Heterogeneous Map Merging in Structured Indoor Environments

Maximiliaan van Schendel
student #4384644

1st supervisor: Edward Verbree
2nd supervisor: Pirouz Nourian
external supervisor: Robert Voûte

14/09/2021

Contents

1 Introduction

1.1 Relevance of the project

Collaborative SLAM (CSLAM) systems allow several independent agents to collaboratively map a space and to position themselves within it in a common coordinate frame. The advent of powerful mobile computing devices, high speed networks and affordable 3D laser scanners has lead to a surge in research on this topic. The results of this research have already found their way into new technology, such as intelligent robot swarms, multi-player augmented reality and autonomous vehicles.

As of yet most research has focused on cases where agents are homogeneous, meaning they each sense their environment and track their movement in the same way. However, in practice there could be situations where heterogeneous agents are advantageous, with different sensor configurations based on the agents' capabilities and purpose. For example, a human agent carrying lightweight but inaccurate sensors collaborating with a robot carrying heavy but accurate sensors.

Certain environments pose additional challenges to CSLAM systems based on their visual and physical characteristics. This is especially the case for indoor environments like offices or hospitals because external signals are usually blocked. Because humans spend most time indoors, research on heterogeneous CSLAM in indoor environments could have various applications, among which in 3D emergency management information systems, asset management and human-service robot interaction.

1.2 Its place in geomatics

The field of geomatics has historically focused on the outdoors due to limitations in sensing capabilities and ease of data acquisition. Technological advances have made it possible to map the world at ever smaller scales and has opened up the indoors to geomatics. The challenge of combining disparate measurements into a coherent whole that is more than the sum of its parts is also at the core of geomatics.

1.3 Problem statement

The simultaneous localization and mapping (SLAM) problem aims to estimate the posterior probability distribution of a map of an environment and an agent's trajectory within it, based on the agent's observations, control signals and initial pose (see equation 1) (Saeedi, Trentini, Seto, & Li, n.d.). Because the resulting map and trajectory are estimated in the local coordinate frame of the agent's initial pose they are referred to as the *local* map and trajectory.

$$p(x_{1:t}, m_t | z_{1:t}, u_{1:t}, x_0) \tag{1}$$

$x_{1:t}$ is the agent's local trajectory from the start of the session until time t.
m_t is the local map in the agent's coordinate frame at time t.
$z_{1:t}$ and $u_{1:t}$ are the entire set of the agent's observations and control signals.

The collaborative SLAM (CSLAM) problem is an extension of the SLAM problem where a single map and the trajectory of multiple agents are estimated in a common coordinate frame, based on each agent's observations and control signals (see equation 2). The results are referred to as the *global* map and trajectories. Effective solutions to the CSLAM problem exist for *homogeneous* observations, which represent the same physical quantity at a similar accuracy and resolution. However, when observations are *heterogeneous* it becomes unfeasible to find correspondences between observations directly, as they might represent different aspects of the environment, e.g. observations of infrared and visible spectrum light (Saeedi et al., n.d.).

$$p(X^i_{1:t}, M_t | z^i_{1:t}, u^i_{1:t}, x^i_0) \tag{2}$$

$X^i_{1:t}$ and M_t are the global trajectories and map, i identifies agents.

The *map merging* problem is a subproblem of the CSLAM problem where global maps and trajectories are estimated by merging local maps derived from each agents' observations (see equation 3). If the local maps represent comparable aspects of the environment, e.g. its geometry or topological structure, map merging allows us to estimate global maps and trajectories from heterogeneous observations. However, local maps derived from heterogeneous observations might still differ in accuracy and resolution, which makes finding correspondences between them difficult (Andersone, n.d.). We call this specific case the *heterogeneous map merging* problem, it is largely unsolved (Andersone, n.d.).

$$p(X^i_{1:t}, M_t | m^i_t, x^i_{1:t}) \tag{3}$$

Indoor environments are often highly structured, this places topological and geometrical constraints on the global map, e.g. ceilings are generally above 2m, rooms must be reachable. In this thesis we will research how prior knowledge of the topological and geometrical characteristics of structured indoor spaces can be used to improve heterogeneous map merging estimates (see equation 4).

$$p(X^i_{1:t}, M_t | m^i_t, x^i_{1:t}, e_{indoor}) \tag{4}$$

e_{indoor} are the defining characteristics of structured indoor environments.

