Map Summarization for Tractable Lifelong Mapping

Marcin Dymczyk, Igor Gilitschenski, Roland Siegwart, and Elena Stumm
Autonomous Systems Lab, ETH Zurich

***Abstract*—Accurate pose estimation has become a core capability for modern robotics, required in a multitude of applications, where precise navigation or collaboration is involved. In beacon-free and GPS-denied environments, visual-based odometry and localization systems are a popular solution for obtaining reliable, high-frequency pose estimates. The current state-of-the-art approaches, however, are limited to rather short timespans and small scales due to the excessive amount of data which is being produced in each mapping session. This issue is particularly relevant if a wireless online communication between mapping peers is required, where bandwidth is limited.**

This paper investigates map summarization and reducing the data flow in visual feature-based localization systems. We examine the data storage and transmission requirements of state-of-the-art localization systems and identify potential remedies to existing limitations. We discuss choices and methods related to the localization map representation, lifelong map maintenance and landmark selection approaches. Finally, we evaluate some of the discussed methods in a real-life scenario, using an autonomous mobile robot repeatedly mapping an ever-changing office space over a period of 2 months. We prove it is possible to reduce the data transfer by a factor of 5 while maintaining good localization performance.

I. Introduction

Accurate estimation of the pose of a robot has become one of the fundamental components of modern robotic systems. These systems generally require precise and high-frequency pose information, while keeping the cost and the weight of the required hardware as low as possible. If we assume no beacons or external signals to be available, a visual-inertial localization and mapping system seems to be a good fit for many pose estimation applications, and is therefore the focus of this work.

Regardless of whether we would like to estimate a pose in a global frame or relative poses between robots in a multiagent mapping scenario, we need to provide an agent with some prior data with which to localize against. An agent can localize itself against prior observations, a previously recorded map, or a map that is currently being recorded by another peer. In many cases, especially in collaborative and large-scale scenarios, the map needs to be transmitted to the local agent, and is subject to often strict bandwidth and storage limitations.

This paper considers the problem of reducing the data flow that is induced by large-scale, long-term localization systems. Particularly, we assume that the entire setup contains multiple mapping agents and a backend, which maintains a consistent factor-graph representation of the environment. The map is built upon sparse 3D visual landmarks, represented by binary 2D image descriptors such as BRISK [18]. The considered localization system establishes 2D-3D matches between a query keyframe and the prior 3D model of the environment [13, 32] (in contrast to the techniques related to image retrieval, such as [7] or [14] and their variants). Initially, as in [22], an efficient nearest neighbor search in the descriptor space is performed and then pruned based on the covisibility graph [35], built from the past landmark observations. The remaining 2D-3D matches are then passed to a perspective-n-point RANSAC procedure yielding pose information and inlier matching pairs.

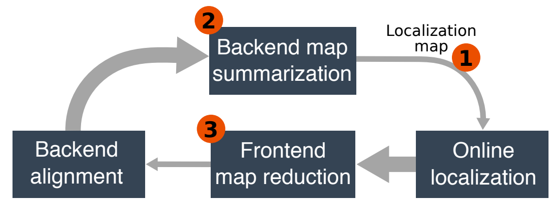

Fig. 1. A typical data flow diagram of a lifelong localization system. Multiple agents perform mapping and online localization, while the backend collects local maps, co-registers them and creates a single, consistent representation. The data can be then sent back to agents for localization. In this paper, we discuss multiple methods of reducing the data flow: **1)** a compact localization map representation, **2)** landmark selection algorithms in the backend and **3)** feature selection methods in the frontend.

Our main goal is to provide a brief survey of existing approaches and present directions of future research for map summarization in these types of systems. Particularly, we consider three different aspects:

1) A summarized localization map representation, e.g. by reducing descriptor dimensionality.
2) Long-term map data maintenance in the backend, e.g. by deleting redundant, outdated, and unused data.
3) Agent-side data selection, e.g. by choosing landmarks particularly suitable for localization.

