Proposal of International Voluntary Activities on Establishing Benchmark Test Schemes for AR/MR Geometric Registration and Tracking Methods

Hideyuki Tamura[†], Hirokazu Kato[‡], and the TrakMark Working Group

[†] Ritsumeikan University [‡] Nara Institute of Science and Technology

1 Introduction

This is a proposal to the ISMAR community from Japanese AR/MR researchers for the future progress of AR/MR technology. We hope to expand our activities over the ISMAR community and call for international participants who would take part in a number of voluntary works. At the same time, this paper presents a current view of the outcomes of these activities may have be in due course.

We will focus here on the various tracking methods, one of the most active themes at the annual ISMAR symposiums. The main goal of our activities is to build a framework to comprehensively evaluate a variety of existing and future tracking methods. Strictly speaking, our targets should include all the geometric registration methods that merge the real and virtual world seamlessly. They can also be termed as real-time 3D matchmove. After initial registration (calibration) of two spaces is achieved, either object tracking or camera tracking is required when the subject or camera moves. In addition to methods suited to static registration, there exist also methods that focus solely on improving the performance of tracking, without calibration. In this paper, when we use the terms "registration and tracking" or simply "tracking" we will be referring to the general definition of tracking as given above.

We were much inspired with the Tracking Competition held at ISMAR 2008 last year. The competition was a highly appealing event in which tracking performance was tested under specific conditions and in a limited number of locations. With an ever increasing number of participants every year, this competition doubtless has the potential to excite interest in active young researchers eager to draw attention to their work through a competition win. Spurred on by this event, in fact, we decided to attempt to build a technical basis and framework that would allow a more comprehensive evaluation, in other words, a benchmark test scheme, of AR/MR registration and tracking methods.

The tracking research field is highly active, and numerous methods appear on a regular basis. Hence, it is necessary to create a benchmark test that permits objective and accurate evaluation. To conduct such an investigation, it is necessary to select evaluation items, set evaluation standards and consolidate test data. If the foundations and the test scheme are established, it should be possible to contribute to industries seeking to use AR/MR technology in practical settings. Moreover, this could help provide a spur to researchers aiming to improve on current tracking methods.

A working group formed based on the above philosophy as a lower branch of the Special Interest Group on Mixed Reality (SIG-MR), the Virtual Reality Society of Japan, has already been established in Japan and is working towards realization of these goals. The following is an outline of the group's activities and current status.

2 Working Group (WG) Organization and Current Action Plan

2.1 Working group members

We collectively term the framework for the proposed benchmark test, as well as evaluation software now in the pipeline and the body of evaluation data as *TrakMark*, with the group assigned to work thereon as the *TrakMark WG*. This title borrows from terms often found in the PC field, such as PCMark and 3DMark.

Establishment of the TrakMark WG was approved at the 33rd SIG-MR Committee Meeting held on May 23, 2009. The current active members of WG are shown in Table 1.

2.2 Call for international participants

By presenting our activities at ISMAR 2009, we hope to recruit international participants to our cause. Specifically, we are looking for the following kinds of participants.

(a) Program contributors: Persons able to provide fully developed algorithms in program code. Those programs do not have to use methods devised by the group in question; programs that employ improvements to methods found in the paper

Table 1. WG Members

General Chair	Hideyuki Tamura	Ritsumeikan University
General Vice Chair	Hirokazu Kato	Nara Institute of Science and Technology (NAIST)
Secretariat	Tomokazu Sato Fumihisa Shibata	Nara Institute of Science and Technology (NAIST) Ritsumeikan University
Working Members	Yoshinari Kameda Itaru Kitahara Kiyoshi Kiyokawa Takeshi Kurata Yuko Uematsu	University of Tsukuba University of Tsukuba Osaka University National Institute of Advanced Industrial Science and Technology (AIST) Keio University
Senior Advisors	Yuichi Ohta Hideo Saito Haruo Takemura Naokazu Yokoya	University of Tsukuba Keio University Osaka University Nara Institute of Science and Technology (NAIST)

proposed by other researchers or the like are also acceptable.

(b) Commentators: Persons able to provide comments by closely inspecting our draft versions of TrakMark which is periodically distributed by the authors. To provide comments affecting the larger picture of the overall action plan, we wish to recruit a veteran researcher in a class with the Steering Committee. On the other hand, to investigate the details of the evaluation methods, specialists in tracking are expected.

