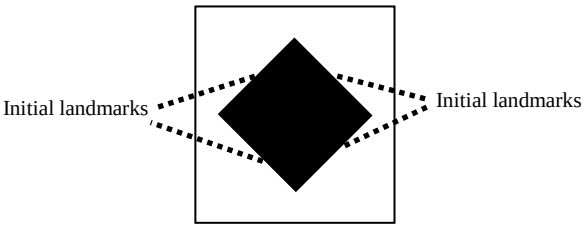


Fig.2 Rectangle marker.

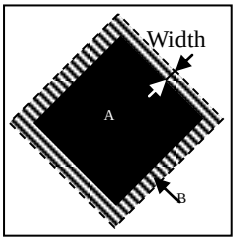


Fig.3 Calculation regions of rectangle marker.

4.3 Line Detection

Line features on an image are one kind of edge feature, which indicates pixels with a large gradient of grayscale. To ascertain the line features, the edge points are extracted from the image using image processing (undistortion, contrast enhancement, Gaussian smooth, Canny operator) first. Then the line feature can be detected using iterative end point fitting (IEPF) [15]. As shown in Fig.4, assuming a point group with two endpoints P_1 and P_2, a point P of the point group which has the longest distance d to the line P_1P_2 is sought. If d is larger than a threshold (2 pixels in this study), then divide the group into two group at P, and repeat this step on each. When d is smaller than the threshold, the corresponding group is then regarded as a line feature. Least-squares error optimization is applied to deduce the optimal line (dashed line ***l***). Finally, some lines might be mutually connected if the difference of their inclinations (<1 °in this study) and the distance (≤3 pixels in this study) between their endpoint are both very small.

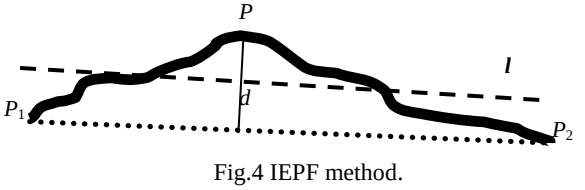


Fig.4 IEPF method.

4.4 Line Matching

4.4.1 Invariant moment

Because of the instability of Canny operator and IEPF, the same 2D line might be detected with different lengths at different frames. It is difficult to produce a template of the line with a determined size of a window as a matching patch. Therefore, a general patch method using normalized cross correlation for matching features is difficult to apply. The histogram invariant moment [16] of landmarks is used for tracking in this study. The invariant moment represents the average grayscale of an image. Therefore we need not define a window of fixed size and shape in advance. Moreover, it is invariant when the image is zoomed or rotated. The definition of the invariant moment of a line is the following.

For a 2D line, a window (dash frame) with 15 pixel width was defined as presented in Fig.5. If the absolute value of the slope to the x axis is greater than 1, then the width is along the x direction. Otherwise it is along the y direction.

In the window, n is the total number of pixels. And n_r is the number of pixels with grayscale r ($0 \leq r \leq 255$). The probability function $P(r)$ is defined as $P(r) = n_r/n$. Then the k-order moment m_k is $m_k=\sum r^k P(r)$, and the k-order central moment is $u_k=\sum(r-r_a)^k P(r)$, where $r_a=m_1/m_0$. Finally, invariant moment h is defined as $h=\eta_5/\eta_2\eta_3$, where $\eta_k= u_k/ u_0^{k+1}$.

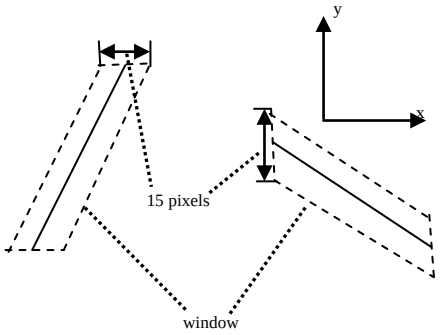


Fig.5 Calculation window of invariant moment.

4.4.2 Line matching with that of previous frame

When the camera moves slowly and smoothly, and the capturing frame rate is sufficiently high, the movement of a line between consecutive frames would be sufficiently small. It can therefore be matched in the previous frame in a small zone around the line, and will not be confused with other lines. In this study, the camera pose is estimated from the correspondence between a registered landmark and its projection at every frame. Therefore, after image processing, the corresponding 2D-lines of registered landmarks must be found from extracted line features. At the first frame, the initial correspondences are created as described in section 3.2. Then we can match every extracted line feature of the current frame with that of previous frame to find the correspondences between landmarks and detected 2D-lines at the current frame. As depicted in Fig.6, **l** and **l'** are two detected line segments (Thick solid line segments. Dashed lines are their extended lines) in the prior frame and current frame. If their difference of invariant moment is sufficiently small (<20% in this study), then the minimum distance d between two points located respectively on segments **l** and **l'** and the angle θ between them are checked. If they are both sufficiently small (d<20 pixel, θ<5 ° in this study), then **l'** is matched with **l**. If **l** is the corresponding 2D-line of a landmark, then **l'** will be the corresponding 2D-line of the landmark too. If more than one line is matched with **l**, then the line is considered more or less unreliable. Therefore, only the lines which have a single matched line are selected as matched lines. Following the method described above, the correspondence between a landmark and its projection are ascertained at every frame.

