CZECH INSTITUTE OF INFORMATICS,
ROBOTICS AND CYBERNETICS
Dr. Torsten SATTLER
Senior Researcher

June 28, 2021, Prague
Ref. No. 81/21/37241/TSat

**Evaluation of the PhD Thesis "Visual Localization in Natural Environments" by Jan Brejcha**

To Whom It May Concern:

The topic of Mr. Jan Brejcha's PhD thesis is visual localization in natural, and more specific mountainous, environments. Visual localization is the problem of precisely estimating the position and orientation from which a given image was taken in a known scene, typically using some form of 3D model as a scene representation. Visual localization is an important problem in the research fields of computer vision and robotics, with practical applications including, but not limited to, augmented and mixed reality, virtual reality, and autonomous robots such as self-driving cars and drones. As such, advances in visual localization are not only of interest to the academic community but also to companies such as Microsoft, Google, Facebook, and companies working on autonomous driving.

Most of the literature on visual localization focuses on urban environments, with a few vegetated environments being recently included in commonly used benchmarks. In contrast, there is very little research on localization in mountainous environments, which are the focus of Mr. Brejcha's thesis. In many ways, localization in mountainous terrain is much harder than localization in urban scenes: (1) urban scenes typically contain many static and well-textured surfaces, which can be used to robustly and reliably detect local features. (2) the set of admissible viewpoints is much smaller in urban environments, making it much easier to sufficiently cover scenes with images. In particular, images are taken much closer to the scene in urban environments compared to mountainous scenes. (3) in general, much less data is available for mountainous environments, preventing the use of commonly applied machine learning techniques such as deep learning due to a lack of training data. At the same time, there are interesting applications for visual localization in mountainous areas, as evident from the applications developed in this thesis. Overall, Mr. Brejcha's thesis thus tackles an important and challenging problem that has not received a lot of attention previously. The topic is appropriate for the area of the dissertation.

The thesis consists of three main parts that follow an introduction that clearly introduces the problem setting considered in the thesis, describes the challenges of visual localization in mountainous terrain, and summarizes the main contributions of the thesis.

- Part I consists of a survey of state-of-the-art visual localization techniques (Chapter 2), including a classification of these approaches based on where they are applied (urban scenes, natural scenes, etc.) and which data modalities they rely on, and an introduction to datasets and evaluation measures commonly used in the literature (Chapter 3). Chapter 2 is based on a paper published by Mr. Brejcha in the Pattern Analysis and Applications journal in 2017.

- Part II focuses on the problem that there are no suitable large-scale datasets for visual localization in mountainous environments available in the literature and develops multiple approaches to create such datasets:
    - Chapter 4, based on a paper published in the Image and Vision Computing journal in 2017, introduces the GeoPose3K dataset that is later on used for evaluating the localization approaches developed in this thesis. The dataset is the first dataset for visual localization in mountainous terrain that provides both camera positions and orientations for evaluating localization accuracy. The chapter describes the construction of the dataset, details its characteristics, and compares it against previous datasets by evaluating a baseline method on all datasets. I consider the GeoPose3K dataset to be one of the main contributions of the thesis and a great benchmark for future work on localization in mountainous scenes.
    - During the creation of the GeoPose3K dataset, it was assumed that the geo-positions of the images in the dataset are known. Chapter 5, based on parts of a publication at the European Conference on Computer Vision (ECCV, one of the top-tier conferences in the field of computer vision) in 2020, presents two automated strategies to remove this assumption. Both are based on Structure-from-Motion (SfM), with the first aligning the 3D model computed using SfM to a digital elevation model (DEM) and the second using virtual renderings of the DEM (with aerial images projected to the terrain model to provide texture), with known poses with respect to the DEM, during the reconstruction process to allow geo-registration and prevent drift. Compared to the approach used in Chapter 4 to create the GeoPose3K dataset, both methods should scale better to larger dataset and are easier to automate. As such, the chapter contributes important methods for creating datasets for visual localization in mountainous environments.

