Online Adaptive Integration of Observation and Inpainting for Diminished Reality with Online Surface Reconstruction

Taiki Kato[1*] Naoya Isoyama[1†] Norihiko Kawai[2‡] Hideaki Uchiyama[1§] Nobuchika Sakata[3¶]

Kiyoshi Kiyokawa[1‖]

[1]Nara Institute of Science and Technology [2] Osaka Institute of Technology [3] Ryukoku University

Abstract

Diminished reality (DR) is one of the frameworks in augmented reality, which visually makes obstructive objects transparent and shows their background. The techniques in DR can be divided into observation-based and inpainting-based methods. The observation-based method completes the region of interest (ROI) with the color and shape observed as the camera moves, while the inpainting-based method completes the region by inferring it from the surrounding pixels. In this paper, we propose a method that adaptively integrates the advantages of both methods with online surface reconstruction. The inpainting-based method completes the ROI at the beginning, while the observation-based method gradually completes the region using the reconstruction with RGB-D SLAM as the camera moves. In this way, our proposed method adaptively utilizes both the inpainting-based and observation-based methods according to the camera movement online. The evaluation with various scenes was conducted to demonstrate the effectiveness of our proposed integration method by comparing it with each method.

Index Terms: Human-centered computing—Visualization—Visualization techniques—Treemaps; Human-centered computing—Visualization—Visualization design and evaluation methods

1 Introduction

Diminished Reality (DR) is a framework that hides or erases visually unnecessary objects or makes obstructive objects transparent when it is difficult to erase the objects that physically exist in the real world [3, 27]. DR is the new augmentation of Augmented Reality (AR), which visually adds computer-generated objects to the real world. In contrast to AR, DR is the opposite concept because it visually erases physical objects and shows their background in the images. As well as AR, the process in DR should be performed in real time, rather than being processed offline like visual effects in motion pictures. The ultimate goal of DR is to achieve high-quality visual removal in real time without making users aware of the removal of objects or the process. Owing to the recent advance in the practical use of AR in many fields, the usefulness of DR technology has also been presented [2, 5–9, 11–15, 17, 19, 21–23, 25, 26, 28, 30, 32–38].

An early framework for DR was similar to AR, such as overwriting obstructive objects existing in the real world with computer-generated objects to make them invisible [23]. In contrast, many technical issues must be solved when superimposing a real-world background to erase the objects. For instance, this framework involves a combination of technical challenges, including camera pose estimation, hidden background estimation, and region-of-interest (ROI) estimation through recognition and tracking of the object to be removed.

The recent techniques in DR can be divided into two major categories: observation-based and inpainting-based methods [27]. The observation-based methods use object's background images observed previously or simultaneously with multiple cameras to complete the ROI, thus allowing the physical background to be shown. This method is valid only when the background can be observed by moving the camera or obstacles. The inpainting-based method completes the image by inferring the texture from the surrounding pixels for the ROI, thus eliminating any preliminary preparation. The inpainting result cannot always be precise because it is based on inference.

The integration of both methods has been performed to efficiently and effectively complete the ROI [14, 20]. The method in [14] initially applies inpainting to the entire ROI. In subsequent frames, a part of the ROI is replaced with the actual background when the camera moves and the actual background is visible. However, manual cutting of the ROI by the user is required. The method in [20] automatically detects the ground plane region and completes it with the inpainting-based method, while other regions are completed with the observation-based method using the reconstructed 3D model. However, this method must reconstruct the background 3D model before performing DR.

In this paper, we propose a method that adaptively integrates the advantages of observation-based and inpainting-based methods in an online manner without the user's manual instruction for switching the results of both methods. At the beginning of the DR experience, the inpainting-based method completes the ROI because no observation is available. As the camera moves, the observation-based method adaptively completes the region using the online reconstruction with dense RGB-D SLAM. In other words, our method gradually switches the primary method from the inpainting-based to the observation-based methods online. From the technical point of view, our method is similar to [20] such that the observation-based method completes the ROI by using the observed object's background and the inpainting-based method completes the region that cannot be completed by the observation-based method. Specifically, we focus on the online adaptive integration of the two methods by running the DR processing and surface reconstruction simultaneously. In the evaluation, the effectiveness of our adaptive integration is demonstrated by comparing it with each method.

