Mobile Augmented Reality Based on Invisible Marker

Changmin Lim, Chanran Kim, Jong-Il Park*

Department of Computer and Software
Hanyang University

Hanhoon Park

Department of Electronic Engineering
Pukyong National University

ABSTRACT

This paper proposes an approach for implementing marker-based augmented reality (AR) on smartphone. Specifically, to resolve the obtrusiveness of visual markers, use of infrared (IR) markers that are not visible to the human eye is studied. The main idea is to use an additional external camera with IR functionality to track IR markers that are not detectable in smartphone camera. Additionally is to put a visual fiducial object at the place where the fields of view of the external camera and the smartphone camera are overlapping, which enables the external camera to track the geometric transform between the fiducial object and the IR markers. As a result, since the fiducial object is trackable with the smartphone camera, the smartphone camera pose relative to the IR markers can be computed by using the transform. To validate the feasibility of the proposed approach, a proof-of-concept system is implemented where a visual marker is used as fiducial object for the convenience of implementation. The system accuracy mainly depends on the transform accuracy. Thus, to improve the transform accuracy, two constraints are defined and evaluated: one is that both markers lie on the same plane and the other is that the 3D marker data is available. Through experiments, with the constraints, it is verified that virtual contents can be stably augmented on IR markers in smartphone camera images.

Keywords: mobile augmented reality, invisible marker, fiducial object, external camera with IR functionality.

Index Terms: J.0.2 [Human computer interaction (HCI)]: Interaction paradigms—Mixed/ augmented reality; J.3.0 [Ubiquitous and mobile computing]: Ubiquitous and mobile computing theory, concepts and paradigms—Mobile computing

1 INTRODUCTION

Our aim is to implement a mobile augmented reality (AR) system that augments moving objects with virtual contents on smartphone camera images. For this purpose, various approaches for estimating the camera pose relative to the objects have been studied. However, when target objects has less texture and their structure models are unavailable or difficult to measure, common marker-less approaches [7, 9, 10] cannot be employed. In turn, use of markers can be an excellent alternative. Binary markers (e.g., ARToolkit [1], ARTag [2]) are easy to detect and can be robustly tracked regardless of poor scene conditions. Therefore, we also use markers, i.e., attaching markers to target objects.

However, the obtrusiveness of markers, i.e., breaking the sense of immersion, has been a major obstacle for their use in practical applications. There have been several efforts for making markers less obtrusive [6, 8]. As the most effective solution, infrared (IR) markers which are invisible to the human eye have been proposed [3]. In the environment where IR cameras are available, there have appeared good AR applications using IR markers. In [4], IR markers were invisibly embedded into a picture book and virtual contents could be plausibly displayed using a potable projector, whose pose relative to the book could be easily estimated by detecting the IR markers using an IR camera stuck to the projector.

Unfortunately, commodity smartphones do not have the infrared camera, there is no way to detect the IR marker. This makes difficult to use IR markers in our system. Therefore, we propose a roundabout approach to use IR markers in our system. The main idea is to add an external camera with a wide field of view and an IR functionality and put a visual fiducial object within its field of view. Once the geometric relationship between the IR marker on the target object and the fiducial object is estimated, the pose of smartphone camera relative to the IR marker can be computed by estimating the pose relative to the fiducial object instead. In this approach, the only requirement is that the fiducial object should be robustly trackable in both the smartphone camera and the external camera. Objects with rich texture are good candidates because they can be robustly tracked by using feature tracking methods [6, 7]. However, they are not always available. First of all, in this paper, we focus on implementing a proof-of-concept system for validating the feasibility of the proposed approach. Thus, for the convenience of implementation, we use a visual marker as fiducial object. The configuration of our proof-of-concept system is given in Fig. 1.

In the proposed approach, when both the markers are at a distance, their geometric relationship cannot be reliably estimated and the final pose (smartphone camera to the IR marker) can be unstable. Actually, due to the noisy IR images, the pose error could be more severe. To resolve this problem, we enforced two constraints in setting up the experimental environment. One is that both markers lie on the same plane and the other is that the 3D data of both markers is available. For each constraint, it will be verified how the virtual object is stably augmented.

