Interactive Editing of Monocular Depth

Obumneme Stanley Dukor
Simon Fraser University

S. Mahdi H. Miangoleh
Simon Fraser University

Mahesh Kumar Krishna Reddy
Simon Fraser University

Long Mai
Portland State University

Yağız Aksoy
Simon Fraser University

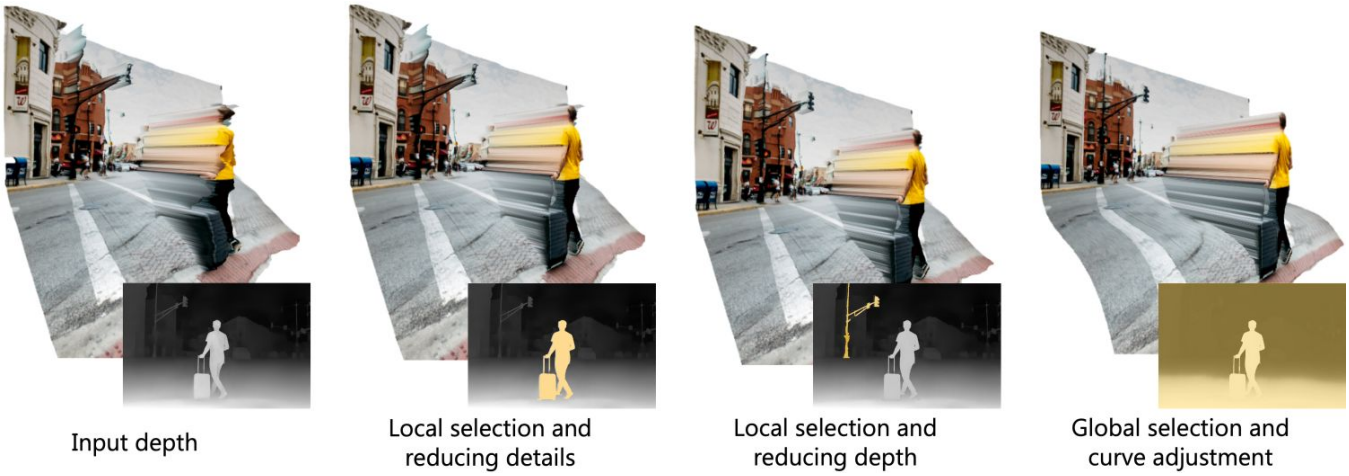

Figure 1: We propose an interactive web-based depth editing and visualization tool to perform local and global depth editing operations. From left to right, we apply iterative edits using our tool on the input depth to refine its 3D geometric properties.

ABSTRACT

Recent advances in computer vision have made 3D structure-aware editing of still photographs a reality. Such computational photography applications use a depth map that is automatically generated by monocular depth estimation methods to represent the scene structure. In this work, we present a lightweight, web-based interactive depth editing and visualization tool that adapts low-level conventional image editing operations for geometric manipulation to enable artistic control in the 3D photography workflow. Our tool provides real-time feedback on the geometry through a 3D scene visualization to make the depth map editing process more intuitive for artists. Our web-based tool is open-source and platform-independent to support wider adoption of 3D photography techniques in everyday digital photography.

ACM Reference Format:
Obumneme Stanley Dukor, S. Mahdi H. Miangoleh, Mahesh Kumar Krishna Reddy, Long Mai, and Yağız Aksoy. 2022. Interactive Editing of Monocular Depth. In *Interactive Editing of Monocular Depth (SIGGRAPH '22 Posters).* ACM, Vancouver, BC, Canada, 2 pages. https://doi.org/10.1145/nnnnnnn.nnnnnnn

SIGGRAPH '22 Posters, August 08–11,2022, Vancouver, BC

ACM ISBN 978-x-xxxx-xxxx-x/YY/MM…$15.00
https://doi.org/10.1145/nnnnnnn.nnnnnnn

1 INTRODUCTION

Recent advances in computer vision and generative networks have opened up many new applications in computational photography. Among these new applications is the structure-aware rendering of photographs such as 3D Photography [Shih et al. 2020], 3D Ken Burns effect [Niklaus et al. 2019] and Synthetic depth-of-field [Wadhwa et al. 2018]. All these applications require a *depth map* as input, which encodes the distance of objects in the image from the camera center, providing partial geometric information. These depth maps are typically computed by *monocular depth estimation* methods automatically from a single photograph.

Monocular depth estimation is typically achieved by data-driven methods that estimate the structure from a single photograph through a neural network. Currently, state-of-the-art approaches [Miangoleh et al. 2021; Ranftl et al. 2021] can generate high-resolution depth estimations with high edge accuracy that enables convincing photography applications. Despite the wide range of possible applications, however, current pipelines lack any way for an artist to make an interactive edit to the depth map which limits the adaptability of this new medium.

