Reproduction of Specular Reflection Using 3D Gaussian Splatting in Diminished Reality for AR Marker Hiding

Issa Narutomi[1,*], Norihiko Kawai[1]

[1]Osaka Institute of Technology (OIT), 1-79-1 Kitayama, Hirakata-shi, Osaka, 573-0196 Japan

Abstract

In augmented reality (AR) games, markers are often used to synthesize and present virtual objects on camera images. However, if the AR markers remain visible on the images, it reduces the quality of the AR experience. This problem can be solved by using diminished reality (DR), which visually removes real objects from real-time images and presents the background. However, conventional methods of diminished reality have difficulties with presenting natural background images in areas containing local luminance changes such as specular reflections. In this study, we propose a method for AR marker hiding that can handle specular reflections by using 3D Gaussian Splatting (3DGS), which generates free viewpoint images from a set of 2D images. The proposed method synthesizes a texture with appropriate specular reflections in the marker area according to the current camera pose in AR. Experiments demonstrate the effectiveness of the proposed method by comparing results with conventional methods.

Keywords

AR marker, Specular reflection, 3D Gaussian Splatting, Diminished reality, Loss function

1. Introduction

Games using augmented reality (AR) provide realistic visual experiences by synthesizing and presenting 3D virtual objects in camera images, and markers are often used to estimate camera poses and determine the positions of virtual objects in such games. However, if the AR markers remain visible when 3D virtual objects are displayed, the quality of the AR experience is reduced. This problem can be solved using diminished reality (DR) [1], which visually removes real objects from real-time images and presents the background. However, most of the conventional methods of diminished reality have the problem that local luminance changes, such as specular reflection, are not reflected in the background image synthesized on the AR markers when the camera moves.

In this study, we focus on specular reflections whose positions change according to camera movement and propose a novel method of diminished reality to improve the quality of background images synthesized on AR markers placed on planes where specular reflections occur. The proposed method uses 3D Gaussian Splatting (3DGS) [2], which generates free viewpoint images from a set of 2D images. In this study, as a pre-process, by excluding the loss of the AR marker area in images during 3DGS training, we can reconstruct a 3D scene in which the AR marker is removed and specular reflections are reproduced in the AR marker area according to the camera pose. In the process of performing AR, the current camera pose is estimated from the marker, and the image is rendered using the camera pose in the scene generated by 3DGS, and the rendered image is synthesized into the marker area to achieve marker hiding with specular reflection. In the following sections, we first provide an overview of conventional methods and then describe the proposed method.

2. Related work

Diminished reality can be classified into three main categories: the first method uses multiple cameras set at different viewpoints, the second method uses a previously captured background image, and the third method uses inpainting to estimate the background texture based on the information around the area.

In the method using multiple cameras at different viewpoints [3, 4, 5], unwanted objects are removed by transforming and copying the background texture captured from another camera. This category is not related to the proposed method because these methods cannot be applied to AR marker hiding situations, where an AR marker is placed on a surface such as a wall or floor, completely hiding its background.

In the method using pre-captured background images[6, 7, 8, 9], unwanted objects are removed by combining pre-captured background images with image-based rendering according to the current camera pose. However, to accurately reproduce specular reflections, which shift as the camera's pose change, background images need to be densely captured from various camera

APMAR'24: The 16th Asia-Pacific Workshop on Mixed and Augmented Reality, Nov. 29-30, 2024, Kyoto, Japan

*Corresponding author.

m1m24a35@st.oit.ac.jp (I. Narutomi); norihiko.kawai@oit.ac.jp (N. Kawai)

0000-0002-7859-8407 (N. Kawai)

angles. Additionally, these methods require capturing background images in advance in scenes without unwanted objects. In contrast, the proposed method does not have to capture the background of the AR marker to be removed.

Methods using inpainting [10, 11, 12, 13, 14, 15, 16] remove unwanted objects by estimating the background of the object from the surrounding area. These techniques can be further divided into two types: one in which inpainting is performed on every frame [10, 12, 16], and one in which inpainting is performed on a single frame and the inpainted texture is geometrically and photometrically adjusted on every frame [11, 13, 14, 15]. The former methods require high-speed inpainting, and even the latest deep learning-based method [16] has limited removal quality. On the other hand, the latter methods give better inpainting quality because it can take more time. However, it is important to adjust the inpainted image every frame.