2 Related Work

2.1 Offline-Observation-based DR

Among the observation-based methods, the methods in which static background images are acquired in advance are referred to as offline-observation-based methods. The methods erase objects by transforming the previously captured background images into an appropriate

*e-mail: kato.taiki.kp7@is.naist.jp
†e-mail: isoyama@is.naist.jp
‡e-mail: norihiko.kawai@oit.ac.jp
§e-mail: hideaki.uchiyama@is.naist.jp
¶e-mail: sakata@rins.ryukoku.ac.jp
‖e-mail: kiyo@is.naist.jp

2771-1110/22/$31.00 ©2022 IEEE
DOI 10.1109/ISMAR-Adjunct57072.2022.00069

appearance according to the camera motion and merging it into the ROI. For example, Cosco et al. acquire background images and shapes in advance and present a background image with an appropriate appearance for each frame using image-based rendering [4]. In such a method, it is necessary to obtain accurate and dense geometry in advance to deal with complex background shapes. DR using 3D scanning with mobile structure sensors has also been used in recent years [1, 31]. Another method generates background images with an appropriate appearance by synthesizing many images from the Internet [21]. This method compensates for the appearance of images on the Internet using only Homography transformation to match the appearance of the DR user's camera image under the assumption that images that were taken very close to the current camera position exist on the Internet.

2.2 Online-Observation-based DR

The online-observation-based methods generate a background image to complete the ROI by observing the background with cameras located at different positions from the user's viewpoint, in parallel with the DR processing. As a result, it is possible to reflect background changes in the results in real time. This type of method has been proposed to be used to see through walls [2], through cars driving in front of a user [32], and through hands and tools in manual work [28], where you need to see beyond obstacles in real time.

In other approaches to achieve this, for example, Enomoto et al. use multiple hand-held cameras to erase obstructions from an image by estimating multiple camera poses using markers and generating a background image with an appropriate appearance using a homography transformation under the assumption that the background is planar [5]. For relaxing restrictions on background shape, an RGB-D camera [24, 28] and a stereo camera [32] have been used for DR to enable the elimination of objects and adapt to changes in the scene even when the background is not flat.

2.3 Inpainting-based DR

The inpainting-based methods estimate a plausible background from the surroundings without observing the actual background of obstacles to complete the ROI for DR. This method was initially developed to remove markers used in augmented reality [33]. Since this method straightforwardly generates textures on the markers, it does not produce good results when the textures around the ROI are complex. More plausible marker removal was achieved using a patch-based inpainting method [15]. Since typical patch-based inpainting does not run in real time, this method inpaints the ROI in the initial frame, and the inpainted result is transformed by homography and composited in subsequent frames, assuming that the markers are placed on a plane. On the other hand, a method has been proposed to achieve DR in real time by inpainting every frame by speeding up the patch-based method, targeting both markers and general objects [7]. However, this method also works well only with objects on a plane.

As a solution to alleviate the limitation of the background shape, several methods have been proposed to perform inpainting on keyframes, and to composite the inpainted images by performing geometric transformations on each frame [13, 14, 25]. A method has been proposed to extract feature points by SLAM [18] and plausibly erase 3D objects by inpainting under the assumption that the background is composed of multiple planes [14]. As another approach, a method has been proposed to handle curved background shapes by deforming the background texture generated by inpainting to match the texture around the ROI [13]. A method that fuses the results of inpainting for multiple keyframes has also been proposed to deal with uneven background shapes [25]. Another method that performs inpainting every frame using fast neural network inference without any restriction on the background shape has been proposed.

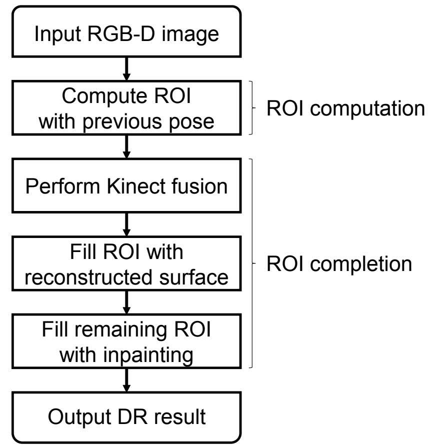


Figure 1: Flow of our DR processing

However, it assumes that a new virtual object is synthesized from above, and the quality of the inpainting is not still so high [12].

3 Proposed Method

3.1 Overview

Before starting the DR processing, the users specify 2D ROI in the first image by manually selecting the rectangle around an object to be erased on a 3D plane. The 3D ROI is then generated from the 2D ROI with a depth image to define the region to erase in the 3D space. In our method, online surface reconstruction is conducted by masking the 3D ROI in the incoming images to avoid reconstructing the objects to be erased.

The flow of our DR processing is illustrated in Figure 1. First, the 3D ROI is projected onto an incoming RGB-D image using the previous camera pose to determine the 2D ROI in the image. We assume small relative motion, as described in Section 3.2. Next, the RGB-D image is masked with the 2D ROI to remove the object in the image. Then, the masked RGB-D image is input to Kinect fusion (Kinfu) [10] to estimate the camera pose and reconstruct the scene. Finally, the method to complete each pixel inside the 2D ROI is adaptively selected. The RGB value in the image rendered with the reconstructed scene is primarily used when available. Otherwise, the inpainting-based method is applied to the remaining pixels. From a timeline point of view, the inpainting-based method completes the ROI at the beginning because the object's background has not been observed yet. The observation-based method adaptively completes the region with online surface reconstruction as the camera moves. In summary, our method gradually changes the background image obtained by the inpainting-based method to that obtained by the observation-based method.

