Paper

Design and Construction of Data Acquisition Facilities for Diminished Reality Research

Shohei Mori[†1], Yuya Eguchi[†2], Sei Ikeda[†3], Fumihisa Shibata[†3], Asako Kimura[†3] and Hideyuki Tamura[†3] (member)

Abstract   Diminished reality (DR) refers to interactive techniques for deleting or diminishing undesirable objects from a perceived environment, whereas augmented/mixed reality seamlessly merges a real and virtual scene. In this paper, we introduce data acquisition facilities and evaluation workflow towards DR method benchmark. In the proposed data acquisition facilities, simulated indoor and outdoor scenes are constructed, illumed, and photographed, using full-scale and miniature sets, a cinematography-based lighting system, and camera attached 6 degrees of freedom industrial robot arm respectively. Consequently, it facilitates acquisition of paired image sequence with and without target objects of interest, i.e., source and ground truth image sequences, to evaluate DR methods in indoor and outdoor scenarios. Through operations test, several datasets are recorded and test benches of DR methods are evaluated using the dataset to show that such data is usable for qualitative and quantitative evaluation of DR methods.

Keywords: Diminished Reality, Ground Truth Acquisition, Benchmark.

1. Introduction

Diminished reality (DR) is a systemized set of techniques to visually hide, delete, or see through undesirable objects in a perceived environment. The concept of DR is opposite to that of augmented/mixed reality (AR/MR), which seamlessly merges real and virtual environments to enhance user perception of reality, although its methodology is considered an extended form of AR/MR.

In DR, real undesirable objects are removed by covering the objects with background images estimated in the current view. DR methods are therefore categorized into three approaches depending on the hidden view recovery techniques: the observation-based DR (OB-DR), which recovers the hidden view from background images observed in advance (pre-observation-based DR; POB-DR) [1-5] or in real time (real-time observation-based DR; ROB-DR) [6], and the image inpainting-based DR (IB-DR) [7-9], which fills in a region

of interest (ROI) using images estimated through pixels or image patches around the ROI. OB-DR provides semantically correct results based on the observation. On the other hand, IB-DR provides seemingly plausible results, although the observation step is not required.

After the emergence of multi-view paraperspective projection approach [1], various object removal techniques have been proposed. Although most of the evaluations in DR literature describe observable facts regarding their resultant images, they rely on qualitative evaluations, such as visual verifications [1-4] with a comparative evaluation between the existing and proposed methods [5]. The main reason of this background is that it is difficult to obtain the ground truth of their results, i.e., an image of a scene with and without the target objects. In particular, it is physically difficult to capture such data as consecutive frames, especially outdoor. Accordingly, DR methods are presented to visually remove the undesirable objects from a scene.

To address this issue, we introduce data acquisition facilities to enable DR researchers ground truth data acquisition and evaluation workflow for quantitative evaluation of DR methods using the data. In this paper, our main target is OB-DR because IB-DR problems do not have a single solution represented as ground truth, i.e., solutions of IB-DR, in principle, are viewer-

dependent.

The major contributions of this research can be summarized as follows.

· Design and construction of data acquisition facilities

· Specific workflow for capturing input frames, frame-by-frame corresponding ground truth frames, and other relevant data for evaluating DR methods

· Clarification of a quantitative evaluation workflow using ground truth data for OB-DR

The reminder of this paper is organized as follows. Section 2 describes the requirements and definitions for the ground truth of DR results. Section 3 describes the design and construction of the data acquisition facilities to obtain the ground truth and relevant data. Section 4 presents the flow of data acquisition using the facilities. In Section 5, we demonstrate the data acquired by the facilities sufficiently satisfies the ground truth requirements of DR research. Limitations are detailed in Section 6 and we summarize future work in Section 7.

2. Ground Truth and Evaluation Workflow of Diminished Reality

2.1 Overview

The goal of this study is to acquire a pair of consecutive frames with the target objects to be diminished I^s_i and corresponding frames without the target objects I^g_i. The image sequence pair I^s_i and I^g_i must be captured without any differences except the existence of the target objects, i.e., the scene is mutual in terms of geometry and illumination and is captured with the same camerawork. Thus, the proposed system must use an accurate motion-controlled camera and mechanisms for controlling scene complexity and illumination. In addition, this must be achieved regardless of indoor or outdoor scenarios. Figure 1 shows three photography phases in the proposed facilities.

Details of these phases are described in the following sections.

2.2 Evaluation Criteria

Using an input dataset such as a source view image sequence I^s_i ($1 \le i \le M$), hidden view images I^b_j ($1 \le j \le N$), and 6 degrees of freedom (DoF) camera pose $\mathbf{M}_i$, results of a DR method are generated as an image sequence I^{dr}_i. Given the result I^{dr}_i and the ground truth image sequence I^g_i, errors or scores in i^{th} frame ε_i are calculated (Eq. 1).

$$\varepsilon_i = R(I^{dr}_i, I^g_i) \tag{1}$$

We assume that R is a measure such as mean squared error (MSE) or peak signal to noise ratio (PSNR) for geometric measurement, or structural similarity (SSIM) [10] or high dynamic range visual difference predictor (HDR-VDP) [11] for perceptual metric to compare I^{dr}_i and I^g_i. We consider such metric is selected in accordance to the goal of the evaluated DR method, i.e., user-dependent.