As for the photometric adjustment, while many of the latter methods adjust the luminance of the background image for global luminance changes [11, 13, 14], the method in [15] deals with specular reflection. In this method, specular reflections are added to the inpainted background texture by detecting specular reflections in different brightness levels around the target area and fitting ellipses to them. Due to this mechanism, if most of the specular reflections are included within the target area and do not appear in the surrounding area, the accuracy of the specular reflection area estimation may decrease significantly.

The proposed method falls into a hybrid area between the second category using images captured in advance, and the third category using inpainting. However, unlike the conventional methods in the second category, the images captured in advance includes the target marker to be removed, and it is not necessary to capture the background behind the marker. Also, unlike the conventional methods in the third category, instead of removing the target object from the image using patch-based or deep learning-based inpainting, we construct a 3D scene without the target marker by manipulating the loss during 3DGS training. These enable to achieve diminished reality that reproduces appropriate specular reflections in the AR marker area according to the camera pose.

3. Proposed method

3.1. Overview

The proposed method is divided into two processes: preprocess and process during the AR experience. The flowchart of the entire proposed method is shown in Figure 1. In the pre-process stage of the proposed method,

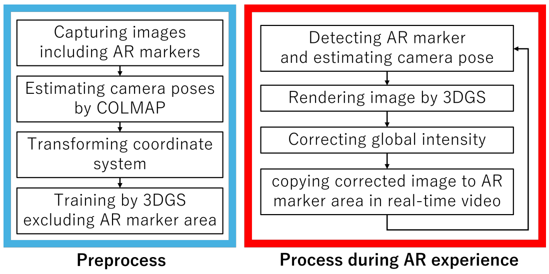


Figure 1: Flowchart of the entire proposed method.

a set of images including AR markers are captured, and the camera poses are estimated by COLMAP [17], a type of Structure from Motion (SfM) software, and the coordinate system is transformed. 3D Gaussian Splatting [2] is then trained using these as input. During training, the losses of the AR marker areas are excluded, so that the 3D scene is reconstructed as if the AR markers in the image do not exist, enabling to generate images without AR markers from free viewpoints.

In the AR experience process, the image is rendered from the reconstructed 3D scene according to the current camera pose estimated from an AR marker. The texture of the area corresponding to the AR marker area in the rendered image is copied frame by frame to the same area in the real-time image after global intensity correction, thereby achieving the reproduction of natural specular reflections. In the following sections, an overview of 3D Gaussian Splatting and the details of the proposed method are described.

3.2. Overview of 3D Gaussian Splatting

3D Gaussian Splatting (3DGS) reconstructs a 3D scene represented by a 3D Gaussian distribution from a 2D image set through training and can generate images from new viewpoints that are not included in the 2D image set via volume rendering. Before training, COLMAP is used to estimate the intrinsic and extrinsic parameters of the cameras, as well as the 3D coordinates of the feature points for each image in the 2D image set. 3DGS uses the 2D image set along with the estimated camera poses and the 3D coordinate of each feature point as training data.

In the training of 3DGS, the parameters of the 3D Gaussian function representing the Gaussian distribution in 3D space are initialized, and the loss between an image selected at random from a set of 2D images and the rendered image with the same camera pose is calculated by the difference in pixel values at each pixel and the DSSIM. The Gaussian distribution representing the 3D scene is updated according to the gradient calculated beased on the loss.

(a) Input image (b) Rendered image

Figure 2: Training image after marker area is filled.

3.3. Pre-process of the proposed method

In the pre-processing stage, many images are captured from various camera poses, ensuring that AR markers are included in the images. These images are used as input for the training of 3D Gaussian Splatting. In this study, the training is conducted in such a way that no loss occurs in the AR marker area, allowing free viewpoint images to be generated as if the AR markers were not present in the scene.

Specifically, the four corners of the AR marker in the input image are obtained, and the inner area containing these corners and its surroundings are filled with a fixed pixel value, as shown in Figure 2(a). The same area is filled with the same pixel value in the rendered image, as shown in Figure 2(b), to reduce the loss in the AR marker area to zero. This results in reconstructing the 3D scene as if the AR markers did not exist. In this case, as shown in Figure 3, in the space constructed by 3D Gaussian Splatting, a virtual light source is generated at a position symmetrical to the actual light source with respect to the plane on which the marker exists, based on the reflection of specular light. By rendering an image in this constructed environment according to the camera pose, specular reflections appear in the marker area at appropriate positions.

