

 Open access • Proceedings Article • DOI:10.1109/ICRA.2018.8461155

Data-Efficient Decentralized Visual SLAM — [Source link](#)

[Titus Cieslewski](#), [Siddharth Choudhary](#), [Davide Scaramuzza](#)





Institutions: [University of Zurich](#), [Georgia Institute of Technology](#)

Published on: 21 May 2018 - [International Conference on Robotics and Automation](#)

Topics: [Simultaneous localization and mapping](#), [Pose](#) and [Robot](#)

Related papers:

- [NetVLAD: CNN Architecture for Weakly Supervised Place Recognition](#)
- [Distributed mapping with privacy and communication constraints: Lightweight algorithms and object-based models:](#)
- [ORB-SLAM2: An Open-Source SLAM System for Monocular, Stereo, and RGB-D Cameras](#)
- [Bags of Binary Words for Fast Place Recognition in Image Sequences](#)
- [Collaborative monocular SLAM with multiple Micro Aerial Vehicles](#)

Share this paper:    

View more about this paper here: <https://typeset.io/papers/data-efficient-decentralized-visual-slam-4setbn0ky8>



**University of
Zurich**^{UZH}

**Zurich Open Repository and
Archive**

University of Zurich
University Library
Strickhofstrasse 39
CH-8057 Zurich
www.zora.uzh.ch

Year: 2018

Data-Efficient Decentralized Visual SLAM

Cieslewski, Titus ; Choudhary, Siddharth ; Scaramuzza, Davide

Abstract: Decentralized visual simultaneous localization and mapping (SLAM) is a powerful tool for multi-robot applications in environments where absolute positioning is not available. Being visual, it relies on cheap, lightweight and versatile cameras, and, being decentralized, it does not rely on communication to a central entity. In this work, we integrate state-of-the-art decentralized SLAM components into a new, complete decentralized visual SLAM system. To allow for data association and optimization, existing decentralized visual SLAM systems exchange the full map data among all robots, incurring large data transfers at a complexity that scales quadratically with the robot count. In contrast, our method performs efficient data association in two stages: first, a compact full-image descriptor is deterministically sent to only one robot. Then, only if the first stage succeeded, the data required for relative pose estimation is sent, again to only one robot. Thus, data association scales linearly with the robot count and uses highly compact place representations. For optimization, a state-of-the-art decentralized pose-graph optimization method is used. It exchanges a minimum amount of data which is linear with trajectory overlap. We characterize the resulting system and identify bottlenecks in its components. The system is evaluated on publicly available datasets and we provide open access to the code. A narrated video presentation can be found at: <https://youtu.be/zEBfCA5tVOK>

DOI: <https://doi.org/10.1109/ICRA.2018.8461155>

Posted at the Zurich Open Repository and Archive, University of Zurich

ZORA URL: <https://doi.org/10.5167/uzh-150442>

Conference or Workshop Item

Published Version

Originally published at:

Cieslewski, Titus; Choudhary, Siddharth; Scaramuzza, Davide (2018). Data-Efficient Decentralized Visual SLAM. In: IEEE International Conference on Robotics and Automation (ICRA), 2018., Brisbane, 1 May 2018 - 25 May 2018. IEEE, 1-8.

DOI: <https://doi.org/10.1109/ICRA.2018.8461155>

Data-Efficient Decentralized Visual SLAM

Titus Cieslewski¹, Siddharth Choudhary² and Davide Scaramuzza¹

Abstract—Decentralized visual simultaneous localization and mapping (SLAM) is a powerful tool for multi-robot applications in environments where absolute positioning is not available. Being visual, it relies on cheap, lightweight and versatile cameras, and, being decentralized, it does not rely on communication to a central entity. In this work, we integrate state-of-the-art decentralized SLAM components into a new, complete decentralized visual SLAM system. To allow for data association and optimization, existing decentralized visual SLAM systems exchange the full map data among all robots, incurring large data transfers at a complexity that scales quadratically with the robot count. In contrast, our method performs efficient data association in two stages: first, a compact full-image descriptor is deterministically sent to only one robot. Then, only if the first stage succeeded, the data required for relative pose estimation is sent, again to only one robot. Thus, data association scales linearly with the robot count and uses highly compact place representations. For optimization, a state-of-the-art decentralized pose-graph optimization method is used. It exchanges a minimum amount of data which is linear with trajectory overlap. We characterize the resulting system and identify bottlenecks in its components. The system is evaluated on publicly available datasets and we provide open access to the code.

SUPPLEMENTARY MATERIAL

Data and code are at: https://github.com/uzh-rpg/dslam_open

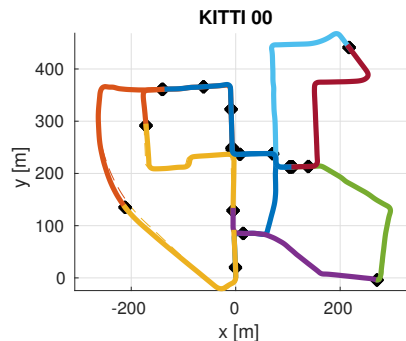
I. INTRODUCTION

Using several robots instead of one can accelerate many tasks such as exploration and mapping, or enable heterogeneous teams of robots, where each robot has a specialization. Multi-robot simultaneous localization and mapping (SLAM) is an essential component of any team of robots operating in an absolute positioning system denied environment, as it relates the current state estimate of each robot to all present and past state estimates of all other robots. Because cameras are cheap, light-weight and versatile sensors, we seek to implement a visual SLAM system. Visual Multi-Robot SLAM can be solved in a centralized manner, where a single entity collects all data and solves SLAM for all robots, but that relies on a central entity to always be reachable, to never fail and to scale to the size of the robot team, both in

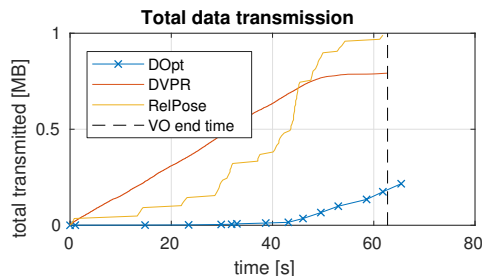
¹T. Cieslewski and D. Scaramuzza are with the Robotics and Perception Group, Dep. of Informatics, University of Zurich, and Dep. of Neuroinformatics, University of Zurich and ETH Zurich, Switzerland—<http://rpg.ifi.uzh.ch>.

²S. Choudhary is with the College of Computing, Georgia Institute of Technology, Atlanta, GA, USA, siddharth.choudhary@gatech.edu

This research was funded by the DARPA FLA Program, the National Center of Competence in Research (NCCR) Robotics through the Swiss National Science Foundation, the SNSF-ERC Starting Grant and the U.S. Army Research Lab (ARL) through the MAST-CTA project 329420.



(a) Ten sub-trajectories of KITTI 00 after running our method. Each color represents an individual trajectory, place matches are marked in black, dashed lines indicate the ground truth.



(b) Data transmission over time for the system components: decentralized optimization (DOpt), decentralized visual place recognition (DVPR) and relative pose estimation (RelPose).

Fig. 1. The proposed decentralized visual SLAM system is able to build a globally consistent map from ten sub-trajectories of KITTI 00 by exchanging only around 2MB in total – for decentralized optimization, place recognition and visual relative pose estimation, for all ten robots.

computation and bandwidth. Decentralized systems do not have this bottleneck, but are more challenging to implement.

