TopoTag: A Robust and Scalable Topological Fiducial Marker System Supplementary Material

Guoxing Yu, Yongtao Hu, Jingwen Dai, *Member, IEEE*

1 Detection Accuracy of Previous Work of Default Version vs. Our Fine Tuned Version

Its worthy noting that segmentation is crucial for marker detection and pose estimation for all marker systems. Thus, for the fair comparison, we fine tune the segmentation parameters for each marker algorithm unless it already uses advanced approaches like adaptive thresholding, line detection, etc. We try to fine tune previous work, specifically ARToolKit [1] and ArUco [2], to achieve the best detection accuracy (recall and precision) on the dataset we collecte.

1.1 ARToolKit

We use threshold of 60 instead of default 100. See Tab. 1 for the result comparison. Our fine tuned version achieves better recall and precision over its default version.

TABLE 1
Detection accuracy comparison of ARToolKit of its default version vs. our fine tuned version. Best result shown in bold.

	Recall (%)	Precision (%)
Default	72.079	99.529
Fine Tuned	**99.990**	**99.880**

1.2 ArUco

We use 15 and 2 for `AdaptiveThresholdWindowSize` and `AdaptiveThresWindowSize_range` instead of default -1 and 0. See Tab. 2 for the result comparison. Except detection precision slightly drops a little for one tag family, our fine tuned version achieves better recall and precision over its default version for all its three tag families.

TABLE 2
Detection accuracy comparison of ArUco of its default version vs. our fine tuned version. Best result for each tag family shown in bold.

Tag Family		Recall (%)	Precision (%)
16h3	Default	96.542	**100.000**
	Fine Tuned	**100.000**	99.910
25h7	Default	97.046	**100.000**
	Fine Tuned	**99.009**	**100.000**
36h12	Default	97.012	**100.000**
	Fine Tuned	**99.470**	**100.000**

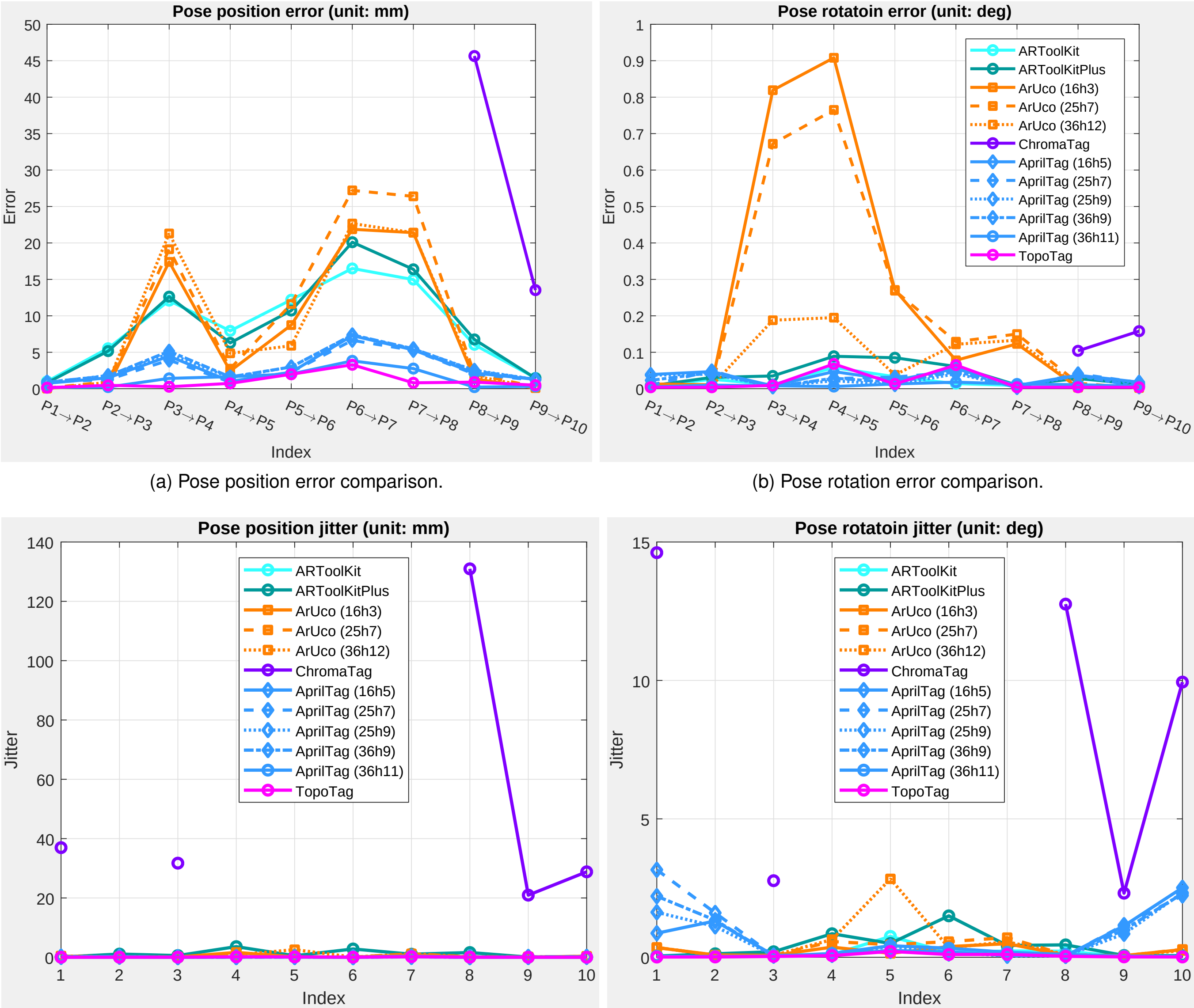
2 Full Comparison of Pose Error and Jitter

In our paper, we trim the figures of pose error and jitter for better visualization. Here we show the full figures for references, see Fig. 1.

