THESIS SUPERVISOR'S REPORT

I. IDENTIFICATION DATA

Thesis title:	**Combining Monocular Depth Estimation with 2D-3D Correspondences**
Author's name:	**Martin Koudelka**
Type of thesis :	master
Faculty/Institute:	Faculty of Electrical Engineering (FEE)
Department:	Department of Cybernetics
Thesis reviewer:	RNDr. Zuzana Kúkelová PhD.
Reviewer's department:	Department of Cybernetics

II. EVALUATION OF INDIVIDUAL CRITERIA

Assignment **extraordinarily challenging**

How demanding was the assigned project?

The project required the student to familiarize himself with multiple areas of computer vision, ranging from classical topics such as camera models, calibration, coordinate system transformations, and Structure-from-Motion, to modern deep learning-based techniques (monocular depth estimation). The student needed to familiarize himself with the literature on monocular depth estimation, select an approach suitable for the task, modify the implementation of the approach for the purposes of the project, find and modify a suitable dataset, train the modified approach on the resulting data, and evaluate and interpret the results. This combination of getting acquainted with a range of topics, modifying an approach, creating suitable data, and designing and running experiments can require some time and can be challenging even for PhD students. It certainly is challenging for MSc students.

Fulfilment of assignment **fulfilled**

How well does the thesis fulfil the assigned task? Have the primary goals been achieved? Which assigned tasks have been incompletely covered, and which parts of the thesis are overextended? Justify your answer.

The assigned tasks were: 1) Familiarize yourself with the relevant literature. 2) Investigate approaches for using 2D-3D matches for fine-tuning monocular depth estimators. 3) Investigate approaches for combining 2D-3D matches with monocular depth estimation for camera geometry estimation. 4) Investigate the approaches on real-data.
As evident from the detailed discussion of related work, the student clearly fulfilled task 1. The experimental results on real data show that the chosen approach for using 2D-3D correspondences for fine-tuning a monocular depth predictor has great promise. Thus, it is clear that the student fulfilled tasks 2) and 4). The experimental results show that it is possible to fine-tune a monocular depth predictor to improve depth prediction performance from only a few 2D-3D correspondences. We recently published a range of methods for camera geometry estimation that use monocular depth predictions. We observed that better depth predictions lead to better geometry estimations. Some of these solvers furthermore require metric or scale-invariant (instead of affine-invariant) depth predictions. The method developed in this thesis clearly improves the depth predictions, and potentially upgrades affine-invariant to metric/scale-invariant depth estimates. It can hence be used to improve the performance of these solvers for camera geometry estimation. Thus, the student also fulfilled task 3.

Activity and independence when creating final thesis **A - excellent.**

Assess whether the student had a positive approach, whether the time limits were met, whether the conception was regularly consulted and whether the student was well prepared for the consultations. Assess the student's ability to work independently.

Martin worked very independently and required very little guidance. He chose the depth predictor used in the thesis himself after a detailed review of monocular depth predictors. Based on higher-level discussions in meetings, he was able to design and implement modifications to the chosen approach, as well as design and carry out experiments. He was well-prepared for our meetings and not afraid to contact me outside scheduled meeting times if he needed help or feedback. For our meetings, he typically had a set of recent results prepared that we could go over and discuss. In addition, he typically had a set of questions prepared. He showed the ability to work independently to a level typical of PhD students and beyond what I typically observe for Master's students.

Technical level	**A - excellent.**
Is the thesis technically sound? How well did the student employ expertise in his/her field of study? Does the student explain clearly what he/she has done?	
The thesis is technically sound and builds upon the current state-of-the-art in the field. The thesis clearly explains what the student was doing and why he was doing it. The design of the experiments presented in the thesis is very good, and the student provides good interpretations of the results. Overall, the student was clearly able to use the expertise on camera models, camera calibration, coordinate transformations, and learning-based camera pose and monocular depth estimation, which he acquired at the beginning of the project. The technical level of the thesis is clearly excellent.	

Formal level and language level, scope of thesis	**A - excellent.**
Are formalisms and notations used properly? Is the thesis organized in a logical way? Is the thesis sufficiently extensive? Is the thesis well-presented? Is the language clear and understandable? Is the English satisfactory?	
The thesis is well-organized, following a standard structure. It is well-written and easy to follow. The student clearly motivates each part and nicely explains the relation between the thesis and prior work. The part on related work is very detailed, with a nice explanation of the main directions of existing methods, a detailed review of recent state-of-the-art methods, as well as of evaluation metrics typically used. The experimental evaluation is extensive, and the thesis provides a detailed discussion of the results.	

Selection of sources, citation correctness	**A - excellent.**
Does the thesis make adequate reference to earlier work on the topic? Was the selection of sources adequate? Is the student's original work clearly distinguished from earlier work in the field? Do the bibliographic citations meet the standards?	
The thesis contains a detailed description of prior work (11 pages), covering a wide range of topics (including the theory behind camera models, traditional and very recent monocular depth estimation, a discussion of metric vs. affine-invariant depth predictions, applications and datasets for depth prediction, evaluation metrics, and a detailed description of recent state-of-the-art depth prediction algorithms). The references to earlier work and the selection of sources are more than adequate. The bibliographic citations meet the standards, and the student clearly described the differences between his approach and prior work.	

Additional commentary and evaluation (optional)
Comment on the overall quality of the thesis, its novelty and its impact on the field, its strengths and weaknesses, the utility of the solution that is presented, the theoretical/formal level, the student's skillfulness, etc.
See comments below.

III. OVERALL EVALUATION, QUESTIONS FOR THE PRESENTATION AND DEFENSE OF THE THESIS, SUGGESTED GRADE

Summarize your opinion on the thesis and explain your final grading.

Overall, this is an excellent thesis. Martin was able to clearly understand a complex field and derived a simple solution that allowed him to significantly improve the quality of monocular depth prediction from only a few 2D-3D correspondences. His results show that existing approaches still struggle to predict metric depth for previously unseen scenes. His thesis proposes a simple solution that can improve performance by fine-tuning the depth predictor using only few data. This opens up a wide range of applications, e.g., in the fields of robotics and augmented reality. In these applications, it is common to have a sparse 3D model of the scene obtained from images. Using the approach developed in this thesis, this sparse model can be used to improve a predictor for monocular dense depth estimates. These depth predictions can then be used for collision avoidance and path

planning for robots or for occlusion handling in augmented reality. As such, I believe that Martin's results have the potential for significant impact. Certainly, the results will inform my future research, and I believe that the thesis forms the basis for a good publication at a top-ranked conference on 3D computer vision.

One potential question for the thesis defense is: Can you think of a way to select a small subset of 2D-3D correspondences from a Structure-from-Motion model that are of high quality and can thus be used to fine-tune the depth predictor?

The grade that I award for the thesis is **A - excellent.**

Date: **13.6.2025** Signature: