An Intermediate Report of TrakMark WG

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

Fumihisa Shibata [1], Sei Ikeda [2], Takeshi Kurata [3], Hideaki Uchiyama [4], and the TrakMark WG *

[1] Ritsumeikan University, [2] Nara Institute of Science and Technology,
[3] National Institute of Advanced Industrial Science and Technology, [4] Keio University

ABSTRACT

In the study of AR/MR field, tracking and geometric registration methods are very important topics that are actively discussed. Especially, the study on tracking is flourishing and many algorithms are being proposed every year. With this trend in mind, we, the TrakMark WG, had proposed benchmark test schemes for geometric registration and tracking in AR/MR at ISMAR 2009 [1]. This paper is an intermediate report of the TrakMark WG, which describes its activities and the first proposal on benchmarking image sequences.

KEYWORDS: Geometric Registration, Tracking, Benchmark

1 INTRODUCTION

In realizing augmented reality (AR), in which the real world is augmented electronically, and mixed reality (MR), in which the real world is mixed with the virtual world, tracking and geometric registration are important subjects. Assuming that an objective assessment scheme will be required for various geometric registration methods being proposed to promote further development of AR/MR, we have created the TrakMark WG as a subsidiary organization of the Special Interest Group on Mixed Reality (SIG-MR), the Virtual Reality Society of Japan, to construct "benchmark test schemes for geometric registration and tracking in AR/MR," of which the first proposal had been presented at ISMAR 2009 held last year [1]. This paper reports the past activities of the TrakMark WG, introduces TrakMark 1.0, the first version of benchmarking image sequence, and describes the benchmarking results based on the evaluation of these sequences.

2 OBJECTIVES OF THE WG AND PAST ACTIVITIES

The TrakMark WG was first proposed and officially approved for its establishment as a subsidiary organization of SIG-MR at the 33rd Committee held on May 23, 2009 (at Future University-Hakodate).

Upon the establishment of the WG, the authors have been meeting once in one or two months and working to develop benchmark test schemes by sharing responsibilities among the members. Table 1 summarizes past agendas discussed in the WG meetings. The WG has already started to introduce the first version of benchmark image sequences [2] and started to use them for testing some geometric registration algorithms on a trial basis.

Henceforth, the WG intends to release the second proposal (TrakMark 2.0) by establishing a certain level of benchmarking standard before ISMAR 2011 and making some revisions based on feedback before terminating its operation.

The following sections describe the policy adopted to develop TrakMark 1.0 together with some discussions made in the WG meetings.

* info@trakmark.net

3 DEVELOPING TRAKMARK 1.0

3.1 Applicable Geometric Registration Methods

Geometric registration/tracking methods intended for use in AR/MR are generally classified as follows.

- Ones using physical sensors (sensor-based methods)
- Ones using camera images (vision-based methods)
- Hybrids of the above two types (hybrid methods)

In recent years, a drastic advancement in computational power has pushed the use of camera images to the mainstream position because of its compatibility with video see-through systems.

Therefore, TrakMark focuses on geometric registration/tracking methods using camera images. To be more precise, with an assumption that the internal parameters of the camera are known, methods that estimate external parameters (the position and posture) based on some indices that exist in the camera images are considered. Currently, the WG is concentrating on methods using monocular camera.

Table 1. Past Activities of the TrakMark WG

Date	Agendas
2009/07/30	- Confirmation of the WG objectives - Proposal at the special session of ISMAR 2009 - Fundamental policies on the development of benchmark
2009/09/30	- Description of the special session of ISMAR 2009 - Policies for making image sequences
2009/11/10	- Variations in image sequences - Public relations activity for TrakMark
2009/12/15	- Proposed scenarios for image sequences - Algorithms to be benchmarked
2010/01/23	- Publishing methods for image sequences - Development of the first image sequence
2010/03/09	- Verification of the first image sequence - Proposal of a website for releasing image sequences - Activities performed at KJMR 2010
2010/05/21	- Website for releasing image sequences - Policies for future public relations activity
2010/06/24	- Plans for the workshop just prior to ISMAR 2010 - Public relations on a global scale - Trial benchmark results using image sequences

With regard to the use of other physical sensors, there was a question raised about the use of a gyroscope at the special session of ISMAR 2009. In the future, if measurable, the WG will consider recording data from such a physical sensor at the same time benchmark image sequences are taken. However, such data will be used for reference purposes only because outputs from physical sensors tend to vary depending on various conditions including ambient temperature and internal states. The WG consider it necessary to prepare guidelines for using such data.