Consequently, the frame-by-frame analysis of DR methods, which has not been attempted in DR research community, is achieved. Figure 2 summarizes the flow of evaluation of a DR algorithm.
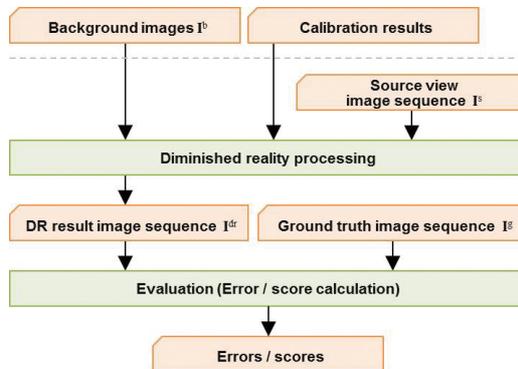

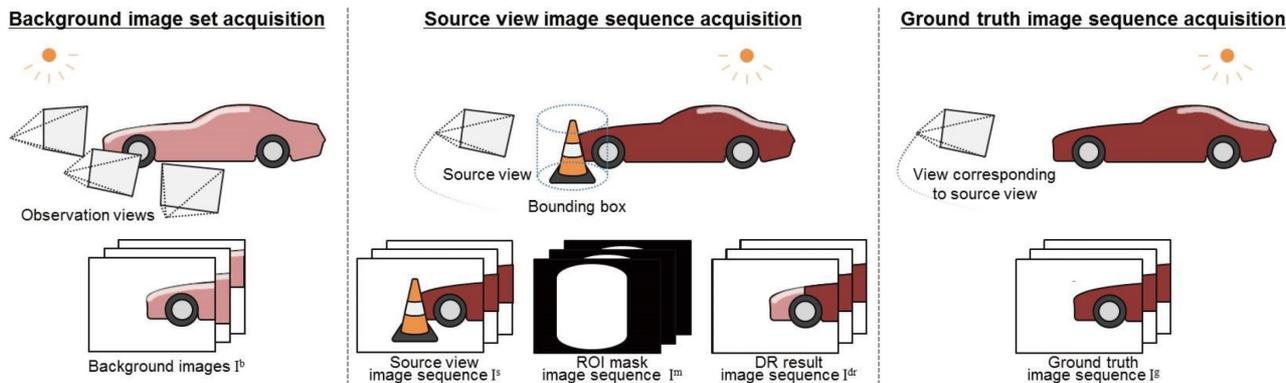**Fig.2** Workflow of evaluation of a DR method.

Fig. 1 Three photography phases in the proposed facilities. The real view and ground truth image sequence are referred to as "the image sequence set," and the data shown in this figure except DR result image sequence is referred to as "the dataset."

2.3 Geometric and Photometric Measure

Geometric and photometric errors should be separately measured for rational evaluation. First, this section briefly describes geometric and photometric consistency problems in DR.

Geometric consistency problem: The geometric consistency problem causes geometric gaps between the current real view and the synthesized view primarily due to errors in camera pose estimation, hidden area reconstruction, and target object detection.

Photometric consistency problem: The photometric consistency problem is evident after geometric consistency problems are solved. POB-DR must absorb illumination changes due to two data acquisition steps, i.e., the hidden area image and the source view image sequence acquisition stages. ROB-DR must resolve the differences in camera optics. Although the dataset should include illumination changes during real view image acquisition, only a few attempts have been made to address this problem [4, 8]. Thus, this issue remains as future work in this paper.

The following two possible data acquisition patterns are considered to isolate these two problems. The first pattern uses a hidden area image set I^{b1}_j under lighting #1, a source image sequence I^{s1}_i under the same lighting, a source image sequence I^{s2}_i under lighting #2, and corresponding ground truth image sequences I^{g1}_i and I^{g2}_i. Geometric errors $\varepsilon^{geo\#1}_i$ are then calculated (Eq. 2).

$$\varepsilon^{geo\#1}_i = R(I^{dr1\text{-}1}_i, I^{g1}_i) \tag{2}$$

where $I^{dr1\text{-}1}_i$ is the ith DR result image based on I^{b1}_j and I^{s1}_i.

Photometric errors $\varepsilon^{pho\#1}_i$ are then calculated (Eq. 3).

$$\varepsilon^{pho\#1}_i = \left| R(I^{dr1\text{-}2}_i, I^{g2}_i) - \varepsilon^{geo\#1}_i \right| \tag{3}$$

where $I^{dr1\text{-}2}_i$ is the ith DR result image based on I^{b1}_j and I^{s2}_i.

The second pattern uses two hidden view image sets I^{b1}_j and I^{b2}_j under different lightings, a source view image sequence and I^{s1}_i, and I^{g1}_g (Eq. 4 and 5).

$$\varepsilon^{geo\#2}_i = R(I^{dr1\text{-}1}_i, I^{g1}_i) \tag{4}$$

$$\varepsilon^{pho\#2}_i = \left| R(I^{dr2\text{-}1}_i, I^{g1}_i) - \varepsilon^{geo\#2}_i \right| \tag{5}$$

where $I^{dr2\text{-}1}_i$ is the ith DR result image based on I^{b2}_j and I^{s1}_i. Our facilities assume to use the second data acquisition pattern to reduce number of image sequences.

2.4 Real vs. Synthetic Datasets

Among existing ground truth acquisition research and benchmarks, some datasets offer real data [11, 12] and others provide synthetic data [12, 13]. A real dataset includes unavoidable factors, e.g., motion blur, noise, optics differences, etc. However, with real data acquisition, it is sometimes difficult to handle coordinate systems and control environments, especially in outdoor scenarios. Synthetic data can handle such problems but the abovementioned real factors with real datasets tend to be ignored. In recent years, some immersive quality computer graphics datasets [12, 13] have been created; however, this study focuses on real data, which is essential for DR research.

