A prototype tool in TrakMark to generate benchmark data sets for visual tracking methods using virtualized reality models

Koji Makita* Tomoya Ishikawa† Takeshi Kurata‡

Center for Service Research, AIST, Japan

ABSTRACT

This paper describes a prototype tool to generate benchmark data sets for visual tracking methods using virtualized reality models.

In TrakMark working group, benchmark tests that permit objective and accurate evaluation of the tracking methods have been created. Ground truth data of camera parameters and feature points are needed for benchmarks of visual tracking methods. But making ground truth data is costly in real environment. Therefore, we have been developing a tool to generate data sets for benchmark using virtualized reality models. The data sets contribute to evaluation and simulation of visual tracking methods.

The data sets are composed of images generated from models and ground truth data. The ground truth data includes intrinsic and extrinsic camera parameters, and tracking data of interest points. In the tracking data, 3D-2D correspondences of interest points are shown. Generated images and all materials of ground truth data are synchronized. In this paper, the prototype tool is and generated ground truth data are shown.

1 INTRODUCTION

In TrakMark working group, various activities about evaluations of tracking methods for augmented reality and mixed reality are held. One of main activity is to generate benchmark data sets for visual tracking methods. For constructing augmented reality or mixed reality applications, an estimation of camera parameters is essential. Therefore, the benchmark data sets are composed of camera images and ground truth data that includes intrinsic and extrinsic camera parameters of each image. Moreover, for visual tracking methods, ground truth data of feature points (interest points, edge lets, and so on) are useful for accuracy evaluations.

But making ground truth data is costly in real environment. In order to make ground truth data, for example, we have to use camera with accurately controlled devices like a robot arm to get movement data. Moreover, to make ground truth data of feature points, we have to measure lots of distances between the measurement point and feature points in real world. Measurements can be held with range sensors, but it is highly time-consuming work.

Therefore, we have been developing a tool to generate data sets for benchmark using virtualized reality models. The aim of using virtualized reality models is to make ground truth data with low cost. In case virtualized reality models are able to be obtained, this tool can generate benchmark data sets that are composed of images generated from models and ground truth data. The ground truth data includes intrinsic and extrinsic camera parameters, and tracking data of interest points. In the tracking data, 3D-2D correspondences of interest points are shown. Generated images and all materials of ground truth data are synchronized.

*e-mail: k.makita@aist.go.jp
†e-mail: tomoya-ishikawa@aist.go.jp
‡e-mail: t.kurata@aist.go.jp

The rest of the paper is organized as follows. In Section 2, a design of our proposed tool is explained. In Section 3, the prototype tool is introduced, and details are shown. Section 4 is about generated data sets that are shown in TrakMark web site. Finally, Section 5 is conclusion and about future works.

2 DESIGN OF THE TOOL

Figure 1 shows an outline of the data sets. The data sets are composed of time (t), extrinsic camera parameters (EP), intrinsic camera parameters (IP), generated image (I) and tracking data (TD). The data are time synchronized. The tracking data is composed of 3D-2D correspondences of interest points. Figure 2 describes the model space and generated images $I_0, I_1, ..., I_n$. In this sample, 3D position data of interest point (x, y, z) and 2D position data (u_0, v_0) are included in tracking data TD_0, (x, y, z) and (u_1, v_1) are included in TD_1, and finally (x, y, z) and (u_n, v_n) are included in TD_n.

For the purpose of making the data shown in Figure 1, the tool is composed of four functions, **F1 : Model rendering**, **F2 : Function to generate camera parameters**, **F3 : Function to generate interest points and its tracking data**, **F4 : Data output**. In the following, 2.1 describes **F2** (Function to generate camera parameters), and 2.2 describes **F3** (Function to generate interest points and its tracking data).

2.1 Function to generate camera parameters

When we generate the data sets, there are many type of supposed scenarios. For example, human navigation, tabletop or desktop AR/MR, and so on. And the type of camera motions we want to generate depends on the scenario. In human navigation fields, for example, the camera is often supposed as hand-held type or head-mounted type that include shake. But it takes many time to independently set all extrinsic parameters with considering shake of the hand or the head.

