Augmented and Diminished Reality: Computational Imaging of Existence and Non-Existence

Shohei Mori

Keio University, : 3-14-1 Hiyoshi Kohoku-ku, Yokohama 223-8522, Japan

Keywords: Diminished Reality, Augmented Reality, Free-Viewpoint Image Generation, Technical Summary

ABSTRACT

This article presents a technical summary of a research area called diminished reality (DR). DR is described from its principle to open problems, with a comparison with its opposite concept known as augmented reality to highlight their differences.

1. INTRODUCTION

Once the real world is recorded in a digitized format [1, 2, 3], the digitized scene can be transferred, played, rewinded, and modified to present people with new experiences. The technologies that enables such are well studied in a research field called virtual reality (VR). While VR encloses its users in a fully digitized environment, augmented reality (AR) brings the virtual world partially to the real world to enhance real world-based experiences [4, 5, 6]. Augmenting real scenes with digital objects will support future technologies, such as traveling with direct 3D annotations, simulation of in-house lighting and furniture arrangement, virtual teleportation, virtual TV of any size anywhere in one's home, and location-based video games in which creatures appear in the real world.

However, one might notice that only adding information does not lead to true real-world augmentation. In practical scenarios, for example, virtual annotations can conflict with the real ones such as signboards, an augmented curtain does not shut off the real lights, and virtual creatures do not break real objects. Ideally, unnecessary information should be removed from the environment. To achieve this concept in reality, technologies called diminished reality (DR) have been studied [7, 8]. In this article, basic and advanced topics are summarized to compare DR with AR and highlight their differences.

2. REPRESENTING INVISIBILITY

2.1 Principles and Basic Functions

DR is different from AR that superimposes virtual objects on the real world to enhance reality. The basic idea in the virtual removal of real objects is to overlay the real scene with background information. Fig. 1 shows the distinctive differences between AR and DR. AR overlays virtual objects to add some positive information to the real world. DR also overlays virtual objects, but these objects are negative information (i.e., the background information) to diminish the reality. Therefore, we could say that DR is a visualization method by diminishing.

Fig. 2 shows schematic images that describe the differences of the real world, VR, AR, and DR in terms of the visible light rays in each reality. In (a), the observer sees real objects in the environment. In VR (b), the head-mounted display (HMD) occludes the light rays of the real environment and presents the lights from the virtual world. In AR (c), the real light rays are visible through the HMD, and virtual rays are also presented. Thus, the virtual object (black star) looks like it is existing in the real environment. In DR (d), real rays are selectively occluded by the HMD. Rays initially occluded by objects are recovered and therefore visible through the HMD.

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. Each function can be described as follows: To *diminish* objects, the objects of interest are degraded in colors or textures so that they receive less attention. To make objects *see-through*, the backgrounds of objects are digitized beforehand or in real time in a manner similar to that in free-viewpoint image generation [9], 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 from the pixels except for region of interest pixels on the fly.

2.2 Head-Mounted Displays

Once the real world is digitized via multi-view observations, the digitized world can be transferred, modified, and played on a computer from any viewpoint and time points [10]. This basic idea is called free-/arbitrary-viewpoint image generation, which was proposed 20 years ago. With this technology, staying in a 2D display with a fixed viewpoint is no longer necessary. As typified by head-mounted displays (HMDs), displays are now portable and therefore arbitrary view images are presented in front of the eyes. Similar to the processes in AR, combined with head pose estimation, the virtual view is presented with the head motion in real time.

3. TECHNICAL ISSUES IN DIMINISHED REALITY

3.1 Dealing with Viewpoint Changes

In DR, an arbitrary viewpoint image generated from

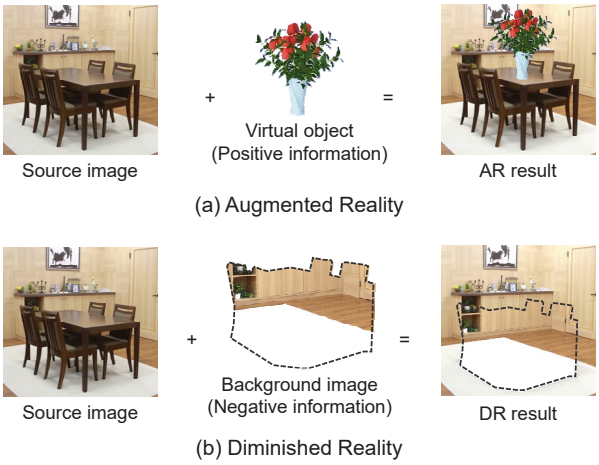


Fig. 1 Conceptual differences in AR and DR
AR overlays arbitrary virtual objects onto the real scene whereas DR creates virtual objects in the background by using free-viewpoint image generation.

