3D PixMix: Image Inpainting in 3D Environments

Shohei Mori* Graz University of Technology
Jan Herling facebook Inc.
Wolfgang Broll Ilmenau University of Technology
Norihiko Kawai SenseTime Japan
Hideo Saito Keio University
Dieter Schmalstieg Graz University of Technology
Denis Kalkofen Graz University of Technology

Abstract

State of the art methods for diminished reality (DR) propagate pixel information from keyframes to later frames to achieve real-time image inpainting in 3D spaces. However, this approach assumes a planar scene and produces artifacts, if the scene geometry is not sufficiently planar. In this paper, we present 3D PixMix, a new real-time inpainting method that addresses non-planar scenes by considering both color and depth information in the inpainting process. We define cost functions for both the color and the geometric appearance in the inpainting scheme and use an RGB-D sensor for depth fusion using a SLAM. Comparison results against the conventional PixMix show that 3D PixMix obtains the equivalent or even better quality in 3D scenes with additional depth information.

Index Terms: H.5.1 [Information Interfaces and Presentation]: Multimedia Information Systems—Artificial, Augmented, and Virtual Realities ; I.4.5 [Image Processing and Computer Vision]: Reconstruction—

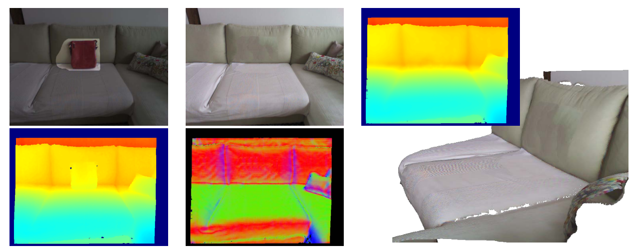
1 Introduction

Diminished reality (DR) allows removing objects from the user's perspective view and uncovering otherwise hidden structure in the user's physical environment [5]. Occluded pixels are restored from either multi-view observations or by duplicating pixels in the vicinity of the region of interest (ROI). Multi-view approaches directly observe the background at different points of view, either in preprocessing or online, using additional cameras. The resulting reconstructions represent the real situation very accurately, providing a high level of confidence and high quality renderings [1, 6]. However, multi-view DR cannot restore unobserved areas. Inpainting approaches allow to overcome this problem. They use pixels in the vicinity of the ROI and, therefore, do not require additional cameras or pre-recorded observations. If "hallucinated" pixels are acceptable, inpainting has considerable benefits over observation-based methods, in particular for mobile applications, where offline preparation is not feasible.

Inpainting naturally occurs in image space, but this can limit the supported application cases. For example, applications that require temporal or spatial coherence between frames, such as the rendering of stereoscopic images or relighting of the background, are not possible without information about the underlying 3D structure. For this reason, some inpainting systems assume a 3D space, for example, by estimating a dominant plane and performing inpainting operations on a plane embedding. If the dominant plane can be tracked throughout a sequence of frames, the inpainted images can be projected back into the user's perspective view. Such an approach is sufficient for providing plausibly DR, but only if the scene is flat and occlusions can be safely ignored. There are attempts to deform the inpainted result [3]. However, deformation changes the appearance only in the inpainted plane and never fits the geometry. As a results, inpainted regions suffer from inconsistencies.

*e-mail: s.mori.jp@ieee.org

Figure 1: 3D diminished reality by 3D inpainting and relighting.



(a) Input RGB-D frame (b) Inpainted color and normal (c) Inpainted depth and 3D reconstruction

Figure 2: 3D Inpainting. (a) Given an RGB-D frame (b), the proposed method finds globally optimized pixels in the color and normal maps (c), while PixMix and other inpaint-based DR achieves inpainting only in the color space.

In this paper, we present 3D PixMix, a system for inpainting color and depth information in occluded regions. Our DR application enables to change the visualization of the physical environment in addition to adding virtual objects (Figure 1). In this example, we removed the Stanford bunny (magenta color) from the scene by inpainting its color and depth values. Instead of the bunny, a lantern is placed to demonstrate the ability to relight the inpainted color and depth information.