A generic data flow diagram of the considered *localization loop* is shown in Figure 1. In order to reduce the overall amount of data transfer, we can deploy various algorithms at multiple stages of the workflow, indicated by corresponding numbers in the figure. These approaches are discussed in the subsequent sections below. To validate our claims, we present a preliminary evaluation involving a subset of the discussed methods that considers all three aspects in a multisession mapping framework. We prove that even in a challenging, dynamic office space we can significantly reduce the amount of data transfer, maintain a map of a constant size in the backend, and still retain high place recognition performance.

This paper proceeds by posing the lifelong large-scale localization problem in Section II and discusses state-of-the-art approaches in Section III. A selection of the map summarization methods is evaluated in Section IV, using an autonomous mobile robot in a dynamic office space. Finally, in Section V, we present our views on the future of map compression and summarization.

II. The Problem of Large-Scale Lifelong Mapping

A growing interest in robot collaboration, long-term autonomy, and increased robustness has sparked intensive research in the field of vision-based pose retrieval. The ability to transfer the localization data on-the-fly to the agents and perform pose retrieval onboard to tightly couple the estimates with a visual-inertial odometry estimator [24, 21] is of crucial importance.

But due to bandwidth and computational power constraints, the existing multiagent systems usually need to delegate the loopclosure and relocalization task to the backend, e.g. [10], [24]. We believe these limitations can be attributed to two major factors which lifelong, large-scale localization approaches need to tackle:

- *spatial extent:* with the growing sizes of covered areas, the effort needed to store and maintain maps is quickly becoming prohibitive,
- *time scales:* multiple visits to the same area result in redundant data or even contradictory measurements (e.g. when the environment has changed).

In this paper, we deal with visual feature-based localization systems. Following this assumption, we can specify the sources of excessive data much more precisely:

- a suboptimal map representation, containing data that is unnecessary for the localization system (e.g. raw IMU data is not necessary for actual pose retrieval),
- a considerable size of descriptors that are used to perform place recognition (holds also for binary descriptors),
- an excessive number of 3D map landmarks, some of which might:
  - have poor position estimates,
  - belong to temporary objects,
  - be prone to changing conditions such as lighting,
  - have a low probability of redetection and matching due to the specific detector and descriptor properties,
  - be relevant for the visual odometry pipeline, but are not necessarily useful for localization.

As a result, in general, raw maps have a considerable size which makes it prohibitive to exchange them in realtime over wireless communication. With these potential sources of extraneous data in mind, we consider how to construct a framework where agents and the backend are able to efficiently exchange localization data on-the-fly. Therefore our goal is to find methods that will reduce data transfer requirements to make them tractable in typical robotics scenarios while maintaining good localization performance. A diagram of such dataflow is depicted in Figure 1.

III. Data reduction in localization systems

We present several recent and novel general concepts in map compression and summarization to minimize bandwidth and storage requirements of lifelong visual feature-based mapping systems. Altogether, the proposed approaches address all the concerns that were mentioned in Section II. We believe the existing work related to map summarization can be subdivided into three major categories (all of which are visualized in Figure 1 according to the following numbering):

1) ways to represent the localization map efficiently,
2) methods to maintain a global, consistent map over long periods of time,
3) methods to reduce the number of landmarks directly after the detection stage.

The sections below discuss the approaches that belong to each of these categories. In section III-A, we discuss different representations of compact localization maps, that are meant to reduce the storage and bandwidth requirements of the localization system. In section III-B we describe ways to maintain a global consistent map in the backend over extended periods of time and use it to build compact localization maps. In general, the process is based upon landmark selection in the backend while trying to guarantee a good feature coverage of the entire environment. Finally, in section III-C we discuss approaches for selecting features directly after the detection stage, on the agent side. This results in a significant reduction of the data flow from the agents to the backend, even before the global map is created.

