SIP (2018), vol. 7, e12, page 1 of 9 © The Authors, 2018.

doi:10.1017/ATSIP.2018.13

OVERVIEW PAPER

An overview of augmented visualization: *observing the real world as desired*

SHOHEI MORI[1] AND HIDEO SAITO[2]

Over 20 years have passed since a free-viewpoint video technology has been proposed with which a user's viewpoint can be freely set up in a reconstructed three-dimensional space of a target scene photographed by multi-view cameras. This technology allows us to capture and reproduce the real world as recorded. Once we capture the world in a digital form, we can modify it as augmented reality (i.e., placing virtual objects in the digitized real world). Unlike this concept, the augmented world allows us to see through real objects by synthesizing the backgrounds that cannot be observed in our raw perspective directly. The key idea is to generate the background image using multi-view cameras, observing the backgrounds at different positions and seamlessly overlaying the recovered image in our digitized perspective. In this paper, we review such desired view-generation techniques from the perspective of free-viewpoint image generation and discuss challenges and open problems through a case study of our implementations.

Keywords: Overview, Diminished reality, Augmented reality, Free-viewpoint image generation, Human–computer interaction

Received 7 February 2018; Revised 24 September 2018

I. INTRODUCTION

Once the real world is visually digitized via multi-view observations, the digitized world can be transferred, modified, and played on a computer from any viewpoint and time point [1, 2]. Such a basic idea is called free-viewpoint image generation, and it has been over two decades since it was proposed. Using this technology, it is no longer necessary to stay in a fixed viewpoint. With a near-eye display such as head-mounted displays (HMDs), we can observe the digitized world from our egocentric viewpoint as we see the world in our daily life. Today, real-world digitization technology is within the scope of real-time processing [3], and such interactive manipulation brings us a new visualization technique: augmented visualization.

Here, we define augmented visualization as computational techniques for visualizing what cannot be seen with raw image input. While free-viewpoint image generation changes the viewpoint in a recorded virtual scene, in augmented visualization, the digitized scene is registered to the real world in accordance with our viewpoint to the presently modified vision. For example, augmented reality (AR) is considered to be an augmented visualization technique since AR overlays computer graphics (CG) (possibly CG of digitized scenes) onto the real world to change the appearance of our field of view. With augmented visualization techniques, we can modify the way we see the world.

In this paper, we give an introductory review of such visualization technologies emerging from free-viewpoint image-generation techniques that go one step beyond conventional AR. Specifically, we focus on diminished reality (DR), which has been a missing part of augmented visualization due to the lack of image and computational resources. Here, we use a case study approach to examine an easy step-up from the well-known computer vision technique (i.e., free-viewpoint image generation) to augmented visualization.

First, we briefly review the history of free-viewpoint image capture and generation (Section II). Then, we point out that abundant multi-view resources have recently become available to the public that can potentially empower one of the key technologies of augmented visualization known as DR (Section III). After giving a brief summary of [4] to describe the basics of DR (Section IV), we introduce case studies in an attempt to give readers ideas regarding how to use such resources for DR and how DR changes the way of seeing the world (Section V). Within the section, we also explain open problems in multi-view-based DR and finally summarize the discussions. Due to this, which is contrary to a recently published DR survey [4] aimed at readers in AR communities, we believe that this paper

[1] Graz University of Technology, Rechbauerstraße 12, Graz, Austria
[2] Keio University, 3-14-1 Hiyoshi Kohoku-ku, Yokohama, Japan

Corresponding author:
Shohei Mori
Email: s.mori.jp@ieee.org

is more introductory but more convincing for readers in non-AR communities.

II. FREE-VIEWPOINT IMAGE/VIDEO

Measuring and reconstructing the three-dimensional (3D) shape of target scenes have been studied for nearly 50 years as a basic technology in the fields of image processing and measurement [5–7]. In the fields of CG and virtual reality (VR), such digitized 3D shapes have been used for taking pictures of scenes from arbitrary viewpoints where input from real-world cameras did not exist. In Virtualized Reality [1, 8], Kanade *et al.* demonstrated that 3D shape reconstruction techniques can be applied for dynamic scenes in time-space for generating free-viewpoint videos. This technology has been studied as a new type of image presentation technology and has been used in practical situations such as sports[1].

The essence of this technology lies in how accurately one can obtain the shape and texture (colors) of the target object. However, the accuracy of the abovementioned 3D shape reconstruction-based approaches is limited and detracts from the final output image quality. On the other hand, the ray space theory-based approaches [9] can generate an arbitrary viewpoint images without acquiring the 3D shape of the scene.

Ray space theory [9], light fields [10], and lumigraphs [11] represent an image as a group of light rays in a space filled with light rays coming from every direction and has been studied since it was first proposed 20 years ago. This method is implemented by photographing the target scene with a large number of cameras (a camera array), providing a variety of viewpoints [12–14]. Recently proposed methods using such image data structures can computationally change camera parameters, such as focus and aperture, *after* the shooting whereas previously we had to change the parameters *before* the shooting when using conventional cameras [15, 16]. Also, methods for photographing pictures in higher resolutions, in terms of space and time beyond the limitations determined by sampling theory, have been proposed [17, 18]. Thus, future developments in computational photography technology are expected.

III. FREE-VIEWPOINT TO CONTROLLABLE REALITY

The techniques in the previous section are used to collect image data photographed and acquired by sensors, such as cameras, or to generate desirable images for observers after the data acquisition. For example, when free-viewpoint video technology is used in broadcasting what *free* usually means is that the observer is free from the observation position that the photographer or the director of the broadcast had been predetermined [21–23]. Therefore, free-viewpoint videos allow the audience actually watching the video to move their viewing positions to where they desire (see Figs 1(a)–1(c)).

