High-Resolution Depth Estimation for 360◦ Panoramas through Perspective and Panoramic Depth Images Registration

Chi-Han Peng
National Yang Ming Chiao Tung University
www.pengchihan.co
pengchihan@nycu.edu.tw

Jiayao Zhang
King Abdullah University of Science and Technology / ByteDance
jiayao.zhang@bytedance,com

TL;DR: better stitching of multiple perspective depth maps to form a panoramic depth map

We generate high-resolution and highly detailed depth maps for panoramas through merging many perspective depth maps (generated by existing methods such as LeReS and MiDaS) and a panoramic depth map (generated by existing methods such as SliceNet and UniFuse) serving as the *global "reference" map* to guide the merging. Our method outperformed a previous stitching-based method quantitatively, qualitatively, and by speed.

MOTIVATION

We find state-of-the-art perspective depth estimation methods can outperform panoramic methods by large margins (see right insert). But if we simply stitch them together (in equirectangular view), the results are unusable as they would be inconsistent to each other:
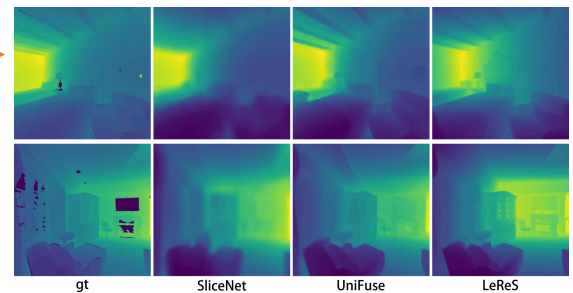
360MonoDepth (CVPR 2022) proposed a Poisson-blending based approach to merge the perspective maps with success, but suffered from huge computational costs, and has occasional qualitative quality issues.
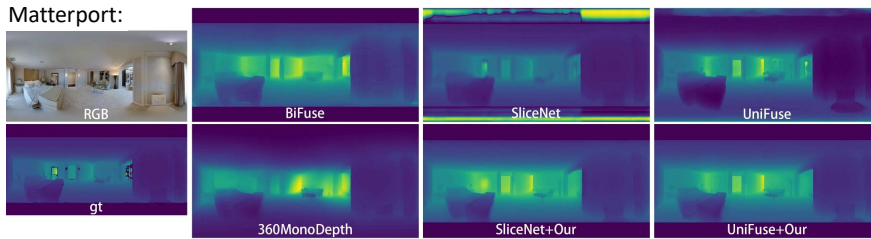
OUR METHOD

We propose a simple solution to the consistency problem: *leveraging a panoramic depth map as the common target for the perspective depth maps to register to*. Note that such "reference" maps can be generated in real-time by an existing panorama-based method, which is known to produce globally consistent values. Our pipeline is as follows.
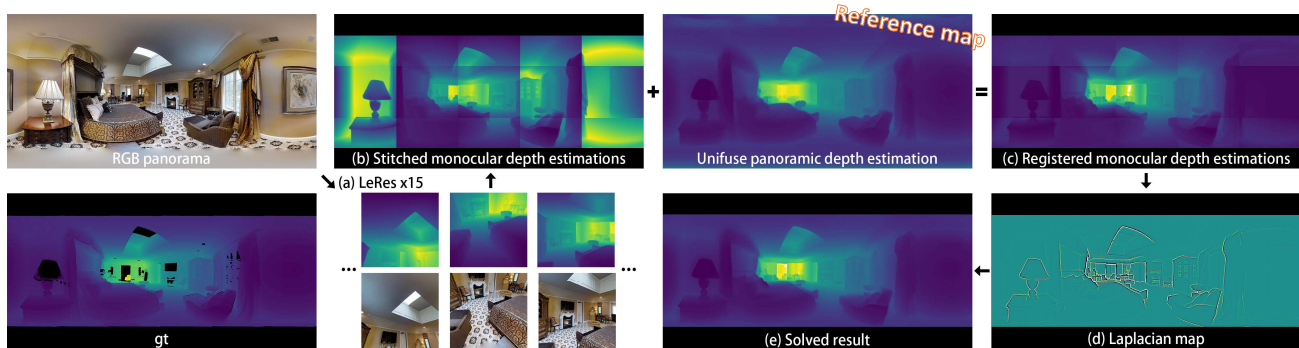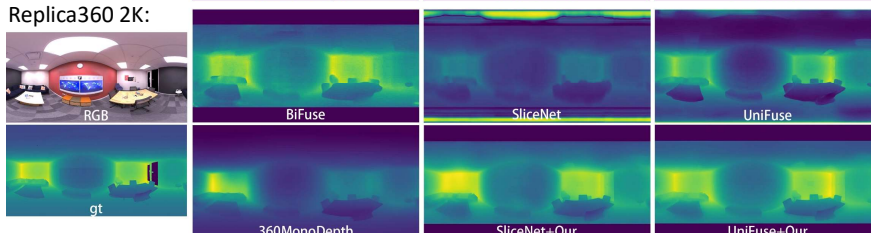

RESULTS

Our results outperform 360MonoDepth quantitatively, qualitatively, and is significantly faster, for both 2K (Matterport dataset) and 4K (Replica360 dataset) outputs. We also outperformed existing end-to-end methods (up to year 2021).

Matterport:

Replica360 2K:

BEING "FUTURE-PROOF"

Our method essentially merges perspective and panoramic depth maps generated by external methods. It is straightforward to switch to newer methods for both kinds of inputs.

ACKNOWLEDGMENTS