Visual SLAM systems typically consist of three components: 1) a visual odometry algorithm which provides an initial state estimate, 2) a place recognition system which is able to relate the currently observed scene to previously observed scenes, and 3) an optimization back-end which consistently integrates the place matches from the place recognition system with the full stream of state estimates. The end product is a map, and that map feeds back to place recognition and optimization. It is this feedback which makes decentralized SLAM challenging, especially if one is concerned about communication bandwidth. Visual odometry is not involved in the feedback loop and is thus trivial to distribute – it can be run on each robot independently and feed its output to the rest of the decentralized SLAM system.

In previous work we have proposed state-of-the-art decentralized place recognition [1] and optimization [2] systems separately. Both systems focus on data efficiency: they

achieve performance similar to a centralized system while minimizing the data that needs to be exchanged. In this work, we integrate both systems along with a data-efficient method for visual feature association [3] into a full decentralized visual SLAM system. The system is evaluated on publicly available datasets and the code is provided open-source.

II. RELATED WORK

In the following, we independently consider related work pertaining to decentralized optimization and to decentralized place recognition. We conclude with a review of existing integrated decentralized SLAM systems.

A. Decentralized Optimization

Decentralized estimation in multi-robot systems is an active field of research, with special attention being paid to communication constraints [4], heterogeneity [5], [6], consistency [7], and robust data association [8]. The literature offers distributed implementations of different estimation techniques, including Kalman filters [9], information filters [10], particle filters [11], [12], and distributed smoothers [2], [13]

In multi-robot systems, maximum-likelihood trajectory estimation can be performed by collecting all measurements at a centralized inference engine, which performs the optimization [5], [8], [14]–[16]. However, it is not practical to collect all measurements at a single inference engine since it requires a large communication bandwidth. Furthermore, solving trajectory estimation over a large team of robots can be too demanding for a single computational unit.

These reasons triggered interest towards *decentralized trajectory estimation*, in which the robots only exploit local communication, in order to reach a consensus on the trajectory estimate [17]–[21]. Recently, Cunningham et al. [13], [21] used Gaussian elimination, and developed an approach, called DDF-SAM, in which robots exchange Gaussian marginals over the *separators* (i.e., the variables observed by multiple robots).

While Gaussian elimination has become a popular approach, it has two major shortcomings. First, the marginals to be exchanged among the robots are dense, hence the communication cost is quadratic in the number of separators. This motivated the use of sparsification techniques [4]. Second, Gaussian elimination is performed on a linearized version of the problem, hence these approaches require good linearization points and complex bookkeeping to ensure consistent linearization across the robots [13]. The need of a linearization point also characterizes gradient-based techniques [20]. An alternative approach to Gaussian elimination is the Gauss-Seidel approach of Choudhary et al. [2], which requires communication linear in the number of separators. The Distributed Gauss-Seidel approach only requires the latest estimates for poses involved in inter-robot measurement to be communicated and therefore does not require consistent linearization and complex bookkeeping. This approach also avoids information double counting since it does not communicate marginals.

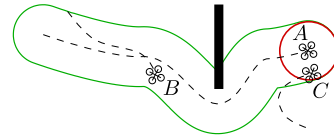


Fig. 2. Restricting relative pose measurements to robots that are in line of sight (red circle) severely limits the recall of relative localization: Robots A and C can establish relative localization but A and B cannot. This often applies to direct relative measurements. By ensuring that relative localization can occur from all past measurements (green outline) we can increase recall and put less restrictions on the high-level application that controls the robots.

B. Decentralized Place Recognition and Pose Estimation

In order to solve decentralized SLAM, measurements that relate poses among different robots (*inter-robot measurements*) need to be established. A popular type of inter-robot measurements are direct measurements of the other robot [22], such as time-of-flight distance measurements [4] or vision-based relative pose estimation [15]. The latter is typically aided with markers that are deployed on the robots. To the best of our knowledge, most types of direct measurements require specialized hardware (which can precisely measure time-of-flight for example, or visual markers). Furthermore, many types of direct measurements require the robots to be in line of sight, which, in many environments, imposes a major limitation on the set of relative poses that can be established, see Fig. 2. Limiting relative measurements in this way translates into limitations for higher-level applications that would like to use decentralized SLAM as a tool.

We prefer, instead, to use indirect relative measurements. These are established through registration of observations that two robots A and B make of the same scene [3], [23]. This requires no additional hardware: the sensors already used for odometry can be re-used for inter-robot measurements. Since establishing inter-robot measurements in this way is equivalent to establishing loop closures in single-robot SLAM, indirect measurements are widely used in centralized visual SLAM algorithms [24]–[26]. The establishment of indirect measurements is still limited by the communication range of the robots, but in practice it is feasible to establish a communication range that exceeds line-of-sight. Furthermore, if two robots cannot communicate directly, they might still communicate through a multi-hop network. A disadvantage of indirect measurements, then, is that they require bandwidth to communicate. If all robots share all data with every other robot, the amount of data exchanged is quadratic in the number of robots. One way to mitigate the bandwidth is to exchange data only with immediate neighbours, but this has the same reduced recall as discussed in Fig. 2. Another way to reduce bandwidth is to use compact representations that allow to establish inter-robot measurements with a minimal amount of data exchange. Visual feature-based maps can be compressed by compressing feature descriptors [27], substituting feature descriptors for corresponding cluster centers in a *visual vocabulary* [3], removing landmarks that prove not to be necessary for localization [28], [29] or using compact full-image descriptors [30]. In previous work, we have shown another way to reduce the bandwidth of place recognition:

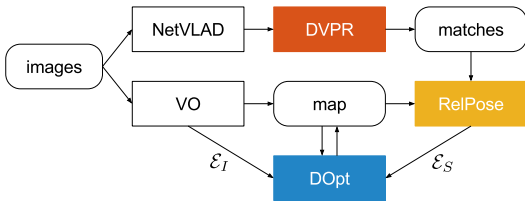


Fig. 3. Components and interactions in our system. The diagram shows the components that run on each robot. The colored blocks indicate components that communicate with other robots. Sharp corners indicate software modules, rounded corners indicate data. *DVPR*: Decentralized visual place recognition, *VO*: visual odometry, *RelPose*: Geometric verification and relative pose estimation, *DOpt*: Decentralized optimization.

data exchange can be reduced by a factor of the robot count if the problem can be cast to key-value lookup [31]. In this work, we use [1], an advanced version of [31] which at the same time makes use of a compact, state-of-the-art place representation, NetVLAD [30]. [3] is used in association of visual features for relative pose estimation.

C. Integrated Decentralized SLAM

Several systems have previously combined decentralized optimization and relative pose estimation into decentralized SLAM. Several optimization works cited in Section II-A, such as [2], [4] use direct relative measurements and can thus be considered full decentralized SLAM systems with the recall limitation illustrated in Fig. 2. DDF-SAM [21] has been extended to a decentralized SLAM system in [32] by adding a data association system which matches planar landmark points. This system has been validated in the real world, but the landmarks consisted of manually placed poles that were detected by a laser range finder. [16] establishes relative poses by exchanging and aligning 2D laser scan edges. [2] has been extended to a decentralized SLAM system in [33] where data association is provided by identification and pose estimation of commonly observed objects. This approach relies on the existence of unique, pre-trained objects in the environment. The majority of decentralized SLAM systems we have found in our literature review are not vision-based. A first decentralized visual SLAM system has been proposed in [34], but it relies on exchanging the full maps among robots. Furthermore, it is only evaluated on $50m \times 50m$ L-shaped trajectory and with only two robots, with images obtained from a simulated environment. In contrast, we evaluate our system on a large scale scenario, with more robots, on real data, and our system uses state-of-the-art algorithms which exchange far less than the full maps.