A. The localization map structure

When vision-based mapping algorithms started to be deployed in large-scale and long time horizon settings, the amount of produced data became difficult to store, transfer, and maintain. Additionally, the factor-graph optimization has become computationally expensive. Therefore, most map reduction related contributions were not looking into the problem of localization, but mainly focused on pruning the factor-graphs to reduce the computational complexity of the refinement operations [5, 27].

An approach explicitly meant to maintain lifelong localization maps is presented by Walcott-Bryant *et al.* [38]. The authors propose a method to fuse local maps into a single global map, remove the outdated factor-graph nodes, and try to isolate dynamic objects. An extensive evaluation of an information-based graph-pruning method is presented by Ila *et al.* [12], where both reduced memory footprint and faster computations are reported. In both works the gain is a result of removing the factor-graph nodes, which in our opinion is too coarse a granularity for visual feature-based approaches. In the context of place recognition, it in general makes more sense to reason over the underlying environment (3D landmarks), rather than the robot trajectory (keyframes).

The problem of localization maps can be approached from a different direction. Instead of pruning the factor-graph to obtain a sparse representation, it might be worth defining a separate data container that suits the needs of the localization engine and is aimed to reduce the data size. In [21], Lynen *et al.* are using a 2D-3D pose retrieval pipeline and therefore define a representation that contains:

- triangulated 3D landmarks,

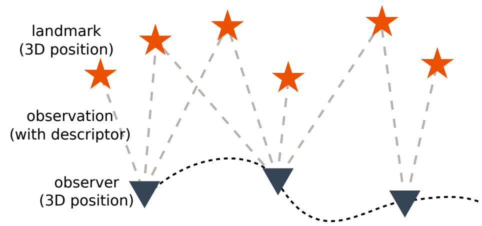

Fig. 2. A compact localization map representation, presented in [21], suitable for localization systems that rely on 2D-3D matching. This type of matching means, that for each keypoint of the query frame we try to match a corresponding triangulated 3D landmark of the prior localization map. The additional covisibility information stored in this compact map representation is used to filter the descriptor matches, as in [35].

- observations of these landmarks with corresponding compressed descriptors,
- observer pose information to perform covisibility filtering in the metric space.

Such a map representation is presented in Figure 2. All the other data (raw images, IMU data, raw descriptors etc.) is retained only in the backend or completely discarded depending on the particular requirements.

An additional step that can be used to further reduce the localization map size is to compress the visual descriptor information. A rather straightforwad method was suggested by Lynen *et al.* [22]. They propose to project the binary descriptors into a lower dimensional Euclidean space using PCA, which only marginally reduces the performance (but makes the place recognition queries easier). This technique was originally applied by Bosse *et al.* to laser keypoints [3], but can be easily adapted to any type of high-dimensional data. The research on the descriptor compression yields even more advanced approaches, such as [1] where a Set Compression Tree is used to store large sets of descriptors and perform an efficient nearest neighbor search.

In general, we believe the localization map representation can either be a sparser version of the global map or rather a separate structure, carefully crafted for compactness and specific localization engine needs. The additional means of data compression, such as novel data containers or descriptor projection can not only further reduce the data size, but also make the nearest neighbor search more efficient.

B. Lifelong mapping and building the localization maps

Apart from deciding upon the localization map representation, it is necessary to choose a way to populate it. The question that needs to be answered is: do we need to fill such map with all the data from all the mapping sessions? Is it possible to select only a subset of landmarks while retaining satisfying place recognition performance?

One of the first works aimed to handle the visual-based lifelong mapping problem was presented by Konolige *et al.* [16]. The authors proposed a method to remove outdated views in the map and to keep only exemplars of places that were visited multiple times. It is therefore possible to avoid data redundancy in the global map.