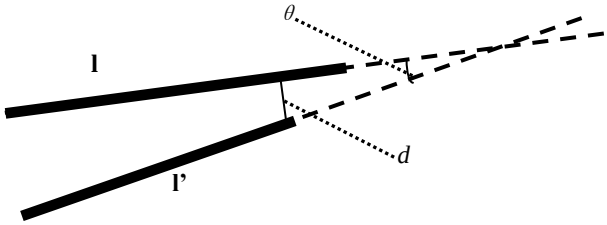


Fig.6 Line matching with that of previous frame.

4.4.3 Relocalization using key frames

If line matching with the previous frame fails (the number of 2D-lines matched with landmarks is ≤3), then the tracking also fails at the current frame. A relocalization method to estimate the camera pose is necessary after tracking failure. In this study, line features are matched between the current frame and some previous frames which are saved as key frames in the relocalization. In a key frame, the coordinates and invariant moments of the 2D lines which have been matched with the 3D landmarks registered in the database are saved, so the similarity of the lines

between the current and the key frame can be tested by checking the difference of their coordinates and invariant moments. If two lines have high similarity, they will be matched. So the matched line in the current frame corresponds to a 3D landmark in the database. Therefore, if more than three lines in the current frame are matched with the lines of the key frame, the camera pose can be estimated using RANSAC-based method (section 3.5), which is the same as with the tracking method.

A 2D infinite line can be decided uniquely using a point that is located on the line and with the smallest distance to the origin on the image. Therefore, the line can be represented by the point. The similarity between two lines can be test by checking the distance between their corresponding points. As shown in Fig.7, for every point $\mathbf{p}_i$ (i=1,2,3…) corresponding to a landmark in a key frame, a circle with radius of 40 pixels, and center at point $\mathbf{p}_i$ is defined. Then in the current frame, the points located in the circle are matched with point $\mathbf{p}_i$ by checking their difference of the invariant moment. If multiple points are matched with $\mathbf{p}_i$ as candidates, then the best point would be selected from the matched points by checking their distance from point $\mathbf{p}_i$ and other candidates located in other circles. Assuming two circles with center $\mathbf{p}_0$ and $\mathbf{p}_1$, with distance between them as $l=\|\mathbf{p}_0\mathbf{p}_1\|$, their respective candidates are $\mathbf{p}'_0$ and $\mathbf{p}'_1$, and $l_i=\|\mathbf{p}'_0\mathbf{p}'_1\|$. If the angle between the lines $\mathbf{p}_0\mathbf{p}_1$ and $\mathbf{p}'_0\mathbf{p}'_1$ is smaller than 5°, and $|\, l_i - l \,|$ is smaller than 20 pixels, then points $\mathbf{p}'_0$ and $\mathbf{p}'_1$ are regarded as matched respectively with $\mathbf{p}_0$ and $\mathbf{p}_i$. If multiple candidates still exist, then the point which is nearest to the center of the circle will be selected finally.

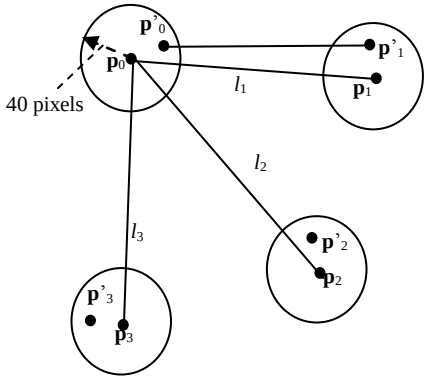


Fig.7 Line matching with that of key frame.

4.5 RANSAC-based Method for Solving P3L Problem

In general, the P3L problem has a maximum eight solutions. In this study, four equations was obtained from the P3L problem, in which each equation has a maximum of two solutions. Therefore, they are solvable in parallel to improve the efficiency. Details of the RANSAC-based method are the following steps:

1. Randomly choose three line correspondences (2D lines detected in the current frame and their corresponding 3D landmarks registered in the database).
2. Use correspondences to estimate the camera pose by solving the P3L problem.
3. Based on the result obtained from step 2, check the re-projection error of all line correspondences. The line correspondences for which the re-projection error is smaller than a threshold are inliers.
4. Repeat step 1 to step 3 by min(n,100) times, where n is the number of line correspondences. Choose the result with the maximum number of inliers.

4.5.1 Coordinate system

We use subscripts to represent a vector or matrix in different coordinate systems. For example, $\mathbf{t}_A$ is a vector in system A, and $\mathbf{t}_{AB}$ and $\mathbf{R}_{AB}$ respectively denote the translation vector and rotation matrix from system A to system B. Subscripts W and C respectively denote the world system and camera system.

4.5.2 P3L problem

It is extremely convenient to represent a 3D line using Plücker coordinates [17]. A Plücker line is the following, as presented in Fig.8.

$$\mathbf{L}_{6\times1}=\begin{pmatrix}\mathbf{n}\\ \mathbf{u}\end{pmatrix}=(n_1,n_2,n_3,u_1,u_2,u_3)^T \tag{1}$$

(1)

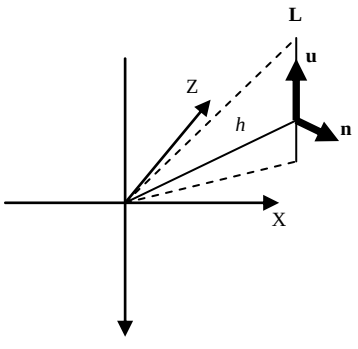


Fig.8 Plücker line.