III. METHODOLOGY

The proposed decentralized SLAM system consists of $|\Omega|$ robots $\Omega = \{\alpha, \beta, \gamma, \dots\}$ which each execute the system illustrated in Fig. 3. The images taken by a camera are fed to both NetVLAD [30] and a visual odometry (VO). NetVLAD produces a compact image representation used for decentralized visual place recognition (DVPR) [1], which itself produces candidate place recognition matches. The VO produces a sparse feature map that can be used for localization. The relative pose estimation module (RelPose)

extracts the parts of the sparse map that are observed at the candidate place matches and uses them to establish relative poses between the robots trajectories, or to reject candidate matches. The decentralized optimization module (DOpt) [2] obtains the initial guess from the map, intra-robot relative pose measurements \mathcal{E}_I from the VO and the inter-robot relative pose measurements \mathcal{E}_S from RelPose, and updates the map. The system works continuously as new images are acquired and assumes that every robot can communicate with every other robot.

A. Intra-Robot Measurements

Intra-robot measurements are obtained from a visual odometry (VO) algorithm. Any VO algorithm can be used in our system as long as it fulfills the following requirements:

1) *It produces a pose graph.* The VO of robot α defines frames $\mathcal{F}_\alpha = \{(\alpha, 1), (\alpha, 2), (\alpha, 3), \dots\}$ and provides the intra-robot measurements

$$\mathcal{E}_{I,\alpha} \doteq \{\bar{z}_{\alpha_{i+1}}^{\alpha_i} \mid \forall (\alpha, i), (\alpha, i+1) \in \mathcal{F}_\alpha\} \quad (1)$$

$$\bar{z}_{\alpha_{i+1}}^{\alpha_i} \doteq (\mathbf{R}_{\alpha_{i+1}}^{\alpha_i}, \bar{\mathbf{t}}_{\alpha_{i+1}}^{\alpha_i}) \in \text{SE}(3) \quad (2)$$

describing pose transforms between subsequent frames.

2) *Each pose is associated to an image to be used for place recognition.* We denote this image by I_{α_i} . Our current system does not yet take full advantage of a multi-camera setup, which could be a useful future extension. The image capture time is denoted as t_{α_i} ¹.

3) *The VO provides absolute scale,* as implied by (2). Monocular VO typically estimates the pose up to a scale due to scale ambiguity. Scale can be obtained from a stereo camera or by fusing with inertial measurements.

In our implementation, we use for visual odometry ORB-SLAM [36] in stereo configuration.

B. Inter-Robot Measurements

The inter-robot measurements are

$$\mathcal{E}_S \doteq \{\bar{z}_{\beta_j}^{\alpha_i} \mid \alpha \neq \beta\} \quad (3)$$

with $\bar{z}_{\beta_j}^{\alpha_i}$ defined like (2). They are established in two phases: Given its latest image I_{α_i} , robot α first tries to determine whether there is any image I_{β_j} of the same scene previously captured by another robot. This is achieved using decentralized visual place recognition, detailed in Section III-C. Then, if $\exists I_{\beta_j}$, α tries to establish $\bar{z}_{\beta_j}^{\alpha_i}$ using the method detailed in Section III-D. This two-phase approach is pursued because, as we will see, establishing relative poses consumes more data than recognizing whether $\exists I_{\beta_j}$. The first phase is capable of reducing the amount of data to be exchanged by determining which robot to send the relative pose query to, and by determining whether it is worthwhile to send it in the first place, using a minimal amount of data.

¹A common time reference is hard to establish, and, strictly speaking, not well defined, so consider t_{α_i} a Lamport timestamp [35]. t_{α_i} is not actually used in the implementation, so this detail is not particularly important.

C. Decentralized Visual Place Recognition

The first phase is accomplished by matching image I_{α_i} to the images previously captured by the other robots, $\{I_{\gamma_k} | t_{\gamma_k} < t_{\alpha_i}\}$, using the full-image descriptor NetVLAD [30]. NetVLAD uses a neural network to calculate a vector $\mathbf{v}_{\alpha_i} \in \mathbb{R}^{D_{\text{NetVLAD}}}$ from I_{α_i} . In rough terms, it is trained such that $\|\mathbf{v}_{\alpha_i} - \mathbf{v}_{\gamma_k}\|_{\ell_2}$ is lower if I_{α_i} and I_{γ_k} observe the same scene than if they observe different scenes. For more details, see [30]. NetVLAD is well-suited for decentralized place recognition, as \mathbf{v}_{α_i} is all that robot α needs to send to robot γ to establish whether a matching image exists among $\{I_{\gamma_k}\}$. Concretely, we seek the image $I_{\beta_j}^*$ captured by another robot whose NetVLAD vector $\mathbf{v}_{\beta_j}^*$ has the shortest ℓ_2 distance to \mathbf{v}_{α_i} and which satisfies

$$\|\mathbf{v}_{\alpha_i} - \mathbf{v}_{\beta_j}\|_{\ell_2} < \tau_{\text{NetVLAD}} \quad (4)$$

where τ_{NetVLAD} is a threshold parameter. $I_{\beta_j}^*$ can be found by sending \mathbf{v}_{α_i} to all other robots, but we use a method that requires $|\Omega|$ times less data exchange [1], which is particularly interesting for scaling to large robot teams. In this method, each robot γ is pre-assigned a cluster center \mathbf{c}_γ and each robot knows the \mathbf{c} of all other robots. The query \mathbf{v}_{α_i} is sent only to robot $\delta = \arg \min_{\gamma} \|\mathbf{v}_{\alpha_i} - \mathbf{c}_\gamma\|_{\ell_2}$, and δ replies with the identifier of I_{β_j} ,

$$(\beta, j) = \arg \min_{(\gamma, k): \mathbf{v}_{\gamma_k} \in \mathcal{V}_\delta} \|\mathbf{v}_{\alpha_i} - \mathbf{v}_{\gamma_k}\|_{\ell_2}, \quad (5)$$

if \mathbf{v}_{β_j} also satisfies (4), where \mathcal{V}_δ is the Voronoi cell $\mathcal{V}_\delta = \{\mathbf{v} | \mathbf{c}_\delta = \arg \min_{\mathbf{c}_\gamma} \|\mathbf{v} - \mathbf{c}_\gamma\|_{\ell_2}\}$. The query is executed for every frame of every robot, so at t_{α_i} this query has already been executed $\forall \gamma, k: t_{\gamma_k} < t_{\alpha_i}$, and δ is aware of all $\mathbf{v}_{\gamma_k} \in \mathcal{V}_\delta$ by virtue of having received the previous queries. It can thus provide (β, j) without any extra data exchange. This method approximates $I_{\beta_j}^* \sim I_{\beta_j}$; $I_{\beta_j}^* = I_{\beta_j}$ if $\exists \delta: \mathbf{v}_{\beta_j}^*, \mathbf{v}_{\alpha_i} \in \mathcal{V}_\delta$. In the method, the cluster centers are determined using k-means clustering on a training dataset, in our case the Oxford Robotcar Dataset [37]. The properties of the method are discussed in more detail in [1].

