# A Systematic Literature Review on Long-Term Localization and Mapping for Mobile Robots*

Ricardo B. Sousa Héber M. Sobreira António Paulo Moreira

September 2, 2022

## Abstract

**Keywords:** simultaneous localization and mapping (SLAM), lifelong SLAM, long-term autonomy, mobile robots.

# 1 Introduction

# 2 Purpose of the study

## 2.1 Limitations of current studies

Table 1: Existent Literature Reviews, Surveys, and Tutorials on SLAM.

| **Topic** | **Reference** |
|---|---|
| Probabilistic approaches and data association | Bailey and Durrant-Whyte 2006; Durrant-Whyte and Bailey 2006 |
| SLAM back end | Grisetti et al. 2010 |
| Multi-robot SLAM | Saeedi et al. 2016 |
| Visual odometry | Fraundorfer and Scaramuzza 2012; Scaramuzza and Fraundorfer 2011 |
| Overview of challenges in SLAM | Cadena et al. 2016 |
| Trends in SLAM for autonomous vehicles | Bresson et al. 2017 |
| **Completar tabela!** | |

## 2.2 Motivations and goals

Research question: What is the current state of the art of long-term localization and mapping using mobile robots?

Goals of this review:

- which are the main strategies for accomplishing long-term operations with mobile robots;
- how to deal with varying conditions of the environment;
- how do autonomous robots deal with the dynamics of the environment;
- which are the main strategies to deal with the limited computational resources of a mobile robot on long-term operations;
- how the methods evaluate their results;
- which are the public datasets more used for evaluating long-term localization and mapping.

PICO framework (Population–Intervention–Comparison–Outcome) helps to frame the research questions of this systematic review into a more structured framework:

- **Population:** mobile robots;
- **Intervention:** localization, mapping, SLAM;
- **Comparison:** *not applicable to this study*;
- **Outcome:** long-term operation, lifelong autonomy, robust.

# 3 Methodology

A systematic literature review uses explicit, rigorous, and reproducible systematic methods to synthesize the findings of studies related to a particular research question, topic area, or phenomenon of interest. This type of review assures the quality and trustworthiness of the review's findings by presenting a complete, organized, and summarized analysis of all works considered while allowing others to replicate or update the reviews. The most common standard for performing a systematic review is the Preferred Reporting Items for Systematic reviews and Meta-Analysis (PRISMA) (Page et al. 2021) statement. Although the PRISMA statement has been designed originally for evaluating the effects of health interventions, the checklist items of the methodology are general and applicable to other subject areas. Thus, the methodology used in this systematic review follows the PRISMA (Page et al. 2021) guidelines.

This section presents the detailed methodology used in this study. First, the eligibility criteria decide which studies to include in the review. Next, the search strategy details the information sources considered in the review and the base string and search fields used for inquiring these sources. Furthermore, the selection process focuses on describing its stages and the quality evaluation criteria used to select works for the synthesis and analysis phase of the review. Lastly, the data extraction process details the relevant data collected for synthesis and analysis. Parsifal (Freitas 2014) is the online tool used to support the literature review in designing the methodology protocol, removing duplicates, screening and selecting works including their quality assessment. Additional documentation and scripts developed within the scope of this review related to removing duplicates, checking and processing the bibliographic references, and data extraction are available in a public GitHub repository[1].

*This work is financed by National Funds through the Portuguese funding agency, FCT – Fundação para a Ciência e a Tecnologia, within scholarship 2021.04591.BD.

Ricardo B. Sousa[1, 2]
up201503004@edu.fe.up.pt

Héber M. Sobreira[2]
heber.m.sobreira@inesctec.pt

António Paulo Moreira[1, 2]
amoreira@fe.up.pt

[1] Faculty of Engineering of the University of Porto, Electrical Engineering Department, Porto, 4200-465 Porto, Portugal
[2] INESC TEC – Institute for Systems and Computer Engineering, Technology and Science, CRIIS – Centre for Robotics in Industry and Intelligent Systems, Porto, 4200-465 Porto, Portugal

[1] https://github.com/sousarbarb/slr-ltlm-mr

### 3.1 Eligibility criteria

Table 2 presents the exclusion criteria used to determine the eligible studies for the selection process. These eligibility criteria focus mainly on the type of paper and availability. The index criterion rejects all publications not indexed in a scientific publication venue. This rejection guarantees that the eligible works were peer-reviewed by the scientific community. Also, the exclusion criteria reject short papers and gray, secondary, and tertiary literature. Short papers do not usually present a detailed methodology of their scientific contribution. As for only considering primary literature in the review, this criterion increases the relevance of search results by favoring original articles and simultaneously guaranteeing peer-revision of the works. In terms of language, only considering studies with English full-texts increases the scope and visibility of the review. Similarly, the eligibility criteria reject studies not available in digital libraries for reproducibility and accessibility reasons.

Table 2: Exclusion criteria for the selection process.

| E# | Criteria | Statement |
|---|---|---|
| E1 | Index | Papers not indexed in a scientific publication venue |
| E2 | Language | Full-text of the papers not published in English |
| E3 | Subject Area | Papers not classified in the databases as Computer Science, Engineering, Mathematics, or Multidisciplinary |
| E4 | Short Papers | Papers classified as short papers accordingly to the publication venue |
| E5 | Gray, Secondary, and Tertiary Literature | Books, preprints, reports, reviews, thesis, ... |
| E6 | Availability | Full-text of the papers not available in digital libraries |
| E7 | Dataset | Papers that focus only on data collection |
| E8 | Coverage | Papers using only odometry for localization |
| E9 | Scope | Papers that focus on different and not related subjects |

Another exclusion criterion considered in the review is relative to the studies' categorization of their subject areas by bibliographic databases. The ones considered in the review are Computer Science, Engineering, Mathematics, or Multidisciplinary areas. In the list provided by the Clarivate's Journal Citation Reports[2], these four subject areas include the artificial intelligence, interdisciplinary applications, electrical and computers engineering, robotics, and applied mathematics categories, among others. These categories are intrinsically related to the localization and mapping problem for long-term operation of mobile robots.

The final three criteria presented in Table 2 focus on the scientific contribution of the studies. The dataset criterion rejects all works that focus only on sharing a data collection. Although these works are important for the evolution of localization and mapping algorithms in providing a benchmark for comparison and reference purposes, their scientific contribution is not directly comparable to research articles. Odometry-only approaches are unusable over long distances invalidating their use for long-term operations with mobile robots. As for the scope criterion, this review does not consider eligible for selection papers not related to long-term localization and mapping.

### 3.2 Search strategy

The search phase consists of identifying the data sources that could be relevant for this literature review, and defining the base string and which search fields considered to obtain the results for the review. *Web of Science* and *Scopus* are traditionally the two most widely used bibliographic databases. However, previous studies demonstrate that different databases differ significantly in their scientific coverage (Mongeon and Paul-Hus 2016; V. K. Singh et al. 2021). Thus, the data sources considered in this review are the following ones: *ACM Digital Library*, *Dimensions*, *IEEE Xplore*, *INSPEC*, *Scopus*, and *Web of Science*.

Moreover, May 17, 2022, is the date of the last full inquiry. Future reviews on the topic of this study should consider this final date as theirs initial one. As for inquiring the data sources, the base string used is the following one:

```
(robot* OR vehicle*)                               AND
((locali* AND map*) OR "slam")                     AND
("long term" OR "life long" OR lifelong)
```

The first terms, `robot* OR vehicle*`, attempt to focus the search results to the desired population. These two terms have multiple synonyms within the scope of autonomous mobile robots: mobile robots, autonomous vehicles, robotics, agricultural robots, intelligent robots, service robots, unmanned aerial/ground/underwater vehicles, among other terms. Therefore, by adding the asterisk to the end of the terms robot and vehicle (`robot*` and `vehicle*`, respectively), and by only considering the terms with asterisk in the inquiry, all the synonyms are covered for the desired population. Given the incompatibility of the *Dimensions* database with wildcards (e.g., using the asterisk), the first part of the base string becomes as follows when searching in this database: `robot OR robots OR robotics OR vehicle OR vehicles`.

The next part of the query focus on the intervention side of the systematic review. Given the interest of this review on searching for localization and mapping algorithms, `locali*` and `map*` summarize all the synonyms for the localization and mapping terms, respectively. For example, `locali*` not only is agnostic to the US versus UK spelling differences (localization vs localisation, respectively) but also resumes several synonyms: localization, localize, or localizing. The term `map*` also attempts to cover its respective synonyms such as map, maps, or mapping. Also, the acronym `"slam"` is another alternative to search for localization and mapping algorithms. Even though its definition is compatible with `locali* AND map*`, some authors only refer to SLAM. Similarly to the inquiry's first part, the second one becomes as follows for searching in *Dimensions*: `((localize OR localization OR localizing OR localise OR localisation OR localising) AND (map OR maps OR mapping)) OR "slam"`.

As for `"long term" OR "life long" OR lifelong`, this part of the base string is relative to the outcome of the PICO framework, presented in Section 2. The reason for having both `"life long"` and `lifelong` terms is the existing confusion in which term is grammatically the correct one.

Furthermore, the Title, Abstract, and Keywords are the fields considered for obtaining the search results. The third one includes the author keywords, the indexed terms by the databases, and the uncontrolled ones if they are available. The selection of these search fields for this review improves the relevance of the results compared to using all fields and the full text by focusing the search on the summary items of the works. Indeed, the main contributions of scientific works should be summarized in at least the title, abstract, or the author keywords. The indexed terms also help in obtaining records only related to the base string used

[2]https://jcr.clarivate.com/jcr/browse-categories

in this review. However, not all data sources have available the search fields considered in the review or some of them require an adaptation when performing the search. Although the *ACM Digital Library* allows searching within multiple search fields, including the ones considered in this review, the advanced search query on this library sets by default an AND operator between the different fields. This setting must be changed manually in the query syntax to the desired OR operator. Also, there are two options to search items in the *ACM Digital Library*: *The ACM Full-Text Collection* and *The ACM Guide to Computing Literature*. Given that the latter includes all the content from the former, the identification process in this source performs the search using *The ACM Guide to Computing Literature* option. Other than searching in the publications' full data, *Dimensions* only has the title and abstract search fields compatible with this review. Given the limitation of *IEEE Xplore* to 7 wildcards, the search results of this digital library using the base string for the inquiry are the grouping of different searches considering only a search field at a time, importing each search results to Parsifal and removing the duplicates. As for *INSPEC*, *Scopus*, and *Web of Science*, these databases have available all the search fields considered in the review.

In terms of the publication date, this review does not restrict it to avoid ignoring important works and to improve the discussion. Indeed, to best of the authors knowledge, there is not available a systematic review on long-term localization and mapping for mobile robots to provide an initial date for rejecting older publications. Even though the number of publications per year could indicate an initial date on when the topic gained relevance, the date filtering could still reject important works.

## 3.3 Selection process

The selection process of this review summarized in Figure 1 has three phases: identification, screening, and quality assessment. The first phase consists of inquiring each data source discussed previously with the base string and adapting it if needed. The second phase requires screening the papers. In this review, screening is equivalent to reading the publications' title and abstract and deciding whether the study is eligible or not based on the exclusion criteria. Then, a set of evaluation criteria assesses the quality of the eligible records. The records obtained after the three phases of the selection process are for the data extraction phase.

### 3.3.1 Identification

In the identification phase of this review, the search strategy is applied to all data sources. *ACM Digital Library*, *Dimensions*, *INSPEC*, *Scopus*, and *Web of Science* data sources only require a single inquiry to obtain the search results. Given the limitation of the *IEEE Xplore* for using wildcards mentioned in Section 3.2, the number of records for this source presented in Figure 1 represents the results of 7 inquiries (using the fields title, abstract, author keywords, IEEE terms, INSPEC controlled terms, and the INSPEC uncontrolled ones, respectively) after removing the duplicates with the support of Parsifal. Although the total number of search results found is 2160, Parsifal is used to remove duplicates from different data sources, excluding 1339 records. Following the duplicates removal, the exclusion criteria defined in Section 3.2 exclude 232 works from the review. This exclusion is possible due to *INSPEC*, *Scopus*, or *Web of Science* having filters related to the publication's type, subject area, and language.

The works excluded from the search results also include the ones that do not meet the exclusion criteria E4 and E7. For the first one, a Python script available in the GitHub repository of this review searches studies with a number of pages lower or equal to 4. Even though short papers have a maximum number of 3 pages, the papers with 4 pages do not usually present a detailed methodology. As for the E7 exclusion criterion, some works are possible to remove from the review by searching in their title for the term "dataset". All excluded articles of this review are double-checked to certify if the exclusion criteria are correctly applied. For example, articles published in the Remote Sensing journal from MDPI do not meet the E3 criterion. Indeed, the Journal Citations Reports from Clarivate classifies it by the following categories: Remote Sensing, Geosciences Multidisciplinary, Environmental Sciences, and Imaging Science & Photographic Technology. However, most search results from this journal found in the identification phase are directly related to the topic of this review and the respective subject areas. Thus, in these cases and in other ones related to the remaining exclusion criteria, the decision is reverted to consider the initially rejected studies for the next phase of the review.

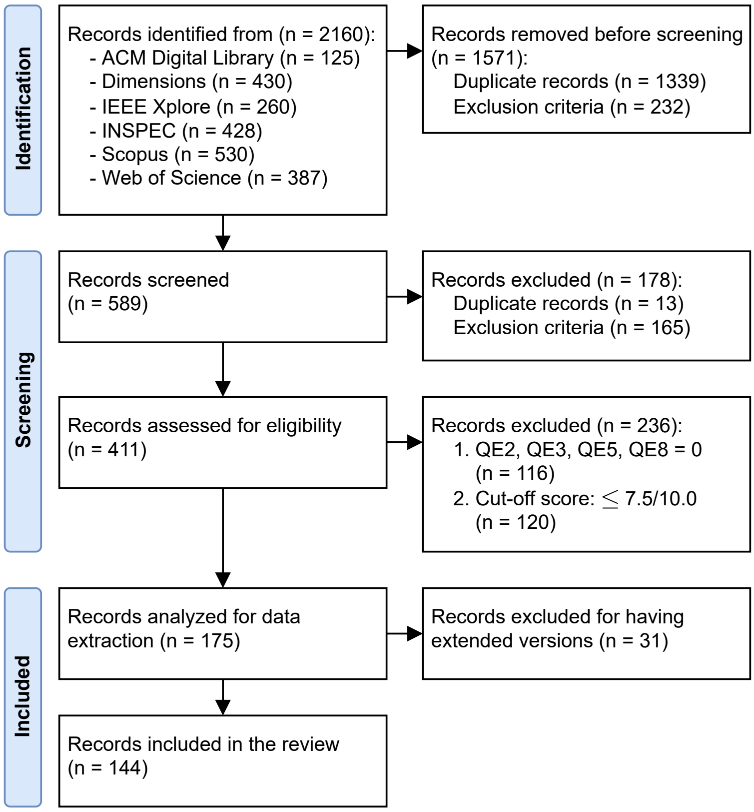



Figure 1: PRISMA flow diagram for the selection process.

### 3.3.2 Screening

Next, the screening phase in this review consists of reading the title and abstract of the publications and rejecting the ones that meet the exclusion criteria. However, the initially rejected papers have another assessment for validating the exclusion. The analysis of the results and conclusions of these publications considering the exclusion criteria either confirms the exclusion decision or reverses it to eligible works for quality assessment. As a result of the screening phase, 178 studies are rejected from the initial identified 589 works. The duplicate records found in screening and removed manually are due to titles with invalid characters originated by exporting the search results from the *Dimensions* database.

### 3.3.3 Quality assessment

The quality evaluation in this review of the selected works from screening follows the 8 Quality Evaluation (QE) criteria presented in Table 3. All of them are subjective criteria derived from the analysis of the eligible works. The score column establishes the possible values for the QE criteria, in which the minimum, intermediate, and maximum values correspond to none, partial, and full compliance, respectively. Furthermore, QE1, QE2, QE4, and QE8 focus on the details provided in the papers, specifically, if the discussion of the related work, the proposed methodology, the experimental setup, and the results are detailed and thoroughly analyzed in the publication, respectively. The possible scores for QE3 are twice the value of QE1, QE2, QE4, and QE8 due to this criterion being directly related to the topic of the review. A work focusing on both localization and mapping problems will have a score of 2.0 (full compliance). If the study only focuses on one of these problems or none of them, the scores will be 1.0 or 0.0, i.e., partial or no compliance, respectively. QE5 evaluates the long-term results of the eligible studies and is either 2.0 (full) or 0.0 (no compliance). This criterion has the same range as QE3 for similar reasons, given the focus of this review on long-term localization and mapping algorithms. The definition of long-term experiments for assigning full compliance in QE5 is the following one: dynamic changing environments (e.g., dynamic elements or semi-static ones), increasing environments or feature maps in terms of their size, redundant data removal, or varying conditions (e.g., different seasons of the year or lighting conditions). QE6 and QE7 can only be 1.0 or 0.0. The former criterion intends to highlight works that compare themselves to the state of the art and/or ground-truth data. The latter emphasizes the importance of having available either the implementation of the proposed methodology or the data used in the experiments for other works to be able to compare the proposed methodologies. Lastly, considering the possible scores for the QE criteria in Table 3, each work can only have a maximum score of 10.0.

Table 3: Quality evaluation criteria and score range.

| QE# | Criteria | Score |
|---|---|---|
| QE1 | Does the paper have an updated state of the art on long-term localization and mapping? | {0.0, 0.5, 1.0} |
| QE2 | Is the methodology appropriate and detailed? | {0.0, 0.5, 1.0} |
| QE3 | Does the methodology consider both localization and mapping problems? | {0.0, 1.0, 2.0} |
| QE4 | Is the hardware and/or software used in the experiments detailed? | {0.0, 0.5, 1.0} |
| QE5 | Does the paper presents any kind of long-term experimental results? | {0.0, 2.0} |
| QE6 | Does the paper presents comparative results with other methods and/or ground-truth data? | {0.0, 1.0} |
| QE7 | Does the work's implementation and/or the data used in the experiments are publicly available? | {0.0, 1.0} |
| QE8 | Is the discussion of the results and conclusions appropriate and detailed? | {0.0, 0.5, 1.0} |

After evaluating the 411 eligible works accordingly to the previously discussed QE criteria (the scores of each record are available in the GitHub repository), the first conclusion of the authors is that works with a non-detailed or not appropriate methodology, results' discussion, or conclusions should not be included in the review. Another conclusion is relative to rejecting works that do not consider either localization or mapping problems, or do not present any long-term experimental results, given the focus of this review on the long-term localization and mapping problem for mobile robots. Furthermore, the quality assessment phase should consider a cut-off score to filter works with low quality scores. Consequently, the assessment phase considers the following two reasons to reject a record:

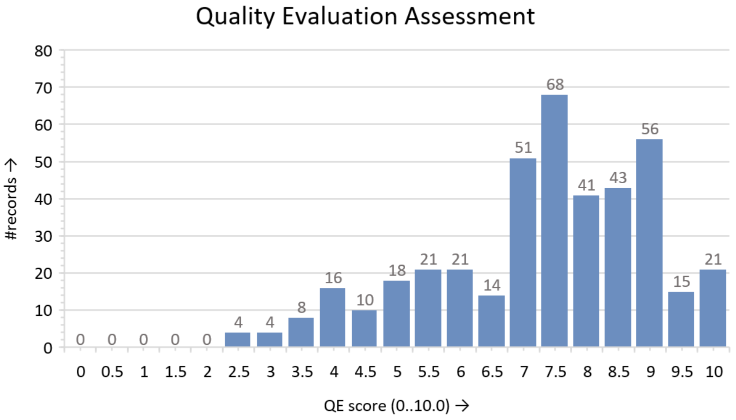
1. QE2, QE3, QE5, QE8: reject works with a 0.0 (no compliance) score;
2. cut-off score: reject works with a score lower or equal to 7.5/10.0.

The distribution of the evaluation scores and the QE criteria itself justify the selection of a 7.5/10.0 cut-off score. Figure 2 illustrates the scores distribution for all eligible works versus the scores of the ones that pass the first criterion defined previously for the QE phase (related to the compliance on the QE2, QE3, QE5, and QE8 criteria). The assessment of this criterion rejects 116 records (28%) of the 411 eligible works (see Figure 1). Even though the distribution of the evaluation scores changes significantly in the range of scores lower or equal to 7.5/10.0, as observed between Figures 2b and 2a, only one work with a score higher than 7.5 is rejected due to not having a detailed and appropriate discussion of the results. This result indicates that interesting works are associated with high scores, as intended when using a quality assessment methodology, while also suggests that the range between 8.0 and 10.0 have the most interesting and quality works compatible with the focus of this review on long-term localization and mapping. Although only assessing the eligible works would seem to lead to the same results in terms of records included in the review, the rejection criterion on QE2/3/5/8 prevents outliers related to the quality assessment. From the remaining 295 eligible works, cut-off scores from 7.5 up to 8.5 have the following corresponding rejection rates:

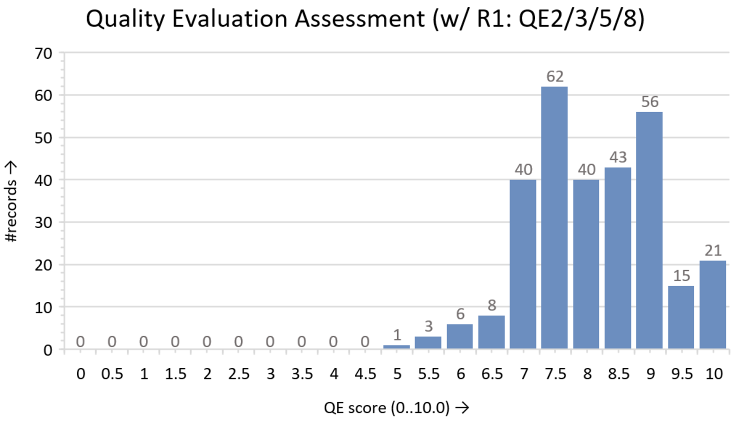
- 7.5/10.0 120 records (40.7%) 175 records
- 8.0/10.0 $\xrightarrow{\text{reject}}$ 160 records (54.2%) $\xrightarrow{\text{include}}$ 135 records
- 8.5/10.0 203 records (68.8%) 92 records

The 8.5 cut-off score would not be suitable because methods that focus only on localization or mapping, or not having either the implementation or the experimental data publicly available would be obligated to have maximum scores in the other criteria to be included in the review. In these cases, a work would have a maximum score of 9.0 due to partial compliance on QE3 or no compliance on the QE7 criteria. Likewise, a cut-off score of 8.0 would only leave a margin for having a single partial compliance on QE1, QE2, QE4 or QE8 criteria in similar cases, even though it would reject 160/295 (54%) records. Therefore, the 7.5/10.0 cut-off score is more appropriate for the quality assessment phase in this review by leaving margin for works to have partial compliance in more than one criterion. Indeed, this cut-off score allows an article with no public data and/or implementation (e.g., due to confidentiality agreements) to have up to four criteria with partial compliance, depending on the criterion's maximum score or if the work has available the experiments data and/or implementation. Another example is articles that only focus on localization or mapping. In these cases, the work could have no public implementation, even though requiring a maximum score on all other criteria, or, if the work has public data or implementation available, two other criteria could have partial compliance.