Therein, $\mathbf{u}$ is the unit directional vector of the 3D-line and $\mathbf{n}=\mathbf{p}\times\mathbf{u}$, where $\mathbf{p}$ is an arbitrary point on the 3D line. $h=\|\mathbf{n}\|$ is the distance from origin to the 3D line. Considering the constraints as (2) and (3), only 4 degrees of freedom (DOF) exist.

$$\mathbf{n}\cdot\mathbf{u}=0 \qquad (2)$$

$$\|\mathbf{u}\|=1 \qquad (3)$$

The line projection is depicted in Figure 9. Assuming that $\mathbf{p}$ is an arbitrary point on line $\mathbf{L}$. For convenience, we define $\mathbf{p}=\mathbf{u}\times\mathbf{n}$. $\mathbf{v}$ is the unit normal vector of the plane (the shadow plane in Fig.9), which contains the origin of the camera system and the 3D-line. The 2D-projection of the line is also placed on this plane, so $\mathbf{v}$ is obtainable as $\mathbf{v}=\mathbf{a}\times\mathbf{b}$, where $\mathbf{a}$ and $\mathbf{b}$ are the two endpoints of the detected 2D line. The directional element of $\mathbf{L}$ and the vector $\mathbf{p}$ in the camera system are the following.

$$\mathbf{u}_C=\mathbf{R}_{WC}\mathbf{u}_W \qquad (4)$$

$$\mathbf{p}_C=\mathbf{R}_{WC}(\mathbf{p}_W-\mathbf{t}_{WC}) \qquad (5)$$

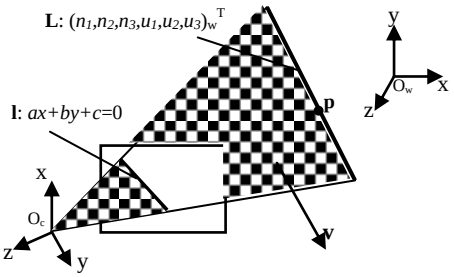


Fig.9 Line correspondence.

Because $\mathbf{u}_C\perp\mathbf{v}_C, \mathbf{p}_C\perp\mathbf{v}_C$, (6) and (7) are deduced as shown below.

$$(\mathbf{R}_{WC}\mathbf{u}_W)\cdot\mathbf{v}_C=0 \qquad (6)$$

$$\mathbf{R}_{WC}(\mathbf{p}_W-\mathbf{t}_{WC})\cdot\mathbf{v}_C=0 \qquad (7)$$

Equation (6) includes only three unknown parameters related to rotation. Therefore the rotation information is solvable first if there are three line correspondences when the three lines are not parallel. It is a P3L problem.

Assuming that the unit direction vectors of the three lines are $\mathbf{u}_{1W}$, $\mathbf{u}_{2W}$, and $\mathbf{u}_{3W}$, respectively, in the world system, then their corresponding unit vectors $\mathbf{v}_{1C}$, $\mathbf{v}_{2C}$, and $\mathbf{v}_{3C}$ in the camera system are obtainable from the two detected endpoints of their corresponding 2D lines and camera intrinsic parameters. For convenience, we define a local system A in which $\mathbf{v}_{1A}=(1,0,0)^T$, $\mathbf{v}_{2A}=(m,0,n)^T$, and $\mathbf{v}_{3A}=(r,s,t)^T$, where $\mathbf{v}_{2A}=\mathbf{R}_{CA}\mathbf{v}_{2C}$, $\mathbf{v}_{3A}=\mathbf{R}_{CA}\mathbf{v}_{3C}$, and $\mathbf{R}_{CA}=\left[\mathbf{v}_{1C}, \frac{(\mathbf{v}_{1C}\times\mathbf{v}_{2C})}{\|\mathbf{v}_{1C}\times\mathbf{v}_{2C}\|}, \frac{(\mathbf{v}_{1C}\times(\mathbf{v}_{1C}\times\mathbf{v}_{2C}))}{\|(\mathbf{v}_{1C}\times(\mathbf{v}_{1C}\times\mathbf{v}_{2C}))\|}\right]^T$. $\mathbf{v}$ is a unit vector. Therefore, it is readily apparent that $m^2+n^2=r^2+s^2+t^2=1$.

4.5.3. Method for solving P3L problem

In the proposed method, four independent equations were solved in parallel threads to improve the efficiency. It is assumed that $\mathbf{u}_{1A}=(0,\cos\theta,\sin\theta)^T$, $\mathbf{u}_{2A}=(x,y,z)^T$, and $\mathbf{u}_{3A}=(x',y',z')^T$. According to the geometry relation, $\mathbf{u}_{2A}$ and $\mathbf{u}_{3A}$ is solvable by being represented using the unknown parameter θ (there are a maximum two solutions for each of $\mathbf{u}_{2A}$ and $\mathbf{u}_{3A}$. Details are described in **Appendix 1**.) Then, according to the constraint $\mathbf{u}_{2A}\cdot\mathbf{u}_{3A}=\mathbf{u}_{2W}\cdot\mathbf{u}_{3W}=c$ (c is a constant decided by $\mathbf{u}_{2W}$ and $\mathbf{u}_{3W}$), a nonlinear Equation (8) with one unknown parameter θ is obtainable.

$$f(\theta)=\mathbf{u}_{2A}\cdot\mathbf{u}_{3A}-c_3=0 \qquad (8)$$

Equation (8) has four independent types (**Appendix 1**). Each type has a maximum of two solutions. Therefore they are solvable in parallel threads. The curve of one type of $f(\theta)=\mathbf{u}_{2A}\cdot\mathbf{u}_{3A}-c_3$ is depicted in Fig.10. The upper panel of Fig.10 presents a case in which both $\mathbf{u}_{2A}$ and $\mathbf{u}_{3A}$ have two solutions. The lower panel of Fig.10 is a more general case in which $\mathbf{u}_{2A}$ and $\mathbf{u}_{3A}$ would have no solution in some field. The shapes of the other three types are similar. From Fig.10, it is known that the maximum number of possible solutions of each type of Equation (8) is four. Because $f(\theta)=f(\theta+\pi)$, we can only consider two solutions in the field [0, π]. No analytical solution exists for Equation (8), so an iterative method is applied for solving the equation. In this study, the bisection method [18] is applied to search the solution of Equation (8) because it is an extremely simple and robust method.

