Multimedia Tools and Applications manuscript No.
(will be inserted by the editor)

Addressing Temporal Inconsistency in Indirect Augmented Reality

Fumio Okura · Takayuki Akaguma · Tomokazu Sato · Naokazu Yokoya

Received: 31 May 2015 / Accepted: 29 December 2015

Abstract Indirect augmented reality (IAR) employs a unique approach to achieve high-quality synthesis of the real world and the virtual world, unlike traditional augmented reality (AR), which superimposes virtual objects in real time. IAR uses pre-captured omnidirectional images and offline superimposition of virtual objects for achieving jitter- and drift-free geometric registration as well as high-quality photometric registration. However, one drawback of IAR is the inconsistency between the real world and the pre-captured image. In this paper, we present a new classification of IAR inconsistencies and analyze the effect of these inconsistencies on the IAR experience. Accordingly, we propose a novel IAR system that reflects real-world illumination changes by selecting an appropriate image from among multiple pre-captured images obtained under various illumination conditions. The results of experiments conducted at an actual historical site show that the consideration of real-world illumination changes improves the realism of the IAR experience.

Keywords Indirect augmented reality · Temporal inconsistency · Spatial inconsistency · Cultural heritage

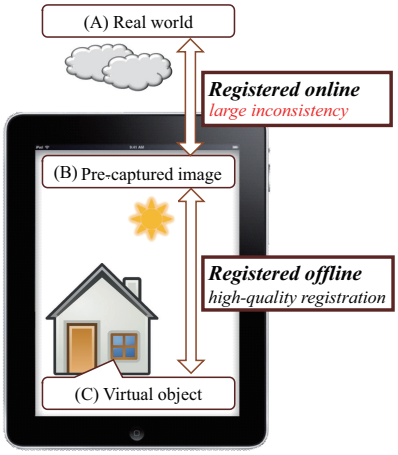
1 Introduction

Video see-through augmented reality (AR) [Azuma (1997), Zhou et al. (2008)] using mobile devices (e.g., smartphones and tablets) has various applications such as

This research was partially supported by JSPS KAKENHI 23240024, 26330193, 15H06362, 25-7448, and by the NAIST Advanced Research Partnership Project.

F. Okura
Department of Intelligent Media, ISIR, Osaka University, Mihogaoka 8-1, Ibaraki, Osaka, Japan
Tel.: +81-6-6879-8422
Fax: +81-6-6877-4375
E-mail: okura@am.sanken.osaka-u.ac.jp

T. Akaguma, T. Sato, N. Yokoya
Vision and Media Computing Lab., Nara Institute of Science and Technology, Takayama 8916-5, Ikoma, Nara, Japan



(a) How inconsistency occurs in IAR.

(b) Example of inconsistency between the real world and a pre-captured image.

Fig. 1 Inconsistency in IAR. (a): Virtual objects (C) are superimposed on pre-captured images (B) in the offline process. Significant inconsistencies between the real world (A) and pre-captured images (B) are expected. (b): An example shows inconsistencies due to changes in illumination condition in the real world. In the IAR scene shown in the mobile display, an historic building is superimposed on a pre-captured image. The real world is cloudy, whereas the IAR scene is sunny.

the visualization of lost buildings at cultural heritage sites [Zoellner et al. (2009)]. In order to achieve highly realistic AR, it is necessary to obtain geometric registration, i.e., geometric alignment between the real world and virtual objects, and photometric registration, i.e., consistency in appearance, including lighting and shadowing. In most mobile AR applications, geometric and photometric registration should be performed in real time with limited resources. In general, geometric registration is achieved by estimating the pose (position and orientation) of cameras used for virtual object superimposition. Numerous methods have been proposed for geometric registration using mobile devices, which can be classified into sensor-based [Hollerer et al. (1999), Tenmoku et al. (2003)] and vision-based [Klein and Murray (2009), Ventura and Hollerer (2012), Schops et al. (2014)] approaches. It is known that such real-time methods often lead to misalignment (e.g., jitter and drift) due to errors in camera pose estimation, which could detract from the immersion and naturalness of AR. On the other hand, for photometric registration, the illumination environment of the real world is often estimated from captured images in order to apply it to virtual objects. However, owing to limited resources, it is difficult for mobile systems to employ resource-hogging techniques [Gruber et al. (2012), Kán and Kaufmann (2012), Lensing and Broll (2012)], which allow high-quality photometric registration.

Recently, indirect augmented reality (IAR), which differs considerably from conventional registration techniques, has been proposed [Wither et al. (2011)]. In IAR, as shown in Figure 1, a pre-captured image (B) of the real world and the virtual object (C) are initially registered geometrically and photometrically with

high quality in the offline process. In the online process, an appropriate part of the pre-rendered omnidirectional AR image is presented to a user on a mobile device, based on the roughly estimated orientation of the device. Unlike conventional AR methods that superimpose virtual objects in real time, IAR principally achieves

- geometric registration without any perceptible misalignment between pre-captured images (B) and virtual objects (C), and
- high-quality photometric registration by offline superimposition of virtual objects,

while incurring low computational costs on mobile devices in the online process. In this regard, IAR has an inherent potential for super-photorealistic AR in certain applications such as architectural simulation using mostly static virtual objects. In particular, cultural heritage applications [Liestol and Morrison (2013)] have leveraged the advantages of IAR, which achieves ultra-stable and occlusion-free superimposition of static virtual objects.

An important concern with regard to IAR approaches is the inconsistency between the pre-captured image (B) and the real world (A). In practice, some inconsistencies are known to have a relatively minor effect on the IAR experience. For example, to track the device orientation, the original IAR [Wither et al. (2011)] employs only a small inertial sensor, the accuracy of which is often insufficient for traditional AR applications; however, most users do not notice the misalignment due to the orientation error. Nevertheless, there are several sources of IAR inconsistencies, including geometric misalignment, dynamic objects, and illumination changes. In order to exploit IAR effectively in the AR field, it is necessary to analyze IAR inconsistencies between the pre-captured image and the real world.

Toward this end, the contributions of the present study can be summarized as follows:

- We classify IAR inconsistencies into spatial and temporal inconsistencies. Our experiment shows that temporal inconsistencies, which have not been investigated extensively thus far, have a significant effect on the IAR experience.
- We address temporal inconsistencies caused by changes in real-world illumination, and we experimentally verify the resulting improvement in the IAR experience at an actual historical site.

This paper is an extended version of two short conference papers [Akaguma et al. (2013), Okura et al. (2014)]. As compared to the previous versions, the inconsistency classification (Section 3), proposed system (Section 5), and experimental analysis (Section 6) are described in greater detail. Moreover, an additional experiment (Section 7) is described herein.

The remainder of this paper is organized as follows. Section 2 briefly summarizes IAR techniques. Section 3 describes the classification of IAR inconsistencies into spatial and temporal inconsistencies. Section 4 presents the results of a public experiment conducted using a historical IAR application, which shows that temporal inconsistencies, especially those caused by real-world illumination changes, have a significant effect on the realism of the IAR experience. Section 5 describes the proposed IAR system for addressing inconsistencies caused by illumination changes. Section 6 describes an experiment conducted to show that the consideration of real-world illumination improves the IAR experience in practical situations. Section 7 describes an experiment conducted to derive the relationship between

realism and illumination changes. Finally, Section 8 summarizes our findings and concludes the paper with a brief discussion on the scope for future work.