3.2 ROI computation

At the beginning of the DR processing, the users specify the 2D ROI on a 3D plane by using a rectangle in the image, as illustrated in Figure 2. In this figure, the users tried to set the 3D ROI for a doll. Next, the 3D coordinate of each corner on a 3D plane in the world coordinate system is computed by using its corresponding depth value. Then, the rectangle surrounding the four corners is computed. Finally, the 3D ROI is generated by setting the height to the rectangle. The generated 3D ROI corresponds to the 3D cuboid to erase an object in the world coordinate system. Since the depth image is available, the 3D cuboid can be generated from a single view image.

At each incoming frame in the DR processing, the eight vertices of the 3D ROI are first projected onto the image at time t by using

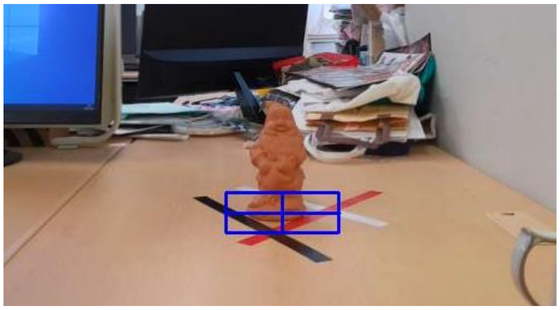
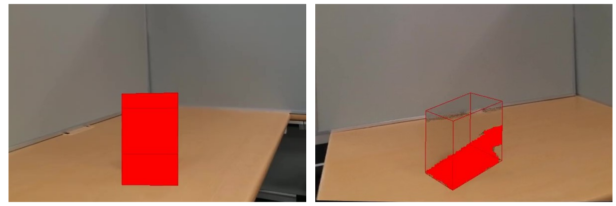


Figure 2: User's ROI input for a doll



(a) Initial view (b) View after camera movement

Figure 3: Obsevation-based ROI completion

a previous camera pose at time $t-1$. Next, the convex hull of the projected vertices is computed to generate the 2D ROI in the image. Then, the 2D ROI is used to mask the input RGB-D image, which becomes the input to Kinfu. Finally, the method to complete each pixel inside the 2D ROI is adaptively selected frame by frame according to the conditions described in Section 3.3.

We use the previous camera pose in the projection process for the following reason. Our method avoids reconstructing the object to be erased in Kinfu. Ideally, an incoming image is first masked by projecting the 3D ROI with the current camera pose. Then, the masked image is used in Kinfu to reconstruct the scene without the object. However, this process is not feasible because the current camera pose is acquired after performing Kinfu. For this reason, we use the previous camera pose by assuming that the relative motion between successive frames is small. This assumption does not cause a severe problem in practice when the 3D ROI is set up large enough. Also, we can use motion prediction by assuming uniform linear motion to estimate the 2D ROI accurately [29].

3.3 ROI completion

Our method adaptively performs two ROI completion methods according to the camera motion online. The observation-based method completes pixels where the color and depth are observed as the camera moves [10]. The remaining pixels are completed with the inpainting-based method [16], which iteratively searches for and synthesizes texture patterns to minimize an energy function based on patch similarity.

First, Kinfu is performed with a masked RGB-D image to reconstruct the scene without the object being erased and compute the current camera pose. Then, the synthesized image is rendered at the computed camera pose with the reconstructed surface and color obtained by colored Kinfu. The rendered image is close to the input RGB image when the computed camera pose is accurate.

The process of the ROI completion is as follows. First, the same 2D ROI is cropped in the rendered image. The cropped image contains the RGB values at the pixels where the object's background is observed. The black is rendered at the pixels where the surface is not reconstructed. As the first DR processing, the cropped image is inserted into the ROI in the input RGB image. However, the rendered texture is not always available. For instance, the object's background to be erased is not observed at the beginning. Therefore, the inpainting-based method completes the black pixels in the ROI as second DR processing.

The example is explained with Figure 3. Figure 3(a) and Figure 3(b) shows the results of observation-based method before and after the camera moves, respectively. Red pixels represent the 2D ROI to be completed by the inpainting-based method. In Figure 3(b), the upper left part of the ROI is completed while the camera moves to the right. This result is because the left part is previously reconstructed in Kinfu before the camera moves and is filled with the rendered image. The remaining ROI in red is the region to be completed with the inpainting-based method. By using inpainting-based methods only for the region that cannot be completed with the observation-based method, DR is achieved as plausible as possible.

