IEIE Transactions on Smart Processing and Computing, vol. 4, no. 6, December 2015
http://dx.doi.org/10.5573/IEIESPC.2015.4.6.407

IEIE Transactions on Smart Processing and Computing

Subjective Evaluation on Perceptual Tracking Errors from Modeling Errors in Model-Based Tracking

Eun Joo Rhee[1], Jungsik Park[1], Byung-Kuk Seo[2], and Jong-Il Park[1]

[1] Department of Computer and Software, Hanyang University / Seoul, South Korea
{heckism, nangsik}@mr.hanyang.ac.kr, jipark@hanyang.ac.kr
[2] Department of Virtual and Augmented Reality, Fraunhofer IGD / Darmstadt, Germany
byung-kuk.seo@igd-extern.fraunhofer.de

***** Corresponding Author: Jong-Il Park

Received December 14, 2015; Accepted December 27, 2015; Published December 31, 2015

** Short Paper: None*

** Invited Paper: This paper is invited by Donggyu Sim, the associate editor.*

** Review Paper: This paper reviews the recent progress possibly including previous works in a particular research topic, and has been accepted by the editorial board through the regular reviewing process.*

** Extended from a Conference: Preliminary results of this paper were presented at the IEEE VTC Fall 2012. This present paper has been accepted by the editorial board through the regular reviewing process that confirms the original contribution*

***Abstract*:** In model-based tracking, an accurate 3D model of a target object or scene is mostly assumed to be known or given in advance, but the accuracy of the model should be guaranteed for accurate pose estimation. In many application domains, on the other hand, end users are not highly distracted by tracking errors from certain levels of modeling errors. In this paper, we examine perceptual tracking errors, which are predominantly caused by modeling errors, on subjective evaluation and compare them to computational tracking errors. We also discuss the tolerance of modeling errors by analyzing their permissible ranges.

***Keywords*:** Tolerance analysis, Human perceptual error, Modeling error, Model-based tracking

1. Introduction

Vision-based tracking is a well-known technique in computer vision for estimating six degrees of freedom (6DOF) poses of target objects or scenes. In general, vision-based tracking can be classified into two approaches: marker-based tracking and markerless tracking. Marker-based tracking is a simple, fast, and reliable approach [1], so that it has popularly been used for handling pose estimation problems in a cost effective manner. On the other hand, markerless tracking has actively been issued in many applications based on tracking framework because artificial features of visual markers easily disturb users' experience and immersion. Natural feature tracking is the most common approach to markerless tracking, which extracts distinctive features on target objects or scenes and identifies them using various types of descriptors such as scale-invariant feature transform (SIFT) [2] or speeded up robust features (SURF) [3]. Obviously, natural feature tracking is visually friendly, but sufficient texture is strongly in needed for robust tracking.

Given 3D models of target objects or scenes, model-based tracking is a promising approach for 6DOF pose estimation without using any artificial markers even on poorly textured and non-planar surfaces. As prior knowledge, the 3D models can be created using optical measuring devices such as 3D scanners, and they can also be reconstructed by vision-based 3D modeling [4, 5]. In particular, vision-based 3D modeling (or reconstruction) has recently been more attractive because special devices with high precision and resolution are not always available in many situations. For instance, 3D models of target objects or scenes can be reconstructed using visual features such as silhouette [6] or feature points [7]. Instead of fully automatic and expensive tasks, users' intervention can

