DES1-1
Invited

Benchmarking Indicators of Spatial Registration and Tracking Methods for Mixed and Augmented Reality

Koji Makita[1]

koji.makita.jp@ieee.org
[1]National Institute of Advanced Industrial Science and Technology, Japan
Keywords: augmented reality (AR) and mixed reality (MR), benchmarking indicators, ISO/IEC

ABSTRACT

This paper proposes benchmarking indicators of multiple camera tracking methods for mixed and augmented reality (MAR). We have been proposing AR/MR application-oriented individual indicators such as position and posture error, projection error of virtual objects, and frame rate. Fig.1 shows an outline of the projection error of virtual objects, and the concept of the indicators are included in ISO/IEC 18520:2019 [5].

1 INTRODUCTION

Recently, many kinds of camera tracking methods (mainly depend on image processing, or multiple sensors) have been proposed for a wide variety of MAR applications such as navigation, maintenance support, game, and so on. Therefore, "Benchmarking" of multiple camera tracking methods is very important both for researchers/developers of camera tracking methods and users of camera tracking methods. For researchers or developers of camera tracking methods, benchmarking results can be referred when they construct new methods. Moreover, one tracking method can be quantitatively evaluated with other methods. On the other, for users of camera tracking methods, benchmarking results can be applied as information for choosing one from multiple camera tracking methods. For realizing an efficient environment of benchmarking multiple camera tracking methods for augmented reality and mixed reality, there are mainly two kinds of issues shown in section 2 below.

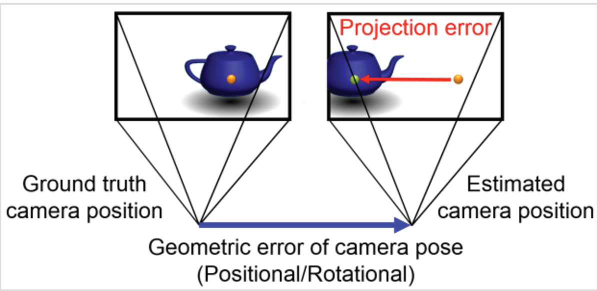


Fig.1 A sample of projection error of a virtual object. The error occurs with geometric error of camera tracking methods for mixed and augmented reality (MAR).

2 ISSUES

The first issue is about a dataset treated as an input data of camera tracking methods. By sharing common datasets that composed of camera or CG images and ground truth or reference data of camera position and orientation, researchers and developers can create estimation results of camera position and orientation by using their methods, and compare the results to the ground truth or the reference data. For instance, a dataset obtained in outdoor environment with handheld camera is proposed by K. Daniel et al [1]. However, only with the dataset, benchmarking activity is not so efficient. Because how to evaluate the results is not standardized, and each tracking method is evaluated with different indicators. In this case, when we want to evaluate our method, we have to implement several traditional methods separately from our own research efforts.

The second issue is about indicators for evaluations of the tracking results. If common datasets and common indicators for the benchmarking are available, we can evaluate the result without any additional implementations. Because in this case we can simply compare our scores to other scores reported by other researchers or developers. Generally, in order to design the indicators for benchmarking, application-oriented indicators are important. For example, J. Fritsch et al proposed a new indicator for benchmarking road detection algorithms with considering ego-lane boundaries, because most of road detection algorithms intend to achieve Advanced Driver Assistance Systems (ADAS) for passenger cars [2].

3 BENCHMARK INDICATORS

In this research, we have been proposing designs of indicators for camera tracking methods to be applied for AR/MR applications. We have been proposing AR/MR application-oriented individual indicators such as position and posture error, projection error of virtual objects, and frame rate [3]. Projection error of virtual objects is designed with considering positions of virtual objects in real world and its displacement on 2D images. Frame rate is also important indicator because it is highly related to the usability of AR/MR application. However, only with the individual indicators, each tracking method has a possibility to be optimized or fine-tuned for one or a part of datasets. In this case, flexibility of the tracking method to various situations is unclear.

To overcome this, we propose variety indicators to evaluate flexibility of the tracking method as new individual indicators. We argue that the number of datasets and variety of properties in datasets used for benchmarking can be applied to calculate the variety indicators. Moreover, we also argue that a comprehensive indicator is effective to show a summary of the benchmarking result of multiple camera tracking methods. It is difficult to theoretically design an optimum formula for calculating the variety indicators and the comprehensive indicator. For the purpose of conducting evaluations in various scenes, we have developed datasets shared in TrakMark web site [4]. In this paper, we focus on making explanations about projection error of virtual objects as especially important and characteristic indicator included in ISO/IEC 18520:2019 [5].