3. Data Acquisition Facilities

3.1 Full-scale and 1/12 Miniature Sets

The data acquisition facilities (Fig. 3) are constructed as a full-scale set (Fig. 4) and a 1/12 miniature set (Fig. 5). Practically, DR methods effectively work for outdoor scenes because, generally, real outdoor objects are not allowed to be replaced or arbitrarily controlled. However, it is virtually impossible to acquire appropriate outdoor datasets due to dynamic geometric and illumination changes during an image sequence
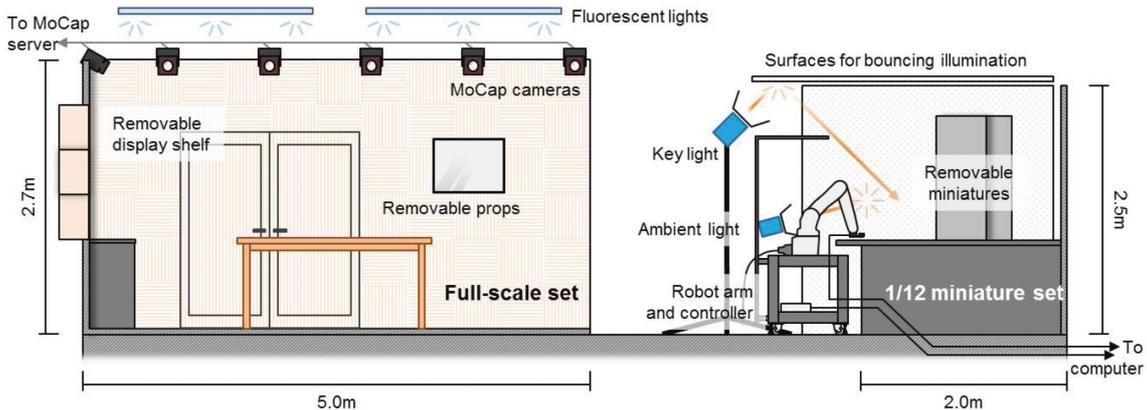

Fig.3 Sketch of the full-scale and 1/12 miniature sets.

Fig. 4 Full-scale set and scene arrangement example.

Fig. 5 1/12 miniature set.

Fig. 6 Illumination control in the proposed studio (left to right: sunny, cloudy, and sunset).

Fig. 7 Photographing system and coordinate system in hand-eye calibration.

capture. Therefore, in this study, a miniature set and controllable illumination devices are used to imitate outdoor data capture, i.e., illumination is static during each data capture and it can be changed between the captures.

The data acquisition facilities comprise an illumination control system and a photography system at a full-scale set (Fig. 4) and a 1/12 miniature set (Fig. 5) in a room with blacked-out windows, as shown in Fig. 3. The geometric complexity of the sets can be changed by adding, removing, and arranging props. The photometric complexity of the sets is controlled by the illumination control system described in Section 3.2.

The $5{\times}5m^2$ full-scale set has a motion capture system (six Vicon Bonita3 and six Bonita10) to track and record head motion (Fig. 4), which can be used for robot arm manipulation (Section 3.3). The scene geometry is changed using props and display shelves with openable lids (Fig. 4). The $2{\times}2m^2$ miniature set is 1/12 scale. This size was determined based on the quality of miniature models and the movable range of the photography system described in Section 3.3. All miniature models are replaceable, which facilitates changing the geometric complexity of the scene. The acrylic plates of the buildings are also removable, which allows the specular reflection of the scene to be changed.

3.2 Illumination Control System

The scene is illumined based on cinematography illumination techniques to represent outdoor illumination in the miniature set. The illumination control system comprises one key light (ARRI HMI575W) as sun light and several ambient light

apparatus (ARRI 650PLUS) as ambient light. These lamps are attached to stands and can be manually positioned and directed in the environment. The power and diffusivity of the lamps are manipulated by bouncing light off silk surfaces (Fig. 3) or attaching diffusers. The colors are also changeable using color filters. Figure 6 shows illumination examples obtained using the illumination control system.

Note that this lighting system is also used to illume the full-scale set basically lit by fluorescents when additional effects are necessary, e.g., light coming through a window.

3.3 Photographing System

A computer-controlled 6DoF industrial robot arm with an RGB camera allows the image sequence set to be captured with frame-by-frame coherent camera motion. The proposed photographing system is composed of a computer and a Denso Wave VS-087 (position repeatability: ±0.03mm, maximum motion area: 905mm) (Fig. 7). The system converts pre-recorded camera motion data (e.g., head motion data recorded in the full-scale set) in the form of an AIST dataset of TrakMark[12] into a language to control robot arm (PacScript) and transfers it to the robot arm controller (RC8) via local area network. Note that positions must be scaled down by 1/12 when used in the miniature set.

An RGB camera is attached to the robot arm using a custom attachment mechanism. Camera parameters, such as shutter timing, exposure time, and white balance, are changeable using the computer. In this study, the system uses Point Grey Flea3 GigE or Canon EOS 5D Mk. III cameras depending on the required

image resolution. The camera control systems for the Flea3 and EOS are implemented using the FlyCapture SDK and the EOS SDK, respectively, and control signals are transported via Ethernet and USB 2.0 cable, respectively.

Note that the industrial robot arm guarantees position repeatability only on predetermined control points. In addition, the robot arm can shake during imaging. To prevent these problems, the robot arm moves its hand to the next control point, stays for n s (1 s by default), captures an image, and then repeats these steps until it reaches the end of the predetermined control points. Although this photographing method prevents motion blur, the photographing system prioritizes the coherence of the image sequence set. Manual photography is also available using a teaching pendant attached to the robot arm.

The photography system requires hand-eye calibration to transform the robot hand pose to the attached camera pose. Such camera pose is recorded at every frame and used for, for example, hidden view recovery and placing 3D bounding objects. In our setup, it is assumed that the robot arm is accurately placed at a known position and orientation of the base and the attached camera is pre-calibrated (e.g., using Zhang's method[16]). Figure 7 shows transformations used in our hand-eye calibration setup.

