

# Registration for Stereo Vision-based Augmented Reality Based on Extendible Tracking of Markers and Natural Features

Masayuki Kanbara †, Naokazu Yokoya † and Haruo Takemura ‡

<sup>1</sup>Graduate School of Information Science, Nara Institute of Science and Technology

8916-5 Takayama-cho, Ikoma-shi, Nara, 630-0101, JAPAN

{kanbara, yokoya}@is.aist-nara.ac.jp

‡Cybermedia Center, Osaka University

takemura@cmc.osaka-u.ac.jp

## Abstract

*This paper describes a method to extend the registration range of a vision-based augmented reality (AR) system. We propose to use natural feature points contained in images captured by a pair of stereo cameras in conjunction with pre-defined fixed fiducial markers. The system also incorporates an inertial sensor to achieve a robust registration method which can handle user's fast head rotation and movement. The system first uses pre-defined fiducial markers to estimate a projection matrix between real and virtual coordinate systems. At the same time, the system picks up and tracks a set of natural feature points from the initial image. As a user moves around in an AR environment, the initial markers fall out from the camera frame and natural features are then used to recover the projection matrix. Experiments for evaluating the feasibility of the method are carried out and show the potential benefits of the method.*

## 1. Introduction

Augmented reality (AR) produces an environment in which virtual objects are superimposed on user's view of the real environment. Augmented reality has received a great deal of attention as a new method for displaying information or increasing the reality of virtual environments. A number of applications have already been proposed and demonstrated [1, 2, 3]. To implement an augmented reality system, we must solve some problems. Geometric registration is especially the most important problem because virtual objects should be superimposed on the right place as if they really exist in the real world.

One of the major approaches to the registration between the real and virtual worlds is the vision-based method [4, 5, 6, 7, 8]. The methods, which are sometimes referred to as vision-based tracking or registration, estimate the position and an orientation of user's viewpoint from images

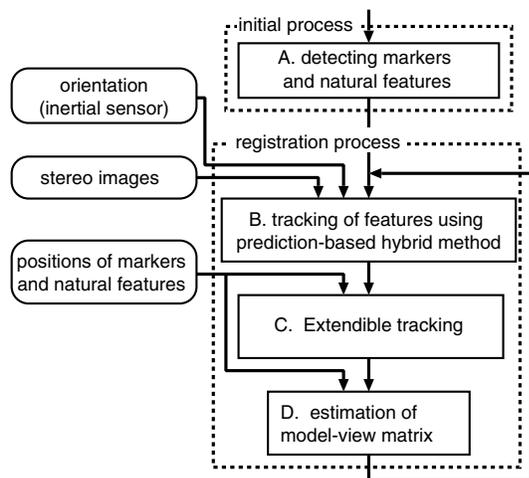
captured by a camera attached at the user's viewpoint. Because the method usually uses fiducial markers placed in the environment, the measurement range is actually limited.

To overcome this limitation, we propose a stereo vision-based augmented reality with a wide range of registration, which is based on using pre-defined fiducial markers and natural features. The system first uses pre-defined fiducial markers to estimate a projection matrix between real and virtual coordinate systems. At the same time, the system picks up and tracks a set of natural feature points from the initial image. As a user moves around in an AR environment, the initial markers fall out from the camera frame and natural features are then used to recover the projection matrix. Even if the markers and natural features fall out from the frame, the method can achieve the registration between the real and virtual environments by extendibly detecting new markers and natural features in sequence. In addition, to realize robust feature tracking, we apply a prediction-based hybrid method[9] which estimates the positions of markers and natural features with support of an inertial sensor.

The following part of the paper is structured as follows. Section 2 describes the stereo vision-based registration method using both markers and natural feature points. In Section 3, experimental results with the proposed method and discussion about a prototype system are described. Finally, Section 4 summarizes the present work.