(c) Active working members: Persons able to participate in the development of TrakMark. We wish to recruit individuals who conduct evaluations using newly developed software/TrakMark data, or provide/consolidate TrakMark data set.

(d) Public relations staff: As the TrakMark concept begins to take form, we wish to recruit individuals who can help in setting up a website to distribute information about the concept, and to actively disseminate information to the ISMAR community and industry.

There is no need to say that this work will be on a volunteer basis without remuneration. However, participants will have the right to know of the work before non-participants. In addition, the above (a)-(d) are not exclusive categories, rather, persons who can contribute to several categories are more than welcome.

For WG participants for overseas, we will have little choice but to use email as our base for communication, however, we plan to hold workshops for face-to-face interaction as the need arises.

2.3 Activity period

The following activities are scheduled for about two and a half year period following ISMAR 2009. This period has three phases.

- [1st Phase] Compile the first version of TrakMark (TrakMark 1.0) with the assistance of international participants recruited at ISMAR 2009. Firstly, it will be necessary to select minimum essential tracking algorithms and evaluation criteria in order to find the overall path. Since the ISMAR 2010 will be held in Korea, the workshop related to TrakMark will be held the week before it in Kyoto, Japan to allow presentation of the first draft and final editing.
- [2nd Phase] The edited first draft and test data will be presented at ISMAR 2010. Then, the framework of TrakMark 2.0 will be built with feedback from researchers in Japan and overseas. The aim will be to produce a benchmark test scheme of a certain level in time for ISMAR 2011.
- [3rd Phase] Following publication at ISMAR 2011, the next six months or so will be spent on including comments from all quarters, bug fixing, and the like.

3 Towards Establishment of TrakMark

To formulate the evaluation method referred to in this paper will entail investigation of evaluation subjects, evaluation items, evaluation criteria and procedure for implementation of the evaluation method, as well as consideration and construction of an overall scheme that evaluates influential tracking algorithms. Selection of test data will also form a part of this work.

The following plan has been drawn up as a result of meetings of the WG to date in Japan. What follows is no more than a tentative sketch, but, we hope it gives a rough outline of the proposed activities. The diagrams also serve merely to give an idea of the activities.

3.1 Fundamental policy on formulation of TrakMark

The aim of the WG is to formulate and publicize standard methods for evaluation of geometric registration and tracking methods in the AR/MR field. We hope to make this information widely available not only to researchers working on new methods of tracking, but also to others wishing to make use of AR/MR technology. Figure 1 shows the process of evaluation at the 1st phase. We will collect leading tracking algorithms and prepare image sequence set for evaluation. Then, we will release the result of the evaluation via the TrakMark website.

To formulate TrakMark, the following three matters should be discussed:

(1) What kind of registration and tracking algorithms will be evaluated (evaluation subject)
(2) What kind of items will be evaluated (evaluation items)
(3) How the algorithms will be evaluated (evaluation method)

3.2 Evaluation subject

Registration methods for AR/MR are classified into following three groups:

- (Physical) sensor-based methods
- Vision-based methods
- Hybrid methods combining the above

