Deep Relative Pose Estimation for Stereo Camera
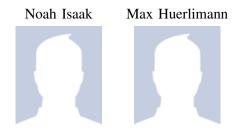
3D Vision Project Proposal
Supervised by: Zhaopeng Cui
March 9, 2018

GROUP MEMBERS

Noah Isaak   Max Huerlimann

I. DESCRIPTION OF THE PROJECT

This project aims at designing a neural network which can can estimate the depth and relative pose through pictures attained with a stereo camera. It is inspired by the work done with an unsupervised network for single-view monocular cameras [2]. The depth estimation has been shown to work well with a stereo configuration [1]. For training the KITTI dataset will be used, which provides scenes captured with stereo cameras. If the time allows, there will be attempted to fit a mask as well.

II. WORK PACKAGES AND TIMELINE

To get familiar with the recent work, at first familiarization with the state-of-the-art algorithms will take place. Then, adaption of the algorithm used in [2] to a stereo camera, using DispNet [1] will be done. Training of the network through the KITTI dataset (and possibly others) will be done, preferably without groundtruth in an unsupervised manner.

Detailed descriptions of work packages you planned, their outcomes, the responsible group member and estimated timeline. Specify the challenges that will be tackled and considered solutions with possible alternatives, citing related documents if applicable. Mention the platform (Android, PC etc.) and the language (C++ etc.) you plan to use.

III. OUTCOMES AND DEMONSTRATION

The main expected outcome is an improvement of depth perception with respect to the single view monocular algorithm. This could lead to a general improving of identification of relative pose estimation and possibly semantics. This can be demonstrated on available datasets or possibly live, if there is a stereo camera available. Give detailed information on the expected outcome of your project and the experiments you plan to test your implementation. If applicable, describe the online or offline demo you plan to present at the end of the semester.

Instructions:
- The document should not exceed two pages including the references.
- Please name the document **3DVision_Proposal_Surname1_Surname2.pdf** and upload it via the moodle.

REFERENCES

[1] Benjamin Ummenhofer, Huizhong Zhou, Jonas Uhrig, Nikolaus Mayer, Eddy Ilg, Alexey Dosovitskiy, and Thomas Brox. Demon: Depth and motion network for learning monocular stereo. *CoRR*, abs/1612.02401, 2016.
[2] Tinghui Zhou, Matthew Brown, Noah Snavely, and David G. Lowe. Unsupervised learning of depth and ego-motion from video. In *CVPR*, 2017.