2 Indirect AR: A Brief Overview

This study is based on IAR, which was originally introduced by Wither et al. [Wither et al. (2011)]. Several web-based applications for the interactive exploration of the virtualized real world have traditionally employed omnidirectional images [Chen (1995), Uyttendaele et al. (2004)], resulting in the emergence of some of the most well-known virtual reality applications, e.g., Google Street View [Anguelov et al. (2010)]. Although some applications superimpose virtual objects onto omnidirectional images [Debevec (1998), Grosch (2005), Okura et al. (2015)], they are not used at the location where the images were captured. In contrast, IAR uses such augmented omnidirectional navigation applications on-site, and it functions as an AR application that does not require real-time registration between real scenes and virtual objects. Owing to its inherent advantages, such as stability and low computational cost, IAR has been employed for the visualization of cultural heritage sites [Liestol and Morrison (2013)] and construction sites [Côté et al. (2013)].

Unlike conventional mobile AR, where all registration processes are performed in real time, IAR is divided into online and offline processes. In the offline process, images are captured by an omnidirectional camera at the site where users will experience the IAR application. Further, virtual objects are rendered with high quality using the pre-captured images. Any type of manual editing is possible, and thus, we can create *cinema-quality* IAR scenes in principle. In the online process, a mobile device shows planar perspective images obtained from the augmented omnidirectional image according to the pose of the mobile device, which is usually acquired by embedded sensors.

3 IAR Inconsistency Classification

Inconsistencies in traditional AR are often categorized into geometric and photometric inconsistencies [Kanbara and Yokoya (2002)]. However, this well-known classification, which has been proposed for inconsistencies between real scenes and virtual objects, is not suited for IAR inconsistencies. Therefore, we introduce a new classification for IAR consistencies, i.e., we classify them into spatial and temporal inconsistencies. In this section, we review previous works that have analyzed IAR inconsistencies.

3.1 Spatial Inconsistency

If users experience an IAR application at a location different from that *where* the omnidirectional images were captured, they may find differences in viewpoints. Similarly, if the presented IAR scene includes orientation errors due to errors in device orientation measurement, users may notice the misalignment.

Most IAR studies deal with spatial inconsistencies, and they have experimentally confirmed that the impact of spatial inconsistencies is much smaller than that of geometric inconsistencies in traditional AR between real scenes and virtual objects, if the viewpoint difference and the orientation error are sufficiently small. This is because such misalignments are accommodated through traditional mobile AR and digital photographic interfaces owing to the position difference between the center of the device camera and the human eye. The original IAR study [Wither et al. (2011)] investigated how the distances among the user's location, the virtual object, and the position where the omnidirectional images were captured affects the IAR experience. Madsen et al. [Madsen and Stenholt (2014)] compared the effect of the device orientation error on IAR and traditional AR. These works showed the robustness of IAR against spatial inconsistencies. Liestol and Morrison [Liestol and Morrison (2013)] employed large viewpoint changes (e.g., bird's eye view) for IAR and tried to determine which view is preferred in an application for visualizing historical architecture.

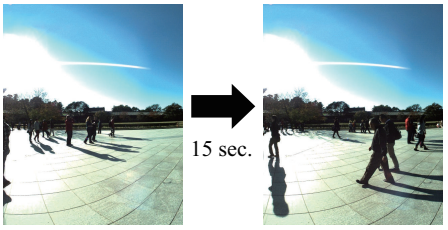
Some recent IAR applications have addressed spatial inconsistencies by presenting viewpoint changes along with the user's behavior. An IAR application developed by Yamamoto et al. [Yamamoto et al. (2014)] presents a small disparity simulated using the device orientation. Further, 3D reconstruction techniques allow viewpoints to be changed freely in an IAR environment [Waegel (2014)].

3.2 Temporal Inconsistency

When the appearance (e.g., illumination, dynamic objects) of the real scene is different from that *when* the omnidirectional images were captured, the inconsistencies become more severe. In typical IAR applications, temporal inconsistencies can be more problematic than spatial ones for providing a natural IAR experience. We can control the user's location in many cases, e.g., by placing a sign indicating the location for the experience. However, especially in outdoor environments, most temporal factors (e.g., weather) are uncontrollable. We further categorize temporal inconsistencies in terms of the duration for which the appearance change occurs, as shown in Figure 2.

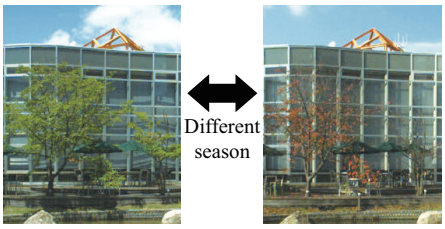
a) Order of seconds: Dynamic objects In pre-captured images, dynamic objects, e.g., pedestrians (see Figure 2(a)), are captured, and users of the IAR system may perceive large differences from the real-world scene. Capturing and presenting real-time appearances of the dynamic objects are straightforward ways to overcome this problem; however, they cause real-time inconsistencies between the objects and the pre-captured image, as in the case of traditional AR.

The original IAR study [Wither et al. (2011)] provided an interesting insight for dynamic objects. In an environment with dynamic objects, traditional AR was compared with an IAR application on the basis of an omnidirectional image including no dynamic objects. It was found that users perceive greater degradation in traditional AR owing to the occlusion of virtual objects by dynamic objects. For landscape simulation, which is considered as the main application of IAR, using occlusion-free pre-captured images is advantageous as it improves the IAR experience by removing undesired occluders that hide virtual contents from the users.



(a) Order of seconds: the appearance of pedestrians varies over 15 seconds.

(b) Order of minutes/hours/days: the real world is cloudy, whereas the IAR scene is sunny.



(c) Order of months: the appearance of trees changes depending on the season.

Fig. 2 Examples of temporal inconsistencies.

b) Order of minutes/hours: Photometric inconsistency Real-world illumination varies according to changes in the weather and the position of the sun. The illumination globally changes the appearance of scenes, e.g., sky color, shading, and brightness. Such inconsistencies can have a significant effect on the IAR experience.

One way to overcome this problem is to utilize omnidirectional images captured under various weather and illumination conditions. This paper reports the effect of considering illumination changes by using an actual prototype system that selects and displays an image similar to the real scene captured in real time.

c) Order of months: Changes in scene structure The appearance of vegetation and buildings can vary not only with their age but also with the season or construction works. When such long-term changes occur, the images may be captured once again. The IAR approach proposed in this paper does not deal with this problem directly.

To the best of our knowledge, this study is the first to focus on photometric issues in IAR, i.e., the second category of temporal inconsistency. Most IAR studies have dealt with spatial inconsistencies. Nevertheless, we found that temporal inconsistencies, especially those of the photometric type, have a significant effect on the IAR experience, through the experiment described in Section 4. Therefore, this paper proposes an IAR approach that addresses photometric inconsistencies, and it focuses on real-world illumination issues through subjective evaluations. The experimental results indicate that the prototype system that considers illumination changes improves the realism of IAR.

(a) Visualization of lost buildings.

(b) Visualization of re-colored statues in the Great Buddha Hall.

Fig. 3 IAR scenes presented to participants.

4 Impact of IAR Inconsistencies

To investigate how inconsistencies affect the IAR experience, we developed an application based on the original IAR for a virtual history experience at the Todaiji temple, a UNESCO World Heritage site in Nara, Japan, and we performed a public experiment in collaboration with Todaiji. We focused on the temporal inconsistency caused by illumination changes, which is expected to be a major issue for typical IAR applications in outdoor environments. In these experiments, virtual models of ancient buildings and heirloom-quality statues, which no longer exist or have aged considerably[1], were superimposed on the pre-captured omnidirectional images in the offline process semi-manually. This section is a detailed version of the experiment section of a previous conference paper [Akaguma et al. (2013)].