D. Relative Pose Estimation

Once and if (β, j) has been established, α sends a relative pose estimation request to β . The relative pose estimation (RelPose) can and should heavily depend on the visual odometry (VO) since it can benefit from re-using by-products of the VO such as 3D positions of landmarks. In our method, we use ORB-SLAM [36] for VO and accordingly, our RelPose implementation imitates its loop closure method. However, it does not imitate [36] exactly, since the latter makes use of the data surrounding the match, which in decentralized RelPose would translate into a lot of data exchange. We want to avoid this and thus, we combine the relative pose estimation method from [36] with the visual data association method from [3], which reduces the data to be sent from α to β , $d_{\text{RelPose}}^{\alpha \rightarrow \beta}$ to a set of landmarks and visual word identifiers:

$$d_{\text{RelPose}}^{\alpha \rightarrow \beta} = (\{(w_k, \mathbf{p}_k) \forall k \in K_{\alpha_i}\}, \alpha, i, j) \quad (6)$$

where K_{α_i} are the keypoints observed in I_{α_i} , w_k is the identifier of a visual word associated to keypoint k and \mathbf{p}_k is the 3D position of the landmark corresponding to keypoint k , expressed in the camera frame of I_{α_i} . Visual words represent Voronoi cells in the keypoint descriptor space, here the space of ORB descriptors [38]. They provide a quantization of the high-dimensional descriptor space, and since similar descriptors are likely to be assigned to the same word, word identifiers can be used for matching keypoints between images [3]. Since this is less precise than using raw descriptors, we provide an option to exchange full descriptors instead. Given $d^{\alpha \rightarrow \beta}$, β establishes pairs of matching keypoints $\{(k, k') | w_k = w_{k'}, \nexists l \in K_{\alpha_i} : w_l = w_k, \nexists l' \in K_{\beta_j} : w_{l'} = w_{k'}\}$. Given the corresponding pairs of landmark positions $\{(\mathbf{p}_k, \mathbf{p}_{k'})\}$, RANSAC is used to determine an initial estimate of $\bar{\mathbf{z}}_{\beta_j}^{\alpha_i}$ and the corresponding inlier matches $M = \{(k, k')^*\}$. A place match is rejected if $|M| < \tau_{\text{inliers}}$. Otherwise, the inlier landmark position pairs $\{(\mathbf{p}_k, \mathbf{p}_{k'})^*\}$ are used to refine $\bar{\mathbf{z}}_{\beta_j}^{\alpha_i}$ to the pose that minimizes a robust 3D registration error

$$\sum_{\{(\mathbf{p}_k, \mathbf{p}_{k'})^*\}} \rho_{\tau_{\text{loss}}}(\|\mathbf{p}_k - (\bar{\mathbf{R}} \cdot \mathbf{p}_{k'} + \bar{\mathbf{t}})\|_{\ell_2}^2). \quad (7)$$

We use a robust weight loss function $\rho_{\tau_{\text{loss}}}(s) = \arctan \frac{s}{\tau_{\text{loss}}}$ to reduce the weight of remaining keypoint match outliers.

To avoid relative pose $\bar{\mathbf{z}}_{\beta_j}^{\alpha_i}$ outliers we use a consistency check, where $\bar{\mathbf{z}}_{\beta_j}^{\alpha_i}$ is only accepted if either: *a)* there is another already accepted $\bar{\mathbf{z}}_{\beta_{j'}}^{\alpha_{i'}}$ such that (α, i') and (α, i) are within a distance of τ_{cdist} and $\bar{\mathbf{z}}_{\beta_j}^{\alpha_i}$ and $\bar{\mathbf{z}}_{\beta_{j'}}^{\alpha_{i'}}$ are consistent; *b)* no accepted $\bar{\mathbf{z}}_{\beta_{j'}}^{\alpha_{i'}}$ exists, but a previous candidate which fulfills the same conditions. $\bar{\mathbf{z}}_{\beta_j}^{\alpha_i}$ and $\bar{\mathbf{z}}_{\beta_{j'}}^{\alpha_{i'}}$ are considered consistent if the two ways of calculating the relative position between (α, i') and (β, j) – with $\bar{\mathbf{z}}_{\beta_{j'}}^{\alpha_{i'}}$ and $\mathcal{E}_{I, \beta}$ or with $\mathcal{E}_{I, \alpha}$ and $\bar{\mathbf{z}}_{\beta_j}^{\alpha_i}$ – result in positions that are within a distance $< \tau_{\text{tol}}$.

RelPose still constitutes a significant amount of data, so we introduce a parameter τ_{mdg} (minimum distance geometric verifications) which allows to throttle geometric verifications. Geometric verification are skipped for matches which are close to already established relative poses. $d^{\alpha \rightarrow \beta}$ is only sent for geometric verification to β if there is no other (α, i') matched to a frame of β such that $\|\mathbf{t}_{\alpha_i} - \mathbf{t}_{\alpha_{i'}}\|_{\ell_2} < \tau_{\text{mdg}}$. This assumes that the VO and RelPose are consistent enough such that the resulting subset of retrievable \mathcal{E}_S suffices to produce a globally-consistent state estimate.

E. Decentralized Optimization

In our decentralized optimization, each robot estimates its own trajectory using the available measurements, and leveraging occasional communication with other robots. The pose of frame (α, i) is denoted with \mathbf{x}_{α_i} . The trajectory of robot α is then denoted as $\mathbf{x}_\alpha = [\mathbf{x}_{\alpha_1}, \mathbf{x}_{\alpha_2}, \dots]$, where $\mathbf{x}_{\alpha_i} = (\mathbf{R}_{\alpha_i}, \mathbf{t}_{\alpha_i}) \in \text{SE}(3)$, $\mathbf{R}_{\alpha_i} \in \text{SO}(3)$ represents the 3D rotation, and $\mathbf{t}_{\alpha_i} \in \mathbb{R}^3$ represents the 3D translation.

While our classification of the measurements (inter vs intra) is based on the robots involved in the measurement process, all relative measurements can be framed within the

same measurement model. Since all measurements correspond to noisy observations of the relative pose between a pair of poses, say \mathbf{x}_{α_i} and \mathbf{x}_{β_j} , a general measurement model is:

$$\bar{\mathbf{z}}_{\beta_j}^{\alpha_i} \doteq (\bar{\mathbf{R}}_{\beta_j}^{\alpha_i}, \bar{\mathbf{t}}_{\beta_j}^{\alpha_i}), \quad \text{with: } \begin{cases} \bar{\mathbf{R}}_{\beta_j}^{\alpha_i} = (\mathbf{R}_{\alpha_i})^\top \mathbf{R}_{\beta_j} \mathbf{R}_\epsilon \\ \bar{\mathbf{t}}_{\beta_j}^{\alpha_i} = (\mathbf{R}_{\alpha_i})^\top (\mathbf{t}_{\beta_j} - \mathbf{t}_{\alpha_i}) + \mathbf{t}_\epsilon \end{cases} \quad (8)$$

where the relative pose measurement $\bar{\mathbf{z}}_{\beta_j}^{\alpha_i}$ includes the relative rotation measurements $\bar{\mathbf{R}}_{\beta_j}^{\alpha_i}$, which describes the attitude \mathbf{R}_{β_j} with respect to the reference frame (α, i) , “plus” a random rotation \mathbf{R}_ϵ (measurement noise), and the relative position measurement $\bar{\mathbf{t}}_{\beta_j}^{\alpha_i}$, which describes the position \mathbf{t}_{β_j} in the reference frame (α, i) , plus random noise \mathbf{t}_ϵ .

