ViRC: Vision-Based Robust Calibration for Augmented Reality Using Optical See-Through Head-Mounted Displays

Naoya Makibuchi [†], Haruhisa Kato (member)[†], Hiromasa Yanagihara (member)[†]

Abstract

Augmented reality using optical see-through head-mounted displays (OSTHMDs) provides the user with a highly realistic experience compared to those using smartphones or tablet devices. It is necessary for the positional relationship between the user's eye and a virtual screen to be calibrated using input from the user. However, conventional calibration methods are highly sensitive to input errors. In this paper, we propose a vision-based robust calibration (ViRC) method using a fiducial marker, which can be used for any OSTHMD equipped with a camera. The ViRC method decomposes 11-DoFs calibration parameters into device-dependent parameters and user-dependent parameters. Once the device-dependent parameters are calculated, the user only has to perform a calibration phase for estimating the 4-DoFs user-dependent parameters. Experiments show that the ViRC method can decrease reprojection error by 83% compared with the conventional method. Consequently, users can observe correctly aligned superimpositions of computer graphics with little distortion.

Key words: Augmented Reality, Optical See-Through Head-Mounted Display (OSTHMD), HMD Calibration

1. Introduction

Head-mounted displays (HMDs) are apparatuses that are attached to a user's head and allow the user to enjoy videos while leaving the hands free. HMDs fall into two categories: (1) video see-through type and (2) optical see-through type. Video see-through head-mounted displays (VSTHMDs) provide screens that cover the entire area of the user's field of view. The user can see the real world through camera images on the screen. Augmented reality (AR) systems on VSTHMDs can be developed easily just by using a standard image registration method such as camera pose estimation based on marker detection¹⁾. However, VSTHMDs have two image-related problems from the user's viewpoint. One is that the visibility of the real world is restricted according to the camera capability such as viewing angle and resolution. The other is the display delay of the real world that is captured by the camera. Even a slight delay can give the user motion sickness. On the other hand, optical see-through head-mounted displays (OSTHMDs) use semi-transparent virtual screens that occupy a part of the user's field of view. OSTHMDs

Fig. 1 Two types of AR systems using OSTHMDs: Only see-through AR systems require calibration for image registration.

do not have the above problems that afflict VSTHMDs because the user can see the real world directly.

AR systems using OSTHMDs are also categorized into (1) video types and (2) see-through types. **Figs. 1(a)** and **(b)** show the respective explanatory illustrations of annotation systems, where the user can see the prices of commodities. Video AR systems display camera images on the virtual screen and then superimpose computer graphics (CGs) on the camera images. Thus, the image registration required is between the CGs and the camera images. This can be realized just by using standard image registration¹⁾²⁾, the same

Received March 31, 2016; Accepted June 2, 2016

[†] KDDI R&D Laboratories, Inc.

(2-1-15 Ohara, Fujimino-shi, Saitama, Japan)

as AR systems on VSTHMDs.

However, the user experience using video AR systems on OSTHMDs also has some issues. One is that the user sees actual objects twice, first in the real world and secondly on the virtual screen, which is shown in **Fig. 1(a)**. The other is that the CGs are displayed away from the real world in the user's field of view. Actually, the user is forced to take his or her gaze away from the real world when checking the CGs.

Meanwhile, see-through AR systems do not display camera images on the virtual screen and superimpose only CGs onto the appropriate positions in the virtual screen where the user sees the actual objects. Consequently, users are able to have a highly realistic AR experience as if the CGs really exist in the real world. However, see-through AR systems have some specific and challenging problems. These systems require image registration between the CGs and the real world, not camera images. In other words, these systems need to find the positions of actual objects in the virtual screen from the user viewpoint. However, it is difficult to detect the positions directly because only the user himself or herself can see the positions in his or her field of view. To realize this image registration, the positional relationships between a user's eye and a virtual screen is preliminarily calibrated.