2 COMPUTATION OF INVISIBLE MARKER POSE

In Fig. 1, the visual marker pose $\mathbf{E}_A$ and the invisible marker pose $\mathbf{E}_I$ are first estimated by detecting the markers from the external camera images. Then, the transform $\mathbf{E}_{AtoI}$ is computed by $\mathbf{E}_A^{-1}\mathbf{E}_I$. Once the visual marker pose $\mathbf{E}_{A'}$ to the smartphone camera is estimated, the invisible marker pose $\mathbf{E}_{I'}$ to the smartphone camera is simply computed by $\mathbf{E}_{A'}\mathbf{E}_{AtoI}$. However, when $\mathbf{E}_{AtoI}$ is unreliably estimated, the virtual object is augmented on the incorrect location as shown in Fig. 2.

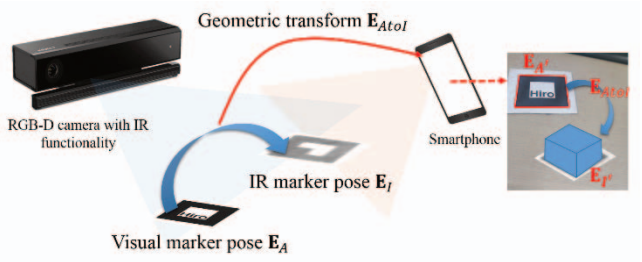

Figure 1: A proof-of-concept system for the pose computation of the IR marker in smartphone.

* jipark@hanyang.ac.kr (corresponding author)

978-1-5090-3740-7/16 $31.00 © 2016 IEEE
DOI 10.1109/ISMAR-Adjunct.2016.38

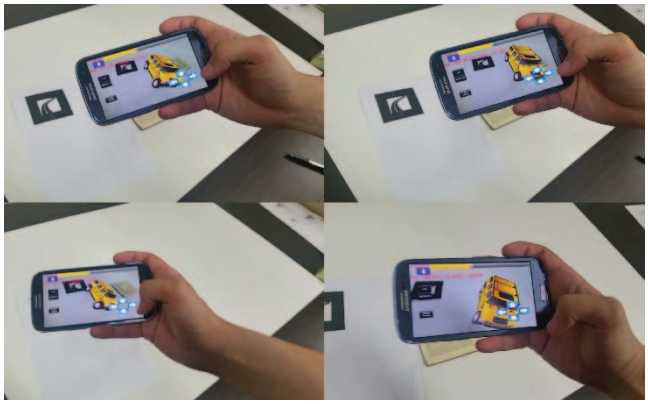

Figure 2: The problem that virtual object is augmented on the incorrect location.

2.1 Using the constraint that both markers lie on the same plane

When the marker is detected in the infrared image which is usually noisy, the marker pose error is increased. Also, the computation error of $\mathbf{E}_{AtoI}$ is increased in proportion to the pose difference between $\mathbf{E}_A$ and $\mathbf{E}_I$. To reduce the errors, a constraint that both markers lie on the same plane is used. With this constraint, $\mathbf{E}_{AtoI}$ can contain only the rotation along the Z-axis and the translations along the X and Y axes.

First, the visual marker is used to compute the common plane equation, since it is detected more accurately than invisible marker. Specifically, the plane equation is as follows:

$$ax + by + cz + d = 0, \tag{1}$$

where a, b, c, and d are coefficients of the plane equation, which can be computed from 3D coordinates $(x_i,\ y_i,\ z_i)$ of four corners of the marker. Then, the 2D vertices (u_i, v_i) of marker in camera images are back-projected to rays by the intrinsic camera matrix $\mathbf{K}$ as follows:

$$\mathbf{K}^{-1}\begin{bmatrix} u_i \\ v_i \\ 1 \end{bmatrix} = \begin{bmatrix} \alpha_i \\ \beta_i \\ \gamma_i \end{bmatrix}, \tag{2}$$

where $[\alpha_i, \beta_i, \gamma_i]^{\mathrm{T}}$ denotes the ray vector. Next, the line-plane 3D intersection points (x_i^p, y_i^p, z_i^p) are computed. Finally, the method for computing the optimal rigid-body transformation [5] is used to compute $\mathbf{E}_{AtoI}$ between four pairs of $(x_i,\ y_i,\ z_i)$ and (x_i^p, y_i^p, z_i^p).

2.2 Using the constraint that the 3D data from RGB-D camera is available

The coplanarity constraint is useful once the infrared camera is available. However, in practical AR applications, the constraint is quite strict. In addition, even with the constraint, the large translation error can be occurred when both markers are at a distance. Therefore, to augment the virtual object more accurately, the other constraint is used: accurate 3D coordinates of markers are available. To get the 3D coordinates, an RGB-D camera is used.