4.1 Experimental Condition

Experimental Platform This application consists of five IAR scenes in the Todaiji temple, where an omnidirectional image was pre-captured for each scene. The appearance of lost buildings was shown at two scenes, and statues inside the existing building (The Great Buddha Hall) were superimposed on the other scenes, as shown in Figure 3. The IAR scenes were displayed on the following tablet devices in full-screen mode: iPad2, iPad 4th generation, and iPad mini (Apple Inc.). Ladybug3 (Point Grey Research, Inc.), an omnidirectional multi-camera unit, was used to capture the omnidirectional images. In order to enhance user experience of the application, some sound effects were added. The virtual objects were rendered using 3ds Max (Autodesk, Inc.) with manual post-editing. The public experiment was performed from 3 pm to 5 pm under a cloudy sky; the omnidirectional images had been captured around noon under a sunny sky.

[1] These models were created with the cooperation of the Graduate School of Media Design, Keio University.

We used omnidirectional images without dynamic objects to overcome inconsistencies caused by the dynamic objects, as in the case of a previous study [Wither et al. (2011)] (see Section 3 for details). Because it was difficult to capture omnidirectional images without any dynamic objects, we removed them from the pre-captured images by applying a temporal median filter to an omnidirectional sequence [Arai et al. (2010)]. The dynamic object removal process is described in Section 5.1.1.

Participants Forty-six participants, whose ages are summarized in Table 1, experienced our application. All the participants were recruited from among the general public.

Task and Procedure Groups of a few participants navigated to each location within 3 m of the location where the omnidirectional image was captured. During their experience, a navigator explained the historical background of the content presented. After they experienced all the IAR contents, the users were asked to fill out a questionnaire consisting of the following questions:[2]

Q1 Did you perceive consistency between the virtual objects and the real scenes displayed on the device?
Q2 Did you perceive consistency in location between the displayed scenes and the real scene that you saw with your eyes?
Q3 Did you perceive consistency in appearance and illumination between the displayed scenes and the scene that you saw with your eyes?

Q1 evaluates the consistency between the pre-captured images (B) and the virtual objects (C) shown in Figure 1. Q2 and Q3 evaluate the spatial and temporal inconsistencies, especially photometric ones, between the real world (A) and pre-captured images (B). For each question, the participants selected a Likert scale value from 1 (very inconsistent) to 5 (very consistent).

4.2 Result and Discussion

Figure 4 shows the results of the questionnaire. In the case of spatial inconsistency, over 70 % of the participants answered "consistent" (4 or 5) for Q1 and Q2, with regard to the inconsistency between the virtual objects and the pre-captured images. In our experimental situation, the temporal inconsistency between the real world and the pre-captured image was found to have a greater effect than the

Table 1 Age distribution of participants.

Age	20's	30's	40's	50's	60's	>70
Number of participants	5	4	6	12	12	7

[2] The English translation of the earlier paper [Akaguma et al. (2013)] did not preserve the meanings of the original questions (which were written in Japanese) accurately; therefore, we have modified the English sentences according to the original meaning.

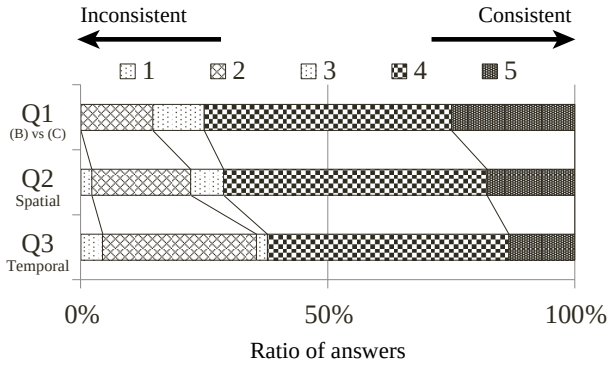


Fig. 4 Stacked bar graphs showing the degree of consistency. Q1 measures the inconsistency between the pre-captured scene and virtual objects. Q2 and Q3 respectively evaluate the spatial and temporal inconsistencies between the real scene and the pre-captured scene.

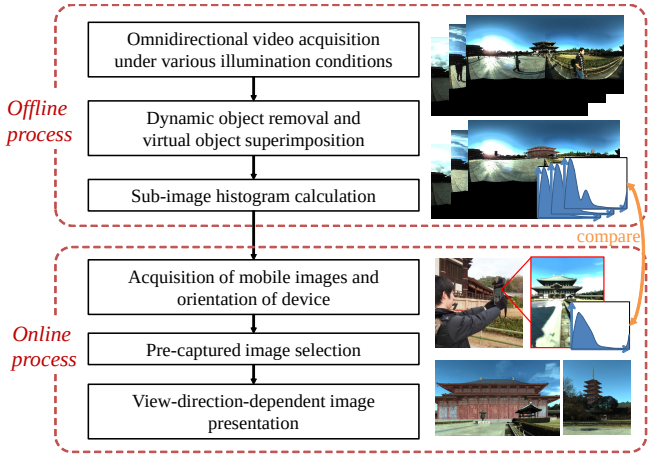


Fig. 5 Flow of the prototype system.

other inconsistencies according to the results of Q3, although around 60 % of the participants perceived consistency.

5 IAR Addressing Temporal Inconsistency Caused by Illumination Change

Now, we observe that temporal inconsistency, particularly photometric inconsistency, is a practical issue in IAR. Therefore, we propose a novel IAR approach to address photometric inconsistency, i.e., the second category of temporal inconsistency in our classification (see Section 3). We leverage knowledge from a previous study [Wither et al. (2011)] to exploit dynamic objects, which constitute another source of temporal inconsistency.

Our IAR approach is divided into online and offline processes (see Figure 5) as in the case of conventional IAR approaches. In contrast to the other approaches,

during the offline process, omnidirectional images are captured under various illumination conditions at the same position. First, we create omnidirectional images by removing dynamic objects. Then, virtual objects are superimposed onto the omnidirectional images using a high-quality rendering method that considers the real-world illumination. Omnidirectional augmented images and sub-image histograms corresponding to pre-captured images are stored into a database. In the online process, a pre-generated augmented image, whose illumination condition is the most similar to that of the real scene acquired by a mobile device, is selected from the database by comparing the sub-image histograms. Finally, the appropriate part of the selected image is displayed on the mobile device depending on its orientation.

5.1 Offline Process

In the offline process, first, omnidirectional videos are captured at predefined positions multiple times under various illumination conditions. A database consisting of augmented omnidirectional images is generated, from which dynamic objects are removed, and the corresponding sub-image histograms are obtained. The histograms will be used in the online process to efficiently determine a similar image in terms of the illumination condition.

5.1.1 Dynamic Object Removal and Virtual Object Superimposition

We employ omnidirectional images without dynamic objects for improving the visibility of the contents. In contrast to [Wither et al. (2011)], we remove dynamic objects from omnidirectional images because it is difficult to capture omnidirectional images without any dynamic objects under cloudy conditions, in practice. First, the dynamic objects are roughly removed by computing the temporal median of the input video. The duration of the video is several tens of seconds and it is captured at a fixed position [Arai et al. (2010)]; then, the remaining dynamic objects, whose regions are manually determined, are removed by an image inpainting technique [Kawai et al. (2009)]. Figure 6 shows the result of dynamic object removal.

