ÉCOLE CENTRALE DE NANTES UNIVERSITY OF GENOA

MASTER ERASMUS MUNDUS
EMARO+ "European Master in Advanced Robotics"

2016 / 2017

Master Thesis Report

Presented by

Ernest Skrzypczyk, BSc

On 24.01.16

UGV and UAV collaboration in an autonomous infrastructure scenario

Jury

President:	Olivier Kermorgant	Professor (Institution)
Evaluators:	Ina Taralova	Professor (Institution)
	Name	Position (Institution)
	Name	Position (Institution)
Supervisors:	Olivier Kermorgant	Professor (IRCCYN)
	David Marquez-Gamez	Professor (IRT Jules)

Laboratory: Laboratoire des Sciences du Numérique de Nantes LS2N

Contents

List of Figures

List of Tables

Abstract

Abbreviations

UAV Unmanned Air Vehicle

UGV Unmanned Ground Vehicle

SLAM Simultaneous Localization And Mapping

VO Visual Odometry

GVG Generalized Voronoi Graph

LSD-SLAM Large-Scale Direct Monocular SLAM

DSO Direct Sparse Odometry

KF Kalman Filter

EKF Extended Kalman Filter

UKF Unscented Kalman Filter

IR Infra-Red

IMU Inertial Measurement Unit

GPS Global Positioning System

LIDAR Light Detection And Ranging

RGB Red, Green, Blue

RGBD Red, Green, Blue, Depth

MAP Maximum-A-Posteriori [estimation]

MR-SLAM Multi-Robot SLAM

BoW Bag of Words

BA Bundle Adjustment

ICP Iterative Closest Point

CML Concurrent Mapping and Localization

SODAR SOnic Detection And Ranging

Sonar SOund Navigation And Ranging (original acronym)

API Application Programming Interface

UUID Universally Unique IDentifier

BFS Breadth-first Search

ORB Oriented FAST and Rotated BRIEF

FAST Features from Accelerated Segment Test

BRIEF Binary Robust Independent Elementary Features

LC Loop Closure

PnP Perspective-n-Problem

RRR Realizing, Reversing and Recovering

Symbols

R_i – Robot unit i

Chapter 1

Introduction

Introduction

Autonomous infrastructure, even though not yet formally defined, has seen a high interest and development in recent years in certain fields related to security, data transmission [1] and transport. In the scope of global production, personal and products transport as well as services it is highly dependent on robot collaboration. Autonomy of the individual units is desired, since the possible working scenarios are usually non trivial. For very complex problems the collaboration of robots that have complementary capabilities could achieve group autonomy. The master thesis deals with an Unmanned Ground Vehicle (UGV) and Unmanned Air Vehicle (UAV) collaboration for a decentralized multi SLAM (Simultaneous Localization and Mapping) process with the help of visual servoing within the autonomous infrastructure aspect.

The proposed framework is a decentralized approach with units of different type. The robots conduct SLAM in an unknown environment, starting from an unknown location and store the map locally in memory. The relevant data is shared between units once in communication range and both local maps are optimized based upon virtual loop closures using Bundle Adjustment (BA) or similar techniques. In order to provide additional information and utilize rendezvous for data exchange, visual servoing and visual communication can be utilized. While the UAV follows the UGV marked with a QR code or similar marker, it can provide additional information from another perspective about the environment. It is of particular interest to involve these two types of robots, because of their different capabilities and very different points of view into the environment.

By using LEDs or other light source devices around the marker, an unidirectional communication between the mobile platform and drone could be established. This can help with establishing rendezvous points between the robots, once a unit is ready to optimize it's local map, it's memory is full or it requires assistance in establishing next move. In the last case the UAV could be instructed to perform an area sweep in order to supply additional map information. The LEDs also provide a possibility of pose estimation using DeMenthon's POSIT algorithm [29].

One possible problem in this approach is the communication of limited range between units. Therefore data transmission optimizing measures should be performed. The basic data will consist of key frames coming from the output of an monocular SLAM approach like Large-Scale Direct Monocular SLAM (LSD-SLAM) [12] or Direct Sparse Odometry (DSO) [13].

Context of the work

An autonomous infrastructure is one with the capability of independent self-adjustment and self-maintenance. Today this term is mostly associated in robotics and in a very limited way, in general with aspects of security, data networking and transport. The master thesis formalizes the term therefore in more detail 2.

Motivation

The reason are current developments in the technological sector as well as on the level of society. According to the United Nations Department of Economic and Social Affairs report, human population will reach approximately 9 billion in the year 2050 and 11 billion by the end of the century with a probability of 80% [34][15]. It can be observed that the current premise of the socio-economic structure is that of an ever growing economy with little regard to related physical processes. Yet the demand put onto the system grows exponentially, accelerated by the rapid population growth in developing countries, the market demand and finite resources availability.

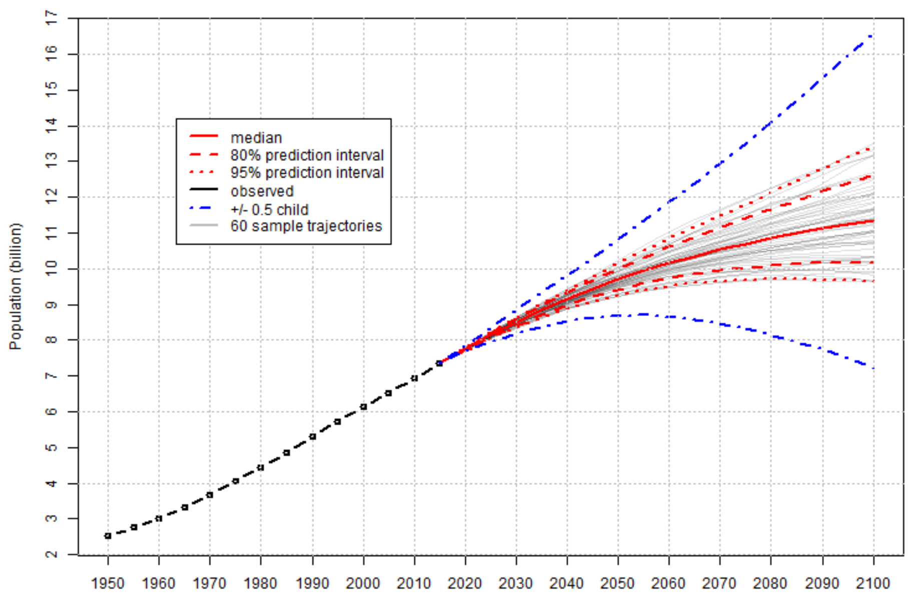


Figure 1.1: Total world population projection using Bayesian filtering [34].

Recent technological developments in the infrastructure lead to highly complex system, that are difficult to control. The power outage in Canada that occurred in 2003, caused by a chain of events, mostly related to human error [14], clearly indicated that current state of control in the infrastructure is not satisfactory and mostly still a human responsibility.