In other words, this is a technology that allows us to change our viewpoint in the digitized 3D world, regardless of the intention of the direct acquisition of the image data. The technique for such desirable images, regardless of the intention or situations at the time of actual image capture, becomes more meaningful as the amount of acquired image data increases. When the camera was invented, "taking a picture" required considering every property every time (camera position, posture, subject, shutter, etc.) since the cost required for one photograph was large. Therefore, even a single photo conveyed details inexpressible by words to many people.

In the modern worlds, plenty of surveillance cameras are installed in societies, and images are collected at every moment (e.g., "EarthCam"[2] and "CAM^2"[3]). Similarly, many images are taken by individuals with their personal cameras (e.g., Instagram[4]). However, most of the surveillance and personal camera images are not seen by humans. That is, a lot of images are stored that are not seen by anyone. Therefore, there is a demand for techniques to re-create and present images from these huge image datasets in desired forms [24–26]. As an example of such forms, the authors have been conducting research on image generation of DR [4], which is one of the missing components of augmented visualization. In contrast to AR, in DR such image resources are synthesized from the user's perspective in accordance with the user's position to virtually reveal the hidden parts of the scene (Fig 1(d)). As we explained in Section I, with a combination of AR and DR, we can *freely* modify the captured environment. In other words, combining AR and DR completes augmented visualization.

IV. DIMINISHED REALITY: PRINCIPLES

A) Comparing DRs

DR is different from AR [27–29], which superimposes virtual objects on the real world to enhance reality. AR overlays virtual objects to add positive information to the real world. DR overlays virtual objects as well, but the objects are negative information to offset the reality. Therefore, we could say that DR is a visualization method based on diminishing.

Figure 2 shows schematic figures describing differences in the real world, VR, AR, and DR in terms of visible light rays in each reality display. In (a), the observer sees real objects in the environment. In VR (b), HMD occludes light rays of the real environment and presents ones from the virtual world. In AR (c), real light rays are visible through the HMD and, at the same time, virtual rays are presented. Thus, the virtual object (black star) looks like it is existing in the real environment. In DR (d), real rays are

[1] https://youtu.be/Bse7YXWdP-c

[2] https://www.earthcam.com/
[3] https://cam2.ecn.purdue.edu/
[4] https://www.instagram.com/

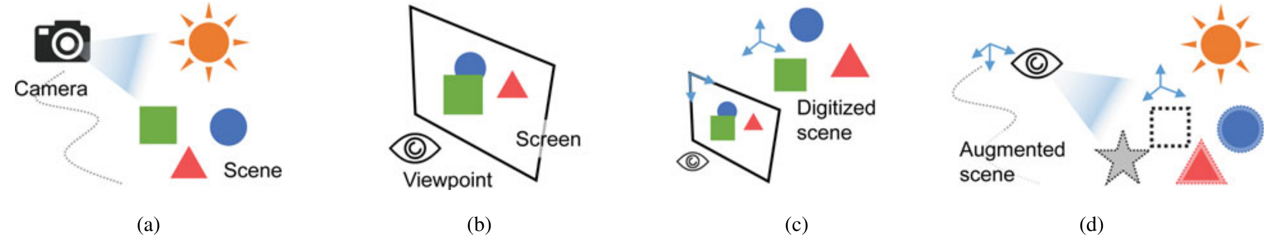

Fig. 1. Differences in conventional photos or videos, free-viewpoint image generation, and augmented visualization. After recording the scene (a), the observer's viewpoint is fixed in the conventional photos or videos (b). Free-viewpoint image generation enables the observer to move around in the recorded virtual space through a physically fixed display (c). In augmented visualization (d), with wearable displays such as near-eye displays [19, 20], the observers can move around in the real space to which the recorded virtual scene (objects with dotted lines) is registered as proxies for visual augmentation. In other words, we could achieve *free* scene modifications using both AR and DR technologies via the proxies (e.g., the star-shaped object is added and the box is removed, respectively, in this example).

selectively occluded by the HMD. Rays initially occluded by objects are recovered and therefore visible through the HMD. To achieve this, the digitized scene must be registered to the real one and tracked according to the observer's head motion. Here, DR requires free-viewpoint image-generation technology.

B) Background resources

Figure 3 shows typical settings in DR. DR methods assume that the background resources are obtained by (1) directly observing the background from different viewpoints (e.g., surveillance cameras and other users' cameras), (2) copying and pasting pixels of patches from the surrounding regions or previous frames, or (3) fetching images from databases such as Google Street View. (4) Of course, we can combine these approaches. Based on the available background resources, we can categorize DR methods [4]. Note that DR relies on free-viewpoint image-generation techniques when it employs these approaches (except the second one).

C) Functions

DR technology is defined as a set of methodologies (*diminishing*, *replacing*, *in-painting*, or making something *see-through*) for altering real objects in a perceived environment in real time to "diminish" reality [4]. Each function can be described as follows: to *diminish* objects, the objects of interest are degraded in colors or textures to get less attention; to make objects *see-through*, the backgrounds of the objects are digitized beforehand or in real time in a similar manner to free-viewpoint image generation and overlaid onto the observer's view in accordance with the observer's head motion; to *replace* objects, alternative virtual objects are overlaid onto the real objects to hide them; to *in-paint* objects, plausible background images are generated on the fly from all pixels except for those in the region of interest (ROI).