Assuming that the translation noise is distributed according to a zero-mean Gaussian with information matrix $\omega_t^2 \mathbf{I}_3$, while the rotation noise follows a Von-Mises distribution with concentration parameter ω_R^2 , it is possible to demonstrate [39] that the ML estimate $\mathbf{x} \doteq \{(\mathbf{R}_{\alpha_i}, \mathbf{t}_{\alpha_i}), \forall \alpha \in \Omega, \forall i\}$ can be computed as solution of the following optimization problem:

$$\min_{\substack{\mathbf{t}_{\alpha_i} \in \mathbb{R}^3, \mathbf{R}_{\alpha_i} \in \text{SO}(3) \\ \forall \alpha \in \Omega, \forall i}} \sum_{(\alpha_i, \beta_j) \in \mathcal{E}} \omega_t^2 \left\| \mathbf{t}_{\beta_j} - \mathbf{t}_{\alpha_i} - \mathbf{R}_{\alpha_i} \bar{\mathbf{t}}_{\beta_j}^{\alpha_i} \right\|^2 + \frac{\omega_R^2}{2} \left\| \mathbf{R}_{\beta_j} - \mathbf{R}_{\alpha_i} \bar{\mathbf{R}}_{\beta_j}^{\alpha_i} \right\|_{\mathbb{F}}^2 \quad (9)$$

As proposed in [33], we use a two-stage approach to solve the optimization problem in a decentralized manner: we first solve a relaxed version of (9) and get an estimate for the rotations \mathbf{R}_{α_i} of all robots, and then we recover the full poses and top-off the result with a Gauss-Newton (GN) iteration.

In both stages, we need to solve a linear system where the unknown vector can be partitioned into subvectors, such that each subvector contains the variables associated to a single robot in the team. For instance, we can partition the vector \mathbf{r} in the first stage, as $\mathbf{r} = [\mathbf{r}_\alpha, \mathbf{r}_\beta, \dots]$, such that \mathbf{r}_α describes the rotations of robot α . Similarly, we can partition $\mathbf{p} = [\mathbf{p}_\alpha, \mathbf{p}_\beta, \dots]$ (\mathbf{p} represents the linearized poses) in the second stage, such that \mathbf{p}_α describes the trajectory of robot α . Therefore, the linear systems in the first two stages can be framed in the general form $\mathbf{H}\mathbf{y} = \mathbf{g}$:

$$\begin{bmatrix} \mathbf{H}_{\alpha\alpha} & \mathbf{H}_{\alpha\beta} & \dots \\ \mathbf{H}_{\beta\alpha} & \mathbf{H}_{\beta\beta} & \dots \\ \vdots & \vdots & \ddots \end{bmatrix} \begin{bmatrix} \mathbf{y}_\alpha \\ \mathbf{y}_\beta \\ \vdots \end{bmatrix} = \begin{bmatrix} \mathbf{g}_\alpha \\ \mathbf{g}_\beta \\ \vdots \end{bmatrix} \quad (10)$$

where we want to compute the vector $\mathbf{y} = [\mathbf{y}_\alpha, \mathbf{y}_\beta, \dots]$ given the (known) block matrix \mathbf{H} and the (known) block vector \mathbf{g} , partitioned according to the block-structure of \mathbf{y} .

The linear system (10) can be rewritten as:

$$\sum_{\delta \in \Omega} \mathbf{H}_{\alpha\delta} \mathbf{y}_\delta = \mathbf{g}_\alpha \quad \forall \alpha \in \Omega$$

Taking the contribution of \mathbf{y}_α out of the sum, we get:

$$\mathbf{H}_{\alpha\alpha} \mathbf{y}_\alpha = - \sum_{\delta \in \Omega \setminus \{\alpha\}} \mathbf{H}_{\alpha\delta} \mathbf{y}_\delta + \mathbf{g}_\alpha \quad \forall \alpha \in \Omega \quad (11)$$

The set of equations (11) is the same as the original system (10), but clearly exposes the contribution of the variables associated to each robot.

The distributed Gauss-Seidel algorithm [40] starts at an arbitrary initial estimate $\mathbf{y}^{(0)} = [\mathbf{y}_\alpha^{(0)}, \mathbf{y}_\beta^{(0)}, \dots]$ and applies the following update rule, for each robot $\alpha \in \Omega$:

$$\mathbf{y}_\alpha^{(k+1)} = \mathbf{H}_{\alpha\alpha}^{-1} \left(- \sum_{\delta \in \Omega_\alpha^+} \mathbf{H}_{\alpha\delta} \mathbf{y}_\delta^{(k+1)} - \sum_{\delta \in \Omega_\alpha^-} \mathbf{H}_{\alpha\delta} \mathbf{y}_\delta^{(k)} + \mathbf{g}_\alpha \right) \quad (12)$$

where Ω_α^+ is the set of robots that already computed the $(k+1)$ -th estimate, while Ω_α^- is the set of robots that still have to perform the update (12), excluding node α (intuitively: each robot uses the latest estimate). We can see that if the sequence produced by the iterations (12) converges to a fixed point, then such point satisfies (11), and indeed solves the original linear system (10).

F. Making Optimization Work Continuously

Since the map estimate handled by DOpt is not a consistent map state between iterations, it cannot be considered as best estimate at any time. Furthermore, the optimization is not laid out to incorporate new measurements between iterations. As a consequence, the map estimate is concurrently modified by the VO and DOpt, which would lead to inconsistencies if unchecked. This can be solved using optimistic concurrency control (OCC). In OCC, each of the processes that concurrently operate on some data operate on their own copy. Once a process is done, its copy is merged back to the reference state. OCC is a central concept in version control tools such as SVN and GIT, and has recently been applied to distributed robotic systems in [41] and [42]. In our method, the handling of VO and place recognition is considered one process and decentralized optimization another. We let VO and place recognition operate directly on the reference state. Decentralized optimization, as described in Section III-E is executed episodically. At the beginning of each episode, the robots consent on a reference time t_e . During the episode, only the poses \mathbf{x}_{α_i} which satisfy $t_{\alpha_i} < t_e$ are optimized. Once the optimization has finished, the \mathbf{x}_{α_i} of the reference state are replaced with the newly optimized estimates \mathbf{x}'_{α_i} . The remaining \mathbf{x}_{α_j} for which $t_{\alpha_j} > t_e$ are corrected with:

$$\mathbf{x}_{\alpha_j} \leftarrow \mathbf{R} \cdot \mathbf{x}_{\alpha_j} + \mathbf{t} \quad (13)$$

which satisfies $\mathbf{x}'_{\alpha_e} = \mathbf{R} \cdot \mathbf{x}_{\alpha_e} + \mathbf{t}$ with (α, e) referring to the frame recorded just before t_e .

IV. EXPERIMENTS

In the experiments, we seek to find out how much data the individual components and the overall system use and how accurate the state estimate is as time progresses. This will give us an idea of the applicability and scalability of the system, and will let us identify which components are most worthwhile to be optimized in the future.