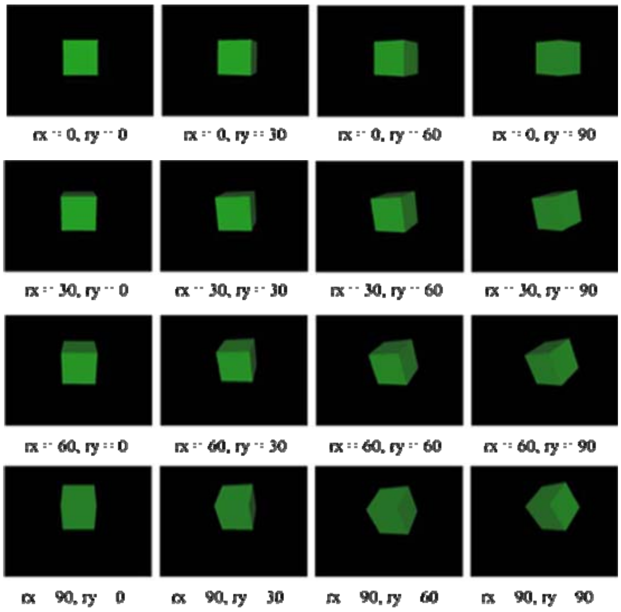

Fig. 1. Virtual cubic models rendered with different poses (rx, ry: rotation angles (degree) along the x- and y-axes).

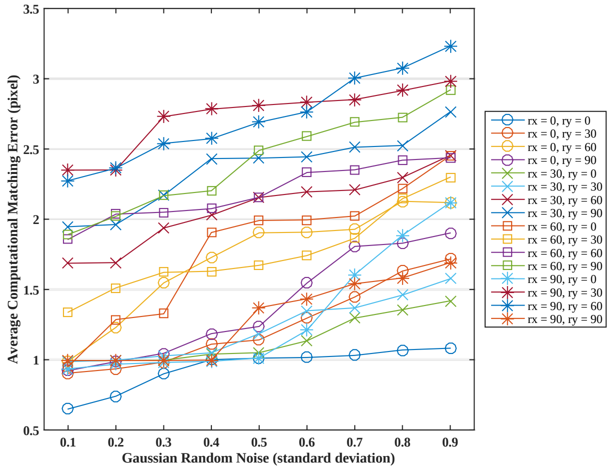

Fig. 2. Average computational matching errors from modeling errors (rx, ry: rotation angles (degree) along the x- and y-axes).

Table 1. Permissible ranges of modeling errors (standard deviation) in regard to computational matching errors* (rx, ry: rotation angles along the x- and y-axes).

	rx (degree)			
ry (degree)	**0**	**30**	**60**	**90**
0	**< 0.5**	**< 0.4**	**< 0.2**	**< 0.5**
30	< 0.4	< 0.3	< 0.1	< 0.1
60	< 0.2	< 0.1	< 0.1	< 0.1
90	< 0.3	< 0.1	< 0.1	< 0.5

*It is assumed that the tracking succeeds when the matching error is less than one pixel.

often be helpful for efficient 3D modeling [8, 9]. Furthermore, recent RGBD cameras (or 3D depth cameras) such as Kinect allow 3D modeling to be much easier and faster [10].

Despite such modeling capabilities, the accuracy of the model can be an important issue for achieving accurate tracking even though it is mostly assumed in many approaches based on model-based tracking. In the context of user experience, on the other hand, end users are not highly distracted by tracking errors from certain levels of modeling errors. From that fact, perceptual tracking errors can be a more interesting and valuable measure on application domains. In this paper, therefore, we examine perceptual tracking errors, which are predominantly caused by modeling errors, on subjective evaluation and compare them to computational tracking errors. In the subjective evaluation, two groups of participants (expert group and user group) watched test video sequences where model errors occurred with Gaussian random noise and responded to the corresponding model errors when they perceived tracking failures. With evaluation results, we also discuss the tolerance of modeling errors by analyzing their permissible ranges.

2. Computational Tracking Errors

In our evaluation, a virtual cubic object (75 mm × 75 mm × 75 mm) was assumed as a 3D target object, and its model was created by a wireframe structure with 8 vertices and 18 lines. Note that, in this study, we focus on tracking errors, which are predominantly caused by modeling errors; thus, we do not consider challenging cases such as occlusions, fast motions, illumination changes, and background clutter. With the wireframe model, the virtual object was rendered on a black homogeneous background with different poses, as shown in Fig. 1. The poses were prepared with 16 rotation angles along the x- and y-axes with 30-degree increments from 0 to 90 degrees. Since the cubic model has a symmetric shape, the rotation angles were limited to between 0 and 90 degrees. Finally, model errors were generated by adding Gaussian random noise to each vertex in all 16 cases.