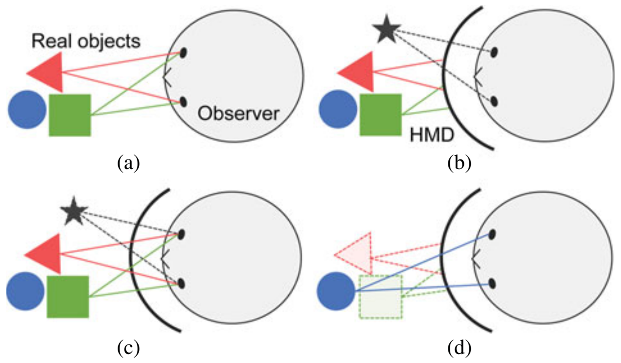

Fig. 2. Observable light rays in (a) the real space, (b) VR, (c) AR, and (d) DR. Light rays are selectively presented to the eyes via an HMD. In DR, occluded light rays (the blue rays) from the user's viewpoint must be observed and reproduced based on the other image resources. Here, we need free-viewpoint image-generation techniques.

Figures 4 and 5 show example results of *see-through* and *in-paint*, respectively. Note that *in-paint* does not require resources for different perspectives (review the second approach in Section IV B), and therefore, *in-paint* [30–32] does not visualize the "real" background. The resultant visualization is computationally generated, and there is no guarantee that the actual hidden areas will be visualized. On the other hand, *see-through* uses the free-viewpoint image-generation technology to visualize the hidden areas as they are [33–37].

D) Display devices

Display devices are one of the important factors in DR. Most DR researches implicitly assume video see-through systems since optical see-through devices cannot perfectly occlude incoming lights from the real environments due to their see-through nature [38–40].

Egocentric displays, such as HMDs, are ideal in terms of portable and hands-free experiences. Tablet systems

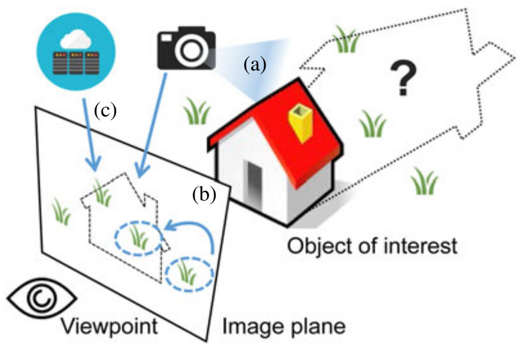

Fig. 3. Background resources in DR. The goal of DR is to estimate background from (a) different perspectives, (b) surrounding pixels, (c) database resources, and these combination.

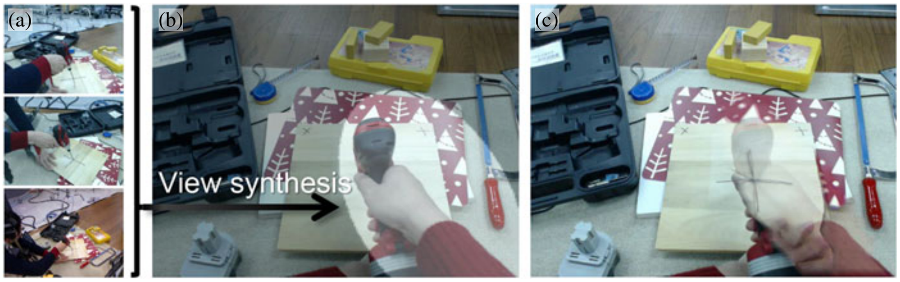

Fig. 4. *See-through* approach. Multi-view image resources (a), user view (b), and DR view (c).

can also be used as egocentric displays, although user-perspective distortion should be corrected to compensate for differences in viewpoint between the recording camera and the eyes [41–44]. Since the tablet surface looks transparent following distortion correction, user-perspective distortion correction is considered to be a DR application.

V. CASE STUDY OF FREE-VIEWPOINT IMAGE-GENERATION-BASED DR

A) Overview

In this section, we introduce four DR implementations as case studies where free-viewpoint image-generation techniques are necessary to address challenges in DR. Here, we show how augmented visualization changes the way of seeing the world and how free-viewpoint image-generation techniques contribute to augmented visualization. Keeping in mind the background of this paper, here we show *see-through* approaches only. We additionally discuss practical issues related to each case to highlight open problems in DR. Table 1 provides a summary of the case studies.

B) Filmmaking support [45]

1) Background and challenges

Special effects (SFX) or visual effects (VFX) help to make films popular, but the filmmaking process is complicated. Therefore, pre-visualized movies created in the early stage of the filmmaking, referred to as PreVis, have played an important role over the last several decades. PreVis movies are usually shot with low-cost cameras or generated using simple CG to share a creator's vision for directing staff and to estimate camera movements, necessary personnel, and other potential expenditures.

However, newly placed vending machines or modern signboards in the environment can ruin the image, especially when filmmakers create a historical movie. This is where we apply DR to remove such objects in PreVis. In this filmmaking scenario, we have numerous photos as resources recorded in advance in the location-hunting stage (e.g., several weeks ago). The scenes in such photos should be similar to the views to be filmed in PreVis, while the only problem is that the pictures were taken at different time points. In this example, we had to augment images in multiple monitors to share the visualizations among crews members. One had to be a portable monitor attached to a cinematographic camera for a cameraman. Fig. 6(a) illustrates this issue.

2) Example solution

Since the appearance of the photos is quite similar to the views in PreVis, we do not generate a new image from the photo resources, although we need to align the photos to the views in PreVis via homography warping. We can simply select and warp one of these images onto the video stream from a cinematography camera to hide the newly placed visually annoying object.

