Flexible Marker-based Augmented Reality Based on Estimation of Object Pose With RGB-D Sensor

Naoki Ibe
Department of Precision Engineering
The University of Tokyo
Tokyo, Japan
Email: ibe@robot.t.u-tokyo.ac.jp

Atsushi Yamashita
Department of Precision Engineering
The University of Tokyo
Tokyo, Japan
Email: yamashita@robot.t.u-tokyo.ac.jp

Hajime Asama
Department of Precision Engineering
The University of Tokyo
Tokyo, Japan
Email: asama@robot.t.u-tokyo.ac.jp

***Abstract*—Augmented Reality (AR) is one of the important technologies in the field of computer graphics and is utilized for many applications. When marker-based AR systems are utilized for applications in maintenance of plants, markers need to be allowed to be placed at any points on the target objects because of a lot of occlusions by the complicated piping network in plants. Therefore, we propose a method to overlay the virtual graphic models exactly onto the corresponded physical objects in an RGB image, by the extension of the marker-based AR. In our proposed method, the camera pose is tracked by the detection of the marker on the target object from the RGB image. It is also possible to recognize the object by using the information included in the marker. On the other hand, the 3D pose of the objects is estimated based on the alignment of 3D point clouds instead of marker-based AR method. We present the validity of the proposed method by conducting experiments with an RGB-D sensor and physical objects on which a marker is installed.**

I. Introduction

Augmented Reality (AR) is one of the important technologies in the field of computer graphics and is utilized for many applications. Among those applications, the demand to use AR for applications in maintenance of plants has been growing in these years. Even though inspectors must have expert skill to satisfactorily perform the maintenance in power plants, it is difficult to keep this level of expertise in the staff because of the difficulty of passing on these skills to non-expert and the reduction of veterans due to their aging in recent years. In order to solve this problem, it is required to develop support systems for non-expert inspectors using AR technologies.

When an AR system is utilized in the maintenance of nuclear power plants, the precise registration of virtual imagery is essential, because there are millions of equipments in nuclear power plants. Therefore, it is difficult for non-expert inspectors to find the correspondences. Many AR applications in maintenance that address this issue have been presented in recent years. Sczwald and de Lavel developed an AR system to assist maintenance using an optical see-through Head Mounted Display (HMD) [1]. Leutert and Schilling proposed an AR system to support telemaintenance in which inspectors use robot manipulators to maintain the plant remotely [2]. These systems can improve the safety and efficiency of maintenance based on AR, but these systems use special 3D positioning sensors, such as an electromagnetic tracker and the robot's odometory. As a result, they are difficult and impractical to apply in maintenance by human inspectors in large environments, such as the nuclear power plants.

Therefore, we develop an AR system using cameras, which are able to obtain large amounts of information and easy to use by inspectors during the patrol. To overlay virtual imagery on the real world using AR, one of the most important thing to be considered is the precise registration of virtual objects to the real world to keep the geometric consistency between them. Therefore, we propose a method for vision-based registration in AR, for its application in nuclear power plant's patrol.

II. Related Work

In recent years, a significant number of approaches have been developed in vision-based AR to address the issue of the camera tracking and virtual object registration. Present methods of AR can be classified into two categories: marker-based AR and marker-less AR.

In the marker-based AR, some markers are installed in the real world, and the estimation of camera pose and the registration of virtual imagery are performed based on these markers. ARToolKit [3] and Chilitags [4] utilize 2D images as markers. This approach allows the stable camera pose tracking, and object recognition by adding information of object to the marker. Olson [5] develops *AprilTag*, which is a marker-based AR system with the greater robustness to occlusion, warping, and lens distortion. However, these studies have difficulties to perform precise registration of the virtual imagery in plants because a lot of occlusions do not always allow some equipments to be seen in a desired appearance due to the plant's complicated piping network. As a result, the points where the markers should be placed based on a priori information cannot sometimes be seen from inspectors. Hence markers need to be allowed to be placed at any points on the target objects.