In model-based tracking, computational tracking errors can be considered as matching errors between a 3D model and its corresponding features detected in an image; thus, in the evaluation, computational matching errors were evaluated by mean distance errors between boundary edges of the 3D model projected on the rendered image with estimated poses and their corresponding edges detected in the rendered image [11]. Here, the camera calibration was performed offline and an initial pose was given in advance. Fig. 2 shows the average computational matching errors from modeling errors (the entire process was repeated 1000 times in each case). From the results, the tolerance of modeling errors can be estimated by analyzing their permissible ranges as shown in Table 1.

3. Perceptual Tracking Errors

Perceptual tracking errors were evaluated by perceived

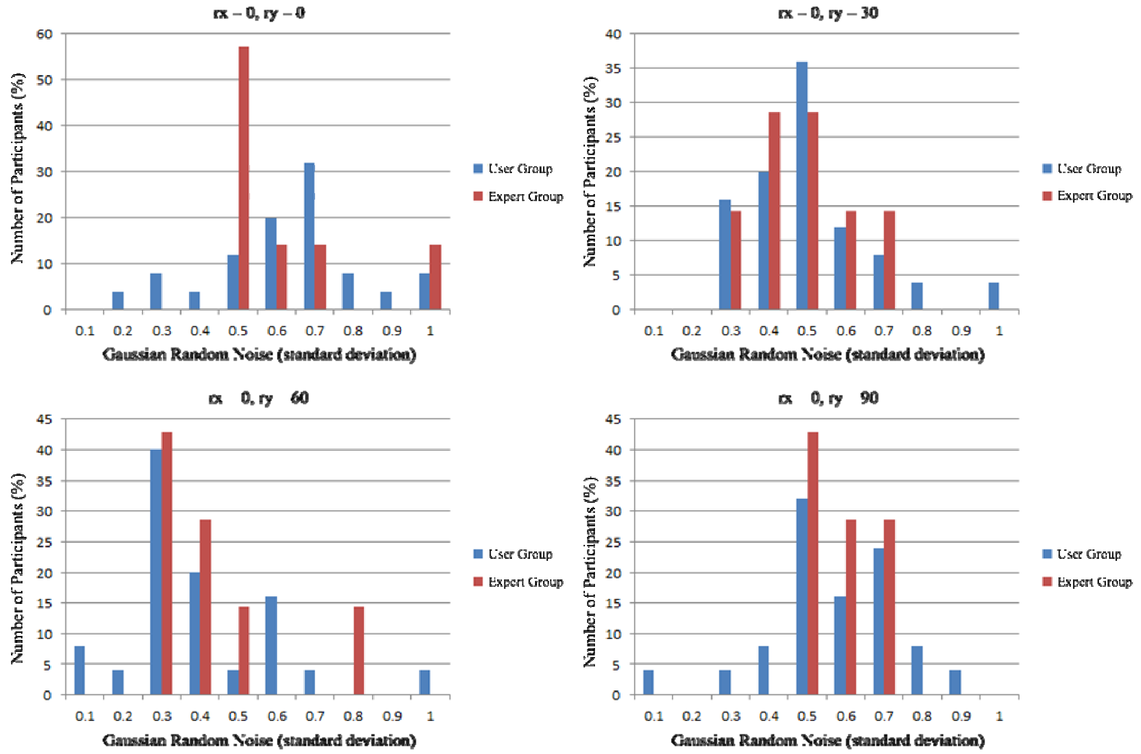

Fig. 3. Perceived matching errors from modeling errors (rx, ry: rotation angles (degree) along the x- and y-axes)

