Collaborative SLAM over Mobile Networks

Han My Do, Marton Kajo*
*Chair of Network Architectures and Services, Department of Informatics
Technical University of Munich, Germany
Email: han-my.do@tum.de, kajo@net.in.tum.de

***Abstract*—Simultaneous Localization and Mapping (SLAM) in general is a problem that is key to the path planning of autonomous robots. The tasks of generating a map of an unknown environment while keeping track of its position are accomplished more accurate in a system with multiple robots. Such collaborative SLAM systems can be found in modern warehouses, where the logistics chain is performed by Automated Guided Vehicles (AGVs). With example industry use cases, this paper gives an overview on the main topics of collaborative SLAM and analyzes the different approaches to its components, architecture and communication. SLAM communication methods over mobile networks are analyzed and provide insights to the synergy potential 5G and SLAM has to offer.**

***Index Terms*—SLAM, Simultaneous Localization and Mapping, mobile networks, collaborative, 5G, logistics, visual, autonomous**

1. Introduction

Due to the need for more automated and flexible logistics systems, Automated Guided Vehicles (AGVs) are gaining foothold in the industry and enable a more efficient way of modernizing the industry. 5G has a lot of industrial focus with its Ultra Reliable and Low Latency Communications (URLLC) and broadband use-cases, and is the perfect fit for the communication solution for these AGVs.

1.1. Visual SLAM

For autonomous robots to function and navigate in a secure and robust way in a highly complex environment such as warehouses, the topic of Simultaneous Localization and Mapping (SLAM) is of high interest. SLAM involves the problem of simultaneously determining the position of a robot and the generation of a map of its environment. The interdependence of those two is made clear when keeping in mind that for path planning of a robot, not only its own position and orientation, but also obstacles such as humans and other robots play a role [1]. Therefore, to generate the map of the robot's environment different kinds of sensors are used.

Visual SLAM describes those systems that use cameras as the only exteroceptive sensor [2]. Cameras are lightweight, inexpensive and offer a lot of visual information. Thanks to the fast development and improvements of visual SLAM, as well as the growing computer performance, cameras have become an increasingly popular sensor for SLAM applications. Especially since inertial measurement units were integrated into visual SLAM entities, the system profits from better robustness and accuracy thanks to the additional inertial information such as acceleration and angular rate [3].

Typically, a visual SLAM system has two task areas. The front-end takes care of processing the image and extracting features to match and track those across different video frames. The back-end computes the camera poses and 3D coordinates. This geometric computation is often done with a filter or a nonlinear least squares optimizer. Further important SLAM issues are loop closure, re-localization, outlier rejection and the architecture [3].

Figure 1 shows the categorization of visual SLAM tasks in front-end and back-end.

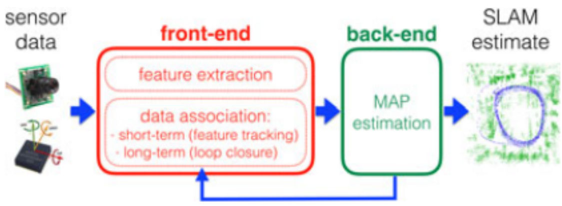

Figure 1: The two task areas of visual SLAM [1]

As SLAM research over the years has mainly developed visual-inertial algorithms, visual SLAM represents the state-of-the-art [1]. Nonetheless, we will take a short look at other systems with a different main sensor than video cameras.

1.2. Nonvisual SLAM

Besides cameras, other exteroceptive sensors in SLAM systems include sonars, range lasers, and global positioning systems (GPS). Even though sonars and range lasers are very precise and offer dense information of the environment's setting, they are limited for automated robots in logistic facilites, since they are heavy and have large pieces of equipment which makes them unsuitable for aerial or smaller robots [2]. Furthermore, they do have difficulties with highly cluttered environments, which makes it difficult for correctly mapping warehouses or comparable facilites. GPS sensors face similar complications when the signal is not available indoors at all times.

To ensure an accurate and robust estimation of the position of the robot, it is of advantage to combine the gained information from multiple exteroceptive sensors and proprioceptive ones. The latter are for example en-

doi: 10.2313/NET-2021-05-1_02

coders, accelerometers and gyroscopes which measure velocity, position and acceleration [2].