There are well known approaches for estimating the necessary offset from the robot hand to the camera $\mathbf{M}_{f,\,c}$[17] and simultaneously estimating $\mathbf{M}_{f,\,c}$ and the external offset from the known calibration object to the robot origin $\mathbf{M}_{m,\,r}$ for further error minimization[18]. In this paper, the robot arm is assumed to be attached to the known base; therefore, $\mathbf{M}_{m,\,r}$ is known, and it is only necessary to estimate $\mathbf{M}_{f,\,c}$. Unlike Tsai's method[17], wherein three-dimensional (3D) space errors are minimized, the projection errors of known points of the calibration object are minimized to estimate $\mathbf{M}_{f,\,c}$.

First, the photography system collects a set of calibration object images and corresponding hand poses $\mathbf{M}^{(i)}_{r,\,f}$. Next, $\mathbf{M}_{r,\,f}$ is calculated based on the following equations:

$$h\tilde{\mathbf{x}}^{\mathrm{T}} = \mathbf{CM}_{f,c}\mathbf{M}^{(i)}_{r,f}\mathbf{M}_{m,r}\tilde{\mathbf{X}}^{\mathrm{T}} \qquad (6)$$

$$h\tilde{\mathbf{x}}^{T} = \mathbf{CM}_{f,c}\tilde{\mathbf{X}}'^{\mathrm{T}}, \text{ where } \tilde{\mathbf{X}}' = \mathbf{M}^{(i)}_{r,f}\mathbf{M}_{m,r}\tilde{\mathbf{X}} \qquad (7)$$

where h is a scale factor, $\tilde{\mathbf{x}}$ are homogeneous screen coordinates detected in the captured images, $\mathbf{C}$ is a set of intrinsic parameters, and $\tilde{\mathbf{X}}$ is a set of the known


Fig. 8 Basic workflow of the proposed data acquisition method (boxes with heavy lines are the three phases in Fig. 1 and boxes with thin lines are preparation phases).

homogeneous object coordinates of the calibration object. Therefore, solving the perspective n-point (PnP) problem gives $\mathbf{M}_{f,\,c}$.

4. Data Acquisition Workflow

Figure 8 shows the basic workflow of the data acquisition in the full-scale and the miniature set. First, hand-eye calibration is executed to calculate camera poses (Section 3.3). Next, props are arranged in the set to adjust the geometric complexity of the scene depending on the DR methods to be analyzed or evaluated (Section 3.1). Then, the key light and ambient lights are set to illuminate the scene (Section 3.2). Backgrounds are recorded as a set of images using the photography system based on the determined photographing points (Section 3.3). Note that the scene lighting and hidden area image set acquisition stages are repeated as required. A ground truth image sequence is captured based on pre-fetched camera motion for a source view image sequence. Similarly, the corresponding source view image sequence is captured based on the same camera motion after the target objects are placed in the scene.

Note that placing the target objects is only allowed once to maintain geometric and photometric consistency between the image sequence pair I^{s}_{i} and I^{g}_{i}. However, illumination changes are allowed when capturing the background images. Consequently, the proposed photographing scheme can capture background images under a variety of lighting conditions and the image sequence pair under the same lighting conditions. Note that on-off switching of the lamps at the same pose can be tolerated when capturing the image sequence pair.
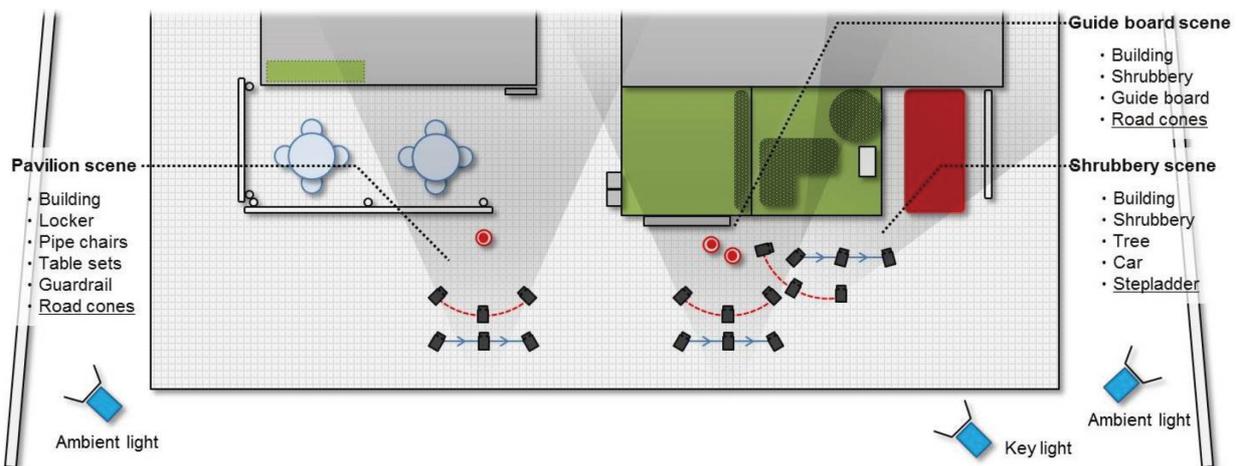
Fig. 9 Props and illumination settings of the experiments. Red double circles depict target objects. Blue arrows represent the camera path of real image sequences. Red dotted lines show camera paths of the hidden area image acquisition.

5. Operations Test and Dataset Evaluation

5.1 Outline

This section demonstrates that the proposed data acquisition facilities can provide sufficient ground truth datasets for analysis and quantitative evaluation of DR methods. Figure 9 shows props and illumination settings in the 1/12 miniature set. Given the calibration results (Section 3.3), a hidden area image set I^b_j, a source view image sequence I^s_i, and a ground truth image sequence I^g_i are captured; a mask image sequence I^m_i is generated for each scene. Through an operations test, it is shown that these datasets can be used for DR methods in the following and their quantitative evaluation. As a quantitative measure, we used MSE as a reference.

