Hybrid Indoor Localization Based on Ranges and Video

Maria Beatriz Alves de Sousa Quintino Ferreira

Thesis to obtain the Master of Science Degree in

Electrical and Computer Engineering

Supervisors: Prof. Doctor João Pedro Castilho Pereira Santos Gomes
Prof. Doctor João Paulo Salgado Arriscado Costeira

Examination Committee

Chairperson: Prof. Doctor João Fernando Cardoso Silva Sequeira
Supervisor: Prof. Doctor João Pedro Castilho Pereira Santos Gomes
Member of the Committee: Prof. Doctor José Manuel Bioucas Dias

October 2014

Acknowledgments

This document is the result of more than six months of commitment and hard-work in order to solve the challenging posed problem. This was a long and sometimes arduous road, which I wouldn't be able to surpass without the valuable support of a large group of people.

Firstly I genuinely thank my advisers, Professor João Pedro Gomes and Professor João Paulo Costeira, who, despite their ultra busy agendas, could always find the time to help and guide me throughout this work, reassuringly answering to my doubts and anxieties. Moreover, I feel that they believed in this work victory from the very first moment, which meant a great deal to me. From that confidence, enthusiasm and engagement expressed during the whole process, I am extremely thankful.

Secondly, my appreciation goes to my colleagues, for all the good discussions about (not only) our work and materials and experiences share, thus providing a very dynamic environment for my research. I shall personally thank to Cláudia Soares, for all the readily support given during the development of the sensor network localization method and João Carvalho for the hours spent calibrating cameras and bumping on hardware limitations and incompatibilities, among others. From the VisLab I have to thank Ricardo Nunes, who was most helpful and efficient regarding the use of the 3D printer to build the reflector modules for the cricket nodes. From the LRM people I point out my good friend Maria Braga, who had always a cheerful word for me.

This long and sometimes hard journey would not have been possible for me to cross without the underlying support provided by my parents, Beatriz and Carlos. Without your wise advice and enlightening talks about the scientific method and the right attitude to have towards a MSc thesis in Engineering, this would have been incomparably more painful. Thank you deeply for always be so effective in keeping me calm and cheering me up in the difficult moments. Furthermore, you were a fundamental help during the data acquisitions carried out in the deployed set-up. Without your help such would have been an even more onerous task to be performed solo.

Of course this journey would not have been the same without the ever present, even thought long distance, support of Nuno Diegues. Very frequently did your rational judgement comforted me and gave me strength to surpass the current frustration till the next one was in place. I am exceedingly thankful that I could rely on your assistance throughout this whole time.

To all my friends from Univerisdade de Aveiro and from Técnico I express my sincere gratitude. I shall not forget to mention my special partner for everything, Francisco Ruivo, my Técnico adventure partner Miguel Griné and the always kind-hearted colleague Diogo Miranda.

Last but not least I address my so-called second family, namely Teresa, Beatriz, Margarida, Mafalda, Inês, Mariana and Inês Carolina, to whom I must thank for their unconditional support.

Lisboa, October 14, 2014

Beatriz Quintino Ferreira

Resumo

A localização é, actualmente, um tópico de investigação essencial e premente devido às numerosas aplicações e sistemas que necessitam de conhecimento relativo à localização. Sistemas do tipo GPS (*Global Positioning System*) são actualmente a solução mais popular para o problema da localização, contudo estes não podem ser usados em ambientes interiores ou sub-aquáticos, uma vez que a propagação do sinal é bloqueada. Por esta razão, vários sistemas alternativos, baseados em distâncias, ângulos e potência de sinal têm sido propostos para localização nesses ambientes.

Nesta dissertação propõem-se dois métodos alternativos, híbridos, para paradigmas de fonte única e colaborativo, com vista à localização interior usando uma rede de sensores sem fios. Tais métodos conjugam informação de distâncias (obtida por via acústica ou rádio) e visual (adquirida por uma câmera vídeo), sendo estes dois tipos de informação usados em sinergia.

Apresenta-se o estado da arte relevante, em particular, trabalhos relativos a sistemas de localização baseados em distância, visão e, ainda, trabalhos recentes que se focam na fusão de ambos.