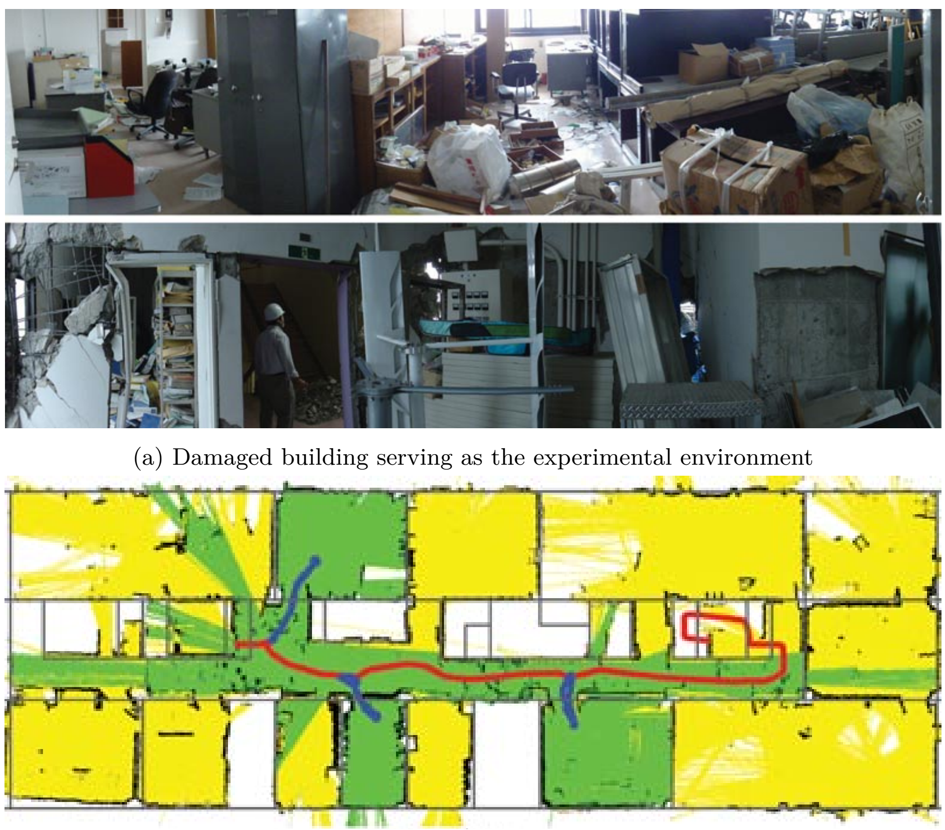
On the other hand, there rapid development can be observed in various fields of technology and research. Most promising technologies, that would allow for increase of the system's overall efficiency, like nanotechnology and technological singularity, can also have detrimental effects by simply increasing the complexity of the system, for example. As per projections of current developments [22][19], the planetary power output generated could elevate the Kardashev scale rating, shown in (1.1), of humanity to a civilization type I [20], ergo approximately $10^{16} \div 10^{17} W$.

$$K = \frac{log_{10} P - 6}{10} \tag{1.1}$$

K – Kardashev rating, P – power output

From these observations it can be stated that the current socio-economic system, in particular the infrastructure, is not prepared for future demand mentioned earlier, also. Therefore the master thesis introduces the idea of autonomous infrastructure and deals with a specific type of robot collaboration that is currently an open problem [6].

Robot collaboration offers a variety applications, from simple delivery tasks to complex rescue scenarios in dangerous or damaged environments [27] shown in (1.2). Other applications for robot collaboration in search and rescue like forest fires, in other fields like cleaning operations, space and underwater exploration, military applications, security and surveillance, as well as maintenance investigations are also of interest [32]. Current state of robot collaboration has a lot of potential for improvement, especially in the context to an autonomous infrastructure level 4 defined in *"Definitions"*. Therefore is it the motivation of the master thesis to contribute in a problem scenario that can benefit further from the collaboration aspect. It is also of interest to explore methods that can improve collaboration for solving the SLAM problem like visual servoing, visual communication and others.



(a) Damaged building serving as the experimental environment

(b) 2D occupancy grid maps

– UGV1, – UGV2, – UAV

Figure 1.2: Collaborative mapping of a damaged building performed by two UGV and one UAV [27].

Problem formulation

Based on previous work [9][8] development, implementation and evaluation of a decentralized approach for collaborative SLAM with the possibility of further expansion, are the desired goal of the master thesis.

Contribution and innovation

The master thesis project aims at developing and implementing a working multi SLAM algorithm between different type of robots, in particular between UAV and UGV. Furthermore the proposed solution uses a decentralized approach, where data communication is under a constraint related to the limited range of transmission. Additionally aspects like visual servoing for the purpose of optimization of the SLAM process and visual communication are to be analysed and evaluated. The whole work occurs in context of autonomous infrastructure.

Chapter 2

Background theory

Definitions

Robot collaboration Process of coordinated actions performed by robots towards a specified goal.

Autonomy The ability of making independent decisions.

Unmanned vehicle Vehicle designed without support systems for humans, capable of similar performance as a manned vehicle.

Autonomous infrastructure An infrastructure with the capability of independent self-adjustment and self-maintenance.

To further formalize the term *autonomous infrastructure*, it can be classified by 4 levels:

0 The autonomy is local in very limited aspects, there is no autonomous chain. The majority of processes are human controlled and monitored.

1 Branches of the economy hold at least one full chain of product or services. Specific needs of the community are satisfied by the infrastructure. Autonomy is regional in limited aspects. The majority of processes are human controlled and monitored.

2 Major branches of the economy, like nutrition production and personal transport, hold full autonomous chains of product or services. Majority of the needs of the community are satisfied by the infrastructure with little human interaction, mostly limited to monitoring. Autonomy is national and pseudo-continental in many aspects, self-maintenance achieved.

3 Full economy autonomy, virtually no human interaction required. Complex dynamic demand is satisfied after infrastructure adjustment.

4 Planetary autonomy. Infrastructure extrapolates current and adjusts itself for future developments.

Simultaneous localization and mapping

The process of mapping of the environment and using it for localization, while at the same time localizing with the help of that map has been first defined as Simultaneous Localization And Mapping (SLAM) problem in the 1986 [3]. It still is one of the fundamental challenges in robotics and spreads into many other fields of research. It is being further developed in the robotics field with respect to current open problems [6]. SLAM is also known as the Concurrent Mapping and Localization (CML) [2]. The complexity in the SLAM problem arises from the interdependence of both mapping and localization [32]. Artificial and natural features of the environment can be utilized for this purpose using a variety of sensors, whose measurements can be then fused and used for the localization and map building process.

SLAM is most often defined as a maximum-a-posteriori (MAP) estimation problem [30][6], that can be then expressed as probability distribution (2.1)

$$P(x_i, l_a | y_{0:i}, u_{0:i}) \tag{2.1}$$

where:

- x_i is the pose in the state vector at instance i;
- m_a is a set containing the locations of all landmarks, $m = m_0, m_1, \ldots, m_n$ with n being the number of landmarks;
- $z_{0:i}$ represents all measurements or observations of landmark locations l up to time instance i;
- $u_{0:i}$ holds all the control inputs up to time instance i, ergo the history of control input u.

(2.1a) illustrates the SLAM problem.

Additionally, often the formalism of factor graphs, shown in (2.1b), is used to express the relations among relevant variables [21][6].

Visual SLAM

Visual SLAM (V-SLAM) corresponds to solutions of the SLAM problem using vision sensors as the main tool . Cameras are often combined with other types of sensors in order to increase the confidence of measurements. This variant of the SLAM problem often can be formulated as mapping environment and determining camera pose at given instance, ergo trajectory of poses across data stream. Depending on the architecture, representations of the map and overall approach, the solutions for V-SLAM can be very different.

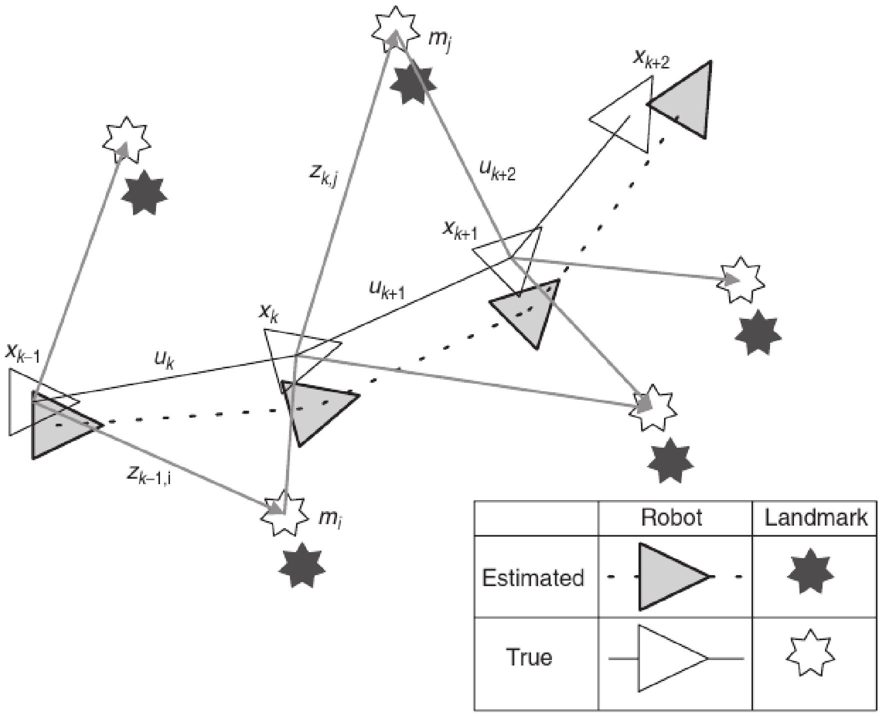
As with conventional SLAM approach, one of the more important aspects is the information certainty, which can be increased by revisiting already explored areas. This can be interpreted by the system as a Loop Closure (LC), once detected. Visual SLAM systems usually have a LC detection block in their algorithms.