4 Evaluation

The effectiveness of the proposed method is investigated for scenes with various characteristics: uniform color planes, textured planes, and 3D structure. We compare the results of the observation-based method by Kinfu [10], the inpainting-based method [16], and our proposed method from a timeline point of view. We used default parameters in Kinfu through the evaluation. Since the inpainting-based method [16] takes several seconds for each image, image sequences were taken in advance, and the methods were applied to them online in the experiments. In each figure, the input image is shown in (a), the result of the inpainting-based method and its close-up are shown in (b) and (e), the result of the observation-based method is shown in (c), and the result of our method and its close-up are shown in (d) and (f), respectively.

4.1 Uniform color planes

Figures 4 to 6 show the comparison results in the uniform color planes scene. The camera was not moved in the first frame, as shown in Figure 4. Since the object background was not observed, the observation-based method was not able to complete the pixels. Instead, the inpainting-based method completed all the pixels by inferring the texture from the surrounding pixels. Therefore, the results of our method are identical to the result of the inpainting-based method in this frame.

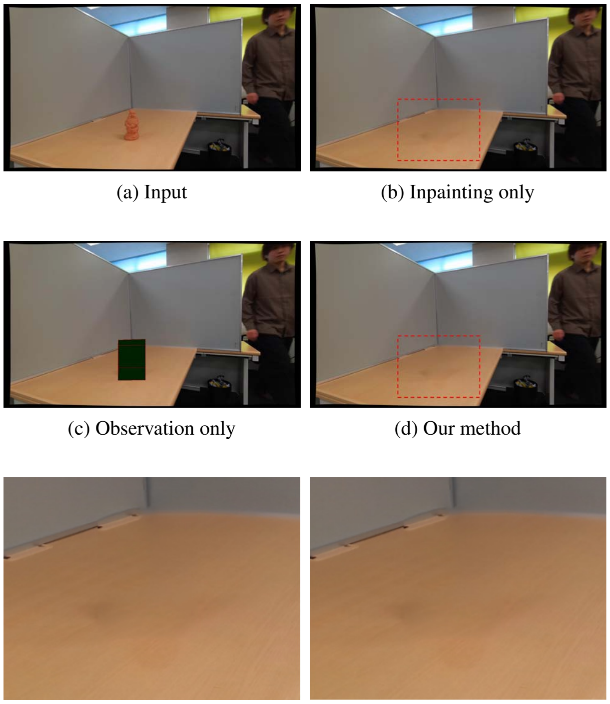
In the 150th frame, as shown in Figure 5, the observation-based method partially completed the ROI as the camera moved. In the 301st frame, as shown in Figure 6, the observation-based method was applied to almost all of the ROI by moving the camera. Since only the bottom of the ROI was not able to be observed by the camera, the inpainting-based method was applied there. Note that we sometimes had insufficient inpainting results around the boundary parts of the ROI in our result due to the artifacts in Kinfu in our implementation.

These results show that the inpainting-based method alone can produce a plausible image in such simple scenes with uniform color. Therefore, the effect of integrating observation and inpainting is not obvious for simple scenes.

4.2 Textured planes

Figures 7 to 9 show the comparison results in the textured planes scene. As with the experiment in the previous section, the initial frame results shown in Figure 7 are from the inpainting-based method only.

From the 179th to the 299th frame, the observation-based method was partially applied as the camera moved, as shown in Figures 8 and 9. In the results of the inpainting-based method shown in (c) in the figures, the wall texture extends to the floor, but in the results of the proposed method shown in (f) in the figures, the wall and

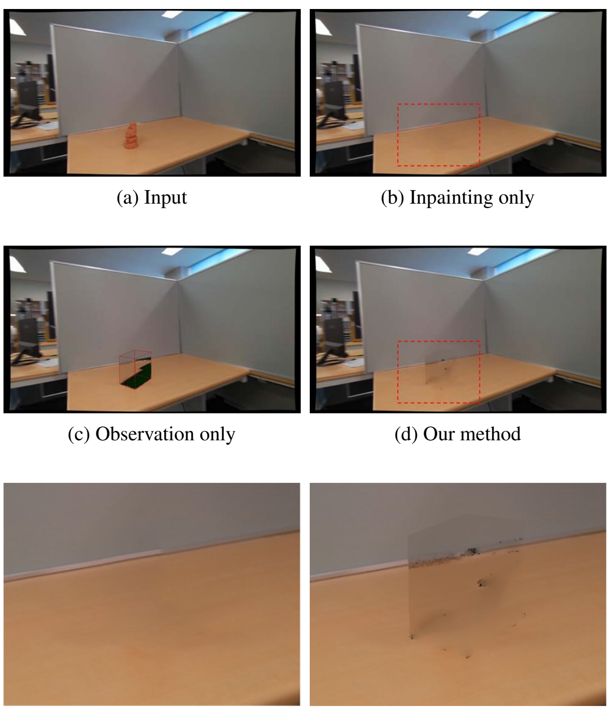


(a) Input (b) Inpainting only

(c) Observation only (d) Our method

(e) Dotted rectangle in (b) (f) Dotted rectangle in (d)