Neste trabalho são desenvolvidos, implementados e testados métodos de localização híbridos para espaços interiores e também um procedimento capaz de realizar uma auto-calibração entre duas redes de sensores separadas. O primeiro método proposto apresenta uma formulação conjunta, não convexa, baseada na função de máxima verosimilhança para ruído gaussiano, para informação de distância e orientação para a qual uma relaxação convexa semidefinida positiva bastante eficaz é aplicada. Uma extensão deste trabalho é a exploração do paradigma colaborativo com a proposta de um algoritmo baseado numa relaxação para discos que funde, igualmente, distâncias com direcção medida.

Mostra-se que os métodos propostos apresentam uma performance semelhante ou mesmo superior, em alguns cenários, aos métodos da literatura. Tal resultado vem sustentar a aposta em sistemas de localização híbrida de larga escala, potenciando investigação futura com vista a superar algumas das limitações e evoluir para paradigmas distribuídos e utilizando novas tecnologias.

Palavras-chave: localização, optimização não convexa, relaxação convexa, rede de sensores sem fios, visão, distância e orientação.

Abstract

Location awareness is, currently, a key and urgent research topic due to the various applications and systems that require localization information. The GPS is the most popular localization system, yet it is not available in indoor or underwater environments, since its signal is blocked. Hence, alternative systems, based on distances, angles or signal strength have been proposed to perform localization in such environments.

In this dissertation, two alternative hybrid methods for single-source and sensor network indoor localization using a wireless sensor network are presented. These methods fuse information from range (obtained acoustically or via radio) and vision (gathered by a video camera), which will be used in synergy.

The relevant state-of-the-art is presented, more specifically, works related to range-based systems (single-source or cooperative), the use of visual information, and some novel work in fusing both methods.

An indoor hybrid localization method able to perform self-calibration between two detached wireless sensor networks was fully developed, implemented and tested. The first proposed method performs a non-convex joint formulation, based on the maximum likelihood functions for range and bearing information, for which a tight convex relaxation is applied to obtain a semidefinite program. An extension to this work is the exploration of the cooperative approach, for which an algorithm based on a disk relaxation method fusing, again, both range and incident streaks measurements is devised.

It is shown, both in simulation and experimentally, that the two proposed methods have comparable performance to the state-of-the-art methods, even outperforming them in some scenarios. This result suggests that investment should be made on large-scale hybrid localization systems, fostering future research not only to surpass current limitations but also to evolve to distributed paradigms and using novel technologies.

Keywords: localization, nonconvex optimization, convex relaxation, wireless sensor networks, vision, range and orientation.