The most common calibration method requires user input¹⁾. The user is required to move a cross-hair cursor in the virtual screen and then click the mouse button at the position where the cross-hair cursor coincides with a certain point of a fiducial marker (e.g., the original point of the marker). This process is called alignment. This method has an advantage in that it can be used for all OSTHMDs equipped with a camera simply by preparing a printed fiducial marker. However, if the user performs alignments carelessly, this can cause calibration errors. Recently, an automatic calibration method without user input has been proposed³⁾. Attaching an eye tracker to an OSTHMD enables calibration parameters to be calculated from a tracked eye position. This study has reported that eye tracking is more stable than user input. However, there are currently few commercial OSTHMDs equipped with eye trackers.

The motivation of this study is to propose a calibration method based on a fiducial marker, which is highly robust against alignment errors caused by the user. The proposed method does not require any additional sensor or instruments and can be used easily for most com-

mercial OSTHMDs. The structure of this paper is as follows. **Section 2** presents the mechanisms of the conventional calibration methods. **Section 3** explains the proposed method called vision-based robust calibration (ViRC). **Section 4** shows that the proposed method is able to achieve significant improvements. **Section 5** concludes this paper.

We have already developed the basic algorithm of ViRC method⁴⁾. This paper adds the detailed descriptions about the features of VSTHMD and OSTHMD applications in order to highlight the difficulty of OSTHMD calibration in **Section 1**. In addition, this paper clarifies the advantage of our evaluation system to provide additional explanations about conventional evaluation systems in **Section 4.1**. Furthermore, this paper gives a new experimental result that reveals another advantage of ViRC method in **Section 4.3**.

2. Related Work

OSTHMD's calibrations have been conducted with various types of tracking systems: vision-based systems¹⁾, magnetic-based systems⁵⁾⁶⁾, ultrasonic-based systems⁷⁾ and infra-red-based systems⁸⁾⁹⁾. All these systems with the exception of vision-based systems require additional external aids such as an IR camera, an ultrasonic transmitter and a magnetic transmitter, respectively. In contrast, the vision-based system only requires a camera device, which is built into most of the recently released commercial OSTHMDs. To achieve a low introduction cost, this study also adopts a vision-based system.

Kato *et al.* have proposed a vision-based calibration system using a fiducial marker¹⁾. Tuceryan *et al.* have proposed a system using a magnetic tracker, called SPAAM⁶⁾. Both systems use the same algorithm called direct linear transformation (DLT) to estimate calibration parameters. Here, the calibration parameters are defined by a transformation matrix from a 3-dimensional camera coordinate system to a 2-dimensional virtual screen coordinate system and is composed of 11 DoFs. In the calibration phase, two constraints about 2D-3D point correspondences are obtained for every alignment. Therefore, the alignment process needs to be repeated at least six times. However, the DLT algorithm is very sensitive to mismatches¹⁰⁾. Consequently, in actual practice, a user is forced to perform alignments more than six times, more often ten or twenty times.

In this paper, we propose a vision-based calibra-

Fig. 2 Configuration of the ViRC method: P_2 is a projection matrix composed of calibration parameters.

Fig. 3 Flowchart of the ViRC method

tion method that is robust against user misalignments. The novelty of the proposed method is that it applies the perspective-n-point (PnP)¹¹⁾ algorithm to OSTHMD calibration. In general, the PnP algorithm provides a more accurate camera pose than the DLT algorithm. However, the PnP algorithm cannot be applied directly to OSTHMD calibration because accurate intrinsic parameters about the positional relationship between the user's eye and a virtual screen are not known in advance. In the proposed method, the device-dependent parameters are first estimated using the PnP algorithm with the approximate values of the intrinsic parameters. Next, the user-dependent parameters are estimated through an optimization calculation that corrects the intrinsic parameters. Then, merging the device-dependent parameters and the user-dependent parameters results in all the calibration parameters being obtained.