The scale ambiguity is a problem predominantly in monocular SLAM systems. However, the formulation used by approaches like Large-Scale Direct SLAM (LSD-SLAM) that operate on *Sim(3)* allows to detect the scale drift [12]. It can hold 3 dimensional similarity transforms $\mathbf{S} \in Sim(3)$ consisting of rotation matrix $\mathbf{R} \in SO(3)$, translation vector $t \in \mathbb{R}^3$ and the scaling factor $s \in \mathbb{R}^+$ defined as:

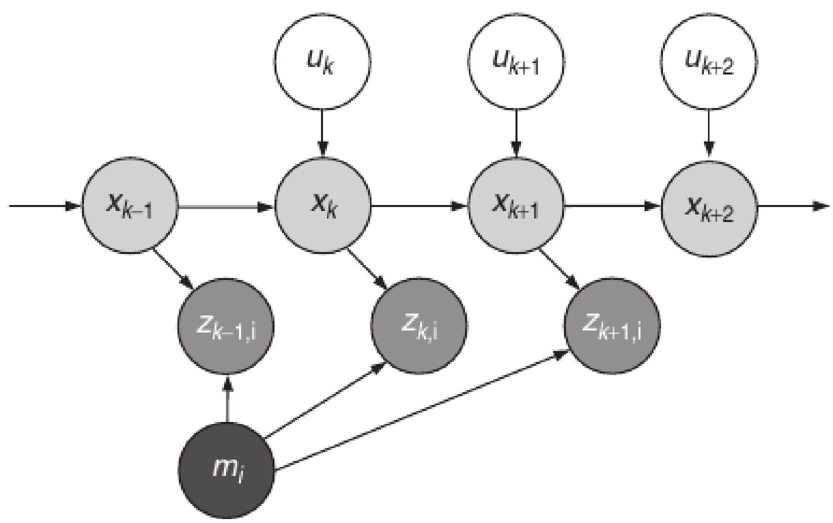
$$\mathbf{S} = \begin{bmatrix} sR & t \\ 0 & 1 \end{bmatrix} \tag{2.2}$$

This allows the SLAM system to operate within 7 DOFs, where 6 DOFs are for the position and orientation screw, and the last DOF is for the scale of environment.

The approaches of V-SLAM typically are either dense, semi-dense or semi-sparse, or sparse [6]. Figure (2.2) shows rendered raw and processed frames for a specific view of the environment in a semi-sparse approach, where lines and edges are Regions Of Interest (ROI) and therefore serve as information holders of the scene.



(a) Simultaneous mapping and localization problem

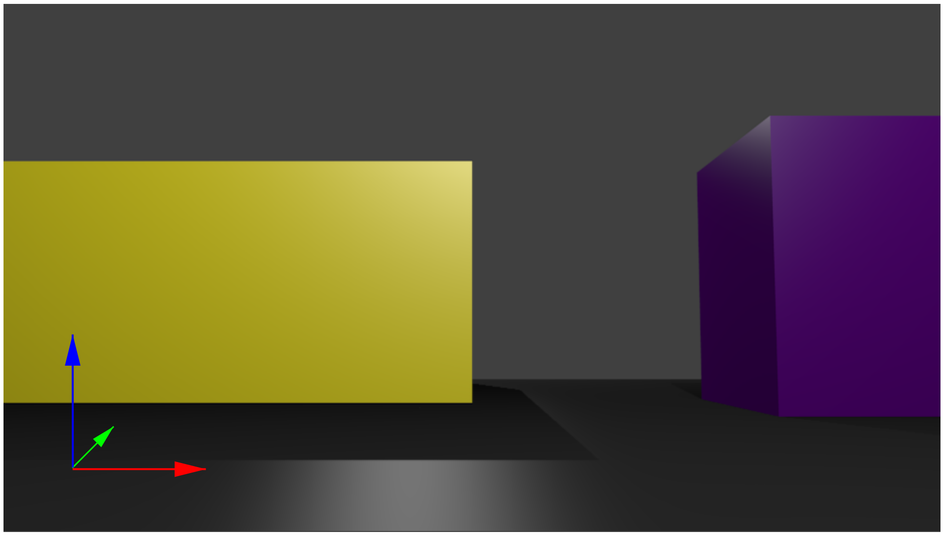


(b) Factor graph

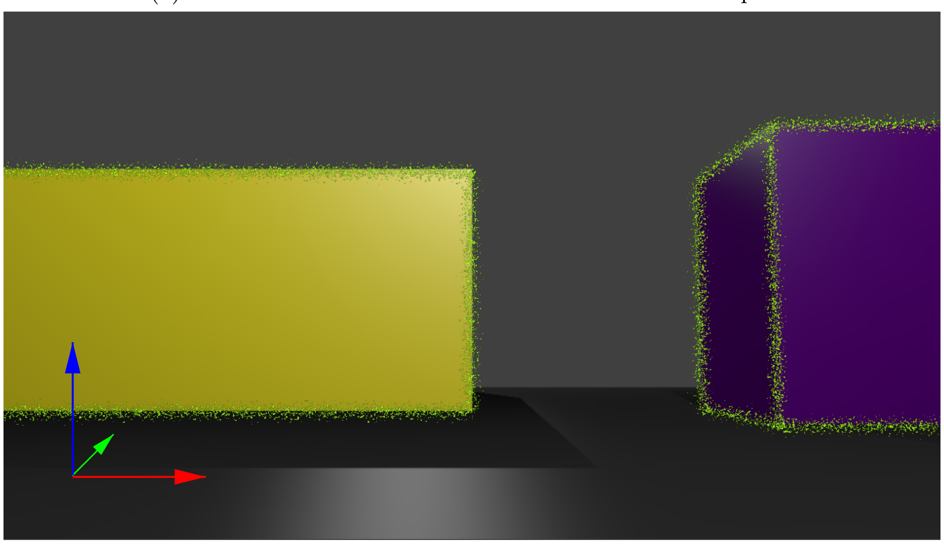
Figure 2.1: SLAM problem and factor graph representation [11].

Collaborative SLAM

In contrast to multi robot SLAM (MR-SLAM), also known as multiple robot SLAM [32], that can be defined as SLAM performed by multiple units, a collaborative SLAM has an active effort to use capabilities



(a) Raw frame with and indicator of the current unite pose.



(b) Processed frame with an sparse approach detecting edges of objects.

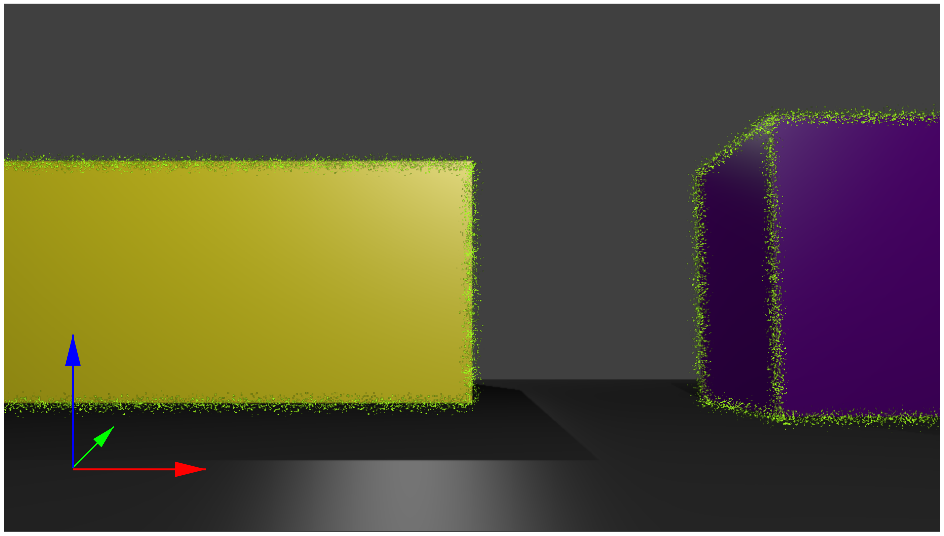
Figure 2.2: Visual SLAM using camera as sensor and a sparse/feature based approach for processing.

of all units *actively* in order to solve the SLAM problem.