In the process during the AR experience mentioned below, the AR marker in the real-time video is detected, and the camera pose is calculated in a coordinate system where the center of the AR marker is the origin (marker coordinate system). Based on this information, a virtual camera is set up and the same view is rendered in the reconstructed 3D scene. However, since the coordinate system of the 3D scene (COLMAP coordinate system) is established during processing by COLMAP and differs from the marker coordinate system, simply using the current camera pose as the virtual camera one in the reconstructed 3D scene does not yield the same view with the camera image.

Therefore, before training with 3D Gaussian Splatting, the COLMAP coordinate system is transformed to the marker coordinate system using the model aligner, which is one of the functions of COLMAP. Specifically, the poses of all cameras are transformed so that all camera positions in the COLMAP coordinate system match the camera positions in the marker coordinate system, which are obtained by detecting the AR marker. Examples of the coordinate system and camera poses before and after the transformation are shown in Figures 4(a) and 4(b), respectively.

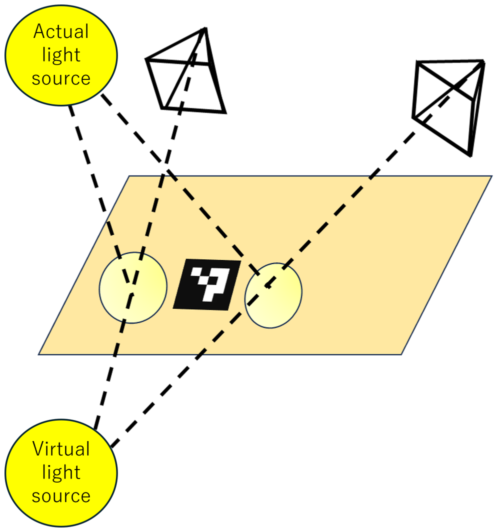
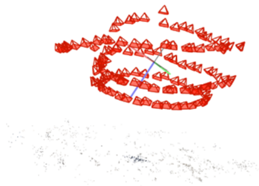
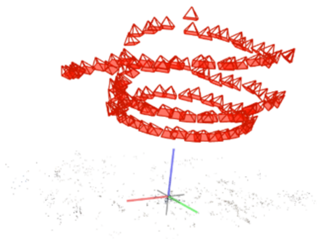


Figure 3: Light sources in virtual space.



(a) Before transformation



(b) After transformation

Figure 4: Coordinate system transformation.

(a) Before adjustment (b) After adjustment

Figure 5: Comparison before and after texture intensity adjustment process.

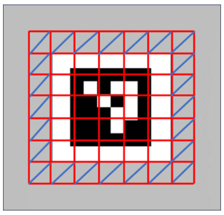


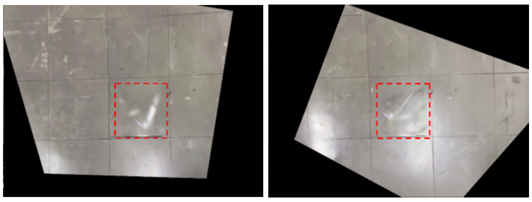
Figure 6: AR marker's surrounding area.

3.4. Process during AR experience

In the process during the AR experience, an image is rendered according to the current camera pose and synthesized into the AR marker area of the real-time video as the background image. However, if the rendered image is copied directly, the difference in intensity of images between the camera used for training data collection in 3DGS and the AR camera may cause the inconsistency between the texture of the AR marker area and its surrounding texture in the real-time video as shown in Figure 5(a).

To eliminate the inconsistency caused by the difference in intensity between the textures, we employ the intensity adjustment methods in [11, 13]. Specifically, both the real-time image and the rendered image are first transformed as if the scene is captured by a camera in front of the marker. Luminance change coefficients of grids of the marker's surrounding area as shown in Figure 6 between the rendered image and real-time camera image are then calculated. The coefficients in the marker area are then interpolated using those around the marker. Finally, the pixel values of the rendered image are multiplied by the coefficients and copied to the marker area of the current frame, as shown in Figure 5(b).