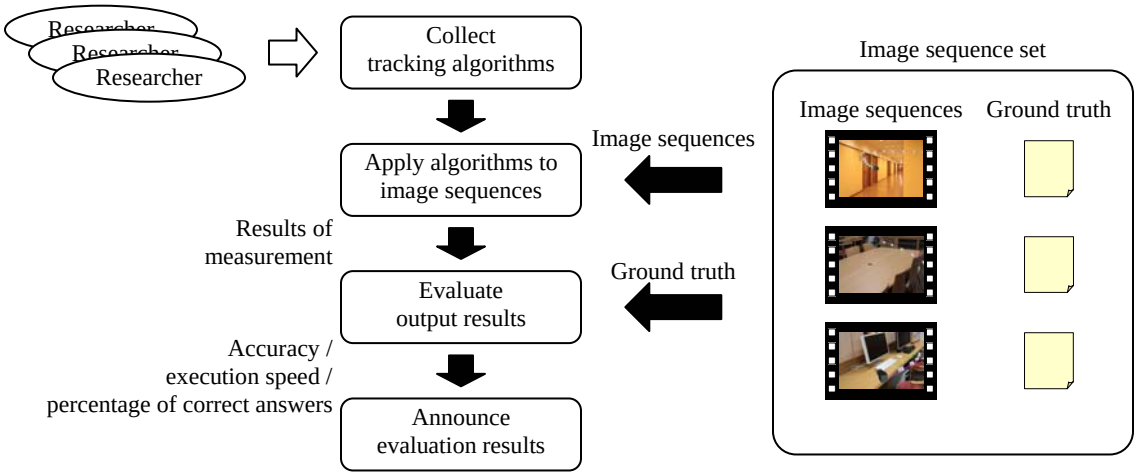
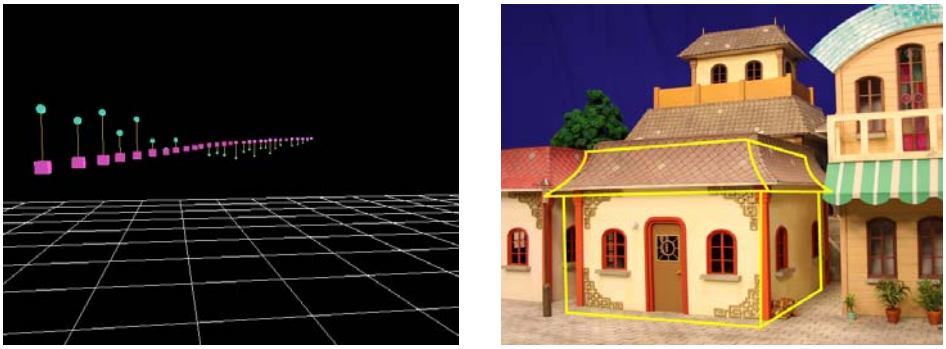


Figure 1. The process of evaluation at the 1st phase.



(a) (b)

Figure 2. Examples of the results of accuracy measurement. Fig. (a) shows the differences of camera position and orientation,and Fig. (b) indicates the reprojection error of the CG image.

In recent years, the vision based methods are becoming the mainstream because of compatibility with video see-through system as well as drastic improvement in computing power. Thus, for this evaluation method, only those tracking methods that use images captured on camera will be used as subjects.

Specifically, assuming use of cameras whose intrinsic parameters are known, the subjects for evaluation will be methods that estimate extrinsic parameters of camera (position, orientation) using features somewhere in the camera images. Initially, evaluation will be carried out of only those methods that use images from a monocular camera, with methods employing stereo images or other physical sensor information to be added at a later date.

Furthermore, regarding tracking methods using artificial markers, given the problems with image sequence preparation methods, to be discussed below, evaluation results will only be published for representative methods such as ARToolKit, in the interests of evaluating the practicality of such techniques.

3.3 Evaluation items

Of the registration and tracking methods that have been proposed to date, some have been self-evaluated and compared with other research from various angles. However, with TrakMark, we will rather select quantitative items common to most tracking algorithms than aim to comprehensively evaluate all aspects of each algorithm. Specifically, we will only adopt accuracy, execution speed, and percentage of correct answers as the evaluation items, and characteristics of each method will be recorded in a comments column.

Meanwhile, we will enrich the variation in the image sequences detailed in the section 3.4 to measure applicability to various specific tasks.

In case of measuring accuracy, there are the following two alternatives:

(a) Difference between the ground truth and measured value for the transformation matrix of the world coordinate system and camera coordinate system (3D evaluation)

(b) Reprojection error between real objects and superimposed CG objects (2D evaluation)

For example, for (a), it will be possible to express the difference between the ground truth and measured values in the world coordinates as shown in Figure 2(a). While for (b), the reprojection error can be expressed as the difference between the real objects and superimposed CG objects based on estimation of the extrinsic parameters in the image as shown is Figure 2(b).

When the primary objective is realization of AR/MR composition, evaluation of reprojection error is important, whereas the position and orientation of the camera are not. Yet, this is not sufficient reason to entirely disregard evaluation of position and orientation. It also depends on the situation in which the tracking algorithm is used. Therefore, we will give consideration to both measurement methods. Moreover, it will be necessary to investigate how to measure the ground truth (the actual position and orientation of the camera) for (a).