In section 4, definitions of the projective indicator are described. Moreover, benchmarking result examples with the indicator are described.

4 PROJECTIVE INDICATORS

Many of camera tracking methods minimize re-projection error of feature points. The re-projection error can be an index to evaluate tracking accuracy. However, on AR/MR literature, it is important to measure the projection error of a virtual object that may not be close to the feature points. Therefore, projection error of virtual points has been introduced as benchmark index [3].

Positions of virtual points in 3D space are crucial on evaluating camera tracking methods in AR/MR. They should be placed within the field of view of the camera. Two positioning strategies have been defined. First is relative placement: to place the virtual point at a relative position from the ground truth camera position. In other words, the virtual point is always at in front of the camera. Second is absolute placement: to place the virtual point at a fixed point in the world coordinate. In relative placement strategy, we place the virtual points on a virtual plane that is parallel to the image plane of the ground truth camera at a certain distance.

We first examine the 2D projection errors of nine virtual 3D points placed by the relative strategy. As transitions of the errors are similar among the nine points, we placed more virtual points within the field of view of ground truth camera and examine the mean projection error.

In evaluating a camera tracking method for AR/MR, we concern about the visibility of a virtual point at a certain distance from the camera. If it is visible, the amount of 2D projection error is also important. Therefore, projective indices should be defined on two stages below.

1. Is a virtual point visible and within the frame of the estimated camera?
2. If yes on (1), 2D Euclidean distance between the projections of virtual point that is placed in front of the ground truth camera and the corresponding projection of virtual point in front of the estimated camera.

As benchmark supporting tool, we made a program in R language. To run the program, a user needs to prepare estimated extrinsic camera parameters, ground truth of extrinsic camera parameters, position data of virtual points, intrinsic camera parameters, resolution of camera images, and parameters for creating graphs.

In current implementation, we use index number Id from zero to four to indicate the projective index of the first stage. A virtual point P is judged as “IN” if projected position: (u,v) is in the camera image and if P is in front of the camera, otherwise P is judged as “OUT”. P’ is also judged with following the same process with position: (u’,v’). Then, in case (P, P’) = (IN, IN), Id = 0. Also Id = 1 for (P, P’) = (IN, OUT), Id = 2 for (P, P’) = (OUT, IN), and Id = 3 for (P, P’) = (OUT, OUT). Finally, Id = 4 in case there is not an estimated camera parameter for the frame.

We created benchmarking results with the R program using “NAIST Campus Package 01” dataset shared in TrakMark web site [4]. The dataset includes both monocular camera sequence and omnidirectional camera sequence. We applied monocular camera sequence that includes intrinsic camera parameters computed by Tsai's method [6], and two types of extrinsic camera parameters as reference data. One is a camera path estimated from known points made by hand work (this one can be treated as ground truth). The other is a camera path estimated by a landmark-based tracking method [7]. Fig.2 shows positions of virtual points on a virtual plane. In current implementation, we set nine virtual points from A to I with Tsai's intrinsic parameters: *(f, sx, dx, dy)*, the resolution of the camera image: *(h, w) = (480, 720)*, and a distance between the camera position of ground truth and the virtual plane: *a*.

Fig.3-6 shows benchmarking results automatically created by the program. In this test, we applied relative positions of virtual points, and set the distance *a* = 1000 and 5000[mm]. When the distance is short, position error is dominant for projection error. In the results as shown in Fig.4, Id = 1 were observed at several frames when *a* = 1000, but were not observed when *a* = 5000. Fig.5 and Fig.6 show projection errors of nine points. Both when *a* = 1000 and 5000, transitions of the errors are similar among the nine points. Therefore, the tracking method [7] is supposed to be balanced to overlay virtual objects for this scene.