Overall, as illustrated in Figure 1, the quality assessment of the 411 eligible works considering the two rejection criteria previously mentioned leads to rejecting a total of 236 (57%) records. As a result, the remaining 175 records will be analyzed for data extraction.




(a)




(b)

Figure 2: Distribution of the quality evaluation scores obtained from assessing the eligible works considered in the review: (a) all eligible works; (b) works that pass the rejection criterion during the QE assessment related to QE2/3/5/8 = 0.0 (no compliance).

### 3.4 Data extraction

The data extraction process analyzes the records selected after the quality assessment phase and extracts information from these works. In the scope of this review, the Data Extraction (DE) items required for each record are the following ones:

- [**DE1**] **Long-term considerations** – long-term factors the works consider in their proposed approach and experiments. Considering the knowledge obtained in the previous phases of this review's methodology, the authors considered the following factors for categorizing the included works:
  - appearance: varying conditions, appearance changes;
  - dynamics: environment dynamics, dynamic elements;
  - sparsity: map pruning, redundant data removal;
  - multi-session: map management;
  - computational: memory management, efficiency.
- [**DE2**] **Localization** – how the robot localizes itself and the type of localizer;
- [**DE3**] **Mapping** – type of the map;
- [**DE4**] **Multi-robot** – if the proposed methodologies consider multi-robot systems;
- [**DE5**] **Execution mode** – offline, online, if requires both, or if no information on this item;
- [**DE6**] **Environment and domain** – type of environment (indoor, outdoor) and domains (air, ground, water) tested with the proposed methodologies;
- [**DE7**] **Sensory setup** – which sensors considered in the methodologies;
- [**DE8**] **Non-public experiments** – if the authors performed experiments or tests with non-public data;
- [**DE9**] **Ground-truth** – how ground-truth for non-public data is obtained or its type, if available;
- [**DE10**] **Distance and time characteristics** – relative to the non-public experiments if available, as follows:
  - total distance (km) of the non-public experiments;
  - path (km), in the case of repetitive paths;
  - total time (h) in terms of continuous operation;
  - time interval (day/week/month/year, or d/w/m/y) between the first and the last run.
- [**DE11**] **Datasets** – if and which public datasets are used in the experiments;
- [**DE12**] **Evaluation metrics** – which metrics are used for evaluation.

In Section 5, a comparison table of the public datasets identified by the DE11 will contain the sensory setup, ground-truth data availability from the datasets, and the distance and time characteristics, similar to the data extraction items for non-public data, among other aspects. As a result, the distinction between public and non-public data availability represented in DE8, DE9, and DE10 allows to understand the distance and time characteristics of non-public data independently from the public datasets.

Although the data extraction phase in a systematic literature review usually does not remove any records, 31 of the analyzed 179 works have extended versions of the proposed methodologies, more detailed ones, or equivalent methods applied in different conditions. Thus, these records are not included in the review to improve the discussion section in terms of singularity and originality of proposed approaches for the long-term localization and mapping problem. The extracted information helped identifying the corresponding extended and more complete versions of these works. A document containing the association of the removed versions to the records included in the review is available in the public GitHub repository, including their bibliographic references.

Consequently, 144 original works are included in this review for an overview of these records in Section 4, and their synthesis and discussion in Section 5. The information relative to the 12 data items for each of the included records is available in Appendix A and also in the repository. The included works represent 35% of the 411 eligible records for this review. This result indicates that the methodology followed in this review led to a high percentage of quality results.

## 4 Results Overview

In this section, the main goal is to overview the results not in terms of their scientific contribution but in terms of their bibliographic data for presenting an overview of the included records in the review. First, statistic results of the data sources in which the 144 included records could be identified in the methodology allow the evaluation of the coverage between the sources. Next, the tool VOSviewer (van Eck and Waltman 2010, 2014) is used to obtain the co-occurrence analysis for the keywords and the authors. The former focus on the keywords recency and their occurrence in the sources, while the latter discusses the research networks between the authors, and the ones with more publications in long-term localization and mapping. Lastly, two analysis are presented relative to the evolution of the publication year and most relevant publication venues.

## 4.1 Data source

The results on the identification phase are exported to BibTeX files from each data source. This exportation considers all the information available in the data sources, such as citation (e.g., author, title, publication venue, and type of record) and bibliographic (e.g., affiliation and the publisher) information of each record, the abstract, and author and indexed keywords. Next, using the `bibtexparser`[3] Python library, the BibTeX files are processed to identify uncompleted records. For example, the DOI must be specified and, if not available, the record's information must be manually completed with a corresponding URL. Then, considering the 144 included records in this review, a Python script searches each record in the BibTeX files corresponding to each data source. This search uses the DOI, URL, and title data to identify if a data source had in its identification results the searched record. Given that these three fields can contain lower and upper letters, the respective strings must be compared only after converting them to lower cases. As a result, the number of identified records by each data source of the 144 included ones in the review are the following ones:

- *ACM Digital Library*: 25 records (17.4%);
- *Dimensions*: 85 records (59.0%);
- *IEEE Xplore*: 68 records (47.2%);
- *INSPEC*: 104 records (72.2%);
- *Scopus*: 122 records (84.7%);
- *Web of Science*: 105 records (72.9%).

The database *Scopus* is the source that identified the greatest number of included records. This result was expected given that *Scopus* is considered as one of the largest curated databases (V. K. Singh et al. 2021), indexing more than 25000 active titles (e.g., conferences proceedings, journals) and 7000 publishers[4]. Two other sources with more than 70% of identified records are *INSPEC* and *Web of Science*. Similarly to *Scopus*, these two databases index also records from thousands of journals, conferences, and publishers[5,6]. Although *Dimensions* is also a bibliographic database covering millions of publications from thousands of sources, this database is the newest one (created in 2018) relative to the other three considered in this review (*INSPEC*, *Scopus*, and *Web of Science*) and could be a factor to why it obtained a lower percentage (59.0%) than the other three databases. Another possible reason is that *Scopus* and *Web of Science* have the majority of their coverage in Life Sciences, Physical Sciences, and Technology Area (including the Engineering subject area related to the topic of this review), while *Dimensions* has better coverage in Social Sciences and Arts & Humanities (V. K. Singh et al. 2021). Even though *IEEE Xplore* is a digital library and only indexes works published by IEEE and its partners, this data source returns 47.2% of the include records in the review. The main reason is that this library indexes publications related to electrical engineering and computer science, subject areas related to long-term localization and mapping[7]. Finally, the *ACM Digital Library* using *The ACM Guide to Computing Literature* collection only finds published records by ACM and possible links to other records focused exclusively on computing[8] and not directly related to the Computer Science or Engineering subject areas, explaining why this source obtained a lower coverage percentage of the included results than the other sources for this review.

Furthermore, Table 4 presents a coverage analysis of the identified results from each data source for the 144 included records in this review. Table 4a presents the pairwise overlap between sources. The corresponding percentage is the ratio of records identified by both sources to the one between the two that has the smallest number of results: $\#\{A \cap B\}/\min\{\#A, \#B\}$, where $\#A$ and $\#B$ is the number of results for a data source $A$ and $B$, respectively, and $\#\{A \cap B\}$ is the intersection results between the two sources. For example, if the pairwise results is 100%, it means that the data source with more records found was capable of obtaining all the results, i.e., had full coverage over the other source. Table 4b reports the percentage of records identified by at least one of two data sources over all 144 included records: $\#\{A \cup B\}/144$, where $A \cup B$ is the union correspondence results of the sources $A$ and $B$. This percentage represents the joint coverage of two databases over the 144 included records.

Table 4: Pairwise coverage analysis of the data sources considered in the review over the 144 included records: (a) identification only on both pairwise sources ($\#\{A \cap B\}/\min\{\#A, \#B\}$); (b) on either ones ($\#\{A \cup B\}/\#$records). Legend: dim – *Dimensions*, ieee – *IEEE Xplore*, insp – *INSPEC*, scop – *Scopus*, wos – *Web of Science*.

(a)

| $A \cap B$ | **acm** | **dim** | **ieee** | **insp** | **scop** | **wos** |
|---|---|---|---|---|---|---|
| **acm** | – | 96.0% | 44.0% | 88.0% | 96.0% | 96.0% |
| **dim** | – | – | 69.1% | 77.6% | 97.6% | 95.3% |
| **ieee** | – | – | – | 89.7% | 91.2% | 73.5% |
| **insp** | – | – | – | – | 87.5% | 68.3% |
| **scop** | – | – | – | – | – | 89.5% |
| **wos** | – | – | – | – | – | – |

(b)

| $A \cup B$ | **acm** | **dim** | **ieee** | **insp** | **scop** | **wos** |
|---|---|---|---|---|---|---|
| **acm** | – | 59.7% | 56.9% | 74.3% | 85.4% | 73.6% |
| **dim** | – | – | 73.6% | 85.4% | 86.1% | 75.7% |
| **ieee** | – | – | – | 77.1% | 88.9% | 85.4% |
| **insp** | – | – | – | – | 93.8% | 95.8% |
| **scop** | – | – | – | – | – | 92.4% |
| **wos** | – | – | – | – | – | - |

Analyzing the coverage results in Table 4, the first observation is that the pairwise union results of two sources increase the independent coverage of each source. This observation validates the need identified in the methodology discussed in Section 3 to consider several data sources in the identification phase of a review. Moreover, the pairwise union coverage of *INSPEC*, *Scopus*, and *Web of Science* is greater than 90% of the included records. When evaluating the joint coverage of these three databases, they identify all 144 of the included records, i.e., a 100% coverage. Although this result could indicate that those three sources guarantee full coverage of the long-term localization and mapping research topic, it is always advisable to consider as most as possible sources in the methodology. Another observation is relative to the overlap of *Scopus* with the other sources, which is greater than 85%. This overlap indicates that *Scopus* covers results not only on the topic of this review but also the results obtained by the other sources considered in the methodology. Lastly, *IN-*

[3]https://bibtexparser.readthedocs.io/en/master/
[4]https://www.elsevier.com/solutions/scopus/how-scopus-works
[5]https://www.elsevier.com/solutions/engineering-village/content/inspec
[6]https://clarivate.com/webofsciencegroup/solutions/web-of-science/
[7]https://ieeexplore.ieee.org/Xplorehelp/overview-of-ieee-xplore/about-ieee-xplore
[8]https://libraries.acm.org/digital-library/acm-guide-to-computing-literature

*SPEC* and *Web of Science* achieve a pairwise overlap percentage of 68.3% between themselves, while their union represents 95.8% of the included records. This discrepancy indicates that these two sources identify unique results between themselves. Indeed, *INSPEC* identifies 33/144 records not found by *Web of Science*, and vice-versa for *Web of Science*, with 34/144 unique records.

## 4.2 Keywords co-occurrence

Next, VOSviewer (van Eck and Waltman 2010, 2014) is used to analyze the co-occurrence of keywords in the included articles. This co-occurrence is the relatedness of items determined based on the number of documents in which the keywords occur together. For this analysis, first, a Python script processes the BibTeX file containing the citation and bibliographic information, the author and the indexed keywords, and the abstract of the records to join the author with the indexed keywords in the same `keywords` field. Then, an online tool[9] converts this processed BibTeX to a RIS file. Even though VOSviewer supports file types directly exported from *Dimensions*, *Scopus*, or *Web of Science* as input, none of these data sources obtained all the 144 included records of the review in the identification phase. Given that VOSviewer does not support BibTeX files, the conversion to RIS file is required for using as input. The disadvantage of using this file format in VOSviewer is only allowing to perform co-occurrence of items (e.g., keywords or authors), while bibliographic data from *Dimensions*, *Scopus*, or *Web of Science* in CSV files would allow other analysis such as citation, co-citation, or bibliographic coupling. However, the creation of these CSV files follow different templates depending on the data source. So, RIS files allow the integration of all 144 included records for obtaining the two co-occurrence analysis presented in this review (namely, keywords and co-authorship).

In Figure 3a, the network presents the overlay visualization of the keywords co-occurrence in the included records weighted by the number of occurrences of each term, using full counting for the links' strength. The latter computes the strength of the links directly by the number of co-occurrences of the respective two terms. The overlay visualization colors the keywords differently according to the average publication year of the included records in which each of the keywords appears. This coloring allows analyzing which are the ones that are associated with the most recent publications. As for the keywords' weighting, the number of occurrences dictates the size of the circles. Furthermore, the minimum number of occurrences of a keywords set in VOSviewer for obtaining the graph is 5 originating the 35 keywords illustrated in Figure 3a. This parameter was selected for visualization purposes while also filtering uninteresting keywords. Similarly, setting the attraction and repulsion parameters to 2 and 0, respectively, distances the terms more from each other than using the values recommended in the VOSviewer manual[10] (2 and 1, respectively). These two parameters only interfere in the localization of the terms in the map, not in the graph connections. Lastly, a thesaurus of the keywords (available in the repository) is used to join similar terms: spelling differences (e.g., localization – localisation), full terms versus abbreviations (simultaneous localization and mapping – SLAM), while also allowing the concatenation of long keywords for visualization reasons.

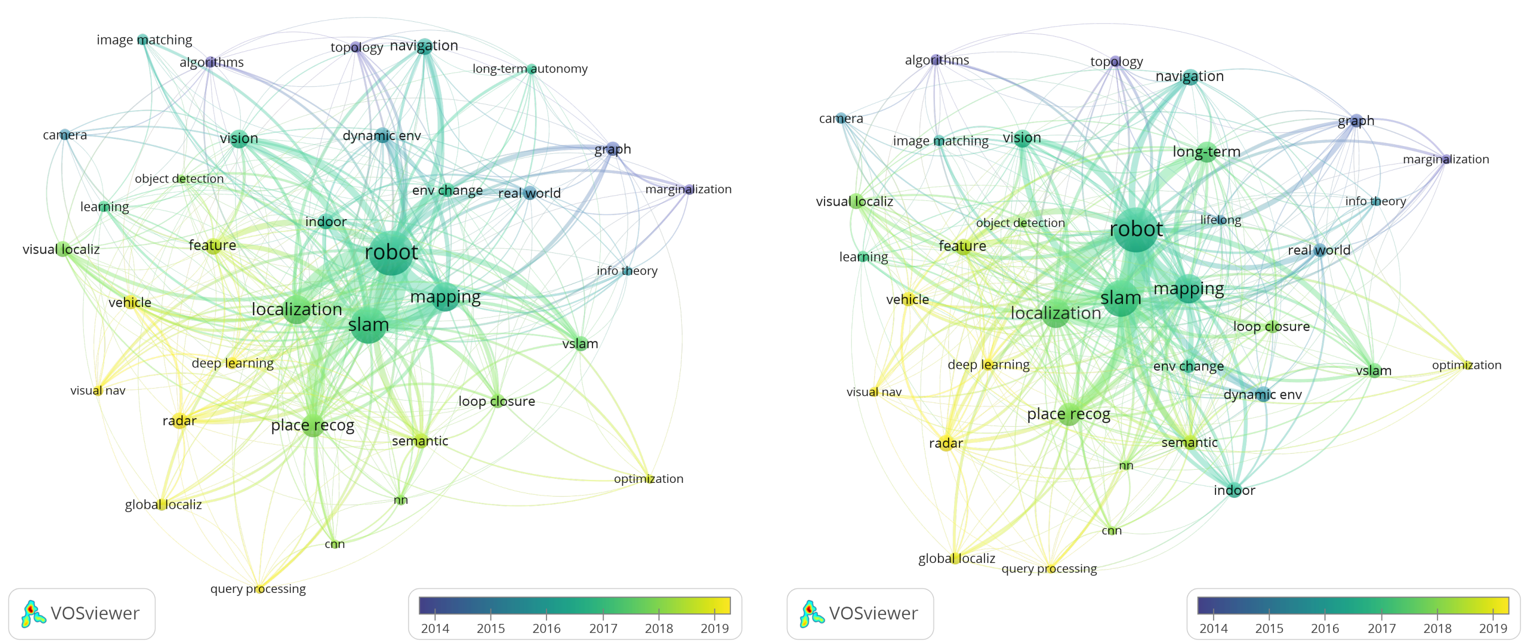
Overall, the keyword `robot` is the one that appears more times in the included records: 111 occurrences, links with 34 other terms, and has a total link strength of 403 (sum of co-occurrences of all of its links). This result is expected due to the relation of this review's topic to robotics. Similarly, three other keywords in the network related to long-term localization and mapping topic with high values of occurrence, number of links, and total link strength are `slam` (75, 34, and 288), `mapping` (48, 33, and 204), and `localization` (47, 32, and 194, respectively). The methodology for the search strategy discussed in Section 3.2 considers all of these four keywords. Thus, the significant influence of `robot`, `slam`, `mapping`, and `localization` in the keywords co-occurrence analysis indicates that, after the all the phases executed in this review's methodology, the 144 included records have a high correlation with the keywords considered in the search query. Given that the keywords are usually selected or indexed to capture the essence of the document, this correlation indicates that the search query is appropriate to obtain the search results, even considering only the keywords as search fields.

As for keywords related to the outcome of the PICO framework, `long-term autonomy` occurs only 6 times in the included records, linking with 16 other keywords and having a total link strength of 27. This low occurrence could indicate that the term `long-term autonomy` is not usually used by the authors nor indexed by the databases. However, the specific term of `long-term autonomy` does not summarize all the possibilities for the outcome of the PICO framework (see Section 2). Indeed, for this reason, the search query for the identification phase uses only the following single terms: `"long term"` and `"life long"` (resumes the possibility of having a space or a hyphen), and `lifelong`. Figure 3b presents the keywords co-occurrence analysis using the same parameters for obtaining Figure 3a. The difference to the latter network is using a thesaurus that summarizes all the keywords that contain `long-term` and `lifelong` into the terms themselves, obtaining 36 keywords with a minimum of 5 occurrences in the 144 included records. In terms of occurrences, number of links, and total link strength, the impact of the thesaurus keyword `long-term` is 25, 28, and 105, and for lifelong 6, 17, and 31, respectively. These values are much higher than the ones respective only to `long-term autonomy` from Figure 3a. The reason is that `long-term` in Figure 3b compiles the occurrences of keywords such as `long-term autonomy`, `long-term mapping`, and `long-term localization` (6, 2, and 2 occurrences, respectively), and `lifelong` sum up, for example, three different versions of `lifelong learning` (using `lifelong`, `life-long` and `life long` with 2, 1, and 2 occurrences, respectively) and `lifelong slam` (1 occurrence). Hence, these results proves that the third `AND` part of the search query (`"long term" OR "life long" OR lifelong`) covers well the PICO framework's outcome. Plus, they also show no consensus among the authors and by the databases indexation on how to define a keyword for the topic of long-term localization and mapping.

In terms of the average year of publication, analyzing the diagrams in Figure 3 on its colorization, the first observation is the recency of terms related to visual localization. The keywords visual SLAM (`vslam`), visual navigation (`visual nav`), and visual localization (`visual localiz`) have all an average publication year higher than 2017. This recency indicates that recent approaches related to the topic of this review, long-term localization and mapping, are more inclined to use vision as a sensorization input. Another sensor that appeared with high relevance in the network is `radar`, with 15 occurrences and an average publication year of 2019.20. This sensor is agnostic to the environment changes such as illumination and season changes intrinsically associated with vision and could be the reason why the recent works related to

[9] https://www.bibtex.com/c/bibtex-to-ris-converter/

[10] https://www.vosviewer.com/documentation/Manual_VOSviewer_1.6.8.pdf




(a) (b)

Figure 3: Keywords co-occurrence analysis on the 144 included records generated by VOSviewer with overlay visualization by the average publication year: (a) original keywords; (b) all keywords containing long-term and lifelong summarized by the terms themselves. Parameters used for generating the co-occurrence network: minimum number of occurrences = 5, attraction = 2, repulsion = 0, scale = 1.49, circles size variation = 0.5, lines size validation = 1.0. Legend: `cnn` – Convolutional Neural Networks, `env` – environment, `localiz` – localization, `nav` – navigation, `nn` – Neural Networks, `recog` – recognition, `vslam` – visual SLAM.

long-term localization and mapping are using it. Moreover, place recognition (`place recog`) stands out not only by its recency but importance. The keyword itself (`place recog`) occurs 31 times and an average publication year of 2017.77, with terms related to place recognition such as `loop closure` and global localization (`global localiz`) with recent average publication years (2017.82 and 2018.75, respectively) and strong link to place recognition (5 co-occurrences for each of the links between `loop closure` and `global localiz` with `place recog`). Lastly, machine learning also seems to be used in recent works included in this review. The keywork learning occurs 7 times with an average publication year of 2017.00. Neural Netowrks (`nn`), Convolutional Neural Networks (`cnn`), and `deep learning` have a similar number of occurrences (6, 5, and 8) and publication years higher than 2017 (2017.83, 2018.00, and 2019.12, respectively). These results could mean another trend of using machine learning to improve the long-term autonomy of mobile robots.

Although the recency of keywords related to dynamic environments is lower than 2017 (2015.50 and 2016.75 for `dynamic env` and `env change`), they have a high occurrence (14 and 12, respectively), located close to each other in the network, and have a strong link between them (4 co-occurrences). Three keywords also located near each other are `graph`, `marginalization`, and information theory (`info theory`) while having similar average publication years (2014.36, 2014.00, and 2015.50, respectively). Even though the number of occurrences of these terms is low (11, 6, and 5 for `graph`, `marginalization`, and `info theory`, respectively), their map proximity could indicate a focus in the past on the topic of graph sparsity, i.e., maintaining the graph in the long-term to only depend on the environment size and not on the robot's operation time. However, all of these tendencies indicated in this keywords co-occurrence analysis among other ones will be discussed in Section 5 in further detail.

## 4.3 Co-authorship analysis

The other analysis obtained using VOSviewer is the co-authorship network presented in Figure 4. Similar to the keywords network illustrated in Figure 3, the co-occurrence of the authors' names creates links among them in the graph. The strength of these links is dictated by the number of documents the two authors of a link are co-authors in the same record, and the number of co-authored works determines the size of the circles respective to each author in the graph. In contrast to Figure 3, the network in Figure 4 does not have any overlay specific to coloring depending on the average publication year. Instead, the main goal of the co-authorship analysis in this review is to present possible research networks detected in the 144 included records. Thus, the coloring in Figure 4 represents the clusters of authors detected by VOSviewer. This network only considers authors with a minimum of 3 works for relevance and visualization reasons, resulting in 29 authors. Also, authors identified only by the initial of the first name and by the surname can lead to incorrect correspondences in terms of co-authorship. VOSviewer detects 392 authors in the 144 included records using the original RIS file used in Section 4.2 compared to 413 after checking the authors names. Indeed, a manual check is performed on all authors of the included records to guarantee no false correspondences for the co-authorship analysis with VOSviewer. This manual check ensures each author has its full first and surname and any middle initials while also using the same name for an author in different records.

Analyzing Figure 4, the co-authorship network presents 8 clusters. These clusters are separated from each other, i.e., no link exists between authors from different clusters. However, this separation does not mean that there is not any co-authorship between authors from different clusters only indicating that for a miminum of 3 co-authored documents there is not a connection between these 8 clusters. Even so, the graph presented in Figure 4

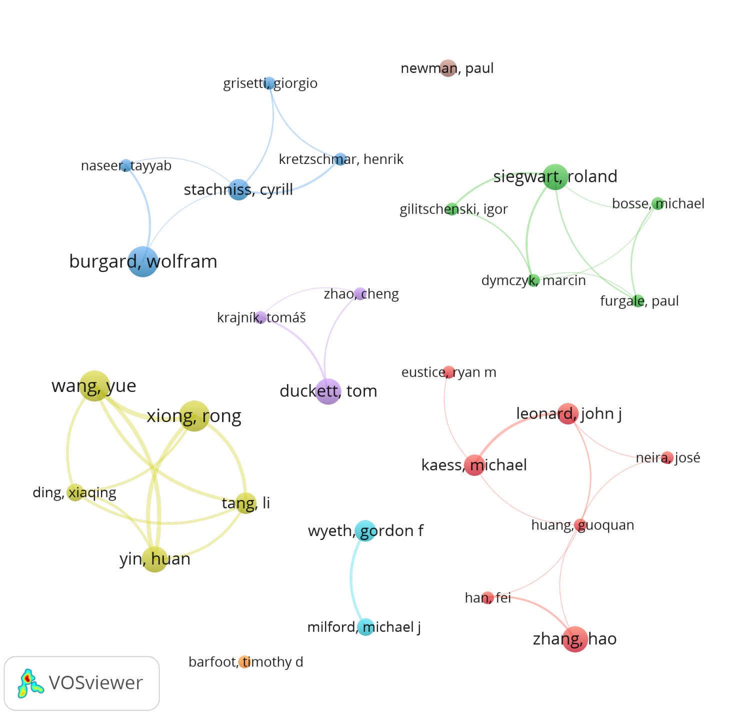



Figure 4: Co-authorship analysis on the 144 included records generated by VOSviewer. Parameters used for generating the co-occurrence network: minimum number of occurrences = 3, attraction = 2, repulsion = -3, scale = 1.49, circles size variation = 1.0, lines size validation = 1.0.

allows the identification of the most relevant research networks in terms of number of co-authored documents and in the context of long-term localization and mapping, considering the 144 records included in this review. As a results, the following enumeration presents the authors that belong to each cluster in the format of author (number of co-authored documents):

1. Rong Xiong , (7), Yue Wang , (7), Huan Yin , (6), Li Tang (5), and Xiaqing Ding , (4);
2. Hao Zhang (6), John J. Leonard , (5), Michael Kaess (5), Fei Han (3), Guoquan Huang (3), José Neira , (3), and Ryan M. Eustice , (3);
3. Wolfram Burgard , (7), Cyrill Stachniss , (5), Giorgio Grisetti , (3), Henrik Kretzschmar (3), and Tayyab Naseer , (3);
4. Roland Siegwart , (6), Igor Gilitschenski , (3), Marcin Dymczyk , (3), Michael Bosse (3), and Paul Furgale , (3);
5. Tom Duckett , (6), Cheng Zhao , (3), and Tomáš Krajník , (3);
6. Gordon F. Wyeth , (5) and Michael J. Milford , (4);
7. Paul Newman (4);
8. Timothy D. Barfoot , (3).

When analyzing the affiliations of the authors mentioned previously at the time of publication, all authors of the first cluster belonged to the State Key Laboratory of Industrial Control and Technology (SKLICT) and the Institute of Cyber-Systems and Control at Zhejiang University in China. Even though Huan Yin, Yue Wang, Xiaqing Ding, Li Tang, and Rong Xiong mention their affiliation to the Joint Centre for Robotics Research between Zhejiang University, China, and the University of Technology Sydney, in the work (H. Yin, Y. Wang, et al. 2020), this specific affiliation only appeared in this article. The total link strength (sum of all links weights) of each of the authors in that cluster is higher than 16, meaning a high co-authorship between them. Indeed, all five authors have links between all of them. Similar to the first cluster, the third, fourth, fifth, and sixth clusters have common affiliations within each one: the Autonomous Intelligent Systems at the University of Freiburg in Germany, the Autonomous Systems Lab (ASL) at ETH Zürich in Switzerland, the Lincoln Centre for Autonomous Systems (LCAS) at the University of Lincoln in UK, and the School of Electrical Engineering and Computer Science at Queensland University of Technology (QUT) in Australia. However, the interlinking between the authors is not as strong as in the first cluster, as shown in Figure 4 by the authors of these clusters not being connected between all the ones within each cluster. Even so, the common affiliation shows there is considerable interest by these research units in the long-term localization and mapping topic.

The affiliation analysis in the second cluster is more complex given that there was no affiliation common to all authors at the time of the records' publication. Instead, the following affiliations were found: Fei Han and Hao Zhang with the Department of Computer Science at Colorado School of Mines in the USA, Guoquan Huang with the Department of Mechanical Engineering at the University of Delaware in the USA, John J. Leonard and Michael Kaess with the Computer Science and Artificial Intelligence Laboratory (CSAIL) at the Massachusetts Institute of Technology (MIT) in the USA, Ryan M. Eustice with the Perceptual Robotics Laboratory (PeRL) at the University of Michigan in the USA, and José Neira with the Instituto Universitario de Investigación en Ingeniería de Aragón (I3A) at the Universidad de Zaragoza in Spain. Although there are 5 different affiliations to which the 7 authors stated in the respective records, 4 of the research institutions noted for the second cluster are in the USA, indicating a possible reason for facilitating the linkage between these authors from different research units.

In terms of the clusters composed by single authors, the affiliations of Paul Newman and Timothy D. Barfoot are the Oxford Robotics Institute at the University of Oxford in UK and the Autonomous Space Robotics Laboratory (ASRL) at the University of Toronto Institute for Aerospace Studies (UTIAS) in Canada, respectively. Even though these two authors are not linked with any others in the network, the co-authorship analysis indicates that they have an interest in long-term localization and mapping. This interest is shown by their number of co-authored records: 4 and 3 by Paul Newman and Timothy D. Barfoot, respectively.

As for the number of co-authored publications, considering the 144 included records, the authors that appeared to have more research on the review's topic are Rong Xiong, Yue Wang, and Wolfram Burgard, given the 7 co-authored publications of each one. However, Rong Xiong and Yue Wang have co-authored the 7 documents attributed to each of them. This relation and similar ones can biase the analysis of which authors are having more impact in the review's topic. The clustering shown in Figure 4 allows a more unbiased analysis relative to the co-authorship links between authors. Thus, based on the clustering and which author from each cluster has the most co-authored publications, the most influential authors in long-term localization and mapping are the following ones: Rong Xiong (or Yue Wang), Hao Zhang, Wolfram Burgard, Roland Siegwart, Tom Duckett, Gordon F. Wyeth, Paul Newman, and Timothy D. Barfoot.

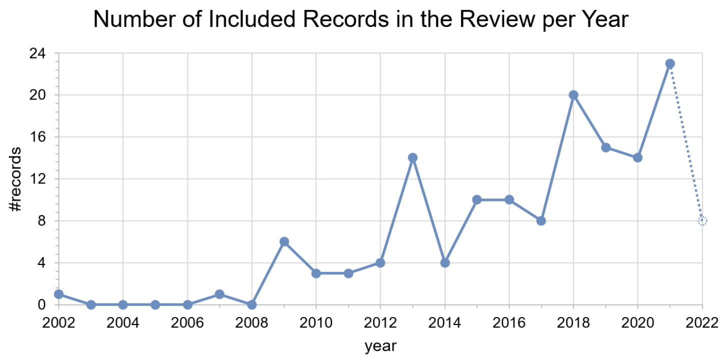



Figure 5: Evolution of published records per year considering the 144 included records in this review

## 4.4 Year of publication

The relevance of the long-term localization and mapping topic can be evaluated by the evolution of the number of publications. Figure 5 presents this evolution from the earliest year of publication of the included records to the year at the time of writing this article. The latter has its respective data dashed to indicate that the last year is not completed at the time of writing. Analyzing Figure 5, this review's topic seems to have gain relevance in 2009 with 6 works, compared to only one publication in 2007 and another in 2002 in the previous years to 2009. From that year onwards, the graph has an almost linear tendency reaching a maximum of 23 records in 2021, while already having 8 publications in 2022 until May 17, 2022. This tendency shows that long-term localization and mapping is gaining interest throughout the years and, consequently, supports the importance and relevance of this review for the scientific community.

## 4.5 Publication venue

Finally, the last overview of the 144 included records in the review is relative to the publication venue. Table 5 presents the venues with more than 1 publication, separating the journals and conferences in two different tables (Tables 5a and 5b, respectively). The columns $\mu$ present the average year of publication of the records associated to a certain venue, while max columns display the publishing recency by the year of the most recent publication in the venue. For comparing to the average value ($\mu$), the third column ($\sigma$) of each table presents the standard deviation based on the publication year data. The last column state the number of records published in the venue from the 144 records included in the review for discussion.

In terms of journals, the Robotics and Autonomous Systems, IEEE Robotics and Automation Letters, and the International Journal of Robotics stand out with more than 10 publications. Also, these journals have a high standard deviation (greater than 1.5), indicating that the publications spread out throughout the years. In the case of the IEEE Robotics and Automation Letters, these results gain more relevance indicating a recent trend on publishing on this journal, considering that its creation was only on 2015[11]. With more than 5 publications, the Journal of Field Robotics and the Autonomous Robots have recent average of publication (2017) with a high standard deviation (greater than 2.0), similarly indicating that authors have been publishing in these two journals along the years. In contrast, the IEEE Transactions on Intelligent Transportation Systems and Sensors journals have a standard deviation lower than 1 year, with an average publication year of at least 2019. The recency of publication on these two journals with a very low deviation suggests a recent interest of the authors to publish in these two journals works related to long-term localization and mapping.

As for conferences, the data in Table 5b shows a high discrepancy in the number of publications related to this review's topic in ICRA and IROS compared to the other venues. Indeed, all the other conferences have only a maximum of 3 records published in them, compared to 22 and 18 papers in ICRA and IROS, re-

[11]https://www.ieee-ras.org/publications/ra-l

Table 5: Publication venues of the included records in this review with more than one record published in the venue: (a) journals; (b) conferences. Legend: $\mu$ – average year of publication, $\sigma$ – standard deviation of the publication year, max – maximum year of publication, # – number of records published at a certain venue

(a)

| | Year | | | |
|---|---|---|---|---|
| **Journal** | $\mu$ | $\sigma$ | max | # |
| Robotics and Autonomous Systems | 2016 | 3.9 | 2021 | 13 |
| IEEE Robotics and Automation Letters | 2019 | 1.7 | 2022 | 12 |
| International Journal of Robotics Research | 2014 | 3.2 | 2022 | 11 |
| Journal of Field Robotics | 2017 | 3.5 | 2022 | 8 |
| Autonomous Robots | 2017 | 2.2 | 2020 | 7 |
| IEEE Transactions on Intelligent Transportation Systems | 2021 | 0.8 | 2022 | 4 |
| Sensors | 2019 | 0.8 | 2020 | 4 |
| IEEE Transactions on Robotics | 2017 | 3.1 | 2022 | 4 |
| IEEE Sensors Journal | 2020 | 1.5 | 2021 | 2 |
| International Journal of Advanced Robotic Systems | 2020 | 1.5 | 2021 | 2 |

(b)

| | Year | | | |
|---|---|---|---|---|
| **Conference** | $\mu$ | $\sigma$ | max | # |
| IEEE International Conference on Robotics and Automation (ICRA) | 2016 | 3.9 | 2021 | 22 |
| IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS) | 2017 | 3.8 | 2021 | 18 |
| IEEE International Conference on Robotics and Biomimetics (ROBIO) | 2019 | 2.1 | 2021 | 3 |
| IEEE International Intelligent Transportation Systems Conference (ITSC) | 2018 | 2.4 | 2021 | 3 |
| European Conference on Mobile Robots (ECMR) | 2014 | 0.9 | 2015 | 3 |
| IEEE Intelligent Vehicles Symposium (IV) | 2019 | 0.5 | 2019 | 2 |
| International Conference on 3D Vision (3DV) | 2018 | 1.5 | 2019 | 2 |
| International Conference on Advanced Robotics (ICAR) | 2011 | 2.0 | 2013 | 2 |

spectively. When considering that 62 of the 144 included records are published in conferences, ICRA and IROS with a total of 40 published works related to this review's topic represent 65% of works published in conferences and 27.8% of all included records. This result expresses the high relevance of ICRA and IROS in the topic of long-term localization and mapping.

# 5 Discussion

The keywords co-occurrence analysis discussed in Section 4.2 identifies terms with large occurrence and strong links between each other, that can be related to the categorization considered for the data extraction item DE1 (see Section 3.4). The terms place recognition, global localization, and loop closure are associated to varying appearance of the environment, equivalent to the first category of DE1 (appearance). The second one (dynamics) is related to the group of keywords corresponding to dynamic environments. Although methods focused on varying appearance and conditions can possibly deal with dynamic elements in the scene, these methods do not identify specifically those elements neither model the dynamics of the environment. As for the other group of keywords related to graph and information theory, these works focus on removing uninformative data from the map (Kretzschmar and Stachniss 2012), which are related to map sparsification and to the third category of DE1 (sparsity). These relations between the categories appearance, dynamics, and sparsity to the semantic analysis of the keywords co-occurrence supports the categorization of DE1, while also indicating that the discussion on the proposed methodologies should focus on each one of the categories. Even though the two remaining categories of DE1 (multi-session and computational) are not represented in the keyword analysis, the execution of the data extraction phase identified the need for having these two categories, given the importance of multi-session handling and computational efficiency for long-term localization and mapping.

Furthermore, the discussion section should focus not only on the proposed methodologies but also on the evaluation of experimental results. Indeed, the data extraction items DE8–12 are all intended to retrieve information on this matter. These data items help identify the evaluation metrics used to assess the methods performance, which datasets were used, and the characteristics of the experiments performed by the authors.

Accordingly, this section is organized as follows. First, the included works in the review are discussed by the categories of DE1 separately, focusing on the proposed methodologies and identified trends. Section 5.1 discusses methods related for dealing with varying appearance of the environment and place recognition. Section 5.2 reviews articles focused on modeling the environment dynamics or identifying dynamic objects within the environment. Section 5.3 focus on methods for removing redundant data of the map or identifying novelty data to keep the map size constrained to the environment size. Section 5.4 discusses how methods handle multi-session in terms of mapping. Section 5.5 reviews works related to computation concerns over long-term localization and mapping, in addition to the ones relative to map sparsification discussed in Section 5.3. Then, the evaluations metrics used in the included works are discussed in Section 5.6. Section 5.7 analyzes the experimental data used for evaluating the proposed methodologies, including the datasets used and characteristics of the experiments. Finally, Section 5.8 is reserved to final observations of the discussion relative to the localization and mapping algorithms, and the sensorization more used by the authors of the included records.

## 5.1 Appearance variance

Probably put here an intro more detailed for the appearance variance subsection!!!

### 5.1.1 Experience maps

One way to deal with the appearance variance of environments is by treating different conditions as multiple experiences. The biologically inspired RatSLAM (Ball et al. 2013) introduces the experience map as a semi-metric topological map, where each experience is a view of the environment at a certain position and wheel odometry provides the relative pose for the links. New experiences are created when none of the previous ones saved in the map are sufficiently similar in appearance to the current scene. Glover et al. 2010 combines the mapping of RatSLAM with the place recognition of FAB-MAP (Cummins and Newman 2008). The latter improves the loop closure detection of the original RatSLAM due to FAB-MAP having light invariant characteristics for data association by learning a generative model for the Bag of Words (BoW) model (Sivic and Zisserman 2003). Both RatSLAM and the hybrid RatSLAM+FAB-MAP systems uses visual data to retrieve information from the environment. Although Martini et al. 2020 uses also experience-based mapping, the main sensor is a radar, where an experience is represented by a point cloud from the sensor and the point descriptors retrieved from it. Radar is known for being less affected by environment changes such as different illumination or weather conditions compared to vision sensors (Hong et al. 2022).

The concept of adding the environment changes to the map identified by the degradation in localization is also employed by Konolige and Bowman 2009 and Tang, Y. Wang, Ding, et al. 2019. The former implements a keyframe SLAM system created from the Visual Odometry (VO) module, where each keyframe represents a view of the environment, while a place recognition module tries to match the current frame to similar views already in the map for loop closure. The latter applies a similar idea to experience maps based on the 2D manifold assumption for locally smooth navigation. Even though the proposed topological local-metric framework encodes geometric information in the edges, the nodes do not require global pose, i.e., no restriction for global consistency. New nodes are trigered either from localization failure or after a certain length is traveled by the robot. The goal is to restrict the erroneous alignment computed from odometry locally.

Instead of considering an experience as a location or a view of the current scene, Churchill and Newman 2013 defines it as a whole sequence of the saved poses and related features directly obtained from VO. In this case, the topological mapping links experiences not geometrically but instead if two experiences observe the same space. However, the method does not implement a specific place recognition module for loop closure, assuming that the robot will subsequently return to a place that can have successful localization. Gadd and Newman 2016 builds on the work of Churchill and Newman 2013 for multi-robot systems. This method adds FAB-MAP for place recognition in the existing map maintained by a centralized versioning framework. The selection of the most relevant experiences by the centralized framework for localizing multiple agents in the system assumes that appearance change is only driven by the passage of day time.

Another example of experience maps is Visual Teach & Repeat systems using spatial-temporal pose graphs, as implemented in MacTavish et al. 2018 and N. Zhang et al. 2018. Similar to Churchill and Newman 2013, an experience is the output of the VO module defining the appearance of a scene throughout a path. In the teaching phase, the robot is teleoperated by humans creating privileged experiences in the graph. Autonomous experiences are the ones relative to the repetition phase. These experiences are linked either temporally or spatially if they are sequential in time or related metrically by multi-experience matching, respectively. Unlike Churchill and Newman 2013, new experiences have a known metric pose relative to the others in the pose graph.

In general, experience-based navigation methods try to generate new experiences if the environment changes, expecting that at a certain point in time the robot will be able to localize itself relative to previous experiences, not requiring new ones to be added to the map. However, these approaches are not scalable in the long-term timeframe nor to deal with dynamic elements, even using central servers as in Gadd and Newman 2016 with more computational resources than the robots. Pruning algorithms would be required to remove redundant or outdated information, as in Konolige and Bowman 2009 or Tang, Y. Wang, Ding, et al. 2019. Also, other methods should be employed to deal not only with long-term appearance changes (weather conditions or seasonal changes) but also with dynamic elements in the scene.

### 5.1.2 Illumination transformations

As a preprocessing step, illumination invariant transformations can be applied to color images for increasing the robustness of visual localization to changing lighting conditions and shadows. One example is the illumination invariant space that combines the log-responses of the 3 color channels into an one-dimensional space with a weighting parameter conditioned by the peak spectral responses of each channel, usually available in the camera specifications. This one-dimensional space is only dependent on the sensor and elements in the scene, while being independent of the intensities and colors. Both works of Arroyo et al. 2018 and Z. Yang et al. 2021 uses this transformation for preprocessing the color images into grayscale ones demonstrating the robustness of the illumination invariant space when lighting changes appear.

An alternative to using predefined illumination invariant transformations is to learn them. Clement et al. 2020 learns a nonlinear transformation mapping function from the RGB color space to grayscale also combining the three-channel log-responses, but relaxing the constraints of the one-dimensional space due to the original weighting parameter used in Arroyo et al. 2018 and Z. Yang et al. 2021. Instead of using the same parameters independently of the image content, Clement et al. 2020 trains an encoder to predict the optimal transformation weighting parameters of the three-channel log-responses. The objective function chosen for maximization is based on the number of inlier feature matches from a vision localization pipeline. The learned nonlinear RGB to grayscale transformation helped achieving a full-day cycle using a single mapping experience and the applying the optimized transformation to the color images.

Even though the Gamma correction does not transform an image to an invariant color space, this transformation can be used to strengthen low-illumination changes. Li Sun et al. 2021 uses the Gamma transform to synthesize low-illumination night-time images from daytime ones. Applying the transformation in the HSV (Hue, Saturation, Value) space, the gamma parameter adjusts the value channel without distorting the colors. Then, the synthesized images are used for training the DarkPoint descriptor proposed by Li Sun et al. 2021 to improve day-to-night matching.

### 5.1.3 Handcrafted features

Many localization and mapping algorithms rely on detection and extraction of features. The designation of handcrafted features refers to properties derived from the sensors data as a two-step process: a keypoint detector to locate the features and their characterization by computing a descriptor capable of distinguishing each feature from the others (Nanni et al. 2017). Algorithms for long-term localization and mapping using handcrafted feature should be robust to changing conditions such as illumination, appearance, weather and seasonal changes.

**Visual features** A way to improve long-term feature-based visual localization is to enhance the descriptiveness of visual feature descriptors and their long-term stability. Kawewong et al. 2013 defines the Position Invariant Robust Features (PIRF). In a sliding window framework, PIRF tracks the motion of local features such as Scale-Invariant Feature Transform (SIFT) or Speeded Up Robust Features (SURF) selecting the stable ones. Using an incremental tree-like PIRF (with inverted index as in BoW) dictionary, the method has shown robustness to viewpoint variance and unstable features. Also, PIRF-based localization improved the recall over FAB-MAP in the experiments.

Moreover, Histogram of Oriented Gradients (HOG) features have been used in different works to improve robustness to appearance variance, given that HOG descriptors capture local gradient information robust to seasonal changes (Naseer, Suger, et al. 2015). Li et al. 2015 computes local HOG descriptors from visually-salient image patch features in an underwater environment. Using a trained Support-Vector Machine (SVM) to classify the matching between corresponding patches, the proposed method achieved approximately 80% accuracy with dramatic appearance changes. Although Naseer, Suger, et al. 2015 computes HOG descriptors from each cell of a partitioned image, a global descriptor for the whole image joins all the ones respective to each cell. The global descriptor proved to be robust to foliage color changes, occlusions, and seasonal changes. Vysotska et al. 2015 uses the same global HOG descriptor as in Naseer, Suger, et al. 2015, but applied to image sequence matching requiring a rough global pose estimation for the images (e.g., GPS) for efficient matching.

Local Difference Binary (LDB) features also include gradient comparisons. These features are used in the Able for Binary-appearance Loop-closure Evaluation (ABLE) (Arroyo et al. 2018) approach to achieve higher descriptiveness power for appearance invariance. ABLE outperformed FAB-MAP in terms of precision-recall evaluation metrics. An advantage of using binary features such as LDB is the possibility of using the Hamming distance to compute descriptor similarity, improving the computational efficiency of this process over cosine similarity or Euclidean distance.

Another work from the included records focused on improving the long-term performance of handcrafted visual features is from Karaoğuz and Bozma 2016. Their approach uses bubble descriptors for preserving the relative $S^2$ geometry of visual features, being rotationally invariant. The experimental results demonstrated improvements on viewpoint and illumination invariance of bubble features-based localization.

Even though Cao, Yan, et al. 2021 requires a 2D or a 3D laser for place recognition and not a visual sensor, the proposed method transforms a 3D point cloud (acquired either from accumulating a sequence of 2D laser scans or directly from the 3D laser) to its 2D polar image representation. This transformation uses the centroid of the point cloud to ensure viewpoint invariance.

Then, using Gabor filters to detect and describe the contours of the images, Cao, Yan, et al. 2021 generates Binary Robust Independent Elementary Features (BRIEF) descriptors for matching images using an Approximate Nearest Neighbors (ANN) search. In addition to showing the seasonal appearance variance in laser data (e.g., different foliage in the scene), the proposed methodology outperforms SeqSLAM (Milford and Wyeth 2012) (sequential place recognition) and PointNetVLAD (Uy and Lee 2018) (CNN-based place recognition for 3D point clouds) on precision-recall.

**Environment structure features** The structure of the environment defined by its geometry is more robust to appearance variance than the appearance itself. Common structure features extracted from sensors data are line and edge features. Biswas and Veloso 2013 extracts 2D line segments corresponding to the walls from depth and 2D laser sensors. The line segment-based localization had a low failure rate on an over-a-year long-term indoor deployment even in areas with movable objects, due to the long-term stability of the line segment features. Nuske et al. 2009 extracts 3D edge features of the scenes using a monocular camera to get the edges of the buildings in the environment, while employing an exposure control to maximize the strength of edges corresponding to the mapped ones. The proposed method was able to successfully track the edges of the buildings along an all-day outdoor experiment. Instead of using the walls of the buildings, An et al. 2016 formulates a visual node descriptor based on ceiling salient edge points. Even though the method achieved good results in lighting changing conditions, the method's performance decreases using low and inclined ceilings, due to the image perspective effect that may lead to matching failure in the implemented Iterative Closest Point (ICP) framework.

Furthermore, Meng et al. 2021 extracts edge and planar features by evaluating the large and small values of the local surface smoothness over the points of a 3D laser, respectively. ICP estimates the laser odometry while the histogram cross-correlation of the Normal Distribution Transform (NDT) that computes local probability density functions of the surface smoothness identifies the loop closures. The proposed methods outperformed an ICP-based SLAM approach on Absolute Trajectory Error (ATE) in the experiments. As for Bosse and Zlot 2009, 2D point clouds segmented into connected components are clustered at regions of high curvature to get high curvature keypoints from multiple scans. The proposed descriptor based on the moment grid improves outdoor place recognition relative to SIFT or Hough transform peaks due to the moment grid descriptor includes higher order of moments relative to other descriptors.

Poles are structures also used for long-term localization. Schaefer et al. 2021 retrieves the 2D coordinates of poles registered with a 3D laser. Results demonstrated the ability of reliable long-term localization over more than one year. In addition to poles, Berrio, Ward, et al. 2019 extracts also corner features from the 3D laser point cloud, being able to localize over a 6 month experiment at different times of the day.

Another possible application of environment structure features found in the included works is in crop fields for agriculture. Chebrolu et al. 2018 formulates an aerial image registration algorithm based on the positions of the crops and the gaps between them remaining the same over time. The method computes a vegetation mask by exploiting the Excess Green Index (ExG) of RGB images. Using the Hough transform to find lines between vegetation, the center of the crops are the peaks on vegetation histograms perpendicular to the rows. The testing results demonstrated invariance of the registration algorithm to changing conditions caused by weather and crop growth over one month.

#### 5.1.4 Convolutional Neural Networks (CNN)

Possibly, discuss the recency of these methods as an introduction to this subsubsection.

MISSING HERE INTRO - COMPARE THESE TO HANDCRAFTED IN TERMS OF RECENCY E.G.

**CNN-based features** zhu-et-al:2018:8500686 VGG16, global feature Zhu uses the VGG16 network and its layers conv3_3, conv4_3, and conv5_3 for feature extraction from the images. Forming a global feature for the image and then normalizing the descriptor from float to a binary vector, images are compared using the Hamming distance. Performance of every layer outperforms FAB-MAP, and the fusion of the layers outperforms the results of using features from individual layers.

yang-et-al:2021:12054 VGG16, global feature, max pooling Yang extracts local features using VGG16 and evaluated the performance from different layers of the network. Fully connected layers suffer from poor accuracy due to spatial information increasing when deep layers in the network, choosing pool 5 (max pooling due to dimension reduction) due to its better accuracy and feature size. PLSAV (Parallel loop searching and verifying for loop closure detection) had better results than BoW-based approach and FABMAP. Feature map from layer flattened to vector, normalized using L2-norm, and used as high-level feature description of the image.

sun-et-al:2021:9635886 VGG, triplet loss, local feature, DarkPoint Sun uses the Gamma transform as a non-linear method to strengthen low-illumination (night-time) changes, applying it to normal daytime images to generate low-illumination night-time images. Given that transforming RGB images introduces distortions in the colour channels of the correlated image, for preserving colour information, transform first to HSV (Hue, Saturation, Value) space, and the gamma parameters adjusts the value of the lightness channel exponentially, based on the maximum lightness values. The adjusted channel value together with original hue and saturation will be converted to grey image. Sun uses VGG-like architecture for keypoint feature detection. Applying random illumination transforms (translation, scale, in-plane rotation and symmetric perspective distortion) to original training images to generate paired images with different illuminations, geometry consistency computes dense pixel-to-pixel correspondences. The keypoint detector from SuperPoint, MagicPoint, is used to generate keypoints in training. Then, cross-entropy loss applied on detection scores to learn the keypoint detector and triplet (feature positive and negative pairs) loss on uniformly sampled features for contrastive descriptor learning. DarkPoint achieves approximetly 1.7x more inliers during navigation than SuperPoint in day-night experiments.

zhang-et-al:2022:3086822 Key.Net + HardNet, ASMK, local features Zhang uses Key.Net (combines handcrafted and learned filters to detect keypoints at different scale levels, helping reduce the number of learnable parameters) and HardNet (applies novel loss to L2Net generating a compact descriptor of 128 dimensions) to extract the keypoints and corresponding descriptors, given that it was demonstrated that Key.Net with HardNet can achieve impressive performance in both repeatability results and matching scores regardless of viewpoint and illumination changes. Selecting a candidate frame via Aggregated Selective Match Kernel (ASMK) to search semantically similar images in the reference database. Then, using Locality-driven Accurate motion field Learning (LAL) for loop closure verification concentrating on searching for compatible planar geometry between candidate pairs meaning that motion field between them should satisfy the smoothness prior (due to loops detected within few meters). LAL has a good performance on precision curves, due to its using extra neighboring information to construct accurate motion field. Problems in experiment with variations in orientation and velocity, due to have wide baselines and dynamic objects.

xin-et-al:2017:8310121 AlexNet conv3, global + local features, global for candidates ¿ local for accurate matching Xin also uses features extracted with AlexNet because it was shown to have invariance against challenging environments, but from the third convolutional layer (conv3), due to fully connected layers (fc) not being as effective as convolutional ones because of the loss of spatial information. Both global (image-wise) and local (region-wise) are extracted with conv3 from AlexNet to improve the robustness to viewpoint changes compared to using only global features. Edge Boxes applied to extract reliable features to describe the scene sorting all candidate boxes according to objectness score, similar to Taishko. Image-wise features find small set of potential places, and then stable region-wise features considering both spatial and descriptor distance of the features for a more accurate result. Compared to HOG, BoW, and to conv3 global features, the proposed method outperforms them in precision-recall, but HOG BOW and conv3 glboal+local features proposed methodology had a worsen performance when testing with all day sequences (due to region detail dependent features being affected by day to night appearance changes) compared to global conv3 AlexNet. Studies the possibility of using a random selection technique for feature dimension reduction, given that requires no need for further training nor significant loss in efficiency and effectiveness compared to Local Sensitive Hushing (LSH) and Principal Component Analysis (PCA). Cosine distance for evaluate query image compared to database (global descriptor).

camara-et-al:2020:9196967 global + local features, global for initial set candidates ¿ local accurate matching, VGG16 Similar to Xin, Camera uses global features of the conv5-2 layer from VGG16 to compute global features to get an initial set of candidates based on nearest neighbor distance, and then compares geometrically the N candidates to the query image, based on the activations from the conv4-2 layer and their spatial location of the features created from them. Two databases: image filtering database to save global descriptors of reference images for the first step, and a spatial matching database to save the spatial arrangement of their containing vectors. Spatial matching by identifiying matching pairs by closest distance and keeping track of their position. PCA to reduce each vector of the 16 cubes to 100 dimensions.

taisho-kanji:2016:7866383 AlexNet fc6, instead of BoW uses a cross-domain library of images, global descriptor (60 nearest neighbor), reformulate visual place recog as classification problem BoW methods describe a scene image as a collection of visual words using a pre-trained library of feature descriptors. However, domain specific-learning of the library has poor performance in cross-domain scenarios. Thus, Taisho uses a cross-domain library consisting of images collected in different routes and seasons than the database images. From each library image, extract a set of 100 bounding boxes with the highest objectness score and select the 20 ones with the largest area as scene parts. Then extract 4096-dimensional DCNN features from the boxes' regions using the AlexNet (fc6, fully connect layer) network. Place recognition also extracts the 20 scene parts as for library images, defining the images as a set of 60 nearest neighbor library features (find 3 for each bounding box), and then employing the image-to-class distance with the Naive Bayes Nearest Neighbor (NBNN). The method outperformed in localization rate the works BoW and FAB-MAP. Principal Component Analysis (PCA) can be used for dimensionality reduction (4096 to 128-dimensional features), defining the proposed descriptor PCA-NBNN.

chen-et-al:2018:2859916 VGG16, place recognition as classification, local features used for the classification problem Chen uses a pre-trained VGG16 network to generate local features in a spatial location and then applies an attention mask for weighting the importance of each spatial location in the image. Local descriptor and global latent context fused to estimate the attention score at each spatial location to use global-level context information to guide attention spent in each region. The attention mask is parameterized by a CNN which takes the feature map as input, learned from different layers of the network to consider features from early layers (e.g., edges or corners) and deeper ones (semantic structures). They formulate place recognition as a classification problem, i.e., classifying each image to its correct place instead of comparing pairs or triplet of images. Proposed methodology outperforms FABMAP, SeqSLAM, and without attention mask in precision-recall metrics on seasonal changing conditions.

liu-et-al:2021:9561126 global + local features, global coarse estimation ¿ local accurate, MobileNetV2 Liu selects the MobileNetV2 network for global feature extraction due to its computational efficiency, in using triplet images for training, while also using Grid-based Motion Statistics (GMS) with ORB features for local geometrical verification. Compared to FABMAP, iBoW, and SeqSLAM, performed similar or even better maximum recall rate at 100% precision and lower execution times.

yin-et-al:2020:2905046 3D laser, LocNet, semi-handcrafted features (instead of "fully" CNN features) Yin uses LocNet as the feature extraction from a 3D laser scan, with LocNet following a semi-handcrafter feature learning architecture and embedding the feature in the Euclidean space, while reducing complexity of the network and improving the efficinecy on similarity evaluation. Indeed, a designed handcrafted rotational invariant feature is the input of the network. Siamese architecture and contrastive loss employed in training LocNet. LocNet improves AUC and f-score compared to M2DP.

piasco-et-al:2021:6 CNN-based features, triplet loss, side information, global features Piasco also employs triplet loss training within a CNN feature extractor framework. The CNN encoder aggregates local features to produce a global image-wise descriptor, while a decoder uses the deep representation of the image to reconstruct the scene geometry from the features obtained by the encoder. Then, training uses the fusion of image and depth map descriptors in the triplet loss function. Depth information is only used as side information during training given that could not be available at test time providing interesting building edges understanding. Compared to a domain adaptation method in an all-day experiment, the proposed method does not need to know in advance the source and target domains given that depth maps are invariant to the image domain. Reduce dimension of the descriptors by applying PCA and whitening.

yu-et-al:2019:8961714 DenseNet, global features using WVLAD Yu chooses DenseNet for feature extraction due to DenseNet reusing feature maps - earlier / lower layers contain more structural info and measure fine-grained similarity (similar to handcrafted features), while higher / deeper layers care more about semantic information and measure semantic similarity. Using the output of the last dense block with 7x7 feature maps and 1024 channels, decoumpling can be made by feature maps - 49 local descriptors with 1024 dimensions - or by channel - 1024 local descriptors with 49 dimensions. Feature maps decoupling chosen due to having better performance than by channel decoumpling. Then, uses Weighted Vector of Locally Aggregated Descriptor (WVLAD) for obtaining a global descriptor of the image, i.e., encode the 49x246 (256 after 4 max-pooling by channel) local descriptors into a single one. Compared with other networks such as ResNet50, VGG, etc., improves precision-recall. Plus, compared to ORG and SIFT encoded by BoW and VLAD, WVLAD obtains better precision-recall. The 4 max-pooling by channel proposed to reduce the descriptors dimensions with minimal accuracy reduction. 1024-dimension divided into 256 groups and maximum of ech group used as final descriptor. Compared to PCA, 4 max-pooling by channel has less computational complexity but similar performance.

martini-et-al:2020:s20216002 NetVLAD + VGG16, radar, global feature by NetVLAD encoding Martini uses NetVLAD with VGGG16 as a front-end feature extractor from polar radar images (range azimuth images). Modifications on VGG architecture to add circular padding along the azimuth direction to improve the rotation invariance of the feature descriptors to the input radar scans. Plus, application of a Gaussian blur before downsampling to reduce aliasing effect of the convolutional operations. Then, use a triplet loss training function (anchor radar scan, positive example from same location, and negative example from dissimilar location) to enforce metric space. Compared to Scan Context adapted to radar, proposed approach had improved precision-recall metrics.

yin-et-al:2021:3061375 VLAD, 3D point cloud, global feature, top-down + spherical views, VGG16 Yin proposes the FusionVLAD as a global descriptor for place recognition from a 3D point cloud with a parallel fusion network structure to learn point cloud representations from multi-view projections: project the 3D point cloud from an octree map into a top-down view and a spherical view. Using two separated 2D CNN following the convolutional layers in VGG16 to encode local features, the VLAD layer extracts 512-dimension place features from each view (top-down and spherical) separately. Then tightly-coupled fusion network between each feature of the 2 views, with a novel supervision objective lazy viewpoint-free loss metriuc to learn viewpoint-invariant and distinguishable descriptors, by combining a triplet loss for orientation-invariance with a transitional triplet loss to reduce feature difference, the later similar to NetVLAD.

yin-et-al:2018:8593562 3D laser, adversarial learning, SeqSLAM, bird-view view, global feature Yin proposes the Place Feature Learning (PFL) method for 3D laser-based features. Using Dynamic Octree Mapping (DOM) that is updated based on raw points and motion error model of the robot to model the local mapping around the robot, the method encodes the bird-view images of the DOM into a low dimensional feature vector using adversarial feature learning, which is a variant of the Generative Adversarial Network (GAN). The encoder maps data space to latent code space and the decoder generates synthetic data from latent code. This encoding improves the generalization of feature inference and unique mapping from raw data to latent code space. Using the extracted features for place recognition within the SeqSLAM framework, the proposed features improved the precision-recall metrics over the original SeqSLAM in changing conditions. Plus, compared to a pure Adversarial Feature Learning (AFL), the latter sometimes loses original geometry while reducing mapping uniquess, while the proposed method was able to maintain the global geometry structure.

kim-et-al:2019:2897340 3D laser, SCI (scan context-based), LeNet for feature + place classification, augmentation, new place detection Kim formulates a point cloud descriptor named Scan Context Image (SCI) and introduces a classification-based place recognition using the SCI. First, the 3D point cloud is converted to a Scan Context (SC) matrix containing the maximum heigh of points around a scene, with the rows and columns defined by the radial and azimuthal directions of the point cloud planar region (cylindrical representation). Using the jet colormap to transform the SC into a three-channel image suitable for the CNN inputs, use a LeNet network for feature extraction and place classification, being the output label the corresponding place index. The proposed method also uses augmentation to tackle viewpoint variance and is able to detect un-learn places using the entropy of the output vector of the network - if entropy higher than a certain threshold, place consider new one. SCI outperforms PointNetVLAD and the handcrafter M2DP in precision-recall.

xu-et-al:2021:3060741 u-net, scan context based, 3D laser, DiSCO, place recognition for global localization, FFT Xu proposes the Differentiable Scan Context with Orientation (DiSCO) downsamples a 3D point cloud from a 3D laser to its bird-eye view representation. Being compatible with normal Scan Context (single-layer height bird-eye view), multi-layer density BEV counting number of points in each voxel, and multi-layer BEV occupied BEV, use polar transform then CNN to learn features in polar domain. Finally, apply Fourier Transformation FFT to convert the polar BEV image representation to the frequency domain, given that frequency spectrum is translation-invariant (note rotation is transformed to translation by polar transform), thus, descriptor becoming rotation invariant. Quadruplet loss in training as PointNetVLAD to force the network to learn a close distance between vs taken at similar places and far apart distance between different ones. Superior performance than ScanContext (Kim), PointNetVLAD.

yin-et-al:2021:661199 LIDAR + radar, U-Net, ScanContext, triplet loss, FFT Yin proposes a shared network to extract features of lidar and radar data. Collecting lidar and radar scans at the same pose, use ScanContext of Kim to extract their respective representations (radar as polar representation, and lidar as occupied representation). Using a shared U-Net architecture, extract feature embeddings from the lidar and radar. Then, apply FFT to the polar bird's eye view representation given that theoretically the rotation of the vehicle results in a translation in polar representation and, given that the magnitude of frequency is translation-invariant, the final signatures would be rotation-invariant to the vehicle heading, similar to DiSCO assumption. Training uses the triplet loss and mixes and combinations in it (Radar2Radar, Lidar2Radar, and Lidar2Lidar), even though, only training the network once. Compared to DiSCO and ScanContext from Kim, joint learning in the proposed method improved or had similar in the 3 recognition tasks.

**Semantic segmentation** naseer-et-al:2017:7989305 qin-et-al:2020:9340939 berrio-et-al:2021:3094485 singh-et-al:2021:9564866 wang-et-al:2021:9739599

**Predict environment changes** neubert-et-al:2015:005 hu-et-al:2022:1003907

**Appearance-content disentanglement** qin-et-al:2020:103561 oh-eoh:2021:app11198976 tang-et-al:2021:17298814211037497 hu-et-al:2022:1003907

#### 5.1.5 Feature relevance

konolige-bowman:2009:5354121 murphy-sibley:2014:6907022 dymczyk-et-al:2016:66 nobre-et-al:2018:8461111 egger-et-al:2018:8593854 luthardt-et-al:2018:8569323 bürki-et-al:2019:21870 derner-et-al:2021:103676 berrio-et-al:2021:3094485

#### 5.1.6 Multi-modal features

filliat:2007:364080 neubert-et-al:2015:005 han-et-al:2017:2662061 latif-et-al:2017:016 han-et-al:2018:3 zhang-et-al:2018:8460674 siva-zhang:2018:8461042 siva-et-al:2020:9340992

#### 5.1.7 Geometric consistency

xin-et-al:2017:8310121 stable region-wise features considering both spatial and descriptor distance of the features for a more accurate result.

camara-et-al:2020:9196967 spatial matching by identifiying matching pairs by closest distance and keeping track of their position.

**Map geometry** li-et-al:2015:7139706 xin-et-al:2017:8310121 cao-et-al:2018:2815956 mactavish-et-al:2018:21838 martini-et-al:2020:s20216002 camara-et-al:2020:9196967 liu-et-al:2021:9561126 singh-et-al:2021:9564866 hong-et-al:2022:02783649221080483 zhang-et-al:2022:3086822

**Graph embedding** li-et-al:2015:7139706 han-et-al:2018:2856274 gao-zhang:2020:9196906 hong-et-al:2022:02783649221080483

#### 5.1.8 Temporal consistency

nguyen-et-al:2013:004 murphy-sibley:2014:6907022 naseer-et-al:2015:7324181 vysotska-et-al:2015:7139576 griffith-pradalier:2017:21664 arroyo-et-al:2018:7 han-et-al:2018:3 ouerghi-et-al:2018:s18040939 zhu-et-al:2018:8500686 cao-et-al:2021:2962416 yin-et-al:2021:3061375

#### 5.1.9 Sensor modalities

laser bosse-zlot:2009:009 cao-et-al:2018:2815956 egger-et-al:2018:8593854 yin-et-al:2018:8593562 berrio-et-al:2019:8814289 kim-et-al:2019:2897340 yin-et-al:2020:2905046 cao-et-al:2021:2962416 meng-et-al:2021:3062647 schaefer-et-al:2021:103709 siva-et-al:2020:9340992 wang-et-al:2021:9739599 xu-et-al:2021:3060741 yin-et-al:2021:3061375

radar martini-et-al:2020:s20216002 hong-et-al:2022:02783649221080483

magnetometer coulin-et-al:2022:3136241

laser + radar yin-et-al:2021:661199

camera + laser biswas-veloso:2013:0278364913503892 pérez-et-al:2015:y ding-et-al:2020:2942760 berrio-et-al:2021:3094485

camera + laser + UWB nguyen-et-al:2022:3094157

### 5.2 Dynamic environments

### 5.3 Map sparsification

### 5.4 Multi-session

### 5.5 Computational

taisho-kanji:2016:7866383 Principal Component Analysis (PCA) can be used for dimensionality reduction (4096 to 128-dimensional features), defining the proposed descriptor PCA-NBNN.

xin-et-al:2017:8310121 Studies the possibility of using a random selection technique for feature dimension reduction, given that requires no need for further training nor significant loss in efficiency and effectiveness compared to Local Sensitive Hushing (LSH) and Principal Component Analysis (PCA). Cosine distance for evaluate query image compared to database (global descriptor).

yu-et-al:2019:8961714 The 4 max-pooling by channel proposed to reduce the descriptors dimensions with minimal accuracy reduction. 1024-dimension divided into 256 groups and maximum of ech group used as final descriptor. Compared to PCA, 4 max-pooling by channel has less computational complexity but similar performance.

camara-et-al:2020:9196967 PCA to reduce each vector of the 16 cubes to 100 dimensions.

piasco-et-al:2021:6 Reduce dimension of the descriptors by applying PCA and whitening.

yang-et-al:2021:12054 select max-pooling layer instead of convulutional one

### 5.6 Evaluation metrics

### 5.7 Long-term experimental data

### 5.8 Final observations

## 6 Challenges and Future Directions

## 7 Limitations of the Study

Section to discuss possible limitations of the study (timeframe, wider approach, etc.).

- only one query for discussion: e.g., for searching datasets, possibly, a different query should have been used
- overview long-term SLAM vs in-depth analysis and discussion of each type of techniques: our review synthesizes all types of techniques, if the reader wants an in-depth analysis, different reviews should be performed
- related to previous one, each category appearance dynamics sparsity multi-session and computational have all different aspects related to each other, that probably should be treated differently in different data extraction items to improve the data organization – however the main goal of this review was to overview all trends and not focus on one specifically, so this categorization should be done separately and probably in singular literature reviews.
- limited information on the experiments conditions (traveled distance, duration, etc.) given by the authors
- some works such as PointNetVLAD, FAB-MAP and SeqSLAM (+LOAM, and LEGO-LOAM) did not appeared in the identification phase. However, the ones included have improvements over all of these...

## 8 Conclusions

## Acknowledgments

This study was supported by FCT – Fundação para a Ciência e a Tecnologia and INESC TEC – Institute for Systems and Computer Engineering, Technology and Science.

## ORCID

*Ricardo B. Sousa* https://orcid.org/0000-0003-4537-5095
*Héber M. Sobreira* https://orcid.org/0000-0002-8055-1093
*António Paulo Moreira* https://orcid.org/0000-0001-8573-3147

## References

*Acm digital library.* (n.d.). Retrieved May 6, 2022, from https://dl.acm.org/

Ali, A. J. B., Hashemifar, Z. S., & Dantu, K. Edge-SLAM: Edge-assisted visual simultaneous localization and mapping. In: 2020, 325–337. https://doi.org/10.1145/3386901.3389033.

Ali, W., Liu, P., Ying, R., & Gong, Z. (2021). A life-long SLAM approach using adaptable local maps based on rasterized LIDAR images. *IEEE Sensors Journal*, *21*(19), 21740–21749. https://doi.org/10.1109/JSEN.2021.3100882

An, S.-Y., Lee, L.-K., & Oh, S.-Y. (2016). Ceiling vision-based active SLAM framework for dynamic and wide-open environments. *Autonomous Robots*, *40*(2), 291–324. https://doi.org/10.1007/s10514-015-9453-0

Arroyo, R., Alcantarilla, P. F., Bergasa, L. M., & Romera, E. (2018). Are you ABLE to perform a life-long visual topological localization? *Autonomous Robots*, *42*(3), 665–85. https://doi.org/10.1007/s10514-017-9664-7

Bacca, B., Salví, J., & Cufí, X. (2013). Long-term mapping and localization using feature stability histograms. *Robotics and Autonomous Systems*, *61*(12), 1539–58. https://doi.org/10.1016/j.robot.2013.07.003

Bailey, T., & Durrant-Whyte, H. (2006). Simultaneous localization and mapping (SLAM): Part II. *IEEE Robotics Automation Magazine*, *13*(3), 108–117. https://doi.org/10.1109/MRA.2006.1678144

Ball, D., Heath, S., Wiles, J., Wyeth, G. F., Corke, P., & Milford, M. J. (2013). OpenRatSLAM: An open source brain-based SLAM system. *Autonomous Robots*, *34*(3), 149–176. https://doi.org/10.1007/s10514-012-9317-9

Berrio, J. S., Ward, J., Worrall, S., & Nebot, E. Identifying robust landmarks in feature-based maps. In: *2019-June*. 2019, 1166–1172. https://doi.org/10.1109/IVS.2019.8814289.

Berrio, J. S., Worrall, S., Shan, M., & Nebot, E. (2021). Long-term map maintenance pipeline for autonomous vehicles. *IEEE Transactions on Intelligent Transportation Systems*. https://doi.org/10.1109/TITS.2021.3094485

Bescos, B., Fácil, J. M., Civera, J., & Neira, J. (2018). DynaSLAM: Tracking, mapping, and inpainting in dynamic scenes. *IEEE Robotics and Automation Letters*, *3*(4), 4076–4083. https://doi.org/10.1109/LRA.2018.2860039

Biber, P., & Duckett, T. (2009). Experimental analysis of sample-based maps for long-term SLAM. *International Journal of Robotics Research*, *28*(1), 20–33. https://doi.org/10.1177/0278364908096286

Biswas, J., & Veloso, M. M. (2013). Localization and navigation of the CoBots over long-term deployments. *International Journal of Robotics Research*, *32*(14), 1679–1694. https://doi.org/10.1177/0278364913503892

Biswas, J., & Veloso, M. M. (2017). Episodic non-Markov localization. *Robotics and Autonomous Systems*, *87*, 162–176. https://doi.org/10.1016/j.robot.2016.09.005

Boniardi, F., Caselitz, T., Kümmerle, R., & Burgard, W. (2019). A pose graph-based localization system for long-term navigation in CAD floor plans. *Robotics and Autonomous Systems*, *112*, 84–97. https://doi.org/10.1016/j.robot.2018.11.003

Bosse, M., & Zlot, R. (2009). Keypoint design and evaluation for place recognition in 2D lidar maps. *Robotics and Autonomous Systems*, *57*(12), 1211–1224. https://doi.org/10.1016/j.robot.2009.07.009

Bouaziz, Y., Royer, E., Bresson, G., & Dhome, M. (2022). Map management for robust long-term visual localization of an autonomous shuttle in changing conditions. *Multimedia Tools and Applications*. https://doi.org/10.1007/s11042-021-11870-4

Bresson, G., Alsayed, Z., Yu, L., & Glaser, S. (2017). Simultaneous localization and mapping: A survey of current trends in autonomous driving. *IEEE Transactions on Intelligent Vehicles*, *2*(3), 194–220. https://doi.org/10.1109/TIV.2017.2749181

Bürki, M., Cadena, C., Gilitschenski, I., Siegwart, R., & Nieto, J. (2019). Appearance-based landmark selection for visual localization. *Journal of Field Robotics*, *36*(6), 1041–1073. https://doi.org/10.1002/rob.21870

Cadena, C., Carlone, L., Carrillo, H., Latif, Y., Scaramuzza, D., Neira, J., Reid, I., & Leonard, J. J. (2016). Past, present, and future of simultaneous localization and mapping: Toward the robust-perception age. *IEEE Transactions on Robotics*, *32*(6), 1309–1332. https://doi.org/10.1109/TRO.2016.2624754

Camara, L. G., Gäbert, C., & Přeučil, L. Highly robust visual place recognition through spatial matching of CNN features. In: 2020, 3748–3755. https://doi.org/10.1109/ICRA40945.2020.9196967.

Cao, F., Yan, F., Wang, S., Zhuang, Y., & Wang, W. (2021). Season-invariant and viewpoint-tolerant LiDAR place recognition in GPS-denied environments. *IEEE Transactions on Industrial Electronics*, *68*(1), 563–74. https://doi.org/10.1109/TIE.2019.2962416

Cao, F., Zhuang, Y., Zhang, H., & Wang, W. (2018). Robust place recognition and loop closing in laser-based SLAM for UGVs in urban environments. *IEEE Sensors Journal*, *18*(10), 4242–4252. https://doi.org/10.1109/JSEN.2018.2815956

Carlevaris-Bianco, N., Kaess, M., & Eustice, R. M. (2014). Generic node removal for factor-graph SLAM. *IEEE Transactions on Robotics*, *30*(6), 1371–1385. https://doi.org/10.1109/TRO.2014.2347571

Chebrolu, N., Läbe, T., & Stachniss, C. (2018). Robust long-term registration of UAV images of crop fields for precision agriculture. *IEEE Robotics and Automation Letters*, *3*(4), 3097–3104. https://doi.org/10.1109/LRA.2018.2849603

Chen, Z., Liu, L., Sa, I., Ge, Z., & Chli, M. (2018). Learning context flexible attention model for long-term visual place recognition. *IEEE Robotics and Automation Letters*, *3*(4), 4015–4022. https://doi.org/10.1109/LRA.2018.2859916

Churchill, W., & Newman, P. (2013). Experience-based navigation for long-term localisation. *International Journal of Robotics Research*, *32*(14), 1645–1661. https://doi.org/10.1177/0278364913499193

Clement, L., Gridseth, M., Tomasi, J., & Kelly, J. (2020). Learning matchable image transformations for long-term metric visual localization. *IEEE Robotics and Automation Letters*, *5*(2), 1492–1499. https://doi.org/10.1109/LRA.2020.2967659

Coulin, J., Guillemard, R., Gay-Bellile, V., Joly, C., & de La Fortelle, A. (2022). Tightly-coupled magneto-visual-inertial fusion for long term localization in indoor environment. *IEEE Robotics and Automation Letters*, *7*(2), 952–959. https://doi.org/10.1109/LRA.2021.3136241

Cummins, M., & Newman, P. (2008). FAB-MAP: Probabilistic localization and mapping in the space of appearance. *International Journal of Robotics Research*, *27*(6), 647–665. https://doi.org/10.1177/0278364908090961

Davison, A. J., & Murray, D. W. (2002). Simultaneous localization and map-building using active vision. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, *24*(7), 865–880. https://doi.org/10.1109/TPAMI.2002.1017615

Dayoub, F., Cielniak, G., & Duckett, T. (2011). Long-term experiments with an adaptive spherical view representation for navigation in changing environments. *Robotics and Autonomous Systems*, *59*(5), 285–295. https://doi.org/10.1016/j.robot.2011.02.013

Derner, E., Gomez, C., Hernandez, A. C., Barber, R., & Babuška, R. (2021). Change detection using weighted features for image-based localization. *Robotics and Autonomous Systems*, *135*, 103676. https://doi.org/10.1016/j.robot.2020.103676

*Dimensions*. (n.d.). Retrieved May 6, 2022, from https://app.dimensions.ai/discover/publication

Ding, X., Wang, Y., Tang, L., Yin, H., & Xiong, R. Communication constrained cloud-based long-term visual localization in real time. In: 2019, 2159–2166. https://doi.org/10.1109/IROS40897.2019.8968550.

Ding, X., Wang, Y., Xiong, R., Li, D., Tang, L., Yin, H., & Zhao, L. (2020). Persistent stereo visual localization on cross-modal invariant map. *IEEE Transactions on Intelligent Transportation Systems*, *21*(11), 4646–4658. https://doi.org/10.1109/TITS.2019.2942760

Du, Z.-J., Huang, S.-S., Mu, T.-J., Zhao, Q., Martin, R. R., & Xu, K. (2022). Accurate dynamic SLAM using CRF-based long-term consistency. *IEEE Transactions on Visualization and Computer Graphics*, *28*(4), 1745–57. https://doi.org/10.1109/TVCG.2020.3028218

Durrant-Whyte, H., & Bailey, T. (2006). Simultaneous localization and mapping (SLAM): Part I. *IEEE Robotics Automation Magazine*, *13*(2), 99–110. https://doi.org/10.1109/MRA.2006.1638022

Dymczyk, M., Lynen, S., Cieslewski, T., Bosse, M., Siegwart, R., & Furgale, P. The gist of maps - summarizing experience for lifelong localization. In: *2015-June*. (June). 2015, 2767–2773. https://doi.org/10.1109/ICRA.2015.7139575.

Dymczyk, M., Schneider, T., Gilitschenski, I., Siegwart, R., & Stumm, E. Erasing bad memories: Agent-side summarization for long-term mapping. In: *2016-November*. 2016, 4572–4579. https://doi.org/10.1109/IROS.2016.7759673.

Dymczyk, M., Stumm, E., Nieto, J., Siegwart, R., & Gilitschenski, I. Will it last? Learning stable features for long-term visual localization. In: 2016, 572–581. https://doi.org/10.1109/3DV.2016.66.

Egger, P., Borges, P. V. K., Catt, G., Pfrunder, A., Siegwart, R., & Dubé, R. PoseMap: Lifelong, multi-environment 3D LiDAR localization. In: 2018, 3430–3437. https://doi.org/10.1109/IROS.2018.8593854.

Einhorn, E., & Gross, H.-M. Generic 2D/3D SLAM with NDT maps for lifelong application. In: 2013, 240–247. https://doi.org/10.1109/ECMR.2013.6698849.

Einhorn, E., & Gross, H.-M. (2015). Generic NDT mapping in dynamic environments and its application for lifelong SLAM. *Robotics and Autonomous Systems*, *69*(1), 28–39. https://doi.org/10.1016/j.robot.2014.08.008

Filliat, D. A visual bag of words method for interactive qualitative localization and mapping. In: 2007, 3921–3926. https://doi.org/10.1109/ROBOT.2007.364080.

Fraundorfer, F., & Scaramuzza, D. (2012). Visual odometry. part II: Matching, robustness, optimization, and applications. *IEEE Robotics Automation Magazine*, *19*(2), 78–90. https://doi.org/10.1109/MRA.2012.2182810

Freitas, V. (2014). *Parsifal*. Retrieved May 12, 2022, from https://parsif.al/

Gadd, M., & Newman, P. Checkout my map: Version control for fleetwide visual localisation. In: *2016-November*. 2016, 5729–5736. https://doi.org/10.1109/IROS.2016.7759843.

Ganti, P., & Waslander, S. L. Network uncertainty informed semantic feature selection for visual SLAM. In: 2019, 121–128. https://doi.org/10.1109/CRV.2019.00024.

Gao, P., & Zhang, H. [Hao]. Long-term place recognition through worst-case graph matching to integrate landmark appearances and spatial relationships. In: 2020, 1070–1076. https://doi.org/10.1109/ICRA40945.2020.9196906.

Glover, A. J., Maddern, W. P., Milford, M. J., & Wyeth, G. F. FAB-MAP + RatSLAM: Appearance-based SLAM for multiple times of day. In: 2010, 3507–3512. https://doi.org/10.1109/ROBOT.2010.5509547.

Griffith, S., & Pradalier, C. (2017). Survey registration for long-term natural environment monitoring. *Journal of Field Robotics*, *34*(1), 188–208. https://doi.org/10.1002/rob.21664

Grisetti, G., Kümmerle, R., Stachniss, C., & Burgard, W. (2010). A tutorial on graph-based SLAM. *IEEE Intelligent Transportation Systems Magazine*, *2*(4), 31–43. https://doi.org/10.1109/MITS.2010.939925

Han, F., Beleidy, S. E., Wang, H., Ye, C., & Zhang, H. (2018). Learning of holism-landmark graph embedding for place recognition in long-term autonomy. *IEEE Robotics and Automation Letters*, *3*(4), 3669–3676. https://doi.org/10.1109/LRA.2018.2856274

Han, F., Wang, H., Huang, G., & Zhang, H. (2018). Sequence-based sparse optimization methods for long-term loop closure detection in visual SLAM. *Autonomous Robots*, *42*(7), 1323–1335. https://doi.org/10.1007/s10514-018-9736-3

Han, F., Yang, X., Deng, Y., Rentschler, M., Yang, D., & Zhang, H. (2017). SRAL: Shared representative appearance learning for long-term visual place recognition. *IEEE Robotics and Automation Letters*, *2*(2), 1172–1179. https://doi.org/10.1109/LRA.2017.2662061

Hochdorfer, S., Lutz, M., & Schlegel, C. Lifelong localization of a mobile service-robot in everyday indoor environments using omnidirectional vision. In: 2009, 161–166. https://doi.org/10.1109/TEPRA.2009.5339626.

Hochdorfer, S., & Schlegel, C. Towards a robust visual SLAM approach: Addressing the challenge of life-long operation. In: 2009, 1–6. https://ieeexplore.ieee.org/document/5174794

Hong, Z., Petillot, Y., Wallace, A., & Wang, S. (2022). RadarSLAM: A robust simultaneous localization and mapping system for all weather conditions. *International Journal of Robotics Research*. https://doi.org/10.1177/02783649221080483

Hu, H., Wang, H., Liu, Z., & Chen, W. (2022). Domain-invariant similarity activation map contrastive learning for retrieval-based long-term visual localization. *IEEE/CAA Journal of Automatica Sinica*, *9*(2), 313–328. https://doi.org/10.1109/JAS.2021.1003907

Huang, G., Kaess, M., & Leonard, J. J. Consistent sparsification for graph optimization. In: 2013, 150–157. https://doi.org/10.1109/ECMR.2013.6698835.

*Ieee xplore*. (n.d.). Retrieved May 6, 2022, from https://ieeexplore.ieee.org/Xplore/home.jsp

Ikeda, K., & Kanji, T. Visual robot localization using compact binary landmarks. In: 2010, 4397–4403. https://doi.org/10.1109/ROBOT.2010.5509579.

Ila, V., Polok, L., Solony, M., & Svoboda, P. (2017). SLAM++ - a highly efficient and temporally scalable incremental SLAM framework. *International Journal of Robotics Research*, *36*(2), 210–230. https://doi.org/10.1177/0278364917691110

*Inspec*. (n.d.). Retrieved May 6, 2022, from https://www.engineeringvillage.com/search/quick.url

Johannsson, H., Kaess, M., Fallon, M., & Leonard, J. J. Temporally scalable visual SLAM using a reduced pose graph. In: 2013, 54–61. https://doi.org/10.1109/ICRA.2013.6630556.

Karaoğuz, H., & Bozma, H. I. (2016). An integrated model of autonomous topological spatial cognition. *Autonomous Robots*, *40*(8), 1379–1402. https://doi.org/10.1007/s10514-015-9514-4

Karaoğuz, H., & Bozma, H. I. (2020). Merging of appearance-based place knowledge among multiple robots. *Autonomous Robots*, *44*(6), 1009–1027. https://doi.org/10.1007/s10514-020-09911-2

Kawewong, A., Tongprasit, N., & Hasegawa, O. (2013). A speeded-up online incremental vision-based loop-closure detection for long-term SLAM. *Advanced Robotics*, *27*(17), 1325–1336. https://doi.org/10.1080/01691864.2013.826410

Kim, C., Cho, S., Sunwoo, M., Resende, P., Bradaï, B., & Jo, K. (2021). A geodetic normal distribution map for long-term LiDAR localization on earth. *IEEE Access*, *9*, 470–484. https://doi.org/10.1109/ACCESS.2020.3047421

Kim, G., Park, B., & Kim, A. (2019). 1-day learning, 1-year localization: Long-term LiDAR localization using scan context image. *IEEE Robotics and Automation Letters*, *4*(2), 1948–1955. https://doi.org/10.1109/LRA.2019.2897340

Konolige, K., & Bowman, J. Towards lifelong visual maps. In: 2009, 1156–1163. https://doi.org/10.1109/IROS.2009.5354121.