5.2 Operations Test

Hidden view recovery methods of existing OB-DR methods can be classified into two approaches. One is a geometric approach that attempts to reconstruct 3D geometry and the color of hidden areas[1]. The other is an image-based approach that uses image-based rendering (IBR) or similar approaches[2, 4, 5], which is known to be more effective. Accordingly, the following three DR methods were implemented as test benches.

Method A: This test bench uses geometric approach to recover hidden views. Backgrounds are reconstructed as a textured 3D model. The model is projected onto the current view to cover the target region based on the given camera poses. When hidden area images are given, the corresponding textured 3D model is created using Autodesk 123D Catch.

Method B: This test bench uses IBR to recover hidden views. Background are reconstructed as a spherical light field (SLF)[19] focusing on the positions target objects to be placed. The hidden view is recovered using the SLF according to the camera motion. Since a SLF is a modeless IBR method, the SLF is expected to generate sophisticated DR results compared to existing DR methods[1-9].

Method G: This test bench uses ground truth image I^g_i to fill in a target ROI determined by binary masks I^m_i in the current frame I^s_i.

Note that these DR methods use the camera pose $\mathbf{M}_i$ from the photography system and a common pre-fetched mask image sequence I^m_i for each scene. The mask image I^m_i is generated using a 3D bounding box manually placed in the environment. The 3D bounding box is transformed based on a given camera pose $\mathbf{M}_i$ and is projected into the current view as a binary mask to separate the ROI from other regions.

5.3 Dataset Evaluation

The generated datasets are evaluated by comparing the results of Method G against Methods A and B. Errors in Method G are equivalent to errors due to the proposed data acquisition system. Therefore, errors in Method G must be lower than those of Methods A and B, which are accomplished under better conditions than those of existing studies, i.e., no illumination changes and use of known camera poses.

5.4 Test Datasets

All data was captured using the Point Grey Flea 3 GigE camera (640×480 resolution, 24-bit RGB color, with Myutron Inc. FV0622 40.5° (V) FoV lens). 200 source and ground truth image pairs are recorded in the following scenes.

Pavilion scene: Objects with complex shape, such as pipe chairs, tables, and guardrails are placed in the

environment. Compared to target scenes in existing studies, the scene is significantly complex. Methods A, B, and G are compared using a dataset created in the scene. The SLF of Method B comprises 6,000 images at horizontally and vertically equal intervals such that 60° and 7.5° is divided into 200 and 30 sections at 0.3° and 0.25°, respectively. Fifty images are selected among these and input to Autodesk 123D Catch to generate a textured 3D model. The 3D model is manually registered in the scene. Hereafter, a textured 3D model of a scene is created in the same manner.

Guide board scene: Target objects placed in front of a guide board are visually removed. Since the geometry of the scene is simple enough as to be approximated to a set of planes, Methods A and G are compared.

Shrubbery scene: A stepladder placed in the shrubbery is visually removed. The size of the ROI is the largest compared to the other scenes. Since the scene geometry and reflectance are more complex due to the shrubbery and the car, the degree of difficulty is higher than the guide board scene. However, the geometry is

sufficiently reconstructed as a textured 3D model. Accordingly, Methods A and G are compared for this scene. The real view image sequence is captured away from where the hidden area image set is acquired.