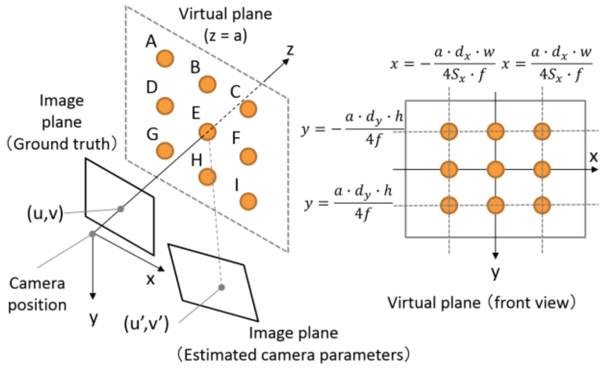


Fig.2 Positions of virtual points

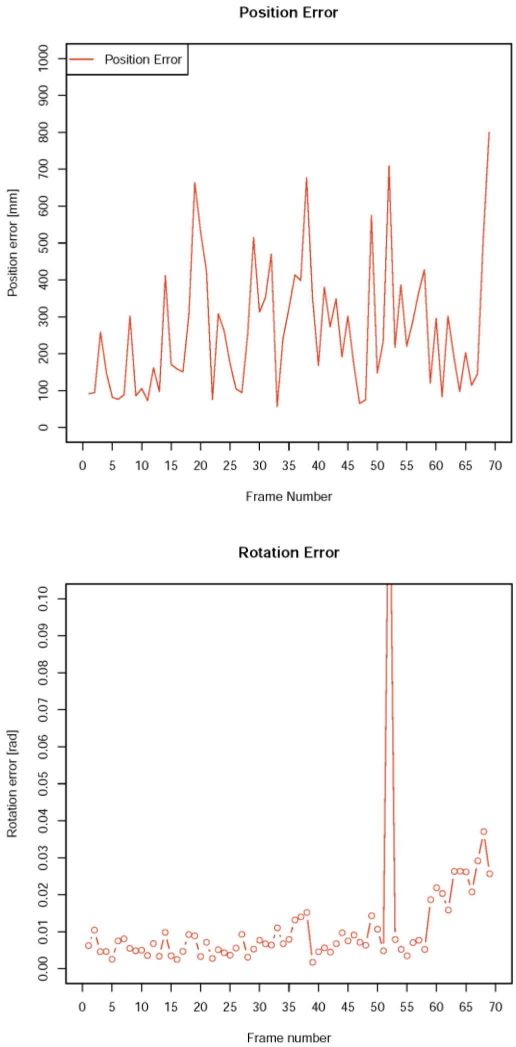
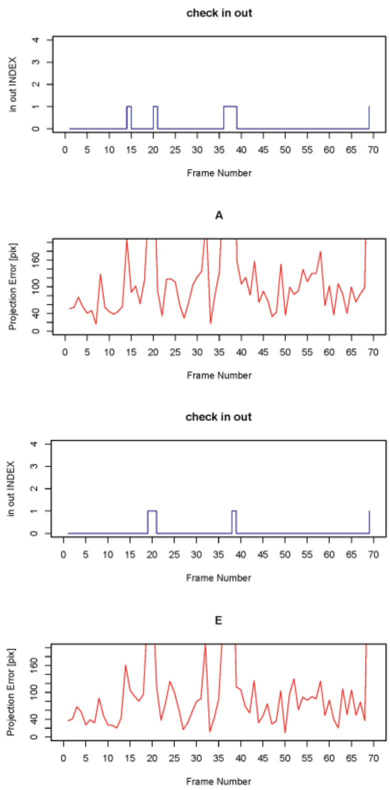
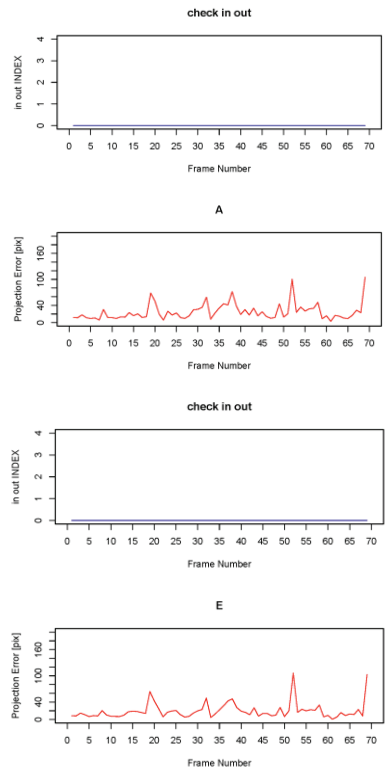


Fig.3 Position and rotation error



Distance: 1000 [mm]



Distance: 5000 [mm]