3. Proposed Method: ViRC

We have developed the vision-based robust calibration (ViRC) method for the AR system using monocular OSTHMDs equipped with a camera. The configuration of the ViRC method is shown in **Fig. 2**. The virtual screen is positioned a focal length away from

Fig. 4 Alignment using a cross-hair cursor: The cross-hair cursor is moved by a mouse. Regarding the inside area, refer to **Section 4.1**.

the user's eye. The projection matrix P_1 that projects the 3D point X_1 on the marker coordinate system onto the 2D point m_1 on the virtual screen is decomposed as follows:

$$sm_1 = P_1 X_1 \quad (1)$$

$$= P_2 W_{2,1} X_1 \quad (2)$$

where m_1 and X_1 are homogeneous coordinate representations. s is a scale factor. $W_{2,1}$ is a transformation matrix from the marker coordinate system O_1 to the camera coordinate system O_2 and can be calculated through camera pose estimation based on marker detection¹⁾. P_2 is a projection matrix from O_2 to the virtual screen coordinate system and is composed of 11-DoF calibration parameters. The calibration parameters remain constant as long as the user does not move his/her OSTHMD glasses because the calibration parameters are based on the positional relationship between the user's eye and the virtual screen.

Fig. 3 shows the flowchart of the ViRC method. While P_2 is calculated through only one phase in the conventional DLT-based method, it is identified via two phases in the ViRC method: phases I and II. First, phase I estimates the device-dependent parameters that represent the approximate position of the virtual screen. Next, phase II estimates the 4-DoFs user-dependent parameters to obtain the positions of both the user's eye and the virtual screen accurately. Both phases require user alignments. Here, those who perform phase I and the following phase II may be different individuals. In other words, one user can perform phase II using the device-dependent parameters estimated by another user. Once phase I is completed, it does not need to be recomputed as long as the camera devices are fixed securely on OSTHMDs. On the other hand, phase II must be performed whenever a user wears or moves his/her OSTHMD. However, phase II only has to estimate the 4-DoFs user-dependent parameters because the device-

Fig. 5 Phase I calibration of the ViRC method: f^d , c_u^d , c_v^d and $W_{3,2}^d$ are the device-dependent parameters.

dependent parameters are obtained beforehand. Consequently, the ViRC method provides more robust calibration compared with the conventional method that estimates the 11-DoFs calibration parameters directly.

3.1 Alignment

Alignment is the process that collects point correspondences comprising a 3D object point in the camera coordinate system and its projected 2D point into the virtual screen ($W_{2,1}X_1, m_1$). **Fig. 4** shows this process.

3.2 Phase I: Estimating the Device-Dependent Parameters

In phase I, the approximate position of the virtual screen is calculated by the PnP algorithm¹¹⁾, which is generally used for estimating camera poses with camera intrinsic parameters and a minimum of four 2D-3D point correspondences. A diagrammatic representation of this phase is shown in **Fig. 5**.

Suppose that the perspective projection model is applied to the OTHMD, which means the virtual screen is the image plane, the normal from the user's eye to the virtual screen is the optical axis, and the user's eye is the optical center. Then, let A^d be intrinsic parameters used in this phase, and be denoted as follows:

$$A^d = \begin{pmatrix} f^d & 0 & c_u^d \\ 0 & f^d & c_v^d \\ 0 & 0 & 1 \end{pmatrix} \quad (3)$$

where f^d is the distance between the user's eye and the virtual screen, and (c_u^d, c_v^d) is the point on the pixel coordinate system of the virtual screen (u, v) where the normal and the virtual screen intersect. The units of f^d , c_u^d and c_v^d are pixels. Because the accurate values of these three parameters are unknown at first, this phase sets them as approximate values based on the specification of the OTHMD. For example, when using OTH-

Fig. 6 Phase II calibration of the ViRC method: f^u , c_u^u , c_v^u and k^u are the user-dependent parameters.

MDs that place f millimeters ahead the virtual screen whose resolution is $w \times h$ pixels and the length of whose diagonal is d millimeters, f^d , c_u^d and c_v^d are respectively set as $f k^d$, $\frac{w}{2}$ and $\frac{h}{2}$ where the pixel density (the ratio of pixel to millimeter) $k^d = \frac{\sqrt{w^2 + h^2}}{d}$.

Let $W_{3,2}^d$ be a transformation matrix from the camera coordinate system O_2 to the approximate coordinate system O_3 corresponding to the approximate user's eye position. $W_{3,2}^d$ is calculated by the PnP algorithm according to **Eq. 4** with the given A^d and a minimum of four 2D-3D point correspondences

$$sm_1 = A^d W_{3,2}^d W_{2,1} X_1 \quad (4)$$

Thus, the device-dependent parameters calculated in phase I are defined by f^d , c_u^d , c_v^d and $W_{3,2}^d$. The superscripts d of these expressions mean that they are the device-dependent parameters.