First, 2D vertices of the IR marker and visual marker are detected in the infrared image and the 3D metric coordinates of the 2D vertices are computed by using the depth map acquired by the RGB-D camera and its camera intrinsic matrix. Then, $\mathbf{E}_{AtoI}$ is computed as the optimal rigid-body transformation between the 3D vertices of both markers.

3 Evaluation

The stability of the computed $\mathbf{E}_{AtoI}$ was evaluated. Specifically, the standard deviation of each element (rotation and translation parameters) of $\mathbf{E}_{AtoI}$ was computed with markers placed at various locations. In our experiments, the IR marker proposed in [4] was used. To detect/track the marker, the ARToolkit library was used. Kinect v2 and Galaxy S3 were used as the IR/RGB-D camera and smartphone, respectively. The IR marker-based mobile augmented reality was implemented on a server-client system: a server (PC) that is connected to the Kinect camera and computes $\mathbf{E}_{AtoI}$; a client (smartphone) that augments a virtual object on invisible marker using the received $\mathbf{E}_{AtoI}$ from the server.

The IR camera was located at 90 cm away from the marker. At first, $\mathbf{E}_A$ and $\mathbf{E}_I$ were measured and their stability was evaluated. For the purpose, the standard deviation of each element was measured at nine fixed locations as shown in Fig. 3. For the standard deviation computation, each $\mathbf{E}$ was measured 100 times at each location.

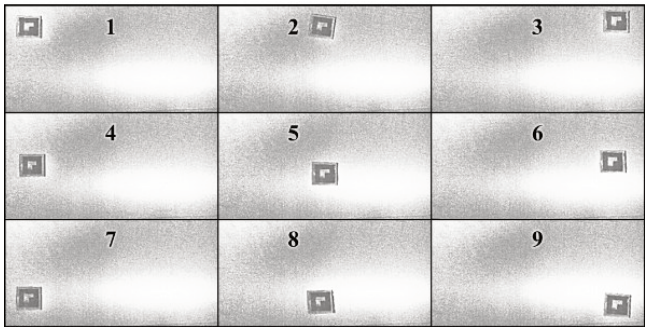

Figure 3: The nine fixed locations for computing $\mathbf{E}$.

Table 1 is the average standard deviation of each element of $\mathbf{E}_A$. The computed average standard deviation was larger than that by the RGB camera in [3], since it used the infrared camera which was located further away from the marker than in [3]. The average standard deviation of each element of the $\mathbf{E}_I$ was also larger than that in [3] because of the distance between the camera and the marker.

Table 1: The average standard deviation of $\mathbf{E}_A$.

Location	θ_x(°)	θ_y(°)	θ_z(°)	t_x(mm)	t_y(mm)	t_z(mm)
1	6.619529	5.070545	1.894814	10.12633	6.282384	18.41202
2	3.008464	1.541081	0.51008	0.719571	0.89685	2.835398
3	2.534579	3.104529	0.98153	0.247525	1.866354	6.384148
4	3.85485	4.519323	0.788088	8.20898	3.520369	15.84253
5	6.511902	3.928045	0.765036	2.004474	1.203766	6.664524
6	1.828591	3.671107	0.549753	0.115794	0.423842	2.457746
7	1.695099	2.282733	0.130933	5.437794	2.425192	14.58463
8	1.303185	1.961014	0.153187	1.674337	0.504528	5.891518
9	1.040718	1.272312	0.150606	0.122274	0.484701	5.410084
Average	3.155213	3.038965	0.658225	3.184119	1.956442	8.720288

Table 2: The average standard deviation of $\mathbf{E}_I$.

Location	θ_x(°)	θ_y(°)	θ_z(°)	t_x(mm)	t_y(mm)	t_z(mm)
1	10.82422	8.680984	4.03052	7.978536	4.820787	14.25504
2	3.835759	4.556959	1.769711	4.051477	5.284113	15.82411
3	4.921462	4.57949	1.527716	0.344332	2.000611	6.660585
4	6.370303	5.967619	0.933659	19.00157	6.821201	36.45441
5	5.289227	3.770396	0.823788	2.926062	2.531032	13.38461
6	13.91977	4.229294	0.743291	0.436848	4.269894	23.04559
7	3.822562	4.876866	0.199263	10.32697	1.811975	18.56447
8	8.35519	3.913727	0.155646	9.274536	1.487305	28.02439
9	3.465197	4.717528	0.369337	0.237699	0.75113	12.5509
Average	6.755965	5.032540	1.172547	6.064225	3.308672	18.75156