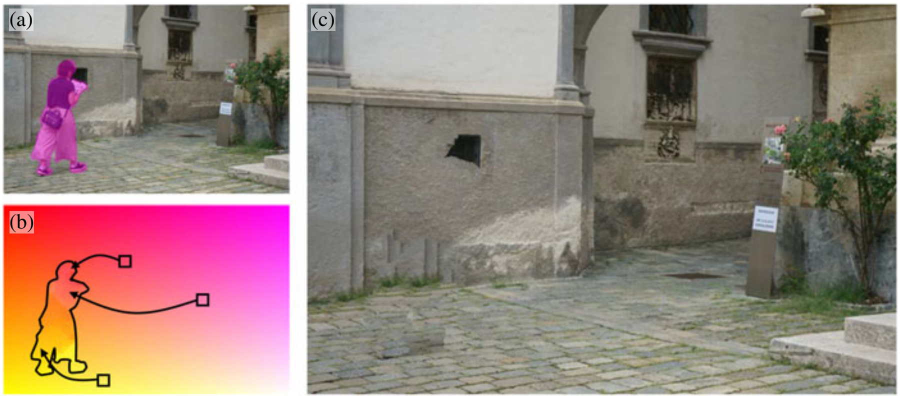

Fig. 5. *In-paint* approach. Input masked image (a), position map (b), and DR view (c). A position map indicates which pixels in the surrounding regions should be mapped to the region of interest. In this example, each color in the region of interest corresponds to pixels of the same color in the surrounding pixels.

Table 1. Summary of the case studies. These four examples are presented in Section V to discuss general issues in DR

Case	Challenge	Example solution Background resource	Background recovery	Potential display device
Filmmaking support [45]	Illumination changes	Pre-obtained image database	Homography warping	Portable monitors
Blind spot visualization [46]	Limited field of view	Wide range RGB-D sensor	Polygon mesh rendering	Smartphones/HUDs
Work area visualization [47]	Dynamic scene visualization	Arbitrary arranged cameras	Detour LFR	HMDs
De-fencing [48]	Fence region restoration	Sweep motion video	Fence-aware LFR	Smartphones

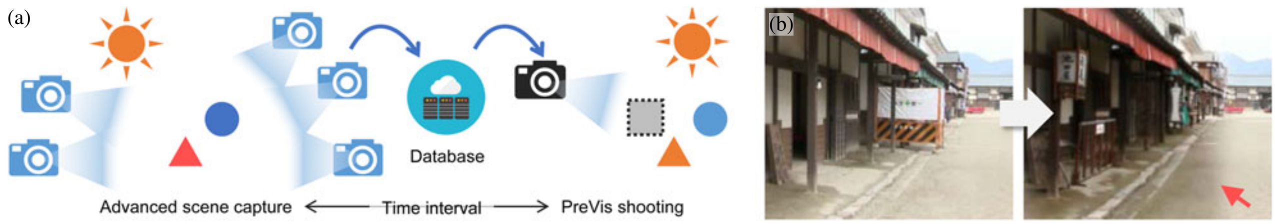

Fig. 6. Filmmaking support. (a) The ideal scene is recorded in advance to be used in the PreVis. The illumination changes and, consequently, the replaced scene appears inconsistent when seen through the object of interest (the square with dotted lines). (b) The construction signboard is replaced with a pre-recorded scene captured in different lighting. The red arrow indicates photometric borders of the ROI [45].

However, when the background image is synthesized to overlay the object, photometric inconsistency between the synthesized and the surrounding regions occurs. Since the illumination conditions are different from each other, the simple overlay generates conspicuous borders around the ROI, as shown in Fig. 6(b). In augmented visualization, keeping visual coherence between the real and synthesized images is one of the most important issues and is known as “photometric consistency”. These borders need to be reduced using a color compensation technique.

DR methods do not estimate illuminations directly on the site but instead use color tone correction processing on the image plane to achieve real-time processing [32, 49]. Even if the light source always moves around within a 3D scene, a simple color tone correction often results in sufficient results with regard to the appearance unless the processing is performed in real time [50].

3) Additional discussion: registration

Every DR methods begin with registering the digitized scene to the real scene (recall Section 3(d)). This filmmaking support scenario was the simplest case with regard to the registration, as explained above. When an object to be removed is small or far from the observer, 2D alignment may be sufficient even for 3D background scenes [49]. When the object to be removed is large or close to the observer (as in the case here), the background region occluded by the object that must be removed increases, and 3D registration is required.

Visual simultaneous localization and mapping (vSLAM) [51] is sufficient for the 3D alignment, but an extension for registration with the background’s and vSLAM’s coordinates is required. There is also an efficient 3D registration method that utilizes the previously [50] or currently obtained [46] textured 3D mesh of the background. Since the result of the synthesis on the screen space is the final output for DR, matching the generated background and the current image should be effective.

C) Blind spot visualization [46]

1) Background and challenges

Blind spot visualization techniques are known as AR X-ray vision [33, 52] or see-through vision [35]. These visualization techniques enable observers to see through walls like Superman, as their names suggest. Using this technology, for example, we can check the cars and pedestrians approaching from behind walls and can check whether the shops behind buildings are open or not. Although we can see behind walls using curved mirrors, see-through technology does not require us to mentally transform the view in 3D like we usually do. To provide such views, we need to show DR results in portable displays, such as a smartphone or a head-up display (HUD) in a car.

We can achieve the see-through ability by overlaying real-time video feeds from cameras placed behind walls onto the observer’s video. However, in this setting, the field of view of the background observation cameras is potentially limited because they are basically placed ahead of the user’s camera, as illustrated in Fig. 7(a). Conventional approaches use multiple surveillance cameras to cover the whole background, although this setup makes things difficult with regard to recording time synchronizations, color compensations, and calibrations between the installed cameras.