It can be assumed that the device-dependent parameters represent the rough position of the virtual screen for the following reason. A frustum is defined by the approximate eye position and the virtual screen, as shown in **Fig. 5**. In this phase, the precise position of the virtual screen cannot be calculated because the exact pixel density k is unknown. Conversely, given the exact pixel density, this frustum provides the precise position of the virtual screen. Therefore, this frustum in this phase is used as a clue to find the precise position of the virtual screen in the following phase.

3.3 Phase II: Estimating the User-Dependent Parameters

In phase II, the positions of both the virtual screen and the user's eye are estimated accurately by non-linear minimization under the given device-dependent parameters. **Fig. 6** shows a diagrammatic representation of this phase. Let $W_{4,3}^u$ be the transformation

matrix from O_3 to the virtual screen coordinate system O_4 , and $W_{5,4}^u$ be the transformation matrix from O_4 to the actual position of the user's eye coordinate system O_5 . Thus, $W_{4,3}^u$ and $W_{5,4}^u$ are introduced as correction terms converting the approximate position of the user's eye into the precise position.

$W_{4,3}^u$ is denoted as follows:

$$W_{4,3}^u = \begin{pmatrix} 1 & 0 & 0 & \frac{c_u^d}{k^u} \\ 0 & 1 & 0 & \frac{c_v^d}{k^u} \\ 0 & 0 & 1 & -\frac{f^d}{k^u} \\ 0 & 0 & 0 & 1 \end{pmatrix} \quad (5)$$

where k^u is the pixel density that has been introduced as one of the user-dependent parameters. Estimating the k^u accurately means identifying the actual position of the virtual screen in the frustum calculated in phase I.

In addition, the position of the user's eye relative to the virtual screen in phase II is, in the same way as in **Eq. 3**, denoted by A^u as follows:

$$A^u = \begin{pmatrix} f^u & 0 & c_u^u \\ 0 & f^u & c_v^u \\ 0 & 0 & 1 \end{pmatrix} \quad (6)$$

where f^u and (c_u^u, c_v^u) are the remaining user-dependent parameters. Then, $W_{5,4}^u$ is denoted as follows:

$$W_{5,4}^u = \begin{pmatrix} 1 & 0 & 0 & -\frac{c_u^u}{k^u} \\ 0 & 1 & 0 & -\frac{c_v^u}{k^u} \\ 0 & 0 & 1 & \frac{f^u}{k^u} \\ 0 & 0 & 0 & 1 \end{pmatrix} \quad (7)$$

As above, the P_2 is composed of **Eqs. 5, 6, 7** and $W_{3,2}^d$ according to **Eq. 8**.

$$P_2 = A^u W_{5,4}^u W_{4,3}^u W_{3,2}^d \quad (8)$$

The projection formula of X_1 onto m_1 in the ViRC method is given by substituting **Eq. 8** into **Eq. 2**. The f^u , c_u^u , c_v^u and k^u of the unknown user-dependent parameters are estimated by the Levenberg-Marquardt algorithm to minimize reprojection error through the projection formula, when the initial values are set as the values of f^d , c_u^d , c_v^d and k^d , respectively. This calculation requires a minimum of four 2D-3D point correspondences because of the number of unknown parameters. This optimization calculation means that the position of the user's eye is estimated accurately while finding the precise position of the virtual screen in the frustum. Thus, the ViRC method estimates the positions of both the user's eye and the virtual screen simultaneously with a small number of unknown parameters.

Fig. 7 Experimental environment

4. Experiments

4.1 Evaluation System

Evaluating an OSTHMD's calibration has also been a troublesome problem because only users wearing it can observe the superimposed scenes from the calibration result. Some studies⁽¹²⁾⁽¹³⁾ employ tablet devices called evaluation boards. First, the evaluation systems display small spheres on the boards at various positions (ground truths). Next, the systems superimpose the corresponding marks onto the virtual screens using the estimated calibration parameters. Then, users are requested to touch the boards at the perceived back-projected positions. The differences between the back-projected positions and ground truth positions on the boards are defined as errors. Another study⁽¹⁴⁾ uses trailable stylus pens and requests users to align the cross-hair cursors on the virtual screens in order to obtain the ground truths of 2D points. In this case, the differences between superimposed positions using the estimated calibration parameters and ground truth positions on the virtual screens are defined as errors. However, above-mentioned evaluation processes depend to a large extent on the alignment skills of individual users.