After all solutions are solved, $\mathbf{u}_{1A}$, $\mathbf{u}_{2A}$ and $\mathbf{u}_{3A}$ are obtainable. Therefore, the rotation matrix is solved as Equation (9), where $\mathbf{R}_{WB}$ and $\mathbf{R}_{AB}$ is obtained as Equation (10) and Equation (11).

$$\mathbf{R}_{WC}=\mathbf{R}_{AC}\mathbf{R}_{WA}=\mathbf{R}_{CA}^T\mathbf{R}_{AB}^T\mathbf{R}_{WB} \qquad (9)$$

$$\mathbf{R}_{AB}=\left[\mathbf{u}_{1A}, \frac{(\mathbf{u}_{1A}\times\mathbf{u}_{2A})}{\|\mathbf{u}_{1A}\times\mathbf{u}_{2A}\|}, \frac{(\mathbf{u}_{1A}\times(\mathbf{u}_{1A}\times\mathbf{u}_{2A}))}{\|\mathbf{u}_{1A}\times(\mathbf{u}_{1A}\times\mathbf{u}_{2A})\|}\right]^T \qquad (10)$$

$$\mathbf{R}_{WB}=\left[\mathbf{u}_{1W}, \frac{(\mathbf{u}_{1W}\times\mathbf{u}_{2W})}{\|\mathbf{u}_{1W}\times\mathbf{u}_{2W}\|}, \frac{(\mathbf{u}_{1W}\times(\mathbf{u}_{1W}\times\mathbf{u}_{2W}))}{\|\mathbf{u}_{1W}\times(\mathbf{u}_{1W}\times\mathbf{u}_{2W})\|}\right]^T \qquad (11)$$

Then Equation (7) becomes a linear equation for translation $\mathbf{t}_{WC}$. After solving it, the best solution is chosen by checking the re-projection error of all solutions.

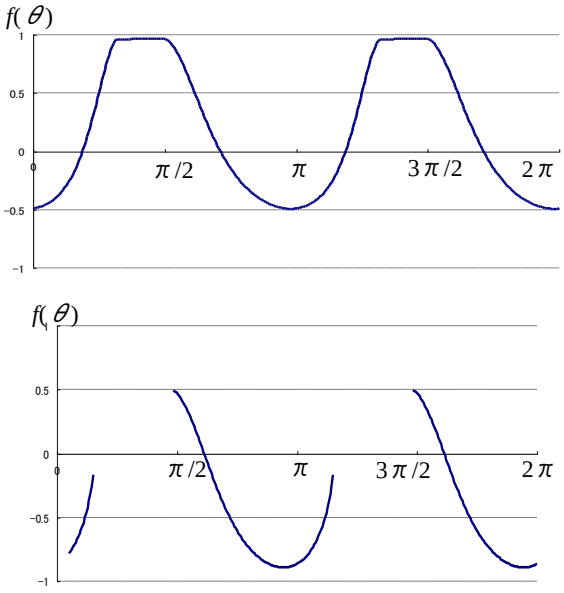


Fig.10 Curve of f(θ).

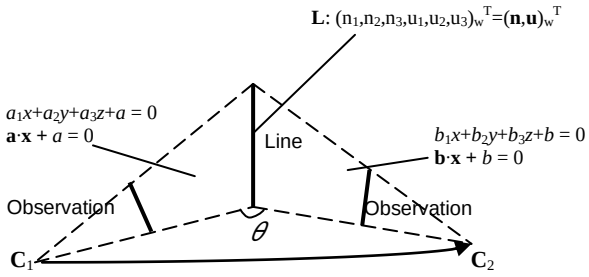
4.6 New Landmark Registration

If only initial landmarks are used for tracking, then it is impossible to conduct tracking in a capacious environment. In such cases, it is necessary to add new features as landmarks. A geometric method is applied to realize this function. In this method, every detected line feature on an image is a candidate or a registered landmark. As presented in Fig.11, two different points C_1 and C_2 denote the camera positions corresponding to different frames. The two planes passing through C_1 and C_2 and a 3D-line are represented respectively as $\mathbf{a}\cdot\mathbf{x}+a=0$ and $\mathbf{b}\cdot\mathbf{x}+b=0$. θ in Fig.11 is defined as the angle between two planes passing respectively through the 3D line and two different camera positions. The 3D landmark is obtainable from two different frames as Equation (12) and Equation (13):

$$\mathbf{u}=\mathbf{a}\times\mathbf{b}/\|\mathbf{a}\times\mathbf{b}\| \qquad (12)$$

$$\mathbf{n}=(a\mathbf{b}-b\mathbf{a})/\|\mathbf{a}\times\mathbf{b}\| \qquad (13)$$