As for the method using artificial markers, the WG is considering to release benchmarking results only for some typical methods, such as ARToolKit, for use in practical reference evaluation because there are some problems in preparing image sequences as described later.

3.2 Evaluation Items

Geometric registration and tracking algorithms proposed in the past have been evaluated from various angles and compared with other relevant studies. However, TrakMark attempts to set certain common quantitative evaluation items only and does not attempt to cover all aspects. Specifically, evaluations shall be performed on "accuracy, execution speed, and percentage of correct results." Characteristics of each algorithm shall be published in the remarks column. For example, if an algorithm is effective in consistently capturing positions and postures in the dark, a statement will be added in the remarks column to that effect. This approach prevents evaluations from becoming cumbersome.

On the other hand, several variations will be provided for the image sequences described later in the evaluation procedure section in order to measure the fitness for various practical applications. For example, by including dark scenes in the image sequence, the sequence can be used to determine algorithms that yield consistent results in dark scenes.

The two ways to measure the accuracy are as follows.

- Evaluate deviations from the ground truth for the transformation matrix between the global and camera coordinate systems (an evaluation in a three-dimensional coordinate system)
- Evaluate re-projection error between real objects and the CG image which is overlaid (an evaluation in a two-dimensional coordinate system)

Although the evaluation of re-projection error of the CG image positioning is more important and the position and posture of a camera in a three-dimensional coordinate system are less important when the main purpose is to realize AR/MR, this does not constitute a sufficient reason for eliminating the three-dimensional evaluation. Because the results also depend on the situation in which a geometric registration algorithm is used, the WG has adopted both evaluation methods.

In performing an evaluation in the three-dimensional coordinate system, it is also necessary to consider how the ground truth (actual position and posture of the camera) is measured. No matter which physical sensor is used, a certain level of errors is inevitable. Therefore, at this point, such measurement results are treated as reference data and are being published as information related to the specifications of the sensor used. Although there were some references during the WG meetings to the use of match move software, which are used in integrating live-action films and CGs in the production of movies, we are yet to reach a conclusion on this subject.

Another solution for the ground truth is to create a full CG image sequence, and it is used in the first proposal of benchmarking image sequence.

As for the execution speed, both latency and throughput are measured on a common computational environment and used as evaluation items. The percentage of correct results is derived from the system's determination on whether the detection of position/posture for each frame has been successful or not.

The results of these evaluation items are expressed using averages, distributions, and maximum and minimum values for simple presentation. When exhaustive results should be displayed, the results shall be presented along a timeline for each image sequence that has been evaluated. This makes it possible for the user to determine the robustness of particular algorithms according to the type of image sequence and scenes.

3.3 Evaluation Procedure

Figure 1 shows the work flow of TrakMark. As depicted in this figure, TrakMark assumes various styles of utilization (scenarios) of AR/MR technology first and then prepares image sequences representing such scenarios along with the ground truth (or reference values) for camera position and posture at the time. In the following sections, these are referred to as "image sequence packages." At the same time, different geometric registration algorithms shall be collected for evaluation from AR/MR researches all over the world. Collected geometric registration algorithms shall be used to capture positions and postures from each image sequence. Finally, captured results shall be compared with the ground truth (or reference values) and published after sorting the results.