Krajník, T., Fentanes, J. P., Santos, J. M., & Duckett, T. (2017). FreMEn: Frequency map enhancement for long-term mobile robot autonomy in changing environments. *IEEE Transactions on Robotics*, *33*(4), 964–977. https://doi.org/10.1109/TRO.2017.2665664

Kretzschmar, H., Grisetti, G., & Stachniss, C. (2010). Lifelong map learning for graph-based SLAM in static environments. *KI - Kunstliche Intelligenz*, *24*(3), 199–206. https://doi.org/10.1007/s13218-010-0034-2

Kretzschmar, H., & Stachniss, C. (2012). Information-theoretic compression of pose graphs for laser-based SLAM. *International Journal of Robotics Research*, *31*(11), 1219–1230. https://doi.org/10.1177/0278364912455072

Kretzschmar, H., Stachniss, C., & Grisetti, G. Efficient information-theoretic graph pruning for graph-based SLAM with laser range finders. In: 2011, 865–871. https://doi.org/10.1109/IROS.2011.6094414.

Kurz, G., Holoch, M., & Biber, P. Geometry-based graph pruning for lifelong SLAM. In: 2021, 3313–3320. https://doi.org/10.1109/IROS51168.2021.9636530.

Labbé, M., & Michaud, F. (2019). RTAB-Map as an open-source lidar and visual simultaneous localization and mapping library for large-scale and long-term online operation. *Journal of Field Robotics*, *36*(2), 416–446. https://doi.org/10.1002/rob.21831

Latif, Y., Cadena, C., & Neira, J. Realizing, reversing, recovering: Incremental robust loop closing over time using the iRRR algorithm. In: 2012, 4211–4217. https://doi.org/10.1109/IROS.2012.6385879.

Latif, Y., Huang, G., Leonard, J. J., & Neira, J. (2017). Sparse optimization for robust and efficient loop closing. *Robotics and Autonomous Systems*, *93*, 13–26. https://doi.org/10.1016/j.robot.2017.03.016

Lázaro, M. T., Capobianco, R., & Grisetti, G. Efficient long-term mapping in dynamic environments. In: 2018, 153–160. https://doi.org/10.1109/IROS.2018.8594310.

Li, J., Eustice, R. M., & Johnson-Roberson, M. (2015). High-level visual features for underwater place recognition. *2015 IEEE International Conference on Robotics and Automation (ICRA)*, 3652–3659. https://doi.org/10.1109/ICRA.2015.7139706

Liu, B., Tang, F., Fu, Y., Yang, Y., & Wu, Y. (2021). A flexible and efficient loop closure detection based on motion knowledge. *2021 IEEE International Conference on Robotics and Automation (ICRA)*, 11241–7. https://doi.org/10.1109/ICRA48506.2021.9561126

Luthardt, S., Willert, V., & Adamy, J. LLama-SLAM: Learning high-quality visual landmarks for long-term mapping and localization. In: *2018-November*. 2018, 2645–2652. https://doi.org/10.1109/ITSC.2018.8569323.

MacTavish, K., Paton, M., & Barfoot, T. D. (2018). Selective memory: Recalling relevant experience for long-term visual localization. *Journal of Field Robotics*, *35*(8), 1265–1292. https://doi.org/10.1002/rob.21838

Maddern, W., Milford, M. J., & Wyeth, G. F. (2012). Capping computation time and storage requirements for appearance-based localization with CAT-SLAM. *2012 IEEE International Conference on Robotics and Automation (ICRA)*, 822–827. https://doi.org/10.1109/ICRA.2012.6224622

Maddern, W., Milford, M. J., & Wyeth, G. F. Towards persistent localization and mapping with a continuous appearance-based topology. In: *8*. 2013, 281–288. https://doi.org/10.15607/rss.2012.viii.036.

Martini, D. D., Gadd, M., & Newman, P. (2020). KRadar++: Coarse-to-fine FMCW scanning radar localisation. *Sensors*, *20*(21), 1–23. https://doi.org/10.3390/s20216002

Mazuran, M., Burgard, W., & Tipaldi, G. D. (2016). Nonlinear factor recovery for long-term SLAM. *International Journal of Robotics Research*, *35*(1-3), 50–72. https://doi.org/10.1177/0278364915581629

Meng, Q., Guo, H., Zhao, X., Cao, D., & Chen, H. (2021). Loop-closure detection with a multiresolution point cloud histogram mode in Lidar odometry and mapping for intelligent vehicles. *IEEE/ASME Transactions on Mechatronics*, *26*(3), 1307–1317. https://doi.org/10.1109/TMECH.2021.3062647

Milford, M. J., & Wyeth, G. F. SeqSLAM: Visual route-based navigation for sunny summer days and stormy winter nights. In: 2012, 1643–1649. https://doi.org/10.1109/ICRA.2012.6224623.

Mohan, M., Gálvez-López, D., Monteleoni, C., & Sibley, G. Environment selection and hierarchical place recognition. In: *2015-June*. (June). 2015, 5487–5494. https://doi.org/10.1109/ICRA.2015.7139966.

Mongeon, P., & Paul-Hus, A. (2016). The journal coverage of Web of Science and Scopus: A comparative analysis. *Scientometrics*, *106*(1), 213–228. https://doi.org/10.1007/s11192-015-1765-5

Mühlfellner, P., Bürki, M., Bosse, M., Derendarz, W., Philippsen, R., & Furgale, P. (2016). Summary maps for lifelong visual localization. *Journal of Field Robotics*, *33*(5), 561–590. https://doi.org/10.1002/rob.21595

Mur-Artal, R., Montiel, J. M. M., & Tardós, J. D. (2015). ORB-SLAM: A versatile and accurate monocular SLAM system. *IEEE Transactions on Robotics*, *31*(5), 1147–1163. https://doi.org/10.1109/TRO.2015.2463671

Murphy, L., & Sibley, G. Incremental unsupervised topological place discovery. In: 2014, 1312–1318. https://doi.org/10.1109/ICRA.2014.6907022.

Nanni, L., Ghidoni, S., & Brahnam, S. (2017). Handcrafted vs non-handcrafted features for computer vision classification. *Pattern Recognition*, *71*, 158–172. https://doi.org/10.1016/j.patcog.2017.05.025

Naseer, T., Oliveira, G. L., Brox, T., & Burgard, W. (2017). Semantics-aware visual localization under challenging perceptual conditions. *2017 IEEE International Conference on Robotics and Automation (ICRA)*, 2614–20. https://doi.org/10.1109/ICRA.2017.7989305

Naseer, T., Suger, B., Ruhnke, M., & Burgard, W. Vision-based Markov localization across large perceptual changes. In: 2015, 1–6. https://doi.org/10.1109/ECMR.2015.7324181.

Neubert, P., Sünderhauf, N., & Protzel, P. (2015). Superpixel-based appearance change prediction for long-term navigation across seasons. *Robotics and Autonomous Systems*, *69*, 15–27. https://doi.org/10.1016/j.robot.2014.08.005

Nguyen, T.-M., Cao, M., Yuan, S., Lyu, Y., Nguyen, T. H., & Xie, L. (2022). VIRAL-Fusion: A visual-inertial-ranging-Lidar sensor fusion approach. *IEEE Transactions on Robotics*, *38*(2), 958–977. https://doi.org/10.1109/TRO.2021.3094157

Nguyen, V. A., Starzyk, J. A., & Goh, W.-B. (2013). A spatio-temporal long-term memory approach for visual place recognition in mobile robotic navigation. *Robotics and Autonomous Systems*, *61*(12), 1744–1758. https://doi.org/10.1016/j.robot.2012.12.004

Nobre, F., Heckman, C., Ozog, P., Wolcott, R. W., & Walls, J. M. Online probabilistic change detection in feature-based maps. In: 2018, 3661–3668. https://doi.org/10.1109/ICRA.2018.8461111.

Nuske, S., Roberts, J., & Wyeth, G. F. (2009). Robust outdoor visual localization using a three-dimensional-edge map. *Journal of Field Robotics*, *26*(9), 728–756. https://doi.org/10.1002/rob.20306

Oberländer, J., Roennau, A., & Dillmann, R. Hierarchical SLAM using spectral submap matching with opportunities for long-term operation. In: 2013, 1–7. https://doi.org/10.1109/ICAR.2013.6766479.

Oh, J., & Eoh, G. (2021). Variational Bayesian approach to condition-invariant feature extraction for visual place recognition. *Applied Sciences*, *11*(19), 8976. https://doi.org/10.3390/app11198976

Opdenbosch, D. V., Aykut, T., Oelsch, M., Alt, N., & Steinbach, E. Efficient map compression for collaborative visual SLAM. In: *2018-January*. 2018, 992–1000. https://doi.org/10.1109/WACV.2018.00114.

Ouerghi, S., Boutteau, R., Savatier, X., & Tlili, F. (2018). Visual odometry and place recognition fusion for vehicle position tracking in urban environments. *Sensors*, *18*(4), 939. https://doi.org/10.3390/s18040939

Ozog, P., Carlevaris-Bianco, N., Kim, A., & Eustice, R. M. (2016). Long-term mapping techniques for ship hull inspection and surveillance using an autonomous underwater vehicle. *Journal of Field Robotics*, *33*(3), 265–289. https://doi.org/10.1002/rob.21582

Page, M. J., Moher, D., Bossuyt, P. M., Boutron, I., Hoffmann, T. C., Mulrow, C. D., Shamseer, L., Tetzlaff, J. M., Akl, E. A., Brennan, S. E., Chou, R., Glanville, J., Grimshaw, J. M., Hróbjartsson, A., Lalu, M. M., Li, T., Loder, E. W., Mayo-Wilson, E., McDonald, S., ... McKenzie, J. E. (2021). PRISMA 2020 explanation and elaboration: Updated guidance and exemplars for reporting systematic reviews. *BMJ*, *372*. https://doi.org/10.1136/bmj.n160

Pan, Z., Chen, H., Li, S., & Liu, Y. (2019). Clustermap building and relocalization in urban environments for unmanned vehicles. *Sensors*, *19*(19), 4252. https://doi.org/10.3390/s19194252

Paul, R., & Newman, P. (2013). Self-help: Seeking out perplexing images for ever improving topological mapping. *International Journal of Robotics Research*, *32*(14), 1742–1766. https://doi.org/10.1177/0278364913509859

Pérez, J., Caballero, F., & Merino, L. (2015). Enhanced Monte Carlo localization with visual place recognition for robust robot localization. *Journal of Intelligent and Robotic Systems: Theory and Applications*, *80*(3-4), 641–656. https://doi.org/10.1007/s10846-015-0198-y

Piasco, N., Sidibé, D., Gouet-Brunet, V., & Demonceaux, C. (2021). Improving image description with auxiliary modality for visual localization in challenging conditions. *International Journal of Computer Vision*, *129*(1), 185–202. https://doi.org/10.1007/s11263-020-01363-6

Pirker, K., Rüther, M., & Bischof, H. CD SLAM - continuous localization and mapping in a dynamic world. In: 2011, 3990–3997. https://doi.org/10.1109/IROS.2011.6048253.

Pomerleau, F., Krüsi, P., Colas, F., Furgale, P., & Siegwart, R. Long-term 3D map maintenance in dynamic environments. In: 2014, 3712–3719. https://doi.org/10.1109/ICRA.2014.6907397.

Qin, C., Zhang, Y., Liu, Y., Coleman, S., Kerr, D., & Lv, G. (2020). Appearance-invariant place recognition by adversarially learning disentangled representation. *Robotics and Autonomous Systems*, *131*, 103561. https://doi.org/10.1016/j.robot.2020.103561

Qin, T., Chen, T., Chen, Y., & Su, Q. AVP-SLAM: Semantic visual mapping and localization for autonomous vehicles in the parking lot. In: 2020, 5939–5945. https://doi.org/10.1109/IROS45743.2020.9340939.

Rapp, M., Hahn, M., Thom, M., Dickmann, J., & Dietmayer, K. (2015). Semi-Markov process based localization using radar in dynamic environments. *2015 IEEE International Intelligent Transportation Systems Conference (ITSC)*, 423–429. https://doi.org/10.1109/ITSC.2015.77

Saarinen, J. P., Andreasson, H., Stoyanov, T., & Lilienthal, A. J. (2013). 3D normal distributions transform occupancy maps: An efficient representation for mapping in dynamic environments. *International Journal of Robotics Research*, *32*(14), 1627–44. https://doi.org/10.1177/0278364913499415

Saeedi, S., Trentini, M., Seto, M., & Li, H. (2016). Multiple-robot simultaneous localization and mapping: A review. *Journal of Field Robotics*, *33*(1), 3–46. https://doi.org/10.1002/rob.21620

Santos, J. M., Krajník, T., Fentanes, J. P., & Duckett, T. (2016). Lifelong information-driven exploration to complete and refine 4-D spatio-temporal maps. *IEEE Robotics and Automation Letters*, *1*(2), 684–691. https://doi.org/10.1109/LRA.2016.2516594

Scaramuzza, D., & Fraundorfer, F. (2011). Visual odometry. part I: The first 30 years and fundamentals. *IEEE Robotics Automation Magazine*, *18*(4), 80–92. https://doi.org/10.1109/MRA.2011.943233

Schaefer, A., Büscher, D., Vertens, J., Luft, L., & Burgard, W. (2021). Long-term vehicle localization in urban environments based on pole landmarks extracted from 3-D lidar scans. *Robotics and Autonomous Systems*, *136*, 103709. https://doi.org/10.1016/j.robot.2020.103709

Schmuck, P., & Chli, M. On the redundancy detection in keyframe-based SLAM. In: 2019, 594–603. https://doi.org/10.1109/3DV.2019.00071.

*Scopus.* (n.d.). Retrieved May 6, 2022, from https://www.scopus.com/search/form.uri

Singh, G., Wu, M., Lam, S.-K., & Minh, D. V. Hierarchical loop closure detection for long-term visual SLAM with semantic-geometric descriptors. In: *2021-September.* 2021, 2909–2916. https://doi.org/10.1109/ITSC48978.2021.9564866.

Singh, V. K., Singh, P., Karmakar, M., Leta, J., & Mayr, P. (2021). The journal coverage of Web of Science, Scopus and Dimensions: A comparative analysis. *Scientometrics*, *126*(6), 5113–5142. https://doi.org/10.1007/s11192-021-03948-5

Siva, S., Nahman, Z., & Zhang, H. Voxel-based representation learning for place recognition based on 3D point clouds. In: 2020, 8351–8357. https://doi.org/10.1109/IROS45743.2020.9340992.

Siva, S., & Zhang, H. [Hao]. Omnidirectional multisensory perception fusion for long-term place recognition. In: 2018, 5175–5181. https://doi.org/10.1109/ICRA.2018.8461042.

Sivic, J., & Zisserman, A. Video Google: A text retrieval approach to object matching in videos. In: 2003, 1470–1477. https://doi.org/10.1109/ICCV.2003.1238663.

Song, Y., Zhu, D., Li, J., Tian, Y., & Li, M. (2019). Learning local feature descriptor with motion attribute for vision-based localization. *2019 IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS)*, 3794–801. https://doi.org/10.1109/IROS40897.2019.8967749

Sun, L. [L.], Yan, Z., Zaganidis, A., Zhao, C., & Duckett, T. (2018). Recurrent-OctoMap: Learning state-based map refinement for long-term semantic mapping with 3-D-Lidar data. *IEEE Robotics and Automation Letters*, *3*(4), 3749–3756. https://doi.org/10.1109/LRA.2018.2856268

Sun, L. [Li], Taher, M., Wild, C., Zhao, C., Zhang, Y., Majer, F., Yan, Z., Krajník, T., Prescott, T., & Duckett, T. Robust and long-term monocular teach and repeat navigation using a single-experience map. In: 2021, 2635–2642. https://doi.org/10.1109/IROS51168.2021.9635886.

Taisho, T., & Kanji, T. Mining dcnn landmarks for long-term visual SLAM. In: 2016, 570–576. https://doi.org/10.1109/ROBIO.2016.7866383.

Tang, L., Wang, Y., Ding, X., Yin, H., Xiong, R., & Huang, S. (2019). Topological local-metric framework for mobile robots navigation: A long term perspective. *Autonomous Robots*, *43*(1), 197–211. https://doi.org/10.1007/s10514-018-9724-7

Tang, L., Wang, Y., Tan, Q., & Xiong, R. (2021). Explicit feature disentanglement for visual place recognition across appearance changes. *International Journal of Advanced Robotic Systems*, *18*(6). https://doi.org/10.1177/17298814211037497

Thomas, H., Agro, B., Gridseth, M., Zhang, J., & Barfoot, T. D. Self-supervised learning of Lidar segmentation for autonomous indoor navigation. In: *2021-May.* 2021, 14047–14053. https://doi.org/10.1109/ICRA48506.2021.9561701.

Tipaldi, G. D., Meyer-Delius, D., & Burgard, W. (2013). Lifelong localization in changing environments. *International Journal of Robotics Research*, *32*(14), 1662–1678. https://doi.org/10.1177/0278364913502830

Tsintotas, K. A., Bampis, L., & Gasteratos, A. (2021). Modest-vocabulary loop-closure detection with incremental bag of tracked words. *Robotics and Autonomous Systems*, *141*, 103782. https://doi.org/10.1016/j.robot.2021.103782

Uy, M. A., & Lee, G. H. PointNetVLAD: Deep point cloud based retrieval for large-scale place recognition. In: 2018, 4470–4479. https://doi.org/10.1109/CVPR.2018.00470.

van Eck, N. J., & Waltman, L. (2010). Software survey: VOSviewer, a computer program for bibliometric mapping. *Scientometrics*, *84*(2), 523–538. https://doi.org/10.1007/s11192-009-0146-3

van Eck, N. J., & Waltman, L. (2014). Visualizing bibliometric networks. In Y. Ding, R. Rousseau, & D. Wolfram (Eds.), *Measuring scholarly impact: Methods and practice* (pp. 285–320). Springer. https://doi.org/10.1007/978-3-319-10377-8_13

Vysotska, O., Naseer, T., Spinello, L., Burgard, W., & Stachniss, C. (2015). Efficient and effective matching of image sequences under substantial appearance changes exploiting GPS priors. *2015 IEEE International Conference on Robotics and Automation (ICRA)*, 2774–9. https://doi.org/10.1109/ICRA.2015.7139576

Walcott-Bryant, A., Kaess, M., Johannsson, H., & Leonard, J. J. Dynamic pose graph SLAM: Long-term mapping in low dynamic environments. In: 2012, 1871–1878. https://doi.org/10.1109/IROS.2012.6385561.

Wang, K., Lin, Y., Wang, L., Han, L., Hua, M., Wang, X., Lian, S., & Huang, B. A unified framework for mutual improvement of SLAM and semantic segmentation. In: *2019-May*. 2019, 5224–5230. https://doi.org/10.1109/ICRA.2019.8793499.

Wang, L., Chen, W., & Wang, J. Long-term localization with time series map prediction for mobile robots in dynamic environments. In: *2020-January*. 2020, 8587–8593. https://doi.org/10.1109/IROS45743.2020.9468884.

Wang, Z., Li, S., Cao, M., Chen, H., & Liu, Y. Pole-like objects mapping and long-term robot localization in dynamic urban scenarios. In: 2021, 998–1003. https://doi.org/10.1109/ROBIO54168.2021.9739599.

*Web of science*. (n.d.). Retrieved May 6, 2022, from https://www.webofscience.com/wos/woscc/basic-search

Williams, S., Indelman, V., Kaess, M., Roberts, R., Leonard, J. J., & Dellaert, F. (2014). Concurrent filtering and smoothing: A parallel architecture for real-time navigation and full smoothing. *International Journal of Robotics Research*, *33*(12), 1544–1568. https://doi.org/10.1177/0278364914531056

Wu, L., & Wu, Y. Deep supervised hashing with similar hierarchy for place recognition. In: 2019, 3781–3786. https://doi.org/10.1109/IROS40897.2019.8968599.

Xin, Z., Cui, X., Zhang, J., Yang, Y., & Wang, Y. (2017). Visual place recognition with CNNs: From global to partial. *2017 Seventh International Conference on Image Processing Theory, Tools and Applications (IPTA)*, 6pp.–. https://doi.org/10.1109/IPTA.2017.8310121

Xing, Z., Zhu, X., & Dong, D. (2022). DE-SLAM: SLAM for highly dynamic environment. *Journal of Field Robotics*. https://doi.org/10.1002/rob.22062

Xu, X., Yin, H., Chen, Z., Li, Y., Wang, Y., & Xiong, R. (2021). DiSCO: Differentiable scan context with orientation. *IEEE Robotics and Automation Letters*, *6*(2), 2791–2798. https://doi.org/10.1109/LRA.2021.3060741

Yang, S., Fan, G., Bai, L., Zhao, C., & Li, D. (2020). SGC-VSLAM: A semantic and geometric constraints VSLAM for dynamic indoor environments. *Sensors*, *20*(8), 2432. https://doi.org/10.3390/s20082432

Yang, Z., Pan, Y., Deng, L., Xie, Y., & Huan, R. (2021). PLSAV: Parallel loop searching and verifying for loop closure detection. *IET Intelligent Transport Systems*, *15*(5), 683–698. https://doi.org/10.1049/itr2.12054

Yin, H., Wang, Y., Ding, X., Tang, L., Huang, S., & Xiong, R. (2020). 3D LiDAR-based global localization using siamese neural network. *IEEE Transactions on Intelligent Transportation Systems*, *21*(4), 1380–1392. https://doi.org/10.1109/TITS.2019.2905046

Yin, H., Xu, X., Wang, Y., & Xiong, R. (2021). Radar-to-Lidar: Heterogeneous place recognition via joint learning. *Frontiers in Robotics and AI*, *8*, 661199. https://doi.org/10.3389/frobt.2021.661199

Yin, P., Xu, J., Zhang, J., & Choset, H. (2021). FusionVLAD: A multi-view deep fusion networks for viewpoint-free 3D place recognition. *IEEE Robotics and Automation Letters*, *6*(2), 2304–2310. https://doi.org/10.1109/LRA.2021.3061375

Yin, P., Xu, L., Liu, Z., Li, L., Salman, H., He, Y., Xu, W., Wang, H., & Choset, H. Stabilize an unsupervised feature learning for LiDAR-based place recognition. In: 2018, 1162–1167. https://doi.org/10.1109/IROS.2018.8593562.