2 Related work

2.1 Simultaneous Localization and Mapping

Simultaneous Localization and Mapping, commonly abbreviated as SLAM, concerns the problem of agents creating a map of an unknown environment while simultaneously positioning themselves within it. There are various approaches to solve this problem depending the resources available to the agent. These approaches are commonly divided into two categories depending on the type of algorithm they use (Yousif, Bab-Hadiashar, & Hoseinnezhad, n.d.).

Filtering SLAM The first category is the filtering approach. Early implementations of the filtering SLAM approach used an extended Kalman filter (EKF) to estimate the agent's current position and map the environment (Smith, Self, & Cheeseman, n.d.). The EKF-SLAM approach later lost popularity in favour of particle filter based approaches such as FastSLAM with improved computational complexity (Yousif et al., n.d.; Bailey, Nieto, Guivant, Stevens, & Nebot, n.d.; Montemerlo, Thrun, Koller, & Wegbreit, n.d.).

Smoothing SLAM In contrast with the filtering approach, smoothing SLAM estimates the current state based on all previous measurements, this is also called the full SLAM problem (Yousif et al., n.d.). A common approach to smoothing SLAM is to consider the agent's path as a graph, where nodes are either a pose or a recognizeable landmark in the environment (Grisetti, Kummerle, Stachniss, & Burgard, n.d.). Edges between nodes represent spatial transformations that are estimated by sensor measurements (dead reckoning and recognition of previously visited landmarks). Due to the noisy nature of real-world measurements edges have an associated uncertainty and may contradict other edges. To solve the SLAM problem error minimization techniques are applied to find the configuration of edges that maximally conforms to their inherent uncertainty (Grisetti et al., n.d.). An essential component of graph-based SLAM systems is loop closure (Grisetti et al., n.d.). If edges are only created by means of dead reckoning it is not possible to optimize the graph because the relationships between nodes are only measured once; the graph contains no loops. To create loops agents needs to be able to recognize previously visited landmarks and position themselves relatively to it. When a landmark has been identified an edge is added between it and the current node, creating a loop and making it possible to optimize the graph. Figure 2 illustrates how loop closure is used to create an optimizable graph.

In the years since its original formulation by Lu and Milios in 1997, advances in computation have made it easier to solve the large minimization problems that graph-based SLAM requires. Currently, this approach is used by state-of-the-art SLAM solutions such as ORB-SLAM and RTAB-Map (Mur-Artal & Tardos, n.d.; Ragot, Khemmar, Pokala, Rossi, & Ertaud, n.d.).

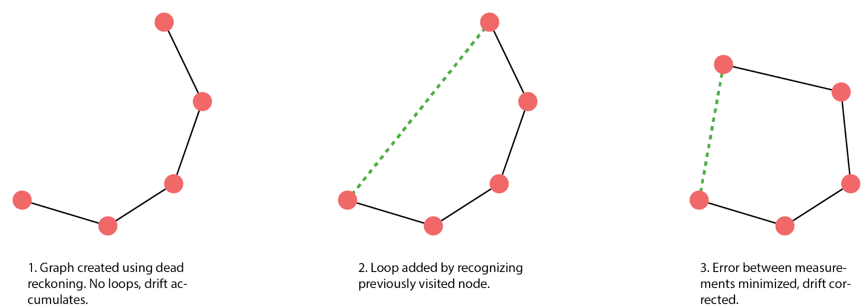

Figure 1: Illustration of loop closure in graph-based SLAM.

2.2 Collaborative Simultaneous Localization and Mapping

The collaborative SLAM (CSLAM) problem is an extension of the SLAM problem that deals with multiple agents mapping and positioning themselves within the same environment. The goal of this is to create a single, global map from each agent's local map. Collaboration allows us to speed up mapping time, map multiple places at once and determine the relative poses of agents.

2.2.1 CSLAM

There are various ways to approach the CSLAM problem that each have their own pros and cons depending on the environment and agent capabilities. Saeedi et al (2016) discusses the advantages of different CSLAM methodologies based on a set of problems that they solve. Based on these results we will discuss the C-SLAM approaches that show the most promise.

GraphSLAM The GraphSLAM approach to CSLAM is an extension of the graph-based SLAM approach discussed in the previous section. It relies on encounters between agents to create edges between their local pose graphs. The resulting connected pose graph is then optimized using the same error minimization techniques as mentioned before (Been Kim et al., n.d.). Encounters can be subdivided into direct and indirect encounters. The former occurs when agents meeting both in time and space and requires agents to be able to recognize eachother. The latter occurs when agents meet in space but not in time, meaning they visit the same location at different moments (Been Kim et al., n.d.).

