Position-Dependent Indirect Augmented Reality Using Multi-Point Omnidirectional Images

Yuta Hamada[1], Norihiko Kawai[1,*]

[1]Graduate School of Information Science, Osaka Institute of Technology, Kitayama 1-79-1, Hirakata, Osaka, Japan

Abstract

Indirect Augmented Reality (IAR), which uses pre-captured omnidirectional images, is less affected by occluders and can provide stable visual presentation compared to AR based on real-time camera images. However, early IAR systems assume that a user experiences AR at fixed viewpoints, and therefore a discrepancy occurs between the presented imagery and the real scene when the user moves. To address this issue, IAR systems that allow movement in 3D space using techniques such as image-based rendering have been developed. Nevertheless, challenges remain, including restrictions on the movable range and high rendering costs. In this study, we propose a method that achieves position-dependent IAR by using a set of omnidirectional images captured at multiple points along a route. For AR content designed to be experienced along a predetermined route, the proposed method provides position-dependent IAR presentation by estimating the device position from real-time camera frames and presenting the omnidirectional image captured at the nearest position. Experimental results in two indoor scenes demonstrated that the proposed method can provide position-dependent IAR presentation along both straight and turning routes.

Keywords

Augmented Reality, Indirect Augmented Reality, Omnidirectional Images, PnP

1. Introduction

Applications based on augmented reality (AR) using real-time camera images have been widely adopted in tourism and educational fields. However, when surrounding tourists or other occluding objects appear in the camera image, camera pose estimation may fail, causing virtual objects to be displayed at positions different from their intended positions. In addition, occluders may conceal parts of the real scene associated with virtual objects, thereby reducing visibility. Although studies on diminished reality, which presents images from which unwanted objects have been removed from real-time video streams, have recently been conducted [1, 2, 3, 4], techniques capable of reliably removing moving objects with high quality in environments where the camera moves freely have not yet been established. Therefore, AR systems based on real-time camera images still face limitations in terms of robustness and visibility.

To address these issues, Indirect Augmented Reality (IAR) [5] has been proposed. Early IAR systems synthesize virtual objects into omnidirectional images captured in advance at fixed viewpoints. During the experience, the system extracts and presents an appropriate view from the omnidirectional image based on the device's orientation obtained from the electronic compass, gyroscope, and accelerometer embedded in the user's device. This approach enables users to experience robust AR that is unaffected by surrounding occluders, unlike AR based on real-time camera images. Due to its stability, IAR has been applied as a presentation method in crowded indoor exhibition environments [6] and has also been put into practical use as a cultural heritage presentation system for archaeological tourism [7]. However, this method assumes that users experience AR from a fixed point of view. As a result, the scene displayed on the device does not change even when the user moves away from the fixed position, leading to a discrepancy between the presented scene and the real scene.

APMAR'26: The 18th Asia-Pacific Workshop on Mixed and Augmented Reality, Aug. 3-5, 2026, Taipei, Taiwan
*Corresponding author.
norihiko.kawai@oit.ac.jp (N. Kawai)
0000-0002-7859-8407 (N. Kawai)

To overcome this limitation, Okura et al. [8] proposed a movable IAR system that reproduces motion parallax by estimating the camera pose through optical flow. In this method, views corresponding to the user's viewpoint are generated by image-based rendering (IBR), in which textures are mapped onto geometry reconstructed using Structure from Motion (SfM) [9] and multi-view stereo. However, IBR based on texture mapping does not provide sufficiently high-quality imagery. In contrast, Ueda et al. [10] proposed a more immersive movable IAR system by combining a 3D Gaussian Splatting-based scene representation with SLAM, achieving high-quality scene reconstruction and camera pose estimation. However, the movable area is limited due to the large data size of 3D Gaussian Splatting and the computational cost of rendering.

In this study, we propose a position-dependent IAR system for environments such as museums and entertainment facilities, where the route of user movement is predetermined. Unlike approaches based on IBR or 3D Gaussian Splatting, the proposed method does not require highly accurate reconstruction of a large-scale space. Instead, it utilizes a set of omnidirectional images captured at multiple locations along a route, thereby achieving an IAR system that can operate on mobile devices while reducing data size and rendering costs. During the experience, the user's position is estimated by establishing correspondences between the device camera frame and a preconstructed 3D point cloud. The omnidirectional image captured at the location closest to the estimated position is then automatically selected for AR presentation.