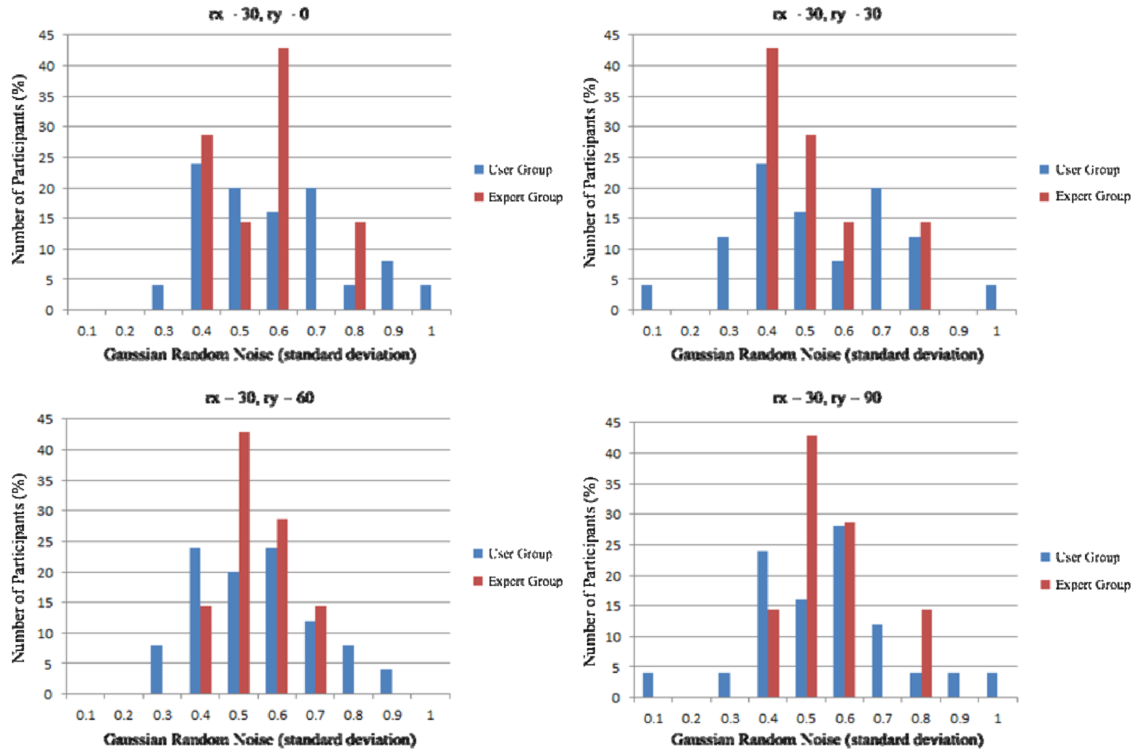

Fig. 4. Perceived matching errors from modeling errors (rx, ry: rotation angles (degree) along the x- and y-axes)

matching errors of 32 participants in two different groups (expert group: 7 researchers in the field of computer vision and user group: 25 general users). During the evaluation, each participant watched test video sequences where model errors occurred with Gaussian random noise and responded to the corresponding model errors when they perceived tracking failures. Here, the tracking setup was the same as ones for computational matching errors.

Perceived matching errors in 16 cases are shown in Fig. 3 to Fig. 6. First, perceived matching errors showed wider