Fig.4 Projection errors of A and E with In/Out indexes

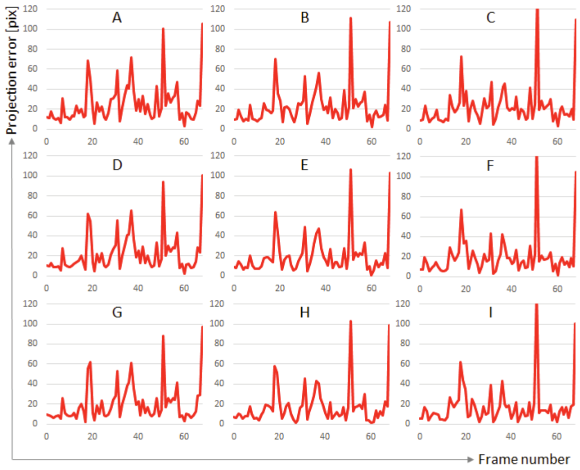


Fig.5 Projection errors of nine points (distance: 1000[mm])

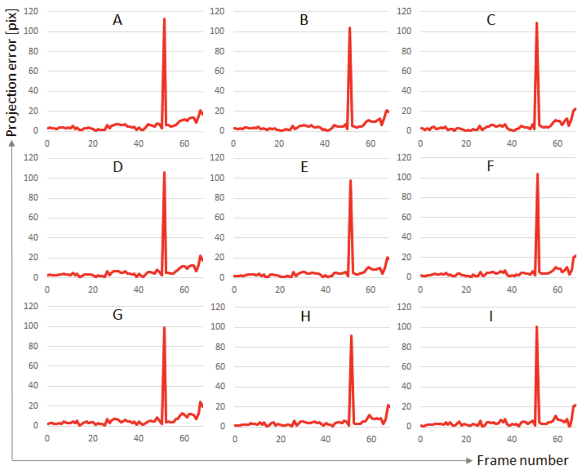


Fig.6 Projection errors of nine points (distance: 5000[mm])

5 CONCLUSION

We discussed benchmarking indicators of multiple camera tracking methods for MAR. We believe that the projective indicator based on projection error of virtual objects is important and characteristic included in ISO/IEC 18520:2019 [5]. In experiments, appropriate parameters of projective indices to compare camera tracking methods for AR/MR were described, and evaluation results on TrakMark dataset were shown. In future, the concept of the projection error of virtual objects would be applied for dynamic scene. For example, for virtual moving objects, virtual information for human [8], and so on.

Future research should be carried in order to benchmarking MAR tracking methods both in research fields and product fields. For example, NIST FRVT [9] have been receiving lots of face recognition methods from various universities and companies. In fact, the results of the FRVT have already great impact for both academic and product fields.

ACKNOWLEDGMENT

We appreciate the members in TrakMark working group giving a basic idea and useful advices to this paper.

REFERENCES

[1] K. Daniel et al, “An Outdoor Ground Truth Evaluation Dataset for Sensor-Aided Visual Handheld Camera Localization”, in Proc. ISMAR2013.

[2] J. Fritsch et al, “A New Performance Measure and Evaluation Benchmark for Road Detection Algorithms”, in Proc. ITSC2013.

[3] K. Makita et al, “Virtualized reality model-based benchmarking of AR/MR camera tracking methods in TrakMark”, in Proc. ISMAR 2012 Workshop on Tracking Methods and Applications (TMA), 4-pages.

[4] http://ypcex.naist.jp/trakmark/

[5] ISO/IEC 18520:2019, Information technology — Computer graphics, image processing and environmental data representation — Benchmarking of vision-based spatial registration and tracking methods for mixed and augmented reality (MAR), https://www.iso.org/standard/66281.html

[6] R. Y. Tsai: "An efficient and accurate camera calibration technique for 3D machine vision," Proc. IEEE Conf. on Computer Vision and Pattern Recognition, pp. 364–374, 1986.

[7] T. Taketomi, T. Sato, and N. Yokoya: "Real-time geometric registration using feature landmark database for augmented reality applications", Proc. SPIE Electronic Imaging, Vol. 7238, 2009.

[8] K. Makita, M. Kanbara, N. Yokoya: "View management of annotations for wearable augmented reality", in Proc. ICME2009.

[9] https://www.nist.gov/programs-projects/face-recognition-vendor-test-frvt