Therefore, for the proposed tool, semi-automatic method to generate camera parameters is introduced. In the method, at first, some control points on the camera path are set by the user of the tool. After that, all extrinsic parameters are automatically generated. In the tool, as a first step of considering the type of camera motions, the camera effect of head-mounted type is introduced. Figure 3 shows the camera path and control points for generating camera parameters. An arrow in Figure 3 (a) indicates camera path the user want to generate. Without the camera effect, as shown in Figure 3 (b), the user has to set lots of control points for considering shake of head-mounted camera. On the contrary, with the camera effect, extrinsic parameters between control points are automatically generated with considering of head-mounted camera. Therefore, the user need to set control points only at start point, end point, and bifurcation as shown in Figure 3 (c).

2.2 Function to generate interest points and its tracking data

In case we manually generate interest points, we can generate and set interest points anywhere we want, but it takes many time only using manually method. On the other hand, in case interest points are automatically generated, lot of interest points can be generated

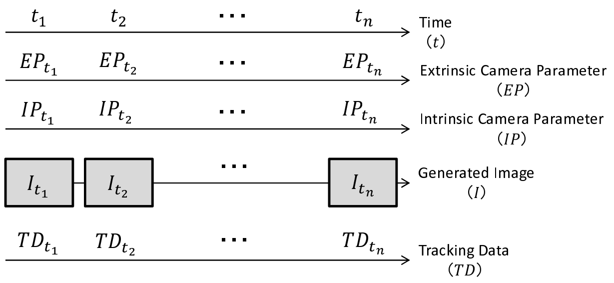


Figure 1: An outline of the data sets.

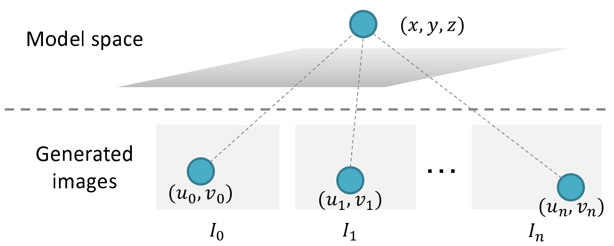


Figure 2: An outline of the tracking data.

in short time, but we can not set the place of interest points. Therefore, for the proposed tool, both manual mode and automatic mode to generate interest points are introduced.

Finally, tracking data, that is composed of 3D-2D correspondences of interest points, is automatically generated. 2D positions of interest points on generated images are calculated by using camera parameters and 3D positions of interest points.

3 Prototype Tool

This section is about a prototype tool we have constructed with following the design shown in section 2. This tool is adapted with 3D models generated by using "In-Situ 3D Indoor Modeler" [1]. For describing about the tool in this section, we use sample 3D model shown in Figure 4. Figure 4 shows a 3D model of the conference venue of ISMAR2009, that was generated in about 7.25 hours with the modeler [1]. Procedures of using the prototype tool are below.

3.1 Model selection

At first, when the tool runs, welcome screen is displayed. The user of the tool can select the 3D model with drag and drop action, and load the 3D model with load button.

3.2 Generation of camera parameters

In the prototype tool, an interface to generate camera parameters is based on control points the user can set anywhere on the ground plane of the model. Figure 5 shows the interface to generate camera parameters. The user of the tool discretely set some control points with mouse click, and set moving speed of the camera. A line in each control point shows a direction (yaw) of a camera. The user can change the direction of the control point by rotating mouse wheel. After setting control points, camera parameters between control points are automatically generated by linear interpolation. The user can select weather shake of head-mounted camera is duplicated or not. Shake of head-mounted camera is defined by some parameters, and the user can edit parameters.

For users to check generated images, we have constructed "Preview mode". The user can edit control points with using preview mode. Figure 6 shows the screen shot of the preview mode. In preview mode, the user can check all generated images with the play button and frame-by-frame step / back buttons.

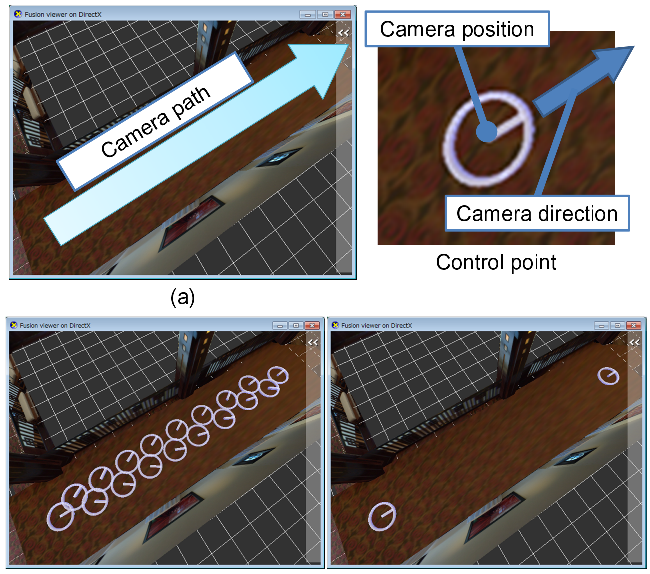


Figure 3: The camera path and control points for generating camera parameters.