After this short introduction to the differences of visual and nonvisual SLAM, the focus will be on collaborative SLAM which enables the interaction between multiple entities and their environment.

2. Collaborative SLAM

In the first chapter we introduced the hardware specifications of a SLAM system and described the different sensors needed for an autonomous robot. In this part of the paper the focus will be on the software side of the system. Different aspects and methods to solve the collaborative SLAM problem are presented.

What makes multiple-agent SLAM more complex than single systems is that robots must process available data to construct a consistent global map while simultaneously localize themselves within the map [4]. In the following we provide an overview on researched approaches to collaborative SLAM.

2.1. Key Components

SLAM exploration and mapping tasks are fulfilled faster and with more accuracy by multiple robots than by just one. This also allows for a heterogenous team of robots with each one having its own specialization [5].

Another advantage is that the whole system is more robust in a distributed system due to the fact that failure of one robot does not crash the whole system [4].

However, the main difference between the elements of single-agent SLAM and multi-agent SLAM is the processing of data from multiple participants [3]. Otherwise, the building blocks are the same to a single SLAM system. In the following we will shortly present specific elements of a multi-robot collaborative SLAM.

There are two prominent approaches to pose estimation: key-frame based and filter-based methods. Key-frame based methods are more suitable and easier to implement for sharing information among the different robots. Depending on the architecture, the key-frames are sent to the server and can be downloaded by the participants. This means that every agent has access to key-frames produced by the others for their own pose estimation.

After synchronization of the participants' video frames a pose estimation between several robots is possible. This approach considers static points as well as moving points and enables a robust localization even with moving obstacles in the environment [3]. Figure 2 shows the method of camera synchronized pose estimation. Even though the moving object blocks the view of Camera A on the static background, Camera B is able to detect the background as well as the moving object.

The key-frame based approaches proved to be an efficient way for visual SLAM systems as it separates computation of real-time pose estimation and the complex mapping tasks. Pose estimation and mapping are calculated rotationally and can therefore resort to the previously calculated results. A key-frame contains the detected feature points and their corresponding coordinates. Aligning those data from the previous and current key-frame allows for the localization of the agent. The mapping task is solved by triangulation of the matching feature points between different key-frames. On a collaborative level the mapping is also done by using image information captured by the different cameras to generate map points.

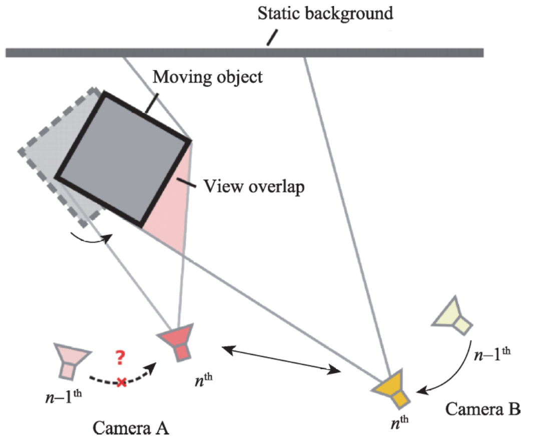

Figure 2: Synchronized pose estimation in collaborative SLAM [3]

The filter-based approaches use the Extended Kalman Filter to estimate the camera pose through iteration. The state vector also includes the 3D coordinates of landmarks in the environment. With every iteration these coordinates and the camera motion are updated and lead to high computation load with an increasing number of landmarks.

Another important task is loop closure, which detects already visited areas to update the map and correct accumulated inaccuracies. Loop closure is done by detecting the overlaps in some specific regions among multiple individual maps for fusion. A globald descriptor is used to check the similarity of two images to detect the overlap. Otherwise, collaborative loop closure follows the same pipeline as in single-agent SLAM algorithms. Such cooperation among the multiple cameras result in more accurate and robust estimations [3].

2.2. Architecture

A major challenge of collaborative SLAM is to distribute the time-consuming computational tasks to different agents with limited onboard computational resources [3]. This involves designing complex distributed algorithms to solve those computational tasks appropriately.