In this paper, we develop a video see-through system instead, similar to circle detection evaluation⁽¹⁵⁾. This system permits the recording of superimposed scenes in addition to the accurate quantitative evaluation of OSTHMD's calibration. As shown in **Fig. 7 (a)**, we use Brother AiRScouter equipped with Logicool Quick-Cam Pro for Notebooks. The AiRScouter places 30 centimeters ahead the virtual screen whose resolution is 800×600 pixels. Additionally, we place Logicool HD Webcam C525 as another camera in the position of the user's eye (hereinafter called "eye camera") so as to be slidable independently of the OSTHMD to simulate an-

Fig. 8 Reprojection error with dataset 2: $\sigma = 5$ and the number of trials is 500.

other user's eye position.

The video see-through system is fixed on a desk as shown in **Fig. 7 (b)**. In practice, alignments are performed by the user input, but in the experiments, they are conducted automatically using marker detection as follows. A fiducial marker is placed somewhere above the desk. Then, the position of its projected 2D point on the virtual screen is identified by applying a marker detection algorithm to images captured by the eye camera. This procedure henceforth is called marker-based alignment. In this way, the accurate ground truths of point correspondences can be obtained that are independent of the user's ability to perform alignment.

First, 100-point correspondences (dataset 1 for calibration) are corrected using marker-based alignment when the eye camera is positioned at a certain position. Then, Gaussian noise with zero mean and standard deviation $\sigma = 5$ pixels is added to the coordinates of the 2D points. This is based on the assumption that in actual practice there are user misalignments. Dataset 1 is only used for phase I of the ViRC method. Next, the additional 100-point correspondences (dataset 2 for calibration) are obtained after the eye camera is slid about 5 millimeters. Dataset 2 is used for both phase II of the ViRC method and the conventional methods. Finally, alternative 100-point correspondences (dataset for test) are collected for use in evaluating the accuracy of each method.

Experiment 1 reveals the dependency of each method on the number of alignments. Each calibration method is performed with n point correspondences with noise of $\sigma = 5$ pixels, which is selected randomly from dataset 2. Then, experiment 2 reveals the dependency of each method on the positions of 2D points included in the point correspondences. From dataset 2, 50-point cor-

Fig. 9 Superimposed cube: DLT (error = 33.4 pixels) and ViRC (error = 5.5 pixels)

respondences are extracted so that the 2D points exist only inside the 400×300 pixel area of the virtual screen (dataset 2-inside), as shown in **Fig. 4**.

4.2 Experimental Result 1: Revealing Dependency On the Number of Alignments

Fig. 8 shows the median of reprojection errors with dataset 2 during 500 trials using each calibration method for n alignments. The reason for using the median is because the re-projection error becomes ridiculously massive according to the configurations of 3D points (e.g., coplanar points). The processing times of the ViRC method and the conventional DLT-based method are 0.40 ms and 0.05 ms on average, respectively. Both methods are executed on Intel Core i7 2600K 3.4 GHz. The ViRC method requires more computational cost than the DLT method because of iterative computation, but the processing time is still sufficiently practical because calibration is executed only once after the user puts on the HMD glasses.

“DLT (dataset 1)” is the calibration result of the DLT method with 100 alignments in dataset 1; this simulates the situation where one user uses the calibration result estimated by another user. “PnP” is the calibration result of the PnP method using the same approximate intrinsic parameters used in phase I of the ViRC method. The ViRC method achieves more positive results overall compared to both the DLT method and the PnP method. Specifically, compared with the DLT method, it can decrease reprojection error by 83% when $n = 6$, from 33.4 pixels to 5.5 pixels. The reason for making a comparison when $n = 6$ is that the minimum value of n

Fig. 10 Reprojection error with datasets 2 and 2-inside when $n = 4$: $\sigma = 5$ and the number of trials is 500.

required for the DLT method is 6 while that required for the ViRC method is 4. **Fig. 9** shows the superimposition of the computer graphic of a cube onto the virtual screen using each of the calibration results when $n = 6$. The shape of the superimposed cube is free from alignment errors in the ViRC method while the DLT method produces some distortion (**Figs. 9(a)** and **(b)**). In addition, the edge of the cube is misaligned at the boundary of the virtual screen in the DLT method, but it is properly aligned in the proposed method (**Figs. 9(c)** and **(d)**).

The DLT method is needed to estimate the positions of both the user's eye and the virtual screen without any prior information. On the other hand, the ViRC method can estimate them with the rough position of the virtual screen, which is calculated beforehand as device-dependent parameters. Therefore, the ViRC method produces a more robust calibration than the DLT method. The superior results for the PnP method compared to the DLT method means that misalignment (simulated by noise) has a more significant impact on the reprojection error than does the approximation error of the intrinsic parameters. The reason the ViRC method gets a better result than the PnP method is that the position of the user's eye is modified from the initial approximate intrinsic parameters to accurate values.

4.3 Experimental Result 2: Revealing Dependency On the Positions of 2D points

Fig. 10 shows the median of reprojection errors using the DLT method and the ViRC method with datasets 2 and 2-inside during 500 trials where $n = 6$. Dataset 2 includes point correspondences where 2D points exist on the overall range of the virtual screen. Meanwhile, dataset 2-inside includes point correspondences where 2D points exist only on the inside area. The result indicates that using only the narrow ranges

of 2D points negatively affects the re-projection errors; specifically, those using the DLT method and the ViRC method increase by 2.50 and 1.18 times, respectively.

Generally, camera pose estimation also become unstable with 2D-3D point correspondences in the small regions of camera images. This suggests that users should perform alignments with 2D points outside the virtual screen for both calibration methods. However, as for the ViRC method, phase I calibration using numerous alignments (100 in this paper) leads to a high-accuracy phase II calibration (error = 6.5 pixel in **Fig. 10**) even when inside points are used.

5. Conclusion

In this paper, we proposed a calibration method called ViRC for augmented reality applications using OSTHMDs. The ViRC method is a simple configuration requiring only a fiducial marker and does not need any additional equipment. The ViRC method divides calibration parameters into device-dependent parameters and user-dependent parameters. Computing the device-dependent parameters beforehand leads to a decrease in the number of unknown parameters in the user calibration phase. Therefore, the ViRC method produces a more robust calibration compared with the conventional DLT-based method. Experiments showed that the ViRC method could decrease reprojection error by up to 83% compared with the conventional method. Consequently, users could obtain correctly aligned superimpositions of the computer graphics with little distortion.

References

- 1) H. Kato and M. Billinghurst : "Marker tracking and hmd calibration for a video-based augmented reality conferencing system," in Proceedings of the 2nd IEEE and ACM International Workshop on Augmented Reality (IWAR '99), pp. 85–94 (1999)
- 2) T. Nöll, Tobias, A. Pagani, and D. Stricker : "Markerless Camera Pose Estimation - An Overview," in Visualization of Large and Unstructured Data Sets - Applications in Geospatial Planning, Modeling and Engineering (IRTG 1131 Workshop), pp. 45–54 (2010)
- 3) Y. Itoh, and G. Klinker : "Interaction-free calibration for optical see-through head-mounted displays based on 3d eye localization," in Proceedings of the IEEE Symposium on 3D User Interfaces 2014, pp. 75–82 (2014)
- 4) N. Makibuchi, H. Kato, and A. Yoneyama : "Vision-Based Robust Calibration for Optical See-Through Head-mounted Displays," in Proceedings of the IEEE International Conference on Image Processing (ICIP '13), pp. 2177–2181 (2013)
- 5) A. L. Janin, D. W. Mizell, and T. P. Caudell : "Calibration of head-mounted displays for augmented reality applications," in Proceedings of the IEEE Virtual Reality International Symposium (VRAIS '93), pp. 246–255 (1993)
- 6) M. Tuceryan, Y. Genc, and N. Navab : "Single-point active alignment method (spaam) for optical see-through hmd calibration for augmented reality," Presence: Teleoperators and Virtual Environments, vol. 11, no. 3, pp. 259–276 (2002)