On the other hand, no markers are utilized in the marker-less AR. One of the major methods of marker-less AR is feature-based AR [6],[7]. This is based on feature detection and tracking algorithms from RGB images or 3D point clouds. Localized features are divided into three categories: feature points (*e.g.* FAST corner, SIFT), feature descriptors (*e.g.* SIFT), and edges. Camera pose can be estimated by applying some alignment algorithm, for instance Iterative Closest Point (ICP) [8], to the set of these features. This kind of methods does not require a priori information such as markers or 3D

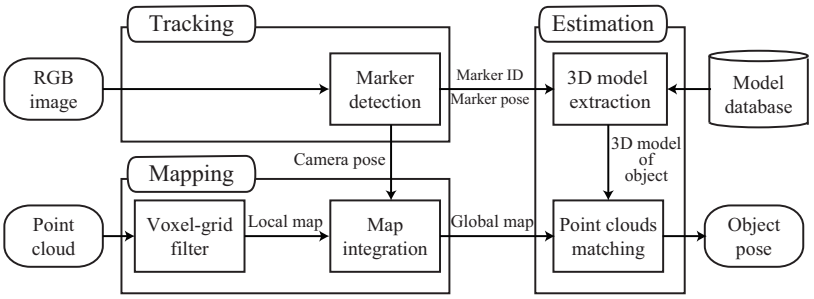


Fig. 1. Overview of our proposed method. Camera pose is tracked based on markers placed on objects in the *Tracking* part, and a global map is built by the integration of point clouds in the *Mapping* part. Finally, in *Estimation* part, the object's pose is estimated by the alignment of the global map and the object's shape model that has been preregistered in a database.

shape models, therefore, no preparation is required and it is possible to align virtual objects to the real world in accordance with their actual shape. However, this method obtains localized feature information so that it cannot recognize objects without any supplementary algorithm.

Marker-based AR, as described before, have a good potential to construct AR systems assisting maintenance in the plant without any special devices other than a camera. However, these methods have difficulties to perform precise registration and the object recognition simultaneously in nuclear power plants because a lot of occlusions do not allow inspectors to always see some equipments in a desired appearance because of the plant's complicated piping network. As a result, the points where the markers should be placed based on a priori information cannot sometimes be seen from inspectors in conventional marker-based AR. Therefore, in this paper, we propose a flexible marker-based AR to overlay the virtual object onto the real scene with the RGB-D camera even if the markers are not placed at the points that are set in advance.

III. Proposed method

A. Overview

In this paper, we propose a method to overlay the virtual graphic model exactly onto the corresponding physical object in an RGB image by a combination of marker-based and marker-less AR. To estimate the object pose, first we capture RGB images and calculate 3D point clouds from an RGB-D sensor. A marker is placed on the target object for camera tracking and object recognition. As shown in Fig. 1, the proposed method is divided into three parts, *Tracking*, *Mapping*, and *Estimation*. In the *Tracking* part, the camera pose is tracked by the detection of the marker on the target object from the RGB image. It is also possible to recognize the object by using the information included in the marker. In the *Mapping* part, a global map is built by integrating local maps built by filtering the voxel grid from the 3D point cloud. The integration is performed with the camera pose that is calculated in the *Tracking* part. In the *Estimation* part, the object pose is estimated by the alignment of the object point cloud that is extracted from the global map and the pre-registered 3D shape model of the target object. These processes automatically adjust the 3D shape model to the real pose.

B. Camera tracking based on marker pose

This research considers a large static environment such as a nuclear power plant in which there are few dynamic objects, and hence it can be assumed that markers placed on the target objects never move. According to this assumption, camera pose in the world coordinates can be calculated based on the marker pose in camera coordinates. First, marker pose in camera coordinates at frame t, denoted as $\mathbf{P}_t^m$, is calculated using an AR library. Camera motion in the world coordinates can then be obtained from the marker poses at two consecutive frames, frame t and $(t+1)$, as follows:

$$\mathbf{P}_t^m = \mathbf{M}_{t+1}^c \mathbf{P}_{t+1}^m, \tag{1}$$

where $\mathbf{M}_{t+1}^c$ denotes the matrix corresponding to the camera motion in the world coordinates between frame t and $(t+1)$. By considering the initial camera pose as the world coordinates, the camera pose in the world coordinates can be calculated at each frame. As the marker pose, the transformation matrix $\mathbf{T}_t$ from camera coordinates to the world coordinates is considered as the camera pose at frame t as follows:

$$\mathbf{T}_t = \prod_{\tau=1}^{t} \mathbf{M}_\tau, \tag{2}$$

Hence, camera pose is obtained in each frame from the marker pose in camera coordinates at each frame. As a result, camera tracking in real time can be performed by the proposed method.