Next, virtual objects are superimposed onto each omnidirectional image from which dynamic objects have been removed. Image-based lighting (IBL) [Debevec (1998)], which employs omnidirectional images for illuminating virtual objects, is adopted for photorealistic superimposition. Note that manual editing (e.g., arrangement of additional light sources and designation of occlusion masks) is applicable in the offline process. If necessary, we can achieve *cinema-quality* synthesis of real and virtual objects, which is difficult to achieve in traditional AR applications, in principle. Figure 7 shows an example of an augmented omnidirectional image, where a virtual model of a lost building, whose shape and color are different from those of the existing one in Figure 6, is rendered by IBL with manually obtained occlusion masks.

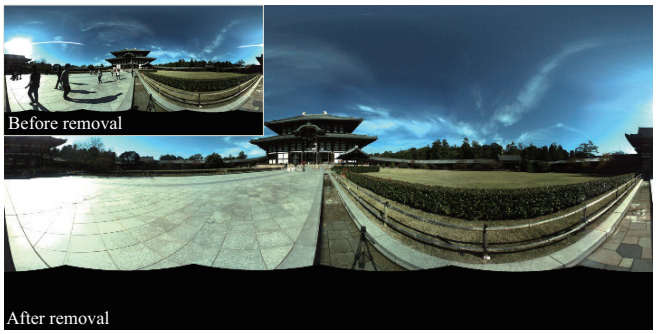


Fig. 6 Dynamic object removal. The inset shows a sampled frame from an omnidirectional image sequence with dynamic objects.

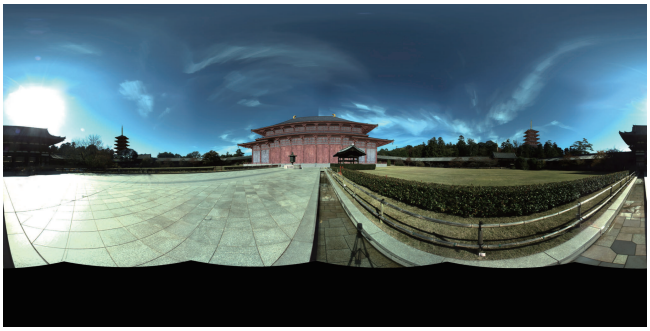


Fig. 7 Augmented omnidirectional image.

5.1.2 Sub-Image Histogram Calculation

Histograms of sub-images (i.e., block regions in the image) are used for efficient and robust online similarity calculation between the pre-captured images and the real scenes captured online. Although vision-based pixel-wise matching methods for mobile and omnidirectional images [Langlotz et al. (2011)] are available, in most IAR implementations, including ours, the orientation of a mobile device is acquired by embedded sensors to achieve light computation. Because sensor-based orientation does not achieve pixel accuracy, pixel-wise similarity measures (e.g., sum of squared differences (SSD) of pixel values) are not suitable for our similarity calculation. Histogram-based similarity is expected to be robust against misalignment; moreover, it is efficient because it is not necessary to store the entire pixel data of the pre-captured images in the database on a mobile device.

First, pre-captured omnidirectional images with no virtual objects or dynamic objects are divided into sub-image blocks. We use sub-images of 100×100 pixels for omnidirectional images consisting of 5400×2700 pixels. Then sub-image histograms of 16 bins are calculated for each color channel (R, G, and B) and are simply stored in the database. Note that missing areas, which are placed in the bottom portion of the omnidirectional images, are ignored for generating histograms.

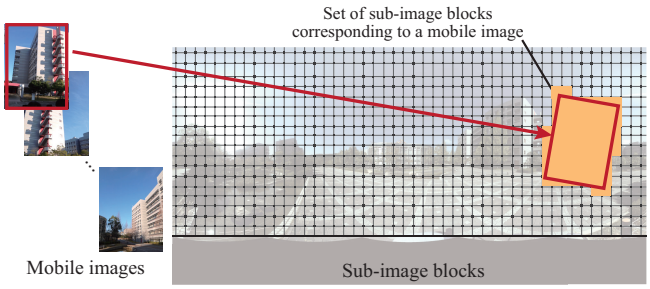


Fig. 8 Orientation of mobile device and corresponding sub-image regions.

5.2 Online Process

In the online process, a pre-captured omnidirectional image, whose illumination condition is considered to be the most similar to that of the real world, is selected according to the differences between the sub-image histograms of each pre-captured image and real images captured by the mobile device (which are referred to as mobile images). The augmented version of the selected image is displayed on the mobile device in the form of a planar perspective image depending on the orientation of the device.

5.2.1 Pre-Captured Image Selection

We compare histograms to find an omnidirectional image similar to the mobile images. For this purpose, a naive approach is to use existing metrics such as Euclidean, χ-square, or Hellinger distances. However, because we use auto-exposure and auto-white-balance to allow mobile devices to handle large variations in real-world illumination, the differences between the luminance and white balance of two images can be a practical issue. Therefore, we introduce a scale factor that should be optimized. Although we employ a metric based on the Euclidean distance, other metrics can be adapted to our purpose similarly. In addition, because users can freely change their view direction during the IAR experience, we unify the histogram differences for various directions into a single difference value for the entire scene, assuming that the real-world illumination does not vary during an IAR experience.

First, a histogram, H_θ^{mob}, of a mobile image is calculated, where θ denotes the three degrees of freedom pose (roll, pitch, yaw) of the device. A set of sub-image regions corresponding to the mobile image is determined using the orientation and pre-defined field-of-view of the device, as shown in Figure 8. Let i denote the ID of a pre-captured image in the database. $H_{i,\theta}^{db}$, a histogram of the set of sub-image regions in the i-th pre-captured image, is calculated as the summation of sub-image histograms belonging to the region.

The difference for a single channel of images, $D_c(i)$, between the i-th pre-captured image and a set of mobile images captured in various orientations is calculated as follows:

$$D_c(i) = \min_s \sum_\theta diff_c(i, \theta, s), \tag{1}$$

where s denotes a scale factor that compensates the difference between the luminance and white balance of different cameras. The white balance of a mobile camera varies dynamically; thus, we treat s as an unknown value required to be estimated. Assuming that the gammas of all images are linearized, the optimal s is searched so as to minimize $\sum_\theta \mathit{diff}_c(i,\theta,s)$, with tiny increments of s in a given domain (in our experiments, s was searched in a domain $0.5 < s < 1.5$ with increments of 0.1). By letting $H(j)$ denote the j-th bin in the histogram H, $\mathit{diff}_c(i,\theta,s)$ is calculated according to the squared difference between the histograms as follows:

$$\mathit{diff}_c(i,\theta,s) = \frac{1}{m(\theta,s)} \sum_{j=0}^{m(\theta,s)} \left(\frac{H_\theta^{mob}(j)}{n^{mob}} - \frac{H_{i,\theta}^{db}(sj)}{n^{db}(i,\theta)} \right)^2, \tag{2}$$

where n^{mob} and $n^{db}(i,\theta)$ denote the number of pixels in the mobile image and the corresponding region of the pre-captured image, respectively; they are used for normalizing the difference in the number of pixels. Further, m is defined as the number of bins used for the difference calculation:

$$m(\theta,s) = \begin{cases} bin(H_\theta^{mob}) & (s \geq 1) \\ s * bin(H_\theta^{mob}) & (s < 1) \end{cases}, \tag{3}$$

where $bin(H_\theta^{mob})$ denotes the number of unsaturated bins in H_θ^{mob}. Saturated pixels are not used for the calculation, and the difference in m varying in accordance with s is normalized in Equation (2).