Map Merging The map merging approach to CSLAM tries to create a global map by finding the transformations between agents' local maps (Saeedi et al., n.d.). This relies on overlaps between local maps to work. When one or multiple

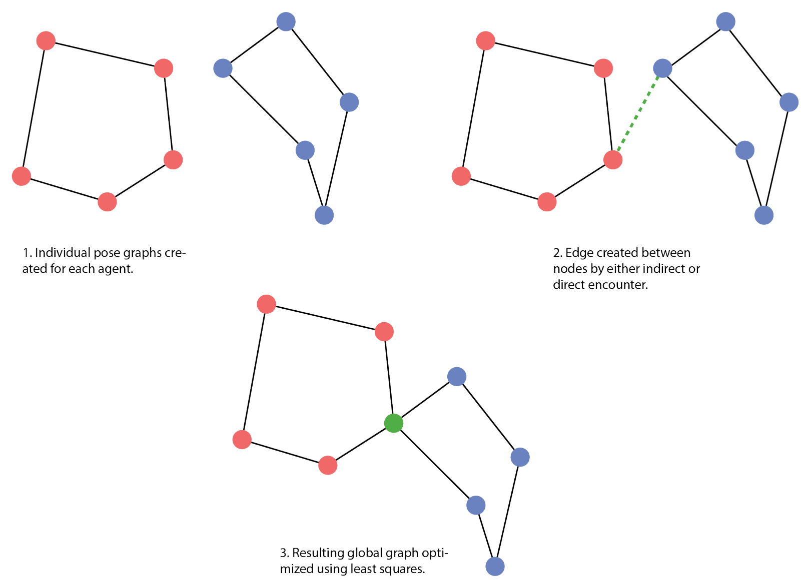

Figure 2: Local pose graph merging using agent encounters.

overlaps have been identified their relative transformations can be found by using scan matching algorithms such as the iterative closest point algorithm (Andersone, n.d.). This is computationally expensive and sensitive to distortions in the local maps (Saeedi et al., n.d.). Other map merging methods use extracted features instead of the map's geometry. In structured environments the Hough transform can be used to extract line or plane features which can be used to determine overlaps instead (Saeedi et al., n.d.). Different kind of features types have also been used to determine overlaps, such as SURF and SIFT, as well as different feature matching approaches, such as RANSAC or neural network matching (Andersone, n.d.; Fischer et al., n.d.).

Most map merging algorithms work on a 2D projection of originally 3D maps to reduce computational complexity. The same algorithms as above can usually also be applied to 3D data but extra care needs to be taken to ensure real-time performance by using a suitable feature matching approach (Saeedi et al., n.d.).

Topological Map Merging In the topological map merging approach local maps are matched by comparing their topological structure. A graph is extracted from each agent's local map that represent the connectivity of the environment. Some approaches try to detect doors and rooms using object detection and region segmentation and use those as the edges and nodes of the graph (Andersone, n.d.). Others extract a Voronoi diagram directly from the map's geometry (Andersone, n.d.). After extracting the graphs they can be merged by finding the transformation that maximizes their overlap using various optimization methods such as maximum subgraph matching (Alami, Chatila, & Asama, n.d.).

2.2.2 Communication

2.2.3 Heterogeneous CSLAM

3 Research questions

3.1 Main question

How can we improve the effectiveness of heterogeneous map merging by exploiting the topological and geometrical characteristics of structured indoor environments?

3.2 Subquestions

1. Which topological and geometrical constraints define structured indoor environments?

2. Which local and global map representation are best suited for our problem?

3. How can we extract this map representation from the local maps produced by existing single-agent SLAM solutions?

4. How can we use the environmental constraints to improve heterogeneous map merging effectiveness?

3.3 Scope

To better delineate the scope of the thesis we provide several things that we will not be doing.

1. Map merging using known relative poses between agents or meeting strategies. Agent behaviour is assumed to be independent and agents are not able to sense eachother.

2. Map merging using observations unrelated to the environment's geometrical and topological characteristics. E.g. the environment's colour or actively transmitted beacon signals.

3. Map merging assisted by a priori knowledge of the environment. E.g. building information models (BIM) or floor plans.

4. Correspondence detection between raw observations. Although research has been done on using machine learning techniques to find correspondences between observations that represent different physical quantities, e.g. lidar depth and monocular RGB, this thesis focuses on the merging of derived local maps.