An approach to feature selection for localization purposes was presented in [9]. A greedy selection procedure was presented, where the landmarks were scored according to the observation statistics while guaranteeing that each region of the map is properly covered. A similar approach was suggested in [30], where the greedy procedure was replaced with an integer linear or quadratic programming problem statement. In [8], a further extension was provided, where the problem is divided into subregions which makes the landmark selection procedure tractable even for very large maps. Both [30] and [8] suggest a similar formulation of the feature selection problem, which corresponds to the *maximum coverage* problem, and solve it using integer linear programming:

$$\begin{aligned} \text{minimize } & \mathbf{q}^T\mathbf{x} + \lambda \mathbf{1}^T \boldsymbol{\zeta} \\ \text{subject to } & \mathbf{A}\mathbf{x} + \boldsymbol{\zeta} \geq b\mathbf{1} \\ & \sum_{i=1}^{N} \mathbf{x}_i = n_{desired} \end{aligned} \tag{1}$$

where $\mathbf{q}$ is a vector of landmark scores, $n_{desired}$ is a target number of landmarks, ζ is a slack variable, $\mathbf{A}$ is a visibility matrix and the constraint $\mathbf{A}\mathbf{x} + \boldsymbol{\zeta} \geq b\mathbf{1}$ guarantees that each keyframe in the map would observe enough landmarks. Each binary variable x_i in $\mathbf{x}$ corresponds to a single landmark that can be either switched on or off.

A similar concept is presented in [26]. There, the authors propose a few scoring functions that might be used to order the map landmarks according to their utility. The evaluation is conducted using numerous mapping sessions in the same outdoor environment, which confirms the lifelong property of the selected landmarks. The insights resulting from these experiments are inline with those presented in [9].

An interesting view on the map reduction was presented by Steiner *et al.* [34]. Instead of attempting to cover the entire environment with landmarks, they look for places that are particularly suitable for localization, introducing *location utility* metric. The landmarks belonging only to the N best places are then kept in the place recognition database. An extension of this work is presented in [33], where the landmarks are sorted based on the *landmark utility* metric.

Yet another approach is presented in [2]. Therein, the aim is to find a minimal set of landmarks that bounds the uncertainty along the robot's trajectory. The method though seems to be particularly suitable for hand-crafted environments and is not robust against scene changes. Additionally, it does not factor in the problem of reliably redetecting the landmark, which we believe to be one of the major causes of unsuccessful localizations.

Combining the information-based landmark selection with application-specific criteria was recently presented by Mu *et al.* [25]. Their method consists of two stages: first we select the landmarks suitable for the specific task and then we subsample them to maximize the information gain. The evaluation, however, is based on the April Tag [29] detection, which have a much higher probability of redetection when compared to the typical local visual features used in robotics.

The different approaches presented above tackle the lifelong mapping and landmark selection problems in various ways. The question we would like to pose is whether the information-gain based landmark selection works well for the real-life keypoint-based approaches? It seems that the probability of the redetection, and not just favorable location, plays a major role in the landmark utility. On the other hand, using just the past mapping experience might lead to degenerate configurations, where the pose cannot be retrieved (or only with a very poor accuracy).

C. Agent-side data selection

Even when using compact localization maps, the system as presented in Figure 1 still needs to transfer the local mapping data from the agents to the mapping backend. A significant fraction of this data transfer is induced by triangulated 3D landmarks and associated descriptors. Would it be possible to transfer only a subset of landmark while still being able to build a global map in the backend that can be used for localization?

The basic solution for the image feature selection are based upon the score from the keypoint detector, often it would be a Harris corner detector or a Difference of Gaussian detector [20]. Many odometry systems use the detector in conjunction with additional criteria that are meant to guarantee a desired distribution of keypoints in the image plane, e.g. in [19]. While these solutions are well suited for the visual odometry purposes, they do not take into account the specific requirements of the localization systems, such as keypoint matchability, viewpoint invariance, or a unique descriptor pattern.

In [4], Buoncompagni *et al.* present a methodology to evaluate keypoints for object detection and matching purposes. They suggest multiple keypoint saliency criteria, trying to grasp the notions of distinctiveness, repeatability and detectability. The approach uses only the raw image to evaluate a keypoint, without taking into account additional available information, such as scene geometry or keypoint tracking performance. Additionally, detectability is only evaluated using the FAST [31] detector score, which is merely a rough approximation of a long-term probability of redetecting the feature.