Contents

List of Figures

List of Tables

List of Abbreviations

AoA Angle of Arrival

EDM Euclidean Distance Matrix

CLORIS Collaborative Localization based on Ranges and Incident Streaks

FLORIS Fusion Localization based on Ranges and Incident Streaks

GPS Global Positioning System

LS Least-Squares

ML Maximum Likelihood

RMSE Root-Mean-Square Error

RSSI Received Signal Strength Indication

SDP Semidefinite Programming

SDR Semidefinite Relaxation

SLCP Source Localization in the Complex Plane

SLNN Source Localization with Nuclear Norm

SR-LS Squared-Range Least-Squares

SVD Singular Value Decomposition

ToA Time of Arrival

TOF Time of Flight

US Ultrasound

WSN Wireless Sensor Network

Chapter 1

Introduction

1.1 Motivation

The "Where am I" problem has always been a key issue in the field of technology, both for human mobility as well as for robots/autonomous vehicles. Currently, the most popular system used to determine the localization of a target is the Global Positioning System (GPS). With the availability of this system, there has been, in the past few years, a widespread adoption of mobile and distributed systems that need precise information on location and positioning. Nonetheless, there are several situations, such as indoors or underwater environments, in which GPS is not available and where location awareness will soon become an essential feature. In fact, accurate indoor localization has the potential to transform the way people navigate indoors in a similar way that GPS transformed the way people navigate outdoors. However, indoor or subaquatic environments present several issues such as multi-path, diffraction due to obstacles and interferences, which lead to over-meter accuracy in the majority of existing systems. Such accuracy might be insufficient for numerous applications such as robot navigation; to overcome this issue it is claimed, in [39], that the solution lies in exploring hybrid schemes. Thus, alternative systems based on distances, angles and signal strength have been proposed [7, 15].
In localization systems there is normally an agent whose position we need to determine, called the target, and a set of anchors (also called beacons) which are sensor nodes that know their coordinates *a priori*. Furthermore, in cooperative scenarios, different agents have access to different location information and will interact in order to jointly estimate their poses in a more precise way than they would do individually, thus enhancing performance.
This cooperative approach, like the single-source approach, can be used not only for humans, but also for robot or autonomous vehicles navigation, since location is essential to make control and navigation decisions.
Focusing on indoor environments, most of the proposed localization systems use only one type of measurement. However, sensor networks are becoming ubiquitous and thus it is commonplace to find different sensors (e.g. Wi-Fi, acoustic and mobile cameras) in the same scenario/space. Hence, the envisaged fusing approach has the ability to combine the strongest points of each technique, paving the way to a more accurate hybrid localization system. This work addresses the use of distances (that can be obtained acoustically or with electromagnetic signals) and visual information (gathered by a video camera) to localize a target. More specifically, range information can be easily estimated using the duration of propagation of an acoustic or electromagnetic signal (time of flight) and usually produces more robust

results for farther regions, whereas the orientation and distance information retrieved by video is more reliable in a near field situation. Therefore the complementary strengths of these two techniques make them extremely appealing to be used in synergy.

The recent hype on *Google Glass*[1] and similar gadgets brings more awareness to pose estimation (localization + orientation) using video and intensifies the motivation for this topic, namely for its use in augmented reality systems. In fact, Google is working, at the present time, on *Project Tango*[2] which aims to develop devices for "human-scale understanding of space and motion". Such new devices will try to capture the 3D motion of the device while creating a map of the environment, targeting localization, navigation and gaming in unknown outdoor/indoor places as main applications. Also, new sensors like *Structure*[3] are being developed in order to add 3D capabilities to existing mobile devices, such as smartphones, to enable similar applications to the ones targeted by *Project Tango*.

Several existing visual localization systems are based on recognizing known patterns, such as artificial markers spread throughout the world in known reference positions. These markers are also known as *tags* and should be fiduciary. However, its use is limiting, thus a method that can recognize natural features and objects is more appealing and is a possible future direction.

The main topic of this work is, therefore, how to fuse these two (heterogeneous) types of sensor information as well as doing so in an efficient way to localize targets and perform self-calibration in an indoor environment.

Regarding the use of acoustic/electromagnetic signals, the specific localization technique addressed in this work is based on the Time of Flight (TOF), which allows to estimate distances simply by measuring delay, thus being preferable for acoustic signals. An alternative technique that measures the signal power, which in ideal propagation conditions has a known mathematical relationship with distance due to signal attenuation, being able to estimate the distance is Received Signal Strength Indication (RSSI) [3]. Although RSSI measurements can be poor surrogates for range due to non-uniform propagation in real environments, it is possible to acquire large sets of data for which a position estimation can be efficiently found, based on the use of Euclidean Distance Matrix (EDM)s (which is an envisaged follow-up to the work presented in this dissertation).

The localization scenario considered throughout this work is presented in Figure 1.1. The scenario comprises wireless and visual anchors, as emphasized in Figure 1.1, in which an individual equipped with a smart device/wearable such as the *Google Glass* can take advantage of both range and pose measurements from the electromagnetic signals and from the in-built camera, respectively, to enhance its location awareness.

Figure 1.1: Envisaged application scenario for single source and network localization

[1] http://www.google.com/glass/start/ (accessed in October 2014)
[2] https://www.google.com/atap/projecttango/ (accessed in October 2014)
[3] http://structure.io/ (accessed in October 2014)

Figure 1.2 depicts the schematic representation of the proposed localization scenario for the sensor network problem, detailing the coordinate systems and transformations as well as the available range and visual pose measurements. Such scenario comprises two different types of anchors: wireless, from which range or pseudo-range measurements can be obtained; and visual, which can be any pattern or object in the environment from which a pose (rotation and translation) can be gathered. Moreover, there will be network nodes with unknown positions. In the paradigm designated as single-source localization only one node does not know its coordinates *a priori* (so-called target node), while in the sensor network paradigm (illustrated in Figure1.2) there may be more than one node with unknown position (also named sensors). In this case sensors usually exchange information among each other in order to jointly achieve a more precise localization than would be possible if they operated individually.