2. Proposed Method

2.1. Overview

The proposed method consists of offline and online processes. During the offline process, multiple omnidirectional images are captured along the route where the AR experience will take place. The camera poses corresponding to the captured images are estimated, and a 3D point cloud of the environment is reconstructed. Correspondences between image feature points and their 3D coordinates are also retained. Subsequently, a virtual environment is constructed by projecting each omnidirectional image onto a spherical shell that reflects the corresponding camera position and orientation, and virtual objects are placed within the environment.

During the online process, the position of the device camera is periodically estimated by establishing correspondences between feature points extracted from the user's device camera frame and the preconstructed 3D point cloud. The system selects the spherical shell corresponding to the omnidirectional image captured at the location closest to the estimated position and sets a virtual camera at the center of the spherical shell. By applying orientation information obtained from the device's internal sensors to the virtual camera, an appropriate field of view from the virtual camera is presented to the user. The details of the offline and online processes are described in the following sections.

2.2. Offline Process

Multiple omnidirectional images are captured at several positions along the route where the AR experience takes place. Offline processing is then performed on the captured images. The details are described below.

2.2.1. Camera Pose Estimation for Pre-captured Images

Camera poses are estimated for the captured omnidirectional images. In this study, the omnidirectional images are first converted into perspective images, and Structure from Motion (SfM) is applied to reconstruct camera poses and a 3D point cloud from feature correspondences across images. In addition, correspondences between the 3D coordinates and SIFT features [11] in each frame are obtained and retained for device position estimation during the online process.

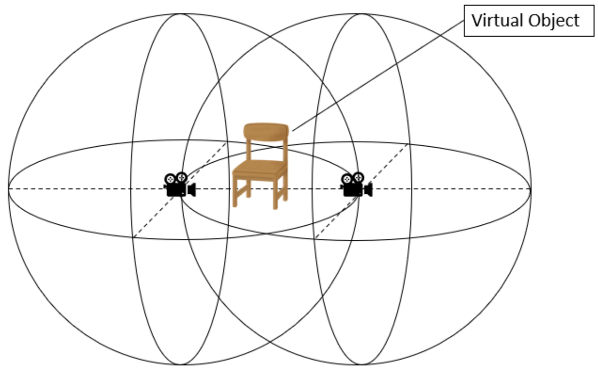

Figure 1: Spatial relationship among virtual objects, virtual cameras, and spherical shells.

2.2.2. Virtual Environment Construction and Virtual Object Placement

After camera pose estimation, spherical shells are placed in a virtual environment, and each omnidirectional image is projected onto the inner surface of the shell. Virtual objects are then placed at arbitrary positions in the environment. In the online process described later, by rendering the scene from a virtual camera located in the center of the spherical shell, users can observe imagery in which virtual objects are integrated into the omnidirectional image.

For AR object authoring, a textured mesh model is generated from the estimated camera poses and the 3D point cloud. Using this mesh model as a reference, the positions and scales of virtual objects are adjusted appropriately. The mesh model is used only as a reference for object placement and is not presented during the experience.

To ensure that virtual objects appear at fixed positions in the real environment, the camera pose parameters estimated by SfM are applied to the virtual cameras and spherical shells as shown in Fig. 1. The position component is applied to the virtual camera, while both the position and orientation components are applied to the spherical shell. It should be noted that the inverse of the estimated rotation is applied to the spherical shell orientation to eliminate the effect of camera rotation during image capture.

2.3. Online Process

During the online process, SIFT features are extracted from the device's real-time camera image, and the device position is estimated based on correspondences with the preconstructed 3D point cloud.

To establish correspondences, a set of perspective images associated with the initial position of the route is used for the first frame. For each subsequent frame, the system selects the set of perspective images associated with the capture position closest to the device position estimated in the previous frame. From this set, the perspective image whose viewing direction is closest to that of the device camera is selected based on orientation information obtained from the device's inertial sensor. The feature descriptors extracted from the selected image and those extracted from the real-time camera image are then matched using brute-force KNN matching ($k = 2$). Incorrect matches are removed using the ratio test, and 2D–3D correspondences between the feature points in the current frame and 3D points associated with the matched feature points in the selected image are established. The device position is estimated by solving a Perspective-n-Point (PnP) problem based on the obtained 2D–3D correspondences.