An interesting idea is presented in [15]. With the increasing accuracy of semantic labelling methods, using the semantic information to reject dynamic or unstable map landmarks is a natural choice, particularly outdoors.

The quality of SIFT features to guarantee robust matching was evaluated in [11]. Based on the tracking information and false positive matches, a Random Forest was trained to classify features in the descriptor space. An evaluation confirms the proposed feature selection method is useful for localization, efficiently rejecting non-distinctive and confusing features such as the ones coming from vegetation.

The final question is how should we combine the above presented approaches. Many of them will output complementary information, as some operate in the image space while others take into account the map geometry, descriptors or the feature matching history. One can therefore expect that combining the methods in a sound statistical approach can bring a significant improvement to the quality of feature selection.

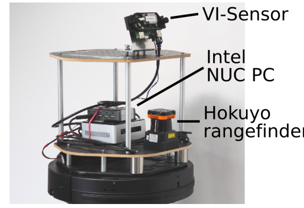

Fig. 3. The Turtlebot robot used to map the office space.

IV. PRELIMINARY EXPERIMENTS

To evaluate a concept of a system incorporating all three categories of approaches presented in Section III, we have set up a mobile robot, see Figure 3, that was autonomously performing mapping and localization using a selection of presented methods. That is, our aim was to verify if the agent-side landmark reduction together with the backend-side map reduction and a localization map representation can provide a reliable localization output.

The robot was equipped with the visual-inertial sensor presented in [28] and a visual-inertial odometry system which was repeatedly mapping an office space, when following exactly the same trajectory. In total, the robot has executed 31 sorties over a period of 2 months. During each run, an up-to-date localization map was used to register the agent against it. Each new local map was then:

- reduced onboard based on the tracking quality and measures presented in [4], a method particularly suitable for computationally-constrained platforms,
- transmitted to the backend, merged into the global map and summarized using the algorithm from [8], a scalable approach robust to highly-dynamic environments,
- and finally used to build a new compact localization map following the representation from [21], see Figure 2, which is a good fit to our localization system needs.

Our preliminary results prove the approach to be useful in reducing the data flow while keeping satisfactory localization results. The constant-size (in terms of number of landmarks) localization map is quickly improving after each of the first few sessions, eventually yielding stable recall values between 75 and 90%. The exact localization retrieval results are presented in Figure 4. In our evaluation, the agent-side summarization reduces the amount of landmarks by 80%. To achieve comparable recall values assuming we would select the landmarks at random, we would be able to reduce this number only by about 15%. In the backend, after receiving a new local map from the agent, we construct a localization map with number of landmarks equal to about 10% of a single mapping session (6,000 features for a trajectory of 150 m).

In terms of data reduction, the proposed approach reduces the size of the local map from the agent from about 90 MB to less than 25 MB. The localization map that is sent back to the agent to localize further reduces this amount to less

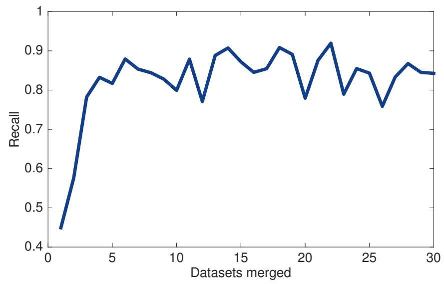

Fig. 4. The recall results of the localization system in an iterative mapping scenario. We have tried to localize each keyframe against the transmitted localization map and compare the result with a ground-truth pose. The threshold for the correct localization was set to 40 cm. The initially improvement of the recall values means that we need a few sessions to build a thorough localization map that will contain persistent elements of the environment. This map size over all the mapping sessions was kept constant and equal to 6,000 features.