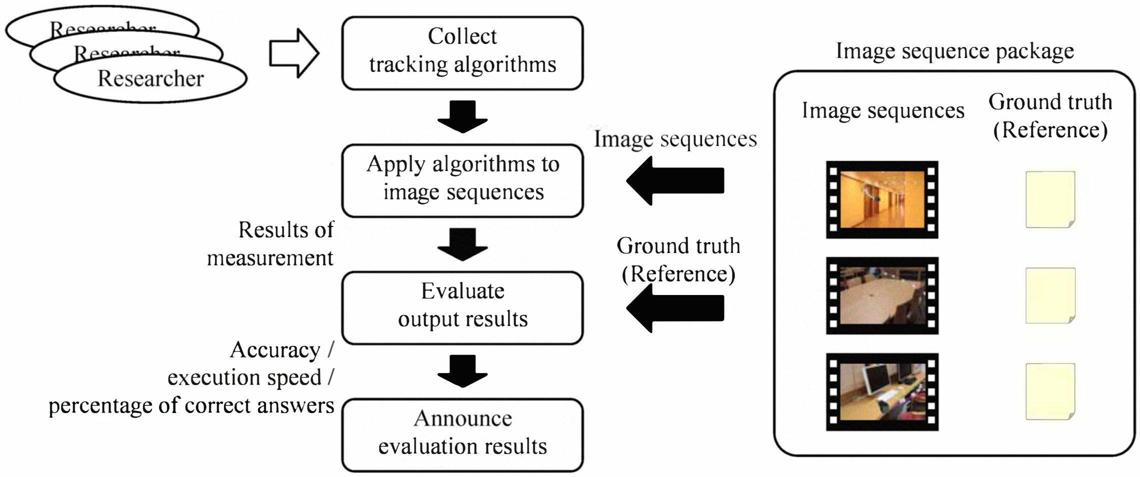

Figure 1. Work Flow of TrakMark

To clarify advantages and disadvantages of various geometric registration/tracking algorithms, it is essential to have well-developed scenarios assuming the scenes in which AR/MR technology is applied. In other words, many practical scenes where AR/MR technology will be used, rather than images for mere evaluation, must be prepared to verify the fitness of particular geometric registration/tracking algorithms for specific situations. By doing this, the WG aims to establish a guideline for selecting algorithms for people considering possible applications of AR/MR technology.

There are several variations to be included in the image sequence package such as the camera movement (linear movement, rotation, and complexity), moving objects, occlusion, shift in the lighting, lens deformation, focus, image quality, white-balance, shutter speed, resolution, viewing angle, the complexity/simplicity/regularity of scenes, etc. These variations will be combined in the package according to the assumed scenarios.

In the 4th WG meeting, the authors considered scenarios to be prepared by collecting about 30 videos that assume the use of AR/MR technology. There are a wide variety of scenarios, for example, a scenario in which certain information is overlaid on an image for the use in magazines, books, and advertisements, one that supports construction of houses, one that supports maintenance work of a bicycle, and one that allows the user to navigate indoors and outdoors. After studying various scenarios, the WG decided to prepare three packages in the first proposal of image sequence package as described in Chapter 4.

3.4 Description of the Package

The image sequence package shall include the following.

- Image sequence
  This is an archived sequence of still images. JPEG, PNG, or Raw shall be used for the image format, and each frame shall be named according to the following convention.

  [Naming convention]
  "Name of the sequence"."hour"."minute"."second".
  "millisecond"."extension"
  Ex., Campus01.12.10.05.0033.jpg

  Although video files may be included in the package for reference, evaluations shall be performed using still images.
- The ground truth (reference values)
  This is a table that contains the ground truth (reference values) of position/posture corresponding to each of the above-described file (frame).
- Package information
  This is a text file that contains information about the image, such as the image size, camera specifications, schematic view of the scene, specifications of the sensor used to measure the reference values, etc.
- Images for calibration
  This is an archive of still images prepared for the purpose of the camera calibration.

4 IMAGE SEQUENCE PACKAGES

Currently, the following three packages are published on the TrakMark website [2] as the first proposal of the benchmark image sequence package. Table 2 summarizes general information for each package.

[Film Studio Package]

The WG has chosen a scenario of a historical play shooting in a studio setting as an example of scenes observed while moving within a limited indoor space (see Figure 2). Because there are physical sensors (Tracking sensor: InterSense IS-900 SCT and Rotary Encoder: SHOTOKU TU-03VR) to measure the position and posture of the camera, reference values were measured using these sensors.

Table 2. General Information on Distributed Packages

Item	Film Studio	Campus	Conf. Venue
Frame rate	24 [fps]	15 [fps]	30 [fps]
Resolution	1920×1080	720×480 / 768×1024	640×480
Format	JPEG	JPEG	JPEG
# of sequences	7	1 / 6	3
Others	w/ ref. values	w/ ref. values	w/ the ground truth

[Campus Package]

The WG has chosen a scenario of navigating/guiding a pedestrian within a university campus as an example of motion in an outdoor setting (see Figure 3). The package includes an image sequence taken by an ordinary camera and another image sequence taken by a spherical vision system (Point Grey Research's "Ladybug") for tutorial purpose. Because it is difficult to measure the position and posture of the camera accurately, reference values are obtained by solving PnP problems after manually assigning three-dimensional positions to natural features measured by using a total station on the image of each frame. These reference values are provided every ten frames.