2) Example solution

To overcome this problem, as a practical solution, we proposed to build a wide field of view RGB-D camera combining a range scanner and an affordable and widely available fish-eye camera [46] (Figs 7(b) and 7(c)). The single wide field of view RGB-D camera straightforwardly mitigates the above-mentioned issues in blind spot visualization. Given such RGB-D images, we can create a wide range of textured meshes in real-time. This is because the reconstructed view is wide enough to have extra regions for the user’s camera

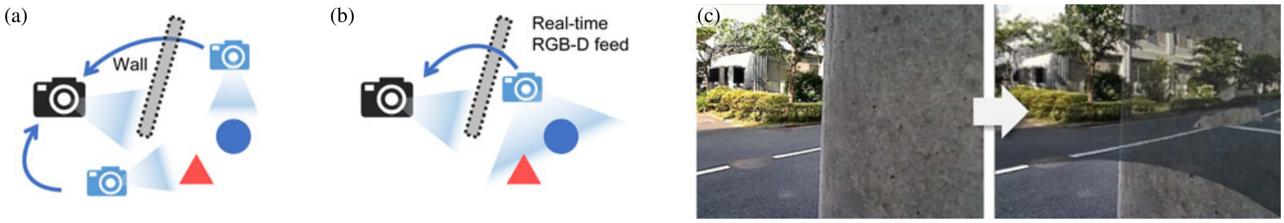

Fig. 7. Blind spot visualization. (a) Blind spot visualization potentially requires multiple cameras to cover a wide area. (b) A single wide field of view RGB-D camera makes the setting simpler and can mitigate a number of issues in (a). (c) Real-time blind spot visualization with the RGB-D camera [46].

to match and register the 3D textured mesh to the user's camera image (Fig 7(c)).

3) Additional discussion: depth perception

Given the user's camera pose, free-viewpoint image-generation techniques provide synthetic views based on the resources. However, simply overlaying the registered free-viewpoint image is not acceptable since all information about the front objects will be lost. A straightforward method to solve this problem is to alpha blend an original input image and a diminished reality image. However, the resulting image will have two types of information (i.e., background and foreground) at once. Therefore, effective representation methods of this information have been discussed in the AR X-ray vision area.

Considering DR as an information filtering method would also involve ethical problems. Therefore, it is a difficult problem to decide what to hide or show and whether to automatically control or trust the user. Although erasing a vehicle in front will give us a wide field of view, it is easy to imagine that we will collide with the invisible vehicle. The perceived appearance varies greatly depending on when the forward vehicle is displayed semi-transparently, when the focus is on the front vehicle, or when the focus is on the far road. Therefore, besides DR displays, taking into account human sensing technologies such as accommodation estimation will become important.

D) Work area visualization [47]

1) Background and challenges

Handcraft operators might see the work area when their move their heads, but they may have difficulty when the holding tool is large or when a third person watching the first person's view video cannot see the work area since he or she cannot move the viewing position. Using DR technology, the operator can change the transparency of hands and tools in the operator's view to check the working process. Here, all of the processes from video capture to visualization should run in interactive rates since the scene is completely dynamic. Such ergonomic views can be provided through HMDs.

2) Example solution

Figure 4 shows an example of visualizing a work area occluded by an operator's hand and a holding tool in the perspective [47, 53]. In this setting, we had no space to place cameras below the hand or the tool, and therefore, we surrounded the workspace to capture the whole area. This setting reproduces a virtual camera with a huge aperture (i.e., we record the light field in real time). During the rendering of the synthetic camera, we give less weight to light rays passing through the hand and the tools that are not to be reproduced in the rendering (Fig 8).

The rendering approach [53] named detour light field rendering (LFR) is computationally efficient since no precise ROI segmentation and tracking are required, which are usually computationally expensive. Instead, a point representing an object to be removed is detected or manually placed. This representative point determines an effective range and excludes light rays passing around the point from the rendering, as illustrated in Fig. 8. In other words, this rendering process mitigates segmentation processing, which is usually computationally expensive.

E) Additional discussion: effects of DR

Pseudo-haptics [54] is a famous perceptual concept suggesting that our haptic sensations are easily affected by a vision in VR. It is known that this perceptual illusion is also present in AR scenarios [55, 56]. In this context, what happens in DR scenarios when a user's hand or held object is removed from the user's perspective has not been well discussed. We obtained some user study results demonstrating that virtually shortened sticks can feel heavier than they actually are [57].

F) De-fencing [48]

1) Background and challenges

De-fencing refers to techniques to diminish the appearance of a fence in an image to create a fence-free view [48]. Such a technique is useful, for example, when a photographer takes a photo of a tourist landmark but the scene is occluded by fences for security reasons. The resultant de-fenced videos

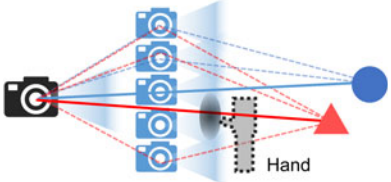

Fig. 8. Work area visualization. The rays are reproduced by the rays with dotted lines that ignore the hand [53].

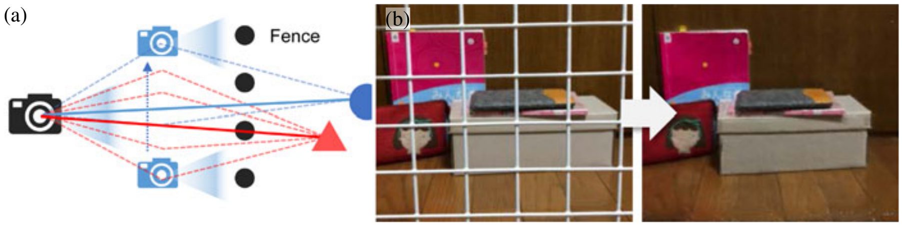

Fig. 9. De-fencing. (a) The rays with solid lines are reproduced by the rays with dotted lines that ignore the fence points. (b) Consequently, the scene without the fence is restored [48].

can be provided on a smartphone display through streaming or in a recorded video format. The challenge here is to segment the fences and ignore them in the free-viewpoint image generation.

2) Example solution

This task can be achieved following a similar approach to the light field rendering-based DR in Section V-D(Fig 9). Here, we set points at the fenceless weights to replace the fence pixels with the pixels in the other images recorded while sweeping a camera. However, in this case, we cannot represent the fence with several points since the fence pixels appear in the whole image. Instead, we segment a 3D scene reconstruction into fence and non-fence points since fence points always appear closer to the camera.