Latency and throughput under a common computing environment will be set as evaluation items in the evaluation of execution speed.

The percentage of correct answers can be determined according to the probability of the loss of position and orientation measurements in each frame as judged by the system.

The results for the evaluation items will be displayed either in a simple way, using average, variance, maximum, minimum and other values, or in detailed way using line graphs in chronological order.

3.4 Evaluation method

Here, the method of evaluation means procedure of assuming various scenarios where AR/MR technology is used, preparing image sequences expressing them as well as the ground truth for camera position and orientation, and comparing the output of the tracking algorithms with the ground truth. Several image sequences will be prepared, and some of them will be used for learning purposes. The image sequences could also be categorized based on degree of difficulty.

The items in Table 2 are currently under consideration to determine actual variation. Image sequences incorporating these variations will be prepared in line with scenarios. The possibility of creating image sequences in full CG was considered, however, actual camera images will be used in the 1st phase. For example, in “navigate through the corridor of a building” scenario, the camera will be moved as shown in Figure 3, and resulting pictures will be made into an image sequence. In “place a CG object on a desk” scenario, an image sequence such as that in Figure 4 will be created.

4 CONCLUDING REMARKS

We wished to present this proposal to all participants at the ISMAR 2009 symposium. In order to do so, the General Chairs, and the Program Committee of ISMAR 2009 have accepted our wish and kindly provided us with an opportunity to make a special

Table 2. Parameters in image sequences.

Category	Details
Movement	Camera path Speed Acceleration
Environment	Presence of moving objects Occlusion Illumination change Complexity/simplicity/regularity of scene
System configuration	View angle, Focus, Lens distortion, Shutter speed Screen resolution, Image quality, White balance

Figure 3. Example of image sequences in navigation. The picture on the left shows an image sequence for methods using markers, that on the right shows an image sequence for methods not using markers.

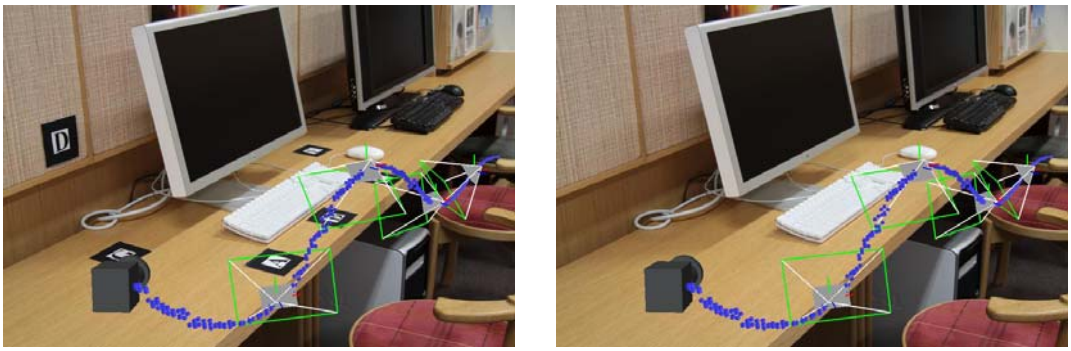


Figure 4. Example of image sequences for a desktop subject. The picture on the left shows an image sequence for methods using markers, that on the right shows an image sequence for methods not using markers.

plenary presentation. We believe that this is because the importance of the theme and necessity of the proposed activities was fully understood by all concerned.

We expect a number of questions and comments to be made when the proposal is presented at the symposium, however there may well be insufficient time to respond to all these questions. We wish that any unanswered questions you may have can be resolved through participation, discussion and effort in the TrakMark International WG.

Some researchers who for some reason were unable to attend the symposium but were able to read the proceedings on subsequent days may come across this article and find it interesting. And this is, in fact, why we wrote the paper. We are very anxious that they would consider participating in the WG.

Those in approval of our activities and able to cooperate in one or more of the ways described in 2.2 are kindly requested to contact us on the email address below. For those wishing more information, the below mentioned Web page should be up and running soon.

Mail to: TrakMarkWG@rm.is.ritsumei.ac.jp
URL: http://www.rm.is.ritsumei.ac.jp/trakmark/