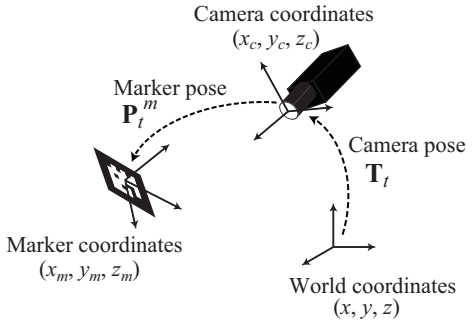


Fig. 2. Camera pose tracking. In this paper, camera and marker pose are defined as the transformation matrix that expresses the transformation between the two coordinates. Camera pose is estimated based on the assumption that markers on objects never move.

C. Map integration with 3D point cloud from RGB-D sensor

In the *Mapping* part, using the RGB-D sensor, the local 3D point cloud in the camera coordinates is generated

from the depth image which includes the depth information for each pixel. According to this information, the 3D point $\mathbf{p} = [x, y, z]^{\mathrm{T}}$ corresponding to a pixel (u, v) on the image can be reconstructed as follows:

$$x = \frac{u - c_x}{fd}, \tag{3}$$

$$y = \frac{c_y - v}{fd}, \tag{4}$$

$$z = d, \tag{5}$$

where d denotes the depth of the pixel (u, v), f denotes the focal length of the camera and (c_x, c_y) is the position of the principal point of the image. The generated point cloud includes too many points to perform any calculation in real time, therefore, the size of the cloud must be reduced for the subsequent processing.

To reduce the number of points, a voxel grid filter is applied to the generated point cloud in our proposed method. A voxel grid filter is a down-sampling of the RGB-D point cloud. At each frame, the space is divided into a set of voxels (volumetric pixels) and all points inside each voxel are approximated with the coordinates of their centroid. After voxel grid filtering, the filtered cloud is transformed into the world coordinates with the camera pose. Each point $\mathbf{p}_t^{\mathrm{voxel}}$ in the filtered cloud is integrated with the global map using the transformation matrix $\mathbf{T}_t$ that indicates the camera pose as follows:

$${}^{G}\mathbf{p}_t^{\mathrm{voxel}} = \mathbf{T}_t \mathbf{p}_t^{\mathrm{voxel}}, \tag{6}$$

where ${}^{G}\mathbf{p}_t^{\mathrm{voxel}}$ denotes the point in the world coordinates.

By integrating these transformed clouds, a global map of the sensing area is built and utilized to estimate the object pose, as described in the next section.

D. Estimation of object pose based on point cloud alignment

In the *Estimation* part, the object pose is estimated using the alignment of the global map that represents the real scene and the 3D shape model corresponding to the object. These 3D shape models are preregistered in the database and associated with the marker on the target objects. In this research, ICP is utilized to align the global map and the point cloud, which is down-sampled on the surface of the 3D shape model of the object. However, this algorithm has a high computational cost and takes too long to finish its calculations, hence it cannot be used in real time. Moreover, if the initial position of the alignment is remote to the ground truth, the performance of ICP is so poor that the solution can converge to a local-minima.

Therefore, in this paper, the area of alignment is limited to the neighborhood of the marker. This restriction is valid because the target object is always close to the marker, as the marker is placed on the object. The point cloud for ICP, called the *object cloud*, is extracted by the selection of points, as shown in Fig. 3. Here, $\mathbf{p} = [x, y, z]^{\mathrm{T}}$ denotes a point in the global map, and $\mathbf{p}_m = [x_m, y_m, z_m]^{\mathrm{T}}$ is the marker's 3D position obtained from the *Tracking* part. The object cloud Ω consists of the points that comply with the relationship as follow:

$$\Omega \equiv \{p \mid \|\mathbf{p} - \mathbf{p}_m\| \leq r\} \tag{7}$$

where, r denotes the maximum distance for extracting points as the object cloud. An example of ICP alignment for an object cloud and 3D shape model is shown in Fig. 4. This represents an alignment of a cube object. The left image shows the point cloud before the alignment, and the right image is its result. While the points extracted from the object's shape model, represented as green points, do not match the object cloud before the alignment, the aligned points, represented as red points, are adjusted to the points corresponding to the physical object by ICP. By extracting the object cloud, the number of points is reduced to less than 10000 and the alignment converges in less than a second.