When position estimation repeatedly fails because the device camera is directed toward textureless regions such as the floor or ceiling, or is occluded by surrounding people, additional perspective images corresponding to positions ahead of the current position are also used for matching. This prevents estimation failure caused by an insufficient number of correspondences and facilitates recovery of position estimation even when the discrepancy between the device camera view and the pre-captured images temporarily becomes large.

Table 1
Specifications of omnidirectional camera.

Image Resolution	6528×3264
ISO Sensitivity	100–3200
Shutter Speed	1/8000–120 s

Table 2
Specifications of tablet device.

OS	Android 11
CPU	Qualcomm Snapdragon 870
Memory	8 GB
Storage	256 GB
Rear Camera	Approx. 13 MP + 5 MP

Finally, the system selects the spherical shell corresponding to the omnidirectional image that was captured at the position closest to the estimated position. The system places a virtual camera at the center of the selected shell and determines its orientation using the device's inertial sensors. By presenting the image captured by this virtual camera, the proposed method achieves IAR presentation corresponding to the user's position along the route.

3. Experiments and Discussion

3.1. Experimental Setup

To evaluate the position estimation accuracy and behavior of the proposed method, experiments were conducted in two indoor environments at the Hirakata Campus of Osaka Institute of Technology: Scene A, a straight corridor, and Scene B, an entrance hall containing a turn along the route.

During the offline process, omnidirectional images were captured at multiple positions along the experimental routes using an Insta360 ONE RS 1-Inch 360 Edition. Camera pose estimation and 3D point cloud reconstruction were performed using COLMAP (v4.02). A virtual environment was constructed in Unity (v6000.3.12f1), and the system was implemented as an Android application running on a LAVIE Tab T12. The specifications of the camera and tablet device are shown in Tables 1 and 2, respectively.

3.2. Settings for Position Estimation and 3D Reconstruction

Camera pose estimation and 3D point cloud reconstruction were performed using the SfM extension of COLMAP (v4.02) for omnidirectional images. This extension enables 3D reconstruction from omnidirectional images by dividing each image into multiple perspective images for feature extraction and jointly processing images originating from the same frame as a camera rig.

For each omnidirectional image, 16 virtual cameras were uniformly arranged around the horizontal circumference to generate perspective images. Adjacent perspective images were configured to have approximately 50% overlap in their fields of view, facilitating the observation of common feature points and improving reconstruction stability. Using this configuration, the camera trajectory and 3D point cloud required for IAR construction were generated in advance.

To facilitate virtual object placement, the camera poses and image set estimated by COLMAP were imported into RealityScan [12] to generate a textured mesh model of the environment.

3.3. Scene A: Experiment

3.3.1. Experimental Conditions

We conducted this experiment to verify the basic operation of the proposed method. The experimental environment was a straight corridor containing continuous structural features such as walls and posted

Figure 2: Examples of omnidirectional images used in Scene A.

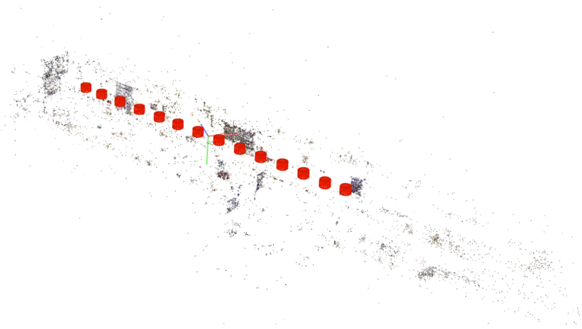

Figure 3: Pose estimation result for Scene A.

notices. We used 14 omnidirectional images captured along the route for the offline process. Several examples of captured omnidirectional images are shown in Fig. 2. The camera pose estimation and 3D point cloud reconstruction results obtained using COLMAP are shown in Fig. 3, and the mesh model generated using RealityScan is shown in Fig. 4. From Fig. 3, we can confirm that the camera poses and environmental structures, including notices posted on the corridor walls, were successfully reconstructed. During the experiment, the user walked along the corridor while holding a tablet device and confirmed the IAR presentation by stopping at arbitrary positions.

3.3.2. Experimental Results

The user experience during the experiment is shown in Fig. 5. As shown in Fig. 5(a), the viewpoint was updated according to the movement of the device, and imagery corresponding to the real environment was presented. Furthermore, Fig. 5(b) shows that the appearance of the virtual object changed according to the user's movement, indicating that viewpoint changes based on the estimated position were

Figure 4: Mesh reconstruction result for Scene A.