The RANSAC method was applied to optimize vectors $\mathbf{n}$ and $\mathbf{u}$ of a new landmark by choosing a model with the most inliers. If a 2D line can be detected more than 20 times in 50 continuous frames, then two frames for which the 2D line is detectable were chosen randomly between which the movement distance of camera is sufficiently large (>40 mm), and a 3D line was estimated from the two frames. Then the re-projection error of the 3D line was checked in the 50 frames. As presented in Fig.12, the dashed line is a re-projection of a landmark, and solid is its corresponding 2D-line detected using IEPF. We defined the re-projection error as $e=d_1+d_2$, where d_1 and d_2 are the distance from the two endpoints of the corresponding 2D-line to the re-projection line. If the error at one frame is sufficiently small (<6 pixels), then this frame is regarded as a good frame. After repeating 50 times, and choosing the result of a 3D line that has the greatest number of good frames, then if the number of good frames is greater than 10, the 3D line would be registered as a new landmark using this result. It can contribute to solving the P3L problem.



θ: the change of view angle.

Fig.11 Geometry method for new landmark registration.

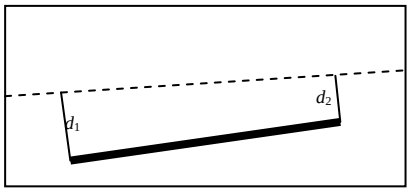


Fig.12 Re-projection error.

4.7 Accuracy Improvement of Landmarks

The registered landmarks might include some errors. To reduce the error of landmarks, bundle adjustment [13] is applied in a parallel thread to update the registered landmark. For this study, assuming n registered landmarks and m positions and orientations of the camera, bundle adjustment is defined as a problem to minimize the total reprojection error with respect to all 3D lines and camera parameters as Equation (14).

$$\min\sum_{j=1}^{m}\sum_{i=1}^{n}v_{ij}(d_{ij1}^2+d_{ij2}^2) \qquad (14)$$

In that expression, v_{ij} is 1 if the ith landmark is detected when the camera is at the jth position, otherwise v_{ij} equals 0. d_{ij1} and d_{ij2} are re-projection

errors of the *i*th landmark when the camera is at the *j*th position. d_{ij1} and d_{ij2} can be represented as a function with respect to $\mathbf{a}_j$ and $\mathbf{b}_i$, where $\mathbf{a}_j$ is a vector representing the *j*th camera pose, and where $\mathbf{b}_i$ is a Plücker vector representing the *i*th landmark. Then Equation (14) can be transferred as shown in Equation (15).

$$\min \sum_{j=1}^{m}\sum_{i=1}^{n} v_{ij}\left(f_1^{\,2}(\mathbf{a}_i,\mathbf{b}_j)+f_2^{\,2}(\mathbf{a}_i,\mathbf{b}_j)\right) \qquad (15)$$

If all $\mathbf{a}_j$ and $\mathbf{b}_i$ are connected as a vector $\mathbf{c}$, where $\mathbf{c}=(\mathbf{a}_1,\mathbf{a}_2,...,\mathbf{a}_m,\mathbf{b}_1,\mathbf{b}_2,...\mathbf{b}_n)^T$, then Equation (16) is obtainable as presented below.

$$\min \sum_{j=1}^{m}\sum_{i=1}^{n} v_{ij}\left(f_1^{\,2}(\mathbf{c})+f_2^{\,2}(\mathbf{c})\right) \qquad (16)$$

Using the Levenberg–Marquardt Algorithm [19], $\mathbf{c}$ is solvable to minimize the series sum of Equation (16). Therefore the registered landmarks are updated.

In this study, sparse bundle adjustment [20] is applied to reduce the computing cost. If all past frames are used every time, then the computing cost would become higher and higher. In our case, we only applied sparse bundle adjustment with a bundle of maximum of the last 50 frames.

5 Evaluation Experiment

5.1 Experimental Purpose

To evaluate the accuracy, stability, and speed of the proposed method, an evaluation experiment was conducted in a pure-water chamber at Fugen NPP.

5.2. Experimental Method

The experimental area is about 8 m × 9 m. Two rectangular markers were pasted in the environment in advance, as presented in Fig.13 (Circular markers were used to define the world system using an automatic marker registration system (MAMS) [21]. They were not used for tracking.). Their four edges were registered as initial landmarks (thick black lines). The edge length was about 20 cm. The positions of the four corners of the rectangle were measured using MAMS beforehand so that their Plücker coordinates were calculated. A small circular marker was pasted on the camera so that its position could be measured by MAMS. The tracking results were compared with the true data measured using MAMS. The experimental environment is depicted in Fig.14. The camera moved along the dash arrow approximately.

The experiment system includes a digital camera (IEEE-1394a; Point Grey Research Inc., Dragonfly2; Table 1) with about 2.1 mm focal length and a computer connected to it. The program was developed using software (Visual C++ 2008; Microsoft Corp.) to realize the proposed method described in Chapter 3. The capture frame rate of the camera was about 10 fps. The camera resolution is 640 × 480. When tracking, the camera was fixed on a tripod, and the tripod was moved in the environment, and stopped at some sample position. When the tripod was stopped, the true data of the camera pose at the sample position were measured using MAMS. The total path length of the camera movement was approximately 5 m. The distance between the camera and landmarks was about 2.5–5 m. To compare the results with true data measured using MAMS, we captured images when moving the camera and saved them on a hard disk. The program processed the images later offline.