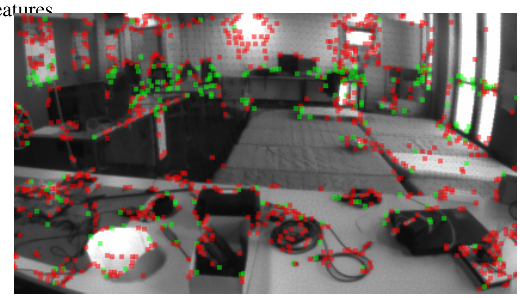

Fig. 5. A preliminary result of the feature selection using convolutional neural networks. The features marked in red are *bad* while the features marked in green are considered valuable for long-term localization. Note that weak features corresponding to *e.g.* the diffuse reflection on the wall are rejected.

than 7 MB. The exact components of the map are presented in Table I.

TABLE I
COMPACT LOCALIZATION MAP DATA SAVINGS

	Single raw map	Localization map
IMU data	4 MB	0
Keyframe data	3 MB	>1 MB
Landmark data	12 MB	1 MB
Descriptor data	78 MB	5 MB

We have additionally investigated a few novel methods for landmark utility prediction. For example, this includes using convolutional neural networks directly on the image patches around the keypoints to classify them into *good* and *bad* subsets. The idea is to adapt the architecture proposed by Krizhevsky *et al.* [17] for ImageNet dataset classification and train it directly on the data collected by our mapping camera. The training data was extracted from multiple mapping sessions in the same area, assuming the *good* features should be detected more often than the *bad* ones. Our initial experiments confirmed that using the image patches directly brings a significant improvement over the Random Forests classification of descriptors as in [11]. An example frame from the CNN-based classifier is presented in the Figure 5.

V. Future Developments

We believe the concept of compact localization maps can see significant improvements, both improving the localization reliability as well as further reducing the data transfer sizes. This progress is of crucial importance for applications, where only low bandwidth connections are available (long range sorties, e.g. for fixed wing planes).

One of the factors that is significantly limiting the current keypoint-based approaches are feature detectors. We have observed in our experiments that the type and direction lighting (direct sunlight or overcast or purely artificial) has a significant influence on the feature redetectability and matching quality. This makes our localization maps larger than could be necessary, as they need to implicitly contain landmarks appropriate for each of the environment conditions, similar to [6] where it was an explicit design choice.

We also believe that using the semantic information, as suggested in [15], can further improve the summarization results. With recent advances in deep learning techniques, semantic information can be retrieved much more reliably and using much more subtle hints. The precise semantic labels can let us immediately reject keypoints that belong to dynamic objects, such as cars in the urban localization scenario.

Furthermore, we see a need for better features, that can get easily matched no matter the viewpoint or lighting conditions. Some significant research was put already into the topic of geometric features, such as lines [37], as they tend to be detected more reliably than corners. But lines, or geometric features in general, still have not made their way into large-scale SLAM robotics systems, mostly due to scalability and computational cost constraints [23].

Another direction of research is looking into the learned features using modern machine learning techniques [36]. The mid-level ConvNet features demonstrate good invariance properties and can be used to extract a topological agent location, before the precise 6DoF posed is retrieved using point or line features.

VI. Conclusions

In this paper, we have discussed the problem of data reduction for localization from maps that were generated by a lifelong mapping system. This problem is of particular importance as data management requirements are often the limiting factor in localization systems, particularly when wireless transmission of map data is required. We discussed emerging and novel directions of research for reducing map data at three levels. First, the choice of the actual map structure. Second, landmark selection based on information from all mapping sessions. Third, client-based landmark quality classification. An example incorporating all three approaches was implemented on an autonomous mapping platform in our office space. We have achieved a significant reduction of the map sizes, reducing the transmission from the agent to the backend by 3 times and from the backend to the agent to only 10% of a single, local map (less than 7 MB). The reduction only marginally affects the precision-recall values of the system, with recall in range between 75 and 90% and 99% precision. We have also shown the system can perform reliably over multiple mapping sessions, gathering *experience* from each consecutive session.

VII. Acknowledgments

The research leading to these results has received funding from Google's project Tango.