## 2. Geometric Registration by Tracking Markers and Natural Features

We assume in this study that a pair of stereo cameras are virtually located at viewer's two eyes in an augmented reality system. Figure 1 illustrates the flowchart of the proposed method. First, to memorize positions of markers and natural features, the markers and natural features, both of which may be simply called features hereafter, are detected



**Figure 1. Flow diagram of the registration method.**

from a pair of stereo images (A in Fig. 1). The features are tracked in the subsequent images using a prediction-based hybrid method[9] with a stereo vision technique and an inertial sensor (B in Fig. 1). Next, the tracking results are evaluated because tracking of natural features contains errors and new features are then detected from the stereo images (C in Fig 1 ). Finally, the results above are used for estimating a model-view matrix which represents a relationship between the real and virtual coordinate systems (D in Fig. 1).

## 2.1. Detecting positions of markers and natural features

In the first frame, markers and natural features are detected from a pair of stereo images. The markers are detected by color matching [4]. On the other hand, the natural features are detected by using Moravec's interest operator [10]. The interest operator can detect characteristic points as natural features that are easily matched between the two consecutive images.

In the proposed method, the natural features are detected from every evenly spaced region by the interest operator to distribute detected features evenly in the images. Next, each natural feature detected from the left image is matched with a corresponding point in the right image using a standard template matching technique with the epipolar constraint. Finally, the 3D positions of the natural features are calculated by triangulation.

## 2.2. Tracking of features

In this section, the tracking of markers and natural features with the prediction-based hybrid method[9], which predicts position of features using an inertial sensor, is briefly described.

**2.2.1. Tracking of markers.** This section describes two cases of features: inside or outside of the current frame. In the case where the predicted position is inside of the image, a search window is determined by the predicted position. Then, the markers are tracked by detecting marker's region in the search window based on color information. Next, the 3D positions of the markers are calculated by a stereo matching algorithm. Note that the farthest marker is tracked without the predictions[9].

In the case where the predicted position is outside of the image, the tracking is realized by assuming that the predicted position is correct in the current frame. Therefore, the markers can be tracked continuously even when markers that have once gone outside come back into sight again.

**2.2.2. Tracking of natural features.** The natural features are tracked by using a standard template matching technique applied to two consecutive images. Note that a template is made from a neighboring region of the natural feature in the previous frame and the similarity measure is a normalized cross correlation. When the cameras rotate in roll direction, the rotation of the template is considered in the matching process. When the position of natural feature in the image is close to the frame of the image, the tracking of the natural feature is discontinued. The 3D positions of natural features that are tracked successfully are calculated in the same way as markers.

## 2.3. Extendible tracking

The range of registration is extended by extendibly tracking visible and stably detectable natural features.

**2.3.1. Evaluation of tracking results.** The tracking results are evaluated by the following criteria because the natural feature positions may be drifted by updating of the template and the tracking may fail owing to occlusion . When at least one of the following criteria are not satisfied, the tracking is discontinued.

· *Correlation between two consecutive images.*

The normalized cross correlation between two consecutive frames is used as a measure to evaluate the tracking error caused by mismatches or occlusions due to camera motion. If the correlation is under a given threshold, the tracking of such a natural feature is discontinued.

· *Epipolar constraint.*

The second criterion is based on the epipolar constraint on binocular stereo images, which means that corresponding points in stereo pair should exist on the epipolar line which is an intersection of two image planes and a plane determined by the feature point in 3D and the two centers of lenses. If the corresponding points do not exist on the epipo-

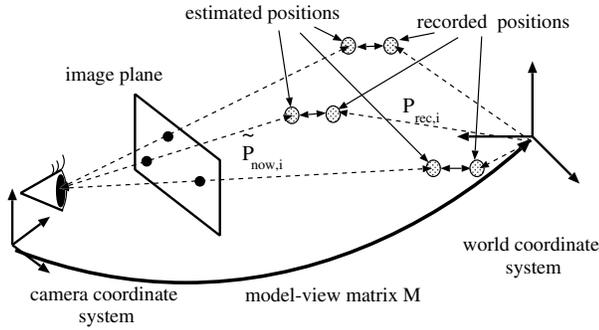


Figure 2. Registration using features.

lar line, the tracking of such a natural feature is discontinued.

· *Displacement of 3D positions.*

The third criterion is that the 3D positions of the features in the current frame should be consistent with their positions recorded at the time of their first detection. If the distance between the positions of natural features in the current frame and the positions at their first detection is larger than a threshold, the tracking of such a natural features is discontinued.