Figure 4: 1st frame in uniform color planes

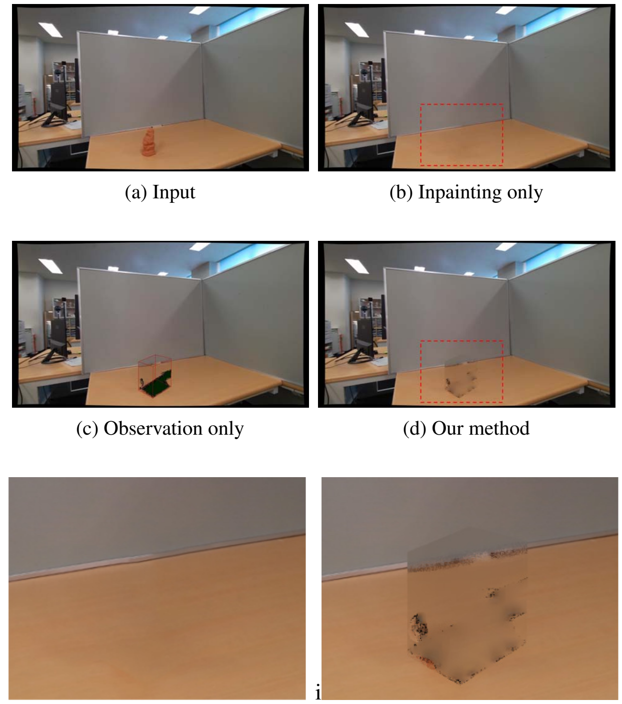


(a) Input (b) Inpainting only

(c) Observation only (d) Our method

(e) Dotted rectangle in (b) (f) Dotted rectangle in (d)

Figure 5: 150th frame in uniform color planes



(a) Input (b) Inpainting only

(c) Observation only (d) Our method

(e) Dotted rectangle in (b) (f) Dotted rectangle in (d)

Figure 6: 301st frame in uniform color planes

floor textures are separated exactly, although the image is rough and blurred.

Our method computed the 3D color reconstruction results with RGB-D images. However, the detail of the texture pattern was not reconstructed and was inferior to that obtained by the inpainting-based method. The reason may be the low resolution of the voxel space in Kinfu. Therefore, the problem can be solved by increasing the voxel resolution. In addition, we need to consider the color adjustment of the image so that the color of the ROI matches that of its surrounding.

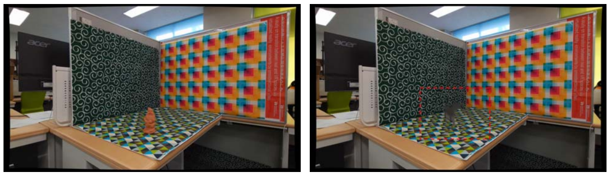
4.3 3D structure

Figures 10 to 12 show the comparison results in the 3D structure scene. In the first frame, as shown in Figure 10, in which the entire ROI was completed by the inpainting-based method, a stuffed penguin is placed in the background behind the obstacle. However, its feet were not reproduced at all. This is because textures that do not exist in the rest of the image are not reproduced in the ROI by patch-based inpainting, which we used in this case.

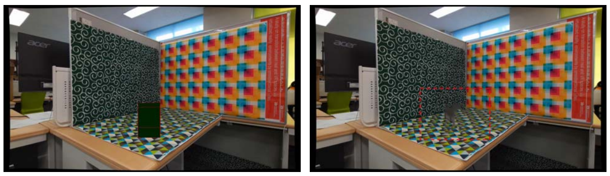
Figures 11 and 12 show the results after moving the camera. Again, the inpainting-based method could not reproduce the background accurately, and the texture varied greatly from frame to frame, causing flickering. On the other hand, the results from the proposed method, which gradually replaces the image in the ROI with that obtained by the observation-based method, show stuffed penguin's feet and books. Although some texture discontinuities between the ROI and its surrounding area are observed due to the camera pose estimation and depth observation accuracy, the background is presented fairly accurately compared to the inpainting-only case.