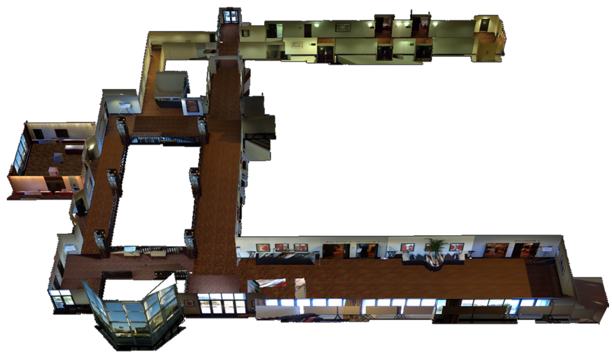


Figure 4: 3D model of the conference venue of ISMAR2009.

3.3 Generation of interest points and its tracking data

For the interface to generate interest points and its tracking data, manual and automatic modes are prepared, because they have both merits and demerits. In the manual mode, the user can generate interest points anywhere on the plane of the virtualized reality models. Figure 7 shows the interface of the manual mode. In Figure 7, interest points are shown using colored circles. After the generation, tracking data of interest points are automatically generated. But it takes long time to generate many interest points. On the contrary, in the automatic mode, the user only has to press a button to generate interest points and tracking data. The automatic mode has two different types, the one is "successive matching" and the other is "majority vote". In the successive matching type, interest points detected in successive frames are automatically accumulated. In the majority vote type, interest points detected in many frames are automatically accumulated. In the method for majority vote type, at first, interest points are independently detected in each frame. Next, interest points that exist at almost the same position in 3D model coordinates are unified by using the clustering method [2].

3.4 Output of data sets

Finally, data sets are generated with camera parameters and tracking data. Generated images are included in the data sets, but for previewing purpose, a movie file is useful. Therefore, the prototype

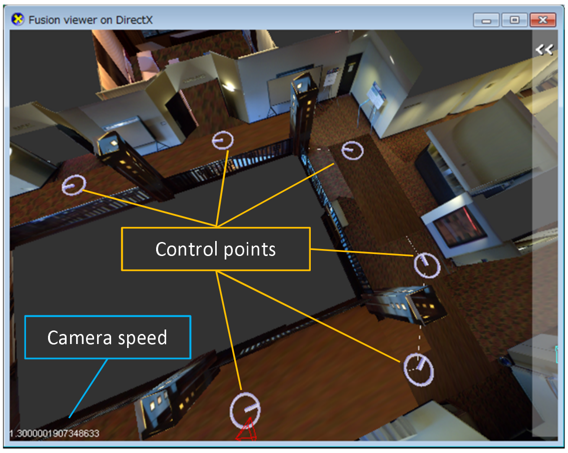


Figure 5: Interface to generate camera parameters

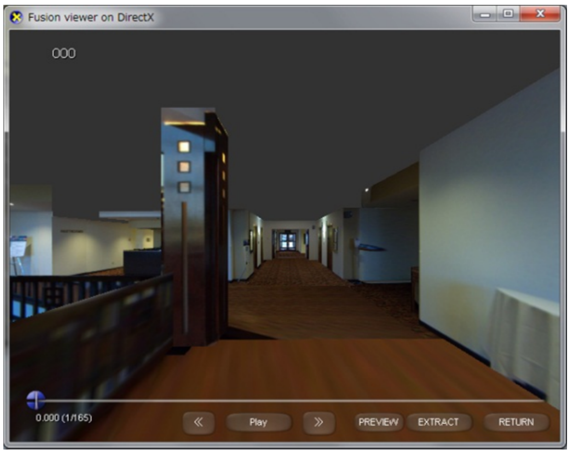


Figure 6: Preview mode

has the additional function to generate a movie file which includes all generated images.

