

Deep Depth Completion of a Single RGB-D Image

Wenjie Niu

July 12, 2018

Abstract

The goal of the paper is to complete the depth channel of an RGB-D image. Commodity-grade depth cameras often fail to sense depth for shiny, bright, transparent, and distant surfaces. To address the problem, the paper train a deep network that takes an RGB image as input and predicts dense surface normals and occlusion boundaries.

1. Introduction

Depth sensing has become pervasive in applications as diverse as autonomous driving, augmented reality, and scene reconstruction. Despite recent advances in depth sensing technology, commodity-level RGB-D cameras like Microsoft Kinect, Intel RealSense, and Google Tango still produce depth images with missing data when surfaces are too glossy, bright, thin, close, or far from the camera. These problems appear when rooms are large, surfaces are shiny, and strong lighting is abundant—e.g., in museums, hospitals, classrooms, stores, etc. Even in homes, depth images often are missing more than 50% of the pixels (Figure 1)

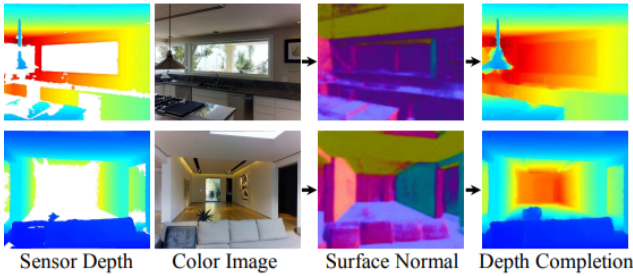


Figure 1. **Depth Completion.** We fill in large missing areas in the depth channel of an RGB-D image by predicting normals from color and then solving for completed depths [5].

Though depth inpainting has received a lot of attention over the past two decades [4], it has generally been addressed with hand-tuned methods that fill holes by extrapolating boundary surfaces [3] or with Markovian images synthesis [1]. Never methods have been proposed to estimate

depth de novo from color using deep networks [2]. The solution is to provide the network with only color images as input (Figure 2).

2. Related Work

Depth estimation. Depth estimation from a monocular color image is a long-standing problem in computer vision. Classic methods include shape-from-shading and shape-from-defocus. Other early methods were based on hand-tuned models and/or assumptions about surface orientations. Newer methods treat depth estimation as a machine learning problem, most recently using deep networks.

Depth inpainting. Many methods have been proposed for filling holes in depth channels of RGB-D images, including ones that employ smoothness priors, fast marching methods, Navier-Stokes, anisotropic diffusion, background surface extrapolation, color-depth edge alignment, low-rank matrix completion, tensor voting, Mumford-Shah functional optimization, joint optimization with other properties of intrinsic images, and patch-based image synthesis. Recently, methods have been proposed for inpainting color images with auto-encoders and GAN architectures. However, prior work has not investigated how to use those methods for inpainting of depth images. This problem is more difficult due to the absence of strong features in depth images and the lack of large training datasets, an issue addressed in this paper.

Depth super-resolution. Several methods have been proposed to improve the spatial resolution of depth images using high-resolution color. They have exploited a variety of approaches, including Markov random fields, shape-from-shading, segmentation, and dictionary methods. Although some of these techniques may be used for depth completion, the challenges of super-resolution are quite different there the focus is on improving spatial resolution, where low-resolution measurements are assumed to be complete and regularly sampled. In contrast, our focus is on filling holes, which can be quite large and complex and thus require synthesis of large-scale content.

Depth reconstruction from sparse samples. Other work has investigated depth reconstruction from color im-

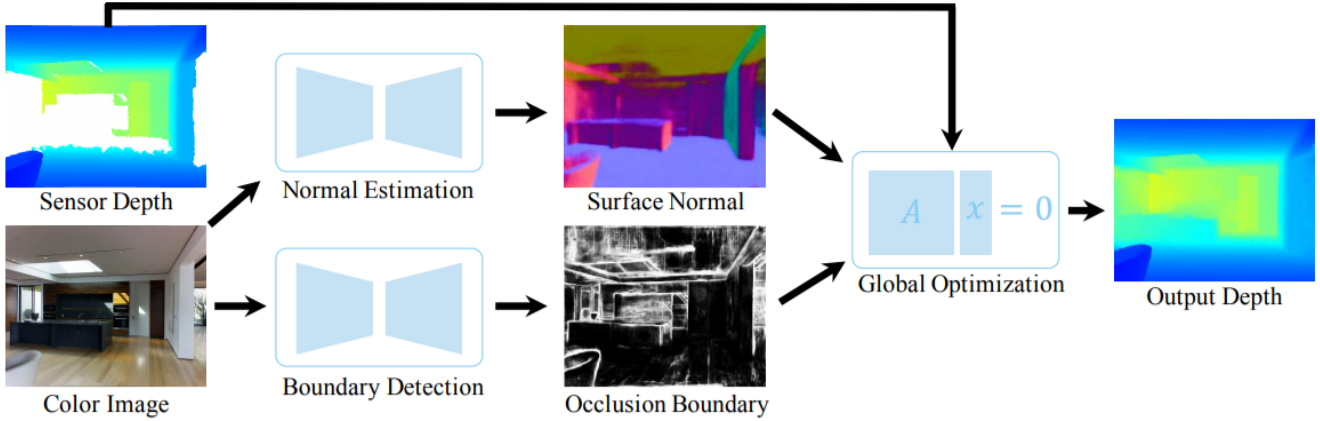


Figure 2. **System pipeline.** Given an input RGB-D image, we predict surface normals and occlusion boundaries from color, and then solve for the output depths with a global linear optimization regularized by the input depth[5].

ages augmented with sparse sets of depth measurements.

References

- [1] D. Doria and R. J. Radke. Filling large holes in lidar data by inpainting depth gradients. In *IEEE Conference on Computer Vision and Pattern Recognition Workshops*, 2012. 1
- [2] D. Eigen, C. Puhrsch, and R. Fergus. Depth map prediction from a single image using a multi-scale deep network. In *Advances in Neural Information Processing Systems*, 2014. 1
- [3] K. Matsuo and Y. Aoki. Depth image enhancement using local tangent plane approximations. In *IEEE Conference on Computer Vision and Pattern Recognition*, 2015. 1
- [4] J. Uhrig, N. Schneider, L. Schneider, U. Franke, T. Brox, and A. Geiger. Sparsity invariant cnns. In *International Conference on 3D Vision*, 2017. 1
- [5] Y. Zhang and T. Funkhouser. Deep depth completion of a single rgb-d image. In *IEEE Conference on Computer Vision and Pattern Recognition*, 2018. 1, 2