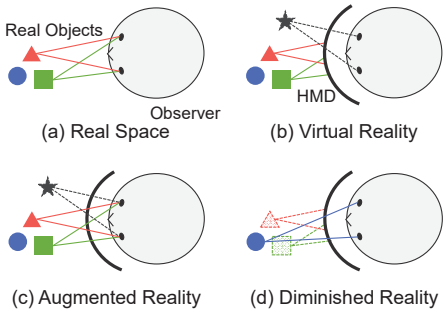


Fig. 2 Differences of the real world, VR, AR, and DR
Light rays are selectively presented to the eyes via an HMD to the user

other viewpoints should overlay only regions where the objects of interest exist. Therefore, when an object to be removed is small or far from the observer, 2D alignment may be sufficient even for 3D background scenes [11]. When the object to be removed is large or close to the observer, the background region occluded by the object to be removed increases, and 3D registration is required. Visual simultaneous localization and mapping (vSLAM) [12] seems sufficient for 3D alignment, but an extension for registration with the coordinates of the background and the vSLAM is required. There is also an efficient 3D registration method that utilizes the previously obtained background data [13]. As the result of synthesis on the screen space is the final output for DR, matching the generated background and the current image should be effective, even if 3D registration is erratic.

3.2 Handling Illumination Changes

In the case of using image data taken in the past, reducing the photometric inconsistency between the background and the current image is necessary because they are imaged at different times. DR methods basically use color tone correction processing on the image plane to achieve real-time processing [3, 11]. Even if the light source always moves around within a 3D scene, a simple color tone correction often results in sufficient results in the image appearance, if the processing is performed in real time [13].

3.3 Recognizing Objects to be Diminished

Objects to be removed are determined in advance or are selected via a user interface, such as a click or a drag. In the case of determining in advance the objects of interest, these objects are limited to specific types, such as walls and cars, or object detection is avoided for simplification. In the selection of an object to be removed with a user interface, the region must be properly enclosed, and even after this, the DR system must keep tracking along with the movement of the viewpoint. Many methods cover objects with a 3D bounding object or track by using image-based tracking methods. One state-of-the-art method enables users to select objects through a category empowered by a combination of a convolutional neural network and vSLAM [14]. For example, the users can remove “trash” or “furniture” from the environment just by selecting the categories (Fig. 3).

4. OPEN PROBLEMS IN DIMINISHED REALITY

4.1 When to Observe Backgrounds

Whether background data should be collected beforehand or in real time depends on the applications and, therefore, is difficult to uniquely determine. Sufficient time and effort before DR processing allow us the generation of high-quality background images that endure observations from any viewing positions. On the other hand, the object arrangement in the scene has evidently changed from that in the pre-observation [13]. Even if the placement of objects is unchanged, illumination will change from moment to moment depending on the weather. To cope with such changes, processing on the image space is used exclusively, and exaggerated models, such as light source information and reflection models of objects, are not used.

The biggest merit of observing the background in real time is that the information changing in real time is included in the background image [15] (Fig. 4). With no difference in the optical system of cameras, the above-mentioned optical inconsistency is often eliminated at the stage when the background is synthesized. However, multi-viewpoint observation in real time makes the system cumbersome, and, it also requires computationally expensive processing. In principle, the background observation viewpoint is located away from the user's viewpoint, so the quality of the synthesis result tends to be lower than that of the method based on the pre-observation.

4.2 What is the Ground Truth Background?

Making the ground truth for DR methods is not an easy task. Similar to free-viewpoint image generation

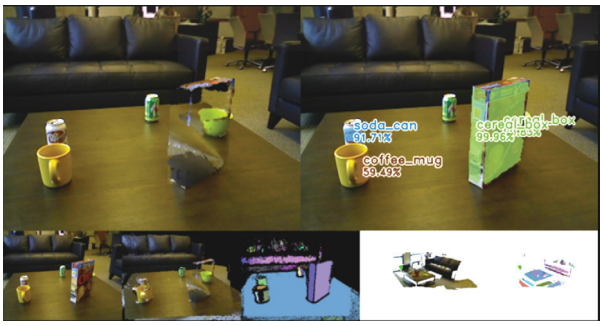


Fig. 3 Category-based object selection in DR
Object selection is one of the important issues in DR. This method selects objects by category [14].

methods, improving the accuracy of the synthesis result is the basic research motivation. However, how much accuracy is adequate for DR? We can calculate errors by comparing a DR-processed image sequence with the corresponding image sequence without a real object to be removed [16] (Fig. 5). However, one must consider when the DR processing result cannot necessarily be defined, or when it is not necessarily a true result. For example, in the case of deleting a manhole on a road, the required quality depends on the purpose of the DR process. Whether the pipe system under the manhole should be visualized or whether asphalt without manholes should be reproduced is unclear to us. Therefore, in this case, performing a user study to evaluate whether the implemented DR method was able to output a visually convincing result is needed.