It is also important to consider the communication load to design the distributed algorithms and consider the strict bandwidth constraints when applying the decentralized architecture [3].

Figure 3 illustrates four main issues that arise in the context of data handling in collaborative SLAM.

The first topic is Data Communication. The SLAM system has to provide communication channels that allow for information sharing between the multiple agents. Central factors are bandwidth and communication network coverage.

The Data Sharing can span from exchanging raw sensor data to refined data. Measurements of exteroceptive and proprioceptive sensors are understood as raw information, while the refined data are those that are processed

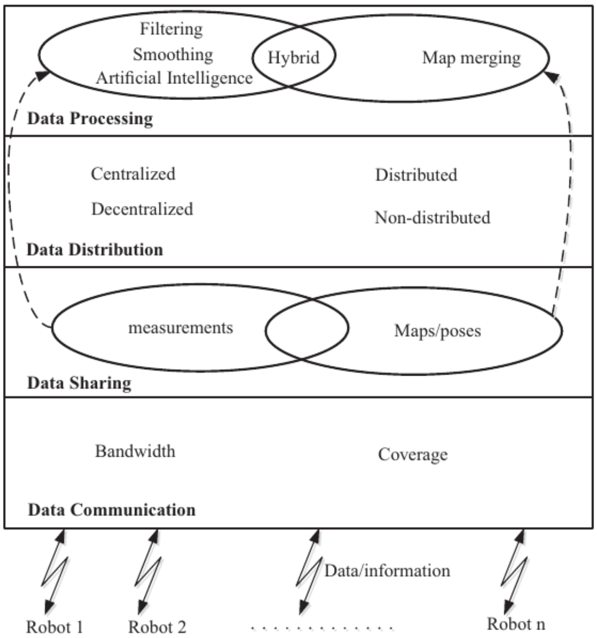

Figure 3: Issues in collaborative SLAM [4]

through filtering or optimization. Examples for processed data are environment maps or robot poses. Even though raw data offers more flexibility, the prerequisites for functioning are high bandwidth and stable communication between the entities. Computational power is essential as well. While processed information does not have the same amount of requirements and even reduces redundancy, the results are determined by the maps' quality.

Data Processing covers a wide range of methods and algorithms with filter-based and key-frame based approaches as their foundation. The choice of the data processing method is again dependent upon several factors such as processing power, type of sensor data and memory space of the entities [4].

For the distribution of data it is possible to deploy a centralized or decentralized architecture. Most collaborative systems use a central powerful server to collect all data and to process the computational-intensive tasks such as map optimization, loop detection and pose graph optimization for each entity [3]. This entity is also responsible for answering requests and providing information. This architecture has the disadvantage that the functioning of the whole system is dependant on the one server to never fail and to always be reachable. It also has to scale to the number of participating robots in computation performance and bandwidth. Decentralized systems do not suffer from such bottlenecks [5], but are much more difficult to deploy as the computational tasks are performed by more than one robot.

2.3. Communication over Wifi and 5G

To fulfil the needs of communication in decentralized SLAM systems, it is advisable to take a look at wireless networks such as Wifi and 5G as they have suitable properties for the wireless and real-time transmission of huge amount of data. Not only is Wifi sufficient for communication, but Wifi sensing can help with the SLAM problem, too. Due to the wide spread of Wifi Access Points in urban environments and the availability of Wifi radios on most robots or mobile devices, [6] and [7] propose to incorporate Wifi sensing into visual SLAM algorithms. A general method for the integration of Wifi into visual SLAM is shown in Figure 4. Similar approaches of using the signal strength of Bluetooth and LTE can be found in [8].

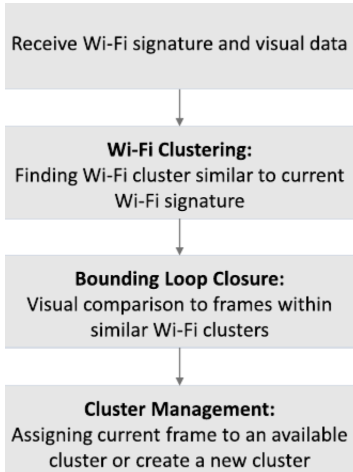

Figure 4: Pipeline of Wifi integration into visual SLAM, proposed by [6]

The use of 5G for SLAM methods, for example for the estimation of angle and delays of mmWave channels [9] or for the use of available multipath signals from landmarks to accomplish the mapping task [10], is promising as many of the required steps of Wifi integration as shown in Figure 4 can be omitted.