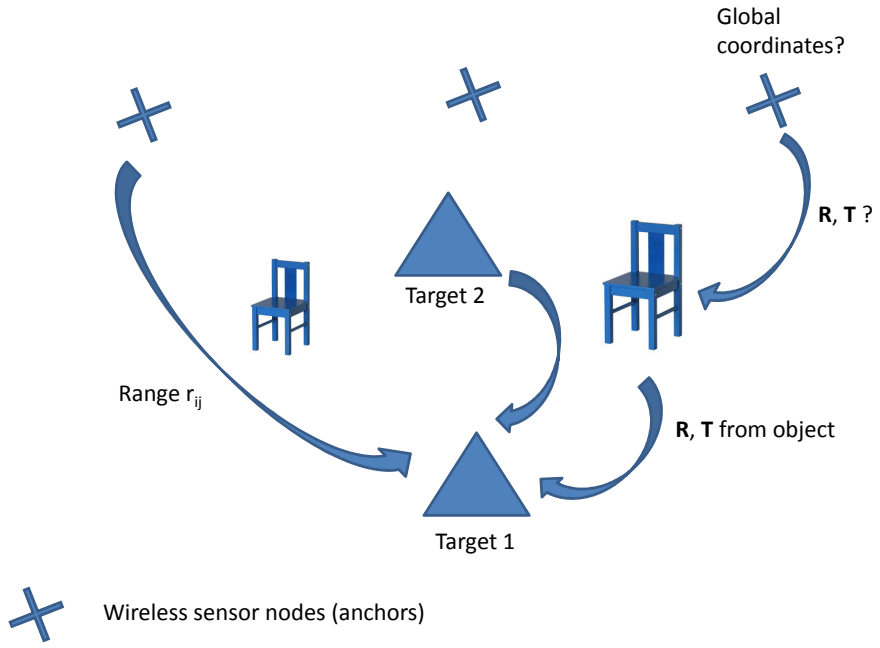

Figure 1.2: Localization scenario scheme considered in this work

This work also envisages the exploration of cooperative scenarios, for which a direct implementation in real navigation systems installed in hospitals, factories, warehouses, airports or shopping malls, is foreseen, as currently such places have already installed various sensors constituting a wireless sensor network. Considering human use, the strong present-day mobile phone usage would easily allow a generalized use of this application, since with a mobile phone one can read signal strength (from Wi-Fi), gather video/images from its camera and communicate over short ranges with low power consumption, thus enabling collaboration with other humans using their phones in the same environment.

1.2 Main contributions of this thesis

Following up with the extremely active research topic of localization (namely indoors), this thesis presents a novel method for single source position estimation based in fusing range and bearing information from image. For this method named Fusion Localization based on Ranges and Incident Streaks (FLORIS) a positive semidefinite relaxation for the optimization problem is derived as well as a self-calibration procedure. The latter procedure stems from the existence of two detached sensor networks (wireless and visual) which must be referred to the same coordinate system. Such set-up differs from the previous fusion methods found in the literature, in which both networks overlap. It is emphasized that

such self-calibration procedure has no precedents in the author's knowledge and is capable of calibrating the networks without any *a priori* information, besides the wireless anchors coordinates.
The proposed method is fully tested with numerical and practical experiments, revealing promising results in both cases. In fact, it is shown that the new hybrid method outperforms, numerically, other localization methods, specially when the measurements are quite noisy. Although several physical setbacks were found, when tested in an experimental set-up, the majority was surpassed, paving the way to demonstrate the validity of the new method in a real scenario. Hence, the new method is fully developed, implemented and tested in this dissertation work, which in addition to the auspicious attained results embodies an important contribution.
Moreover, the work related to the single-source paradigm (FLORIS method) is included in the paper "A unified framework for Hybrid Source Localization based on ranges and video", submitted to the IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP) 2015.
After tackling the single source localization, the sensor network localization problem is also addressed. This paradigm proves to be even more challenging, as fused measurements from multiple targets must be combined so that the position estimation is enhanced. Nevertheless, a new algorithm (Collaborative Localization based on Ranges and Incident Streaks (CLORIS)) based on a disk relaxation method fusing both range and streaks measurements is devised. The latter algorithm is also tested both in simulation and in a real application scenario, achieving very good results (both in accuracy and execution time). It should be noted that in the experimental scenario, the fusion scheme outperforms the single-sensed variable approach. Hence, the introduced solution for the sensor network problem is another relevant contribution of this dissertation to this field of work.