3) Additional discussion: object selection

Considering practical scenarios, the object detection task in DR seems rather challenging compared to those in defencing and work area visualization tasks, as the object to be removed is usually unknown until the observer faces it or it is determined based on the user's preference. Therefore, objects to be removed are selected through user interaction, such as clicking and dragging.

When selecting an object to be removed through user interaction [30, 32, 50], the region must be properly enclosed, and even after that, it must keep tracking along with the movement of the viewpoint. Many methods cover objects with a 3D bounding object [32, 50] or employ image-based tracking methods [30, 49]. One state-of-the-art method enables users to select objects through a category consisting of a combination of a convolutional neural network and vSLAM [58]. For example, the users can remove "trash" or "furniture" from the environment just by selecting the categories.

VI. Conclusions

In this paper, the authors discussed the features of arbitrary viewpoint image generation and extended idea to augmented visualization. All examples shown in Section V are based on or closely related to principles of the two approaches described in Section II (i.e., 3D shape reconstruction and ray space theory).

The essence of these approaches are identical in that their purpose is to describe the 3D space regardless of differences in data representation. In DR, the described 3D space is arbitrarily reproduced in accordance with the real space coordinates to see the scene in the desired way. As described in the case study in Section V, DR has special issues to be solved in terms of image processing, such as viewpoint changes with limited image and computational resources under variable lighting conditions. While some DR methods achieved high-quality object removal results, researchers do not know how the results are perceived. The authors feel that these perceptual issues need further discussion in the near future.

The authors would like to introduce the first international DR survey paper for readers who discovered DR technology through this paper [4]. The authors are the founding members of the "Technical Committee on Plenoptic Time-Space Technology (PoTS)[5]", founded in Japan in 2015. PoTS started to investigate a way of presentation via spatio-time representation based on the idea of extending ray space theory to the time axis. Through such research activities, the authors hope to further advance the technology.

Acknowledgements

The authors are grateful to M. Maezawa, K. Oishi, and C. Lueangwattana at Keio University, L. Jinxia at Ritsumeikan University, and M. Isogawa at Osaka University for their technical assistance with the figures.

Financial support

The images presented in this paper are from works supported in part by the Japan Society for the Promotion of Science (JSPS) KAKENHI Grant Numbers JP17K12729 (Grant-in-Aid for Young Scientists (B)), JP13J09193 (Grant-in-Aid for JSPS Research Fellow (DC-1)), JP16J05114 (Grant-in-Aid for JSPS Research Fellow (PD)), and JP24220004 (Scientific Research (S)).

Statement of interest

The authors declare that they have no competing interests.

[5] http://www.hvrl.ics.keio.ac.jp/pots/index_eng.html

REFERENCES

[1] Kanade, T.; Narayanan, P.J.; Rander, P.W.: Virtualized reality: Concepts and early results, in *Proc. Workshop on Representation of Visual Scenes*, 1995.

[2] Matusik, W.; Buehler, C.; Raskar, R.; Gortler, S.J.; McMillan, L.: Image-based visual hulls, in *Proc. SIGGRAPH*, 2000, 369–374.

[3] Orts-Escolano, S. *et al.*: Holoportation: Virtual 3D teleportation in real-time, in *Proc. SIGGRAPH*, 2016, 741–754.

[4] Mori, S.; Ikeda, S.; Saito, H.: A survey of diminished reality: Techniques for visually concealing, eliminating, and seeing through real objects. *IPSJ Trans. Comput. Vision Appl. (CVA)*, **9**(7) (2017), 1–14. DOI: 10.1186/s41074-017-0028-1.

[5] Henderson, R.L.; Miller, W.J.; Grosch, C.B.: Automatic stereo reconstruction of man-made targets, in *Proc. SPIE 0186, Digital Processing of Aerial Images*, 1979.

[6] Horn, B.K.P.: Shape from shading: A method for obtaining the shape of a smooth opaque object from one view. Technical Report, Massachusetts Institute of Technology Cambridge, MA, USA, 1970.

[7] Shirai, Y.: Recognition of polyhedra with a range finder. *Pattern. Recognit.*, **4**(3) (1972), 243–250.

[8] Kanade, T.; Saito, H.; Vedula, S.: The 3D room: Digitizing time-varying 3D events by synchronized multiple video streams, Carnegie Mellon University, The Robotics Institute, Pittsburgh, PA, USA, CMU-RI-TR-98-34, 1998. http://www.cs.cmu.edu/∼VirtualizedR/3DRoom/CMU-RI-TR-98-34.pdf.gz.

[9] Fujii, T.; Harashima, H.: Coding of an autostereoscopic 3-D image sequence. *Proc. SPIE, Vis. Commun. Image*, **2308** (1994), 930–941.

[10] Levoy, M.; Hanrahan, P.: Light field rendering, in *Proc. Annual Conf. on Computer Graphics and Interactive Techniques (SIGGRAPH)*, 1996, 31–42.

[11] Gortler, S.J.; Grzeszczuk, R.; Szeliski, R.; Cohen, M.F.: The Lumigraph, in *Proc. Annual Conf. on Computer Graphics and Interactive Techniques (SIGGRAPH)*, 1996, 43–52.

[12] Buehler, C.; Bosse, M.; McMillan, L.; Gortler, S.; Cohen, M.: Unstructured lumigraph rendering, in *Proc. Annual Conf. on Computer Graphics and Interactive Techniques (SIGGRAPH)*, 2001, 425–432.

[13] Davis, A.; Levoy, M; Durand, F.: Unstructured light fields. *Comput. Graphics Forum*, **31**(2), Pt. 1, (2012), 305–314.

[14] Levoy, M.: Light fields and computational imaging. *Computer*, **39**(8) (2006), 46–55.