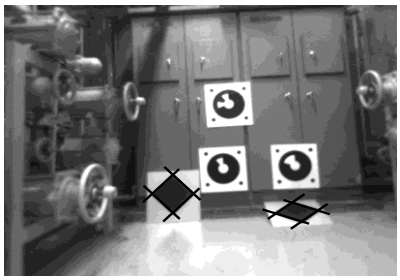


Fig.13 Rectangle markers in the environment.

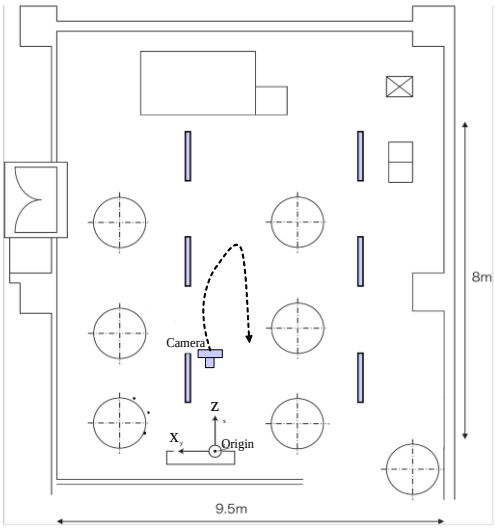


Fig.14 Experimental environment.

5.3 Results and Discussion

The CPU used for this study (Core i7 3930K; Intel Corp.) analyzed total 420 frames that had been

captured in this experiment. Ten camera positions were measured using MAMS. Therefore, the frame number between two adjacent measured positions was about 40. The distance between two adjacent sample positions is about 1 m. The number of registered landmarks was about 200.

Table 1 Specifications of the IEEE-1394a digital camera

Type	Dragonfly2-HIBW
Maker	Point Grey Research Inc.
Sensor	Sony 1/3" progressive scan CCDs, BW
Resolution	640 ×480
Frame Rates	10 fps
Interface	6-pin IEEE-1394a 400Mb/s interface
Pixel Size	4.65 μm square pixel

5.3 Results and Discussion

The CPU used for this study (Core i7 3930K; Intel Corp.) analyzed total 420 frames that had been captured in this experiment. Ten camera positions were measured using MAMS. Therefore, the frame number between two adjacent measured positions was about 40. The distance between two adjacent sample positions is about 1 m. The number of registered landmarks was about 200.

The average processing time of one frame is shown as Table 2. The average running speed of this method is about 10.4 frames per second in the main thread. The speed is sufficient to support some field works in NPP, such as indicating the state of some equipment, or navigating the moving direction of workers. The average cost of executing bundle adjustment once in the parallel thread is about 10 seconds. It is processed in background. Therefore, the speed of the main thread would not be influenced.

In the AR application, the tracking is conducted in real time, which means that the process time on one frame is limited. However, in the RANSAC method, the accuracy depends on the repeat times. From Table 2, it can be inferred that the speed of camera pose estimation in the existing method is about 50% of the proposed method. In other words, if processing one frame in same length of time, the repeat times in the existing method will be about 50% of the proposed method. Table 4 presents errors when the repeat times are 50% reduced in the RANSAC method for solving P3L problems. Compared with Table 3, the errors increased, and the relocalization succeeded later. Therefore, the RANSAC method is effective to improve the tracking result accuracy. Because the proposed line tracking method improves the computing speed for estimating the camera pose, the repeat times of RANSAC can be increased more than in existing methods, and the accuracy is improved.

Table 2 Average processing time of one frame (ms)

Process	Time
Undistortion	2.3
Contrast enhancement	2.2
Gaussian smooth	1.9
Line detection	3.0
Line matching	19.4
Pose estimation	38.6
New landmark registration	28.4
Total	95.8

The position and orientation errors of the camera pose estimation on the 10 sample positions are shown as Table 3. The error increases over time, but the error is sufficiently small (<60 mm) when the tracking time is short using the line-based tracking method. That means that the tracking cannot be applied as a primer tracking method until now, but by combination with the marker tracking method, it is applicable as an assistant tracking method, and reduces the needed marker number at cases which require lower accuracy and stability, for example, indicating the positions of some large component.

Table 3 Errors of pose estimation

No.	Position error (mm)	Orientation error (deg)
1	20.1	5.1
2	55.3	7.6
3	160.4	13.7
4	70.5	10.2
9	180.0	15.3
10	114.7	13.2

In the AR application, the tracking is conducted in real time, which means that the process time on one frame is limited. However, in the RANSAC method, the accuracy depends on the repeat times. From Table 2, it can be inferred that the speed of camera pose estimation in the existing method is about 50% of the proposed method. In other words, if processing one frame in same length of time, the repeat times in the existing method will be about 50% of the proposed method. Table 4 presents errors when the repeat times are 50% reduced in the RANSAC method for solving

P3L problems. Compared with Table 3, the errors increased, and the relocalization succeeded later. Therefore, the RANSAC method is effective to improve the tracking result accuracy. Because the proposed line tracking method improves the computing speed for estimating the camera pose, the repeat times of RANSAC can be increased more than in existing methods, and the accuracy is improved.