The differences $D_c(i)$, which are calculated for each color channel independently, are finally combined as the total difference $D(i)$ for the i-th pre-captured image.

$$D(i) = \sum_c D_c(i). \tag{4}$$

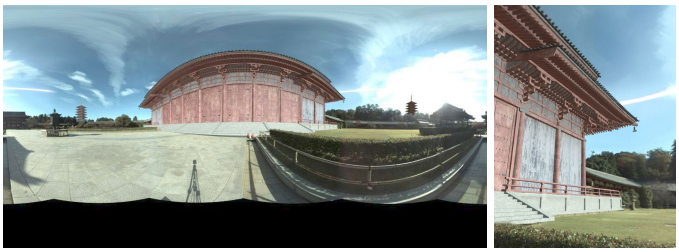
Here, because the optimal scale factor s is calculated independently for each color channel, it compensates for the white balance among multiple channels. The optimal image, which is presented to a user, is selected so as to minimize $D(i)$.

5.2.2 View-Direction-Dependent Image Presentation

The augmented version of the selected omnidirectional image is transformed into a view-direction-dependent planar perspective image according to the orientation of the mobile device, which is acquired by embedded sensors. Although sensor-based pose estimation does often not achieve the same quality as that of geometric registration in traditional AR, most IAR users do not perceive a loss of naturalness, even if the estimated orientations contain errors of a few tens of degrees [Wither et al. (2011)].

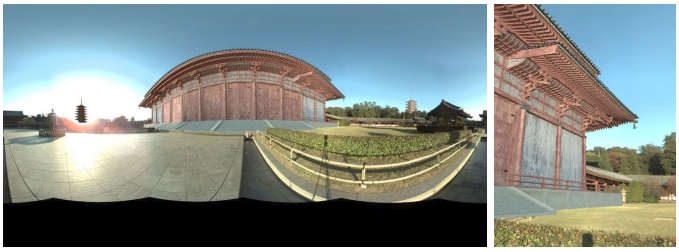
6 Public Experiment Using Cultural Heritage Application

To confirm the effect of illumination consideration on IAR, we developed a cultural heritage application set in the Todaiji temple, and we performed a public experiment with the cooperation of Todaiji, using omnidirectional images captured multiple times (see Figure 9). Note that although the experiment was performed

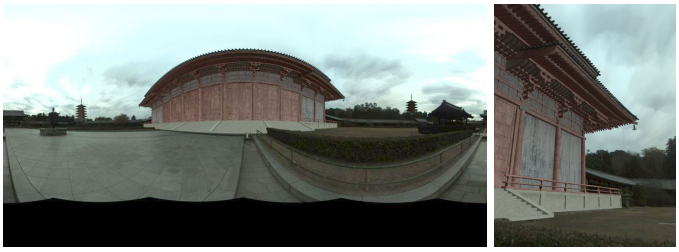


(a) Sunny, 9 am

(b) Sunny, 12 pm



(c) Sunny, 4 pm



(d) Cloudy, 12 pm

Fig. 9 Examples of augmented omnidirectional images under various illumination conditions as well as their perspective transformations for a given orientation.

at similar places to those described in Section 4, some experimental conditions were different. This section is a detailed version of the experiment section of a previous conference paper [Okura et al. (2014)].

6.1 Experimental Condition

Experimental Platform In the application, the appearances of three buildings at the time of the foundation of Todaiji (the Hall of the Great Buddha, the West Tower, and the East Tower)[3] were superimposed on pre-captured omnidirectional images, as shown in Figure 10. The virtual objects were superimposed onto nine pre-captured omnidirectional images with offline manual editing (adjusting light sources and designating occlusion masks). The hardware and software used for capturing and rendering were identical to those used in the experiment described in Section 4. We assumed that the gamma value of both a camera and a mobile device display is 2.2.

This application has three display modes for comparison:

1. Proposed system: Display an IAR scene with the smallest difference $D(i)$.
2. Ordinary IAR: Display a randomly selected IAR scene.
3. Worst case: Display an IAR scene with the largest difference $D(i)$.

The difference $D(i)$ was calculated using mobile images of the entire horizontal orientations that were captured approximately every 30° just before each trial. The IAR scenes were displayed on mobile devices (iPad2 and iPad 4th generation; Apple Inc.) in full-screen mode, except for small buttons for switching the display modes, as shown in Figure 11. The physical locations of the participants during the experiment was within 3 m of the position where the omnidirectional images were captured. The experiment was carried out during the day under sunny or cloudy conditions (depending on the participant).

Participants A total of 87 participants, whose ages are summarized in Table 2, experienced the application. All the participants were recruited from among the general public. Further, 56 participants were male, and the rest were female. Most participants were not familiar with AR/MR.

Task and Procedure Along with the preliminary experiment described in Section 4, groups of a few participants experienced the application, as shown in Figure 12. First, they watched the IAR scenes by switching among three modes (i.e., proposed, ordinary, and worst case), and then, they evaluated each scene by filling out a questionnaire. The ultimate goal of AR (including IAR) is to present virtual objects as if they actually exist in the real world. Thus, we asked the participants about the realism of the IAR scenes:

Q Do the virtual buildings look as if they actually exist in the real world?

Table 2 Age distribution of participants.

Age	10's	20's	30's	40's	50's	60's	>70
Number of participants	1	6	6	11	19	25	19

[3] These models were created with reference to the miniature model of Todaiji with architectural validation by Yumiko Fukuda, Hiroshima Institute of Technology, and patina expression technology validation by Takeaki Nakajima, Hiroshima City University.

(a) The Hall of the Great Buddha (b) The West Tower (c) The East Tower

Fig. 10 Scenes augmented by superimposing virtual buildings at the time of the foundation of Todaiji Temple.

Fig. 11 Application interface for the public experiment.

Fig. 12 Photograph of the public experiment. Left: participants experiencing IAR application. Right: participants filling out the questionnaire.

The participants selected a Likert scale value from 1 (absolutely no) to 7 (absolutely yes). The order in which the display modes were presented for the first time was randomized, and the participants were allowed to switch among them as many times as they wished, even while answering the question. We did not set up any processes for training participants prior to the evaluation. Free commentary was collected after scoring all the images.

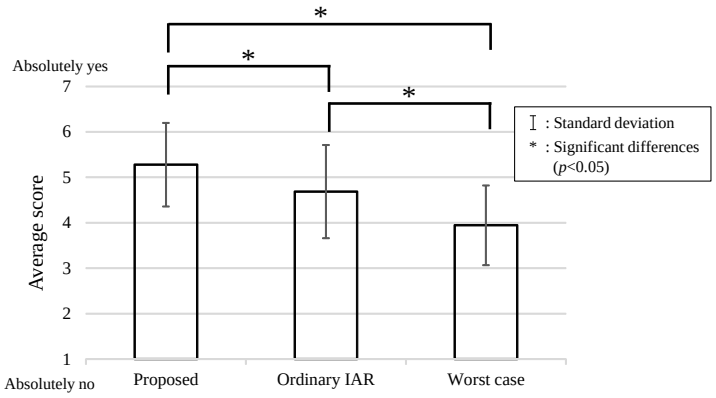


Fig. 13 Improvement in realism by considering real-world illumination changes. (*: significant difference $p < 0.05$ by a t-test with Bonferroni correction).