(a) Before movement.

(b) After moving to a different position.

Figure 5: User experience in Scene A.

appropriately reflected.

In addition, we confirmed that the omnidirectional image corresponding to the user's position was updated in approximately one second when the user stopped to view the IAR content. Although this delay was observed, it was not large enough to cause any noticeable discomfort or perceptual mismatch during the experience. These results suggest that position estimation operated with sufficient stability in Scene A.

3.4. Scene B: Experiment

3.4.1. Experimental Conditions

We conducted this experiment to evaluate the stability of the proposed method in an environment containing changes in the direction of movement. The experimental route included both straight sections and a turning section and contained environmental features such as furniture and posted notices in an entrance hall. We used 32 omnidirectional images captured along the route for the offline process. Several examples of captured omnidirectional images are shown in Fig. 6. The camera pose estimation and 3D point cloud reconstruction results obtained using COLMAP are shown in Fig. 7, and the mesh model generated using RealityScan is shown in Fig. 8. From Fig. 7, we can confirm that both the shape of the entrance hall, including the turning section, and the camera trajectory were

Figure 6: Examples of omnidirectional images used in Scene B.

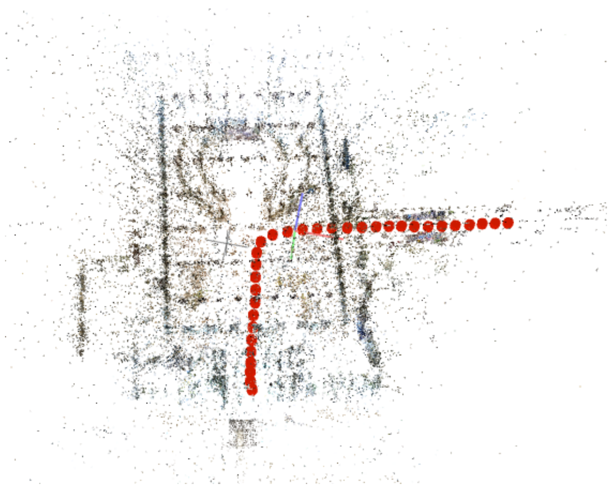

Figure 7: Pose estimation result for Scene B.

successfully reconstructed. During the experiment, the user walked along the route while holding a tablet device and confirmed the IAR presentation by stopping at arbitrary positions.

3.4.2. Experimental Results

The user experience during the experiment is shown in Fig. 9. As shown in Fig. 9, the viewpoint was updated according to the movement of the device even along a route containing turns. Additionally, the rotation of the spherical shell and the orientation of the virtual camera derived from the device's internal sensors were appropriately combined, enabling the system to present imagery corresponding to the user's viewing direction in the real environment. Furthermore, Fig. 9(b) shows that the appearance of

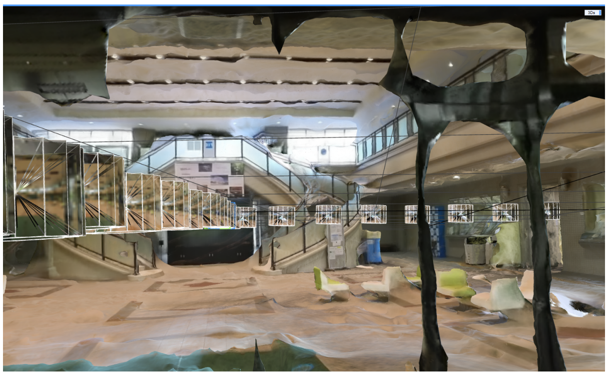

Figure 8: Mesh reconstruction result for Scene B.

(a) Before movement

(b) After moving to a different position

Figure 9: User experience in Scene B.

the virtual object differs from that in Fig. 9(a), indicating that viewpoint changes based on the estimated position were appropriately reflected.

Similar to Scene A, the omnidirectional image corresponding to the user's position was updated in approximately one second when the user stopped to view the IAR content, even in an environment containing turns. Position-dependent IAR presentation was achieved even in sections containing turns.