	Sec.	value		Sec.	value
D_{NetVLAD}	III-C	128	τ_{cdist}	III-D	20m
τ_{NetVLAD}	III-C	0.1	τ_{tot}	III-D	4m
τ_{inliers}	III-D	20	τ_{mdg}	III-D	0m
τ_{loss}	III-D	3m			

TABLE I
DEFAULT PARAMETER VALUES

A. Data Exchange Evaluation

The components of our system that exchange data are decentralized visual place recognition, relative pose estimation and decentralized optimization. Each of these components could potentially be further developed to use less data. Thus, we record data transfer for each component individually. Data transfers for the individual components are:

$$\bigcup_{\text{episode}} d_{\text{DOpt}}^{\alpha \rightarrow \beta} = \left\{ \bigcup_{\text{rot. iters.}} \mathbf{R}_{\beta_j}^{\alpha_i}, \bigcup_{\text{pose iters.}} \bar{\mathbf{z}}_{\beta_j}^{\alpha_i} \right\} \quad (14)$$

$$d_{\text{DVPR}}^{\alpha \rightarrow \delta} = (\alpha, i, \mathbf{v}_{\alpha_i}), \quad d_{\text{DVPR}}^{\delta \rightarrow \alpha} = (\beta, j) \quad (15)$$

$$d_{\text{RelPose}}^{\beta \rightarrow \alpha} = (\bar{\mathbf{z}}_{\beta_j}^{\alpha_i}, \mathbf{x}_{\beta_j}^{-1} \mathbf{x}_{\beta_j}) \quad (16)$$

with $d_{\text{RelPose}}^{\alpha \rightarrow \beta}$ from (6) and where $\mathbf{x}_{\beta_j}^{-1} \mathbf{x}_{\beta_j}$ is sent for the relative pose consistency check. The size of data is calculated accordingly, using $6 \cdot 8$ bytes for poses $\bar{\mathbf{z}}_{\beta_j}^{\alpha_i}$, $9 \cdot 8$ bytes for rotations $\mathbf{R}_{\beta_j}^{\alpha_i}$, $8 \cdot D_{\text{NetVLAD}}$ bytes for NetVLAD vectors, 1 byte for robot indices, 4 bytes for frame indices, 2 bytes for visual word indices w_k and $3 \cdot 4$ bytes for landmarks \mathbf{p}_k . Note that for the first part of the optimization, rotations are parametrized with 9 coefficients, because $\mathbf{R}_{\beta_j}^{\alpha_i} \in SO(3)$ is not enforced in order to have a linear system.

B. Accuracy Evaluation

As it is typically done for SLAM systems, we evaluate the accuracy of our system. Unlike of other SLAM systems, however, we evaluate the evolution of accuracy over time. This allows us to better characterize the system. Accuracy is measured using the average trajectory error (ATE). The ATE requires alignment to the ground truth trajectory, which is only meaningful for connected components. Thus, we report the ATE for different connected components individually.

C. Parameter Studies

A parameter that we find of particular interest is τ_{mdg} , which determines the distance between an established relative pose $\bar{\mathbf{z}}_{\beta_j}^{\alpha_i}$ and the next frame of α to be sent to β for geometrical verification. τ_{mdg} can significantly reduce the data to be transmitted for RelPose and DOpt. We verify that this does not happen at the cost of accuracy.

Another interesting parameter is the used NetVLAD dimension D_{NetVLAD} . Since the last layer of NetVLAD performs principal component analysis, D_{NetVLAD} can be tuned arbitrarily up to 4096 dimensions. The lower D_{NetVLAD} , the lower the DVPR traffic, but also the lower its precision.

V. RESULTS

We evaluate our system on the KITTI dataset [43]. Like in [31], we split the dataset into $|\Omega| = 10$ parts to simulate trajectories recorded by different robots. Fig. 4 shows the

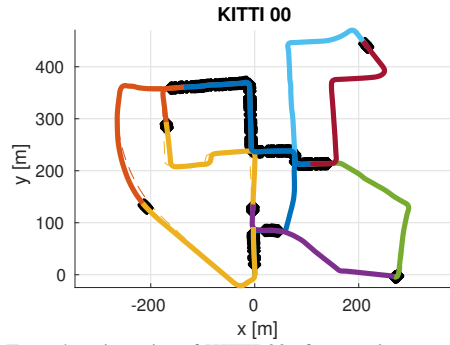


Fig. 4. Ten sub-trajectories of KITTI 00 after running our method with the parameters in Table I. Each color represents an individual robots trajectory, place matches are marked with bold black dots. The aligned ground truth is indicated with dashed lines.

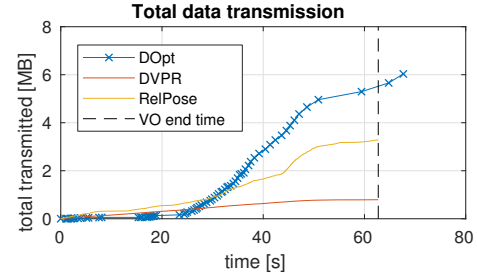


Fig. 5. Data transmission over time for the three system components: decentralized optimization (DOpt), decentralized visual place recognition (DVPR) and relative pose estimation (RelPose), using the parameters in Table I. DOpt continues to run after visual odometry (VO) has ended until it has incorporated all data produced by VO.

used sub-trajectories of KITTI 00 and the end result of running our system with the parameter values shown in Table I and using the visual data association method from [3]. The corresponding data transmission is shown in Fig. 5. Note that we plot the cumulative data transmission up to a given time. This directly exhibits the full amount of transmitted data and makes it easy to see time-varying bandwidth (slope). DOpt and RelPose require the most data transmission, but as we will see, DOpt traffic can be significantly reduced, while RelPose traffic remains high.

Fig. 6 shows the total data transmission among pairs of robots. This data transmission is highly uneven and could require consideration in multi-hop networks, but that is outside of the scope of this paper. DOpt and RelPose highly depend on the trajectory overlap between robots. The higher the overlap, the more data is exchanged by these components. Decentralized visual place recognition traffic is also uneven, but in a different way. It is due to the clustering that assigns NetVLAD cluster centers to robots. What we see in Fig. 6

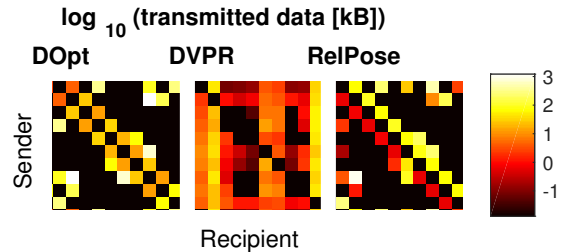


Fig. 6. Data exchanged among individual pairs of robots. Row numbers indicate sender and column numbers receiver. DOpt, DVPR and RelPose traffic is visualized separately and separated by green lines.

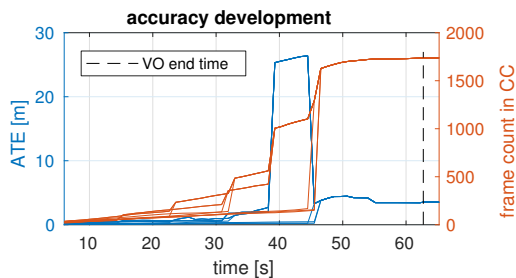


Fig. 7. Average trajectory error (ATE) and size of the connected component (CC) each of the ten robots is in. As the CCs of two robots converge, so does the line representing the frame count and ATE within these CC. Note that accuracy does not change significantly in the optimizations executed after visual odometry (VO) has ended – the state estimate reaches its final accuracy already before all the data is considered for optimization.