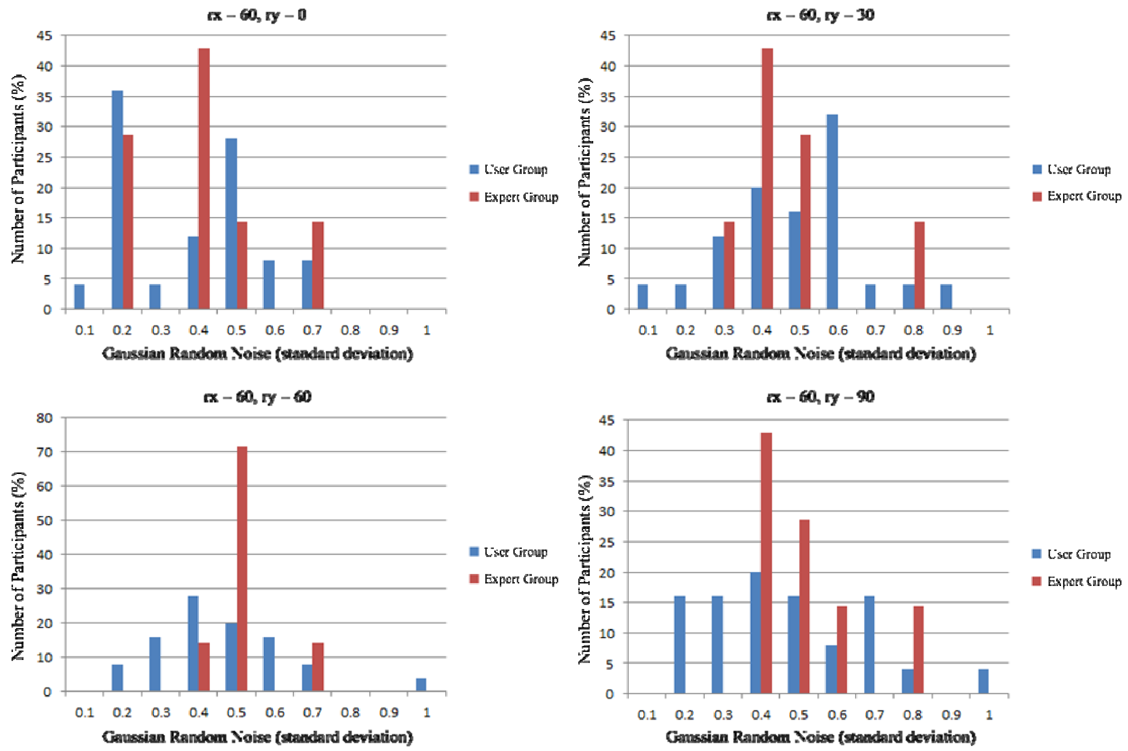

Fig. 5. Perceived matching errors from modeling errors (rx, ry: rotation angles (degree) along the x- and y-axes)

