Hochschule
Bonn-Rhein-Sieg
University of Applied Sciences

Bonn-Aachen
International Center for
Information Technology

R&D Project Proposal

Influence in Performance of Feature Descriptor and Matching in Visual Keypoint-based Tracking for Simultaneous Localization and Mapping (SLAM)

Harley Nelson Lara Alonso

Supervised by

Prof. Dr. Sebastian Houben

November 2023

1 Introduction

1.1 Topic of This R&D Project

For about 35 years the problem of simultaneous localization and mapping has been studied in the area of robotics [4], this problem commonly referred to as SLAM in essence seeks to find a high degree of correlation between localization estimates and different landmarks in the environment [1]. Advances in the area have been outstanding as many methods and algorithms seek to propose a solution, however there is not much conclusive work on how the internal mechanisms of visual keypoint-based SLAM are influenced by extractors, descriptors and matching algorithms. It is well understood and even intuitive to think that better extraction methods, better descriptors or better matching algorithms have a direct influence on SLAM performance, but the question is; *"how much do these components influence SLAM?"*. In this work consider a *component* as one of these three processes; extraction, descriptor or matching.

1.2 Relevance of This R&D Project

This research and development (R&D) project holds relevance for the following reasons:

- **Decision-making**: Provides comprehensive data and contributes to more informed decision-making processes. The outcomes facilitate objective assessments, reducing reliance on subjective biases or speculative choices when selecting a SLAM implementation.
- **Identify performance improvements**: Helps to identify areas where improvement is needed and to take steps to increase quality, efficiency and/or speed, leading to future research directions.
- **Resource-constrained applications**: Presents insights for achieving a balance between accuracy and computational efficiency, particularly in resource-constrained applications. Such information is imperative for optimizing SLAM algorithms to operate effectively within the confines of limited hardware resources.

- **Performance information**: Aims for better performance without degrading other metrics such as matching rate, relative position error (RPE), absolute trajectory error (ATE), and others. This ensures that performance improvements do not come at the expense of degrading other essential aspects of SLAM functionality.

2 Related Work

2.1 Survey of Related Work

- **Extraction, matching and SLAM**: Some works as in [11] [9] aim to evaluate the performance of detectors, matching and SLAM, [11] evaluates extraction and matching methods including; Harris, Features from accelerated segment test (FAST), robust invariant scalable keypoints (BRISK), maximally stable extremal regions (MSER), scale-invariant feature transform (SIFT), speeded up robust features (SURF) and oriented FAST and rotated BRIEF (ORB) while comparing the performance of these algorithms in terms of accuracy, speed, scale invariance, and rotation invariance. [9] Explicitly only evaluates the best landmark detector and descriptor for Visual SLAM using a data-set of aerial images taken with drone and evaluates the highest number of keypoints and matching pairs.

- **Descriptor and matching**: The proposed framework by [8] analyzes feature keypoint descriptors for image matching on three metrics; recall, precision and average number of best matches, the proposed framework applies modifications to the images, these include; distortions such as blur, scale and illumination changes, and image rotations. [10] Compares the performance of SIFT, SURF, binary robust independent elementary features (BRIEF) and ORB on images with various types of transformations and deformations, including rotation, scale, noise, fish eye distortion, and shearing, the authors analyze the performance based on the number of key points in the images, matching rate, and runtime of each algorithm, which is evaluated on a set of images for which the origin or data-set is not specified.

- **Extraction, descriptor and matching**: The assessment protocol in [15] is based on the number of correct and false matches for a given image pair, the accuracy and speed of matching is measured using multiple kd-trees.

- **Extraction and descriptor**: The information present in [2] studies the performance of 25 different descriptors in combinations with matching methods, this analysis measures five parameters; accuracy, time, angle difference between keypoints, number of correct matches, and distance between correctly matched keypoints. *A. M. M. Madbouly et al.* in [13] executes evaluation on multiple detectors including MinEigen in conjunction with descriptors such as fast retina keypoint (FREAK), SURF, and BRISK, the experiments point out that MinEigen detector has the best result in detecting keypoints under rotation, scale, illumination and is not affected by the scene.