[15] Isaksen, A.; McMillan, L.; Gortler, S.J.: Dynamically reparameterized light fields, in *Proc. Annual Conf. on Computer Graphics and Interactive Techniques (SIGGRAPH)*, 2000, 297–306.

[16] Ng, R.; Levoy, M.; Brédif, M.; Duval, G.; Horowitz, M.; Hanrahan, P.: Light field photography with a hand-held plenoptic camera. *Comput. Sci. Tech. Report CSTR*, **2**(11) (2005), 1–11.

[17] Fujii, T.; Tanimoto, M.: Free viewpoint TV system based on ray-space representation. *Proc. SPIC, Three-Dimensional TV, Video, Display*, **4864** (2002), 175–190.

[18] Wilburn, B.: *et al.* High performance imaging using large camera arrays. *In ACM Trans. Graphics (TOG)*, **24**(3) (2005), 765–776.

[19] Huang, F.; Chen, K.; Wetzstein, G.: The light field stereoscope: Immersive computer graphics via factored near-eye light field displays with focus cues. *ACM Trans. Graphics*, **34**(4) (2015), 1–12. https://dl.acm.org/citation.cfm?id=2766922.

[20] Lanman, D.; Luebke, D.: Near-eye light field displays. *ACM Trans. Graphics*, **32**(6) (2013), 1–10. https://dl.acm.org/citation.cfm?id=2508366.

[21] Lee, C.-C.; Tabatabai, A.; Tashiro, K.: Free viewpoint video (FVV) survey and future research direction. *APSIPA Trans. Signal Inf. Process.*, **4** (2015), 1–10. https://www.cambridge.org/core/services/aop-cambridge-core/content/view/0E5F6708FD61193F78CF2BD3D6A58024/S2048770315000189a.pdf/free_viewpoint_video_fvv_survey_and_future_research_direction.pdf.

[22] Shum, H.Y.; Kang, S.B.; Chan, S.C.: Survey of image-based representations and compression techniques. *IEEE Trans. Circuits Syst. Video Technol. (TCSVT)*, **13**(11) (2003), 1020–1037.

[23] Smolic, A.: 3D video and free viewpoint video – from capture to display. *J. Pattern Recognit.*, **44**(9) (2011), 1958–1968.

[24] Kaseb, A.S.; Koh, Y.; Berry, E.; McNulty, K.; Lu, Y.-H.; Delp, E.J.: Multimedia content creation using global network cameras: The making of CAM2, in *Proc. Global Conf. Signal and Information Processing (GlobalSIP)*, 2015, 15–18.

[25] Su, W.-T.; McNulty, K.; Lu, Y.-H: Teaching large-scale image processing over worldwide network cameras, in *Proc. Int. Conf. on Digital Signal Processing (DSP)*, 2015, 726–729.

[26] Su, W.-T.; Lu, Y.-H.; Kaseb, A.S.: Harvest the information from multimedia big data in global camera networks, in *Proc. Int. Conf. on Multimedia Big Data (BigMM)*, 2015, 184–191.

[27] Azuma, R.T.: A survey of augmented reality. *Presence: Teleoperators Virtual Environ.*, **6**(4) (1997), 355–385.

[28] Azuma, R.T.: Recent advances in augmented reality. *IEEE Comput. Graphics Appl.*, **22** (2001), 34–47.

[29] Kruijff, E.; Swan, J.E.; Feiner, S.: Perceptual issues in augmented reality revisited, in *Proc. Int. Symp. on Mixed and Augmented Reality (ISMAR)*, 2010, 3–12.

[30] Herling, J.; Broll, W.: High-quality real-time video inpainting with PixMix. *IEEE Trans. Visual. Comput. Graphics (TVCG)*, **20**(6) (2014), 866–879.

[31] Iizuka, S.; Simo-Serra, E.; Ishikawa, H.: Globally and locally consistent image completion. *ACM Trans. Graphics (TOG)*, **36**(4) (2017), 107:1–107:14.

[32] Kawai, N.; Yamasaki, M.; Sato, T.; Yokoya, N.: Diminished reality for AR marker hiding based on image inpainting with reaction of luminance changes. *ITE Trans. Media Technol. Appl.*, **1**(4) (2013), 343–353.

[33] Avery, B.; Sandor, C.; Thomas, B.H.: Improving spatial perception for augmented reality x-ray vision, in *Proc. IEEE VR*, 2009, 79–82.

[34] Cosco, F.; Garre, C.; Bruno, F.; Muzzupappa, M.; Otaduy, M.A.: Visuo-haptic mixed reality with unobstructed tool-hand integration. *IEEE Trans. Visual. Comput. Graphics (TVCG)*, **19**(1) (2013), 159–172.

[35] Tsuda, T.; Yamamoto, H.; Kameda, Y.; Ohta, Y.: Visualization methods for outdoor see-through vision. *IEICE Trans. Inf. Syst.*, **E89-D**(6) (2006), 1781–1789.

[36] Zokai, S.; Esteve, J.; Genc, Y.; Navab, N.: Multiview paraperspective projection model for diminished reality, in *Proc. Int. Symp. on Mixed and Augmented Reality*, 2003, 217–226.

[37] Enomoto, A.; Saito, H.: Diminished reality using multiple handheld cameras, in *Proc. Asian Conf. on Computer Vision (ACCV)*, 2007, 130–150.

[38] Hiroi, Y.; Itoh, Y.; Hamasaki, T.; Sugimoto, M.: AdaptiVisor: assisting eye adaptation via occlusive optical see-through head-mounted displays, in *Proc. Augmented Human Int. Conf.*, 2017, 9:1–9:9.

[39] Itoh, Y.; Hamasaki, T.; Sugimoto, M.: Occlusion leak compensation for optical see-through displays using a single-layer transmissive spatial light modulator. *IEEE Trans. Visual. Comput. Graphics (TVCG)*, **23**(11) (2017), 2463–2473.