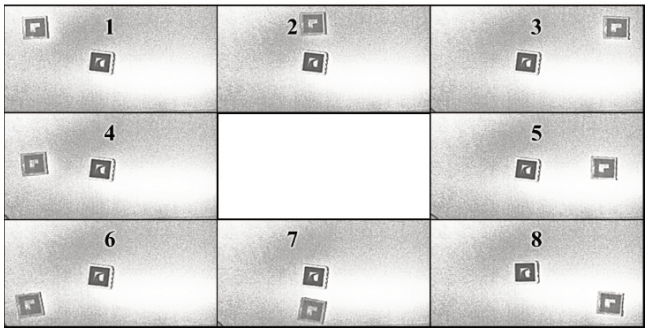

Figure 4: The eight fixed locations for computing $\mathbf{E}_{AtoI}$.

Table 3: The average standard deviation of using the constraint that both the markers lie on the same plane.

Location	θ_x(°)	θ_y(°)	θ_z(°)	t_x(mm)	t_y(mm)	t_z(mm)
1	0	0	0.348829	4.879158	1.801088	0
2	0	0	0.265755	0.383207	3.062539	0
3	0	0	3.006865	8.0808	8.35501	0
4	0	0	0.613185	5.546303	0.665867	0
5	0	0	0.475881	1.806103	1.244989	0
6	0	0	0.334034	10.94196	0.706324	0
7	0	0	0.258874	0.860501	1.675298	0
8	0	0	0.307199	0.719226	1.899943	0
Average	0	0	0.701327	4.152157	2.426382	0

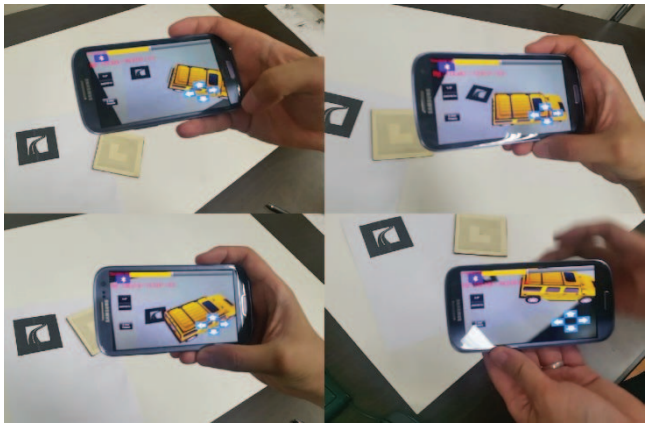

Figure 5: The mobile IR marker-based AR using the coplanarity constraint. It still may suffer from translation errors. The reason why IR markers are not perfectly invisible is because the IR-fluorescent material used in our experiments has its inherent color.

Figure 4 is the eight fixed locations where the average standard deviation of each element of $\mathbf{E}_{AtoI}$ was measured. The visual marker was located at the center of space. $\mathbf{E}_{AtoI}$ was computed while placing the IR marker around the visual marker. First, the average standard deviation when using the coplanarity constraint was computed in Table 3. The rotation along the X-axis and Y-axis and the translation of the Z-axis were zero as expected. The average standard deviation of the other parameters was lower than that in Table 2. Consequently, the virtual object was mostly stably augmented on the IR marker with the coplanarity constraint as shown in Fig. 5. However, in Table 2, when the distance between two markers was increased, the translation errors along the X-axis and Y-axis were increased.

Second, the average standard deviation when using the 3D data availability constraint was computed in Table 4. Because the accurate 3D data was used, the computed average standard deviation was much lower than that in Table 3. Especially, the translation error was dramatically reduced when both markers were at a distance. As shown in Fig. 6, the virtual object was much stably augmented on the accurate location in the smartphone camera images.

Through these experiments, it was verified that:

- When the distance between the visual marker (the fiducial object) and the IR marker was small, both constraints had no difference.
- When the distance was large, the 3D data availability constraint could be preferred.

Table 4: The average standard deviation of using the constraint that 3D data from RGB-D camera is available.

