[POSTER] Design and Implementation of a Common Dataset for Comparison and Evaluation of Diminished Reality Methods

Taiki Morozumi[†1] Shohei Mori[†2] Sei Ikeda[†1] Fumihisa Shibata[†1] Asako Kimura[†1] Hideyuki Tamura[†1]
Ritsumeikan University[†1] Keio University[†2]

ABSTRACT

Diminished reality (DR) is a technique to remove or inpaint real objects in a display. While DR is one of the active topics in the ISMAR community, a fair comparison between existing or emerging DR methods is difficult, that is, many methods rely on subjective evaluation that uses their own results to demonstrate their advantages. We, therefore, present a common dataset as a basis of the research for fair evaluations. In this article, we describe the design and implementation directions of the dataset. Our dataset is open to the public on the Web [†] (as of Oct. 9, 2017).

Keywords: Diminished Reality, Dataset, Ground Truth, Quantitative Evaluation.

Index Terms: [Computing methodologies]: Computer graphics—Graphics systems and interfaces—Mixed/augmented reality.

1 INTRODUCTION

DR is an active area in which many methods have been proposed [1]. Almost all of the methods rely on subjective analysis of the resultant images. One of the main reasons for this approach is that there is no common dataset in this research area. We consider that the difficulties in creating such a dataset lie in: 1) ground truth background generation and 2) variety of the approaches.

The first issue can be resolved using a sophisticated facility [2] (*e.g.*, a robotic arm camera and static lighting systems). Regarding the second issue, we decided to convene with DR experts in Japan (See Acknowledgement) to select the necessary datasets. In this paper, we report our directions based on comments from the meeting while using common datasets [†] that were recorded using our facility [2] and share our experiences.

2 DATASETS AND SCENARIOS

DR can be categorized into two approaches: observation-based and image-inpainting approaches [1]. These approaches use the following data, and we therefore include them in our datasets:

1) Main View Image Sequence. This is an image sequence captured with a user camera during a DR experience.

2) Background Image Set. This is a set of multi-view images of backgrounds. Note that these data are not necessary for the image-inpainting approach.

3) Camera Parameters. These are intrinsic and extrinsic parameters of cameras used for capturing the image sequences described above.

In addition to these data, we obtained ground truth [2].

4) Ground Truth Image Sequence. This is an image sequence without target objects.

It is difficult to create a dataset for evaluating all DR methods without losing generality because DR methods have various approaches depending on scene complexity and input datasets of the user. We therefore convened and discussed with DR experts in Japan to determine scenarios to be included in our datasets. Consequently, the following four major comments from the members were considered in creating four categories of datasets. The members commented that they needed a dataset for:

- Evaluating the performance of their DR method
- Comparing results in outdoor daylight changes
- Evaluating real applications
- Evaluating their DR methods in a dynamic scene

3 DATASET CONTENTS

Our dataset is classified into four categories, and users can choose one of them based on their purpose (**Table 1**).

Category 1: This dataset is used for objective evaluation under eight different conditions in an indoor environment. Almost all of the DR experts commented that they have had difficulties in collecting image data with variations in lighting conditions, textures, and geometries of backgrounds. Our dataset consists of eight scenes containing different shapes, textures, lighting conditions, reflectance properties, and blurring conditions.

In this dataset, we first recorded the user's head motion using a MoCap system (VICON Bonita 3 and 10) and reproduced this motion using a robot arm (DensoWave VS-087) to record main view image sequences under the eight conditions. We used an object with characteristic textures to ease the 3D reconstruction of the object using structures from the motion. One hundred background images for 3D/light fields (LF) reconstructions were recorded in the camera path using grids before the target object was placed in the environment.

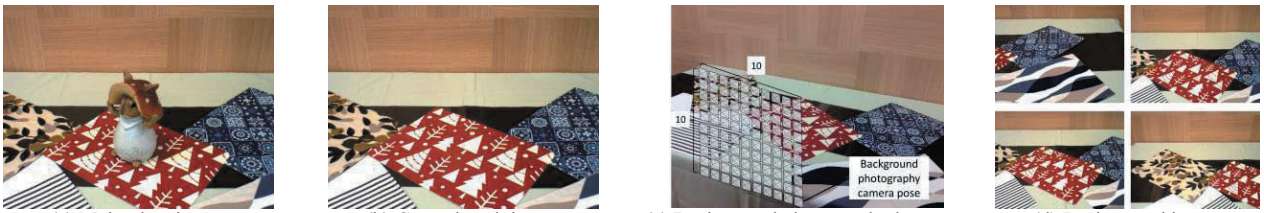
Figure 1 shows an example set of a main view and the

Table 1: Contents of the Dataset

Category	Purpose of Category	Target DR Method	Number of Flame	Target Object	Camera Path	Remarks
1	Comparison of variations in texture, shape, and lighting of the indoor environment	POB-DR IB-DR	1),4) : 509 2) : 100	Rabbit figure	Head motion	Motion was acquired using the motion capture system
2	Comparison of variations in outdoor-simulated lighting	POB-DR	1),4) : 300 2) : 100	Sign	Arc motion	A miniature set was used as an outdoor scene
3	Qualitative evaluation for a real outdoor scene	POB-DR IB-DR	1),4) : 1 2) : 200	Human	Fixed	Ground truth is only available for qualitative evaluation.
4	Evaluation for moving objects	POB-DR	1),4) : 200 2) : 100	Car model	Horizontal rotation	