[40] Kiyokawa, K.; Billinghurst, M.; Campbell, B.; Woods, E.: An occlusion-capable optical see-through head mount display for supporting co-located collaboration, in *Proc. Int. Symp. on Mixed and Augmented Reality (ISMAR)*, 2003, 133–141.

[41] Hill, A.; Schiefer, J.; Wilson, J.; Davidson, B.; Gandy, M.; MacIntyre, B.: Virtual transparency: Introducing parallax view into video see-through AR, in *Proc. Int. Symp. on Mixed and Augmented Reality (ISMAR)*, 2011, 239–240.

[42] Schöps, T.; Oswald, M.R.; Speciale, P.; Yang, S.; Pollefeys, M.: Real-time view correction for mobile devices. *IEEE Trans. on Visualization and Computer Graphic*s, **23**(11) (2017), 2455–2462.

[43] Tomioka, M.; Ikeda, S.; Sato, K.: Approximated user-perspective rendering in tablet-based augmented reality, in *Proc. Int. Symp. on Mixed and Augmented Reality (ISMAR)*, 2013, 21–28.

[44] Tomioka, M.; Ikeda, S.; Sato, K.: Pseudo-transparent tablet based on 3D feature tracking, in *Proc. Augmented Human Int. Conf. (AH)*, 2014, 52:1–52:2.

[45] Li, J., *et al.*: Re-design and implementation of MR-based filmmaking system by adding diminished reality functions. *Trans. Virtual Soc. Japan*, **21**(3) (2016), 451–462, (in Japanese).

[46] Oishi, K.; Mori, S.; Saito, H.: An Instant See-Through Vision System Using a Wide Field-of-View Camera and a 3D-Lidar, in *Proc. Int. Symp. on Mixed and Augmented Reality (ISMAR)-Adjunct*, 2017, 344–347.

[47] Mori, S.; Maezawa, M.; Saito, H.: A work area visualization by multi-view camera-based diminished reality. *Trans. MDPI Multimodal Technol. Interact.*, **1**(3), No. 18, (2017), 1–12.

[48] Lueangwattana, C.; Mori, S.; Saito, H.: Diminishing fence from sweep image sequences using structure from motion and light field rendering, in *Proc. Asia Pacific Workshop on Mixed and Augmented Reality (APMAR)*, 2018.

[49] Li, Z.; Wang, Y.; Guo, J.; Cheong, L.-F.; Zhou, S.Z.: Diminished reality using appearance and 3D geometry of Internet photo collections, in *Proc. Int. Symp. on Mixed and Augmented Reality (ISMAR)*, 2013, 11–19.

[50] Mori, S.; Shibata, F.; Kimura, A.; Tamura, H.: Efficient use of textured 3D model for pre-observation-based diminished reality, in *Proc. Int. Workshop on Diminished Reality as Challenging Issue in Mixed and Augmented Reality (IWDR)*, 2015, 32–39.

[51] Marchand, E.; Uchiyama, H.; Spindler, F.: Pose estimation for augmented reality: a hands-on survey. *IEEE Trans. Visual. Comput. Graphics (TVCG)*, **22**(12) (2015), 2633–2651.

[52] Santos, M.; Souza, I.; Yamamoto, G.; Taketomi, T.; Sandor, C.; Kato, H.: Exploring legibility of augmented reality x-ray. *Multimed. Tools. Appl.*, **75**(16) (2015), 9563–9585.

[53] Mori, S.; Maezawa, M.; Ienaga, N.; Saito, H.: Detour light field rendering for diminished reality using unstructured multiple views. Ditto, 2016, 292–293.

[54] Lecuyer, A.; Coquillart, S.; Kheddar, A.; Richard, P.; Coiffet, P.: Pseudo-haptic feedback: can isometric input devices simulate force feedback, in *Proc. IEEE Virtual Reality?*, 2000, 83–90.

[55] Hachisu, T.; Cirio, G.; Marchal, M.; Lécuyer, A.; Kajimoto, H.: Pseudo-Haptic Feedback Augmented with Visual and Tactile Vibrations, in *Proc. Int. Symp. VR Innovation (ISVRI)*, 2011, 327–328.

[56] Punpongsanon, P.; Iwai, D.; Sato, K.: SoftAR: visually manipulating haptic softness perception in spatial augmented reality. *IEEE Trans. Visual. Comput. Graphics (TVCG)*, **21**(11) (2015), 1279–1288.

[57] Tanaka, M. *et al.* : Further experiments and considerations on weight perception caused by visual diminishing of real objects, in *Proc. Int. Symp. on Mixed and Augmented Reality (ISMAR)-Adjunct*, 2017, 160–161.

[58] Nakajima, Y.; Mori, S.; Saito, H.: Semantic Object Selection and Detection for Diminished Reality based on SLAM with Viewpoint Class, in *Proc. Int. Symp. on Mixed and Augmented Reality (ISMAR)-Adjunct*, 2017, 338–342.

Shohei Mori received his B.S., M.S., and Ph.D. degrees in engineering from Ritsumeikan University, Japan, in 2011, 2013, and 2016, respectively. He was part of the JSPS Research Fellowship for Young Scientists (DC-1) at Ritsumeikan University and PD at Keio University until 2016 and 2018, respectively. He is currently working as a University project assistant at Graz University of Technology.

Hideo Saito received his Ph.D. degree in electrical engineering from Keio University, Japan, in 1992. Since then, he has been part of the Faculty of Science and Technology, Keio University. From 1997 to 1999, he joined the Virtualized Reality Project in the Robotics Institute, Carnegie Mellon University as a visiting researcher. Since 2006, he has been a full professor in the Department of Information and Computer Science, Keio University.