Location	θ_x(°)	θ_y(°)	θ_z(°)	t_x(mm)	t_y(mm)	t_z(mm)
1	2.078307	2.010181	0.611566	0.737193	1.144301	9.075131
2	2.250907	1.993392	0.402624	0.584733	0.273934	3.934114
3	2.039247	1.958548	0.328921	0.465631	1.282901	7.500426
4	1.770503	1.788226	0.499986	0.500916	1.097424	6.065874
5	2.007295	1.890505	0.302008	0.384369	1.269974	6.677059
6	1.792332	1.709509	0.432217	0.433673	0.70255	5.519746
7	1.946284	1.881787	0.303983	0.417848	0.152954	2.73098
8	1.960883	1.773447	0.296573	0.446144	1.203112	8.950029
Average	1.980719	1.875699	0.397234	0.496313	0.890893	6.306669

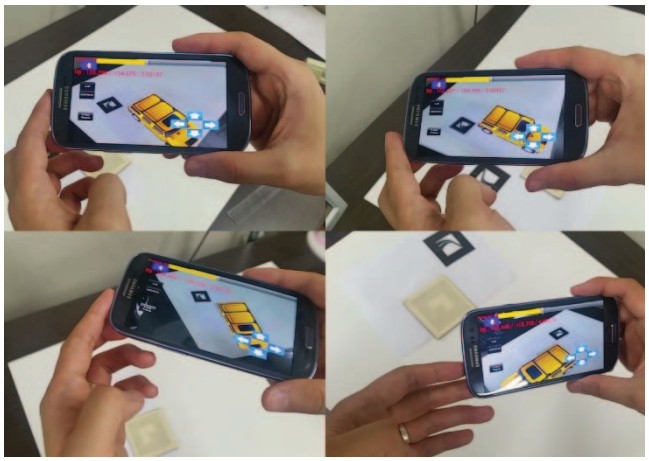

Figure 6: The mobile IR marker-based AR using the 3D data availability constraint.

4 CONCLUSION

This paper proposed an approach for implementing IR marker-based augmented reality on smartphone. The main idea was to add an external camera to estimate the geometric relationship between the IR marker and a fiducial object. To improve the estimation accuracy, two constrains were defined and evaluated. Then, it was verified that the virtual object could be stably augmented on IR marker with the constraints and mobile IR marker-based AR could be successfully implemented.

However, use of IR marker had an inherent problem with inaccurate detection of IR marker due to noisy IR images, which was a main cause of jittering in our results. Thus, we are studying an effective method for enhancing the IR images.

ACKNOWLEDGEMENTS

This research was supported by the MSIP (Ministry of Science, ICT and Future Planning), Korea, under the ITRC (Information Technology Research Center) support program (IITP-2016-H8601-16-1005) supervised by the IITP (Institute for Information & communications Technology Promotion)

REFERENCES

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

[2] M. Fiala, "ARTag, a fiducial marker system using digital techniques," *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, volume 2, pages 590-596, 2005.

[3] H. Park and J. -I. Park, "Invisible marker-based augmented reality system," *In Proceedings of Visual Communications and Image Processing (VCIP)*, pages 59601I-59601I, 2005.

[4] K. D. Willis, T. Shiratori, and M. Mahler, "HideOut: mobile projector interaction with tangible objects and surfaces," *Proceedings of the 7th International Conference on Tangible, Embedded and Embodied Interaction (TEI)*, pages 331-338, 2013.

[5] D. W. Eggert, A. Lorusso, and R. B. Fisher, "Estimating 3-D rigid body transformations: a comparison of four major algorithms," *International Journal of Machine Vision and Applications*, volume 9, pages 272-290, 1997.

[6] D. Wagner, T. Langlotz, and D. Schmalstieg, "Robust and unobtrusive marker tracking on mobile phones," *Proceedings of the 7th IEEE/ACM International Symposium on Mixed and Augmented Reality (ISMAR)*, 2008.

[7] G. Klein and D. Murray, "Parallel tracking and mapping for small AR workspaces," *Proceedings of the 6th IEEE/ACM International Symposium on Mixed and Augmented Reality (ISMAR)*, 2007.

[8] S. Y. Park andJ. W. Choi, "Retrieving and browsing information of building equipment using augmented reality techniques," *Proceedings of CAADRIA*, pages 911–926, 2004.

[9] Vuforia SDK, available at: http://www.vuforia.com

[10] B.-K. Seo, H. Park, J.-I. Park, S. Hinterstoisser, and S. Ilic, "Optimal local searching for fast and robust textureless 3D object tracking in highly cluttered backgrounds," *IEEE Transactions on Visualization and Computer Graphics*, volume 20, number 1, pages 99-110, 2014.