References

[1] Relja Arandjelovic and Andrew Zisserman. Extremely low bit-rate nearest neighbor search using a set compression tree. *IEEE Trans. on Pattern Analysis and Machine Intelligence*, 2014.

[2] Maximilian Beinhofer, Jörg Müller, Andreas Krause, and Wolfram Burgard. Robust landmark selection for mobile robot navigation. In *IEEE/RSJ Int. Conf. on Intelligent Robots and Systems*, 2013.

[3] Michael Bosse and Robert Zlot. Keypoint design and evaluation for place recognition in 2d lidar maps. *Robotics and Autonomous Systems*, 2009.

[4] Simone Buoncompagni, Dario Maio, Davide Maltoni, and Serena Papi. Saliency-based keypoint selection for fast object detection and matching. *Pattern Recognition Letters*, 62, 2015.

[5] Nicholas Carlevaris-Bianco, Michael Kaess, and Ryan M Eustice. Generic node removal for factor-graph slam. *IEEE Transactions on Robotics*, 2014.

[6] Winston Churchill and Paul Newman. Experience-based Navigation for Long-term Localisation. *The Int. Journal of Robotics Research*, 2013.

[7] Mark Cummins and Paul Newman. Fab-map: Probabilistic localization and mapping in the space of appearance. *The Int. Journal of Robotics Research*, 2008.

[8] Marcin Dymczyk, Simon Lynen, Michael Bosse, and Roland Siegwart. Keep it brief: Scalable creation of compressed localization maps. In *IEEE/RSJ Int. Conf. on Intelligent Robots and Systems*, 2015.

[9] Marcin Dymczyk, Simon Lynen, Titus Cieslewski, Michael Bosse, Roland Siegwart, and Paul Furgale. The gist of maps-summarizing experience for lifelong localization. In *IEEE Int. Conf. on Robotics and Automation*, 2015.

[10] Christian Forster, Simon Lynen, Laurent Kneip, and Davide Scaramuzza. Collaborative monocular slam with multiple micro aerial vehicles. In *IEEE/RSJ Int. Conf. on Intelligent Robots and Systems*, 2013.

[11] Wilfried Hartmann, Michal Havlena, and Konrad Schindler. Predicting matchability. In *IEEE Conf. on Computer Vision and Pattern Recognition*, 2014.

[12] Viorela Ila, Josep M Porta, and Juan Andrade-Cetto. Information-based compact pose slam. *IEEE Transactions on Robotics*, 2010.

[13] Arnold Irschara, Christopher Zach, Jan-Michael Frahm, and Horst Bischof. From structure-from-motion point clouds to fast location recognition. In *IEEE Conf. on Computer Vision and Pattern Recognition*, 2009.

[14] Hervé Jégou, Matthijs Douze, Cordelia Schmid, and Patrick Pérez. Aggregating local descriptors into a compact image representation. In *IEEE Conf. on Computer Vision and Pattern Recognition*, 2010.

[15] Nikolay Kobyshev, Hayko Riemenschneider, and Luc Van Gool. Matching features correctly through semantic understanding. In *3DV*, 2014.

[16] Kurt Konolige and James Bowman. Towards lifelong visual maps. In *IEEE/RSJ Int. Conf. on Intelligent Robots and Systems*, 2009.

[17] Alex Krizhevsky, Ilya Sutskever, and Geoffrey E Hinton. Imagenet classification with deep convolutional neural networks. In *Advances in neural information processing systems*, 2012.

[18] Stefan Leutenegger, Margarita Chli, and Roland Y Siegwart. BRISK: Binary robust invariant scalable keypoints. In *Int. Conf. on Computer Vision*, 2011.

[19] Stefan Leutenegger, Simon Lynen, Michael Bosse, Roland Siegwart, and Paul Furgale. Keyframe-based visual–inertial odometry using nonlinear optimization. *The Int. Journal of Robotics Research*, 34(3):314–334, 2015.

[20] David G Lowe. Distinctive image features from scale-invariant keypoints. *International journal of computer vision*, 2004.