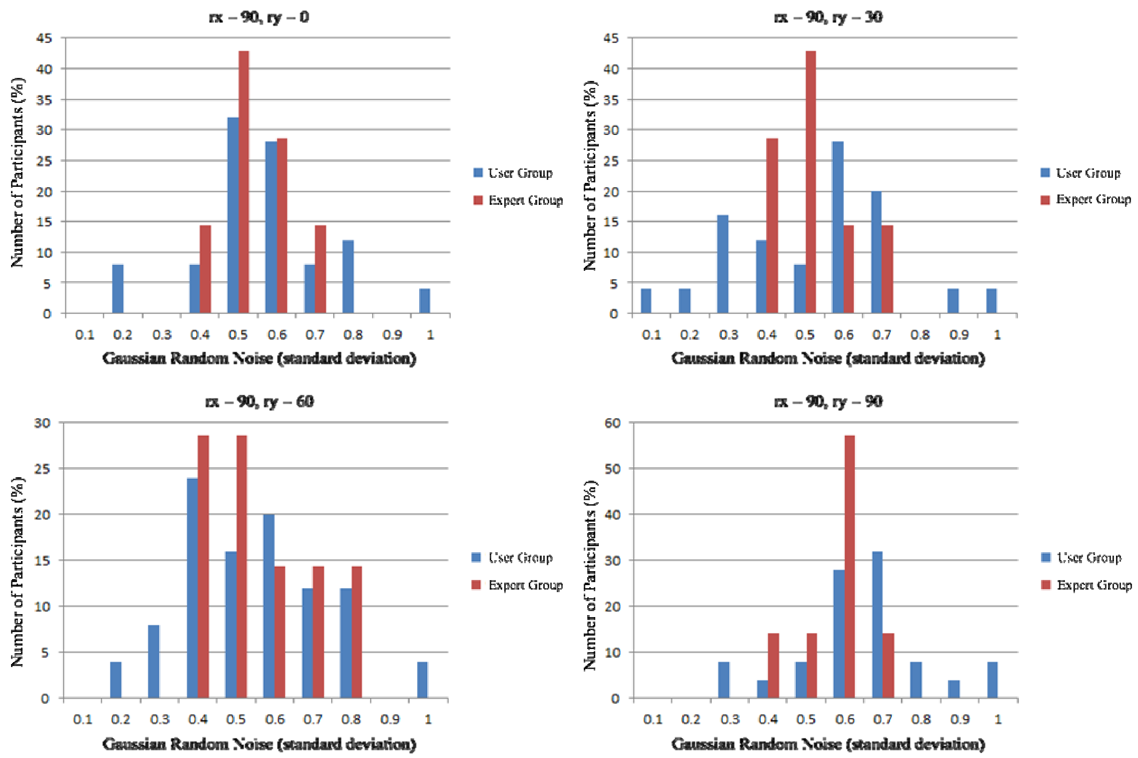

Fig. 6. Perceived matching errors from modeling errors (rx, ry: rotation angles (degree) along the x- and y-axes)

profiles than computational matching errors regardless of which group was involved. When rotating 90 degrees on the x-axis and 60 degrees on the y-axis, for example, the permissible range of modeling errors in regard to computational matching errors was less than 0.1, but most of participants perceived the tracking failures from 0.4 to 0.8 standard deviations of Gaussian random noise. Even researchers were likely to perceive the tracking failures in a wider range than general users in some cases (0 degrees on the x-axis and 90 degrees on the y-axis; 90 degrees on the x-axis and 30 degrees on the y-axis), overall, the human perception had wide profiles about tracking errors from modeling errors.

On the other hand, the expert group tended to perceive the tracking failures more accurately than the user group. In such cases: no rotation on both x- and y-axes; rotating 30 degrees on the x-axis and 0 degrees on the y-axis; and rotating 60 degrees on the x-axis and 0 degrees on the y-axis, researchers' answers coincided with the permissible ranges of modeling errors in regard to computational matching errors (see Table 1). Moreover, the expert group tended to perceive the tracking failures much quicker than the user group. In most cases, researchers perceived the tracking failures before 0.8 standard deviations of Gaussian random noise, whereas some of general users could not perceive the tracking failures until 1.0 standard deviations of Gaussian random noise. Therefore, we can notice that the permissible ranges of modeling errors for general users are wider than ones for users with expert knowledge.

In addition, perceptual matching errors had different profiles according to rotation angles on the x- and y-axes. When rotating 30 degrees on the x-axis and 60 or 90 degrees on the y-axis, for example, the permissible range of modeling errors in regard to computational matching errors was less than 0.1 standard deviations of Gaussian random noise, but the tracking failures were perceived by both of groups with a wide range of standard deviations of Gaussian random noise (from 0.4 to 0.8).

4. Conclusion

In this paper, we evaluated perceptual tracking errors from modeling errors and discussed their permissible ranges compared to computational tracking errors. The human perception had wide profiles about the tracking errors, so that it allowed wider permissible ranges of modeling errors than ones of computational tracking errors. Also, the permissible ranges of modeling errors for general users were wider than ones for expert users. In model-based tracking, therefore, tolerance analysis about modeling accuracy based on human perception would be useful to provide reasonable tracking performance on application domains for end users.

Acknowledgement

This research is supported by Ministry of Culture, Sports, and Tourism (MCST) and Korea Creative Content Agency (KOCCA) in the Culture Technology (CT) Research & Development Program 2015.

