[POSTER] Background Image Registration as a Post-Processing Technique in Diminished Reality Rendering Procedures

Shohei Mori[1*]   Jianing Qie[2†]   Sei Ikeda[2]   Fumihisa Shibata[2]   Asako Kimura[2]   Hideyuki Tamura[2]

[1] Keio University, [2] Ritsumeikan University,

Figure 1: Overview of the proposed framework.

ABSTRACT

Diminished reality (DR) is a technique to remove undesirable objects from a video stream in real time. DR methods calculate a user's camera pose using vision- or sensor-based approaches to recover and overlay a background image to the camera view. Relying on 6DoF camera registration methods, DR results are often ruined due to misregistration. To solve this problem, we propose a registration framework as a *post-processing* technique in diminished reality rendering procedures to reduce the misalignment in an image space. We also reproduce motion blur to improve the image space alignment in the post-processing. The results showed remarkable improvement compared to sensor-based DR results. We argue that this post-processing-based approach can be applied to almost all existing DR methods to improve their quality.

Keywords: Diminished reality, framework, image alignment.

Index Terms: H.5.1 [Information Interfaces and Presentation]: Multimedia Information System—Artificial, augmented, and virtual realities

1 INTRODUCTION

The goal of diminished reality (DR) is to visually remove a real object in accordance with the user's viewpoint [7]. DR methods need to recover and superimpose a hidden background in the user's viewpoint image without misalignment. Many conventional DR methods rely on 6DoF camera pose estimation for the recovery and superimposition processes (e.g., in-advance calibration, fiducial markers, simultaneous localization and mapping, and sensors). However, such methods suffer from erratic results in 6DoF camera tracking leading to misalignment in 2D image space. Even a small error in 6DoF camera pose, especially in rotation, can result in a critical misalignment in image space and, therefore, a high quality camera tracking method is desired.

Contrary to this concept, we propose in this paper to reduce such misalignment in 2D image space and to implement this process at the *post-processing* stage during DR rendering procedures to be easily integrated into existing DR methods. In other words, we transfer 6DoF camera tracking problems to the 2D image alignment phase to be integrated at the end of the rendering pipelines of existing DR methods (e.g., [3–5, 8]). To perform the high-quality 2D alignment, we implement a motion blur rendering scheme as well when recovering backgrounds.

2 TRACKING IN DR

2.1 Vision-Based Tracking

One can describe DR vision-based tracking methods as having two approaches to track the scene, by either the image in the region of
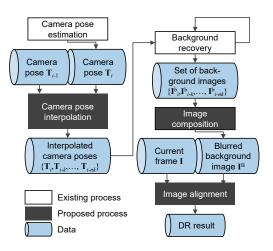
---

*e-mail: s.mori.jp@ieee.org
†e-mail: jianing@rm.is.ritsumei.ac.jp

interest (ROI) or by the surrounding region of interest (sROI; the rest of the image).

Tracking sROI: Most DR studies take this approach, in which the ROI is mainly ignored in a tracking procedure [5, 8]. The disadvantage of this approach is that the tracking may be lost because the ROI occupies most of the scene. DR methods based on pre-captured image datasets often use this approach since they assume that the background is reconstructed in advance in high quality. However, they must compensate for gaps between the recovered image and the image in sROI in terms of lighting and optical systems.

Tracking ROI: Given information about the objects to be removed, we can use them in tracking fiducial objects [3]. However, such objects to be removed are basically unknown until the DR experience starts. Thus, this approach is rarely used.

Consequently, in this study, we assume the DR method of "Tracking sROI" (i.e., only the background image is recovered as a final output of DR rendering procedures).

2.2 Sensor-Based Tracking

One can use a 6DoF sensor for tracking such items as a magnetic, an optical, or an ultrasonic type sensor. While sensor-based approaches are basically robust both indoors and outdoors, one needs to calibrate the coordinates of the background reconstruction and the sensor coordinates. Given this background, this approach is rarely used.

In the proposed method, we assume that calibration errors can be reduced at the final step of the DR rendering procedure.

3 PROPOSED FRAMEWORK

3.1 Overview

Fig. 1 shows the overview of our proposed rendering procedures. Given the previous and current 6DoF poses of the user camera, $\mathbf{T}_{i-1}$ and $\mathbf{T}_i$, respectively, from one of the tracking methods described in Section 2, we first generate n camera poses in between the two poses. Based on these virtual poses, we render the background image using an existing background recovery method. The generated background

Figure 2: Original image (a), a DR result using a raw sensor pose (b), and our DR result (c).

images at each pose are merged as a blurred background image $\mathbf{I}^B$. Finally, the resultant image $\mathbf{I}^B$ is aligned to the current frame $\mathbf{I}$ using an image alignment method.