Table 4 Errors of pose estimation with half repeat times in RANSAC

No.	Position error (mm)	Orientation error (deg)
1	20.1	5.1
2	57.5	7.7
3	170.4	14.3
4	90.0	11.5
9	/	/
10	134.8	14.1

Figure 15 shows the estimated trajectory of the camera movement. The black curve is a smooth connection of ground truth data measured using MAMS at 10 sample positions. Blue points denote the estimated positions of a camera without using bundle adjustment. Red points present results with bundle adjustment. The tracking accuracy is improved when using bundle adjustment. The maximum error (the difference between estimated position and measured position of camera) is about 160 mm with bundle adjustment. The average error is about 100 mm. From the 160th frame to the last frame, most of the initial landmarks (> 4) are almost undetectable. The tracking failed from the 195th frame because insufficient 2D line features were matched with 3D landmarks, and relocalization succeeded from the 364th frame. The average speed of relocalization is about 5 fps. Purple points show estimated results after relocalization when tracking failures occurred. Furthermore, green points are corresponding results obtained with bundle adjustment. In this case, the accuracy is improved when using bundle adjustment. The average error is about 150 mm after relocalization succeeds. From Fig.15, it can be inferred that when the tracking was successful, the estimated result was stable and accurate in short distance tracking. Camera relocalization was achieved.

6 Conclusion

As described in this paper, a line-feature-based tracking method was attempted to be pracital used in an NPP. A RANSAC-based method for solving P3L problem to calculate the camera pose from correspondence lines and a geometry method to add new landmarks were applied in this method. The result of the evaluation experiment shows that estimated results are accurate in short distance tracking. The average error of the camera position is about 100 mm. In some cases of field work in NPP, the accuracy of the proposed method is sufficient to apply AR for supporting field work. For example, the navigation for workers to the destination allows accuracy with 1–2 m errors because the navigation only requires the correct indication of direction. In some cases that require higher accuracy and stability such as temporary placement and conveyance operation of the dismantled components, although the proposed method cannot be used directly, it reduces the preparation work by combination with the marker tracking method. In some cases that require extremely high accuracy, such as the maintenance of small equipment, the accuracy must be improved.

In future work, the line detection stability and the matching algorithm must be improved to make this method feasible for use in some more complex environments. Moreover, evaluation experiments of long-distance tracking in NPP environment will be conducted. The initial landmark registration must be improved to reduce the labor workload. Furthermore, it will be combined with marker tracking.

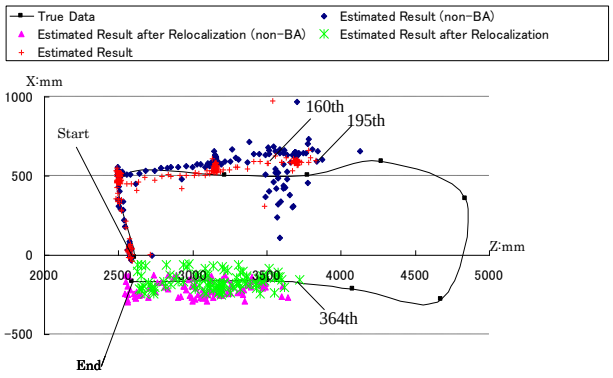


Fig.15 Estimated trajectory of the camera (XZ-plane).

Acknowledgement

This work was partially supported by KAKENHI (No. 23240016).

References

[1] FENG, Z., DUH, H., BILLINGHURST, M.: Trends in Augmented Reality Tracking, Interaction and Display: A Review of Ten Years of ISMAR. In: Proc. Seventh IEEE/ACM International Symposium on Mixed and Augmented Reality, Cambridge (UK), 2008.

[2] ISHII, H.: Augmented Reality: Fundamentals and Nuclear Related Application. International Journal of Nuclear Safety and Simulation. 2010, 1(3): 316–327.

[3] KATO, H., BILLINGHURST, M.: Marker Tracking and HMD Calibration for a Video-based Augmented Reality Conferencing System. In: Proc. Second IEEE/ACM International Workshop on Augmented Reality, San Francisco (USA), 1999.

[4] APPEL, M., NAVAB, N.: Registration of Technical Drawings and Calibrated Images for Industrial Augmented Reality. In: Proc. Fifth IEEE Workshop on Application of Computer Vision, California (USA), 2000.

[5] SKRYPNYK, I., LOWE, D.: Scene Modeling, Recognition and Tracking with Invariant Image Features. In: Proc. Third IEEE/ACM International Symposium on Mixed and Augmented Reality, Arlington (USA), 2004.

[6] GEORG, K., DAVID, M.: Parallel Tracking and Mapping for Small AR Workspaces. In: Proc. International Symposium on Mixed and Augmented Reality, Nara, 2007.

[7] BOSSE, M., RIKOSKI, R., LEONARD, J., TELLER, S.: Vanishing Points and 3d Lines From Omnidirectional Video. In: Proc. IEEE International Conference on Image Processing, New York (USA), 2002.

[8] DAILEY, M.N., PARNICHKUN, M.: Landmark-based Simultaneous Localization and Mapping with Stereo Vision. In: Proc. Asian Conference on Industrial Automation and Robotics, Bangkok (Thailand), 2005.

[9] SOLA, J., VIDAL-CALLEJA, T., DEVY, M.: Undelayed Initialization of Line Segments in Monocular SLAM. In: Proc. IEEE/RSJ International Conference on Intelligent Robots and Systems, St. Louis (USA), 2009.

[10] VABO, R., PIOTROWSKI, L., RINDAHL, G.: 3D Representation of Radioisotopic Dose Rates within Nuclear Plants for Improved Radioprotection and Plant Safety. International Journal of Nuclear Safety and Simulation, 2010, 1(2): 127–133.