5.5 Results

Figure 10 shows example frames of the real view, ground truth, and mask image sequence in the pavilion scene. Figure 11 shows DR results of Methods A, B, and G, and their difference image against the ground truth. The MSEs of these example frames of Method A, B, and G were 199.86, 33.58, and 9.31 respectively. In the difference image of Method A, errors are clearly observed in ROI due to inaccurate 3D reconstruction. The DR results of Method B appear satisfactory; however, edge-like errors are observed in the ROI in the difference image due to the slightly blurred result. The errors of Method G (i.e., errors of the proposed data acquisition facilities) appear outside the ROI. However, only a few errors are observed compared to the other results.
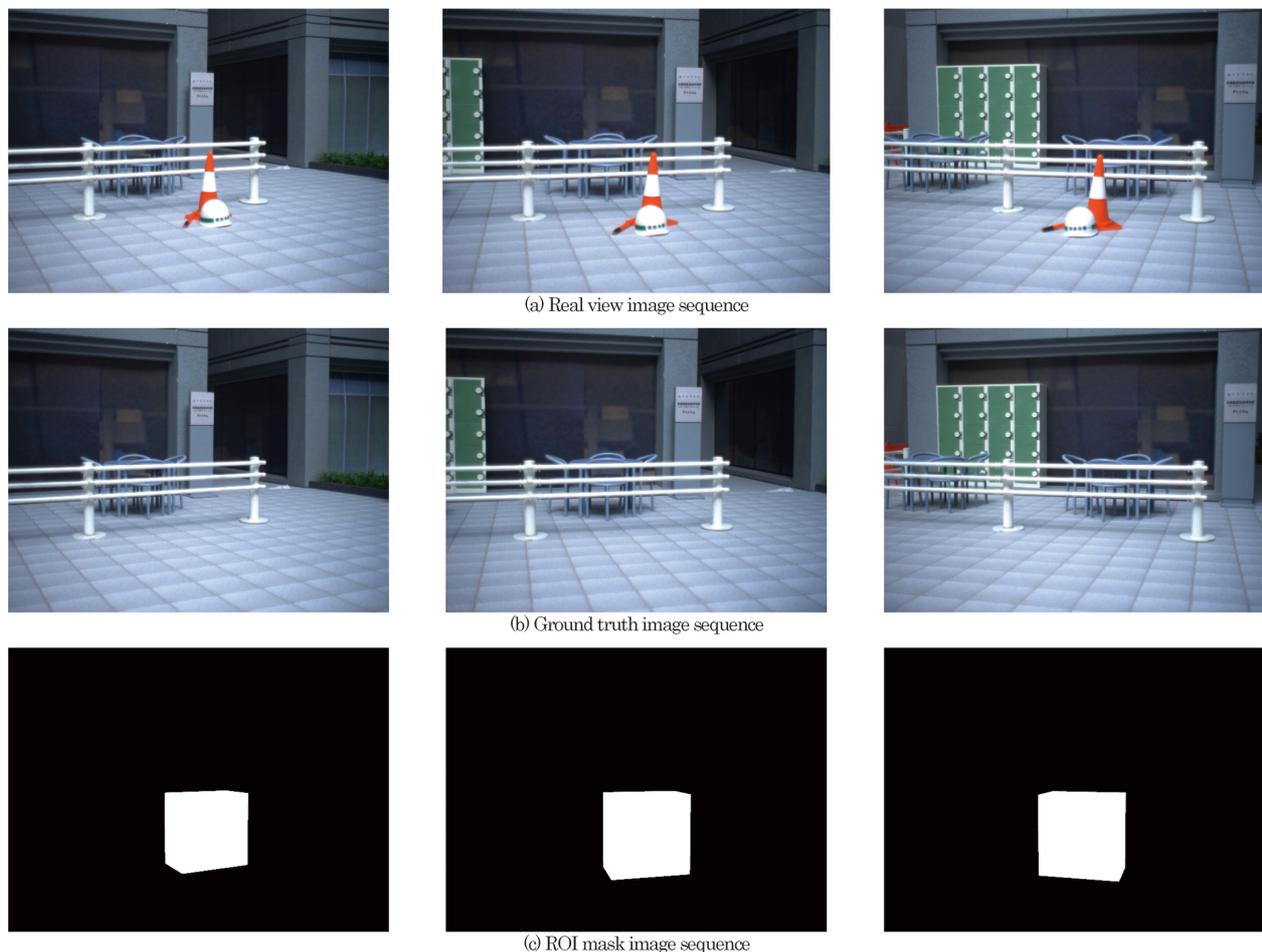
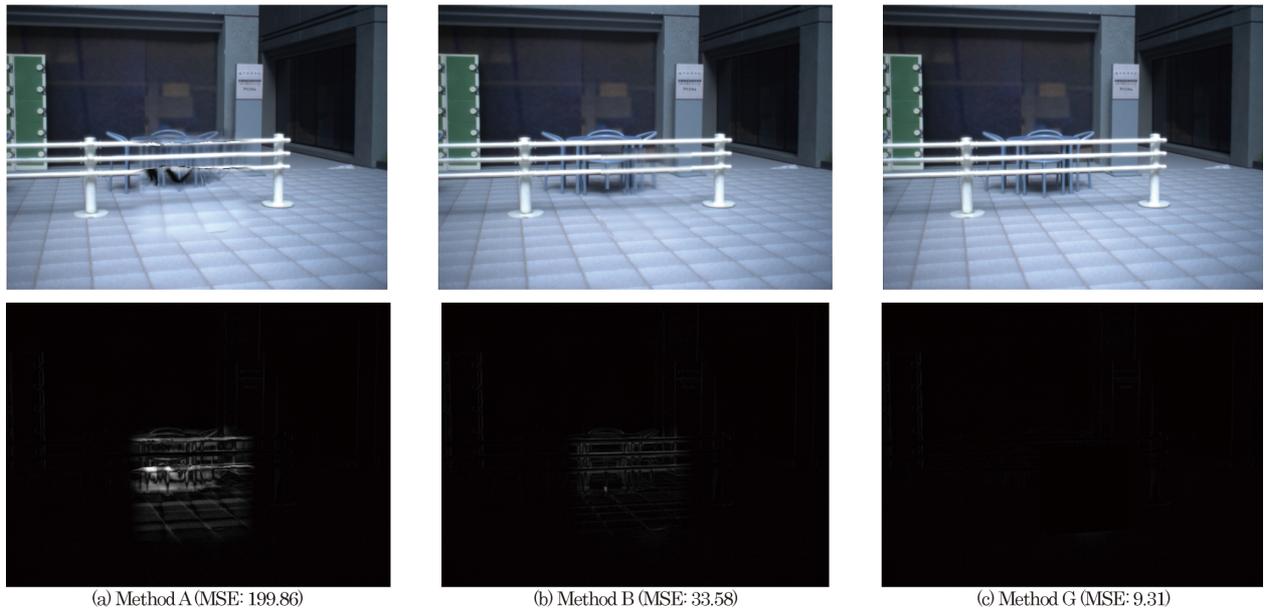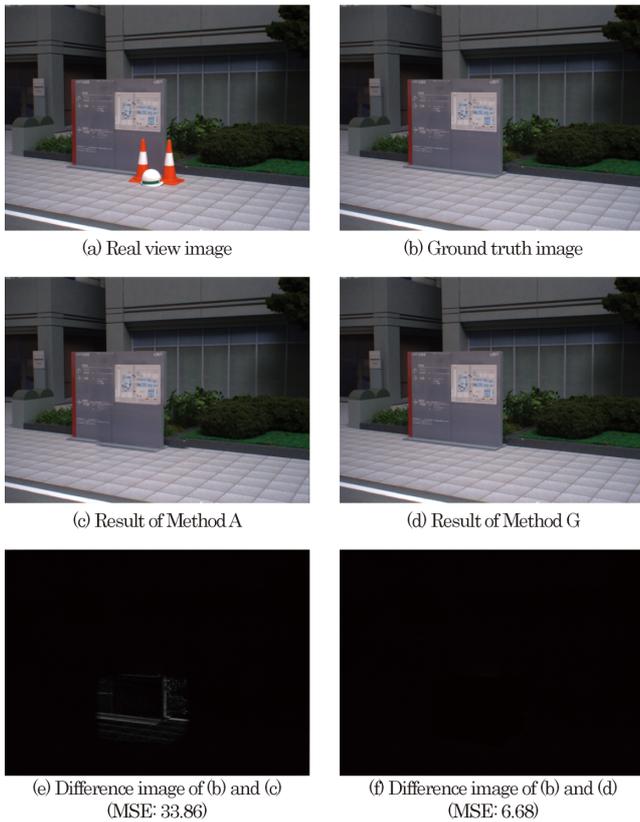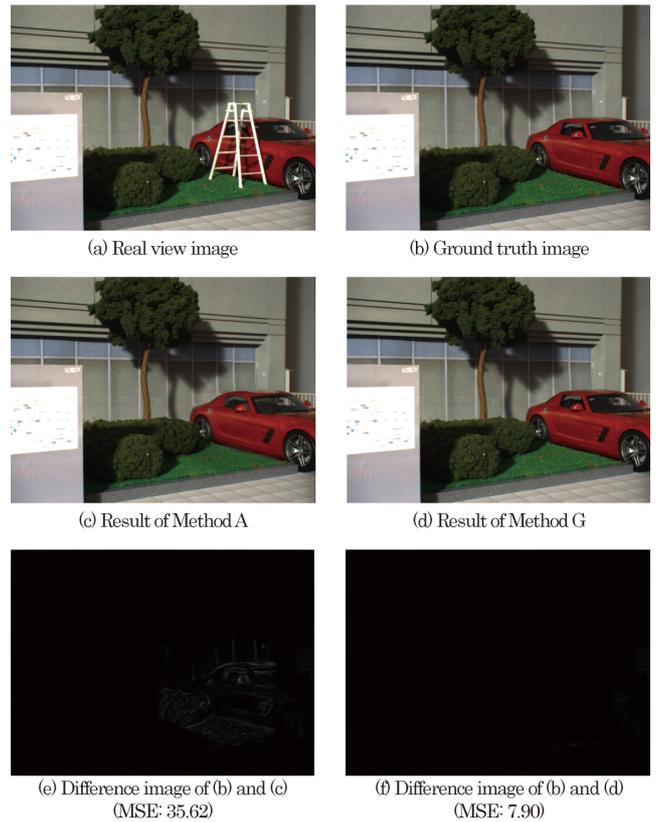Figure 12 and Fig. 13 shows results for the guide