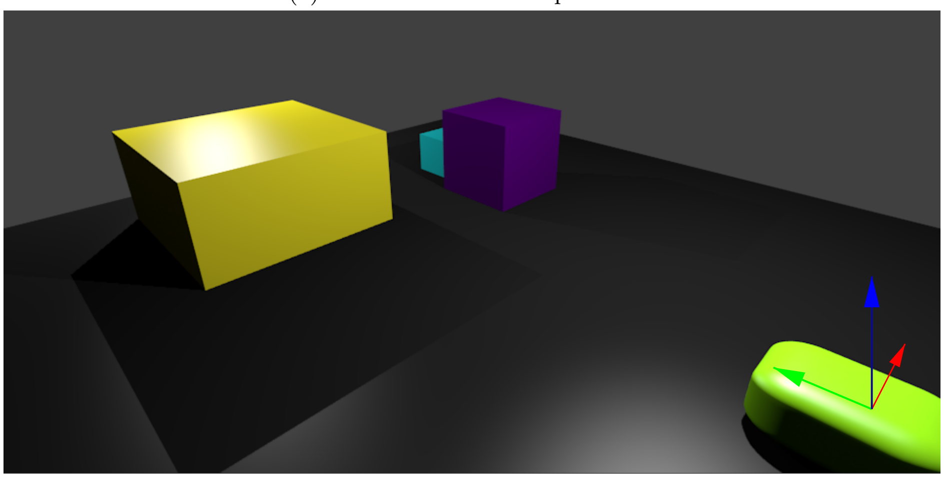
A scenario of mapping the environment shown in (2.3) portrays how a multi or collaborative SLAM can be performed. Frames from both units, an UGV (2.3a) and an UAV (2.3b) are shown in processed and raw state respectively. The camera of the UGV does not have the small cube, right behind the one, in sight, while UAV can perceive it. Through communicating the map to the UGV, it can recognize that what UAV captured with the camera relates to the same perceived scene, which corresponds to a loop closure. The UGV can merge the maps from both units and in that manner perceive the otherwise blocked from view object.

Factor graphs

A factor graph is defined as bipartite graph expressing factorization structure [21]. In other words factor graphs are an abstract holding factorization relations between a global function and local functions,



(a) UGV camera view – processed



(b) UAV camera view – raw

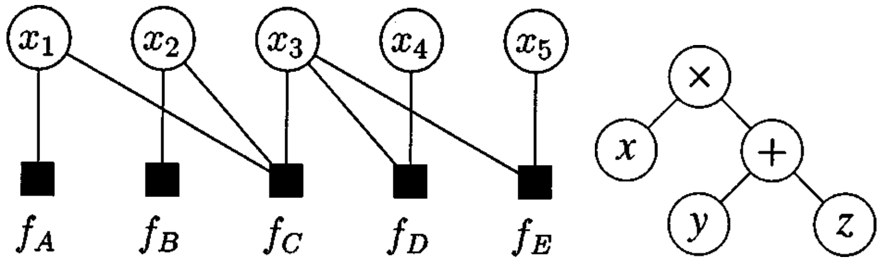
Figure 2.3: Collaborative visual SLAM scenario with UGV and UAV units.
– UGV, – Obstacles

that depend on subsets of variables [21]. A factor graphs consists of

- variables x_i,
- local functions f_j,
- variable nodes $\forall x_i$,
- factor nodes $\forall f_j$,
- edges $e_{i,j}$ connecting $\forall x_i \wedge f_j \; iff \; x_i = arg(f_j)$.

(2.1), (2.5) and (3.3) show examples of factor graphs. One of the more important aspects is the *sum-product update rule* stating:

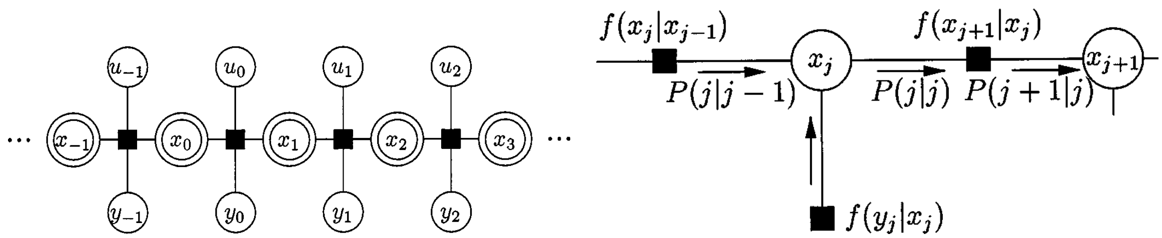
$$g(x_1, x_2, x_3, x_4, x_5) = f_A(x_1) f_B(x_2) f_C(x_1, x_2, x_3) f_D(x_3, x_4) f_E(x_3, x_5) \tag{2.3}$$



(a) Factor graph for the function (2.3) (b) Factor graph for the function $x(y+z)$

Figure 2.5: Factor graphs examples [21].

For the specific application of simultaneous mapping and localization, the factor graphs can be used as a map representation. Kschischang et al. [21] goes over more complex aspects, like probabilistic systems modelling, ergo collections of interacting variables, in detail. Since the factor graph can be used for representation of a posteriori joint probability mass function of variables of the system[21], it can be used directly to solve the SLAM MAP estimation.



(a) State space model represented as a generic factor graph (b) Kalman filter excerpt representing (2.4) in form of a factor graph

Figure 2.6: Systems modelling using factor graphs [21].

$$f(x_1,\dots,x_k|y_1,\dots,y_k) = \prod j = 1kf(x_j|x_{j-1})f(y_j|x_j) \tag{2.4}$$

Data processing

Data fusion and processing

[32] Filtering Bayesian filter Kalman filter Extended Kalman filter Particle filter Smoothing

Data distribution

The computational task mostly refers to generation and optimization of the map and can be classified as shown in 2.1 [32][24].

Active perception

Active perception or sensing can be defined as an application of control theory, such that reasoning, decision making and control are used for optimization of acquisition of desired data [4].

Table 2.1: Data distribution description.

Data distribution	Characteristics
Centralized	Computational task is assigned to one predetermined unit with appropriate capabilities, that can be internal (part of the team) or external (server), processed and output is provided to relevant units.
Decentralized	Computational task is performed by a group of units that are required to provide sufficient capabilities to respond to the demand.
Distributed	Computational task is divided among units.
Undistributed	Computational task is performed by each unit on its own.

Structure from motion

Structure from motion deals with the 3 dimensional reconstruction of environment, in particular objects of interest, by using input image sets that does not necessarily have to be in sequential order of capturing.

Visual servoing

Visual servoing Chaumette Visual Servo Control I and II Derivations? Aim of VS Basics Change of frames Kinematic screw Interaction matrix Robot Jacobian 2D visual features 3D visual features

Exploration and rendezvous tasks

The process of exploration, ergo visiting unmapped or unknown places in the environment, has been in focus of robotics for a long time [6]. It can be combined with the process of exploitation, ergo revisiting places in environment. Active SLAM is a variant, where the decision making process has been integrated in such manner that the robot and data acquisition are controlled with the goal of minimizing the uncertainty of localization and created map. A suitable approach prioritizes both aspects in a balanced manner [33].

The rendezvous aspect applied to multi SLAM deals with multiple units meeting at a specified location. This can be used in decentralized SLAM approaches. At the rendezvous instance the acquired data is shared between the units.

The approach proposed by [25] is combination of exploration and rendezvous task using cost of reaching a location and reward for its characteristics. The cost of reaching the target location is proportional to the distance between the unit and the location, while the reward for reaching the target location is proportional to the estimated gain of relevant information, ergo map at the location.

Assumptions of the method include local sensing information only with a communication range limit. To overcome the latter limitation, a group of hierarchical or role-based robots starting from unknown location, can use it's explorer units to gather data, while constantly sharing information to the main data centre through relay units [18]. The approach from [25] however, deals with the rendezvous problem [31], that derived from the field of game theory and is specified as a search problem. It is defined as meeting at pre-specified time at a rendezvous location in an unknown environment without any communication. The described approach is grid-based, overcomes the range limit of communication and performs minimization of combined exploration and rendezvous time.