Note that, while the conventional methods [11, 13] use the background image obtained by inpainting, which is the just one frame, the proposed method uses the rendered image obtained according to the current camera pose as the background for each frame. The reason for this is that, as shown in Figure 7, the positions of specular reflections also change in the area around the marker depending on the camera pose. By ensuring that the positions of specular reflections around the marker are consistent between the images to be compared, the luminance of the local specular reflection does not affect the global luminance adjustment.



(a) Certain camera pose (b) Different camera pose

Figure 7: Specular reflections appearing in the area around the marker.

Table 1
PC specifications.

OS	Windows 11
CPU	13th Gen Intel(R) Core(TM) i9-13900KF
GPU	NVIDIA GeForce RTX 4090
Memory	64.0GB

Table 2
Cameras and their resolution.

Training data	iPhone13 Pro	1280×720
Real-time image	Logicool C270HD	640×480

4. Experiments

To demonstrate the effectiveness of the proposed method, comparative experiments were conducted between the proposed method and the conventional method [11] in a scenario where an AR markers was placed on three types of planes where specular reflections occur. We used an ArUco marker as an AR marker provided by the aruco module of OpenCV-contrib, an extension of OpenCV. The specifications of the PC, camera, image resolution are shown in Tables 1 and 2, respectively, and the appearance of the experimental environment is shown in Figure 8. In the following sections, we describe the experiments in each plane in turn.

4.1. Experiment 1: floor with specular reflections

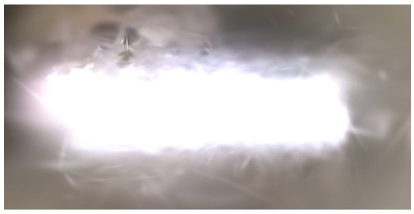
In the experiments conducted on a floor where specular reflections appear, we used 122 images as training data for 3DGS. Figure 9 shows an example of the training data. Figure 10 shows the rendered results obtained by original 3DGS training and the proposed 3DGS training, which excludes the AR marker area. These results demonstrate

Figure 8: Appearance of experimental environment.

Figure 9: An example of training data in Experiment 1.

(a) By original 3DGS (b) By proposed 3DGS

Figure 10: Rendering results by original and proposed 3DGS trainings in Experiment1.



(a) Virtual light source in rendered image when virtual camera is placed under the floor and facing downward

(b) Rendered image when virtual camera is facing upward from below the floor

Figure 11: Virtual light source in environment constructed by 3DGS.

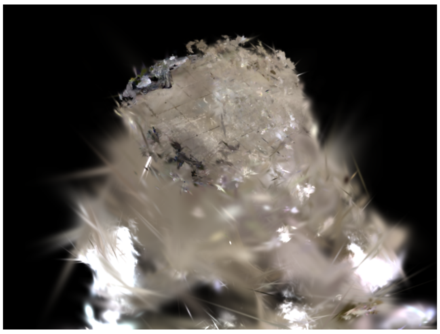


Figure 12: Constructed environment seen from the side.

that the method excluding the marker areas successfully generates a floor without the marker.

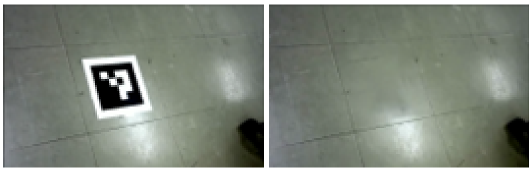
Figure 11 shows rendered images when the virtual camera is placed under the floor, and facing downward and upward. From the figure, we confirmed that a virtual light source was generated not at the actual light source position but under the floor, allowing the light to reach the viewpoint and appear as a specular reflection. Figure 12 shows the constructed environment from the side. From the figure, we can see that the floor in the 3D scene are represented not as a solid plane but as a hierarchical arrangement of Gaussian distributions with information on the color and opacity of the objects. For this reason, not only specular reflections, but also the colors of the floor in areas other than specular reflections are reproduced in the marker area.

Next, we show experimental results during AR experience. Figure 13 shows the input image, the result of the conventional method [11], the results of the proposed method before and after intensity adjustment. From the experimental results, we can see that the proposed method properly represents specular reflection in the area inside the marker, while the specular reflection is disconnected at the boundary of the marker by the conventional method. In addition, the intensity adjustment of the proposed method achieves adequate global luminance adjustment. Figure 14 shows the results from another viewpoint. We can confirm that the shape of specular reflection is slightly difficult to recognize, but compared to the result of the conventional method, the specular reflection can be definitely observed. The frame rate of the proposed method was 8-10 fps.