1.3 Outline

The remainder of this dissertation is organized as follows: in Chapter 2 the state-of-the-art relevant to this work is addressed and discussed; Chapter 3 describes the approaches developed and followed in order to achieve the various proposed goals, namely introducing the new derived algorithms (single-source and cooperative) and procedures (self-calibration) to perform hybrid localization; in Chapter 4 extensive simulation and experimental results regarding performance are presented and compared with other state-of-the art methods. Finally, Chapter 5 states the conclusions of this thesis, pointing also some future research work directions and further improvements.

Chapter 2

Related Work

This Chapter provides an overview of the state-of-the-art related to the problems addressed in this thesis. The main theme is localization, and work associated with range-based positioning, vision, optimization and classification methods is discussed. The remainder of this Chapter is organized in a major Section about Localization 2.1, which is divided into Ranges and Vision approaches. Inside Range Subsection 2.1.1, the optimization based single-source localization methods are addressed and then a special focus on the use of Euclidean Distance Matrices to solve localization in a collaborative paradigm is made. Throughout the Vision Subsection 2.1.2, several methods that are found to be useful tools for localization are discussed. The last Subsection addresses the previous work found in the literature that deals with the same topic as this work, namely, fusion of sensor data to perform localization.

Notation
Throughout this document, both scalars and individual position vectors will be represented by lower-case letters. Vectors of concatenated coordinates and matrices will be denoted by boldface lower-case and upper-case, respectively. The superscript $*$ stands for the conjugate transpose and T for the transpose of the given real vector or matrix. $\mathbf{I}_m$ is the identity matrix with dimension $m \times m$ and $\mathbf{1}_m$ the vector of m ones. For symmetric matrix $\mathbf{X}$, $\mathbf{X} \succeq 0$ means that $\mathbf{X}$ is positive semidefinite.

2.1 Localization

According to Bachrach and Taylor [3], nowadays ad-hoc sensor networks present new trade-offs in system design, as they use very inexpensive nodes rather than globally accessible beacons or GPS. These sensor networks should self define a coordinate system and have enabled emerging applications such as habitat monitoring, smart building failure detection or target tracking. Additionally, GPS is not a valid alternative, for example to be used indoors since the satellite links are blocked or unreadable inside buildings [29]. There are several localization schemes, hardware architectures and methods [3, 44]; in this work only a small part shall be addressed. The diagram in Figure 2.1 shows relevant aspects in localization systems such as the network scheme, ranging methods, computational organization, localization methods and algorithms, based on [3, 35] .
Concerning the hardware implementation, an architecture with anchors (also called beacon nodes), which are simple sensor nodes, is recurrent. These anchors can define relative physical coordinates, known *a*

priori, as they represent an arbitrary rigid transformation - rotation and translation - separately from the global coordinate system.

Localization methods typically comprise two phases [45]: a first phase called *measurement phase*, in which the nodes gather information, communicating with neighbouring nodes (a variety of variables can be used - some of them are addressed in Figure 2.1); and the second phase is the *localization phase*, where the localization itself is performed, as nodes infer their positions from an algorithm whose inputs are the measurements performed and also, in the cooperative paradigm, the state information received from other nodes.

At an algorithmic level, there are multiple approaches for each paradigm, both heuristic and formal [35]. Although there are beaconless algorithms that suffice when the use of nodes is prohibitive [3], this work will be focusing on algorithms based on beacons, which determine location once the distance or angle between the unknown node and the beacon is measured. According to Liberti et al. [26] "the problem of determining the sensor positions using these data was deemed as an important one from the very inception of wireless networks". Hence, this work is specially focused on range-based localization methods (see Figure 2.1), more precisely, using optimization models such as Least-Square (LS) or Semidefinite Programming (SDP).

Among the several existing solutions to ranging, due to its simplicity and reduced cost, the Cricket location system will be used. It uses a clever scheme with mixed radio frequency and acoustic transmissions to measure one-way travel times (hence distances) between nodes without the need for time synchronization of the transmitter and receiver clocks.