Mapping and exploration

For each unit the mapping process is performed in the grid-based approach. From the acquired data Hilditch's algorithm [17] is used to generate a skeleton structure similar to generalized Voronoi graph (GVG)

[10]. The main difference between the two methods is the construction, where GVG uses the edges to close the graph. The reason for using Hilditch's algorithm is that it can create a skeleton structure with a limited sensor range that each unit has. The structure of the skeleton shown in (2.7a) is stored as a graph, where the nodes, representing potential rendezvous points, are then used for exploration. In particular the described approach uses depth-first traversal. Using flood-fill algorithm [23] the adjacent nodes are determined as well as the filled area is calculated. Path planning is performed by an A* search algorithm [16].

Rendezvous

The set of meeting locations is specific for the environment, where each rendezvous place is extended from the conventional point to a finite area. For the problem statement three rendezvous strategies are proposed:

- asymmetric sequential,
- symmetric sequential,
- and exponential.

Each of the rendezvous strategy characterizes a landmark by determining it's distinctiveness measure and ranks it among other landmarks. Three ranking criteria summarized in were proposed.

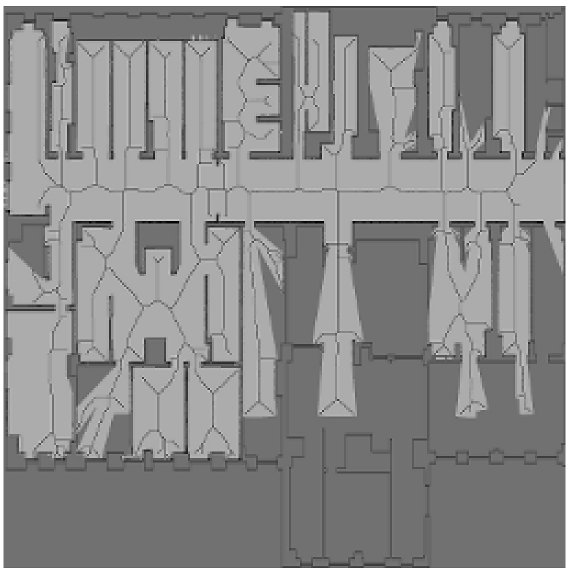
Table 2.2: Ranking criteria and formulations.

Ranking criteria	Formulation
Area	$rank(l_i^R) = area(l_i^R)$
Distance	$rank(l_i^R) = \frac{area(l_i^R)}{distance(l_i^R)}$
Sigmoid distance	$rank(l_i^R) = \frac{area(l_i^R)}{sigmoid(l_i^R)}$

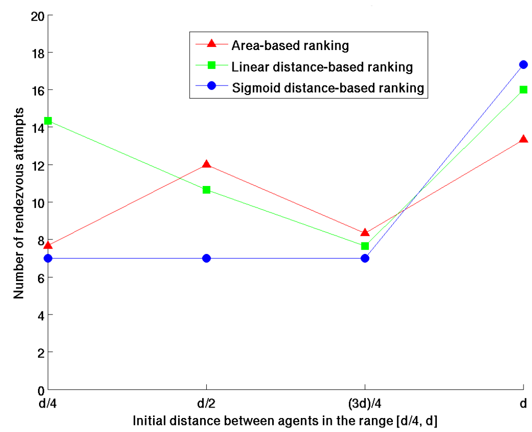
The rendezvous process itself consists of sequential attempts a_i performed at time instances $t(a_i)$ by robots $R \in 1, 2$ at a selected location l_i^R that is determined by each unit based on acquired data. If $l_i^1 = l_i^2$ at time instance $t(a_i)$, then robots 1 and 2 met at the same location, which concludes the rendezvous process, otherwise the robots continue performing previous actions until the next rendezvous time instance $t(a_{i+1})$.

Test results

The provided test results shown in (2.7b) clearly state that the sigmonoid distance based rendezvous strategy obtains best results, ergo minimal number of rendezvous attempts in minimal time for all studied scenarios.



(a) Generated skeleton structure with nodes representing potential rendezvous points.



(b) Test results for the examined rendezvous strategies.

Figure 2.7: Rendezvous and exploration approach [25].

Chapter 3

State of the art

Overview of current state of the art has been presented most recently in Cadena et al. [6], where an current developments and open problems were addressed in detail. Additional information are provided in the single SLAM – Aulinas et al. [2] and multi SLAM – Saeedi et al. [32] surveys.

Simultaneous localization and mapping

Simultaneous mapping and localization Aspects on efficient map building - key frames, high abstract formalism Multi SLAM with same type robots Open problems Multi SLAM with different type of robots

Architecture

The architecture of a typical SLAM system shown in figure (3.1). It consists of a back-end and a front-end. The input is the data from sensors and the output is usually a Maximum-a-posteriori (MAP) estimation [30], the current standard SLAM formulation [6]. There can be feedback coming from the back-end to the front-end.

Sensors

A SLAM algorithm can utilize a wide variety of sensors and tools [6][32], including these shown in 3.1

Table 3.1: List of sensors compatible with current SLAM approaches

Range	Odometry and direct localization	Vision	Sound
• Infra-red • Laser • LIDAR (Light Detection And Ranging) • Sonar	• Encoder (ground vehicles) • Differential pressure (air vehicles) • GPS – Global positioning system • IMU – Inertial measurement unit • Magnetometer • Motion capture system	• Camera – Monochrome – RGB (color) – Camera model • RGBD (color and depth) • Multi camera system • Range camera • Light field camera • Event based camera	• Acoustic • Sonar • Sodar

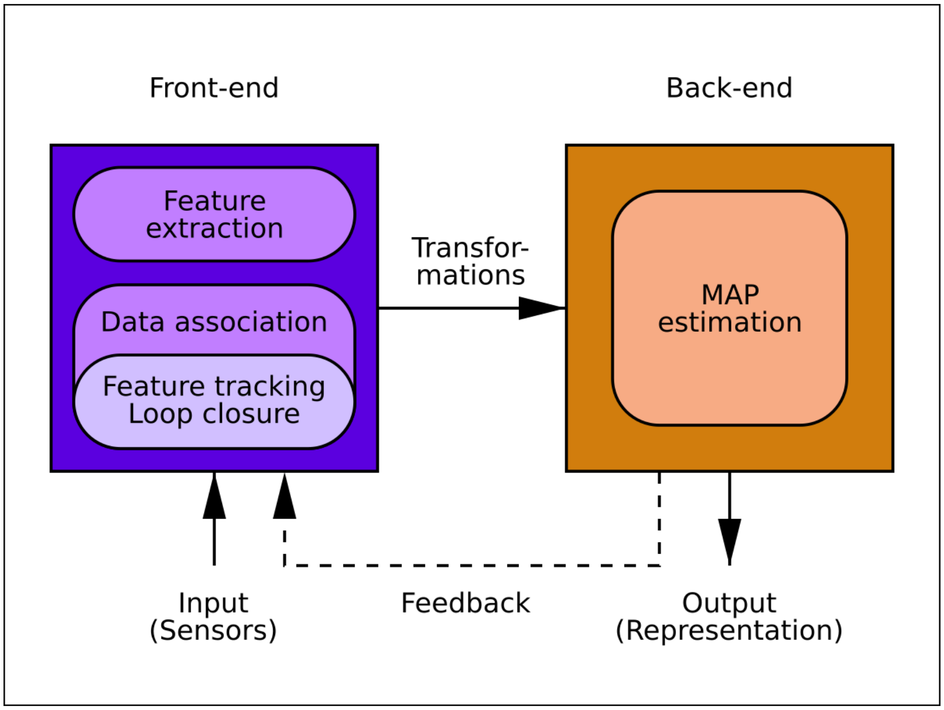


Figure 3.1: SLAM architecture consisting of a back-end and a front-end with sensors data as input and a map as output[6].

Some of these fields intersect. For example RGBD sensors use both, vision and distance from IR (infrared) registered laser reflections. Some are restricted to limited spaces, for example a motion capture system.