6.2 Result

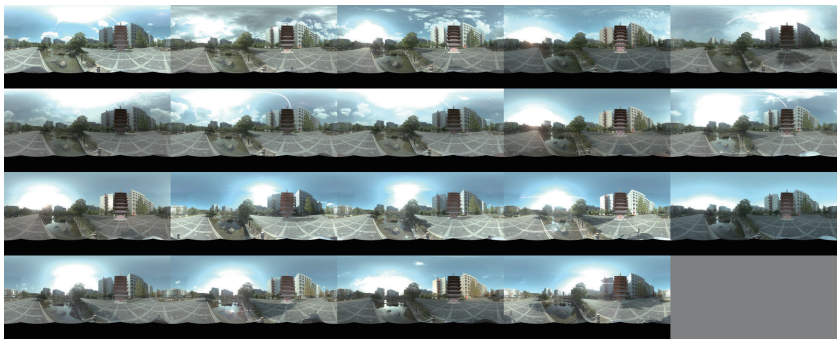
Figure 13 shows the average score for each display mode. The proposed system provided the best score in terms of realism of the IAR scenes. A one-way repeated measures ANOVA test along with a post-hoc paired t-test with Bonferroni correction ($p < 0.05$) show that the scores are significantly different. This result indicates that considering real-world illumination changes significantly improves the realism of IAR scenes.

7 Relationship between Realism and Illumination Difference

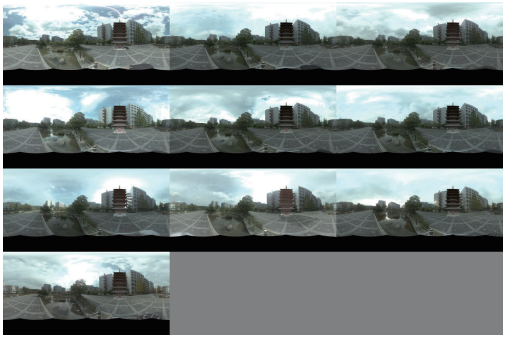
The results of the public experiment (Section 6) show that the consideration of real-world illumination indeed improves the IAR experience in practical situations. We investigated further details, i.e., the relationship between the realism of virtual objects and our illumination difference measure. To derive the relationship, we evaluated multiple IAR scenes captured at various times under different weather conditions instead of selecting an appropriate scene.

7.1 Experimental Condition

Experimental Platform The devices used were identical to those used in the experiment described in Section 6. Twenty-nine omnidirectional images were captured from August to December 2013 at various times of the day under different weather conditions at the same position in our university campus. The images were captured in environments without pedestrians; thus, the dynamic object removal process described in Section 5.1.1 was not adopted. The luminances of the pre-captured omnidirectional images were linearly transformed such that they had the same average luminance. For this experiment, we obtained answers to our questions through a digital interface. The mobile device used in this experiment



(a) Captured under sunny weather.



(b) Captured under cloudy weather.

Fig. 14 Augmented omnidirectional images captured at various times under different weather conditions.

displayed a small region for showing the questions and buttons for selecting the scores, in addition to an IAR scene. Figure 14 shows augmented omnidirectional images on which a virtual tower was superimposed.

Participants A total of 24 participants experienced the prototype system under five different time and weather conditions, as listed in Table 3. All the participants were in their twenties or thirties, and they were recruited from within the university; thus, they were familiar with computer science techniques.

Table 3 Time and weather conditions for the experiments

Date	Time	Weather	Number of participants
Dec. 25	3 pm	Sunny	5
Dec. 28	11 am	Sunny	5
Dec. 28	1 pm	Sunny	4
Feb. 21	2 pm	Cloudy	5
Feb. 26	5 pm	Cloudy	5

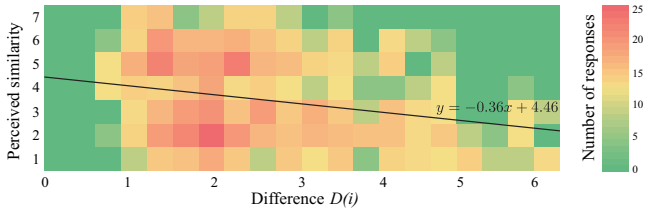


Fig. 15 Result of Q1: A heat-map showing the relationship between the perceived illumination similarity and our difference measure $D(i)$, as well as a regression line ($y = -0.36x + 4.46$). The score distribution shows a correlation of -0.350.

Task and Procedure The participants watched all the IAR scenes generated from 29 pre-captured omnidirectional images, which were presented to each participant in a random order. Then, $D(i)$ was calculated for each pre-captured image just before the experiment, followed by the public experiment. The participants answered the following questions by selecting a Likert scale value from 1 (absolutely no) to 7 (absolutely yes) on the digital interface:

Q1 How similar are the illumination conditions (e.g., weather, shadow, and time) of the real scene and the displayed scene?
Q2 Does the virtual tower look as if it actually exists in the real world?

The first question evaluates the perceived similarity between the real scene and the IAR scene. Through comparison with our difference measure $D(i)$, the score distribution for this question evaluates how well our difference measure describes the perceived illumination condition. Using the second question, we intend to investigate the relationship between the perceived realism and $D(i)$. We acquired the scores for all 29 images from each participant. The order in which the display modes were presented for the first time was randomized, and the participants were allowed to switch among them as many times as they wished, even while scoring each scene. Prior to the actual evaluation, we prepared a training set of five images randomly selected from the IAR scenes. Free commentary was collected after scoring all the images.

7.2 Result

We derived the relationship between the difference measure $D(i)$ calculated by our IAR system and the two human factors acquired through the questionnaires.

Figure 15 shows the relationship between the difference measure $D(i)$, which was computed by our system, and the perceived illumination similarity, which was acquired through the questionnaire (Q1). The score distribution demonstrates a negative correlation (Pearson's correlation coefficient -0.350; linear regression $y = -0.36x + 4.46$). This indicates that our difference measure implemented in the IAR system effectively captures the illumination change perceived by the participants, i.e., the illumination difference $D(i)$, which is automatically calculated, is approximately proportional to the perceptual illumination measurement.

Next, we evaluated the relationship between the perceived realism of the virtual objects and the difference measure $D(i)$. Figure 16 shows the score distribution for

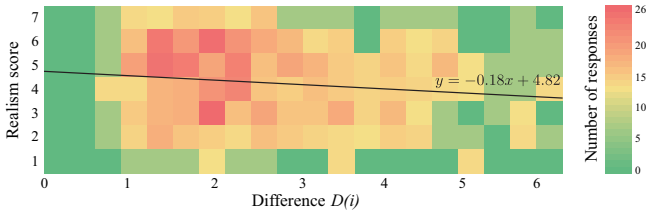


Fig. 16 Result of Q2: A heat-map showing the relationship between the realism score and the difference $D(i)$, as well as a regression line ($y = -0.18x + 4.82$). The score distribution shows a correlation of -0.203.

Q2, which evaluates the relationship between the realism score and the illumination difference $D(i)$. The score distribution shows a negative correlation (Pearson's correlation coefficient -0.203; linear regression $y = -0.18x + 4.82$). Although the correlation was weak owing to the presence of outliers, we confirmed that the realism of the IAR experience bears a direct relation to the similarity between a pre-captured image and the real scene. Section 7.3.2 describes further investigation of the outliers.

7.3 Discussions

7.3.1 Confirmation of the Result in Public Experiment

We performed an experiment to reproduce the results presented in Section 6. In the experiment described in Section 7, we acquired perceptual scores from 24 participants for all 29 IAR scenes, which were captured under various time/weather conditions. From the given score set, we recalculated the scores of the methods compared in Section 6: proposed IAR, ordinary IAR, and the worst case.