[11] ISHII, H., SHIMODA, H., NAKAI, T., IZUMI, M., BIAN, Z., MORISHITA, Y.: Proposal and Evaluation of a Supporting Method for NPP Decommissioning Work by Augmented Reality. In: The 12th World Multi-Conference on Systemics, Cybernetics and Informatics, Florida (USA), 2008.

[12] YAN, W., AOYAMA, S., ISHII, H., SHIMODA, H., SANG, T., INGE, S., LYGREN, T., TERJE, J., IZUMI, M.: Development and Evaluation of a Temporary Placement and Conveyance Operation Simulation System Using Augmented Reality. Nuclear Engineering and Technology, 2012, 44(5): 507–522.

[13] TRIGGS, B., MCLAUCHLAN, P., HARTLEY, R., FITZGIBBON, A.: Bundle Adjustment – a Modern Synthesis. In: Proc. the International Workshop on Vision Algorithms: Theory and Practice, Confu (Greece), 1999.

[14] LINGEBERG, T.: Edge Detection and Ridge Detection with Automatic Scale Selection. In: Proc. IEEE Computer Society Conference on Computer Vision and Pattern Recognition, San Francisco (USA), 1996.

[15] CHOI, Y., LEE, T., OH, S.: A Line Feature Based SLAM with Low Grade Range Sensors Using Geometric Constraints and Active Exploration for Mobile Robot. Autonomous Robot, 2008, 24(1): 13–27.

[16] LIU, H., ZHU, S., WANG, X., WANG, H.: Research on Matching of the Remote Sensing Image Matching Based on Histogram Invariant Moments and Genetic Algorithms. In: IEEE International Conference on Information Engineering and Computer Science, Wuhan (China), 2009.

[17] HOHMEYER, M., TELLER, S.: Determining the Lines through Four Lines. Journal of Graphics Tools, 1999, 4(3): 11–22.

[18] RICHARD, L., DOUGLAS, J.: Numerical Analysis (3rd ed.). PWS Publishers; 1985: 28–31.

[19] MORE, J.: The Levenberg–Marquardt Algorithm: Implementation and Theory. Numerical analysis, 1978, 630: 105–116.

[20] LOURAKIS, M., ARGYROS, A.: SBA: A Software Package for Generic Sparse Bundle Adjustment. ACM Transactions on Mathematical Software, 2009, 36(1): 1–30.

[21] YAN, W., YANG, S., ISHII, H., SHIMODA, H., IZUMI, M.: Development and Experimental Evaluation of Automatic Marker Registration System for Tracking of Augmented Reality. International Journal of Nuclear Safety and Simulation, 2010, 1(1): 52–62.

Appendix 1. Solutions of $\mathbf{u}_{2A}$ and $\mathbf{u}_{3A}$

According to the geometry relation, we can obtain Equations (1) and (2), where c_1 is a constant decided by $\mathbf{u}_{1W}$ and $\mathbf{u}_{2W}$, as shown below.

$$\mathbf{u}_{1A} \cdot \mathbf{u}_{2A} = \mathbf{u}_{1W} \cdot \mathbf{u}_{2W} = c_1 \tag{1}$$

$$\mathbf{u}_{2A} \cdot \mathbf{v}_{2A} = 0 \tag{2}$$

From Equation (1) and Equation (2), consider that $\|\mathbf{u}_{2A}\|$=1; then *x*, *y*, and *z* are solvable respectively as (3), (4), and (5). Two maximum solutions exist (double sign in same order).

$$x = \frac{-mnc_1 \sin\theta \mp n\cos\theta\sqrt{1-n^2\sin^2\theta - c_1^2}}{1-n^2\sin^2\theta} \tag{3}$$

$$y = \frac{c_1\cos\theta \mp m\sin\theta\sqrt{1-n^2\sin^2\theta - c_1^2}}{1-n^2\sin^2\theta} \tag{4}$$

$$z = \frac{m^2 c_1 \sin\theta \pm m\cos\theta\sqrt{1-n^2\sin^2\theta - c_1^{\,2}}}{1-n^2\sin^2\theta} \quad (5)$$

Similarly, $\mathbf{u}_{3A} = (x', y', z')^T$ is solvable as (6), (7), and (8), where c_2 is a constant decided by $\mathbf{u}_{1W}$ and $\mathbf{u}_{3W}$. Two maximum solutions exist (double sign in same order).

$$x' = \frac{\mp(s\sin\theta - t\cos\theta)\sqrt{1-(t\sin\theta + s\cos\theta)^2 - c_2^{\,2}}}{1-(t\sin\theta + s\cos\theta)^2} - \frac{c_2 r(s\cos\theta + t\sin\theta)}{1-(t\sin\theta + s\cos\theta)^2} \quad (6)$$

$$y' = \frac{\pm r\sin\theta\sqrt{1-(t\sin\theta + s\cos\theta)^2 - c_2^{\,2}}}{1-(t\sin\theta + s\cos\theta)^2} + \frac{c_2((1-s^2)\cos\theta - st\sin\theta)}{1-(t\sin\theta + s\cos\theta)^2} \quad (7)$$

$$z' = \frac{\mp r\cos\theta\sqrt{1-(t\sin\theta + s\cos\theta)^2 - c_2^{\,2}}}{1-(t\sin\theta + s\cos\theta)^2} + \frac{c_2((1-t^2)\sin\theta - st\cos\theta)}{1-(t\sin\theta + s\cos\theta)^2} \quad (8)$$