5 Conclusion

We proposed to combine the advantages of two types of DR: observation-based and inpainting-based methods. By integrating



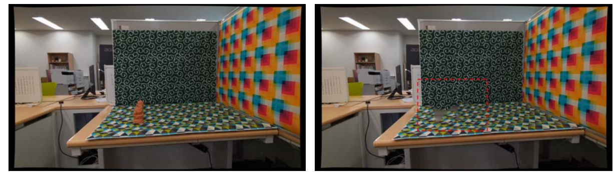
(a) Input (b) Inpainting only



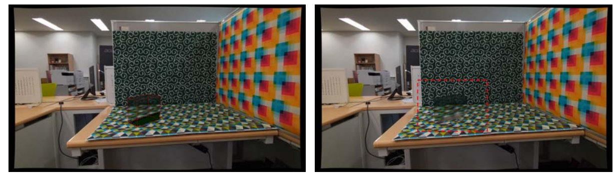
(c) Observation only (d) Our method

(e) Dotted rectangle in (b) (f) Dotted rectangle in (d)

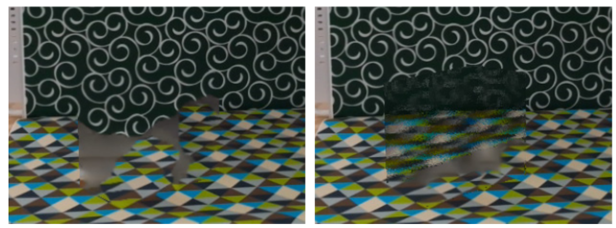
Figure 7: 1st frame in textured planes



(a) Input (b) Inpainting only

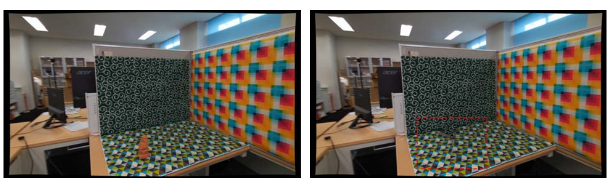


(c) Observation only (d) Our method

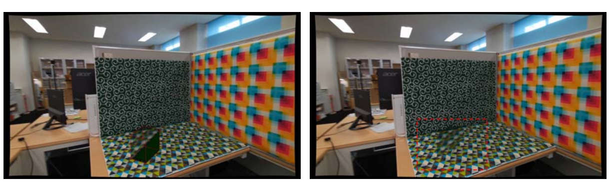


(e) Dotted rectangle in (b) (f) Dotted rectangle in (d)

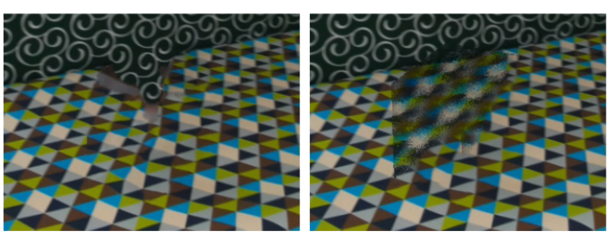
Figure 9: 299th frame in textured planes



(a) Input (b) Inpainting only



(c) Observation only (d) Our method



(e) Dotted rectangle in (b) (f) Dotted rectangle in (d)

Figure 8: 179th frame in textured planes

an observation-based method based on Kinfu, which can generate 3D models from depth and color images, and a patch-based inpainting method, we achieved Diminished Reality that can present more accurate backgrounds online.

In the experiments, we used various scenes, including a simple plane scene with uniform colors, a simple plane scene with textured patterns, and a 3D structure scene, and confirmed that the proposed method gives better results than inpainting alone especially for scenes with textures and 3D structures.

Currently, our implementation does not achieve real-time processing. In the future, we plan to implement the system on a GPU to achieve real-time processing. In addition, to improve the quality of the images, we will examine appropriate voxel size in Kinfu and image color adjustment methods.

Acknowledgments

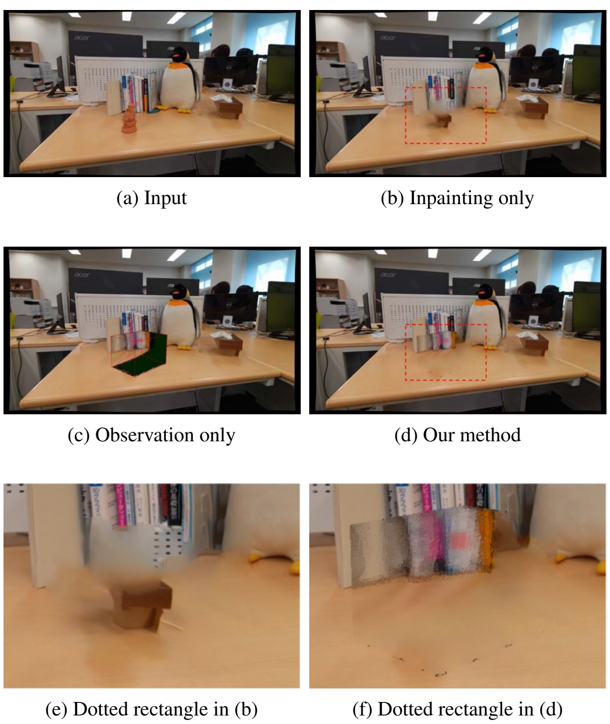
This work was supported by JSPS KAKENHI Grant Numbers JP18H03273 and JP21H03483.

