International Conference on Artificial Reality and Telexistence
Eurographics Symposium on Virtual Environments (2022)
R. Kondo and T. Teo (Editors)

Poster

3D-Aware Image Relighting with Object Removal from Single Image

Yujia Zhang[1] Monica Perusquia Hernandez[1] Naoya Isoyama[1] Norihiko Kawai[2]
Hideaki Uchiyama[1] Nobuchika Sakata[3] Kiyoshi Kiyokawa[1]

[1]Nara Institute of Technology, Japan
[2]Osaka Institute of Technology, Japan
[3]Ryukoku University, Japan

Abstract

We propose a method to relight scenes in a single image while removing unwanted objects by the combination of 3D-aware inpainting and relighting for a new functionality in image editing. First, the proposed method estimates the depth image from an RGB image using single-view depth estimation. Next, the RGB and depth images are masked by the user by specifying unwanted objects. Then, the masked RGB and depth images are simultaneously inpainted by our proposed neural network. For relighiting, a 3D mesh model is first reconstructed from the inpainted depth image, and is then relit with a standard relighting pipeline. In this process, removing cast shadows and sky areas and albedo estimation are optionally performed to suppress the artifacts in outdoor scenes. Through these processes, various types of relighting can be achieved from a single photograph while excluding the colors and shapes of unwanted objects.

CCS Concepts

• ***Computing methodologies*** → ***Image processing;*** • ***Human-centered computing*** → ***Mixed / augmented reality;***

1. Introduction

VR systems using real-world photographs have been spreading. For such systems, two issues must be addressed. (1) When using a photograph taken at a certain time with a certain light source, it is desirable to change the photograph's appearance so that users can experience a VR space that imitates the real world at various conditions. (2) In places such as sightseeing spots, many people can be included in the photographs. It is desirable to use such photographs as they are for the creation of some VR systems.

Each issue has conventionally been solved as follows. For (1), geometry-based and learning-based methods have been proposed for relighting a single image [EZST21] For (2), diffusion-based, patch-based, and learning-based inpainting methods have been proposed to remove unwanted objects from a single image [NNJ*19]. Joint inpainting of color and depth has also been proposed [FHS19]. However, these two issues have been studied separately and are not integrated into one framework. To the best of our knowledge, no framework solves them simultaneously for image synthesis. Such a framework is essential for creating VR systems.

We propose a method for relighting scenes in an image while removing unwanted objects in one framework. Through the framework, various types of relighting can be achieved from a single photograph while excluding the colors and shapes of unwanted objects. Finally, the processed image can be applied to VR applications, image editing, and other applications using image synthesis.

2. Proposed Method

2.1. Overview

The goal of the proposed method is to remove unwanted objects from a photograph and change lighting conditions by considering 3D geometry. The method takes a single RGB image and an object mask that the user wants to remove as input. The inpainted image is used as input for relighting so that the user is free to change the illumination to output the requested image.

The overview of our proposed method is illustrated in Fig. 1. First, the depth image is estimated from the input RGB image by using single-view depth estimation (Fig. 1(a)) with MegaDepth [LS18]. The RGB and estimated depth images are inpainted together with our proposed inpainting network (Fig. 1(c)) after removing target regions masked by the user (Fig. 1(b)). Then, optional processing is performed before relighting for the outdoor environment (Fig. 1(d)). The processing includes shadow removal, sky removal for outdoor scenes, and albedo map estimation. The 3D mesh model is generated from the depth image, and is used for relighiting with a conventional rendering pipeline (Fig. 1(e)). By feeding the image with illumination mapping, we can obtain an image relit with new illumination that users can control freely.