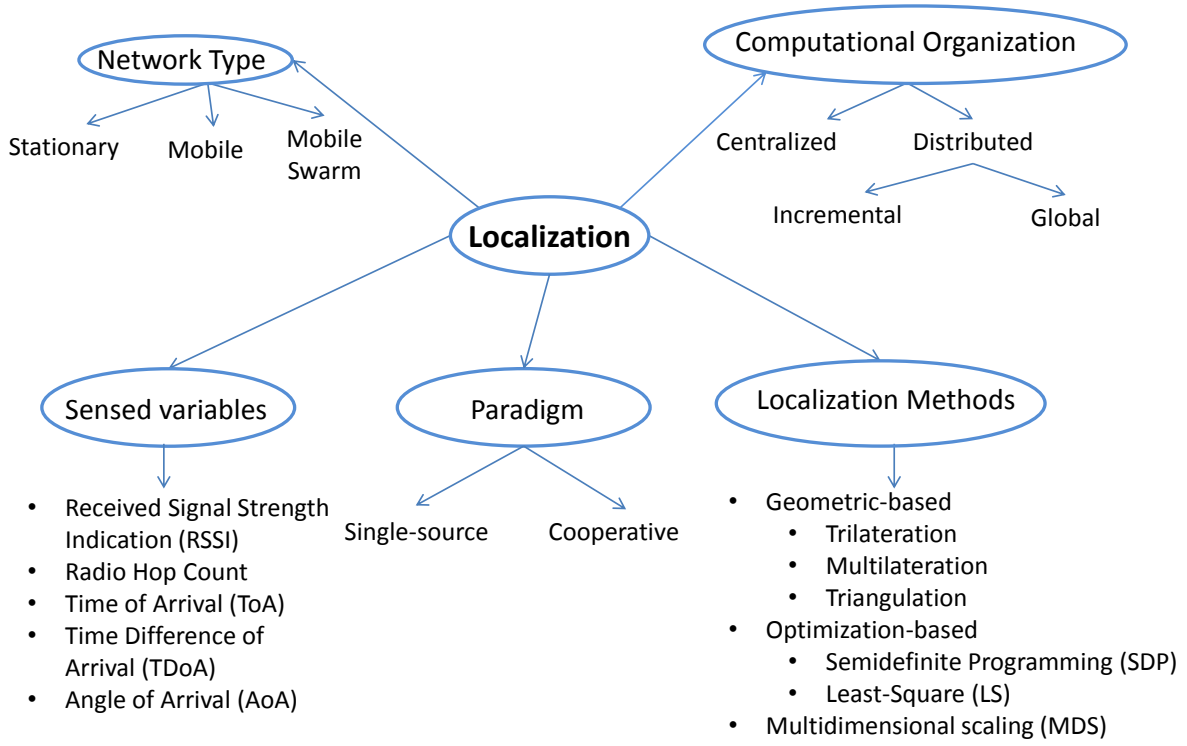

Figure 2.1: Diagram summarizing some of the most important concepts involved in localization

Referring again to Figure 2.1, mobile networks can be further divided into three types [44]. This work

will consider the mobile scheme type where the unknown nodes are moving while the anchors are static. Furthermore, in a non-cooperative sensor network, source nodes (or targets) can communicate only with anchor nodes, whereas in a cooperative localization paradigm source nodes are able to communicate with both anchor nodes and other targets [31], improving the accuracy of location information [45].
A few algorithms used for localization for stationary networks are [44]: Simultaneous Localization and Synchronization (L-S), Large-Scale Hierarchical Localization (LSHL) and Three-Dimensional Underwater Localization (3DUL), where the last two are applied to underwater scenarios. Regarding mobile networks, algorithms that rely on prior information, allowing localization to benefit from knowledge of the dynamics for improved efficiency and accuracy, are frequently used [44], such as the Monte Carlo Localization (MCL), Scalable Localization with Mobility Prediction (SLMP), Simultaneous Localization and Environmental Mapping (SLAM). These prediction based algorithms determine the current location by integrating the prediction from the prior information (prediction phase) with the measurement update from the environment observation (update phase). Filtering methods such as the Extended Kalman Filter (EKF) or Markov methods are usually applied [44]. As for mobile swarm networks, a scheme in which both unknown nodes and beacons are moving (usual in ocean environments) presents several advantages like enhanced coverage and flexibility, efficiency and suitability for cooperation [44]. Some of the algorithms applied are: Motion-Aware Self Localization (MASL), Collaborative Localization (CL) and also Monte Carlo Localization (MCL).
Finally, included in the category of centralized algorithms we may find SDP or MDS [3], whilst in the distributed case algorithms such as Diffusion, Bounding Box or Gradient are used [3]. The design of such localization algorithms has commonly constraints in resources [3], since usually nodes have to be small and cheap. Such restraints have lead to the development of systems characterized by weak processing, sensing and communication capabilities.
In the following Subsections, some algorithms resorting to both distance measurements between nearby sensor nodes and pose estimation of a target will be addressed in further detail. More specifically, we will focus on optimization based methods rather than ad-hoc approaches. The previous methods are intrinsically more robust when dealing with noisy measurements, which is frequent in the low-cost ranging systems that are of interest to this work.