Three usage profiles are defined by the International Telecommunication Union for the International Mobile Telecommunications 2020 requirements for 5G networks. The three key capabilities are enhanced Mobile Broadband (eMBB), massive Machine-Type Communications (mMTC), and Ultra-Reliable and Low-Latency Communications (URLLC) [11].

URLLC lists specifications for seamless interaction between robots in real-time applications. Requirements are robustness, high bandwidth, and low latencies. With such significant advantages, 5G allows for reliable wireless and real-time transportation of high amounts of data, which accelerates the performance and functionalities of mobile robots. In addition, 5G allows to reserve sections of the network with a guaranteed Quality of Service [12].

User localization with 5G offers benefits such as high coverage, high accuracy, low energy consumption and scalability. The improvements in localization of users are possible due to the high concentration of base stations, device-to-device communication and mmWave technology [13]. For users, such as autonomous vehicles in complex settings, one crucial topic is accurate positioning. 5G in combination with collaborative SLAM approaches provide an optimal basis for the positioning task.

3. Use Cases for SLAM

Lastly, we explore applicable use cases of SLAM and show potential future research topics. The approaches to visual SLAM in a collaborative way over wireless networks pave the way to some interesting use cases which are described in the following.

3.1. Logistics

The use of Automated Guided Vehicle (AGV) or autonomous Micro Aerial Vehicles for the automation of modern logistics systems is quite common. Traditionally, SLAM is based on laser reflector and triangulation which is dependant on an established and static structured working environment, which is rarely the case in warehouses. With multi-agent visual SLAM algorithms, entities can localize themselves automatically and trace their path accurately in a dynamic and unstructured environment. Visual SLAM improves working efficiency, system flexibility and reduces constructing cost [14].

3.2. Autonomous Driving

Related to AGV is the use case of self-driving cars. An important aspect is the way data and communication are handled in a centralized or decentralized way. The availability of internet connection in the vehicle also plays a central role. The aspects of real time updates and offloading critical processing aspects onto the cloud spark discussions about safety. A future field of work is the architecture design of software which should be able to handle data flows and to segment updates [15].

3.3. Augmented Reality

Augmented Reality (AR) applications can benefit from SLAM systems, because the gained information enriches the AR experience from a technical aspect. AR systems face important technical challenges which come down to the need of specific information that SLAM can offer. One type of information needed is the current view of the real environment that is supposed to be augmented, while others are the shape of the virtual object and its location within the real world. When combined with other sensors or tracking systems, well-designed user interaction and system design, it is possible to widen the extent of AR to any environment [16]. Such an environment can also be warehouses where AR can be used for information exchange between teams and for prevention of errors and support.

4. Conclusion

In this paper we briefly described the differences between visual and nonvisual SLAM and went on to analyse the characteristics of collaborative SLAM. Focus was also set on the advantages of a decentralized architecture of multi-agent systems and their communication over wireless networks. Based on our findings we referred back to our introductory example use case of collaborative visual SLAM in logistics which was followed by further related use cases of autonomous driving and AR.

Even though visual SLAM in general, as well as in a collaborative way is already discussed thouroughly in existing literature, there is space for further research in the topics of decentralized multi-robot SLAM over wireless networks. Especially 5G in combination with collaborative SLAM is not yet comprehensively researched, but offer many improvements and great potential for synergy as evaluated in chapter 2.3. of this paper.

References

[1] C. Cadena, L. Carlone, H. Carrillo, Y. Latif, D. Scaramuzza, J. Neira, I. D. Reid, and J. J. Leonard, “Simultaneous localization and mapping: Present, future, and the robust-perception age,” *CoRR*, vol. abs/1606.05830, 2016. [Online]. Available: http://arxiv.org/abs/1606.05830

[2] J. Fuentes-Pacheco, J. Ascencio, and J. Rendon-Mancha, “Visual simultaneous localization and mapping: A survey,” *Artificial Intelligence Review*, vol. 43, 11 2015.

