Bachelor Project

Czech
Technical
University
in Prague

F3

Faculty of Electrical Engineering
Department of Cybernetics

3D reconstruction of indoor boulders

Martin Forman

Supervisor: Ing. Jan Čech, Ph.D.
Field of study: Open Informatics
Subfield: Artifitial inteligence
January 2024

Acknowledgements

I would like to thank my supervisor Ing. Jan ƒech Ph.D. for the help he provided with writing and researching my thesis. I would also like to thank the climbing gym Smicho [1] in Prague and Hudy boulder Karlín [2] for letting me collect a dataset of boulders for my work. Many thanks too belongs to the contributors and authors of libraries used in the implementation of this project, namely: Kornia, PyQt, Open CV, sklearn, GCS and more.

Declaration

I declare that the presented work was developed independently and that I have listed all sources of information used within it in accordance with the methodical instructions for observing the ethical principles in the preparation of university theses.

A LLM was used for help with some coding tedious tasks, LaTeXformatting and testing research ideas.

---

[1] Smichof climbing gym website: https://www.lezeckecentrum.cz/cs/

[2] Hudy climbing gym website: https://www.hudysteny.cz/boulderkarlin/

Abstract

In this thesis, the 3D reconstruction of indoor climbing boulders from two images was investigated. An application was developed to measure distances between holds, as well as other objects and angles within the scene. The methodology combined classical Structure from Motion (SfM) techniques with recent learned descriptors for matching, specifically LoFTR.

The application's performance was evaluated for accuracy, revealing that multiple reconstructed scenes achieved measurement errors around 5%. However, some cases exhibited higher errors due to factors such as low texture, poor image quality, and sub-optimal baselines. The evaluation results informed a discussion on potential improvements and identified limitations of the current approach.

Keywords: SfM, Structure from Motion, 3D reconstruction, Boulder 3D reconstruction, boulder, 3D models, point cloud, 2D to 3D, RANSAC, GCS, depth estimation, indoor climbing, LoFTR, feature matching, image-based reconstruction, photogrammetry, measurement accuracy, texture mapping, computer vision, scene reconstruction, distance measurement, angle measurement, image quality, baseline, error analysis, application development

Supervisor: Ing. Jan ƒech, Ph.D.
Fakulta Elektrotechnická ƒVUT,
Na Zderaze 269/4,
121 35 Praha 2

Abstrakt

V této práci byla zkoumána 3D rekonstrukce vnit°ních lezeckých boulder· ze dvou snímk·. Byla vyvinuta aplikace pro m¥°ení vzdáleností mezi chyty, stejn¥ jako jiných objekt· a úhl· ve scén¥. Metodologie kombinovala klasické techniky Structure from Motion (SfM) s nedávnými u£enými deskriptory pro párování, konkrétn¥ LoFTR.

Výkon aplikace byl hodnocen z hlediska p°esnosti, p°i£emº bylo zji²t¥no, ºe více rekonstruovaných scén dosáhlo m¥°icí chyby kolem 5%. N¥které p°ípady v²ak vykazovaly vy²²í chybu kv·li faktor·m jako nízká textura, ²patná kvalita snímk· a neoptimální baseline. Výsledky hodnocení vedly k diskuzi o moºných vylep²eních a identifikaci omezení sou£asného p°ístupu.

Klí£ová slova: SfM, Structure from Motion, 3D rekonstrukce, 3D rekonstrukce boulder·, boulder, 3D modely, mra£no bod·, 2D na 3D, RANSAC, GCS, odhad hloubky, indoor lezení, LoFTR, párování prvk·, rekonstrukce zaloºená na obrazech, fotogrammetrie, p°esnost m¥°ení, mapování textur, po£íta£ové vid¥ní, rekonstrukce scény, m¥°ení vzdáleností, m¥°ení úhl·, kvalita obrazu, základna, analýza chyb, vývoj aplikací

P°eklad názvu: 3D rekonstrukce vnit°ních lezeckých boulder·

Contents

Figures

Tables

Chapter 1

Introduction

This thesis focuses on building a python based application that lets users reconstruct a 3D model from images of a climbing boulder and use the captured model to retrieve data about angles and distances in the scene. To provide a seamless experience and make the application widely available, we opted to reconstruct the 3D model only using two images, that can be captured by any image capturing device (phone camera, DSLR, etc.), that only has to be calibrated once (can be done offline) and allows for hand held image capture and a varied baseline (distance between the capture points of the two images).

The ability to capture accurate 3D data without the use of specialized technologies like LiDAR or stereo camera rigs, is the main goal of this thesis. To achieve this goal, a classical 3D Reconstruction pipeline is employed, with some modern tools from the Kornia [3] library, that uses machine learning to improve on the performance of traditional algorithmic approaches to some tasks.