4.3 How to Obtain Image Resources

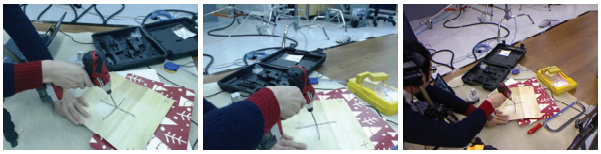
Particular indications of how many photos should be taken will not be clearly presented. In many cases, we will need external resources other than those we have collected ourselves, for example, to save time, human resources, and storage. So far, the Internet [11], surveillance cameras [17], and other users' cameras [18] are listed as such resource candidates, but these methods have not yet been implemented in practical applications.

4.4 How DR Visualizations Affect Our Perception?

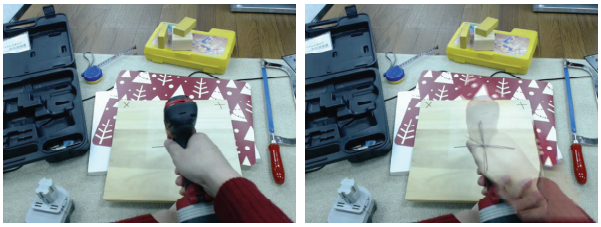
Pseudo haptics [19] is a famous perceptual concept suggesting that our haptic sensations are easily affected by vision in virtual reality. This perceptual illusion is also present in AR scenarios. In this context, what happens in DR scenarios in which a user's hand or objects held are removed in the user's perspective? We obtained some user study results demonstrating that virtually shortened sticks can feel heavier than they actually are. The figure in Fig. 6 shows sticks of the same density, length, and diameter. The three figures show diminished sticks. The participants wielded the same sticks but felt a distinct difference in terms of the weight of these sticks [20].

5. CONCLUSION

In this article, the author presented an introductory overview and the key features of diminished reality by comparing AR with DR. In DR, the described 3D space is



(a) Input frames



(b) User's View (c) User's View w/ DR

Fig. 4 Example of video-stream-based DR results
Removing hands and tools visualizes a cross mark behind them [15].

Fig. 5 Input and ground truth images in DR
Geometry and lighting, except those of the target object must be consistent [16].

arbitrarily reproduced in accordance with the real space coordinates to see the scene in a desired way. As described in Section 3 and 4, DR has special issues that need to be resolved in terms of image processing. While some DR methods achieved high-quality object removal results, researchers need to investigate how the results of DR visual effects influence viewers in terms of other future modalities.

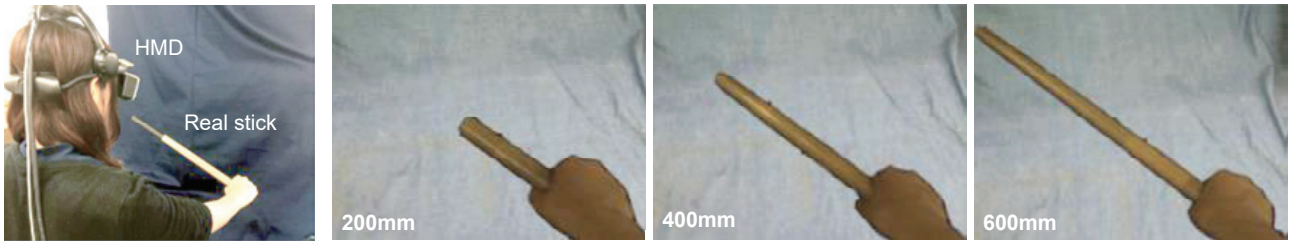
ACKNOWLEDGEMENTS

The work presented in this manuscript is supported in part by 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)).