- **Only descriptors**: The paper [6] conducts a comparative study of 10 local feature descriptors for 3D object recognition, 3D shape retrieval, and 3D modeling applications, the authors propose to measure descriptiveness, compactness, robustness, computational efficiency of the descriptors. Descriptiveness is measured using precision-recall curve (PRC), while compactness is measured using the average value of the area-under-the-curve precision-recall (AUC PR). Robustness is tested with Gaussian noise, shot noise, changes in the grid result, distance to the grid boundary, keypoint localization error, occlusion and clutter.

- **Extraction, descriptor and SLAM**: The material within [5] uses a graph-based RGB-D SLAM method to detect feature points in RGB frames and constructing descriptor vectors, while the 3D coordinates of each feature point are calculated using depth data. The paper evaluates the performance of various feature detectors and descriptors in terms of accuracy and speed. The feature detectors used include SIFT, SURF, BRISK, ORB, FAST, good features to track (GFTT), GFTT HARRIS (GFTT with Harris detector enabled), and STAR. The feature descriptors tested include BRIEF and FREAK, which are tested with all detectors. Additionally, SIFT, SURF, BRISK, and ORB descriptors are also tested with their own detectors. [7]

compares the influence of feature descriptor and feature detectors with RGB-D TUM dataset, the comparison is made in terms of accuracy using the root mean square error of absolute trajectory error (ATE), additionally it measures the runtime in based on processing time per frame.

- **Deep-learning-based approach**: As an alternative to the classical methods, the papers [12] and [16] propose the use of an architecture based on neural networks to extract the features and thus improve matching accuracy while maintaining real-time capabilities. [16] performs matches of a set of two local features by finding correspondences and eliminating non-matchable points, the Deep NN architecture allows SuperGlue [16] to infer information about the underlying 3D scene and feature assignments without explicitly implementing mechanisms to describe invariance, geometric transformations or inconsistencies in the images. In contrast, LightGlue [12] inspired by SuperGlue offers a lightweight yet efficient alternative that prioritizes optimization for memory and computational efficiency, making it particularly suitable for edge devices with limited hardware resources. LightGlue is compared to recent sparse and dense baselines on the Aachen v1.1 dataset, showing that it is far faster than all approaches while predicting a similar amount of correspondences but with higher precision, pose accuracy, and speed than existing sparse matchers. It is also competitive with dense matching methods for a fraction of the inference time.

2.2 Limitation and Deficits in the State of the Art

As presented in section 2.1, many of the research papers focus on specific components without evaluating the impact on SLAM performance, [11] and [9] do not test images that represent scenarios with real-world application, and do not include experiments that evaluate SLAM, but only evaluate metrics related to extraction and matching, while [8] and [10] focus only on descriptors and matching methods. See table 1 below, which shows the components evaluated for each paper in the survey.

		Performance Influence Assessment			
Year	Reference	Extraction	Descriptors	Matching	SLAM
2017	[11]	✓		✓	
2013	[9]	✓		✓	✓
2019	[8]		✓	✓	
2014	[3]		✓	✓	
2017	[10]			✓	✓
2014	[15]	✓	✓	✓	
2017	[2]	✓	✓		
2015	[13]	✓	✓		
2016	[6]		✓		
2005	[14]		✓		
2015	[5]	✓	✓		
2013	[7]		✓		
2023	[12]		✓		✓
2020	[16]		✓		✓
2024	Proposed R&D	✓	✓	✓	✓

Table 1: Comparison of the components evaluated in the papers corresponding to the survey of related works.

3 Problem Statement

This research and development project aims answer the question; *What is the effect of the features properties on the overall SLAM performance?* to explore the core characteristics that degrade or improve the performance of SLAM. The objective is to propose an end-to-end framework to evaluate the characteristics of SLAM components.

The research would consist of conducting a literature survey to determine the metrics of each component and their relationship to SLAM performance, in addition to determining the metrics and the SLAM evaluation process itself, such quantifiers may include Absolute Trajectory Error (ATE) and Relative Position Error (RPE) (for rotation and translation). Subsequently a series of benchmarks will be conducted on a collection of data-sets that are representative of the real-world SLAM.

Additionally explore a learned-based approach in a fully fledged SLAM algorithm by comparing classical keypoint extraction and matching methods with a deep learning based implementation, and including attentional span inspired by [12] [16].