(a) Input (b) Inpainting only

(c) Observation only (d) Our method

(e) Dotted rectangle in (b) (f) Dotted rectangle in (d)

Figure 10: 1st frame in 3D structure



(a) Input (b) Inpainting only

(c) Observation only (d) Our method

(e) Dotted rectangle in (b) (f) Dotted rectangle in (d)

Figure 11: 200th frame in 3D structure

(a) Input (b) Inpainting only

(c) Observation only (d) Our method

(e) Dotted rectangle in (b) (f) Dotted rectangle in (d)

Figure 12: 301st frame in 3D structure

REFERENCES

[1] E. Andre and H. Hlavacs. Diminished reality based on 3d-scanning. In *Joint International Conference on Entertainment Computing and Serious Games*, pp. 3–14. Springer, 2019.

[2] P. Barnum, Y. Sheikh, A. Datta, and T. Kanade. Dynamic seethroughs: Synthesizing hidden views of moving objects. In *Proceedings of IEEE International Symposium on Mixed and Augmented Reality*, pp. 111–114, 2009.

[3] Y. F. Cheng, H. Yin, Y. Yan, J. Gugenheimer, and D. Lindlbauer. Towards understanding diminished reality. In *CHI Conference on Human Factors in Computing Systems*, pp. 1–16, 2022.

[4] F. I. Cosco, C. Garre, F. Bruno, M. Muzzupappa, and M. A. Otaduy. Augmented touch without visual obtrusion. In *IEEE International Symposium on Mixed and Augmented Reality*, pp. 99–102. IEEE, 2009.

[5] A. Enomoto and H. Saito. Diminished reality using multiple handheld cameras. In *Proceedings of ACCV'07 Workshop on Multi-dimensional and Multi-view Image Processing*, pp. 130–135, 2007.

[6] V. Gkitsas, V. Sterzentsenko, N. Zioulis, G. Albanis, and D. Zarpalas. Panodr: Spherical panorama diminished reality for indoor scenes. In *IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pp. 3716–3726, 2021.

[7] J. Herling and W. Broll. High-quality real-time video inpainting with pixmix. *IEEE transactions on Visualization and Computer Graphics*, 20(6):866–879, 2014.

[8] N. Ienaga, F. Bork, S. Meerits, S. Mori, P. Fallavollita, N. Navab, and H. Saito. First deployment of diminished reality for anatomy education. In *IEEE International Symposium on Mixed and Augmented Reality (ISMAR-Adjunct)*, pp. 294–296. IEEE, 2016.

[9] S. Ikeda, I. Takemura, A. Kimura, and F. Shibata. Diminished reality system based on open-source software for self-driving mobility. In *2018 IEEE international symposium on mixed and augmented reality adjunct (ISMAR-Adjunct)*, pp. 354–357. IEEE, 2018.

[10] S. Izadi, D. Kim, O. Hilliges, D. Molyneaux, R. Newcombe, P. Kohli, J. Shotton, S. Hodges, D. Freeman, A. Davison, et al. Kinectfusion: real-time 3d reconstruction and interaction using a moving depth camera. In *ACM symposium on User interface software and technology*, pp. 559–568, 2011.

[11] S. Jarusirisawad, T. Hosokawa, and H. Saito. Diminished reality using plane-sweep algorithm with weakly-calibrated cameras. *Progress in Informatics*, (7):11–20, 2010.

[12] M. Kari, T. Grosse-Puppendahl, L. Falconeri Coelho, A. Fender, D. Bethge, R. Schütte, and H. Christian. Transformr: Pose-aware object substitution for composing alternate mixed realities. In *Proceedings of IEEE International Symposium on Mixed and Augmented Reality*, pp. 69–79. IEEE, 2012.

[13] N. Kawai, T. Sato, Y. Nakashima, and N. Yokoya. Augmented reality marker hiding with texture deformation. *IEEE transactions on Visualization and Computer Graphics*, 23(10):2288–2300, 2017.

[14] N. Kawai, T. Sato, and N. Yokoya. Diminished reality based on image inpainting considering background geometry. *IEEE transactions on visualization and computer graphics*, 22(3):1236–1247, 2016.

[15] N. Kawai, M. Yamasaki, T. Sato, and N. Yokoya. Diminished reality for ar marker hiding based on image inpainting with reflection of luminance changes. *ITE Transactions on Media Technology and Applications*, 1(4):343–353, 2013.

[16] N. Kawai and N. Yokoya. Image inpainting considering symmetric patterns. In *Proceedings of the 21st International Conference on Pattern Recognition (ICPR2012)*, pp. 2744–2747. IAPR, 2012.

[17] D. Kido, T. Fukuda, and N. Yabuki. Diminished reality system with real-time object detection using deep learning for onsite landscape simulation during redevelopment. *Environmental Modelling & Software*, 131:104759, 2020.