[21] S Lynen, T Sattler, M Bosse, J Hesch, M Pollefeys, and R Siegwart. Get out of my lab: Large-scale, real-time visual-inertial localization. In *Robotics: Science and Systems*, 2015.

[22] Simon Lynen, Michael Bosse, Paul Furgale, and Roland Siegwart. Placeless place-recognition. In *3DV*, 2014.

[23] Branislav Micusik and Horst Wildenauer. Descriptor free visual indoor localization with line segments. In *IEEE Conf. on Computer Vision and Pattern Recognition*, 2015.

[24] Sven Middelberg, Torsten Sattler, Ole Untzelmann, and Leif Kobbelt. Scalable 6-dof localization on mobile devices. In *European Conf. on Computer Vision*, 2014.

[25] Beipeng Mu, Ali-akbar Agha-mohammadi, Liam Paull, Mathew Graham, Jonathan How, and John Leonard. Two-stage focused inference for resource-constrained collision-free navigation. 2015.

[26] Peter Mühlfellner, Mathias Bürki, Michael Bosse, Wojciech Derendarz, Roland Philippsen, and Paul Furgale. Summary maps for lifelong visual localization. *Journal of Field Robotics*, 2015.

[27] Esha D Nerurkar, Kejian J Wu, and Stergios I Roumeliotis. C-klam: Constrained keyframe-based localization and mapping. In *IEEE Int. Conf. on Robotics and Automation*, 2014.

[28] Janosch Nikolic, Joern Rehder, Michael Burri, Pascal Gohl, Stefan Leutenegger, Paul T Furgale, and Roland Siegwart. A synchronized visual-inertial sensor system with fpga pre-processing for accurate real-time slam. In *IEEE Int. Conf. on Robotics and Automation*, 2014.

[29] Edwin Olson. AprilTag: A robust and flexible visual fiducial system. In *IEEE Int. Conf. on Robotics and Automation*, May 2011.

[30] Hyun Soo Park, Yu Wang, Eriko Nurvitadhi, James C Hoe, Yaser Sheikh, and Mei Chen. 3d point cloud reduction using mixed-integer quadratic programming. In *IEEE Conf. on Computer Vision and Pattern Recognition Workshops*, 2013.

[31] Edward Rosten, Reid Porter, and Tom Drummond. Faster and better: A machine learning approach to corner detection. *Pattern Analysis and Machine Intelligence, IEEE Transactions on*, 2010.

[32] Torsten Sattler, Bastian Leibe, and Leif Kobbelt. Fast image-based localization using direct 2d-to-3d matching. In *Int. Conf. on Computer Vision*, 2011.

[33] Ted J Steiner, Tye M Brady, and Jeffrey A Hoffman. Graph-based terrain relative navigation with optimal landmark database selection. In *Aerospace Conference, 2015 IEEE*, 2015.

[34] Ted J. Steiner, Guoquan Huang, and John J. Leonard. Location utility-based map reduction. In *IEEE Int. Conf. on Robotics and Automation*, 2015.

[35] Elena Stumm, Christopher Mei, and Simon Lacroix. Probabilistic place recognition with covisibility maps. In *IEEE/RSJ Int. Conf. on Intelligent Robots and Systems*, 2013.

[36] Niko Sünderhauf, Sareh Shirazi, Adam Jacobson, Edward Pepperell, Feras Dayoub, Ben Upcroft, and Michael Milford. Place recognition with convnet landmarks: Viewpoint-robust, condition-robust, training-free. In *Robotics: Science and Systems*, 2015.

[37] Bart Verhagen, Radu Timofte, and Luc Van Gool. Scale-invariant line descriptors for wide baseline matching. In *WACV*, 2014.

[38] Aisha Walcott-Bryant, Michael Kaess, Hordur Johannsson, and John J Leonard. Dynamic pose graph slam: Long-term mapping in low dynamic environments. In *IEEE/RSJ Int. Conf. on Intelligent Robots and Systems*, 2012.