Depth maps are typically estimated and stored as single-channel images, as shown in Figure 1. Despite the depth map's simple representation, conventional image editing tools can not be conveniently used to edit them as it is very challenging to predict the resulting geometric effect of the edit on the structure without visualization. In this work, we present a lightweight web-based interface for interactive editing of monocular depth estimations. Our interface allows the user to make local or global adjustments to the depth map and displays the resulting 3D structure in real-time during the

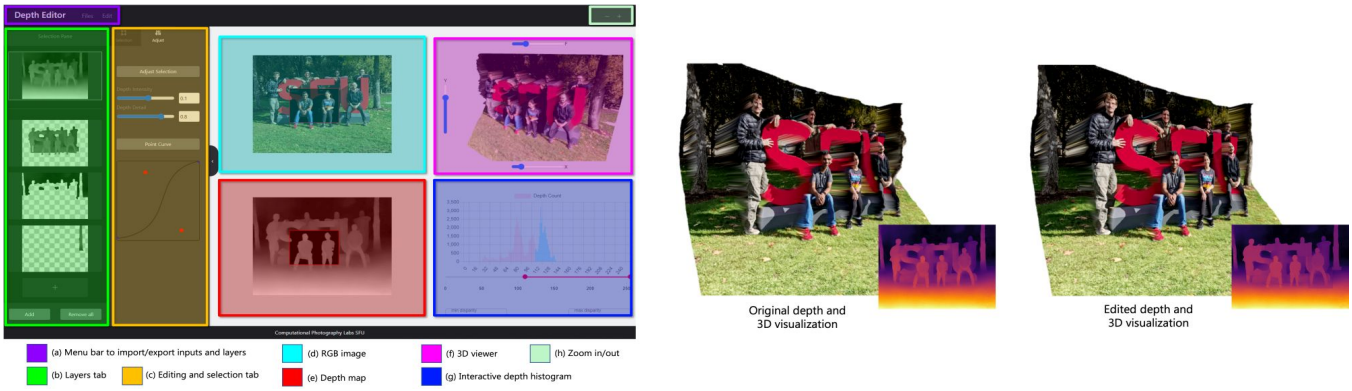

Figure 2: Left: Overview of panels in our depth editing interface. : Menu bar, : Layers tab, : Editing and selection tab, : RGB preview, : Depth preview, : Interactive 3D visualization, : Depth histogram. : Zoom-in/-out. Right: Original and edited depth maps with 3D visualizations.

editing process. The user can create different layers through our depth-based selection interface and edit their depth characteristics and 3D positions in the scene independently.

Our interface is the first tool of its kind that allows intuitive editing of the 3D data provided in the form of a 2D depth map. With the increasing performance of monocular depth estimation methods and the wide variety of down-the-line photography applications depth maps enable, we hope that our open-source interactive tool will help the wider adoption of monocular depth estimation in digital photography.

2 APPROACH

Interface. Our web-based interface allows the user to upload an RGB and monocular depth pair and allows the user to edit the depth map. The input images are displayed together with a 3D rendering of the scene given the current depth map and a histogram of the depth values. We use a layer-based editing interface where the users can generate new layers or edit depth of different layers independently.

An important part of our interface is real-time visualization of any edit on the depth map also in the 3D domain. All the visualizations in our interface get updated as the user works on their edits. This makes it intuitive to design and edit the scene structure on the gray-scale depth map with immediate feedback. An overview of the interface is illustrated in Figure 2.

Depth-based selection. High-resolution depth maps provide rich information on the scene structure and object boundaries. We take advantage of this fact to provide the user with a simple and intuitive depth-based layer selection tool. The user can draw a bounding box on the image to roughly select the object of interest. Then, the range of depth values that define the layer can be selected through our depth histogram visualization in the interface. The user gets immediate visual feedback on the current selection of the RGB image, the depth image, and the 3D visualization. Depth-based selection is a convenient alternative to traditional lasso-based object selection tools that makes use of the available depth information to cut down on total interaction time.

Global adjustment. Monocular depth estimation methods may not produce a realistic scene due to several ambiguities in the estimation process. To give the artist control over the scene structure, we provide a curve tool for the depth map using a cubic Bézier curve. The user can adjust the curve tool to apply global changes to the scene such as moving the background further away from the camera or to create more distinctive depth layering in the scene.

Local adjustments. The user can easily push an object further away or bring it closer to the camera by increasing or decreasing the depth of an individual layer. We also provide a tool for increasing the flatness of a layer in the depth domain

$$z' = \overline{x} + s(z - \overline{x}), \quad \overline{x} = mean(z), \quad s \in [0, 1],$$

where z, s denote the depth and the adjustable parameter. This tool is helpful to mitigate noisy depth estimations in flat image regions and it allows the user to create distinct flat depth layers for artistic effect.

3 CONCLUSION

Monocular depth estimation is becoming an important tool of structure-aware computational photography approaches. We present the first interactive image editing interface specifically designed for editing depth maps. Our tool allows the user to manipulate the 3D structure of the scene with real-time visualization of the geometry. We explore several low-level image editing operations and show their interpretation and use for geometric information coming in the form of a depth map. Our open-source lightweight interactive tool is a first step in integrating depth map manipulation in the traditional photo editing pipeline.

REFERENCES

S. Mahdi H. Miangoleh, Sebastian Dille, Long Mai, Sylvain Paris, and Yağız Aksoy. 2021. Boosting Monocular Depth Estimation Models to High-Resolution via Content-Adaptive Multi-Resolution Merging. In *IEEE Conf. Comput. Vis. Pattern Recog.*

Simon Niklaus, Long Mai, Jimei Yang, and Feng Liu. 2019. 3D ken burns effect from a single image. *ACM Trans. Graph.* (2019).

René Ranftl, Alexey Bochkovskiy, and Vladlen Koltun. 2021. Vision transformers for dense prediction. In *Int. Conf. Comput. Vis.*

Meng-Li Shih, Shih-Yang Su, Johannes Kopf, and Jia-Bin Huang. 2020. 3D photography using context-aware layered depth inpainting. In *IEEE Conf. Comput. Vis. Pattern Recog.*

Neal Wadhwa, Rahul Garg, David E Jacobs, Bryan E Feldman, Nori Kanazawa, Robert Carroll, Yair Movshovitz-Attias, Jonathan T Barron, Yael Pritch, and Marc Levoy. 2018. Synthetic depth-of-field with a single-camera mobile phone. *ACM Trans. Graph.* (2018).