2.1.1 Range

One of the more commonly used types of measurement is range. The measurement phase suffers from several impairments such as noise, multipath blockages, signal interferences, clock drifts and other environmental effects [45]. Thus choosing the adequate subjacent measurement technique is pivotal.
Range can be directly obtained from Time of Arrival (ToA) and Time Difference of Arrival, which comprise two of the most straightforward methods to sense this variable. In fact, the ranges obtained in the experimental part of this work are computed through the ToA of electromagnetic and acoustic signals, knowing the propagation velocities.
Among the other techniques to compute range information, RSSI is based on the way that the energy of a radio signal decreases with distance (following an inverse square law in free space) from the source [3]. Thus, RSSI measurement model is a function of the transmit power of the source node. Knowing this function of magnitude that depends on the distance and the reported power transmitted at the origin, a node listening to a radio transmission should be able to use the strength of the received signal to calculate the range (from the transmitter). Yet, RSSI shows, in practice, some disadvantages, because its ranging measurements are quite noisy (on the order of several meters) leading to uncertainty, as radio propagation tends to be highly non-uniform in real environments (walls and objects reflect and absorb radio waves).

Nonetheless, RSSI is an attractive method mainly because of its low hardware complexity and cost [31].

2.1.1.1 Single-Source Localization Problems - optimization based methods

After range information has been collected, source localization algorithms are invoked to compute the actual position estimates. Within this sort of algorithms, approaches based on the Maximum Likelihood (ML) location estimation are commonly followed. Although the resultant optimization problem is nonconvex, which is problematic since the presence of the local minima makes the search of the global optimal solution hard, several works present in the literature provide exact or fine approximated solutions to this problem, resorting to Semidefinite Relaxation (SDR), namely [13, 5, 34, 16].
Beck et al. [5] use a Least-Squares (LS) approach for locating a radiating source from either range measurements (creating the method R-LS), and range-difference measurements (RD-LS) collected from a network of passive sensors. Theses methods are shown to outperform the seminal work in [13]. Another approach, named SRD-LS applies the same method to the squared range and squared range-difference observations. The authors focused on an efficient way to compute the LS estimates of the source's coordinates by using an assumption that simplifies the problem. The source localization from range measurement problem in [5] is formulated as follows: considering m sensors, $a_i \in \mathbb{R}^n$ denoting the coordinates of the ith sensor and letting $x \in \mathbb{R}^n$ be the source's location to be estimated. The noisy observations between the source and the ith sensor are determined by equation (2.1):

$$r_i = \|x - a_i\| + \varepsilon_i, \quad i = 1, ..., m \tag{2.1}$$

assuming $r_i > 0$, where $\varepsilon = (\varepsilon_1, ..., \varepsilon_m)^T$ is the unknown noise vector.
The R-LS approach to estimate x is via the minimization of the LS criterion (in equation (2.2)), and its solution is the ML estimator for Gaussian noise:

$$\underset{x}{\text{minimize}} \quad \sum_{i=1}^{m} (r_i - \|x - a_i\|)^2. \tag{2.2}$$