- 7) C. B. Owen, J. Zhou, A. Tang, and F. Xiao : “Display-relative calibration for optical see-through head-mounted displays,” in Proceedings of the IEEE and ACM International Symposium on Mixed and Augmented Reality (ISMAR '04), pp. 70–78 (2004)
- 8) S. J. Gilson, A. W. Fitzgibbon, and A. Glennerster : “Spatial calibration of an optical see-through head-mounted display,” *Journal of Neuroscience Methods*, vol. 173, no. 1, pp. 140–146 (2008)
- 9) F. Kellner, B. Bolte, G. Bruder, U. Rautenberg, F. Steinicke, M. Lappe, and R. Koch : “Geometric calibration of head-mounted displays and its effects on distance estimation,” *IEEE Transactions on Visualization and Computer Graphics*, vol. 18, no. 4, pp. 589–596 (2012)
- 10) M. Axholt, M. Skoglund, S. D. Peterson, M. D. Cooper, T. B. Schoen, F. Gustafsson, A. Ynnerman, and S. R. Ellis : “Optical see-through head mounted display direct linear transformation calibration robustness in the presence of user alignment noise,” in Proceedings of the Human Factors and Ergonomics Society 54rd Annual Meeting (HFES '10), 54(28) pp. 2427–2431 (2010)
- 11) V. Lepetit, F. Moreno-Noguer, and P. Fua : “Epnp: An accurate o (n) solution to the pnp problem,” *International Journal of Computer Vision*, vol. 81, no. 2, pp. 155–166 (2009)
- 12) E. McGarrity, Y. Genc, M. Tuceryan, C. Owen, and N. Navab : “A new system for online quantitative evaluation of optical see-through augmentation,” in Proceedings of the IEEE and ACM International Symposium on Augmented Reality, pp. 157–166 (2001)
- 13) N. Navab, S. Zokai, Y. Genc, E. M. Coelho : “An on-line evaluation system for optical see-through augmented reality,” in Proceedings of the IEEE Virtual Reality 2004, pp. 245–246 (2004)
- 14) A. Tang, J. Zhou, and C. Owen : “Evaluation of calibration procedures for optical see-through head-mounted displays,” in Proceedings of the 2nd IEEE and ACM International Symposium on Mixed and Augmented Reality (ISMAR '03), pp. 161–168 (2003)
- 15) P. Maier, A. Dey, C. A. L. Waechter, C. Sandor, M. Tonniss, and G. Klinker : “An empiric evaluation of confirmation methods for optical see-through head-mounted display calibration,” in Proceedings of the 10th IEEE International Symposium on Mixed and Augmented Reality (ISMAR '11), pp. 267–268 (2011)

Naoya Makibuchi received the B.E. and M.E. degrees in engineering from Tokyo Institute of Technology, Japan, in 2009 and 2011, respectively. In 2011, he joined KDDI Co. Ltd. and has been engaged in the research and development of augmented reality. He is an associate research engineer of the Software Integration Laboratory at KDDI R&D Laboratories, Inc.

Haruhisa Kato received the B.E. and M.E. degrees in electrical and electronics engineering from Kobe University in 1997 and 1999, respectively. He also received Ph. D degree from Tokyo Institute of Technology in 2009. In 1999, he joined KDD Co. Ltd. and has been engaged in the research and development of video coding, augmented reality and user interface. He is a R&D manager of the Software Integration Laboratory at KDDI R&D Laboratories, Inc.

Hiromasa Yanagihara received the B.E., M.E., and Ph.D degrees from Nagoya University in electronics engineering, electrical engineering and computer science, in 1988, 1990 and 2014, respectively. In 1990, he joined KDD Co.Ltd. Since 1997, he has been with its R&D division. His current research interests include video coding, video transmission over IP, image processing and related multimedia systems. He is currently executive director in multimedia division in KDDI R&D Laboratories, Inc.