Representations

[32] Map representations: feature, view, appearance, polygon

[6] =Representations - Metric reasoning Landmark based

Boundary and spatial-partitioning dense representations Comparison between sparse and dense methods/feature based and direct feature based methods - depend on feature type - rely on specified thresholds for detection and matching -> ambiguity - dealing with incorrect matches - not optimized for precision + optimized for speed dense methods + in scenes with poor texture, motion blur and defocus can outperform feature based methods - require high computing power - differentiate between structure and motion not possible at the same time => To overcome these shortcomings use semi-dense using dense approach on pixels with high gradients + less computation power wrt dense methods semi-direct + most efficient (proven by [113]) (efficient in what terms?) + allow for joint estimation of structure and motion

High level object based representations Object based reasoning SLAM++, Salas-Moreno [272] Civera [71] Dame [84]

Not yet utilized solid representations Parametrized Primitive Instancing (PPI) Sweep representations Constructive solid geometry Feature based models in CAD Dictionary based models Already some experience in robotics and CV Affordance based models Generative and procedural models Scene graphs [155]

Semantic representations

Open problems Multi SLAM with different types of robots V High level expressive representations in SLAM No general and tractable framework for optimal representation choice Adaptive and automatic representations adjusting to given task and complexity of environment

Solutions

SLAM problem solutions are spread over a variety of fields. Vision is one of the most dominant tools that is used to solve the problem, laser and general range based approaches mostly concentrated around

Filters in SLAM [2] Robotic map-building can be traced back to 25 years ago, and since the 1990s proba- bilistic approaches (i.e. Kalman Filters (KF), Particle Filters (PF) and Expectation Maxi- mization (EM)) have become dominant in SLAM. The three techniques are mathematical derivations of the recursive Bayes rule [2]

Collaborative SLAM

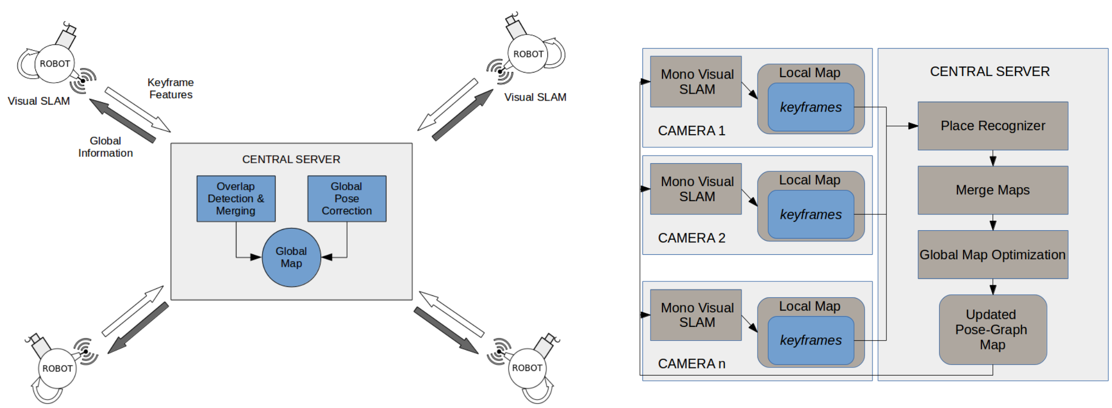
”The multi-robot exploration task is a well-addressed problem in the field of robotics.” Meghjani and Dudek [25, p. 80]

Multi robot approach is a general term for multiple units performing a task.

Collaborative Visual SLAM Framework for a Multi-Robot System

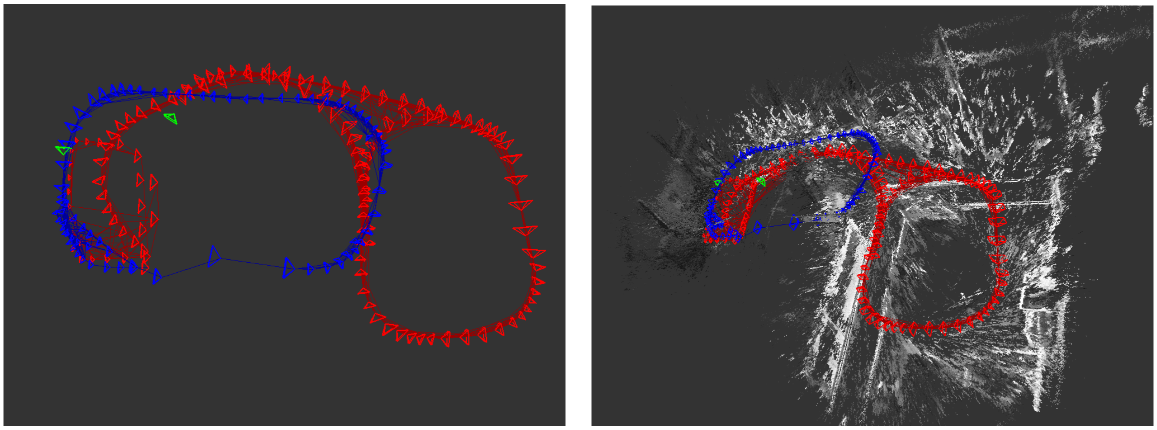
The collaborative visual SLAM is a centralized approach, where each unit equipped with a monocular vision system performs visual SLAM individually and is structured according to (3.2) [9].

The algorithm used for monocular SLAM estimates 7 degrees of freedom, including the scale of the scene using the group $sim(3)$. The environment is mapped to a pose graph of key frames that hold a depth map obtained by a semi-dense approach based on LSD-SLAM [12]. A central server runs a place recognition software that constantly monitors transferred data from mobile units (3.2a). In case of an overlap detection, that is performed in appearance space by comparison of extracted visual features from the key frames with the help of Bag of Words (BoW) technique, the map merging algorithm is executed. RANSAC version of the traditional Horn’s algorithm is utilized for the initial transformation estimate between matched key frames from the overlap detection block (3.2b). Optimization is performed using the calculated estimate as a starting point for similarity transformation estimation, that is additionally processed using and iterative closest point (ICP) algorithm. As next step map merging is conducted into a global map with an additional constraint between the two matched key frames. Finally optimization using bundle adjustment is performed on the global graph and information in form of updated poses is communicated to mobile units as feedback. The robots then perform optimization of the localization of the unit, based on the newly received and already stored information.



(a) Centralized approach architecture (b) Procedural scheme

Figure 3.2: Structure of the collaborative visual SLAM approach [9].



(a) (b)

Figure 3.3: Results of the collaborative visual SLAM approach [9].

Collaborative Visual SLAM

Work for the collaborative visual SLAM is based and a continuation of [9][8][7].

Distributed and Decentralized Cooperative Simultaneous Localization and Mapping for Dynamic and Sparse Robot Networks

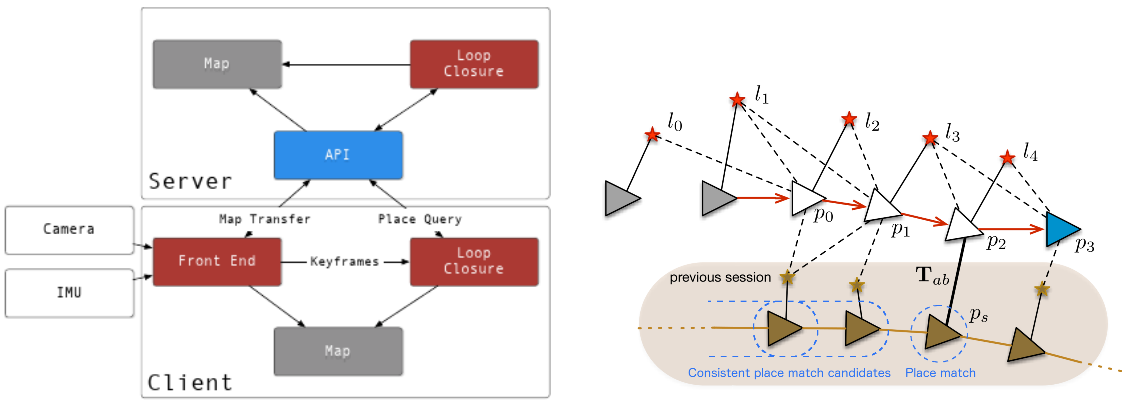
One solution of a collaborative SLAM is the distributed and decentralized approach for dynamic and sparse robot networks described in Leung et al. [24]. It defines the distributed aspect as sharing of computation load among available units and decentralized as that each unit estimates just its own state comprised of all poses and landmark positions. This approach does not focus on specific aspects of communication and provides a guarantee that all robots can recover the centralized equivalent estimate.

Scalable Multi-Device SLAM

Crowdsourced mapping from multiple devices has been proposed by Morrison et al. [28] in [28]. The algorithm can be run on cellphones, which allows an ever growing user base to cooperatively create large scale maps using the internet infrastructure. This is usually a problematic aspect with SLAM, beside the scale ambiguity problem, scaling does not always deliver satisfactory results, since errors are also scaled. Additionally each time a new information occurs the global map usually needs to be optimized in its entirety. That is of concern for maps of large scales.