As mentioned, the limitation of two images was selected to create a user friendly system which could be integrated into a desktop or mobile app. It is also interesting to experiment with the limitations of creating 3D models with limited data and how specific domain knowledge can substitute this data. For boulders, we can assume we are reconstructing a partially planar scene (a wall with mostly neutral colours) that contains easily recognisable texture, like climbing holds, shoe marks or chalk marks (seen in figures 1.1).

This texture is what allows good performance of the selected pipeline, mostly regarding the dense stereo reconstruction phase, where detection of the same physical points in the provided images is performed. The linear interpolation method was also selected based on the planar nature of the boulder.

Since accuracy is part of the goal of 3D reconstruction, 3D information gathered from the scene by different methods will be used for comparison. We show the viability of the more traditional two picture SfM approach against open source tools like COLMAP [16, 17] (a very powerful reconstruction tool-set, however mainly focused on big scenes with many input images). We also compare to real world collected measurements and calculate the error of our reconstruction.

For completeness, it is better to state that 3D reconstruction from a hand-

Figure 1.1: Boulder and its distinct features (chalk and shoe marks).

held camera will never be the most accurate way of obtaining a 3D model, as many steps in its process are based on some form of estimates. Other techniques of scanning 3D space can actually capture the depth information in a scene or more accurately estimate it. LiDAR is specifically a technology that accurately captures the full 3D scene and is not affected by aspects like lighting which can impact photogrammetry results (photogrammetry includes SfM and our approach). Other techniques like infrared 3D scanning or phootgrammetry using a stereo camera (calibrated with fixed distance between two cameras) can also produce more accurate results.

Chapter 2

Related Work

Searching for related work to the specific task of photogrammetric reconstruction of 3D indoor boulders, to the best of our knowledge, no similar research comes up. The closest research project [13] was concerned with reconstructing outdoor climbing walls (natural not artificial) and mainly the individual climbing routes on them. Although the research also uses a fairly traditional tool-set of SfM methods, an approach of using 400 images was taken by this research, and the focus on outdoor climbing walls means there is a difference in the textures and also the scale of the reconstruction process. Another related research article [20, p.7] we found also focused on outdoor boulders of much greater dimensions and used many images, but provides a interesting insight about scaling and model accuracy:

> It was shown that control points are not only necessary to scale the model but also to compensate for the non-linear model misalignment. Indeed, the accuracy can be considerably improved by using control points and performing the optimisation process in Photoscan.

Based on this research, our pipeline will have to overcome these issues in another way, since we do not have control points (manually added elements on the boulder) for the sake of ease of use.
Most of the cited work that was used to build the basis for the 3D reconstruction program is general SfM content (research or otherwise). Even though the theory for 3D reconstruction is generally well defined and easily implemented, many of the methods perform differently for each data set and a lot of experimental work has to be done to get good performance. To get accurate results, tiny mistakes have to be corrected or eliminated, otherwise the error multiplies in the pipeline and the results suffer accordingly.
This could be viewed as one of the benefits of this work, as no other research we found tackles the specific issues encountered in creating a good 3D reconstruction application for indoor boulders while not using a lot of input data but only two images.
Even though the specific problem of boulders is not commonly researched, we need to take into account many general reconstruction tools that try to offer good result for any object (boulder or other). Such tools include COLMAP [17, 16], which is a very powerful tool, but not well suited for our specific task, managing mainly reconstruction from thousands of pictures, videos or

3D scans. However when we try to give it limited data it fails to produce good results. See figure 2.1, where after inputting 5 images, COLMAP only reconstructed around 5000 points of the scene. Using only two images produced even worse results.

There are other commercial tools that reconstruct 3D data from images or video like Capturing reality [1], the distinguishing factor compared to this work being again the amount of images.

Figure 2.1: COLMAP [17, 16] result with 5 images.

Other research it is interesting to compare to is Depth-Anything [22]. This is a mono-depth model trained to estimate depth from a single view. Although the relative results visually seem quite accurate (as seen in figure 2.2), this can only serve as part of the pipeline (providing relative depth information), and we need to follow it up with some SfM processing to retrieve a 3D model. Other mono-depth tools exist and research into them includes [9], however Depth-Anything is the state of the art solution at thge moment of writing.

Figure 2.2: Depth-Anything [22] result from a single image (a model based monodepth algorithm approximates depth without any other data.)

Chapter 3

The solution

In this chapter, the process of getting a 3D reconstruction out of only two images while still providing a accurate result will be discussed together with relevant theory. Parts of the experiments will also be shown through comparisons between available methods and algorithms with examples of the effect on the resulting 3D model.

3.1 Getting the 3D model

Constructing a 3D model is a multi step process that starts with the raw source images and uses SfM theory to get inter-image correspondences, estimate Fundamental and Essential matrices (discussed further) and using triangulation to calculate the 3D points estimation. Along this process many steps can be taken to improve the result, starting with camera calibration, outlier filtration and interpolation. All these steps combined get us a more accurate and more complete result. See the fig. 3.1 for a quick overview or read the description of each step below.