[Conference Venue Package]

The third scenario is for navigating/guiding a person through a site of an academic convention held in a hotel (see Figure 4). This package is special in the sense that all images are created with CG. It reproduces the site of ISMAR 2009 (Marriott Orlando Downtown) as a 3D model utilizing the achievements of Ishikawa et al. [3]. Using the CG data, image sequences were generated with the perspective of a pedestrian. Therefore, the package contains the ground truth and not the reference values of camera position and posture.

At present, this Conference Venue Package contains three image sequences. Naturally, it is possible to generate a variety of image sequences by changing various parameters, such as the camera path and resolution. The WG intends to expand the package by listening to various researchers who are supportive of the TrakMark project. Note that the CG model is published in the COLLADA format (dae file).

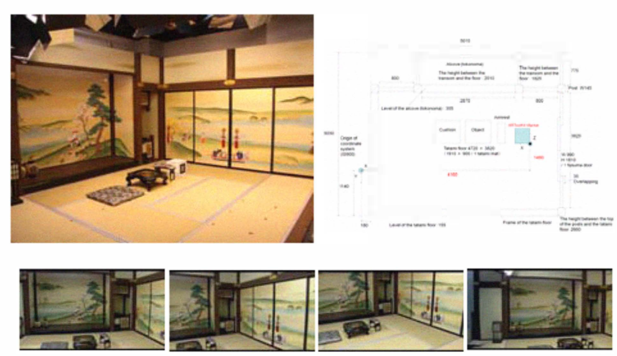

Figure 2. Contents of Film Studio Package.
On the top left is the external view of the studio setting. The floor layout is shown on the top right. Photos at the bottom are parts of the image sequence.

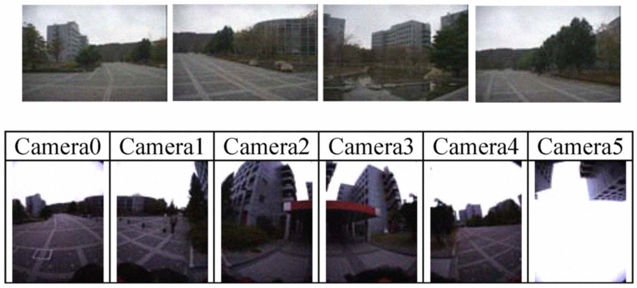

Figure 3. Contents of Campus Package
On the top are parts of the benchmark image sequence. Parts of the image sequence created with the spherical view system are shown at the bottom.

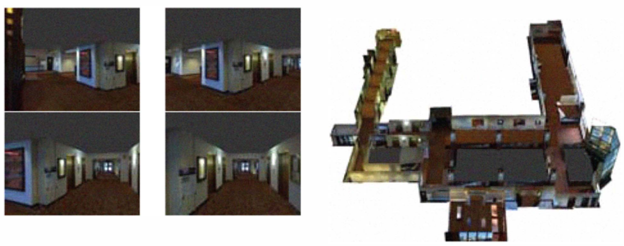

Figure 4. Contents of Convention Room Package.
On the right is the 3D model data. The rest are parts of the image sequence.

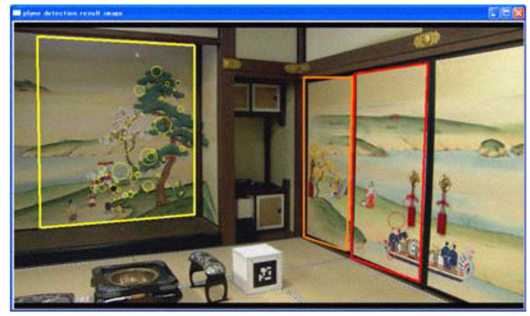

Figure 5. The result of tracking by detecting planes. Each colored rectangle represents a detected plane.