Yu, C., Liu, Z., Liu, X.-J., Qiao, F., Wang, Y., Xie, F., Wei, Q., & Yang, Y. A DenseNet feature-based loop closure method for visual SLAM system. In: 2019, 258–265. https://doi.org/10.1109/ROBIO49542.2019.8961714.

Yue, Y., Yang, C., Zhang, J., Wen, M., Wu, Z., Zhang, H., & Wang, D. (2020). Day and night collaborative dynamic mapping in unstructured environment based on multimodal sensors. *2020 IEEE International Conference on Robotics and Automation (ICRA)*, 2981–7. https://doi.org/10.1109/ICRA40945.2020.9197072

Zeng, T., & Si, B. (2021). A brain-inspired compact cognitive mapping system. *Cognitive Neurodynamics*, *15*(1), 91–101. https://doi.org/10.1007/s11571-020-09621-6

Zhang, H. [Hui], Chen, X., Lu, H., & Xiao, J. (2018). Distributed and collaborative monocular simultaneous localization and mapping for multi-robot systems in large-scale environments. *International Journal of Advanced Robotic Systems*, *15*(3). https://doi.org/10.1177/1729881418780178

Zhang, K., Jiang, X., & Ma, J. (2022). Appearance-based loop closure detection via locality-driven accurate motion field learning. *IEEE Transactions on Intelligent Transportation Systems*, *23*(3), 2350–2365. https://doi.org/10.1109/TITS.2021.3086822

Zhang, M., Chen, Y., & Li, M. SDF-Loc: Signed distance field based 2D relocalization and map update in dynamic environments. In: *2019-July*. 2019, 1997–2004. https://doi.org/10.23919/acc.2019.8814347.

Zhang, N., Warren, M., & Barfoot, T. D. Learning place-and-time-dependent binary descriptors for long-term visual localization. In: 2018, 828–835. https://doi.org/10.1109/ICRA.2018.8460674.

Zhu, J., Ai, Y., Tian, B., Cao, D., & Scherer, S. (2018). Visual place recognition in long-term and large-scale environment based on CNN feature. *2018 IEEE Intelligent Vehicles Symposium (IV)*, 1679–85. https://doi.org/10.1109/IVS.2018.8500686

Zhu, S., Zhang, X., Guo, S., Li, J., & Liu, H. Lifelong localization in semi-dynamic environment. In: *2021-May*. 2021, 14389–14395. https://doi.org/10.1109/ICRA48506.2021.9561584.

**Ricardo B. Sousa** obtained a Master of Science (M.Sc.) degree in Electric and Computers Engineering (ECE) at Faculty of Engineering of the University of Porto (FEUP), in 2020. He is currently working towards the Ph.D. degree in electrical and computer engineering with FEUP, and he has a graduate research scholarship from FCT – Fundação para a Ciência e a Tecnologia at the Centre for Robotics in Industry and Intelligent Systems from INESC TEC. Also, he is an invited assistant lecturing the courses Software Design and Industrial Informatics from the M.Sc. in ECE at FEUP. His research interests include robotics, sensor fusion, and localization and mapping for autonomous robots.

**Héber M. Sobreira** was born in Leiria, Portugal, in July 1985. He graduated with an M.Sc. degree (2009) and a Ph.D. degree (2017) in Electrical Engineering from the University of Porto. Since 2009, he has been developing his research within the Centre for Robotics in Industry and Intelligent Systems at INESC TEC. His main research areas are navigation and control of indoor autonomous vehicles.

**António Paulo Moreira** graduated with a degree in electrical engineering at the University of Oporto, in 1986. Then, he pursued graduate studies at University of Porto, obtaining a M.Sc. degree in electrical engineering – systems in 1991 and a Ph.D. degree in electrical engineering in 1998. Presently, he is Associate Professor with tenure at the Faculty of Engineering of the University of Porto and researcher and head of the Centre for Robotics in Industry and Intelligent Systems at INESC TEC. His main research interests are process control and robotics.

# A Data Extraction Results of the Included Records in the Systematic Literature Review on Long-Term Localization and Mapping for Mobile Robots

Table 6: Data extraction items retrieved from the included records in the review

| DE: | 1: | | | | | 2: | 3: | 4: | 5: | | 6: | | | | | 7: | 8: | 9: | 10: | | | | 11: | 12: |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| Ref. | appearance | dynamic | sparsity | multi-session | computational | localization | mapping | multi-robot | offline | online | indoor | outdoor | air | ground | water | sensor | exp. self-acq. | ground-truth | dist. (km) | path (km) | time (h) | int. (d/w/m/y) | datasets | metrics |
| Davison and Murray 2002 | | | x | | | EKF (2D, 3DoF) | feature (Harris corner detector) | – | | x | x | | | x | | wheel odometry, camera (gray, stereo) | x | manual | – | – | – | – | – | innovation covariance, pose error |
| Filliat 2007 | x | | | | | image classification (location) | dictionary (BoW, location category) | – | x | x | x | | | x | | camera (color, mono) | x | manual | – | – | – | 1d | – | confusion matrix, localization rate |
| Konolige and Bowman 2009 | x | | x | | | visual odometry (3D, 6DoF), vocabulary tree (location) | keyframe (graph, 6DoF edges) | – | | x | x | | | x | | camera (stereo) | x | – | – | – | – | 4d | – | execution time, localization rate, memory |
| Bosse and Zlot 2009 | x | | | | x | point cloud matching (2D, 3DoF) | submap (2D point cloud, graph, 3DoF edges) | – | | x | | x | | x | | laser (2D) | x | SLAM-based | 245.9 | – | 6.8 | 5d | – | pose error, ROC curves |
| Biber and Duckett 2009 | | x | | | | point cloud matching (2D, 3DoF) | submap (2D point cloud, graph, 3DoF edges) | – | x | x | x | | | x | | wheel odometry, laser (2D) | x | – | 9.6 | – | – | 5w | – | average point cloud likelihood, covariance eigenvalues, memory |
| Hochdorfer and Schlegel 2009 | | | x | | | EKF (2D, 3DoF) | feature (SURF) | – | | x | x | | | x | | wheel odometry, camera (omni) | x | position | 0.115 | – | – | – | – | position error, #map points |
| Hochdorfer, Lutz, et al. 2009 | | | x | | | EKF (2D, 3DoF) | feature (SURF) | – | | x | x | | | x | | wheel odometry, camera (omni) | x | position | 0.15 | – | – | – | – | covariance eigenvalues, position error, #map points |
| Nuske et al. 2009 | x | | | | | particle filter (2D, 3DoF) | feature (building edges) | – | | x | | x | | x | | wheel odometry, camera (mono) | x | laser-based | 3.92 | – | 10.5 | 1d | – | execution time, localization rate, pose error |
| Glover et al. 2010 | x | | | | | visual odometry (2D, 3DoF), Bayesian (location) | experience (graph, pose + local views, 3DoF edges) | – | x | x | | x | | x | | camera (mono) | – | – | – | – | – | – | St Lucia 2007 | confusion matrix, precision-recall, #nodes |
| Kretzschmar, Grisetti, et al. 2010 | | | x | | | point cloud matching (2D, 3DoF) | pose graph (graph, 2D point clouds, 3DoF edges) | – | | x | x | | | x | | laser (2D) | – | – | – | – | – | – | FR079, Intel 2003 | execution time, graph connectivity, #edges, #nodes |
| Ikeda and Kanji 2010 | | | | | x | particle filter (location) | dictionary (semantic hashing) | – | x | x | | x | | x | | camera (mono) | x | GPS | 40 | 20 | – | – | – | execution time, localization rate, memory |
| Dayoub et al. 2011 | | x | | | | feature matching (location) | keyframe (graph, 6DoF edges) | – | | x | x | | | x | | camera (omni) | x | initial position, laser-based | – | – | – | 3d | – | orientation error, similarity score |
| Kretzschmar, Stachniss, and Grisetti 2011 | | | x | | | point cloud matching (2D, 3DoF) | pose graph (graph, 2D point clouds, 3DoF edges) | – | | x | x | | | x | | laser (2D) | – | no pruning | – | – | – | – | FR079, Intel 2003 | #edges, #nodes |
| Pirker et al. 2011 | | | x | | | feature matching (3D, 6DoF) | keyframe (graph) | – | | x | x | x | | x | | camera (gray, mono) | x | – | 1.2 | – | – | 2w | – | position error, #map points |
| Walcott-Bryant et al. 2012 | | x | x | | | point cloud matching (2D, 3DoF) | pose graph (graph, 2D point clouds, 3DoF edges) | – | | x | x | | | x | | laser (2D) | x | – | 8.4 | – | – | 5w | – | execution time, position error, #edges, #nodes |
| Kretzschmar and Stachniss 2012 | | | x | | | point cloud matching (2D, 3DoF) | pose graph (graph, 2D point clouds, 3DoF edges) | – | | x | x | | | x | | laser (2D) | – | – | – | – | – | – | FHW, FR079, FR101, Intel 2003 | execution time, #edges, #nodes |
| Maddern et al. 2012 | | | x | | | wheel odometry (2D, 3DoF), particle filter (location) | pose graph (graph, 3DoF edges) | – | | x | | x | | x | | wheel odometry, camera (color, mono) | – | – | – | – | – | – | New College (FAB-MAP) | execution time, memory, precision-recall |
| Latif, Cadena, et al. 2012 | | | | x | | – | pose graph (graph) | – | | x | x | x | | x | | odometry | – | – | – | – | – | – | Bicocca (indoor), Intel 2003, New College | ATE, execution time |
| Kawewong et al. 2013 | x | | | | | vocabulary tree (location) | dictionary (BoW, hierarchical tree) | – | | x | x | x | | x | | camera | – | – | – | – | – | – | City Center (FAB-MAP), New College (FAB-MAP) | execution time, precision-recall |
| Bacca et al. 2013 | | x | x | | | Bayesian (2D, 3DoF) | keyframe (graph) | – | | x | x | | | x | | camera (omni), laser (2D) | x | no pruning | 1.635 | – | – | 1y | – | pose error, precision-recall, #map points |
| Ball et al. 2013 | x | | | | | visual odometry (2D, 3DoF), feature matching (location) | experience (graph, pose + local views, 3DoF edges) | – | | x | | x | | x | | camera (mono) | – | – | – | – | – | – | New College, St Lucia 2007 | pose error, #nodes |
| Einhorn and Gross 2013 | | x | x | | | odometry | pose graph (graph, 2D/3D NDT) | – | | x | x | | | x | | camera (mono, RGBD), laser (2D) | x | – | 7 | – | 3 | 2d | – | execution time, #nodes |
| Tipaldi et al. 2013 | | x | | | | particle filter (2D, 3DoF) | grid (occupancy, 2D) | – | x | x | x | | | x | | laser (2D) | x | manual, SLAM-based | – | – | – | 1d | – | computational complexity, localization rate, pose error |
| Huang et al. 2013 | | | x | | | odometry, point cloud matching (2D, 3DoF) | pose graph (graph, 2D point clouds, 3DoF edges) | – | | x | x | | | x | | wheel odometry, laser (2D) | x | simulation | – | – | – | – | Intel 2003, MIT Kilian Court | pose error, #nodes |
| Johannsson et al. 2013 | | | x | | | odometry (3D, 6DoF), BoW (location) | keyframe (graph, 6DoF edges) | – | | x | x | | | x | | wheel odometry, camera (stereo, RGBD), IMU | – | manual, no pruning | – | – | – | – | MIT Stata Center | execution time, #localization failures, #nodes |
| Oberländer et al. 2013 | | | | x | x | Fourier-Mellin transform matching (2D, 3DoF) | submap (2D occupancy grid, graph, 3DoF edges) | – | | x | x | | | x | | wheel odometry, laser (2D) | – | SLAM-based | – | – | – | – | albert-b-laser-vision, FR079, Intel 2003 | execution time, pose error, precision-recall |
| Saarinen et al. 2013 | | x | | | | – | 3D NDT, grid (occupancy, 3D) | – | | x | x | | | x | | camera (RGBD), laser (3D) | x | – | 5 | – | 17 | – | TUM RGBD | execution time, map similarity |
| Biswas and Veloso 2013 | x | | | | | particle filter (2D, 3DoF) | feature (2D line segments) | – | | x | x | | | x | | wheel odometry, camera (RGBD), laser (2D) | – | manual | – | – | – | – | CoBots | pose error, #localization failures |
| Paul and Newman 2013 | | | x | | | image classification (location) | database (images, semantic visual topics) | – | x | x | | x | | x | | camera (color, mono) | x | GPS, manual | 28 | – | – | – | City Center (FAB-MAP), New College (FAB-MAP) | execution time, f-beta, precision-recall |
| V. A. Nguyen et al. 2013 | x | | | | | feature matching (location) | pose graph (graph) | – | | x | x | | | x | | camera (color, mono) | – | – | – | – | – | – | COLD | computational complexity, execution time, localization rate |
| Maddern et al. 2013 | | | x | | | particle filter (2D, 3DoF) | pose graph (graph, 3DoF edges) | – | | x | | x | | x | | wheel odometry, camera (color, mono) | – | manual | – | – | – | – | New College | execution time, memory, pose error, precision-recall |

*Table 6: continued from previous page*

| DE: | 1: | | | | | 2: | 3: | 4: | 5: | | 6: | | | | | 7: | 8: | 9: | 10: | | | | 11: | 12: |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| **Ref.** | appearance | dynamic | sparsity | multi-session | computational | **localization** | **mapping** | multi-robot | offline | online | indoor | outdoor | air | ground | water | **sensor** | exp. self-acq. | **ground-truth** | dist. (km) | path (km) | time (h) | int. (d/w/m/y) | **datasets** | **metrics** |
| Churchill and Newman 2013 | x | | | | | visual odometry (3D, 6DoF) | experience (graph, local views, observability edges) | – | | x | x | | | x | | camera (color, stereo) | x | RTK-GPS | 37 | 0.7 | – | 3m | – | execution time, #localization failures |
| Pomerleau et al. 2014 | | x | | | | point cloud matching (3D, 3DoF) | point cloud (laser, 3D) | – | | x | | x | | x | | wheel odometry, laser (3D) | x | map, targeted speed | 3.9 | 1.3 | – | 7m | – | execution time, velocity error |
| Murphy and Sibley 2014 | x | | x | | | image classification (location) | keyframe (graph) | – | | x | | x | | x | | camera (color, mono) | x | – | – | – | – | 1w | New College | execution time, confusion matrix, precision-recall, #nodes |
| Carlevaris-Bianco et al. 2014 | | | x | | | odometry, point cloud matching (2/3D, 3/6DoF) | pose graph (graph) | – | | x | x | x | | x | x | camera (mono), laser (2D/3D) | – | no pruning | – | – | – | – | Intel 2003, MIT Kilian Court, NCLT | execution time, KLD, #nodes |
| Williams et al. 2014 | | | | | x | odometry (3D, 6DoF) | pose graph (graph) | – | | x | | x | x | x | | wheel odometry, camera (color, stereo), IMU | x | RTK-GPS, simulation | – | – | – | – | KITTI | execution time, KLD |
| Einhorn and Gross 2015 | | x | x | | | odometry | pose graph (graph, 2D/3D NDT) | – | | x | x | | | x | | camera (mono, RGBD), laser (2D) | x | – | 7 | – | 3 | 2d | – | execution time, #nodes |
| Pérez et al. 2015 | x | | | | | particle filter (3D, 6DoF) | pose graph (graph, 6DoF edges) | – | x | x | | x | | x | | wheel odometry, camera (gray, stereo), laser (2D) | x | SLAM-based | 11.5 | – | 3.5 | – | – | pose error |
| Li et al. 2015 | x | | | x | | feature matching (location) | pose graph (graph) | – | – | – | | x | | | x | camera (gray, mono) | x | manual | – | – | – | 3y | – | confusion matrix |
| Mohan et al. 2015 | | | | | x | BoW (location) | dictionary (BoW) | – | x | x | | x | | x | | camera (color, mono) | – | – | – | – | – | – | Bicocca (indoor), Ford Campus, Malaga 2009, New College, Nordlandsbanen, St Lucia 2007 | confusion matrix, execution time, precision-recall |
| Dymczyk, Lynen, et al. 2015 | | | x | | | feature matching (location) | keyframe (graph) | – | x | x | | x | | x | | camera (mono) | x | no pruning | 1.034 | – | – | 10d | – | f-score, localization rate, #nodes |
| Rapp et al. 2015 | | x | | | | particle filter (2D, 3DoF) | grid (occupancy, 2D) | – | – | – | | x | | x | | odometry, radar | x | – | – | – | – | – | – | pose error |
| Vysotska et al. 2015 | x | | | | | image sequence matching (location) | database (images, sequence) | – | – | – | | x | | x | | camera (color, mono) | x | manual | 3 | – | – | – | – | computational complexity, confusion matrix, precision-recall |
| Neubert et al. 2015 | x | | | | | feature matching (location) | dictionary (translation, winter - summer) | – | – | – | | x | | x | | camera (color, mono) | – | – | – | – | – | – | Nordlandsbanen | precision-recall |
| Mur-Artal et al. 2015 | | | x | | | bundle adjustment (3D, 6DoF) | keyframe (graph) | – | | x | | x | | x | | camera (gray, mono) | – | – | – | – | – | – | KITTI, New College, TUM RGBD | ATE, execution time, pose error, recall, #nodes |
| Naseer, Suger, et al. 2015 | x | | | | | image sequence matching (location) | – | – | x | | | x | | x | | camera (color, mono) | x | GPS | – | – | – | – | New College (FAB-MAP) | f-beta, precision-recall |
| Karaoğuz and Bozma 2016 | x | | | | | feature matching (location) | pose graph (graph, similarity edges) | – | | x | x | x | | x | | camera (color, mono) | x | – | 0.325 | – | – | – | COLD, New College | execution time, precision-recall |
| Santos et al. 2016 | | x | | | | – | grid (occupancy, 3D) | – | | x | x | | | x | | camera (RGBD) | x | simulation | – | – | – | 5d | – | environment model error, execution time |
| Dymczyk, Stumm, et al. 2016 | x | | | | | feature matching (location) | – | – | | x | x | x | | x | | camera (gray, mono), IMU | x | feature labels | 4.05 | 0.15 | – | 3m | NCLT | execution time, f-score |
| Dymczyk, Schneider, et al. 2016 | | | x | | | – | keyframe (graph) | – | | x | x | | | x | | camera, IMU | x | SLAM-based | – | 0.15 | – | – | – | f-score, #map points |
| Gadd and Newman 2016 | x | | x | | | visual odometry (3D, 6DoF) | experience (graph, local views, 6DoF edges) | x | | x | | x | | x | | camera (grey, mono) | x | – | 100 | – | – | 1m | – | memory, #localization failures |
| Mazuran et al. 2016 | | | x | | | – | pose graph (graph) | – | | x | x | | | x | | – | – | – | – | – | – | – | Intel 2003, MIT Kilian Court | KLD, #nodes |
| Ozog et al. 2016 | | | x | x | | particle filter (3D, 6DoF) | pose graph (graph, planar segments, 6DoF edges) | – | | x | | x | | | x | camera (gray, mono), IMU, DVL | x | map model | 10.159 | – | – | 3y | – | KLD, pose error, #nodes |
| Mühlfellner et al. 2016 | | | x | x | | reprojection minimization (3D, 6DoF) | keyframe (graph) | – | x | x | | x | | x | | wheel odometry, camera (gray, mono) | x | RTK-GPS | 22 | – | – | 1y | – | computational complexity, localization rate, pose error, #map points |
| An et al. 2016 | x | x | | | | EKF (2D, 3DoF) | pose graph (graph) | – | | x | x | | | x | | wheel odometry, camera (gray, mono) | x | manual, simulation | 0.254 | – | 0.33 | – | – | pose error, #edges, #nodes |
| Taisho and Kanji 2016 | x | | | | | image classification (location) | database (images, image features) | – | – | – | | x | | x | | camera (color, mono) | x | manual | – | – | – | – | – | localization rate |
| Han, X. Yang, et al. 2017 | x | | | | | feature matching (location) | – | – | x | x | | x | | x | | camera (color, mono) | – | – | – | – | – | – | CMU-VL, Nordlandsbanen, St Lucia 2007 | execution time, precision-recall |
| Biswas and Veloso 2017 | | x | | | | Bayesian (2D, 3DoF) | feature (2D line segments) | – | | x | x | x | | x | | wheel odometry, laser (2D), camera (RGBD) | – | manual, SLAM-based | – | – | – | – | CoBots | pose error, #localization failures |
| Griffith and Pradalier 2017 | x | | x | x | | SIFT flow (3D, 6DoF) | pose graph (graph, 6DoF edges) | – | x | | | x | | | x | camera (color, mono), GPS, IMU | x | manual | 100 | – | – | 1y2m | – | absolute alignment error |
| Naseer, Oliveira, et al. 2017 | x | | | | x | feature matching (location) | – | – | | x | | x | | x | | camera (color, mono) | x | manual | 100 | – | – | 3y | – | f-score, precision-recall |
| Krajník et al. 2017 | | x | | | | – | grid (occupancy, 3D) | – | | x | x | x | | x | | camera (RGBD) | x | external tracking system | – | – | – | 112d | NCLT, Witham Wharf RGB-D | computational complexity, memory, pose error |
| Ila et al. 2017 | | | x | | | odometry (3D, 6DoF) | pose graph (graph) | – | | x | x | x | | x | | – | – | simulation | – | – | – | – | KITTI | pose error, #nodes |
| Latif, Huang, et al. 2017 | x | | | | x | dictionary search (location) | dictionary (incremental) | – | | x | | x | | x | | camera (mono) | – | – | – | – | – | – | Bicocca (indoor), KITTI, New College | confusion matrix, execution time, precision-recall |
| Xin et al. 2017 | x | | | | | feature matching (location) | – | – | – | – | | x | | x | | camera (mono) | – | – | – | – | – | – | CMU-VL, Gardens Point Campus | computational complexity, f-score, precision-recall |
| Bescos et al. 2018 | | x | | | | bundle adjustment (3D, 6DoF) | keyframe (graph) | – | x | | x | | | x | | camera (color, mono, stereo, RGBD) | – | – | – | – | – | – | KITTI, TUM RGBD | ATE, execution time, pose error |
| Opdenbosch et al. 2018 | | | x | | x | – | keyframe (graph, Hamming distance edges) | – | – | – | x | x | x | | | camera (mono, stereo, RGBD) | – | – | – | – | – | – | EuRoC | memory |
| Han, H. Wang, et al. 2018 | x | | | | | image sequence matching (location) | – | – | – | – | | x | | x | | camera (color, mono) | – | – | – | – | – | – | CMU-VL, Nordlandsbanen, St Lucia 2007 | precision-recall |
| Han, Beleidy, et al. 2018 | x | | | | | feature matching (location) | – | – | – | – | | x | | x | | camera (color, mono) | – | – | – | – | – | – | CMU-VL, Nordlandsbanen | precision-recall |