- Part III develops visual localization algorithms for localization in mountainous terrain and applications based on these algorithms:
  - Chapter 6, based on a paper published at the International Conference on 3D Vision in 2018, develops an approach for estimating the camera orientation given a (rough) estimate for the geo-position from which the image was taken. The proposed approach is based on aligning a semantic segmentation of the image with the projection of a semantically-annotated DEM. Intuitively, one would expect that the boundaries of the segmentations contain the most useful information. However, the chapter shows that more information is actually contained within the regions, which allows the proposed method to robustly handle inaccurate segmentations (where the boundaries are not accurately segmented). The chapter also shows that the proposed approach is complementary to previous work based on line segments for orientation estimation. Combining both strategies (semantic segmentation and edge segments) leads to state-of-the-art accuracy for the task.
  - Previous work on localization in mountainous terrain, and the method introduced in Chapter 6, developed localization methods tailored to the fact that little image data is available for building a scene representation. As such, these methods rely on information such as horizon lines, (semantic) edges, and semantic segmentation. Chapter 7, based on the same ECCV paper as Chapter 5, proposes an approach based on feature matching between images and a 3D model, similar to the classical approach used in urban environments. Using techniques from Chapter 5, Chapter 7 shows how to solve the main problem for making this approach work in the considered setting, namely how to generate training data for learning suitable features. The chapter describes how to generate pairs of patches associating real images and renderings of a textured DEM. These pairs are then used to train a feature descriptor that can be used for cross-modal feature matching. The resulting descriptor clearly outperforms state-of-the-art descriptors for the challenging task of visual localization in mountainous terrain. I consider this chapter, in combination with Chapter 5, the second main contribution of the thesis. In particular, it is interesting to see that a traditional pipeline, often used for urban environments, is also applicable in mountainous scenes when one is clever about the available data.
  - Chapter 7 also briefly describes a mobile phone app for localization in mountains based on the approach developed in the same chapter (showing that the proposed approach, in contrast to many state-of-the-art methods, is lightweight enough to run on a mobile device with limited compute capabilities). Chapter 8 describes another application of the localization techniques developed in the thesis, namely an immersive visualization of trips, and analyzes the application through a user study.

- Finally, Chapter 9 summarizes the thesis and discusses potential directions of future work.

A further important contribution of the thesis is that the proposed algorithms and datasets, as well as the training data used in the thesis, have been made publicly available. Thus, the thesis contributes to reproducible research and allows other researchers to easily build on top of Mr. Brejcha's work. Overall, the thesis describes original work that pushes the state-of-the-art for visual localization in mountainous environments, both by introducing new localization algorithms and by proposing the datasets necessary to evaluate such algorithms (as well as the means to create such datasets). The work covers a wide range of techniques, ranging from edge-based information, over semantic segmentation, to modern feature learning. In addition, the thesis demonstrates practical applications for the proposed techniques. In my opinion, the thesis makes significant contributions to its research field and is based on publications at appropriate conferences and journals, including well-respected journals and conferences in the field of computer vision. The quality of the publications reflects well on the quality of Mr. Brejcha's work.

Overall, the thesis is clearly written and structured. The relation between the chapters is clear. Each chapter clearly describes the problem tackled in the chapter and the chapters end with a summary of their main contributions. In general, I found it easy to follow the structure and the argumentation of the thesis. The chapters in Parts II and III also include detailed experimental evaluations of the proposed methods / datasets that are used to generate interesting insights into the characteristics of the datasets and methods, including detailed ablation studies.

On the negative side, I feel that some related work to the topics of this thesis is missing. Concretely, Assia Benbihi's work should be discussed as it clearly constitutes related work on localization in natural scenes (even though it focuses on localization in bucolic and not mountainous environments). In Chapter 7, I am missing a discussion of the work of Aubry et al. [10], who proposed an approach for establishing matches between paintings of a scene and textured 3D models, which seems highly related to the method proposed in that chapter. Chapter 5 states that ``our method is the first to propose a 3D SfM reconstruction jointly using real photographs and rendered imagery to achieve an implicit geo-registration". Yet, Schöps et al., A Multi-View Stereo Benchmark with High-Resolution Images and Multi-Camera Videos, CVPR 2017 also proposed to register real images with renderings of laser scans to align both types of data in one coordinate system. However, the approach from Schöps et al. deals with a much easier situation, where there is a much smaller domain gap between the renderings (of highly detailed laser scans of a scene rather than coarse but textured DEMs). While the claim might not hold, the contribution of this chapter still stands and adjusting the claim would not take away from the impact of Mr. Brejcha's work. Overall, these are minor issues that could be fixed easily.

While the thesis is in general easy to read, I had some trouble following some of the more technical parts, in particular Chapter 6.2.3, which is partially due to the notation used. Similarly, I am confused by seeing negative rotation errors in Tab. 4.2 (I would expect errors to be positive and this is also how they are typically defined in the literature). I am wondering how this affects the mean and median errors in Tab. 4.2. Still, these are minor issues that can be resolved easily. They do not take away from the quality of work of Mr. Brejcha’s PhD thesis.

Overall, as detailed above, I believe that Mr. Jan Brejcha’s dissertation makes important scientific contributions in the area of visual localization in mountainous areas. Given the quality of the research and the thesis, I therefore strongly recommend that the thesis be accepted as it meets the requirements of the proceedings leading to the PhD title conferment.

If you have any questions, please do not hesitate to contact me.

Sincerely yours,

Dr. Torsten Sattler