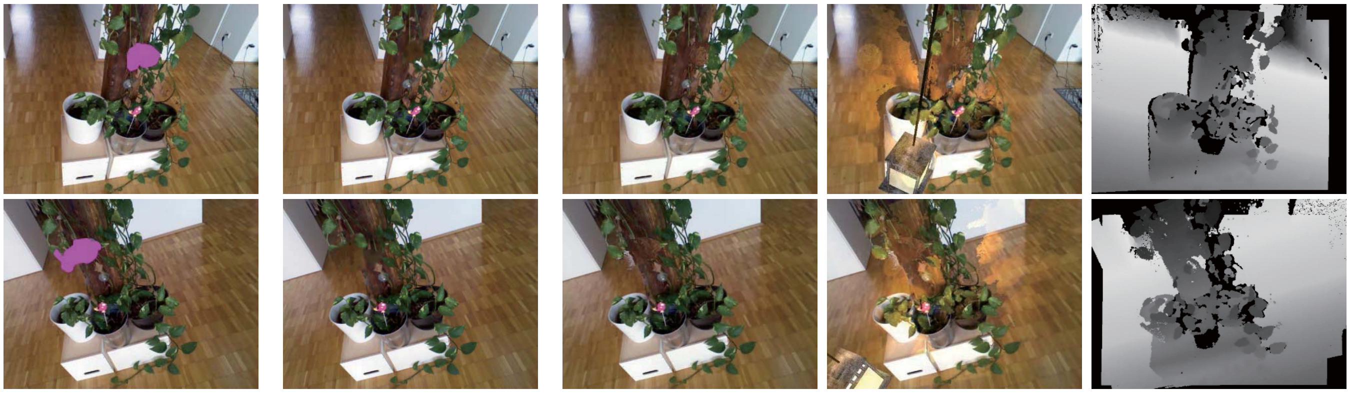
2 Method

Our systems supports per-pixel recovery of color and depth information in the unobservable region T. For a frame $F(i)$, it performs exemplar-based inpainting of a region T_i by copying information from $\overline{T}_i$, in both color and depth domains. Figure 2 illustrates our system outcomes. Our system pipeline has the following stages:

Camera pose estimation. We use simultaneous localization and mapping (SLAM) on the system's RGB-D stream to obtain a frame $F(i) = (\mathscr{C}_i, \mathscr{D}_i, \mathbf{M}_i)$, consisting of color buffer $\mathscr{C}_i$, depth buffer $\mathscr{D}_i$, and 6DOF camera pose $\mathbf{M}_i$, expressed as an $\mathbb{SE}3$ transformation matrix.

Keyframe mapping. A mapping of the transformation function f_k of a previous keyframe $F(k)$ into the current frame $F(i)$ is derived to initialize the transformation function f_i.

Inpainting. An inpainted frame $F^{in}(i) = (\mathscr{C}^{in}, \mathscr{D}^{in}, \mathbf{M}_i)$ is computed for the current frame $F(i)$ by minimizing a cost function over



Input color + mask | PixMix | 3D PixMix (Color) | 3D PixMix + AR | 3D PixMix (Depth)

Figure 3: Comparisons between PixMix [2] and 3D PixMix with additional AR rendering (a virtual lantern illuminates the inpainted area).

all pixels $\mathbf{u} \in T_i$. In addition to the texture and spatial costs [2], we also minimize a the following geometric appearance cost in a normal map $\widehat{\mathscr{N}}$ derived from the depth map $\mathscr{D}^{in}$ to obtain the smooth geometric appearance in the ROI.

$$\rho_g(f,\mathbf{u}) = \sum_{\mathbf{v}\in N_g} (1 - \widehat{n}(\mathbf{u}+\mathbf{v}) \cdot \widehat{n}(f(\mathbf{u})+\mathbf{v})), \qquad (1)$$

denoting N_g as neighboring pixels and $\widehat{n}(\mathbf{u}) \in \widehat{\mathscr{N}}$ as the normal vector at $\mathbf{u}$.

