CTU CZECH TECHNICAL UNIVERSITY IN PRAGUE

THESIS REVIEWER'S REPORT

I. IDENTIFICATION DATA

Thesis title:	**Matching of multimodal features**
Author's name:	**Martin Fischer**
Type of thesis :	master
Faculty/Institute:	Faculty of Electrical Engineering (FEE)
Department:	Department of Cybernetics
Thesis reviewer:	Ruslan Agishev
Reviewer's department:	Department of Cybernetics

II. EVALUATION OF INDIVIDUAL CRITERIA

Assignment — challenging

How demanding was the assigned project?

The assignment of matching of features from RGB images and point clouds is extremely relevant in robotics and remains challenging in computer vision. It requires understanding of the sensory data processing and fusion techniques. The project can find its application in visual localization as well as sensor calibration tasks.

Fulfilment of assignment — fulfilled with major obje

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.

All the project assignment individual steps were covered in the thesis. However, there is still room for improvement. For example, the thesis focuses only on simulated data not addressing the real sensor noise and performance under challenging weather conditions. The limitations are mentioned in the paper, but it is not discussed how one can address them.
The Experiments section could be improved. Dealing with multiple sensors, one must provide more details about the utilized sensor setup (placement, specifications) even though only the simulated data were used. Evaluation of the algorithm at a single trajectory (around 50 m long) does not seem satisfactory given that the simulator allows to generate enormous amount of data. Details of the localization accuracy experiment design should have been provided, for example:
- are evaluation data frames defined by camera frequency?
- is lidar (depth camera) synchronized with the RGB camera? If not how could it be addressed?
- is only the last frame or the map (aligned frames history) is used during the ICP alignment?
I would recommend to add reconstruction error (between the obtained and GT map) as evaluation metric. The resultant path examples must be plotted as well to explain qualitatively the error values (quite large, ~0.5 m) given in tables and figures. It is not clear from the results if the method utilizing different feature extraction methods for neighboring frames is working. Computational complexity is discussed but numerical results of the algorithm performance are not provided.

Methodology — partially applic

Comment on the correctness of the approach and/or the solution methods.

While the feature extraction methods were chosen reasonably and described well, the feature association algorithm requires more attention. It is not obvious and not covered in the thesis if the ICP algorithm will be able to provide correct transformation for features extracted from different sensors (RGB image and a point cloud). It can happen that there are no close enough (in Eucledean distance) correspondences between the obtained features. I assume that was one of the reasons of fairly large errors (RMSE and APE) described in the Experiments section. Probably projection of lidar (depth camera) points to RGB image frame and comparison of them to the

corresponding intensity values can be selected as a distance metric between the multimodal features.

Technical level C - good.

Is the thesis technically sound? How well did the student employ expertise in the field of his/her field of study? Does the student explain clearly what he/she has done?

The chosen methodology is described clearly. However, the results are given "as is" without elaboration on what is the cause and meaning of the obtained metrics. For example, how is the error of 52.2 m in Table 6.4 can be explained for the path of around ~50 m long. A comparison to a baseline from the literature is beneficial, for example:

Mur-Artal, R., Montiel, J.M.M. and Tardos, J.D., 2015. ORB-SLAM: a versatile and accurate monocular SLAM system. IEEE transactions on robotics, 31(5), pp.1147-1163.

Formal and language level, scope of thesis B - very good

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 in general organized well. However, additional details should have been provided, for example the description of the ICP algorithm. It will help the reader to connect the individual components of the work, like feature extraction, matching and obtaining the localization information relative to a map (first frame) in terms of visual SLAM. There are minor English mistakes (annotated in the text). However, it does not affect the readability. Confusing sentences and not-justified assumptions should be avoided.

Selection of sources, citation correctness B - very good

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 Introduction section requires adding references to certain sentences (annotated in the attached pdf-file).
Several sentences are irrelevant to the thesis topic (more are highlighted in the attachment), for instance:
- "The spatial relationship between these modalities allows for creating compelling AR experiences…"
- "Consequently, combining these diverse data sources to derive meaningful semantic insights is a complex and multifaceted task"

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.

Maybe utilization of differentiable algorithm as a way to learn visual features relevant for sensor localization would be beneficial. One can imaging running a feature extraction (function approximator mapping sensor data to visual features) alongside with differentiable SLAM, I.e: https://github.com/gradslam/gradslam. The localization error signal could be back-propagated to learn the function approximator. In this case the simulator benefit (or an accurate GT localization source) is clear.

The annotated thesis is available at:
https://drive.google.com/file/d/1MSUZlfprgIu0uVjvUlN3ZoaxSiPRNFOz/view?usp=sharing

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. Pose questions that should be answered during the presentation and defense of the student's work.

- Explain the ICP and Absolute Orientation algorithms,
- How to improve the algorithm with the presence of noise?

The grade that I award for the thesis is C - good.

Date: **Click here and enter the date.** Signature: