Abstract

Active depth sensors such as Microsoft Kinect and LiDAR have become integral to computer vision and robotics, enabling real-time acquisition of geometric information for 3D scene understanding. Despite their advantages, active depth sensors face fundamental limitations that compromise the reliability and utility of the depth maps. Kinect depth maps suffer from missing values caused by signal absorption on dark surfaces, scattering due to specular and translucent materials, and occlusion due to the physical displacement between the IR projector and camera. LiDAR sensors are precise and suitable for long-range outdoor sensing, but produce geometrically sparse point clouds due to limited horizontal scan lines, with sparsity increasing at distance. Existing depth completion methods present several compounding limitations. Firstly, classical unguided methods are computationally expensive, prone to geometric distortion, and produce excessive smoothing around depth discontinuities. Deep learning-based unguided completion from a single degraded depth map remains a severely ill-posed and underexplored problem. Secondly, while RGB-guided deep learning methods have advanced outdoor LiDAR depth completion, existing approaches adopt a one-stage design that performs dense depth recovery in a single forward pass, requiring computationally intensive fusion schemes to achieve competitive accuracy. Thirdly, both guided and unguided learning-based approaches depend on large annotated domain-specific training datasets, resulting in models that struggle to generalize across domains without retraining.

This thesis addresses the above-mentioned limitations through a progressive series of depth completion frameworks. Initially, in Chapter 2, a lightweight generative adversarial network (GAN) is proposed for unguided indoor depth completion. The framework incorporates a novel variant of the Huber loss, termed the L1-L2 loss, combined with a multi-scale structural similarity (MS-SSIM) perceptual loss within a hybrid objective function. Its practical utility is evaluated in Chapter 3, with two downstream tasks, namely novel view synthesis and augmented reality, demonstrating the potential for scalable depth-aware technologies. In Chapter 4, the problem of outdoor depth completion for LiDAR sensors is addressed by a proposed two-stage deep convolutional neural network (DCNN). Shifting to long-range depth maps introduces additional complexity, which is mitigated by RGB image guidance in this chapter. The first stage produces a coarse, dense depth estimate, and the second stage refines it using a computationally efficient iterative feedback mechanism. Finally, in Chapter 5, a unified zero-shot framework is proposed for depth completion across varied depth ranges without large-scale annotated datasets or domain-specific fine-tuning. In this chapter, a dual-branch convolutional neural network-vision transformer (CNN-ViT) architecture is proposed, which processes RGB image and sparse depth inputs in parallel and is optimized via a joint loss that combines edge-aware smoothness, superpixel-wise planar depth consistency, surface normal alignment, and a cross-modal guidance loss.

The proposed unguided method, in Chapter 2, achieves robust zero-shot generalization with up to 90% missing pixels using approximately 1.4-1.8 million parameters and outperforms several guided and unguided state-of-the-art methods. Integration of the lightweight model into novel view synthesis and augmented reality pipelines yields measurable improvements across all evaluation metrics in Chapter 3. The two-stage LiDAR depth completion framework produces competitive results on the KITTI depth completion benchmark in Chapter 4. The zero-shot framework in Chapter 5 demonstrates superior structural coherence and depth accuracy under stricter evaluation thresholds than supervised methods.

In summary, this thesis advances the state of depth completion through four interrelated contributions, spanning unguided indoor, RGB-guided outdoor, and zero-shot cross-domain settings. A lightweight GAN for unguided indoor depth completion is proposed and validated through novel view synthesis and augmented reality applications, followed by a two-stage RGB-guided framework for outdoor LiDAR completion. Finally, a unified zero-shot CNN-ViT framework is proposed, eliminating the need for domain-specific training. The key findings establish a foundation for efficient depth completion in real-world environments.

Keywords: *Depth map completion*, *Microsoft Kinect sensor*, *LiDAR scanner*, *RGB-D data*, *generative adversarial network (GAN)*, *coarse depth generation*, *iterative depth refinement*, *zero-shot depth completion*.