(a) Real view image sequence

(b) Ground truth image sequence

(c) ROI mask image sequence

Fig.10 Example frames of real view, ground truth, and mask image sequence in the pavilion scene.

(a) Method A (MSE: 199.86)　　　　(b) Method B (MSE: 33.58)　　　　(c) Method G (MSE: 9.31)

Fig.11 Comparison of Methods A, B, and G (Top row: DR results, bottom row: Difference images).

(a) Real view image

(b) Ground truth image

(c) Result of Method A

(d) Result of Method G

(e) Difference image of (b) and (c)
(MSE: 33.86)

(f) Difference image of (b) and (d)
(MSE: 6.68)

Fig.12 Guide board scene.

(a) Real view image

(b) Ground truth image

(c) Result of Method A

(d) Result of Method G

(e) Difference image of (b) and (c)
(MSE: 35.62)

(f) Difference image of (b) and (d)
(MSE: 7.90)

Fig.13 Shrubbery scene.

board and shrubbery scenes, respectively. As can be seen, the proposed data acquisition system provides sufficient datasets. The MSEs of the example frames of Method A and G in Guide board scene were 33.86 and 6.68 respectively. Likewise, in Shrubbery scene, 35.62 and 7.90 respectively.

6. Limitations

As described in Section 3.3, the proposed photographing system cannot represent motion blur because it keeps the camera still until a frame is securely captured. Motion blur is an important problem that should be addressed by DR methods in actual

situations. Motion blur can be addressed by digitally exposing the camera at additional control points between the current control points.

Since all data acquired by the proposed system are real data, real view images and corresponding ground truth images include different noise (e.g., random noise from an image sensor). This problem could be solved using the mean of added images or long exposure for a single image with low gain. However, photographing for a long time has other problems, e.g., we experienced temporal differences in illuminations and the set changes with the heat generated by the lamps. Accordingly, the total photographing time should be limited to a certain fixed time.

Based on the proposed workflow, changes in illumination and props are allowed when capturing background images. On the contrary, variations in the source image sequence capture are rather limited. Overcoming this limitation will require expanding the lighting control systems. For example, it is possible that this problem could be solved by placing controllable objects, such as a digital clock or a display, in the scene.

7. Conclusion

This paper has presented a qualitative evaluation workflow and data acquisition facilities for DR research wherein a ground truth image sequence is captured with sufficient accuracy. The presented facilities include full-scale and 1/12 miniature sets illuminated using a lighting system based on cinematography and photographed using a 6DoF industrial robot arm with a camera. The proposed workflow maintains geometric and photometric consistency between the ground truth and the input data acquisition.

The proposed system can capture these data as consecutive frames, which, to the best of our knowledge, has not been achieved to date. Consequently, frame-by-frame evaluation of DR results is possible using the acquired dataset by comparing DR results and the corresponding ground truth frames. To demonstrate accuracy and effectiveness, three scenes were constructed in the miniature set and three corresponding datasets were created according to the proposed workflow. These datasets were used with three test bench DR methods.

In future, we will attempt to overcome the limitations discussed in Section 6 and publish datasets.