4 Project Plan

4.1 Work Packages

WP*	**Weeks**	**Description**
1	8	Research literature to *identify relevant metrics* for measuring the performance of SLAM, feature descriptors and image matching in the context of visual keypoint-based tracking.
2	2	Identify and select appropriate data-sets to evaluate the influence in performance of the metrics identified in WP 1.
3	2	Extract from literature a set of representative SLAM algorithms based on their capabilities with visual keypoint-based tracking
4	6	Develop an experimental plan to systematically evaluate the selected SLAM algorithms using the identified metrics and data-sets.
5	5	Prepare the necessary infrastructure and software environment for conducting experiments according to the designed experimental plan.
6	7	Conduct experiments based on the established experimental design, using the selected data-sets and SLAM algorithms.
7	2	Process data collected during experiments.
8	6	Analyse experimental results and provide comprehensive transcribe on the findings in the form of a project report

* WP: Work package.

Table 2: Set of work packages for this research and development project.

4.2 Milestones

MS*	Outcome
1	Table summarizing measurable and relevant metrics for performance assessment.
2	Table listing relevant data-sets for experimentation.
3	Table comparing and evaluating delimited SLAM methods.
4	Table, diagrams, structure and description of the experimental design.
5	Configured experimental environment ready for experiments.
6	Set of tables, plots, and diagrams presenting the experimental results.
7	Final project report submitted.

* MS: Milestone.

Table 3: Set of milestones.

4.3 Project Schedule

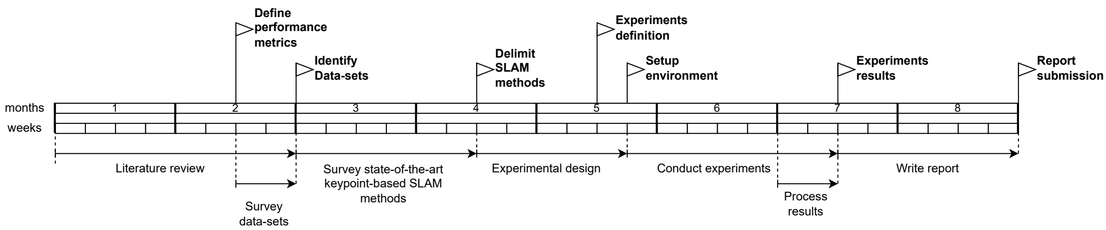

Figure 1: Gantt chart representation of the project schedule

4.4 Deliverables

1. Consolidated data-set from the collection of data-sets used for performance assessment.
2. Docker image as a reproducible experimental environment.
3. Raw experimental results.
4. Jupyter notebook with the data processing pipeline and visualization scripts.

5. Written project report.

References

[1] Tim Bailey and Hugh Durrant-Whyte. Simultaneous Localisation and Mapping (SLAM): Part II State of the Art.

[2] Ertuğrul Bayraktar and Pınar Boyraz. Analysis of feature detector and descriptor combinations with a localization experiment for various performance metrics. *TURKISH JOURNAL OF ELECTRICAL ENGINEERING & COMPUTER SCIENCES*, 25:2444–2454, 2017. ISSN 13000632, 13036203. doi: 10.3906/elk-1602-225. URL `https://journals.tubitak.gov.tr/elektrik/vol25/iss3/66`.

[3] Vijay Chandrasekhar, Gabriel Takacs, David M. Chen, Sam S. Tsai, Mina Makar, and Bernd Girod. Feature Matching Performance of Compact Descriptors for Visual Search. In *2014 Data Compression Conference*, pages 3–12, Snowbird, UT, USA, March 2014. IEEE. ISBN 978-1-4799-3882-7. doi: 10.1109/DCC.2014.50. URL `http://ieeexplore.ieee.org/document/6824408/`.

[4] Hugh Durrant-Whyte and Tim Bailey. Simultaneous Localisation and Mapping (SLAM): Part I The Essential Algorithms.