(a) Input image (b) Conventional method

(c) Before intensity adjustment (d) After intensity adjustment (Proposed method)

Figure 13: Comparison in Experiment 1.



(a) Input image (b) Conventional methods

(c) Before intensity adjustment (d) After intensity adjustment (Proposed method)

Figure 14: Comparison from another viewpoint in Experiment 1.

4.2. Experiment 2: marble-like black tiles

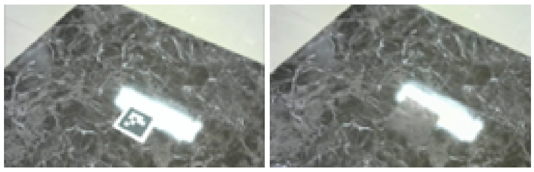
In the experiments conducted on black marble-like tiles, 121 images were used as training data. Figure 15 shows an example of the training data. Figures 16 shows the rendered results obtained by the original 3DGS training and the proposed 3DGS training. In this scene, the proposed method also makes the marker invisible, and black tiles appear on the marker area. However, the fineness of the white texture is slightly different from the surrounding texture.

Figures 17 and 18 show the input images, the results of the conventional method [11], the results of the proposed method before and after intensity adjustment during the AR experience with different viewpoints. In this experiment, the specular reflections, which were not reproduced by the conventional method, are more clearly visible in the result of the proposed method than that in Experiment 1. Although the proposed method produces some artifacts in the results, it can generate specular reflections that are geometrically consistent in the AR marker area. The frame rate of the proposed method was 8-10 fps in this experiment as well.

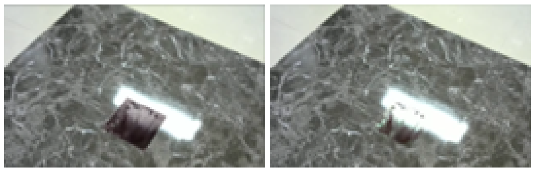
Figure 15: Example of training data in Experiment 2.

(a) By original 3DGS (b) By proposed 3DGS

Figure 16: Rendering results by original and proposed 3DGS trainings in Experiment 2.

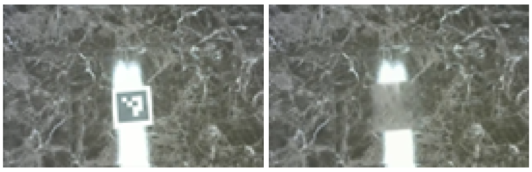


(a) Input image (b) Conventional methods

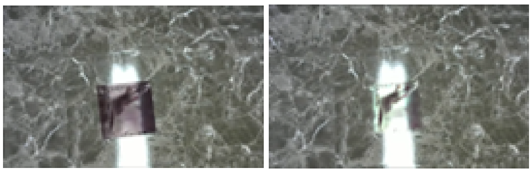


(c) Before brightness adjustment (d) Proposed method(After brightness adjustment)

Figure 17: Comparison in Experient 2.



(a) Input image (b) Conventional methods



(c) Before brightness adjustment (d) Proposed method(After brightness adjustment)

Figure 18: Comparison from another viewpoint in Experiment 2.

Figure 19: Example of training data in Experiment 3.

(a) By original 3DGS (b) By proposed 3GDS

Figure 20: Rendering results by original and proposed 3DGS trainings in Experiment 3.

4.3. Experiment 3: transparent board on a floor

In the experiments conducted with a transparent board on the floor, 122 images were used as training data. Figure 19 shows an example of the training data. Figures 20 shows the rendered results obtained by original 3DGS training and the proposed 3DGS training. In this scene as well, we can confirm that the markers have been naturally erased from the scene.

Figures 21 and 22 show the input images, the results of the conventional method, the results of the proposed method before and after intensity adjustment. In this experiment, the conventional method [11] could not reproduce specular reflection as well as the other experiments. On the other hand, the proposed method rendered the specular reflections. We confirmed that the geometrical consistency of the specular reflections in the AR marker area with the proposed method is better than that observed in Experiment 1. However, we also confirmed the difference in luminance between the specular reflections in the AR marker area and the surrounding area. In this experiment, the frame rate of the proposed method was 8-10 fps.