References

1) S. Zokai, J. Esteve, Y. Genc and N. Navab: "Multiview paraperspective projection model for diminished reality", Proc. ISMAR 2003, pp.217-226 (2003)
2) F.I. Cosco, C. Garre, F. Bruno, M. Muzzupappa and M.A. Otaduy: "Augmented touch without visual obtrusion", Proc. ISMAR 2009, pp.99-102 (2009)
3) P. Barnum, Y. Sheikh, A. Datta and T. Kanade: "Dynamic seethroughs: Synthesizing hidden views of moving objects", Proc. ISMAR 2009, pp.111-114 (2009)
4) S. Mori, F. Shibata, A. Kimura and H. Tamura: "Efficient use of textured 3D model for pre-observation-based diminished reality", Proc. IWDR, pp.32-39 (2015)
5) Z. Li, Y. Wang, J. Guo, L.-F. Cheong and S.Z. Zhou: "Diminished reality using appearance and 3D geometry of internet photo collections", Proc. ISMAR 2013, pp.11-19 (2013)
6) A. Enomoto and H. Saito: "Diminished reality using multiple handheld cameras", Proc. ACCV 2007, pp.130-150 (2007)
7) J. Herling and W. Broll: "PixMix: A real-time approach to high-quality diminished reality", Proc. ISMAR 2012, pp.141-150 (2012)
8) N. Kawai, M. Yamasaki, T. Sato and N. Yokoya: "Diminished reality for AR marker hiding based on image inpainting with reflection of luminance changes", Proc. ITE Trans. on MTA, 1, 4, pp.343-353 (2013)
9) N. Kawai, T. Sato and N. Yokoya: "Diminished reality considering background structures", Proc. ISMAR 2013, pp.259-260 (2013)
10) Z. Wang, A.C. Bovik, H.R. Sheikh and E.P. Simoncelli: "Image quality assessment: from error visibility to structural similarity", IEEE Trans. on Image Processing, 13, 4, pp.600 - 612(2004)
11) R. Mantiuk, K.J. Kim, A.G. Rempel and W. Heidrich: "HDR-VDP-2: A calibrated visual metric for visibility and quality predictions in all luminance conditions", ACM Trans. on Graphics, 30, 4, 40 (2011)
12) H. Tamura, H. Kato and the TrakMark Working Group: "Proposal of international voluntary activities on establishing benchmark test schemes for AR/MR geometric registration and tracking methods", Proc. ISMAR, pp.233-236 (2009)
13) Y. Nakamura, T. Matsuura, K. Satoh and Y. Ohta: "Occulusion detectable stereo-occulusion patterns in camera matrix", Proc. CVPR, pp.371-378 (1996)
14) S. Murtull, M.P. Matorell and K. Fukui: "Realistic CG stereo image dataset with ground truth disparity maps", Proc. PRMU, 111, 430, pp.117-118 (2012)
15) A. Handa, T. Whelan, J. McDonald and A.J. Davison: "A benchmark for RGB-D visual odometry, 3D reconstruction and SLAM", Proc. ICRA, pp.1524-1531 (2014)
16) Z. Zhang: "Flexible camera calibration by viewing a plane from unknown orientations", Proc. ICCV, pp.666-673 (1999)
17) R. Tsai and R. Lenz: "A new technique for fully autonomous and efficient 3D robotics hand/eye calibration", IEEE Trans. on Robotics and Automation, 5, 3, pp.345-358 (1989)
18) F. Dornaika and R. Horaud: "Simultaneous robot-world and hand-eye calibration", IEEE Trans. on Robotics and Automation, 14, 4, pp.617-622 (1998)
19) Z. Wang and Z. Liang: "Sphere light field rendering", Proc. SPIE, 4681, pp.357-365, Medical Imaging (2002)

Shohei Mori received a B.S., M.S., and Ph.D. degrees in engineering from Ritsumeikan University, Japan, in 2011, 2013, and 2016, respectively. He was a Research Fellowship for Young Scientists (DC-1) from the Japan Society for the Promotion of Science until 2016. He is currently a Research Fellowship for Young Scientists (PD). He is a member of IEEE and ACM.

Yuya Eguchi received a B.S. in computer science and a M.S. in computer science and engineering from Ritsumeikan University, Japan, in 2013 and 2015, respectively. He began working at Murata Machinery Inc. in 2015.

Sei Ikeda received a Ph.D. degree from Nara Institute of Science and Technology, Japan, in 2006. He worked as an assistant professor at the Graduate School of Information Science of the same institute until 2011, and at the Graduate School of Engineering Science of Osaka University until 2015. He was also a visiting researcher at the University of Oulu, Finland, in 2010 and 2011. He is currently a lecturer at the College of Information Science and Engineering, Ritsumeikan University. He is a member of IEEE and ACM.

Fumihisa Shibata received a M.E. in computer science and a Ph.D. in engineering from Osaka University, Osaka, Japan, in 1996 and 1999, respectively. In 2003, he joined Ritsumeikan University, Kusatu, Shiga, Japan, as an Associate Professor in the College of Information Science and Engineering. He has been a Professor at Ritsumeikan University since 2013. He is a member of the Virtual Reality Society of Japan, the Information Processing Society of Japan, and the Institute of Electronics, Information and Communication Engineers.

Asako Kimura received B.S., M.S., and Ph.D. degrees in engineering science from Osaka University, Suita, Osaka, Japan, in 1996, 1998, and 2001, respectively. She is a Professor at the College of Information Science and Engineering, Ritsumeikan University, Kusatu, Shiga, Japan. She taught at Osaka University from 2000 to 2003 before coming to Ritsumeikan University. She is a member of the Information Processing Society of Japan, the Institute of Electronics, Information and Communication Engineers, and the Association for Computing Machinery.

Hideyuki Tamura received B.Eng. and Ph. D. degrees in electrical engineering from Kyoto University, Kyoto, Japan, in 1970 and 1984, respectively. During his professional career, which began in 1972, he has been a Senior Research Official at the Electrotechnical Laboratory, MITI, Japan; the Director of Media Technology Laboratory, Canon Inc.; and a member of the executive board of the Mixed Reality Systems Laboratory Inc. In 2003, he joined the Ritsumeikan University, Kusatu, Shiga, Japan, as a Professor at the College of Information Science and Engineering. He is currently an Eminent Professor of Research Organization of Science and Technology, Ritsumeikan University. Professor Tamura has served on the executive boards of several academic societies in Japan and has received several awards from the Information Processing Society of Japan and the Institute of Electronics, Information and Communication Engineers.