[3] D. Zou, P. Tan, and W. Yu, “Collaborative visual slam for multiple agents:a brief survey,” *Virtual Reality & Intelligent Hardware*, vol. 1, no. 5, pp. 461 – 482, 2019, 3D Vision. [Online]. Available: http://www.sciencedirect.com/science/article/pii/S2096579619300634

[4] S. Saeedi, L. Paull, M. Trentini, and H. Li, “Multiple robot simultaneous localization and mapping,” in *2011 IEEE/RSJ International Conference on Intelligent Robots and Systems*, 2011, pp. 853–858.

[5] T. Cieslewski, S. Choudhary, and D. Scaramuzza, “Data-efficient decentralized visual SLAM,” *CoRR*, vol. abs/1710.05772, 2017. [Online]. Available: http://arxiv.org/abs/1710.05772

[6] Z. S. Hashemifar, C. Adhivarahan, A. Balakrishnan, and K. Dantu, “Augmenting visual SLAM with wi-fi sensing for indoor applications,” *CoRR*, vol. abs/1903.06687, 2019. [Online]. Available: http://arxiv.org/abs/1903.06687

[7] J. Huang, D. Millman, M. Quigley, D. Stavens, S. Thrun, and A. Aggarwal, “Efficient, generalized indoor wifi graphslam,” 06 2011, pp. 1038 – 1043.

[8] P. W. Mirowski, T. K. Ho, S. Yi, and M. MacDonald, “Signalslam: Simultaneous localization and mapping with mixed wifi, bluetooth, LTE and magnetic signals,” in *International Conference on Indoor Positioning and Indoor Navigation, IPIN 2013, Montbeliard, France, October 28-31, 2013*. IEEE, 2013, pp. 1–10. [Online]. Available: https://doi.org/10.1109/IPIN.2013.6817853

[9] H. Wymeersch and G. Seco-Granados, “Adaptive detection probability for mmwave 5g slam,” 03 2020, pp. 1–5.

[10] Y. Ge, H. Kim, F. Wen, L. Svensson, S. Kim, and H. Wymeersch, “Exploiting diffuse multipath in 5g slam,” 2020.

[11] Z. Li, M. A. Uusitalo, H. Shariatmadari, and B. Singh, “5g urllc: Design challenges and system concepts,” in *2018 15th International Symposium on Wireless Communication Systems (ISWCS)*, 2018, pp. 1–6.

[12] S. Ludwig, M. Karrenbauer, A. Fellan, H. D. Schotten, H. Buhr, S. Seetaraman, N. Niebert, A. Bernardy, V. Seelmann, V. Stich, A. Hoell, C. Stimming, H. Wu, S. Wunderlich, M. Taghouti, F. Fitzek, C. Pallasch, N. Hoffmann, W. Herfs, E. Eberhardt, and T. Schildknecht, “A5g architecture for the factory of the future,” in *2018 IEEE 23rd International Conference on Emerging Technologies and Factory Automation (ETFA)*, vol. 1, Sep. 2018, pp. 1409–1416.

[13] S. Zügner and M. Kajo, “User localization in 5g mobile networks.”

[14] Y. Chen, Y. Wu, and H. Xing, “A complete solution for agv slam integrated with navigation in modern warehouse environment,” *2017 Chinese Automation Congress (CAC)*, pp. 6418–6423, 2017.

[15] G. Bresson, Z. Alsayed, L. Yu, and S. Glaser, “Simultaneous localization and mapping: A survey of current trends in autonomous driving,” *IEEE Transactions on Intelligent Vehicles*, vol. 2, no. 3, pp. 194–220, 2017.

[16] G. Reitmayr, T. Langlotz, D. Wagner, A. Mulloni, G. Schall, D. Schmalstieg, and Q. Pan, “Simultaneous localization and mapping for augmented reality,” *International Symposium on Ubiquitous Virtual Reality*, vol. 0, pp. 5–8, 07 2010.