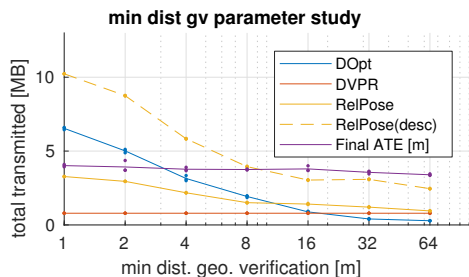


Fig. 8. Effect of changing τ_{mdg} , the parameter that controls the distance between an established relative pose $\tilde{z}_{\beta_j}^{\alpha_i}$ and the next frame of α to be sent to β for geometrical verification. As we can see, τ_{mdg} can be used to significantly throttle the required bandwidth. This happens without a significant effect on the ATE. The dashed line indicates RelPose data transmission if descriptors are used for keypoint association rather than visual word indices.

is that robots 2, 6 and 7 are assigned the v that represent the majority of NetVLAD descriptors in KITTI 00. This unevenness can be mitigated either by randomly assigning several clusters per robot or by training the clustering on a dataset that is very similar to the data at hand, in terms of the scope of visual appearances of keypoints. A detailed study of this problem is provided in [1].

Fig. 7 illustrates how the trajectories of the robots fuse as time progresses, and the accuracy of the resulting connected components. Seven merge events of connected components are apparent around seconds 15, 23, 32, 33, 39, 46 and 47, the remaining two merges occur at the beginning of the experiment and are not visible. At each merge event, there is a jump in connected component size and the corresponding lines in the plot fuse. In early merge events, the ATE grows as the drift of individual components is compounded with noise in relative pose measurements. This is most pronounced at the merge event around second 39. Later, as more relative measurements are added, loops are closed and the ATE decreases again, until it stabilizes around a value of 4 meters, well below 1% of the overall trajectory length.

Fig. 8 shows how increasing the minimum distance between geometric verifications τ_{mdg} can significantly reduce bandwidth requirements without negatively affecting accuracy on KITTI. Fig. 1a shows the place matches that remain and Fig. 1b data transmission over time with $\tau_{\text{mdg}} = 60m$.

Fig. 9 shows how the NetVLAD dimension affects the

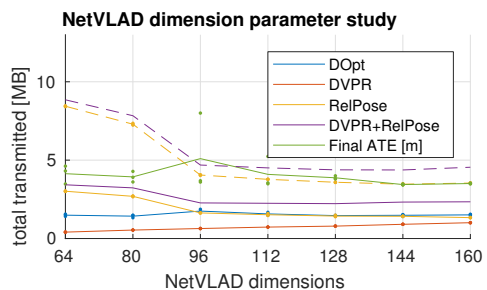


Fig. 9. Data transmission and ATE as the NetVLAD dimensionality is varied, with $\tau_{\text{mdg}} = 10$. Dashed lines indicate RelPose data transmission if keypoints are matched with descriptors, rather than words.

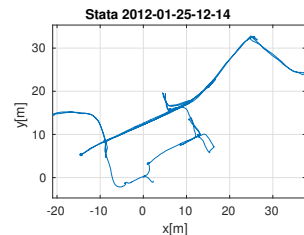


Fig. 10. Groundtruth trajectory for one of the sequences in the MIT Stata Center Dataset [44].

size of transmitted data, with $\tau_{\text{mdg}} = 10m$. Unsurprisingly, DVPR traffic is quasi-linear with respect to the NetVLAD dimension. However, using a too low dimension for NetVLAD leads to more false positives, which then have to be filtered out by the relative pose estimation. This results in a significant increase of RelPose data transmission. Conversely, increasing NetVLAD dimensions beyond a certain point does not increase precision significantly. In Fig. 9 we can see that an efficient trade-off is reached with $D_{\text{NetVLAD}} > 100$.

We also evaluated our system on one of the sequences in MIT Stata Center Dataset [44]. Fig. 10 shows the groundtruth trajectory. Similar to the KITTI 00 dataset, we split the sequence into $|\Omega| = 10$ parts to simulate trajectories recorded by different robots. Fig. 11 shows the sub-trajectories estimated using our method before and after the final loop closure. As we can see, the final estimate given by our method is not visually similar to the groundtruth estimate (Fig. 10). The main reason for this mismatch is that the graph structure in the case of MIT Stata Center dataset is closer to a chain topology. Due to the chain topology, decentralized optimization requires more iterations to reach consensus among robot trajectories on either ends of the chain and therefore finds it difficult to converge to the centralized estimate. Finding sub-trajectories which follow certain graph topologies (chain vs grid) and are favorable for decentralized optimization can be a useful future extension.

VI. CONCLUSION

We have presented a new integrated decentralized visual SLAM algorithm that is based on state-of-the-art components. The system has been characterized using publicly available datasets and we have explored how data transmission can be reduced to a minimum. Based on our results, we believe that future developments of decentralized visual SLAM will focus on one hand on even more data-efficient

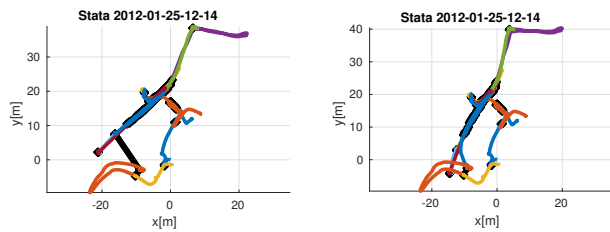


Fig. 11. Ten sub-trajectories of the MIT Stata Center Dataset *before* and *after the final loop closure*, estimated using our method. Each color represents an individual robots trajectory, place matches are marked with bold black lines.

data association and on the other hand on more robust decentralized optimization.