Since the normal map represents a geometric gradient map, the actual depth gradient is sampled from the original pixels. Consequently, depth gradients of the inpainted pixels are calculated from neighbors of sampled pixels in the input depth map, and solving a Poisson equation recovers the depth map.

Depth map fusion. The result of depth inpainting, $\mathscr{D}_i^{in}$, is merged into the global world model $\mathscr{G}$, which is represented as a 3D point cloud. To limit the number of pixels to be inpainted, we categorize pixels into three types. We distinguish between pixels that have been *observed* before, pixels that have been *newly-observed*, and pixels that have been *never-observed*. To always prefer observations over inpainted pixel, we store lastly observed pixel colors in the global map and such pixels will not be masked in the keyframe anymore.

Rendering. The inpainted frame $F^{in}(i)$ is used for rendering. For the final DR rendering, we project the map model into image space, where we can use it for effects such as relighting and stereo rendering of otherwise hidden structures.

3 Results

We have created a dataset of four test sequences, which all include ground truth information. Figure 3 shows one of these sequences, used to compare the results of our system to PixMix [2] and to demonstrate the possibilities of rendering additional 3D visual effects. The left column provides the input RGB image and the mask that identifies the object to be removed. The second left column show the results achieved from using PixMix. The middle column show the resulting color frame from using our approach, and the second column to the right demonstrates a simple relighting effect achieved by adding a virtual lantern to the scene. The results are best scene in color and in a supplemental video.

Figure 3 demonstrates that our approach is able to generate results comparable to state of the art methods in terms of quality. It surpasses previous inpainting by supporting additional 3D visual effects. We believe this is a major improvement for interactive DR applications and interactive image manipulation tools.

Our current implementation achieves 10 - 13 frames per second (fps) on a laptop PC using an Intel Core i7-6567U with 3.3 GHz, 16 GB RAM, and a Intel Iris Graphics 550 GPU running on a Windows 10. This performance will be improved when we combine a multi-threading approach as demonstrated by Kawai *et al.* [4].

4 Conclusion

This paper presents the first system for image inpainting in 3D environments. While previous work provides tools for generating believable DR images, our system additionally enables 3D rendering in DR. This ability highly improves the range of possible use cases for interactive DR applications, for instance, multi-view rendering for stereoscopic display devices. Furthermore, our system supports advanced image editing by its ability to add 3D visual effects to inpainted images. This enables quickly adding 3D visual effects to images and videos and thus, it provides a powerful tool for previewing image and video editing operations.

Acknowledgments

This work is partly supported by JSPS Kakenhi 16J05114, 17K12729, 18H03273, and 18K11375, and the Austrian Science Fund grant P30694, and FFG grant 859208 (MATAHARI). The authors would like to thank Ms. Mariko Isogawa (Osaka University) for her technical support.

References

[1] P. Barnum, Y. Sheikh, A. Datta, and T. Kanade. Dynamic seethroughs: Synthesizing hidden views of moving objects. In *Proceedings of IEEE ISMAR*, pp. 111–114, 2009.

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

[3] N. Kawai, T. Sato, Y. Nakashima, and N. Yokoya. Augmented reality marker hiding with texture deformation. *IEEE Trans. on Visualization and Computer Graphics (TVCG)*, 23(10):2288–2300, 2017.

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

[5] 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*, 9(17), 2017. doi: 10.1186/s41074-017-0028-1

[6] F. Rameau, H. Ha, K. Joo, J. Choi, K. Park, and I. Kweon. A real-time augmented reality system to see-through car. *IEEE Trans. on Visualization and Computer Graphics (TVCG)*, 22(11), 2016.