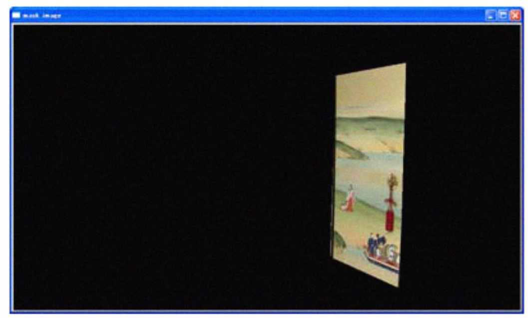

Figure 6. The extracted reference plane. From an image, we extract a referene plane by specifying its four corners.

5 Experiment on Pose Tracking Using Planes

As an example of the usage of movies in our benchmark, we applied a pose tracking using planes to a movie in film studio package. This section describes the detailed procedure of the usage and the suggestions found from the experiment.

5.1 Building plane database

3D metric positions of reference planes are first registered into a plane. In order to reconstruct each reference plane, two images are manually selected from a movie. For each image, rotation and translation parameters of a camera are computed using intrinsic parameters provided in the package and a fiducial marker in the image. Then, keypoint matching using SURF [4] is performed to extract correspondences between two images. For the correspondences inside the reference plane, 3D coordinates are computed by triangulation. This process is applied to each reference plane individually. In the database, each keypoint has a 2D coordinate, a descriptor in SURF, a 3D coordinate and a plane ID.

5.2 Tracking by detecting planes

In on-line tracking, keypoints and their descriptors are computed from an image. For each keypoint in the image, the corresponding point in the database is searched using its descriptor. Then, the searched corresponding points are clusterd by their plane IDs to apply geometric verification using homography with their 2D coordinates. As illustrated in Figure 5, camera pose is computed from 3D coordinates of detected reference planes. This process is individually applied to each input image as tracking by detection.

5.3 Suggestions

When we reconstructed 3D positions of keypoints in a reference plane, the accuracy was low because the size of the marker used for the estimation of extrinsic parameters was small in the image. In this case, we can not utilize simple triangulation for 3D reconstruction. 3D coordinates of a landmark points in each keypoint or extrinsic parameters of each image may need to be provided.

In this experiment, we extracted the texture of each plane manually for reconstruction as illstrated in Figure 6. If the texture of a plane from an anterior view and the location of the plane in a scene are provided, the users are able to avoid its reconstruction.

In the package, images for camera calibration are provided. However, the images capturing a checker board from nearly same views were not compatible for OpenCV [5] because camera calibration on OpenCV needs images captured from several different views. For the users who want to calibrate a camera by themselves, images capturing different types of checker boards from several views may be necessary.

6 Conclusion

At the time of writing this draft, a year has passed since the establishment of the WG, and the development of TrakMark 1.0 is in progress. The project has been received favorably at the special session of ISMAR 2009, which attracted many researchers and some remarks being made to propose collaboration with Tracking Competition, which is undertaken by ISMAR. Currently, the TrakMark WG is studying deficiencies in evaluation data and new evaluation procedures based on the evaluation results obtained by using several algorithms.

Such a laborious task requiring an objective mindset cannot be completed by an effort of a single study group but can only be accomplished with the support of intellectual minds and leading workgroups within the academic society. If readers are interested in the activities of the TrakMark WG, please do not hesitate to contact the WG at info@trakmark.net.

REFERENCES

[1] H. Tamura, H. Kato, and the TrakMark Working Group: Proposal of International Voluntary Activities on Establishing Benchmark Test Schemes for AR/MR Geometric Registration and Tracking Methods, In Proc. 8th Int'l Symp. on Mixed and Augmented Reality 2009, pp. 233 – 236, 2009.

[2] TrakMark, http://www.trakmark.net/

[3] T. Ishikawa, T. Kalaivani, M. Kourogi, A. P. Gee, W. Mayol, K. Jung, and T. Kurata: In-Situ 3D Indoor Modeler with a Camera and Self-Contained Sensors, In Proc. 13th Int'l Conf. on Human-Computer Interaction (HCII2009), LNCS 5622, pp. 454 – 464, 2009.

[4] H. Bay, A. Ess, T. Tuytelaars, L. V. Gool, "SURF: Speeded Up Robust Features," CVIU, vol. 110, No. 3, pp. 346-359, 2008.

[5] OpenCV. http://sourceforge.net/projects/opencvlibrary/