**2.3.2. Detecting new features.** In every frame, the proposed method attempts to detect a new marker from a pair of stereo images based on color information. When natural features are fall out from the frame or are deleted by the criteria described in Section 2.3.1, the method attempts to detect a new natural feature using the method described in Section 2.1. 3D positions of newly detected features are calculated and are memorized.

**2.4. Estimating model-view matrix**

A model-view matrix that represents the relationship between the world and camera coordinate systems is determined using the features. Figure 2 illustrates a relationship between the positions of features in the current frame and their recorded positions. The model-view matrix is calculated by matching their 3D positions. Provided that the position of the *i*-th feature in the world coordinate system recorded at the time of its first detection is  $P_{rec,i}$  and the position of *i*-th feature in the camera coordinate system in the current frame is  $P_{now,i}$ , the model-view matrix  $M$  can be determined by minimizing the following sum of squared differences (SSD) :

$$SSD = \sum_i w_i (P_{rec,i} - MP_{now,i})^2,$$

where  $w_i$  is a parameter that represents the credibility of the *i*-th feature. The credibility is defined according to the satisfaction of the criteria described in Section 2.3.1.

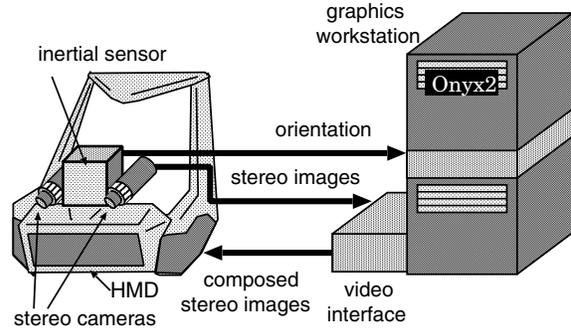


Figure 3. Configuration of prototype system.

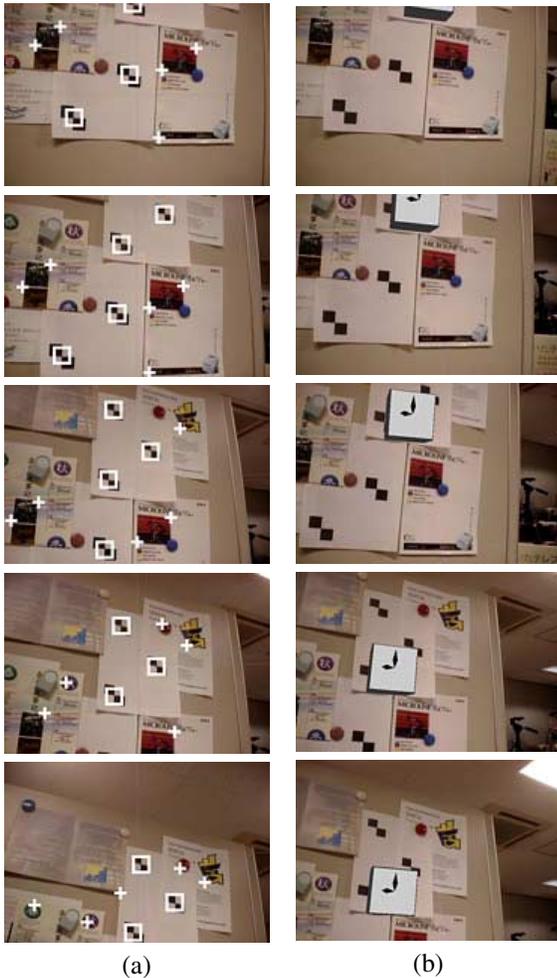


Figure 4. Appearance of the experiment.