(a) Input image (b) Conventional methods

(c) Before brightness adjustment (d) Proposed method(After brightness adjustment)

Figure 21: Comparison in Experiment 3.

(a) Input image (b) Conventional methods

(c) Before brightness adjustment (d) Proposed method(After brightness adjustment)

Figure 22: Comparison from Another viewpoint in Experiment 3.

4.4. Discussion

From the results mentioned above, we confirmed that the proposed method is able to represent specular reflections at geometrically appropriate locations, which could not be reproduced by the conventional method. However, for planes with fine textures, as in Experiment 2, we found it difficult to reproduce the similar fine texture on the marker. In addition, as shown in Experiment 3, we confirmed that the intensity adjustment by the proposed method is effective for global difference in intensity in areas other than specular reflection areas, but it is not sufficient for adjusting the brightness of specular reflections. We consider that this is because the intensity adjustment is performed in the same way for the entire marker area, regardless of the presence or absence of specular reflection. Therefore, one solution to this problem may be to separate the intensity adjustment method for specular reflection areas and other areas.

5. Conclusion

In this study, we proposed a method to reproduce natural specular reflections using 3D Gaussian Splatting in diminished reality for AR marker hiding. The proposed method constructs a marker-free 3D scene by excluding marker areas when training 3DGS. During the AR experience, the images rendered by 3DGS are synthesized into the marker areas while adjusting the intensity to reproduce specular reflection.

Experimental results showed that the proposed method successfully reproduced specular reflections compared to conventional methods. On the other hand, we also confirmed that the specular reflection is not always perfectly consistent with the area around the marker. For example, artifacts in the specular reflection appear, fine textures in the surrounding area are not reproduced in the marker area, and difference in brightness of the specular reflection between the marker area and its surrounding area is observed.

Future work includes the development of a marker hiding method that can simultaneously reproduce specular reflection and fine texture on planes. We will also develop a diminished reality method that can reproduce specular reflections by removing various objects.

Acknowledgments

This study was supported by JSPS KAKENHI Nos. JP23K24875 and JP23K21689.

References

[1] 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) Article 17.

[2] B. Kerbl, G. Kopanas, T. Leimakuhler, G. Drettakis, 3d gaussian splatting for real-time radiance field rendering, ACM Transactions on Graphics 42 (2023) Article 139, 1–14.

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

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

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

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

[7] Z. Li, Y. Wang, J. Guo, L.-F. Cheong, S. Z. Zhou, Diminished reality using appearance and 3d geometry of internet photo collections, in: Proc. International Symposium on Mixed and Augmented Reality, 2013, pp. 11–19.

[8] S. Mori, F. Shibata, A. Kimura, H. Tamura, Efficient use of textured 3d model for preobservation-based diminished reality, in: Proc. International Symposium on Mixed and Augmented Reality Workshops, 2015, pp. 32–39.

[9] 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.

[10] O. Korkalo, M. Aittala, S. Siltanen, Lightweight marker hiding for augmented reality, in: Proc. International Symposium on Mixed and Augmented Reality, 2010, pp. 247–248.

[11] N. Kawai, M. Yamasaki, T. Sato, 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 (2013) 343–353.

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

[13] 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.
[14] S. Mori, J. Herling, W. Broll, N. Kawai, H. Saito, D. Schmalstieg, D. Kalkofen, 3d pixmix: Image inpainting in 3d environments, in: Proc. IEEE International Symposium on Mixed and Augmented Reality Adjunct (ISMAR-Adjunct), 2018, pp. 1–2.
[15] S. H. Said, M. Tamaazousti, A. Bartoli, Image-based models for specularity propagation in diminished reality, IEEE Transactions on Visualization and Computer Graphics 24 (2018) 2140–2152.
[16] 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: Proc. IEEE International Symposium on Mixed and Augmented Reality, 2021, pp. 69–79.
[17] J. L. Schonberger, J.-M. Frahm, X. Gao, Structure-from-motion revisited, in: Proc. IEEE Conference on Computer Vision and Pattern Recognition, 2016, pp. 4104–4113.