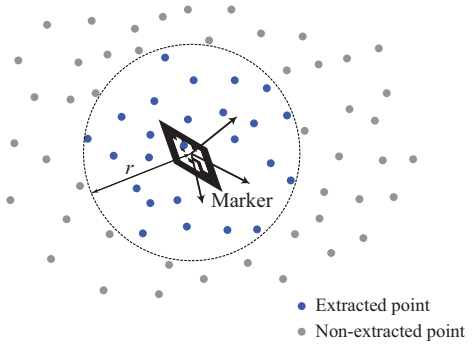


Fig. 3. Extraction of the object cloud from the global map. The points in the neighborhood of the markers are selected as the object cloud.

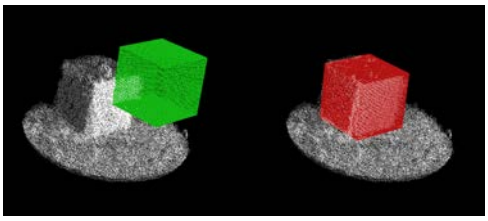


Fig. 4. Point cloud alignment using ICP. Before the alignment, the points extracted from the object's shape model (green points) do not match the object cloud (left image). By applying ICP to these points, the aligned points (red points) are adjusted to the points corresponding to the physical object (right image).

IV. Experiment

We performed an experiment to show the validity of our method. In this experiment, the virtual object, a 3D shape model, is overlaid onto the physical object's RGB image obtained from an RGB-D sensor.

A. Experimental setup

In this experiment, we utilized a Xtion PRO LIVE from ASUS as the RGB-D sensor and OpenNI, an open source library, to implement our proposed method. The Chilitags, which is an AR library, was utilized to track the camera pose and recognize the markers. This library allows stable tracking and recognize markers robustly.

In this experiment, a cube object upon which a Chilitags marker has been placed was utilized as the target object. This cube is 10 cm on each side and the 3D shape model of this object was preregistered in a database. The user holds the RGB-D sensor and portable display and views the target object through the display during the experiment, as shown in Fig. 5. By applying our proposed method to the image obtained from the RGB-D sensor, the pose of the cube object can be estimated

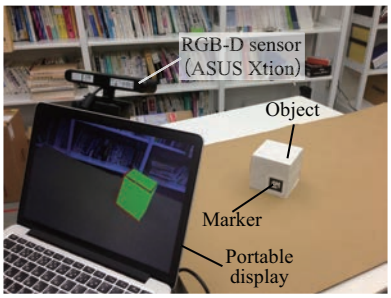


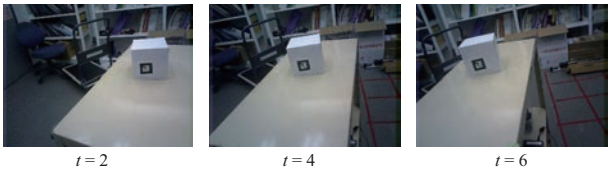
Fig. 5. Experimental setup. An RGB-D sensor for obtaining RGB images and 3D point clouds. An Object on which a marker has been placed is displayed on a portable display with overlaid virtual objects.

and the 3D shape model is overlaid exactly on the object in the RGB image. The pose of the marker in object local coordinates is varied in each case. In case 1, the marker was placed to ensure that the each side of the marker is parallel to the sides of the cube. In case 2, it was installed to enable rotation around the normal line of the cube surface.

