Complementary Depth Integration Using Convolutional AutoEncoder

Mitsuhiro Murata, Chun Xie, Hidehiko Shishido, Takeshi Endo, and Itaru Kitahara

Abstract— **In this paper, we present a novel image-based depth estimation method that synergistically combines the strengths of stereo matching and deep learning-based techniques, offering an integrated solution to overcome their individual limitations. Our method is based on Convolutional AutoEncoder, which complementary integrates the two types of depth data. The proposed method can accurately estimate depth information even in regions where stereo matching struggles to measure depth, such as occluded and texture less regions. It has been observed that depth estimation results using deep learning are inaccurate for objects like building contours, car edges, and road signs. On the other hand, the depth estimation results obtained using the proposed method can predict depth variations in such areas and provide detailed object shapes.**

I. Introduction

Image-based depth estimation has been widely used for many applications such as augmented reality and auto driving due to its low device requirement and high robustness towards dynamic environments. Conventionally, this was usually achieved by stereo vision. However, stereo matching approaches are known to be error prone in occlusion, texture-less, or patterned texture regions. In contrast, deep learning based monocular depth estimation methods have recently attracted attention. However, estimation accuracy decreases when the image appearance differs from the training data. In this paper, we propose a depth estimation method that combines the respective strengths of stereo matching and deep learning based methods, offering an integrated solution to overcome their individual limitations.

II. Proposed Method

This paper proposes a method to integrate Semi Global Matching (SGM) technique [1], a robust method to obtain dense disparity estimates in stereo vision, with Deep Ordinal Regression Network (DORN) [2], a supervised deep learning approach for monocular depth estimation.

SGM, while effective, may encounter estimation errors in areas obscured from one of the stereo cameras, such as occluded regions, or in texture smaller regions, such as building windows. DORN provides better depth information in these challenging areas. However, it has been observed that DORN exhibits reduced depth clarity around object contours.

Figure 1 illustrates the architecture of our proposed deep learning model. The model utilizes depth estimates from both DORN and SGM as inputs. Convolutional Neural Networks (CNNs) then process these inputs, generating feature maps that match the dimensions of each depth map. These feature maps are concatenated along the channel dimension and fed into a CNN Autoencoder. The Autoencoder produces an integrated depth map at its decoder side. In order to further refine depth estimation, the depth image from DORN is incorporated into the output of Autoencoder via a skip connection. Training a model is one by minimizing the Root Mean Squared Error (RMSE) between the output of the network and the ground truth depth map.

Our results demonstrate a lower RMSE for the integrated method compared to the independent SGM or DORN estimates. Moreover, a qualitative evaluation of the depth map (Figure 2) reveals estimations of our method closely aligning with the Ground Truth. This underscores the effectiveness of our approach in harnessing the complementary strengths of stereo matching and deep learning for improved depth estimation.

References

[1] Heiko Hirschmüller "Stereo Processing by Semiglobal Matching and Mutual Information" IEEE TRANSACTIONS ON PATTERN ANALYSIS AND MACHINE INTELLIGENCE 2008

[2] Huan Fu. Mingming Gong. Chaohui Wang. Kayhan Batmanghelich. Dacheng Tao" Deep Ordinal Regression Network for Monocular Depth Estimation" IEEE Conference on Computer Vision and Pattern Recognition (CVPR) 2018.

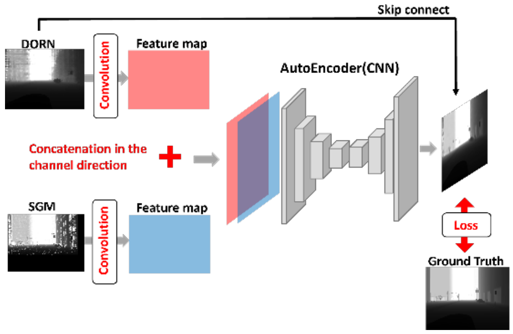

Figure 1: Integration method using depth information estimated by DORN and SGM.

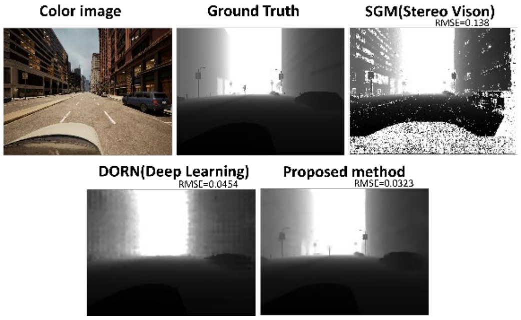

Figure 2: Depth estimation results and RMSE values for each method

M. Murata C.Xie H. Shishido and I. Kitahara are with University of Tsukuba (email: kitahara[at]ccs.tsukuba.ac.jp) T. Endo is Hitachi, Ltd.