5. Implementing a single-agent SLAM solution ourselves. Much research has been done on the single-agent SLAM problem and open-source implementations are widely available.

6. Researching observations besides monocular RGB, monocular infrared, binocular depth and lidar depth. E.g. sonar and radar will not be considered.

7. Creating a feedback mechanism that allows agents to access the global map and trajectories. Although useful for real-world applications this is a matter of implementation and does not add value to the research.

4 Methodology

Overview of the methodology to be used.

5 Time planning

Having a Gantt chart is probably a better idea then just a list.

6 Tools and datasets used

Since specific data and tools have to be used, it's good to present these concretely, so that the mentors know that you have a grasp of all aspects of the project.

References

Alami, R., Chatila, R., & Asama, H. (Eds.). (n.d.). *Distributed autonomous robotic system 6.* Springer. (Meeting Name: International Symposium on Distributed Autonomous Robotic Systems OCLC: ocn144524594)

Andersone. (n.d.). Heterogeneous map merging: State of the art. , *8*(3), 74. Retrieved 2021-07-16, from `https://www.mdpi.com/2218-6581/8/3/74` doi: 10.3390/robotics8030074

Bailey, T., Nieto, J., Guivant, J., Stevens, M., & Nebot, E. (n.d.). Consistency of the EKF-SLAM algorithm. In *2006 IEEE/RSJ international conference on intelligent robots and systems* (pp. 3562–3568). IEEE. Retrieved 2021-09-28, from `http://ieeexplore.ieee.org/document/4058955/` doi: 10.1109/IROS.2006.281644

Been Kim, Kaess, M., Fletcher, L., Leonard, J., Bachrach, A., Roy, N., & Teller, S. (n.d.). Multiple relative pose graphs for robust cooperative mapping. In *2010 IEEE international conference on robotics and automation* (pp. 3185–3192). IEEE. Retrieved 2021-09-28, from `http://ieeexplore.ieee.org/document/5509154/` doi: 10.1109/ROBOT.2010.5509154

Fischer, K., Simon, M., Oelsner, F., Milz, S., Gross, H.-M., & Maeder, P. (n.d.). StickyPillars: Robust and efficient feature matching on point clouds using graph neural networks. Retrieved 2021-09-29, from `http://arxiv.org/abs/2002.03983`

Grisetti, G., Kummerle, R., Stachniss, C., & Burgard, W. (n.d.). A tutorial on graph-based SLAM. , *2*(4), 31–43. Retrieved 2021-06-24, from `http://ieeexplore.ieee.org/document/5681215/` doi: 10.1109/MITS.2010.939925

Montemerlo, M., Thrun, S., Koller, D., & Wegbreit, B. (n.d.). FastSLAM: A factored solution to the simultaneous localization and mapping problem. , 6.

Mur-Artal, R., & Tardos, J. D. (n.d.). ORB-SLAM2: an open-source SLAM system for monocular, stereo and RGB-d cameras. , *33*(5), 1255–1262. Retrieved 2021-09-24, from `http://arxiv.org/abs/1610.06475` doi: 10.1109/TRO.2017.2705103

Ragot, N., Khemmar, R., Pokala, A., Rossi, R., & Ertaud, J.-Y. (n.d.). Benchmark of visual SLAM algorithms: ORB-SLAM2 vs RTAB-map*. In *2019 eighth international conference on emerging security technologies (EST)* (pp. 1–6). IEEE. Retrieved 2021-09-17, from `https://ieeexplore.ieee.org/document/8806213/` doi: 10.1109/EST.2019.8806213

Saeedi, S., Trentini, M., Seto, M., & Li, H. (n.d.). Multiple-robot simultaneous localization and mapping: A review: Multiple-robot simultaneous localization and mapping. , *33*(1), 3–46. Retrieved 2021-09-28, from `https://onlinelibrary.wiley.com/doi/10.1002/rob.21620` doi: 10.1002/rob.21620

Smith, R., Self, M., & Cheeseman, P. (n.d.). Estimating uncertain spatial relationships in robotics. , 26.

Yousif, K., Bab-Hadiashar, A., & Hoseinnezhad, R. (n.d.). An overview to visual odometry and visual SLAM: Applications to mobile robotics. , *1*(4), 289–311. Retrieved 2021-09-24, from http://link.springer.com/10.1007/s40903-015-0032-7 doi: 10.1007/s40903-015-0032-7