In some sections, position estimation temporarily became unstable because sufficient feature correspondences could not be obtained between the real-time camera image and the pre-captured images. However, when estimation failures occurred continuously, the recovery process in the proposed method expanded the matching targets to include perspective images corresponding to capture positions ahead of the current position. As a result, sufficient correspondences were re-established and position estimation recovered promptly. These results suggest that position estimation operated with sufficient stability even in an indoor environment containing turns.

4. Conclusion

In this study, we proposed a method that achieves IAR presentation consistent with the real environment in terms of the user's position, by using omnidirectional images captured at multiple positions along a route, without requiring large-scale space reconstruction or computationally expensive rendering.

The proposed method estimates the device position by establishing correspondences between feature points extracted from the device camera image and a preconstructed 3D point cloud. By presenting imagery aligned with the estimated user position through the omnidirectional image captured at the nearest location, the method enables users to experience IAR at arbitrary positions in environments with predetermined routes.

Experiments conducted in two kinds of indoor scenes demonstrated that position-dependent IAR presentation could be achieved not only along straight routes but also along routes containing turns. The omnidirectional image corresponding to the user's position was updated with a delay small enough to avoid noticeable discomfort during the experience. These results suggest that the proposed method can provide stable position-dependent IAR presentation in indoor environments with predetermined routes.

Future work includes extending the proposed method to support omnidirectional video in addition to omnidirectional images. This extension is expected to provide a richer and more continuous IAR experience by enabling smooth viewpoint transitions that follow user movement while maintaining geometric consistency between video frames. We also plan to improve the stability of position estimation and extend the proposed method to outdoor environments.

Acknowledgments

This work was supported by JSPS KAKENHI Grant Number JP25K15156.

Declaration on Generative AI

During the preparation of this work, the authors used ChatGPT and Copilot in order to: Grammar and spelling check, Paraphrase and reword. After using these tools, the authors reviewed and edited the content as needed and take full responsibility for the publication's content.

References

[1] N. Kawai, T. Sato, N. Yokoya, Diminished reality based on image inpainting considering background geometry, IEEE Transactions on Visualization and Computer Graphics 22 (2016) 1236–1247.
[2] N. Kawai, T. Sato, Y. Nakashima, N. Yokoya, Augmented reality marker hiding with texture deformation, IEEE Transactions on Visualization and Computer Graphics 23 (2017) 2288–2300.
[3] S. Mori, S. Ikeda, H. Saito, A survey of diminished reality: Techniques for visually concealing, eliminating, and seeing through real objects, IPSJ Transactions on Computer Vision and Applications 9 (2017) 1–18.
[4] M. Kari, T. Grosse-Puppendahl, L. F. Coelho, A. R. Fender, D. Bethge, R. Schütte, C. Holz, Transformr: Pose-aware object substitution for composing alternate mixed realities, in: Proceedings of IEEE International Symposium on Mixed and Augmented Reality (ISMAR), 2021.
[5] J. Wither, Y.-T. Tsai, R. Azuma, Indirect augmented reality, Computers & Graphics 35 (2011) 810–822.
[6] J. Gimeno, C. Portalés, I. Coma, M. Fernández, B. Martínez, Combining traditional and indirect augmented reality for indoor crowded environments, Computers & Graphics 69 (2017) 92–103.
[7] Y. Suganuma, M. Oda, K. Nakayama, S. Nishikawa, S. Hata, K. Paul, S. Wada, N. Kawai, Integrated system of augmented and virtual reality for ruins tourism, in: Proceedings of NICOGRAPH International, 2023, p. 85.
[8] F. Okura, Y. Nishizaki, T. Sato, N. Kawai, N. Yokoya, Motion parallax representation for indirect augmented reality, in: Proceedings of the IEEE International Symposium on Mixed and Augmented Reality (ISMAR-Adjunct), 2016, pp. 105–106.
[9] J. L. Schönberger, J. Frahm, Structure-from-motion revisited, in: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR), 2016, pp. 4104–4113.

[10] T. Ueda, Y. Suganuma, N. Kawai, Motion parallax reproduction in indirect augmented reality using 3D gaussian splatting for mobile devices, in: Proceedings of IEEE International Symposium on Mixed and Augmented Reality Adjunct (ISMAR-Adjunct), 2025, pp. 853–854.
[11] D. G. Lowe, Distinctive image features from scale-invariant keypoints, International Journal of Computer Vision 60 (2004) 91–110.
[12] Epic Games, RealityScan, https://www.realityscan.com, 2025. Accessed: 2026-05-27.