The client-server architecture of the approach is shown in (3.4) is capable of simultaneously build and share large scale 3D maps from multiple monocular devices. The maps are stored on the servers and can be generated by clients in multiple sessions. The algorithm running on the cellphones has capability to fully perform SLAM, which allows the clients to work off-line and submit generated maps whenever on-line. Maps are received by the clients from the servers after submitting an query, that uses images and an non complex Application Programming Interface (API), which consists of 3 requests types: queries of places, downloads and uploads. Because of purely image driven queries, accurate global location is not needed. Detected loop closures trigger map merging by sending a query to the servers, however features computed by the front-end are reused instead of recalculating them. Loop consistency is checked by using the map graph.

The map representation introduced by Mei et al. [26] consists of relative pose-graph with metric landmark information. Nodes store key frames, which hold landmarks and their information, edges hold the relative transformations. It used by the server and client side alike and because of its relative framework it's possible to generate large scale maps. This factor graph approach allows for limited resources usage in form of storage and operational memory consumption, which makes it possible to expect constant level long-term performance. Each session is associated by an identified by an Universally Unique IDentifier (UUID) that is then used to label all objects of the system, providing uniqueness among the whole hierarchy. Additionally information about camera calibration is stored in each session. This ensures certainty of relating to the same objects between servers and clients. One of the main characteristics of this representation is the ability to interact with parts of the map directly, without the need to transfer more than necessary information. This is due to the use of multiple frames, which make the information about a global coordinate frame obsolete. It allows for simultaneous update of the map from multiple source.



(a) System architecture outline. (b) Constraints after loop closure.

Figure 3.4

The API allows for the calls:

1. Place Query
   (a) Client provides a query in form of an image
   (b) Server searches for a match in the database
   (c) Key frame is loaded, relative position based on it is estimated
   (d) New edge is connecting the found frame and the query frame and is transmitted to the client
   (e) Access to landmarks based on the provided connection to the global map allows for localization and Breadth-First Search (BFS)
2. Map Download
   (a) Request a map section using parameters frame ID and BFS depth
   (b) Establish nearby connections to the rest of the map
   (c) Send found information, including all related objects, and the potential connections to the client
3. Map Upload
   (a) Client order all objects
   (b) Server receives information from the client
   (c) Map merging occurs if loop closure is detected among stored and received data

The loop closure system works with Oriented FAST and Rotated BRIEF (ORB) descriptor and binary Bags of Words (BoW). The LC detector is used by both, the servers and clients. Places in the environment are described with the help of BoW. The client computes features using the ORB descriptor, which ensures that there is a small number of points that show high repeatability of detection and are distributed across the image. Associated landmarks have their 3D poses estimated. Reuse of features already computed improves calculation time and reduces energy consumption. A search is started using descriptors converted into a BoW vector against the global map. The matches are filtered by using the top 100 candidates, that are grouped if their reference frames are below a threshold in the map. The group with the highest score is used. Additional information from IMU or stereo systems can be used. Loop closure procedure is initiated when minimum 3 points are close to each other as shown in (3.4b). The descriptors are compared, their relative correspondences to the 3D landmarks are computed and used in within the Perspective-n-Problem (PnP) in order to acquire relative transformation that is optimized using other measurements. If the operation was successful a new constraint in form of a new edge is added between the query and matched frames. The relative manifold structure no information is lost. All landmarks are stored and the newly found correlation is represented as a new edge. This way the approach is highly efficient computationally. Bundle adjustment or similar techniques do not take place. The performance data from experiments showed that the tracking of points of interest was far more efficient than calculating features in the image again.

The described approach establishes a working architecture for SLAM dealing with scalability. Suggested improvements include applying Realizing, Reversing and Recovering (RRR) technique to filter out incorrect associations.

Co-operative Localisation and Mapping for Multiple UAVs in Unknown Environments

[5]

CoSLAM: Collaborative Visual SLAM in Dynamic Environments

[36]

Multi-robot simultaneous localization and mapping using D-SLAM framework

[35]

Open problems

Main problems/aspects robust performance, high level understanding, resource awareness, task-driven inference.

Visual communication

Chapter 4

Proposed work

Experimental setup

Collaborative SLAM

Previous work of Nived and Nicole as basis ”Collaborative Visual SLAM Framework for a Multi-Robot System” Propose a decentralized approach on that Min 7 DOF (position, orientation, scale) Other DOFs? Introduce visual communication in aspect of rendezvous

Problem statement

Assumptions

Assumptions Start work with a pair of UGV and UAV, then expand to 3 units, then swarms of both. UGV unit is equipped with high computational power hardware and the energy consumptions aspects are of no concern/negligible UAV unit is equipped with low computational power hardware, low memory capacity and programmed to perform with relatively small amount of computations Decentralized approach assumes exchange of map information between 2 units at a time Experimental hardware might/will be different from these assumptions There will be a limit of data accumulated by drones What about UGVs?

Scenarios

Scenarios Search and rescue Map environment and look for points/features/objects of interest Map dynamic environments Might be too much here

Map characteristics

Map quality/characteristics Allows for active SLAM Performance and quality index definition Proportional to number of frames in a given area Inversely proportional speed of the unit/optical flow magnitude?

Proportional to number of features Proportional to angle/rotational sweep Potential field approach 4 levels with ability to add more, so a general approach For two robots: Map from robot A Coverage of area Performance and quality index Map from robot B Coverage of area Performance and quality index Local A-B map Coverage of area Performance and quality index Areas of interest after merger Merging offline even though it's a decentralized approach Global map Performance and quality index Estimate Potentially difficult areas/Areas of interest Determined how? NN or AI a possible approach? Motion planning based on current state of global map

Algorithm

Collection of data for specific points/features/objects of interest

Multi SLAM Merging maps Rendezvous Random or specified places Visual communication Communicate that unit is ready to merge maps Memory is full, needs optimization Emergency scenario Virtual loops

Work plan

Table 4.1 shows the proposed work plan based on milestone tasks.

Table 4.1: Work plan schedule for the master thesis.

Timeline →	2016	2017							
Milestone task ↓	**Dec**	**Jan**	**Feb**	**Mar**	**Apr**	**May**	**Jun**	**Jul**	**Aug**
State of the art									
Bibliography defence									
Milestone 1									
Milestone 2									
Milestone 3									
Mid term presentation									
Milestone 4									
Milestone 5									
Milestone 6									
Writing thesis									
Master thesis submission									
Master thesis defence									

Bibliography

[1] AIMS 2016 - Autonomous Infrastructure, Management and Security. URL http://www.aims-conference.org/2016/program.html.

[2] Josep Aulinas, Yvan R. Petillot, Joaquim Salvi, and Xavier Lladó. The SLAM problem: a survey. In *CCIA*, pages 363–371. Citeseer, 2008. URL http://citeseerx.ist.psu.edu/viewdoc/download?doi=10.1.1.163.6439&rep=rep1&type=pdf.

[3] T. Bailey and H. Durrant-Whyte. Simultaneous localisation and mapping (SLAM): part . *IEEE Robotics & Automation Magazine*, 2(13):99–110, 2006.

[4] Ruzena Bajcsy. Active perception. *Proceedings of the IEEE*, 76(8):966–1005, 1988. URL http://ieeexplore.ieee.org/xpls/abs_all.jsp?arnumber=5968.

[5] M. Bryson and S. Sukkarieh. Co-operative Localisation and Mapping for Multiple UAVs in Unknown Environments. In *2007 IEEE Aerospace Conference*, pages 1–12, March 2007. doi: 10.1109/AERO.2007.352850.

[6] Cesar Cadena, Luca Carlone, Henry Carrillo, Yasir Latif, Davide Scaramuzza, Jose Neira, Ian D. Reid, and John J. Leonard. Past, present, and future of simultaneous localization and mapping: Towards the robust-perception age. *arXiv preprint arXiv:1606.05830*, 2016. URL https://arxiv.org/abs/1606.05830.

[7] Nicole Camous. Collaborative Visual SLAM. Bibliography, École centrale de Nantes, Nantes, France, 2016.

[8] Nicole Camous. *Collaborative Visual SLAM*. Master thesis, École centrale de Nantes, Nantes, France, 2016.