REFERENCES

[1] S. Orts-Escolano, C. Rhemann, S. Fanello, W. Chang, A. Kowdle, Y. Degtyarev, D. Kim, P. L. Davidson, S. Khamis, M. Dou, V. Tankovich, C. Loop, Q. Cai, P. Chou, S. Mannicken, J. Valentin, V. Pradeep, S. Wang, S. B. Kang, P. Kohli, Y. Lutchyn, C. Keskin, and S. Izadi, "Holoportation: Virtual 3D teleportation in real-time," Proc. SIGGRAPH, pp. 741-754 (2016).

[2] L. Itti, C. Koch, and E. Niebur, "A model of



(a) Setups (b) Diminished sticks

Fig. 6 Wielding a diminished stick
Our user study showed that the shorter was the diminished stick to be wielded, the heavier the stick felt [20].

saliency-based visual attention for rapid scene analysis," IEEE Trans. on Pattern Analysis and Machine Intelligence (TPAMI), Vol. 20, No. 11, pp. 1254-1259 (1998).

[3] N. Kawai, T. Sato, and N. Yokoya: "Diminished reality based on image inpainting considering background geometry", IEEE Trans. on Visualization and Computer Graphics (TVCG), Vol. 22, No. 3, pp. 1236-1247, (2016).

[4] R. T. Azuma, "A survey of augmented reality," Presence: Teleoperators and Virtual Environments, Vol. 6, No. 4, pp. 355-385 (1997).

[5] R. T. Azuma, "Recent advances in augmented reality," IEEE Computer Graphics and Application, Vol. 22, No. 6, pp. 34-47 (2001).

[6] E. Kruijff, J. E. Swan, and S. Feiner, "Perceptual issues in augmented reality revisited," Proc. Int. Symp. on Mixed and Augmented Reality (ISMAR), pp. 3-12 (2010).

[7] S. Mann, "Mediated reality with implementations for everyday life," Presence Connect, 2002.

[8] S. Mori, S. Ikeda and H. Saito, "A survey of diminished reality: Techniques for visually concealing, eliminating, and seeing through real objects," IPSJ Trans. on Computer Vision and Applications (CVA), Vol. 9, No. 17, pp. 1-14 (2017), DOI: 10.1186/s41074-017-0028-1.

[9] T. Kanade, P. J. Narayanan, and P. W. Rander, "Virtualized reality: Concepts and early results," Proc. Workshop on Representation of Visual Scenes, pp. 69-76 (1995).

[10] W. Matusik, C. Buehler, R. Raskar, S. J. Gortler, and L. McMillan, "Image-based visual hulls," Proc. SIGGRAPH, pp. 369-374 (2000).

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

[12] E. Marchand, H. Uchiyama, F. Spindler, "Pose estimation for augmented reality: A hands-on survey," IEEE Trans. on Visualization and Computer Graphics, Vol. 22, No. 12, pp. 2633-2651 (2015).

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

[14] Y. Nakajima, S. Mori, and H. Saito, "Semantic object selection and detection for diminished reality based on SLAM with viewpoint class," Proc. Highly Diverse Cameras and Displays (HDCD4MAR), IEEE ISMAR-adjunct (2017).

[15] S. Mori, M. Maezawa and H. Saito, "A work area visualization by multi-view camera-based diminished reality," Trans. on MDPI Multimodal Technologies and Interaction, Vol. 1, No. 18, pp. 1-12 (2017), DOI:10.3390/mti1030018.

[16] S. Mori, Y. Eguchi, S. Ikeda, F. Shibata, A. Kimura, and H. Tamura, "Design and construction of data acquisition facilities for diminished reality research," ITE Trans. on Media Technology and Application (MTA), Vol. 4, No. 3, pp. 259-268 (2016).

[17] P. Barnum, Y. Sheikh, A. Datta, and T. Kanade, "Dynamic seethroughs: Synthesizing hidden views of moving objects," Proc. Int. Symp. on Mixed and Augmented Reality (ISMAR), pp. 111-114 (2009).

[18] A. Enomoto and H. Saito, "Diminished reality using multiple handheld cameras," Proc. Asian Conf. on Computer Vision (ACCV), pp. 130-150 (2007).

[19] A. Lecuyer, S. Coquillart, A. Kheddar, P. Richard, and P. Coiffet, "Pseudo-haptic feedback: Can isometric input devices simulate force feedback?," Proc. IEEE Virtual Reality, pp. 83-90 (2000).

[20] M. Tanaka, A. Misra, K. Oshima, S. Hashiguchi, S. Mori, F. Shibata, A. Kimura, and H. Tamura, "Further experiments and considerations on weight perception caused by visual diminishing of real objects," Proc. IEEE Int. Symp. on Mixed and Augmented Reality (ISMAR)-adjunct (2017).