**3. Implementation and Experiments**

**3.1. Prototype system**

We have constructed a prototype of video see-through augmented reality system mainly consisting of two small CCD cameras (Toshiba IK-UM42) and an inertial sensor (InterSense IS-300) mounted on an HMD (Olympus Media Mask), as shown in Figure 3, for demonstrating the proposed geometrical registration algorithm. The baseline length between two cameras is set to 6.5 cm. The optical axes of two cameras are set to be parallel to the viewer’s gaze direction (actually the head direction). The images captured by the cameras are fed into a graphics workstation (SGI Onyx2 IR: 16CPU MIPS R10000 195MHz) through the digital video interface (DIVO). The orientation of the camera (head) obtained by the inertial sensor is also fed into the workstation with serial interface. The incoming real world images are merged with virtual objects and output from the DIVO interface to the HMD. The hardware configuration of the whole system is illustrated in Figure 3. Figure 4 shows an appearance of experiment using the prototype system.



**Figure 5. (a) Result of tracking markers (square) and natural features ('+'). (b) Result of registration using both markers and natural features.**

### 3.2. Evaluation of effectiveness of extendible tracking

Figure 5 (a) illustrates the results of natural feature and marker tracking using the proposed method, where the mark '+' and the square represent positions of the natural features and markers, respectively. The marks in the top of the image sequence represents the positions of natural features detected by Moravec's interest operator. The only reliable natural features are then tracked by considering the criteria mentioned in the Section 2.3.1. We can also confirm that newly observed features are correctly detected from the subsequent images for extendible tracking.

Figure 5 (b) shows the results of geometric registration using both markers and natural features. In this experiment, a clock as a virtual object is drawn on the wall by using the estimated model-view matrix. It can be observed in Figure 5

(b), a clock is registered stably even if the tracking features are changed in sequence.

## 4. Conclusion

This paper has proposed a stereo vision-based augmented reality system with a wide range of registration. We have used natural feature points contained in images captured by a pair of stereo camera in conjunction with pre-defined fiducial markers. The method has achieved robust registration by extendible feature tracking with support of prediction by an inertial sensor. The feasibility of the prototype system has been successfully demonstrated through experiments.

### Acknowledgments

This work was supported in part by Grant-in-Aid for Scientific Research under Grant No. 13558035 from the Ministry of Education, Culture, Sports, Science and Technology, and also by Core Research for Evolutional Science and Technology (CREST) of Japan Science and Technology Corporation (JST).

### References

- [1] R. T. Azuma: "A Survey of Augmented Reality," *Presence*, Vol. 6, No. 4, pp. 355–385, 1997.
- [2] S. Feiner, B. MacIntyre and D. Seligmann: "Knowledge-based Augmented Reality," *Commun. of the ACM*, Vol. 36, No. 7, pp. 52–62, 1993.
- [3] T. Ohshima, K. Satoh, H. Yamamoto and H. Tamura: "AR<sup>2</sup> Hockey: A Case Study of Collaborative Augmented Reality," *Proc. VRAIS'98*, pp. 14–18, 1998.
- [4] M. Kanbara, T. Okuma, H. Takemura and N. Yokoya: "A Stereoscopic Video See-through Augmented Reality System Based on Real-time Vision-based Registration," *Proc. IEEE Virtual Reality 2000*, pp. 255–262, 2000.
- [5] T. Okuma, K. Kiyokawa, H. Takemura and N. Yokoya: "An Augmented Reality System Using a Real-time Vision Based Registration," *Proc. ICPR'98*, Vol. 2, pp. 1226–1229, 1998.
- [6] J. Rekimoto: "Matrix: A Realtime Object Identification and Registration Method for Augmented Reality," *Proc. APCHI*, pp. 63–68, 1998.
- [7] M. Uenohara and T. Kanade: "Vision-Based Object Registration for Real-time Image Overlay," *Proc. CVRMed'95*, pp. 13–22, 1995.
- [8] Y. Yokokohji, Y. Sugawara and T. Yoshikawa: "Accurate Image Overlay on See-through Head-mounted Displays Using Vision and Accelerometers," *Proc. IEEE Virtual Reality 2000*, pp. 247–254, 2000.
- [9] M. Kanbara, H. Fujii, H. Takemura and N. Yokoya: "A Stereo Vision-based Augmented Reality System with an Inertial Sensor," *Proc. ISAR2000*, pp. 97–100, 2000.
- [10] H. P. Moravec: "Visual Mapping by a Robot Rover," *Proc. 6th IJCAI*, pp. 598–600, 1979.