[9] Nived Chebrolu, David Marquez-Gamez, and Philippe Martinet. Collaborative Visual SLAM Framework for a Multi-Robot System. 2015. URL http://ppniv15.irccyn.ec-nantes.fr/material/session1/Chebrolu/paper.pdf.

[10] Howie Choset and Joel Burdick. Sensor based planning, part ii: Incremental construction of the generalized voronoi graph. In *IEEE international conference on robotics and automation*, pages 1643–1643. INSTITUTE OF ELECTRICAL ENGINEERS INC (IEEE), 1995. URL https://www-preview.ri.cmu.edu/pub_files/pub1/choset_howie_1995_2/choset_howie_1995_2.pdf.

[11] Hugh Durrant-Whyte and Tim Bailey. Simultaneous localization and mapping: part I. *IEEE robotics & automation magazine*, 13(2):99–110, 2006. URL http://ieeexplore.ieee.org/xpls/abs_all.jsp?arnumber=1638022.

[12] Jakob Engel, Thomas Schöps, and Daniel Cremers. LSD-SLAM: Large-scale direct monocular SLAM. In *European Conference on Computer Vision*, pages 834–849. Springer, 2014. URL http://link.springer.com/chapter/10.1007/978-3-319-10605-2_54.

[13] Jakob Engel, Vladlen Koltun, and Daniel Cremers. Direct Sparse Odometry. *arXiv:1607.02565 [cs]*, July 2016. URL http://arxiv.org/abs/1607.02565. arXiv: 1607.02565.

[14] U.S.-Canada Power System Outage Task Force. Final Blackout Report - Chapters 1-3 - ch1-3.pdf, 2004. URL https://www.ferc.gov/industries/electric/indus-act/reliability/blackout/ch1-3.pdf.

[15] Patrick Gerland, Adrian E. Raftery, Hana Ševcíková, Nan Li, Danan Gu, Thomas Spoorenberg, Leontine Alkema, Bailey K. Fosdick, Jennifer Chunn, Nevena Lalic, Guiomar Bay, Thomas Buettner, Gerhard K. Heilig, and John Wilmoth. World Population Stabilization Unlikely This Century. *Science*, 346(6206): 234–237, October 2014. ISSN 0036-8075. doi: 10.1126/science.1257469. URL http://www.ncbi.nlm.nih.gov/pmc/articles/PMC4230924/.

[16] Peter E. Hart, Nils J. Nilsson, and Bertram Raphael. A formal basis for the heuristic determination of minimum cost paths. *IEEE transactions on Systems Science and Cybernetics*, 4(2):100–107, 1968. URL http://ieeexplore.ieee.org/xpls/abs_all.jsp?arnumber=4082128.

[17] J. Hilditch. *Linear skeletons from square cupboards, Machine Intelligence 4 (B. Meltzer & D. Michie, eds.), 403/420*. Edinburgh Univ. Press, 1969.

[18] J. Hoog, Stephen Cameron, Arnoud Visser, and others. Selection of rendezvous points for multi-robot exploration in dynamic environments. 2010. URL http://dare.uva.nl/record/1/324475.

[19] Michio Kaku. *Physics of the future: how science will shape human destiny and our daily lives by the year 2100*. Doubleday, New York, 2011. ISBN 978-0-385-53081-1. URL http://ebookdownload.3m.com/sites/prototypes/web/media/themes/mmm_patron/img/landing_page/step1.png. OCLC: 708581888.

[20] Nikolai S. Kardashev. Transmission of Information by Extraterrestrial Civilizations. *Soviet Astronomy*, 8:217, 1964. URL http://articles.adsabs.harvard.edu/full/1964SvA.....8..217K/0000219.000.html.

[21] Frank R. Kschischang, Brendan J. Frey, and H.-A. Loeliger. Factor graphs and the sum-product algorithm. *IEEE Transactions on information theory*, 47(2):498–519, 2001. URL http://ieeexplore.ieee.org/xpls/abs_all.jsp?arnumber=910572.

[22] Ray Kurzweil. *The singularity is near: when humans transcend biology*. Viking, New York, 2005. ISBN 978-0-670-03384-3.

[23] Jean-Claude Latombe. Robot Motion Planning (The Kluwer International Series in Engineering and Computer Science). 1990. URL http://www.citeulike.org/group/5970/article/2782334.

[24] Keith YK Leung, Timothy D. Barfoot, and Hugh HT Liu. Distributed and decentralized cooperative simultaneous localization and mapping for dynamic and sparse robot networks. In *Robotics and Automation (ICRA), 2011 IEEE International Conference on*, pages 3841–3847. IEEE, 2011. URL http://ieeexplore.ieee.org/xpls/abs_all.jsp?arnumber=5979783.

[25] Malika Meghjani and Gregory Dudek. Combining Multi-robot Exploration and Rendezvous. pages 80–85. IEEE, May 2011. ISBN 978-1-61284-430-5. doi: 10.1109/CRV.2011.18. URL http://ieeexplore.ieee.org/document/5957545/.

[26] Christopher Mei, Gabe Sibley, Mark Cummins, Paul Newman, and Ian Reid. RSLAM: A system for large-scale mapping in constant-time using stereo. *International journal of computer vision*, 94(2): 198–214, 2011. URL http://link.springer.com/article/10.1007/s11263-010-0361-7.

[27] Nathan Michael, Shaojie Shen, Kartik Mohta, Yash Mulgaonkar, Vijay Kumar, Keiji Nagatani, Yoshito Okada, Seiga Kiribayashi, Kazuki Otake, Kazuya Yoshida, Kazunori Ohno, Eijiro Takeuchi, and Satoshi Tadokoro. Collaborative mapping of an earthquake-damaged building via ground and aerial robots.

Journal of Field Robotics, 29(5):832–841, September 2012. ISSN 15564959. doi: 10.1002/rob.21436. URL http://doi.wiley.com/10.1002/rob.21436.

[28] Jack Morrison, Dorian Gálvez-López, and Gabe Sibley. Scalable Multi-Device SLAM. 2014. URL http://doriangalvez.com/papers/MorrisonRSS14.pdf.

[29] Denis Oberkampf, Daniel F. DeMenthon, and Larry S. Davis. Iterative pose estimation using coplanar points. In *Computer Vision and Pattern Recognition, 1993. Proceedings CVPR'93., 1993 IEEE Computer Society Conference on*, pages 626–627. IEEE, 1993. URL http://ieeexplore.ieee.org/xpls/abs_all.jsp?arnumber=341055.

[30] Hossein Pishro-Nik. Introduction to probability, statistics and random processes. 2014. URL http://scholarworks.umass.edu/ece_ed_materials/1/.

[31] Nicholas Roy and Gregory Dudek. Collaborative robot exploration and rendezvous: Algorithms, performance bounds and observations. *Autonomous Robots*, 11(2):117–136, 2001. URL http://link.springer.com/article/10.1023/A:1011219024159.

[32] Sajad Saeedi, Michael Trentini, Mae Seto, and Howard Li. Multiple-Robot Simultaneous Localization and Mapping: A Review. *Journal of Field Robotics*, 33(1):3–46, 2016. URL http://onlinelibrary.wiley.com/doi/10.1002/rob.21620/full.

[33] Sebastian B. Thrun and Knut Möller. Active exploration in dynamic environments. In *Advances in neural information processing systems*, pages 531–538, 1992. URL http://papers.nips.cc/paper/573-active-exploration-in-dynamic-environments. 00150.

[34] Department of Economic and Social Affairs United United Nations, Population Division. *World population prospects: The 2015 revision*. ES Affairs United Nations, NY, USA, 2015. URL https://esa.un.org/unpd/wpp/.

[35] Zhan Wang, Shoudong Huang, and Gamini Dissanayake. Multi-robot simultaneous localization and mapping using d-slam framework. In *Intelligent Sensors, Sensor Networks and Information, 2007. ISSNIP 2007. 3rd International Conference on*, pages 317–322. IEEE, 2007. URL http://ieeexplore.ieee.org/xpls/abs_all.jsp?arnumber=4496863.

[36] Danping Zou and Ping Tan. Coslam: Collaborative visual slam in dynamic environments. *IEEE transactions on pattern analysis and machine intelligence*, 35(2):354–366, 2013. URL http://ieeexplore.ieee.org/xpls/abs_all.jsp?arnumber=6193110.