REFERENCES

- [1] T. Cieslewski and D. Scaramuzza, "Efficient decentralized visual place recognition from full-image descriptors," *IEEE International Symposium on Multi-Robot and Multi-Agent Systems*, 2017.
- [2] S. Choudhary, L. Carlone, C. Nieto, J. Rogers, H. I. Christensen, and F. Dellaert, "Distributed trajectory estimation with privacy and communication constraints: a two-stage distributed gauss-seidel approach," in *IEEE Int. Conf. on Robotics and Automation*, 2016.
- [3] D. Tardioli, E. Montijano, and A. R. Mosteo, "Visual data association in narrow-bandwidth networks," in *Intelligent Robots and Systems (IROS), 2015 IEEE/RSJ International Conference on*, 2015.
- [4] L. Paull, G. Huang, M. Seto, and J. Leonard, "Communication-constrained multi-AUV cooperative SLAM," in *IEEE Intl. Conf. on Robotics and Automation (ICRA)*, 2015.
- [5] T. Bailey, M. Bryson, H. Mu, J. Vial, L. McCalman, and H. Durrant-Whyte, "Decentralised cooperative localisation for heterogeneous teams of mobile robots," in *IEEE Intl. Conf. on Robotics and Automation (ICRA)*, 2011.
- [6] V. Indelman, P. Gurfil, E. Rivlin, and H. Rotstein, "Graph-based distributed cooperative navigation for a general multi-robot measurement model," *Intl. J. of Robotics Research*, vol. 31, no. 9, August 2012.
- [7] A. Bahr, M. Walter, and J. Leonard, "Consistent cooperative localization," in *IEEE Intl. Conf. on Robotics and Automation (ICRA)*, 2009.
- [8] J. Dong, E. Nelson, V. Indelman, N. Michael, and F. Dellaert, "Distributed real-time cooperative localization and mapping using an uncertainty-aware expectation maximization approach," in *IEEE Intl. Conf. on Robotics and Automation (ICRA)*, 2015.
- [9] S. Roumeliotis and G. Bekey, "Distributed multi-robot localization," *IEEE Trans. Robot. Automat.*, August 2002.
- [10] S. Thrun and Y. Liu, "Multi-robot SLAM with sparse extended information filters," in *Proceedings of the 11th International Symposium of Robotics Research (ISRR'03)*. Sienna, Italy: Springer, 2003.
- [11] A. Howard, "Multi-robot simultaneous localization and mapping using particle filters," *Intl. J. of Robotics Research*, 2006.
- [12] L. Carlone, M. K. Ng, J. Du, B. Bona, and M. Indri, "Simultaneous localization and mapping using Rao-Blackwellized particle filters in multi robot systems," *J. of Intelligent and Robotic Systems*, 2011.
- [13] A. Cunningham, V. Indelman, and F. Dellaert, "DDF-SAM 2.0: Consistent distributed smoothing and mapping," in *IEEE Intl. Conf. on Robotics and Automation (ICRA)*, Karlsruhe, Germany, May 2013.
- [14] L. Andersson and J. Nygard, "C-SAM : Multi-robot SLAM using square root information smoothing," in *IEEE Intl. Conf. on Robotics and Automation (ICRA)*, 2008.
- [15] B. Kim, M. Kaess, L. Fletcher, J. Leonard, A. Bachrach, N. Roy, and S. Teller, "Multiple relative pose graphs for robust cooperative mapping," in *IEEE Intl. Conf. on Robotics and Automation (ICRA)*, 2010.
- [16] M. Lazaro, L. Paz, P. Pinies, J. Castellanos, and G. Grisetti, "Multi-robot SLAM using condensed measurements," in *IEEE/RSJ Intl. Conf. on Intelligent Robots and Systems (IROS)*, 2013, pp. 1069–1076.
- [17] E. Nerurkar, S. Roumeliotis, and A. Martinelli, "Distributed maximum a posteriori estimation for multi-robot cooperative localization," in *IEEE Intl. Conf. on Robotics and Automation (ICRA)*, 2009.
- [18] M. Franceschelli and A. Gasparri, "On agreement problems with Gossip algorithms in absence of common reference frames," in *IEEE Intl. Conf. on Robotics and Automation (ICRA)*, 2010.
- [19] R. Aragues, L. Carlone, G. Calafiore, and C. Sagues, "Multi-agent localization from noisy relative pose measurements," in *IEEE Intl. Conf. on Robotics and Automation (ICRA)*, 2011, pp. 364–369.
- [20] J. Knuth and P. Barooah, "Collaborative localization with heterogeneous inter-robot measurements by Riemannian optimization," in *IEEE Intl. Conf. on Robotics and Automation (ICRA)*, 2013, pp. 1534–1539.
- [21] A. Cunningham, M. Paluri, and F. Dellaert, "DDF-SAM: Fully distributed slam using constrained factor graphs," in *IEEE/RSJ Intl. Conf. on Intelligent Robots and Systems (IROS)*, 2010.
- [22] X. Zhou and S. Roumeliotis, "Multi-robot SLAM with unknown initial correspondence: The robot rendezvous case," in *IEEE/RSJ Intl. Conf. on Intelligent Robots and Systems (IROS)*, 2006, pp. 1785–1792.
- [23] R. Aragues, E. Montijano, and C. Sagues, "Consistent data association in multi-robot systems with limited communications," in *Robotics: Science and Systems*, 2011, pp. 97–104.
- [24] C. Forster, S. Lynen, L. Kneip, and D. Scaramuzza, "Collaborative monocular slam with multiple micro aerial vehicles," in *IEEE/RSJ Intl. Conf. on Intelligent Robots and Systems*, 2013.
- [25] L. Riazuelo, J. Civera, and J. Montiel, "C2tam: A cloud framework for cooperative tracking and mapping," *Robotics and Autonomous Systems*, 2014.
- [26] P. Schmuck and M. Chli, "Multi-uav collaborative monocular slam," in *IEEE Intl. Conf. on Robotics and Automation*, 2017.
- [27] 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.
- [28] M. Dymczyk, S. Lynen, T. Cieslewski, M. Bosse, R. Siegwart, and P. Furgale, "The gist of maps-summarizing experience for lifelong localization," in *IEEE Intl. Conf. on Robotics and Automation*, 2015.
- [29] L. Contreras and W. Mayol-Cuevas, "O-poco: Online point cloud compression mapping for visual odometry and slam," in *IEEE Intl. Conf. on Robotics and Automation*, 2017.
- [30] R. Arandjelovic, P. Gronat, A. Torii, T. Pajdla, and J. Sivic, "Netvlad: Cnn architecture for weakly supervised place recognition," in *IEEE Conf. on Computer Vision and Pattern Recognition*, 2016.
- [31] T. Cieslewski and D. Scaramuzza, "Efficient decentralized visual place recognition using a distributed inverted index," *IEEE Robotics and Automation Letters*, 2017.
- [32] A. Cunningham, K. Wurm, W. Burgard, and F. Dellaert, "Fully distributed scalable smoothing and mapping with robust multi-robot data association," in *IEEE Intl. Conf. on Robotics and Automation (ICRA)*, St. Paul, MN, 2012.
- [33] S. Choudhary, L. Carlone, C. Nieto, J. Rogers, H. I. Christensen, and F. Dellaert, "Distributed mapping with privacy and communication constraints: Lightweight algorithms and object-based models," *arXiv preprint arXiv:1702.03435*, 2017.
- [34] G. Bresson, R. Aufrère, and R. Chapuis, "Consistent multi-robot decentralized slam with unknown initial positions," in *Information Fusion (FUSION), 2013 16th International Conference on*, 2013.
- [35] L. Lamport, "Time, clocks, and the ordering of events in a distributed system," *Communications of the ACM*, 1978.
- [36] R. Mur-Artal and J. D. Tardós, "Orb-slam2: An open-source slam system for monocular, stereo, and rgb-d cameras," *IEEE Transactions on Robotics*, 2017.
- [37] W. Maddern, G. Pascoe, C. Linegar, and P. Newman, "1 year, 1000 km: The oxford robotcar dataset," *The Intl. J. of Robotics Research*, 2017.
- [38] E. Rublee, V. Rabaud, K. Konolige, and G. Bradski, "Orb: An efficient alternative to sift or surf," in *Int. Conf. on Computer Vision*, 2011.
- [39] L. Carlone, D. Rosen, G. Calafiore, J. Leonard, and F. Dellaert, "Lagrangian duality in 3D SLAM: Verification techniques and optimal solutions," in *IEEE/RSJ Intl. Conf. on Intelligent Robots and Systems (IROS)*, 2015.
- [40] D. Bertsekas and J. Tsitsiklis, *Parallel and Distributed Computation: Numerical Methods*. Englewood Cliffs, NJ: Prentice-Hall, 1989.
- [41] T. Cieslewski, S. Lynen, M. Dymczyk, S. Magnenat, and R. Siegwart, "Map api-scalable decentralized map building for robots," in *IEEE Intl. Conf. on Robotics and Automation*, 2015.
- [42] M. Gadd and P. Newman, "Checkout my map: Version control for fleetwide visual localisation," in *IEEE/RSJ Intl. Conf. on Intelligent Robots and Systems*, 2016.
- [43] A. Geiger, P. Lenz, C. Stiller, and R. Urtasun, "Vision meets robotics: The kitti dataset," *The Intl. J. of Robotics Research*, 2013.
- [44] M. F. Fallon, H. Johnsson, M. Kaess, and J. J. Leonard, "The MIT Stata Center dataset," vol. 32, pp. 1695–1699, 2013.