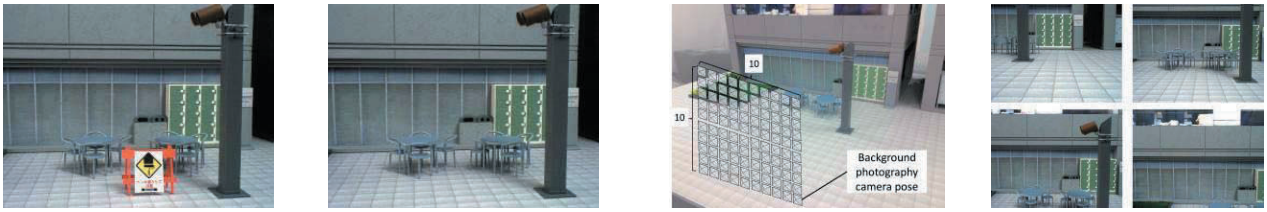
[†] DR Common Dataset (in Japanese/English)
http://www.rm.is.ritsumei.ac.jp/kiban-s/dr-dataset/

978-0-7695-6327-5/17 $31.00 © 2017 IEEE
DOI 10.1109/ISMAR-Adjunct.2017.69



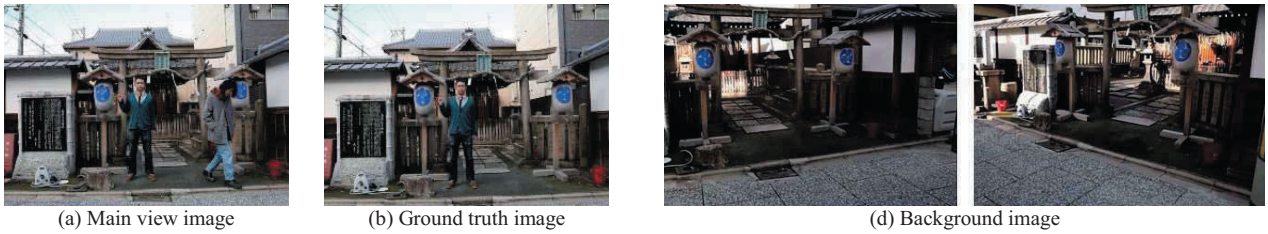
(a) Main view image (b) Ground truth image (c) Background photography image (d) Background image

Figure 1: Category 1



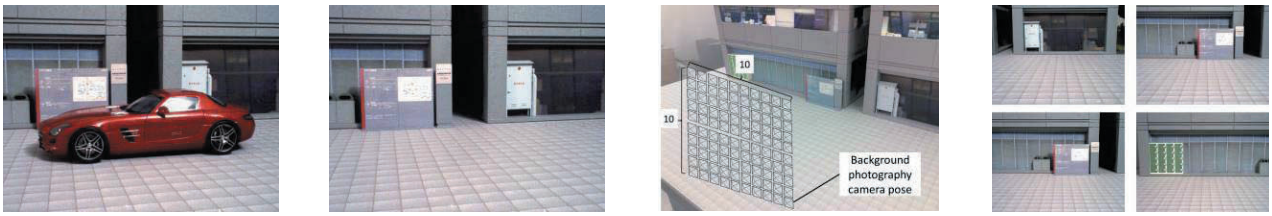
(a) Main view image (b) Ground truth image (c) Background photography image (d) Background image

Figure 2: Category 2



(a) Main view image (b) Ground truth image (d) Background image

Figure 3: Category 3



(a) Main view image (b) Ground truth image (c) Background photography image (d) Background image

Figure 4: Category 4

corresponding ground truth images.

Category 2: This is a dataset for evaluating outdoor daylight changes. It is difficult to obtain ground truth outdoors because the lighting changes while photographing. Thus, we created the dataset indoors using a miniature set under four simulated lighting conditions, such as cloudy sky at 13:00 and sunny sky at 17:00.

We used an ARRI HMI575W and an ARRI 650PLUS as a key light and an ambient light, respectively. During the recording, the main view camera moved in an arc while gazing at the target object. The ground truth images were acquired in the same manner as in Category 1. **Figure 2** shows an example set of a main view and the corresponding ground truth images.

Category 3: In this dataset, we assumed a typical scenario for DR, namely, removing a person from a captured photo. **Figure 3** shows the main view and the ground truth photos. As described above, it is difficult to acquire ground truth outdoors. We therefore acquired the ground truth photo immediately after taking the main view photo to keep the scene as consistent as possible.

Category 4: We moved a miniature car in a miniature set to create a dynamic scene. The main camera was panned during the recording to keep the moving car in the field of view. **Figure 4** shows the main view and the ground truth images. We recorded the ground truth images in the same manner as in Category 1.

4 Discussion

In this paper, we reported a "common dataset" to evaluate DR algorithms and to elevate DR research. Using the valuable comments from DR experts, we selected components, categories, and scenes to be tested for DR methods. In future studies, we could create a dataset of a dynamic scene in which both target objects and backgrounds move.

Acknowledgement

This work was supported in part by JSPS Grant Numbers 24220004 and 16J05114. We are grateful to Professors Yoshinari Kameda; Itaru Kitahara, Ph.D. (Tsukuba University); Tomokazu Sato, Ph.D., Norihiko Kawai, Ph.D. (NAIST); and Hideaki Uchiyama, Ph.D. (Kyushu University) for their helpful discussions.

References

[1] S. Mori *et al.*: "A survey of diminished reality: Techniques for visually concealing, eliminating, and seeing through real objects," IPSJ Trans. on CVA, SpringerOpen, 2017.

[2] S. Mori *et al.*: "Design and construction of data acquisition facilities for diminished reality research," ITE Trans. on MTA, pp. 259 - 268, 2016.