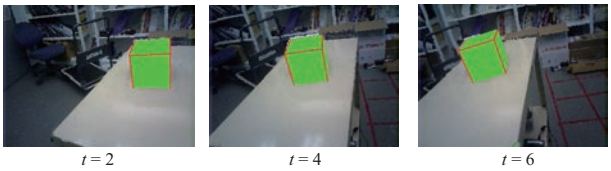
B. Results

Figure 6 shows the result of case 1. Figure 6(a) is the RGB image of a frame during the experiment and the image with an exact overlay of the 3D shape model is shown in Fig. 6(b). The result of case 2, is shown in Fig. 7. Figure 7(a) is the RGB image of a frame during the experiment and the image with an exact overlay of the 3D shape model is shown in Fig. 7(b).

In both cases, it is shown that the 3D shape models are overlaid exactly on the physical objects in the real environment, even if the marker pose is changed in each case. Using not only the marker pose on the object, but also the point cloud of the object, it is possible to estimate the actual pose of the physical object and overlay the virtual object exactly on the real object, despite any variation of the marker pose.



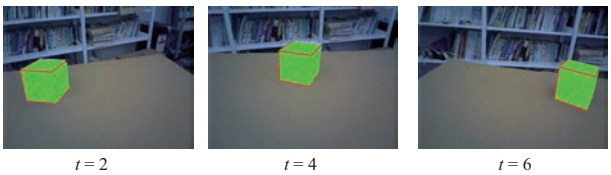
(a) RGB image



(b) Overlay

Fig. 6. Case 1: the marker is placed at the bottom left on the front surface to ensure that the each side of the marker is parallel to the sides of the cube.

(a) RGB image



(b) Overlay

Fig. 7. Case 2: the marker is installed at the bottom right on the front surface for rotation around the normal line of the cube surface

V. Conclusion

In this paper, we proposed a method to exactly overlay the virtual objects on the real world with a combination of object recognition using markers and the alignment of 3D point clouds. In the proposed method, the RGB image and the 3D point cloud were obtained from an RGB-D sensor. The camera pose was tracked based on the marker's pose in the RGB image, and the object was recognized based on the pattern of the marker. We also demonstrated that it is possible to overlay the virtual object on the real object in the RGB image with the point cloud alignment, even if the pose of the marker has some freedom in the object local coordinates. As a future work, we plan to extract the difference from the previous image to the application to the real environment of a power plant.

References

[1] B. Schwald and B. de Laval: "An Augmented Reality System for Training Assistance to Maintenance in the Industrial Context," *Journal of WSCG*, Vol.11, No.3, pp.425-432, 2003.

[2] F. Leutert and K. Schilling: "Support of Power Plant Telemaintenance with Robots by Augmented Reality Methods," *Proceedings of the 2nd International Conference on Applied Robotics for the Power Industry (CARPI2012)*, pp.45-49, 2012.

[3] H. Kato and M. Billinghurst: "Marker Tracking and HMD Calibration for a Video-based Augmented Reality Conferencing System," *Proceedings of the 2nd International Workshop on Augmented Reality (IWAR 99)*, pp.85-94, 1999.

[4] Q. Bonnard, S. Lemaignan, G. Zufferey, A. Mazzei, S. Cuendet, N. Li, and P. Dillenbourg: "Chilitags2: Robust Fiducial Markers for Augmented Reality and Robotics," CHILI, EPFL, Switzerland, http://chili.epfl.ch/software, 2013.

[5] E. Olson: "AprilTag: A Robust and Flexible Visual Fiducial System," *Proceedings of the 2011 International Conference on Robotics and Automation (ICRA2011)*, pp. 3400-3407, 2011.

[6] G. Klein and D. Murray: "Parallel Tracking and Mapping for Small AR Workspaces," *Proceedings of the 6th IEEE and ACM International Symposium on Mixed and Augmented Reality (ISMAR2007)*, pp.225-234, 2007.

[7] R. A. Newcombe, S. Izadi, O. Hilliges, D. Molyneaux, D. Kim, A. J. Davison, P. Kohi, J. Shotton, S. Hodges, and A. Fitzgibbon: "KinectFusion: Real-time Dense Surface Mapping and Tracking," *Proceedings of the 10th IEEE International Symposium on Mixed and Augmented Reality (ISMAR2011)*, pp.127-136, 2011.

[8] S. Rusinkiewicz and M. Levoy: "Efficient Variants of the ICP Algorithm," *Proceedings of the International Conference on 3D Digital Imaging and Modeling*, pp.145-152, 2001.