3.1.1 Creating the data set (capturing images)

For purposes of testing the thesis code, a dataset of images of boulders had to be collected. As no publicly available datasets were found, data collection was performed by the author, by capturing bordering gyms in Prague (mostly Smichoff and Hudy Boulder Karlín).

Important aspects of the images are: a good baseline distance (translation) between the two capture points (for the same scene) was chosen to make the results of SfM more accurate; the same lens (the same camera) was used for capturing the image pair (a camera matrix estimated from calibration was applied to both images); Distinct parts of the scene were captured in both images (the majority of the actual subject - the boulder, was in view for both pictures), while also fitting in a part of the ground for better context.

Elaborating on the baseline choice, we always need to find a compromise between matching accuracy and reconstruction error. Regarding matching accuracy, a smaller baseline enables better performance of LoFTR [18] and subsequent dense stereo GCS [5], which yields more matched points, improving

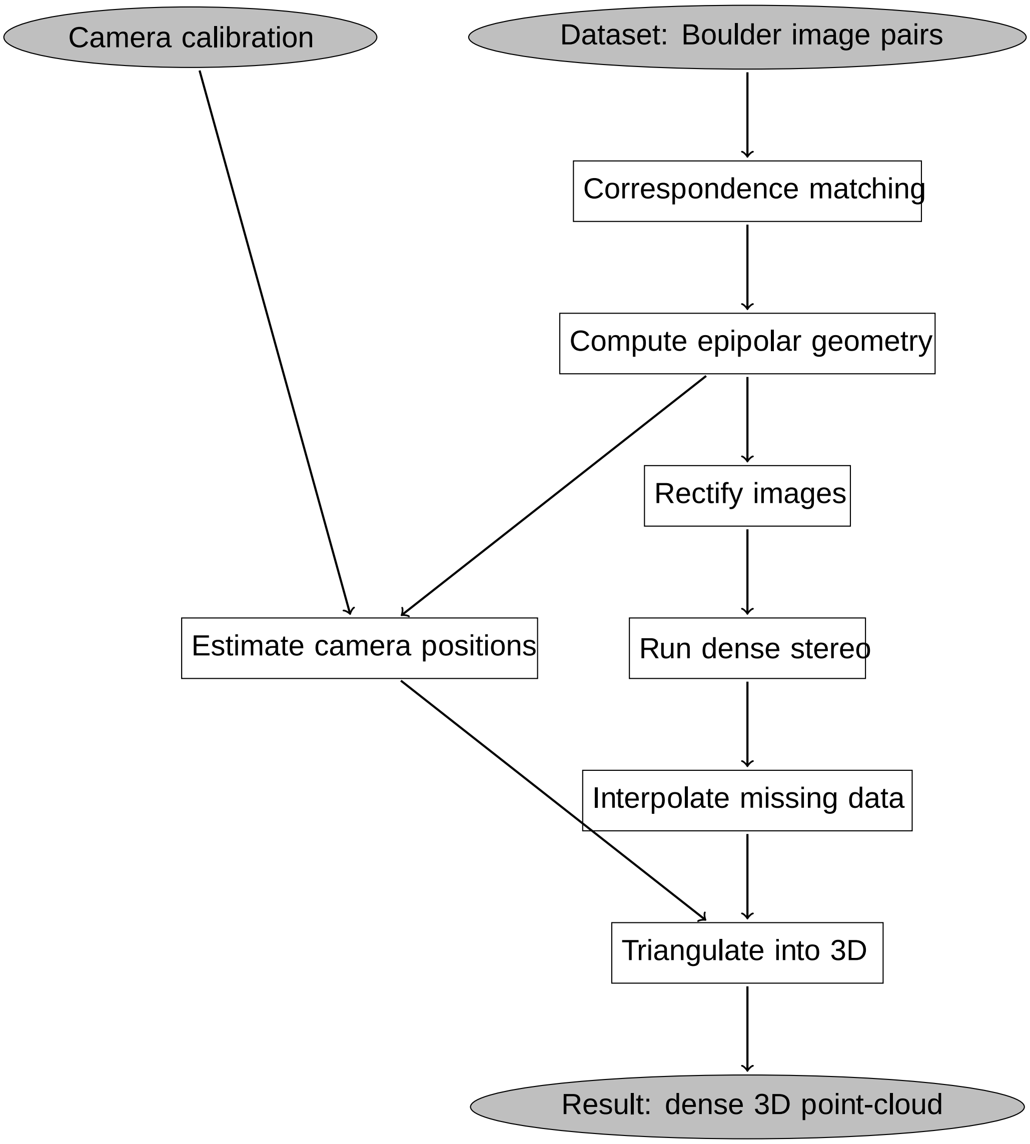

Figure 3.1: 3D reconstruction pipeline overview.