[18] G. Klein and D. Murray. Parallel tracking and mapping for small ar workspaces. In *Proceedings of IEEE International Symposium on Mixed and Augmented Reality*, pp. 225–234, 2007.

[19] O. Korkalo, M. Aittala, and S. Siltanen. Light-weight marker hiding for augmented reality. In *Proceedings of IEEE International Symposium on Mixed and Augmented Reality*, pp. 247–248. IEEE, 2010.

[20] C. Kunert, T. Schwandt, and W. Broll. An efficient diminished reality approach using real-time surface reconstruction. In *International Conference on Cyberworlds*, pp. 9–16. IEEE, 2019.

[21] Y. Li, Zhuwen amd Wang, J. Guo, L.-F. Cheong, and S. ZhiYing Zhou. Diminished reality using appearance and 3d geometry of internet photo collections. In *Proceedings of IEEE International Symposium on Mixed and Augmented Reality*, pp. 11–19, 2013.

[22] P. Lindemann and G. Rigoll. A diminished reality simulation for driver-car interaction with transparent cockpits. In *IEEE Virtual Reality (VR)*, pp. 305–306. IEEE, 2017.

[23] S. Mann and J. Fung. Videoorbits on eye tap devices for deliberately diminished reality or altering the visual perception of rigid planar patches of a real world scene. *EYE*, 3:P3, 2001.

[24] S. Meerits and H. Saito. Real-time diminished reality for dynamic scenes. In *IEEE International Symposium on Mixed and Augmented Reality Workshops*, pp. 53–59. IEEE, 2015.

[25] S. Mori, O. Erat, W. Broll, H. Saito, D. Schmalstieg, and D. Kalkofen. Inpaintfusion: Incremental rgb-d inpainting for 3d scenes. *IEEE Transactions on Visualization and Computer Graphics*, 26(10):2884–3007, 2020.

[26] S. Mori, J. Herling, W. Broll, N. Kawai, H. Saito, D. Schmalstieg, and D. Kalkofen. 3d pixmix: Image inpainting in 3d environments. In *Proceedings of IEEE International Symposium on Mixed and Augmented Reality*, pp. 1–2, 2018.

[27] S. Mori, S. Ikeda, and 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(1):1–14, 2017.

[28] S. Mori, M. Maezawa, and H. Saito. A work area visualization by multi-view camera-based diminished reality. *Multimodal Technologies and Interaction*, 1(3):18, 2017.

[29] R. Mur-Artal, J. M. M. Montiel, and J. D. Tardos. Orb-slam: a versatile and accurate monocular slam system. *IEEE transactions on robotics*, 31(5):1147–1163, 2015.

[30] A. Plopski, A. V. Taylor, E. J. Carter, and H. Admoni. Invisiblerobot: Facilitating robot manipulation through diminished reality. In *IEEE International Symposium on Mixed and Augmented Reality Adjunct (ISMAR-Adjunct)*, pp. 165–166. IEEE, 2019.

[31] G. Queguiner, M. Fradet, and M. Rouhani. Towards mobile diminished reality. In *IEEE International Symposium on Mixed and Augmented Reality Adjunct (ISMAR-Adjunct)*, pp. 226–231. IEEE, 2018.

[32] F. Rameau, H. Ha, K. Joo, J. Choi, K. Park, and I. S. Kweon. A real-time augmented reality system to see-through cars. *IEEE Transactions on Visualization and Computer Graphics*, 22(11):2395–2404, 2016.

[33] S. Siltanen. Texture generation over the marker area. In *Proceedings of IEEE International Symposium on Mixed and Augmented Reality*, pp. 253–254. IEEE, 2006.

[34] S. Siltanen. Diminished reality for augmented reality interior design. *The Visual Computer*, 33(2):193–208, 2017.

[35] S. Siltanen, H. Saraspää, and J. Karvonen. A complete interior design solution with diminished reality. In *IEEE International Symposium on Mixed and Augmented Reality*, pp. 371–372. IEEE, 2014.

[36] A. V. Taylor, A. Matsumoto, E. J. Carter, A. Plopski, and H. Admoni. Diminished reality for close quarters robotic telemanipulation. In *IEEE/RSJ International Conference on Intelligent Robots and Systems*, pp. 11531–11538. IEEE, 2020.

[37] K. Yagi, K. Hasegawa, and H. Saito. Diminished reality for privacy protection by hiding pedestrians in motion image sequences using structure from motion. In *IEEE International Symposium on Mixed and Augmented Reality (ISMAR-Adjunct)*, pp. 334–337. IEEE, 2017.

[38] A. E. Yantaç, D. Çorlu, M. Fjeld, and A. Kunz. Exploring diminished reality (dr) spaces to augment the attention of individuals with autism. In *IEEE International Symposium on Mixed and Augmented Reality Workshops*, pp. 68–73. IEEE, 2015.