3.2 Procedures

Camera Pose Estimation: We assume that the current camera pose $\mathbf{T}_i$ and the previous one $\mathbf{T}_{i-1}$ are given from a camera/scene tracker implemented with the DR method. Various types of trackers, such as positioning sensors, fiducial markers [2], visual simultaneous localization and mapping tools [6], and homography estimator [5], are acceptable.

Camera Pose Interpolation: We generate background image with motion blur $\mathbf{I}^B$ to be overlaid to the current frame $\mathbf{I}$. To integrate the blurring scheme into the existing DR rendering framework, we choose to generate n virtual camera poses with a duration k (= $sf/1000$) to interpolate two frames' poses to render background images at every pose. Here, s is the shutter speed in ms and f is the frame rate. Assuming that an existing DR rendering framework does not reproduce motion blur, we digitally simulate exposure (i.e., motion blur).

Background Recovery: At the generated virtual poses, we generate background images. A background rendering scheme in the existing DR framework, such as image warping [2, 5] and image-based rendering [8], is used for this process.

Image Composition: Generated n background images are compounded with a weighted average as digitally exposed background image $\mathbf{I}^B$.

Image Alignment: We introduce an image alignment method to align $\mathbf{I}^B$ to $\mathbf{I}$. This process can absorb the jitter due to the erratic tracking or the sensor and hand-eye calibration errors (See Section 2). The candidates for this process include a Lucas Kanade (LK)-based image tracker [1] and feature-based homography warping.

4 IMPLEMENTATION AND RESULTS

4.1 Setup

We recorded 30 Hz videos in a room environment to show the effectiveness and to confirm that our rendering framework runs in real time. We built a prototype system composed of a computer, 6DoF sensor (VICON Bonita 3 and 10), and an RGB camera (Canon MREAL HH-A1) as a user camera. We used the sensor for the "camera pose estimation" process. We first created a textured 3D model of the scene from multi-view images using Agisoft PhotoScan. We then rendered the 3D model at given poses using OpenGL without shading as the "background recovery" process. We generated 10 background images ($n = 10$) at each frame to create $\mathbf{I}^B$ and used an LK-tracker [1] as the "image alignment" method.

4.2 Video Results

Fig. 2 shows a frame of an original image (a), a DR result using a raw sensor pose (b), and our DR result (c). The misalignment

due to erratic sensor data in Fig. 2b is aligned in Fig. 2c. Our system ran at 21.5 ms/frame (> 30 fps). In Fig. 2c, the ROI is indistinguishable due to simulated motion blur. We also present a supplemental video recorded in live to show that our framework is applicable under various motions and environments, such as strong translation and rotation with motion blur. The implemented DR method suffered from local illumination changes and strong surface reflections. These can be moderated using an image-based rendering approach (e.g., [8]) before the proposed post-processing.

5 CONCLUSION

In this paper, we proposed a *post-processing*-based rendering framework for DR that can be integrated to existing DR methods to improve the quality. The video results of our prototype showed that the proposed framework can reduce inconsistency between the ROI and the rest of the image in terms of motion blur and geometric misalignment. Our results were promising in the videos but limited to an indoor scenario and a qualitative perspective, which is similar to much of the DR literature [7]. In future work, therefore, we need further examinations.

REFERENCES

[1] J.-Y. Bouguet. Pyramidal implementation of the lucas kanade feature tracker description of the algorithm. 2000.

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

[3] J. Herling and W. Broll. High-quality real-time video inpainting with pixmix. *IEEE Trans. on Visualization and Computer Graphics (TVCG)*, 20(6):866–879, 2014.

[4] N. Kawai, T. Sato, and N. Yokoya. Diminished reality based on image inpainting considering background geometry. *IEEE Trans. Visualization and Computer Graphics (TVCG)*, 22(3):1236–1247, 2016.

[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. In *Proc. Int. Symp. on Mixed and Augmented Reality (ISMAR)*, pp. 11–19, 2013.

[6] C. Mei, E. Sommerlade, G. Sibley, P. Newman, and I. Reid. Hidden view synthesis using real-time visual SLAM for simplifying video surveillance analysis. In *IEEE Int. Conf. on Robotics and Automation (ICRA)*, pp. 4240–4245, 2011.

[7] S. Mori, S. Ikeda, and S. Hideo. A survey of diminished reality: Techniques for visually concealing, eliminating, and seeing through real objects. *IPSJ Trans. on Computer Vision and Applications (CVA)*, 2017. doi: 10.1186/s41074-017-0028-1

[8] S. Mori, F. Shibata, A. Kimura, and H. Tamura. Efficient use of textured 3D model for pre-observation-based diminished reality. In *Int. Workshop on Diminished Reality (IWDR)*, pp. 32–39, 2016.