3 Dataset Variations

We collect a large dataset including in total 169,713 images for evaluation, involving rich modalities. Figures 2, 3, 4, 5 and 6 show the dataset variations in terms of in-plane rotation, out-of-plane rotation, image blur, different distances and various backgrounds respectively.

• *G. Yu, Y. Hu and J. Dai are with Guangdong Virtual Reality Technology Co., Ltd.*
E-mail: {calvin.yu, ythu, dai}@ximmerse.com



(a) Pose position error comparison.

(b) Pose rotation error comparison.

(c) Pose position jitter comparison.

(d) Pose rotation jitter comparison.

Fig. 1. Full comparison of pose error and jitter.

REFERENCES

[1] H. Kato and M. Billinghurst, "Marker tracking and hmd calibration for a video-based augmented reality conferencing system," in *Proceedings 2nd IEEE and ACM International Workshop on Augmented Reality (IWAR'99)*. IEEE, 1999, pp. 85–94. 1

[2] S. Garrido-Jurado, R. Muñoz-Salinas, F. J. Madrid-Cuevas, and M. J. Marín-Jiménez, "Automatic generation and detection of highly reliable fiducial markers under occlusion," *Pattern Recognition*, vol. 47, no. 6, pp. 2280–2292, 2014. 1

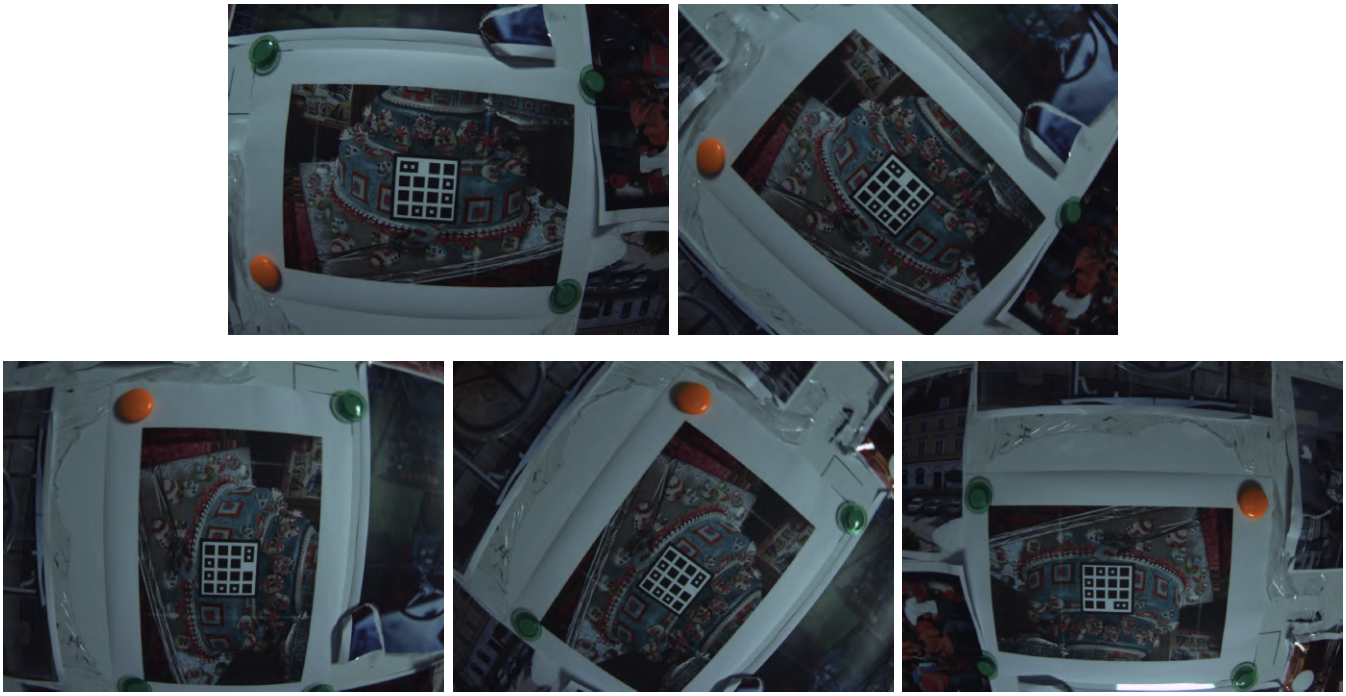


Fig. 2. In-plane rotation variation of our dataset. Images are from Seq #2.

Fig. 3. Out-of-plane rotation variation of our dataset. Images are from Seq #1.

Fig. 4. Blur variation of our dataset. Images are from Seq #1.

Fig. 5. Distance variation of our dataset. Images are from Seq #1.

Fig. 6. Background variation of our dataset. Images are from Seq #3.