*Table 6: continued from previous page*

| DE: | 1: | | | | | 2: | 3: | 4: | 5: | | 6: | | | | | 7: | 8: | 9: | 10: | | | | 11: | 12: |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| **Ref.** | **appearance** | **dynamic** | **sparsity** | **multi-session** | **computational** | **localization** | **mapping** | **multi-robot** | **offline** | **online** | **indoor** | **outdoor** | **air** | **ground** | **water** | **sensor** | **exp. self-acq.** | **ground-truth** | **dist. (km)** | **path (km)** | **time (h)** | **int. (d/w/m/y)** | **datasets** | **metrics** |
| Cao, Zhuang, et al. 2018 | x | | | | | odometry (3D, 6DoF), vocabulary tree (location) | pose graph (graph, 2D point clouds, 3DoF edges) | – | | x | | x | | x | | wheel odometry, laser (2D, 3D), IMU | x | – | – | – | – | 1d | – | execution time, precision-recall |
| Nobre et al. 2018 | x | | | | | Mahalanobis distance minimization (2D, 3DoF) | feature | – | | x | x | | | x | | wheel odometry, camera (color, mono) | x | simulation | – | – | – | – | UTIAS Multi-Robot | precision-recall |
| Hui Zhang et al. 2018 | | | x | | | feature matching (3D, 6DoF) | keyframe (graph) | x | | x | x | x | | x | | camera (mono) | x | – | – | – | – | – | KITTI | localization rate |
| J. Zhu et al. 2018 | x | | | | x | image sequence matching (location) | – | – | x | | | x | | x | | camera (color, mono) | – | – | – | – | – | – | City Center (FAB-MAP), Nordlandsbanen | precision-recall |
| MacTavish et al. 2018 | x | | | | | visual odometry (3D, 6DoF) | keyframe (graph) | – | | x | | x | | x | | camera (stereo) | x | external tracking system | 26.08 | 0.16 | – | 4m | – | execution time, localization rate, pose covariance, pose error |
| L. Sun et al. 2018 | | x | | | | point cloud matching (3D, 6DoF) | grid (occupancy, 3D) | – | x | x | | x | | x | | laser (3D) | x | manual | – | – | – | 2w | KITTI | map accuracy |
| Lázaro et al. 2018 | | x | x | x | | point cloud matching (2D, 3DoF) | submap (2D point cloud, graph, 3DoF edges) | – | | x | x | | | x | | wheel odometry, laser (2D) | – | – | – | – | – | – | MIT Stata Center, Witham Wharf RGB-D | execution time, #edges, #nodes |
| N. Zhang et al. 2018 | x | | | | x | visual odometry (3D, 6DoF) | keyframe (graph) | – | | x | | x | | x | | camera (stereo) | x | – | 25 | 0.25 | – | 4m | – | matching accuracy |
| Chebrolu et al. 2018 | x | | | x | | feature matching (location) | – | – | – | – | | x | x | | | camera (color, mono), GPS | x | manual | – | – | – | 1m | – | matching accuracy |
| P. Yin, L. Xu, et al. 2018 | x | | | | x | feature matching (location) | grid (occupancy, 3D) | – | – | – | | x | | x | | laser (3D) | – | – | – | – | – | – | KITTI, NCLT | precision-recall |
| Egger et al. 2018 | x | | x | | | feature matching (3D, 6DoF), odometry (3D, 6DoF) | submap (surfel, graph) | – | | x | | x | | x | | wheel odometry, laser (3D), IMU | x | RTK-GPS | – | – | – | 1y6m | – | execution time, memory |
| Arroyo et al. 2018 | x | | | | x | image sequence matching (location) | – | – | – | – | | x | | x | | camera (mono, stereo) | – | – | – | – | – | – | KITTI, New College, Nordlandsbanen | confusion matrix, execution time, precision-recall |
| Ouerghi et al. 2018 | x | | | | | image sequence matching (location), visual odometry (3D, 2DoF) | keyframe (graph) | – | | x | | x | | x | | wheel odometry, camera (mono) | – | – | – | – | – | – | KITTI | execution time, pose error |
| Siva and Hao Zhang 2018 | x | | | | | feature matching (location) | – | – | – | – | | x | | x | | camera (omni) | x | GPS | 19.15 | – | – | 1y | – | precision-recall |
| Luthardt et al. 2018 | x | | x | | | visual odometry (3D, 6DoF) | pose graph (graph) | – | x | x | | x | | x | | camera (gray, mono) | x | GPS | – | – | – | – | – | pose error |
| Chen et al. 2018 | x | | | | | image classification (location) | – | – | – | – | | x | | x | | camera (color, mono) | – | – | – | – | – | – | Nordlandsbanen, St Lucia 2007 | precision-recall |
| Yu et al. 2019 | x | | | | | vocabulary hashing (location) | – | – | – | – | | x | | x | | camera (color, mono) | – | – | – | – | – | – | City Center (FAB-MAP), New College (FAB-MAP) | confusion matrix, precision-recall |
| Boniardi et al. 2019 | | x | x | | | point cloud matching (2D, 3DoF) | pose graph (graph, 2D point clouds, 3DoF edges) | – | | x | x | | | x | | laser (2D) | x | external tracking system, map model | 4.657 | – | 2.85 | 4d | – | execution time, pose error, #nodes |
| G. Kim et al. 2019 | x | | | | | feature matching (location) | grid (location, 2D) | – | – | – | | x | | x | | laser (3D) | – | – | – | – | – | – | NCLT, Oxford RobotCar | precision-recall |
| Berrio, Ward, et al. 2019 | x | | x | | | point cloud matching (2D, 3DoF) | feature (pole, corners) | – | x | | | x | | x | | laser (3D) | x | manual | – | – | – | 6m | – | pose covariance, #map points |
| K. Wang et al. 2019 | | x | | | | bundle adjustment (3D, 6DoF) | keyframe (graph) | – | – | – | x | x | | x | | camera (RGBD) | x | simulation | – | – | – | – | TUM RGBD | ATE, precision |
| L. Wu and Y. Wu 2019 | | | | | x | image classification (location) | – | – | | x | | x | | x | | camera (color, mono) | – | – | – | – | – | – | Gardens Point Campus, Nordlandsbanen | execution time, f-score, precision-recall |
| Tang, Y. Wang, Ding, et al. 2019 | x | | x | | | BoW (location), point cloud matching (3D, 6DoF) | submap (graph, manifold) | – | | x | | x | | x | | camera (color, stereo) | – | SLAM-based | – | – | – | – | YQ21 | execution time, localization rate, pose error, #nodes |
| Bürki et al. 2019 | x | | | x | | feature matching (3D, 6DoF) | keyframe (graph) | – | x | x | | x | | x | | wheel odometry, sensor (gray, mono) | x | – | – | – | – | 1y | NCLT | execution time, pose error |
| Labbé and Michaud 2019 | | | | x | x | BoW (location), odometry (2/3D, 3/6DoF) | pose graph (graph) | – | | x | x | x | x | x | | wheel odometry, camera (stereo, RGBD), laser (2D/3D) | – | – | – | – | – | – | EuRoC, KITTI, MIT Stata Center, TUM RGBD | ATE, execution time, #nodes |
| M. Zhang et al. 2019 | | x | | | | point cloud matching (2D, 3DoF) | grid (occupancy, 2D) | – | | x | x | | | x | | laser (2D) | x | SLAM-based | – | – | – | – | – | execution time, pose error |
| Schmuck and Chli 2019 | | | x | | | odometry (3D, 6DoF) | keyframe (graph) | – | | x | x | | x | | | camera, IMU | – | – | – | – | – | – | EuRoC | pose error |
| Ganti and Waslander 2019 | | x | | | | bundle adjustment (3D, 6DoF) | keyframe (graph) | – | – | – | | x | | x | | camera (stereo) | – | – | – | – | – | – | KITTI | pose error |
| Ding, Y. Wang, Tang, et al. 2019 | | | x | | x | EKF (3D, 6DoF) | keyframe (graph) | – | | x | | x | | x | | camera (stereo), IMU | x | laser-based | – | – | 1.32 | 1y | – | ATE, communication constraints |
| Song et al. 2019 | | x | | | | odometry (3D, 6DoF) | keyframe (graph) | – | | x | | x | | x | | camera, IMU | x | RTK-GPS | – | – | – | – | – | pose error |
| Pan et al. 2019 | | x | | | | odometry, reprojection minimization (3D, 6DoF) | feature (point clusters) | – | x | x | | x | | x | | camera (mono), laser (3D) | x | – | – | – | – | 3m | KITTI | execution time, localization rate, pose error, reprojection error |
| A. J. B. Ali et al. 2020 | | | | | x | visual odometry (3D, 6DoF) | keyframe (graph) | – | | x | x | | | x | | camera (RGBD) | x | laser-based | – | – | 0.5 | – | TUM RGBD | communication constraints, execution time, localization rate, memory, pose error |
| C. Qin et al. 2020 | x | | | | | feature matching (location) | – | – | | x | | x | | x | | camera (color, mono) | – | – | – | – | – | – | Alderley, FAS, Nordlandsbanen, Oxford RobotCar, St Lucia 2007 | confusion matrix, execution time, precision-recall |
| Martini et al. 2020 | x | | | | | feature matching (location), point cloud matching (2D, 3DoF) | experience (graph) | – | x | x | | x | | x | | radar | – | – | – | – | – | – | Oxford Radar RobotCar | confusion matrix, localization rate, pose error, precision-recall |
| Karaoğuz and Bozma 2020 | | | x | | | – | pose graph (graph) | x | | x | | x | | x | | camera (mono) | x | – | – | – | – | – | COLD | computational complexity, precision-recall |

*Table 6: continued from previous page*

| DE: | 1: | | | | | 2: | 3: | 4: | 5: | | 6: | | | | | 7: | 8: | 9: | 10: | | | | 11: | 12: |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| **Ref.** | **appearance** | **dynamic** | **sparsity** | **multi-session** | **computational** | **localization** | **mapping** | **multi-robot** | **offline** | **online** | **indoor** | **outdoor** | **air** | **ground** | **water** | **sensor** | **exp. self-acq.** | **ground-truth** | **dist. (km)** | **path (km)** | **time (h)** | **int. (d/w/m/y)** | **datasets** | **metrics** |
| H. Yin, Y. Wang, et al. 2020 | x | | | | x | particle filter (2D, 3DoF), point cloud matching (3D, 6DoF) | pose graph (graph, 6DoF edges) | – | | x | | x | | x | | laser (3D) | – | – | – | – | – | – | KITTI, YQ21 | execution time, f-score, pose error, precision-recall |
| Clement et al. 2020 | x | | | | | feature matching (3D, 6DoF) | – | – | – | – | | x | | x | | camera (color, mono) | – | – | – | – | – | – | Oxford RobotCar | confusion matrix, matching accuracy |
| L. Wang et al. 2020 | | x | | | | particle filter (2D, 3DoF) | grid (occupancy, 2D) | – | | x | x | | | x | | wheel odometry, laser (2D) | x | simulation, SLAM-based | – | – | – | – | – | execution time, memory, pose error |
| Camara et al. 2020 | x | | | | x | feature matching (location) | – | – | | x | | x | | x | | camera | – | – | – | – | – | – | Berlin Kudamm, Gardens Point Campus, Nordlandsbanen | execution time, memory, precision-recall |
| Gao and Hao Zhang 2020 | x | | | | | feature matching (location) | – | – | – | – | | x | | x | | camera (color, mono) | – | – | – | – | – | – | CMU-VL, St Lucia 2007 | precision-recall |
| S. Yang et al. 2020 | | x | | | | visual odometry (3D, 6DoF) | keyframe (graph) | – | | x | x | | | x | | camera (RGBD) | – | – | – | – | – | – | TUM RGBD | ATE, execution time, pose error |
| Siva, Nahman, et al. 2020 | x | | | | | feature matching (location) | – | – | – | – | | x | | x | | laser (3D) | x | simulation | – | – | – | – | NCLT | precision-recall |
| T. Qin et al. 2020 | x | | | | | EKF (2D, 3DoF) | feature (semantic) | – | | x | x | x | | x | | wheel odometry, camera (color, mono), IMU | x | RTK-GPS | 0.324 | – | – | 1m | – | ATE, memory, recall |
| Ding, Y. Wang, Xiong, et al. 2020 | x | x | | x | | bundle adjustment (3D, 6DoF) | point cloud (3D) | – | x | x | | x | | x | | camera (color, stereo), laser (3D) | – | – | – | – | – | – | KITTI, YQ21 | ATE, execution time, pose error |
| Yue et al. 2020 | | x | | | | – | point cloud (3D) | x | | x | | x | | x | | camera (color, mono, thermal), laser (3D) | x | – | – | – | – | – | – | ATE, memory |
| Schaefer et al. 2021 | x | | x | | | particle filter (2D, 3DoF) | feature (poles) | – | x | x | | x | | x | | laser (3D) | – | – | – | – | – | – | KITTI, NCLT | pose error |
| Liu et al. 2021 | x | | | | | EKF (3D, 6DoF), feature matching (location) | keyframe (graph) | – | | x | x | x | | x | | camera (color, mono), IMU | – | – | – | – | – | – | City Center (FAB-MAP), KITTI, New College (FAB-MAP) | execution time, memory, precision-recall |
| C. Kim et al. 2021 | | | | | x | particle filter, point cloud matching (3D, 6DoF) | grid (geodetic, NDT) | – | | x | | x | | x | | laser (3D) | x | RTK-GPS, SLAM-based | – | – | – | – | KITTI | memory, pose error |
| Derner et al. 2021 | x | | | | | feature matching (3D, 6DoF) | database (images, features, pose) | – | x | x | x | | | x | | wheel odometry, camera (RGBD) | x | manual | 0.198 | – | – | – | Witham Wharf RGB-D | execution time, localization rate, pose error |
| Cao, Yan, et al. 2021 | x | | | | | sequence matching (location) | – | – | | x | | x | | x | | laser (2D/3D) | – | – | – | – | – | – | NCLT, Oxford RobotCar | execution time, precision-recall |
| G. Singh et al. 2021 | x | x | | | | feature matching (location) | pose graph (graph, BoW) | – | | x | | x | | x | | camera (stereo, RGBD) | – | – | – | – | – | – | CBD, KITTI | execution time, precision-recall |
| Kurz et al. 2021 | | | x | | | – | pose graph (graph) | – | | x | x | | | x | | wheel odometry, laser (2D), IMU | – | no pruning | – | – | – | – | MIT Stata Center, Witham Wharf RGB-D | execution time, pose error, #nodes |
| H. Yin, X. Xu, et al. 2021 | x | | | | | location matching (location) | – | – | – | – | | x | | x | | laser (3D), radar | – | – | – | – | – | – | MulRan, Oxford Radar RobotCar | confusion matrix, precision-recall |
| Thomas et al. 2021 | | x | | | | point cloud matching (3D, 6DoF) | grid (occupancy, 3D) | – | x | x | x | | | x | | wheel odometry, laser (3D) | x | simulation | – | – | – | – | – | confusion matrix, execution time, precision-recall |
| Berrio, Worrall, et al. 2021 | x | | | | | – | grid (feature, 2D) | – | – | – | | x | | x | | wheel odometry, camera (color, mono), laser (3D), IMU | – | – | – | – | – | – | USyd Campus | pose covariance |
| Oh and Eoh 2021 | x | | | | | feature matching (location) | – | – | – | – | | x | | x | | camera (color, mono) | – | – | – | – | – | – | KAIST, Nordlandsbanen | precision-recall |
| Tsintotas et al. 2021 | | | x | | | BoW (location) | dictionary (BoTW, incremental) | – | | x | x | x | x | x | | camera (mono) | – | – | – | – | – | – | City Center (FAB-MAP), EuRoC, KITTI, Lip6Indoor, Lip6Outdoor, Malaga 2009 | execution time, memory, precision-recall |
| Li Sun et al. 2021 | x | | | | | feature matching (3D, 6DoF) | keyframe (graph) | – | | x | | x | | x | | camera (color, mono) | x | SLAM-based | 0.741 | – | – | 1d | – | ATE, execution time, matching error, pose error |
| Tang, Y. Wang, Tan, et al. 2021 | x | | | | | feature matching (location) | – | – | – | – | | x | | x | | camera (color, mono) | – | – | – | – | – | – | Alderley, Nordlandsbanen, Oxford RobotCar, YQ21 | localization rate, precision-recall |
| Piasco et al. 2021 | x | | | | | feature matching (location) | – | – | x | x | | x | | x | | camera (RGBD) | – | – | – | – | – | – | CMU-VL, Oxford RobotCar | precision-recall |
| P. Yin, J. Xu, et al. 2021 | x | | | | | feature matching, sequence matching (location) | – | – | | x | | x | | x | | laser (3D) | x | – | 132 | 11 | – | – | KITTI, NCLT | execution time, memory, precision-recall |
| Meng et al. 2021 | x | | | | | laser odometry (3D, 6DoF) | pose graph (graph) | – | | x | | x | | x | | laser (3D) | – | – | – | – | – | – | KITTI | ATE, execution time, pose error |
| S. Zhu et al. 2021 | | x | | | | particle filter (2D, 3DoF) | grid (occupancy, 2D) | – | x | x | x | | | x | | wheel odometry, camera (color, mono), laser (2D), IMU | x | manual | – | – | – | – | – | pose error |
| Zeng and Si 2021 | | | x | | | – | pose graph (graph) | – | | x | x | | | x | | wheel odometry, camera (color, mono) | x | no pruning | – | – | – | – | – | #edges, #nodes |
| W. Ali et al. 2021 | | | x | | x | point cloud matching, visual odometry (3D, 6DoF) | keyframe (graph), submap (image, graph) | – | | x | x | x | | x | | camera, laser (3D) | x | – | – | – | – | – | KITTI | CPU usage, memory, pose error, precision-recall |
| X. Xu et al. 2021 | x | | | | | feature matching (location) | – | – | | x | | x | | x | | laser (3D) | – | – | – | – | – | – | MulRan, NCLT, Oxford RobotCar | execution time, precision-recall |
| Z. Yang et al. 2021 | x | | | | x | BoW, feature matching (location) | – | – | | x | | x | | x | | camera (color, mono) | – | – | – | – | – | – | City Center (FAB-MAP), KITTI, Lip6Indoor, Lip6Outdoor, Malaga 2009, New College | confusion matrix, execution time, precision-recall |
| Z. Wang et al. 2021 | x | | x | | | point cloud matching (3D, 6DoF) | feature (poles) | – | – | – | | x | | x | | laser (3D) | x | GPS | 5.52 | – | – | 1m | – | localization rate, pose error |
| Hu et al. 2022 | x | | | | | feature matching (location) | – | – | – | – | | x | | x | | camera (color, mono) | x | RTK-GPS | – | – | – | – | CMU-Seasons, RobotCar Seasons | execution time, precision-recall |
| Coulin et al. 2022 | x | | | | | EKF (3D, 6DoF) | magnetic (vector) | – | x | x | x | | | x | | camera (stereo), IMU | x | SLAM-based | 1.665 | – | – | 1y | – | ATE, execution time |
| K. Zhang et al. 2022 | x | | | | | feature matching (location) | – | – | | x | | x | | x | | camera (color, mono) | – | manual | – | – | – | – | City Center (FAB-MAP), KITTI, Malaga 2009, St Lucia 2007 | execution time, precision-recall |

*Table 6: continued from previous page*

| DE: | 1: | | | | | 2: | 3: | 4: | 5: | | 6: | | | | | 7: | 8: | 9: | 10: | | | | 11: | 12: |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| **Ref.** | **appearance** | **dynamic** | **sparsity** | **multi-session** | **computational** | **localization** | **mapping** | **multi-robot** | **offline** | **online** | **indoor** | **outdoor** | **air** | **ground** | **water** | **sensor** | **exp. self-acq.** | **ground-truth** | **dist.** (km) | **path** (km) | **time** (h) | **int.** (d/w/m/y) | **datasets** | **metrics** |
| T.-M. Nguyen et al. 2022 | x | | | | | bundle adjustment, sensor fusion (3D, 6DoF) | – | – | | x | x | | x | | | camera (mono), laser (3D), IMU, UWB | – | – | – | – | – | – | EuRoC, NTU VIRAL | execution time, pose error |
| Bouaziz et al. 2022 | | | x | x | | feature matching (3D, 6DoF) | keyframe (graph) | – | x | x | | x | | x | | camera (gray, mono) | – | – | – | – | – | – | IPLT, Oxford RobotCar | execution time, memory, #localization failures |
| Du et al. 2022 | | x | | | | reprojection minimization (3D, 6DoF) | keyframe (graph) | – | | x | x | | | x | | camera (RGBD) | – | – | – | – | – | – | Bonn RGB-D Dynamic, TUM RGBD | ATE, execution time, pose error |
| Xing et al. 2022 | | x | | | | feature matching (3D, 6DoF) | keyframe (graph) | – | | x | x | x | x | x | | camera (RGBD), IMU | x | – | – | – | – | – | EuRoC, KITTI, TUM RGBD | execution time, localization rate, pose error |
| Hong et al. 2022 | x | | | | | feature matching (location), point cloud matching (2D, 3DoF) | keyframe (graph) | – | | x | | x | | x | | radar | – | – | – | – | – | – | MulRan, Oxford RobotCar, RADIATE | ATE, execution time, pose error |