4 Released Contents

Benchmark data sets generated with our proposed tool are already released in TrakMark web site [3]. Unfortunately, the data sets do not included the shake because the data sets were generated with the previous version of our tool. Data sets are generated with four models: the venue of ISMAR2009, the tracking competition room of ISMAR2010, nursing home, and japanese restaurant. Moreover, 3D model data of the venue of ISMAR2009 is also uploaded. In future, we are going to upload the data sets that include the shake, and remaining 3D model data to the site.

5 Conclusion

This paper described about our prototype tool to generate benchmark data sets for visual tracking method. Virtualized reality models are introduced for the purpose of effectively generating benchmark data sets. Followings are about future works.

- Improvement of feature points detection method.

  As a first step of constructing the prototype tool, only a function of detecting interest points is introduced. But in future, it is preferable to be introduced a wide variety of feature points (for example, edge lets and so on). Moreover, now there is great bias of the variability of interest points. Therefore, the method to reduce the bias should be introduced.

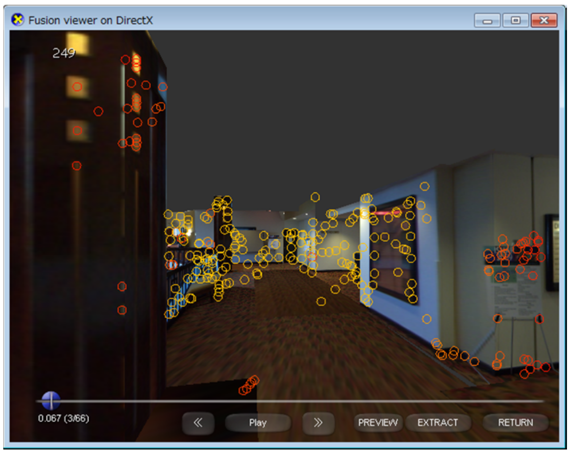
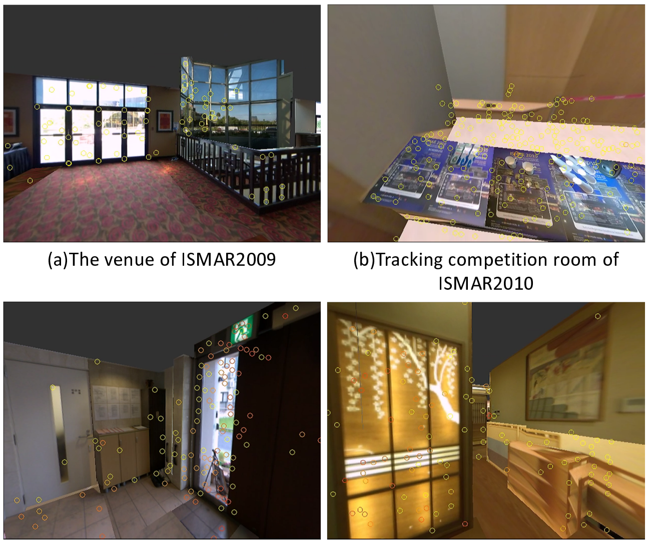


Figure 7: Manual mode (The user can set interest points anywhere on the plane with exploring the virtualized reality model.)



(a)The venue of ISMAR2009 (b)Tracking competition room of ISMAR2010

(c)Nursing home (d)Japanese restaurant

Figure 8: Released contents.

- Additional support for various scenarios.

  The prototype tool has a function to set the shake of head-mounted camera. But there are lots of various scenarios in AR/MR researched. As one of future works, we are planning to introduce the function to set various types of movements.

Acknowledgements

This work was supported by Strategic Japanese-French Cooperative Program on Information and Communications Technology Including Computer Sciences (ANR and JST).

References

[1] T. Ishikawa, K. Thangamani, M. Kourogi, A. P. Gee, W. Mayol, K. Jung, and T. Kurata: In-Situ 3D Indoor Modeler with a Camera and Self-Contained Sensors, Proc. HCII2009, LNCS 5622, pp. 454-464, 2009.

[2] M. Ester, H. P. Kriegel, J. Sander, and X. Xu: A density-based algorithm for discovering clusters in large spatial database with noise, Proc. 2nd Conf. on Knowledge Discovery in Databases and Data Mining (KDD-96), 1996.

[3] TrakMark WEB http://trakmark.net/