Toward this end, we virtually simulated the average scores for the three methods. For each participant, the scores for each method were calculated by selecting one from among all the scores for the 29 IAR scenes according to the strategies of each method:

1. Proposed system: Select the score of an IAR scene with the smallest difference $D(i)$ for each participant.
2. Ordinary IAR: Randomly select a score from among the scores given by each participant.
3. Worst case: Select the score of an IAR scene with the largest difference $D(i)$ for each participant.

Figure 17 shows the average score for each method. The result indeed shows a similar trend to the result of the public experiment, which is shown in Figure 13.

7.3.2 Aesthetics—Another Factor Affecting the IAR Experience

The results clearly show that considering real-world illumination improves the realism of IAR scenes. However, we found outliers who gave perceptual scores

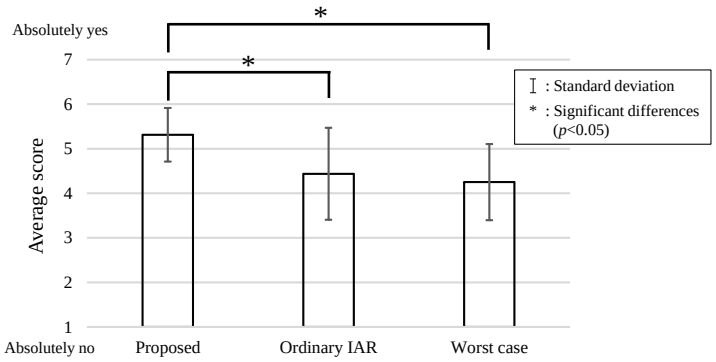


Fig. 17 Confirmation of the results of the public experiment via simulated scores.

that do not directly relate to the difference measure $D(i)$ used in our IAR system. Through further investigation of the outliers, we found significant individual differences among the score distributions, which provide interesting insights.

As a typical example of the individual differences, first, we compared the scores of two different participants (we refer to them as Participant 1 and Participant 2) who had experienced our application under the same conditions (5 pm, Feb. 26, cloudy). The scores given by Participant 1 for all 29 IAR scenes show a strong correlation (-0.737), whereas the score distribution of Participant 2 shows no correlation (0.052; significance level $p > 0.10$ for sample size 29). Free commentary from Participant 2 indicated that there was a hidden additional criterion underlying this distribution. The participant stated that the *aesthetics* of IAR scenes affected the scores in addition to the consistency between the IAR scenes and the real world; in other words, the participant favored sunny IAR scenes (e.g., Figure 18(a)) over cloudy ones (e.g., Figure 18(b)).

To confirm that weather had an impact on the scores of Participant 2, we manually categorized the 29 IAR scenes into two types, sunny (19 scenes) and cloudy (10 scenes), based on whether the sun appears (see Figure 14). Interestingly, although the experiment was performed in cloudy conditions, Participant 2 scored much higher on IAR scenes using sunny pre-captured images (average score, 4.95) as compared to those using cloudy ones (average score, 4.30). Note that the average scores of all the participants who experienced the IAR scenes in cloudy conditions indicated that the scores for cloudy IAR scenes (average score, 4.43) were significantly (t-test, $p < 0.05$) higher than those for sunny ones (average score, 3.76).

Does the weather affect the IAR experience in general? Although we did not observe notable differences among the scores of all the participants for sunny and cloudy IAR scenes, we found that real-world weather conditions affect the IAR experience. We classified the scores of all the participants according to the weather conditions during their experience. Out of 24 participants, 14 participants experienced our system in sunny weather, as indicated in Table 3. In sunny conditions, the average score for the IAR image with the smallest difference $D(i)$, i.e., the image selected by our IAR approach, was 5.64. On the other hand, the average score of the remaining 10 participants who experienced our system in cloudy weather

(a) Sunny. (b) Cloudy.

Fig. 18 Examples of sunny and cloudy IAR scenes. For visualization, the field of view is set larger than that in the actual application.

was 4.50. The scores under sunny conditions were significantly (t-test, $p < 0.05$) higher than those under cloudy conditions, although a small sample size was considered. This result implies that even when we present the users with a proper scene according to the real-world illumination, it is better to experience IAR applications under aesthetic real-world conditions. However, it is difficult to control the real-world illumination in practice.

In addition to the consistency between the real world and the pre-captured images, aesthetics can be an important factor that affects the realism of IAR scenes. This factor can impart a significant advantage to IAR applications when it is investigated in further detail; e.g., depending on the application and/or the user, presenting aesthetic IAR scenes under *aesthetic* real-world conditions improves the IAR experience. We do not believe that aesthetics explains the reasons for outliers in our experiment in all cases. Therefore, an interesting direction for future work is to investigate human factors that affect the IAR experience. For example, the *uncanny valley* problem [Mori et al. (2012)] might be an important factor in decreasing the realism in IAR by using pre-captured images that are similar but not exactly identical to the real-world appearance. Furthermore, we are interested in investigating differences due to user attributes, e.g., age and gender. Based on recent studies that have enhanced the aesthetics of AR scenes [Gruber et al. (2010)] and photographs [Aydin et al. (2015)], we are confident that IAR, as well as AR, can further leverage such human factors in the future.

8 Conclusions and Future Work

In this study, we investigated inconsistencies between pre-captured images and the real world in IAR. We found that temporal inconsistency has a greater impact on user experience than spatial inconsistency. Therefore, we proposed an IAR

approach that addresses a major category of temporal inconsistency by selecting and displaying an IAR scene similar to the real world from among scenes captured at various times of the day under different weather conditions. The results of our experiments showed that our prototype system improves the realism of IAR scenes. Moreover, we identified an additional factor that affects the IAR experience, i.e., aesthetics.

In the future, we plan to develop more efficient methods for preparing the pre-captured images by employing relighting techniques (e.g., Laffont et al. [Laffont et al. (2013)]); thus, it will not be necessary to capture the images multiple times under various illumination conditions. Another direction for future work is to address the spatial inconsistency in IAR by generating a user's viewpoint that is different from the location where the omnidirectional images were captured. The recent success of image-based rendering using multi-view images (e.g., Chaurasia et al. [Chaurasia et al. (2013)]) is expected to inspire photorealistic free-viewpoint navigation for IAR.

Finally, our ultimate goal is to make IAR a technique that is widely adopted as a form of super-light, ultra-stable, and super-photorealistic AR. We firmly believe that our study marks the first step toward achieving this goal, and thus, it should have a significant impact on the AR field.

Acknowledgements We first thank the anonymous reviewers for their constructive comments and suggestions. We thank members of NAIST-Keio joint research group. We also thank Todaiji for giving us the opportunity to conduct public experiments.