DOI: 10.2312/egve.20221293

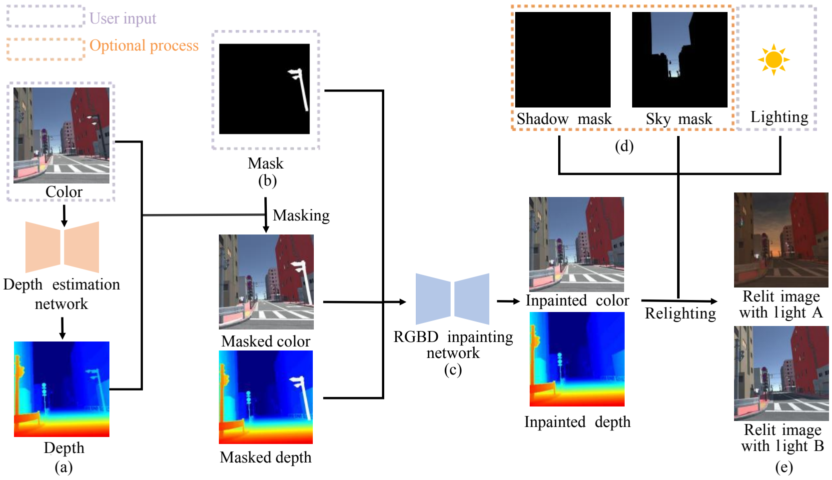


Figure 1: *Method overview.*

3. Inpainting and Relighting In Outdoor Scenes

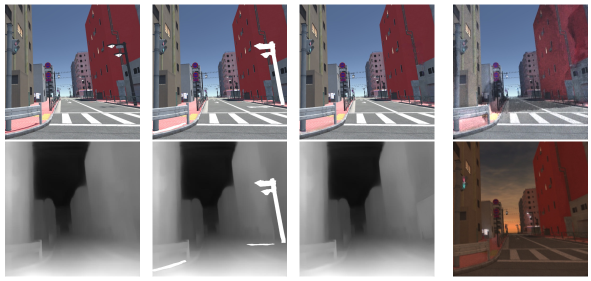
Figure 2 are the RGBD-inpainted result images. For the inpainting part, the street lights along the roadside were deleted in the outdoor scene. Here, we also marked some shadows as target regions for inpainting. We can clearly see that the street light was removed, and the area was successfully filled with the red building texture in the back. The depth map was also naturally inpainted to match the building wall shape.

For relighting, we removed the sky area by the sky mask and added a virtual sky to represent ambient light in Fig. 2. The impact of sky removal is discussed in the next section. Because sunlight can be seen as the only light source during the day in outdoor scenes, relighting is done primarily to redirect sunlight. In Fig. 2(c), by the shadow of the red building on the right, we can see that the sun is shining from almost directly above. The first row of Fig. 2(d) shows the result when the position of the sun was moved to the back of the red building. We can see that the shadow of the red building has become longer. The cast shadow covers the whole road in front of the red building. The second row of Fig. 2(d) shows the image in which the scene is illuminated by dark ambient light and city lights such as street lamps and vending machines in the evening and at night. The sky is the virtual one created by computer graphics. This way, various light source settings can convert a scene from day to evening or night.

Finally, we explain the addition of point lights. When adding a light source, the light source is not added directly to the image but to the generated model.

4. Conclusion

We proposed a method to relight the scene in an image while removing unwanted objects. The input RGB image and the mask image created by the user are the input to our system. First, the depth image is generated by using a neural network. Next, the RGBD is masked with the mask image and is inpainted using our proposed neural network. A mesh model is reconstructed from the inpainted depth image for the relighting process. In some scenes, cast shadow removal, sky region removal, and albedo estimation are selectively performed to suppress the effects of existing light sources. Through these processes, various types of relighting can be achieved from a single photograph while excluding the color and shape of unwanted objects. Future work includes improving the performance of inpainting and considering the use of neural network-based methods for relighting.



(a) *Input* **(b)** *Masked* **(c)** *Inpained* **(d)** *Relit*

Figure 2: *Comparison of relighting results and their ground truth in outdoor scene.*

References

[EZST21] El Helou, Majed, Zhou, Ruofan, Susstrunk, Sabine, and Timofte, Radu. "NTIRE 2021 depth guided image relighting challenge". *CVPR*. 2021, 566–577 1.

[FHS19] Fujii, Ryo, Hachiuma, Ryo, and Saito, Hideo. "Joint inpainting of RGB and Depth images by generative adversarial network with a late fusion approach". *ISMAR Adjunct*. 2019 1.

[LS18] Li, Zhengqi and Snavely, Noah. "Megadepth: Learning single-view depth prediction from internet photos". *CVPR*. 2018, 2041–2050 1.

[NNJ*19] Nazeri, Kamyar, Ng, Eric, Joseph, Tony, et al. "Edge-connect: Generative image inpainting with adversarial edge learning". *arXiv preprint arXiv:1901.00212* (2019) 1.