Pacific Graphics (2021)
M. Okabe, S. Lee, B. Wuensche and S. Zollmann (Editors)

User-centred Depth Estimation Benchmarking for VR Content Creation from Single Images

Anthony Dickson [1] Alistair Knott[1] Stefanie Zollmann[1]

[1]University of Otago, New Zealand

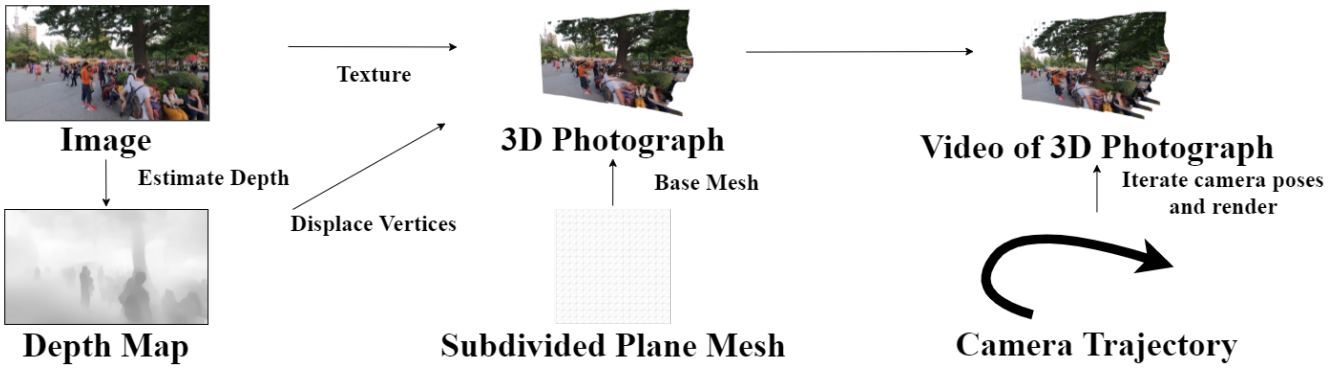

Figure 1: *The process for creating 3D photographs from 2D images. The 3D photographs are created by taking a RGB-D image and texturing and displacing the vertices on a subdivided plane mesh. A 3D photograph dataset is then created by using a predefined camera trajectory.*

Abstract

The capture and creation of 3D content from a device equipped with just a single RGB camera has a wide range of applications ranging from 3D photographs and panoramas to 3D video. Many of these methods rely on depth estimation models to provide the necessary 3D data, mainly neural network models. However, the metrics used to evaluate these models can be difficult to interpret and to relate to the quality of 3D/VR content derived from these models. In this work, we explore the relationship between the widely used depth estimation metrics, image similarly metrics applied to synthesised novel viewpoints, and user perception of quality and similarity on these novel viewpoints. Our results indicate that the standard metrics are indeed a good indicator of 3D quality, and that they correlate with human judgements and other metrics that are designed to follow human judgements.

CCS Concepts

• ***General and reference*** → ***Evaluation;*** • ***Computing methodologies*** → *Computer graphics; Neural networks; Computer vision;*

1. Introduction

Estimating depth from a single image without additional sensors is a challenging task that has applications in many areas ranging from driver-less cars to image editing. Recently, depth estimation from single images has been used for 3D content creation such as for creating 3D photographs [KMA*20].

Depth estimation models are commonly evaluated with the following metrics:

- accuracy under a threshold: the percentage of the estimated depth $\hat{d}_i$ such that $\max(\frac{\hat{d}_i}{d_i}, \frac{d_i}{\hat{d}_i}) = \delta <$ threshold for threshold values of 1.25, 1.25^2 and 1.25^3 (we refer these to as DEL1, DEL2 and DEL3 in the results)
- mean relative error (REL) $\frac{1}{N}\sum_{i=1}^{N}\frac{||d_i-\hat{d}_i||_1}{d_i}$
- root mean squared error (RMSE) $\sqrt{\frac{1}{N}\sum_{i=1}^{N}(d_i-\hat{d}_i)^2}$.

The main reason for using these metrics is that that an early focus of monocular depth estimation (MDE) was on autonomous navi-