References

[Akaguma et al. (2013)] Akaguma T, Okura F, Sato T, Yokoya N (2013) Mobile AR using pre-captured omnidirectional images. In: Proc. ACM SIGGRAPH Asia'13 Symp. on Mobile Graphics and Interactive Applications, pp 26:1–26:4

[Anguelov et al. (2010)] Anguelov D, Dulong C, Filip D, Frueh C, Lafon S, Lyon R, Ogale A, Vincent L, Weaver J (2010) Google street view: Capturing the world at street level. IEEE Computer Magazine 43(6):32–38

[Arai et al. (2010)] Arai I, Hori M, Kawai N, Abe Y, Ichikawa M, Satonaka Y, Nitta T, Nitta T, Fujii H, Mukai M, Hiromi S, Makita K, Kanbara M, Nishio N, Yokoya N (2010) Pano UMECHIKA: A crowded underground city panoramic view system. In: Proc. Int'l Symp. on Distributed Computing and Artificial Intelligence (DCAI'10), pp 173–180

[Aydin et al. (2015)] Aydin T, Smolic A, Gross M (2015) Automated aesthetic analysis of photographic images. IEEE Trans. on Visualization and Computer Graphics 21(1):31–42

[Azuma (1997)] Azuma R (1997) A survey of augmented reality. Presence: Teleoperators and Virtual Environments 6(4):355–385

[Chaurasia et al. (2013)] Chaurasia G, Duchene S, Sorkine-Hornung O, Drettakis G (2013) Depth synthesis and local warps for plausible image-based navigation. ACM Trans. on Graphics 32(3):30:1–30:12

[Chen (1995)] Chen SE (1995) Quicktime VR: An image-based approach to virtual environment navigation. In: Proc. ACM SIGGRAPH'95, pp 29–38

[Côté et al. (2013)] Côté S, Trudel P, Desbiens M, Giguère M, Snyder R (2013) Live mobile panoramic high accuracy augmented reality for engineering and construction. In: Proc. 13th Int'l Conf. on Construction Applications of Virtual Reality (CONVR'13), pp 262–271

[Debevec (1998)] Debevec P (1998) Rendering synthetic objects into real scenes: Bridging traditional and image-based graphics with global illumination and high dynamic range photography. In: Proc. ACM SIGGRAPH'98, pp 189–198

[Grosch (2005)] Grosch T (2005) PanoAR: Interactive augmentation of omnidirectional images with consistent lighting. In: Proc. Computer Vision/Computer Graphics Collaboration Techniques and Applications (Mirage'05), pp 25–34

[Gruber et al. (2010)] Gruber L, Kalkofen D, Schmalstieg D (2010) Color harmonization for augmented reality. In: Proc. 9th IEEE Int'l Symp. on Mixed and Augmented Reality (ISMAR'10), pp 227–228

[Gruber et al. (2012)] Gruber L, Richter-Trummer T, Schmalstieg D (2012) Real-time photometric registration from arbitrary geometry. In: Proc. 11th IEEE Int'l Symp. on Mixed and Augmented Reality (ISMAR'12), pp 119–128

[Hollerer et al. (1999)] Hollerer T, Feiner S, Pavlik J (1999) Situated documentaries: Embedding multimedia presentations in the real world. In: Proc. 3rd IEEE Int'l Symp. on Wearable Computers (ISWC'99), pp 79–86

[Kán and Kaufmann (2012)] Kán P, Kaufmann H (2012) High-quality reflections, refractions, and caustics in augmented reality and their contribution to visual coherence. In: Proc. 11th IEEE Int'l Symp. on Mixed and Augmented Reality (ISMAR'12), pp 99–108

[Kanbara and Yokoya (2002)] Kanbara M, Yokoya N (2002) Geometric and photometric registration for real-time augmented reality. In: Proc. First Int'l Symp. on Mixed and Augmented Reality (ISMAR'02), pp 279–280

[Kawai et al. (2009)] Kawai N, Sato T, Yokoya N (2009) Image inpainting considering brightness change and spatial locality of textures and its evaluation. In: Proc. Third Pacific-Rim Symp. on Image and Video Technology (PSIVT'09), pp 271–282

[Klein and Murray (2009)] Klein G, Murray D (2009) Parallel tracking and mapping on a camera phone. In: Proc. 8th IEEE Int'l Symp. on Mixed and Augmented Reality (ISMAR'09), pp 83–86

[Laffont et al. (2013)] Laffont PY, Bousseau A, Drettakis G (2013) Rich intrinsic image decomposition of outdoor scenes from multiple views. IEEE Trans. on Visualization and Computer Graphics 19(2):210–224

[Langlotz et al. (2011)] Langlotz T, Degendorfer C, Mulloni A, Schall G, Reitmayr G, Schmalstieg D (2011) Robust detection and tracking of annotations for outdoor augmented reality browsing. Computers & Graphics 35(4):831–840

[Lensing and Broll (2012)] Lensing P, Broll W (2012) Instant indirect illumination for dynamic mixed reality scenes. In: Proc. 11th IEEE Int'l Symp. on Mixed and Augmented Reality (ISMAR'12), pp 109–118

[Liestol and Morrison (2013)] Liestol G, Morrison A (2013) Views, alignment and incongruity in indirect augmented reality. In: Proc. 12th IEEE Int'l Symp. on Mixed and Augmented Reality (ISMAR'13)—Arts, Media, and Humanities, pp 23–28

[Madsen and Stenholt (2014)] Madsen JB, Stenholt R (2014) How wrong can you be: Perception of static orientation errors in mixed reality. In: Proc. IEEE Symp. on 3D User Interfaces (3DUI'14), pp 83–90

[Mori et al. (2012)] Mori M, MacDorman K, Kageki N (2012) The uncanny valley [from the field]. IEEE Robotics & Automation Magazine 19(2): 98–100

[Okura et al. (2014)] Okura F, Akaguma T, Sato T, Yokoya N (2014) Indirect augmented reality considering real-world illumination change. In: Proc. IEEE Int'l Symp. on Mixed and Augmented Reality (ISMAR'14), pp 287–288

[Okura et al. (2015)] Okura F, Kanbara M, Yokoya N (2015) Mixed-reality world exploration using image-based rendering. ACM Journal on Computing and Cultural Heritage 8(2):9:1–9:26

[Schops et al. (2014)] Schops T, Enge J, Cremers D (2014) Semi-dense visual odometry for AR on a smartphone. In: Proc. 2014 IEEE Int'l Symp. on Mixed and Augmented Reality (ISMAR'14), pp 145–150

[Tenmoku et al. (2003)] Tenmoku R, Kanbara M, Yokoya N (2003) A wearable augmented reality system using positioning infrastructures and a pedometer. In: Proc. 16th IEEE Int'l Symp. on Wearable Computers (ISWC'03), pp 110–117

[Uyttendaele et al. (2004)] Uyttendaele M, Criminisi A, Kang SB, Winder S, Szeliski R, Hartley R (2004) Image-based interactive exploration of real-world environments. IEEE Computer Graphics and Applications 24(3):52–63

[Ventura and Hollerer (2012)] Ventura J, Hollerer T (2012) Wide-area scene mapping for mobile visual tracking. In: Proc. 11th IEEE Int'l Symp. on Mixed and Augmented Reality (ISMAR'12), pp 3–12

[Waegel (2014)] Waegel K (2014) A reconstructive see-through display. In: Proc. IEEE Int'l Symp. on Mixed and Augmented Reality (ISMAR'14), pp 379–320

[Wither et al. (2011)] Wither J, Tsai YT, Azuma R (2011) Indirect augmented reality. Computers & Graphics 35(4):810–822

[Yamamoto et al. (2014)] Yamamoto G, Lübke AIW, Taketomi T, Kato H (2014) A see-through vision with handheld augmented reality for sightseeing. In: Proc. 16th Int'l Conf. on Human-Computer Interaction (HCI Interational'14), pp 392–399

[Zhou et al. (2008)] Zhou F, Duh HBL, Billinghurst M (2008) Trends in augmented reality tracking, interaction and display: A review of ten years of ISMAR. In: Proc. 7th IEEE/ACM Int'l Symp. on Mixed and Augmented Reality (ISMAR'08), pp 193–202

[Zoellner et al. (2009)] Zoellner M, Keil J, Drevensek T, Wuest H (2009) Cultural heritage layers: Integrating historic media in augmented reality. In: Proc. 15th Int'l Conf. on Virtual Systems and Multimedia (VSMM'09), pp 193–196