References

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

[2] D. G. Lowe, "Distinctive image features from scale-invariant keypoints," International Journal of Computer Vision (IJCV), vol. 60, no. 2, pp. 91-110, 2004. Article (CrossRef Link)

[3] H. Bay, A. Ess, T. Tuytelaars, and L. V. Gool, "Speeded-up robust features (SURF)," Computer Vision and Image Understanding (CVIU), vol. 110, no. 3, pp. 346-359, 2008. Article (CrossRef Link)

[4] M. Pollefeys, L. V. Gool, M. Vergauwen, F. Verbiest, K. Cornelis, J. Tops, and R. Koch, "Visual modeling with a hand-held camera," International Journal of Computer Vision (IJCV), vol. 58, no. 3, pp. 207-232, 2004. Article (CrossRef Link)

[5] M. Brown, T. Drummond, and R. Cipolla, "3d model acquisition by tracking 2d wireframes," in Proc. British Machine Vision Conference (BMVC'00), 2000, pp. 11-14. Article (CrossRef Link)

[6] A. Y. Mulayim, U. Yilmaz, and V. Atalay, "Silhouette-based3-dmodelel reconstruction from multiple images," IEEE Transactions on Systems, Man, and Cybernetics Part B, Cybernetics, vol. 33, no. 4, pp. 582-591, 2003. Article (CrossRef Link)

[7] N. Snavely, S. M. Seitz, and R. Szeliski, "Modeling the world from internet photo collections," International Journal of Computer Vision (IJCV), vol. 80, no. 2, pp. 189-210, 2008. Article (CrossRef Link)

[8] P. Bunnun and W. W. Mayol-Cuevas, "Outlinar: an assisted interactive model building system with reduced computational effort," in Proc. IEEE and ACM International Symposium on Mixed and Augmented Reality (ISMAR'08), 2008. Article (CrossRef Link)

[9] A. van den Hengel, R. Hill, B. Ward, and A. Dick, "In situ image-based modeling," in Proc. IEEE and ACM International Symposium on Mixed and Augmented Reality (ISMAR'09), 2009. Article (CrossRef Link)

[10] R. Newcombe, A. Davison, S. Izadi, P. Kohli, O. Hilliges, J. Shotton, D. Molyneaux, S. Hodges, D. Kim, and A. Fitzgibbon, "KinectFusion: real-time dense surface mapping and tracking," in Proc. IEEE International Symposium on Mixed and Augmented Reality (ISMAR' 11), pp. 127-136, 2011. Article (CrossRef Link)

[11] H. Park, J. Oh, B.-K. Seo, and J.-I. Park, "Automatic confidence adjustment of visual cues in model-based camera tracking," Computer Animation and Virtual Worlds (CAVW), vol. 21, no. 2, pp. 69-79, 2010. Article (CrossRef Link)

Eun Joo Rhee received B.S. and M.S. degrees in electronics and computer engineering in 2011 and 2013, respectively, from Hanyang University, Seoul, Korea. She is currently working at CTO Division, LG Electronics. Her research interests include augmented reality and human-computer interaction.

Jungsik Park received B.S. and M.S. degrees in electronics and computer engineering in 2010 and 2012, respectively, from Hanyang University, Seoul, Korea, where he is currently pursuing his Ph.D. degree. His research interests include 3D computer vision, mobile augmented reality, and general-urpose computing on graphics processing units.

Byung-Kuk Seo received the B.S., M.S., and Ph.D. degrees in electronics and computer engineering from Hanyang University, Seoul, Korea in 2006, 2008, and 2014, respectively. He is currently a postdoctoral fellow at Fraunhofer Institute for Computer Graphics, Darmstadt, Germany. His research interests include 3D computer vision, augmented reality, human-computer interaction.

Jong-Il Park received the B.S., M.S., and Ph.D. degrees in electronics engineering from Seoul National University, Seoul, Korea, in 1987, 1989, and 1995, respectively. From 1996 to 1999, he was with ATR Media Integration and Communication Research Laboratories, Japan. He joined the department of Electrical and Computer Engineering, Hanyang University, Seoul, Korea, in 1999, where he is currently a professor. His research interests include computational imaging, augmented reality, 3D computer vision, and human-computer interaction.