[5] Oguzhan Guclu and Ahmet Burak Can. A Comparison of Feature Detectors and Descriptors in RGB-D SLAM Methods. In Mohamed Kamel and Aurélio Campilho, editors, *Image Analysis and Recognition*, volume 9164, pages 297–305. Springer International Publishing, Cham, 2015. ISBN 978-3-319-20800-8 978-3-319-20801-5. doi: 10.1007/978-3-319-20801-5_32. URL `https://link.springer.com/10.1007/978-3-319-20801-5_32`. Series Title: Lecture Notes in Computer Science.

[6] Yulan Guo, Mohammed Bennamoun, Ferdous Sohel, Min Lu, Jianwei Wan, and Ngai Ming Kwok. A Comprehensive Performance Evaluation of 3D Local Feature Descriptors. *International Journal of Computer Vision*, 116(1):66–89,

January 2016. ISSN 0920-5691, 1573-1405. doi: 10.1007/s11263-015-0824-y. URL http://link.springer.com/10.1007/s11263-015-0824-y.

[7] Jan Hartmann, Jan Helge Klussendorff, and Erik Maehle. A comparison of feature descriptors for visual SLAM. In *2013 European Conference on Mobile Robots*, pages 56–61, Barcelona, Catalonia, Spain, September 2013. IEEE. ISBN 978-1-4799-0263-7. doi: 10.1109/ECMR.2013.6698820. URL http://ieeexplore.ieee.org/document/6698820/.

[8] M. Hassaballah, Hammam A. Alshazly, and Abdelmgeid A. Ali. Analysis and Evaluation of Keypoint Descriptors for Image Matching. In Mahmoud Hassaballah and Khalid M. Hosny, editors, *Recent Advances in Computer Vision*, volume 804, pages 113–140. Springer International Publishing, Cham, 2019. ISBN 978-3-030-02999-9 978-3-030-03000-1. doi: 10.1007/978-3-030-03000-1_5. URL http://link.springer.com/10.1007/978-3-030-03000-1_5. Series Title: Studies in Computational Intelligence.

[9] Herdawatie Abdul Kadir and Mohd Rizal Arshad. Features detection and matching for visual simultaneous localization and mapping (VSLAM). In *2013 IEEE International Conference on Control System, Computing and Engineering*, pages 40–45, Penang, Malaysia, November 2013. IEEE. ISBN 978-1-4799-1508-8 978-1-4799-1506-4. doi: 10.1109/ICCSCE.2013.6719929. URL http://ieeexplore.ieee.org/document/6719929/.

[10] Ebrahim Karami, Siva Prasad, and Mohamed Shehata. Image Matching Using SIFT, SURF, BRIEF and ORB: Performance Comparison for Distorted Images.

[11] Shimiao Li. A review of feature detection and match algorithms for localization and mapping. *IOP Conference Series: Materials Science and Engineering*, 231:012003, September 2017. ISSN 1757-8981, 1757-899X. doi: 10.1088/1757-899X/231/1/012003. URL https://iopscience.iop.org/article/10.1088/1757-899X/231/1/012003.

[12] Philipp Lindenberger, Paul-Edouard Sarlin, and Marc Pollefeys. LightGlue:

Local Feature Matching at Light Speed, June 2023. URL http://arxiv.org/abs/2306.13643. arXiv:2306.13643 [cs].

[13] A M M Madbouly, M Wafy, and Mostafa-Sami M Mostafa. Performance Assessment of Feature Detector-Descriptor Combination. *International Journal of Computer Science Issues*, 12(5), 2015.

[14] K. Mikolajczyk and C. Schmid. A performance evaluation of local descriptors. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 27(10): 1615–1630, October 2005. ISSN 0162-8828. doi: 10.1109/TPAMI.2005.188. URL http://ieeexplore.ieee.org/document/1498756/.

[15] Ondrej Miksik and Krystian Mikolajczyk. Evaluation of local detectors and descriptors for fast feature matching. *Proceedings of the 21st International Conference on Pattern Recognition (ICPR2012)*, pages 2681–2684, 2012.

[16] Paul-Edouard Sarlin, Daniel DeTone, Tomasz Malisiewicz, and Andrew Rabinovich. SuperGlue: Learning Feature Matching with Graph Neural Networks, March 2020. URL http://arxiv.org/abs/1911.11763. arXiv:1911.11763 [cs].