The problem represented by equation (2.2) is nonconvex, hence finding its exact solution is a difficult task. A possible simpler solution is to construct a SDR based on (2.2). The obtained SDR for this problem in [5] can be efficiently solved via, for instance, interior point methods. Although it is not guaranteed that this SDR has the same optimal solution as the original R-LS problems, numerical results show that this SDR can provide an "accurate approximation". Indeed, it is proven, in [5], a Lemma stating that, for any optimal solution of the SDR, a rank one condition must be verified for both matrices of this formulation. However this condition might not be verified, leading this relaxation to poor results.
Applying the LS methodology to the squared range measurements results in the Squared-Range Least-Squares (SR-LS) estimate, which is the solution to equation (2.3).

$$\underset{x}{\text{minimize}} \quad \sum_{i=1}^{m} (r_i^2 - \|x - a_i\|^2)^2. \tag{2.3}$$

Like the R-LS case, the SR-LS approach is nonconvex, however, Beck et al. show that a global solution of (2.3) can be efficiently found by transforming it into a constrained minimization problem. In this last formulation (that belongs to a class that minimizes a quadratic function subject to a single quadratic constraint) the least-square problem is solved making possible the computation of the exact solution of the SR-LS.

In [23], Gomes et al. consider a weighted variant of the SR-LS cost function that allows differentiating the wavefronts (direct or reflected) when computing the source location in an underwater environment. Comparing both methods, R-LS seems to perform better than SR-LS estimate, but no exact method is known to efficiently calculate R-LS solution. Thus, as even the approximate ML-SDR solution can be less accurate than the SR-LS, the latter should be preferred (this article provides a computationally efficient algorithm for this approach).

Supposing now that the range-differences between each sensor i and an additional sensor 0 located at the origin are measured, [5] proposes the SRD-LS estimate, which, likewise SR-LS is not optimal. Yet, an efficient algorithm can be devised to find the global solution of this problem, based on the special structure of the achieved formulation (minimization of a quadratic function subject to two quadratic constraints). The SRD-LS solution also outperforms the R-LS relaxation approach, and it should be used, similarly to the SR-LS, in this type of problems.

More recently, Oğuz-Ekim et al. [33, 34] have been focusing on these optimization-based methods, taking advantage of the optimality properties of ML estimates to increase robustness to noisy range measurements. In [33] and [34] centralized algorithms termed Source Localization in the Complex Plane (SLCP) and Source Localization with Nuclear Norm (SLNN) for range-based source localization under Gaussian noise through ML-SDR approach are proposed. SLCP reformulates the classical range-based in an equivalent problem of determining a set of angles for the source-anchors directions. In that work, the framework of SLCP is expanded to arbitrary dimensions (in which 3D is the relevant case for indoor localization systems - consisting in the SLNN algorithm). Moreover, the relaxation of the nonconvex problem in SLCP is "tighter" (ratio of the first to the second eigenvalues is higher) than previous ones, yielding a matrix solution that will verify more often the rank-1 condition, thus outperforming the aforementioned R-LS. The squaring of the ranges in the cost function leads to degradation in the SR-LS. Regarding the SLNN, as an extension to higher dimensions than 2D, it is shown that the nuclear norm (which equals the sum of the singular values of a matrix) is the appropriate norm to consider in the objective function of this optimization problem. As in the SLCP case, an SDR is obtained by dropping the rank constraint.

A new approach termed $SL-l_1$ brings robustness to outlier ranges modulated as Laplacian variables, and performs well even for non-Laplacian distributions, due to the usage of the l_1 norm instead of l_2, which reduces large discrepancies between predicted and measured ranges (de-emphasizes the contribution of observations corrupted by large noise values). Two derived methods from $SL-l_1$ are further proposed: $SL-l_1$ for multiple dimensions ($SL-l_1$ MD) that outperforms benchmarked algorithms and a simplified formulation $SL-l_1$ SD that is less complex computationally at the expense of a little worse performance. The formulation of these methods conducted in [34], under independent and identically distributed noise (Gaussian or Laplacian), using the same notation of equation (2.1) is in equation (2.4).

$$\underset{x}{\text{minimize}} \quad \sum_{i=1}^{m} (r_i^p - \|x - a_i\|^p)^q \tag{2.4}$$

The SLCP and SLNN algorithms solve (2.4) for the case with $p = 1$ and $q = 2$ (the likelihood function for Gaussian noise), while $SL-l_1$ is used when $p = 1$ and $q = 1$ (the likelihood function for Laplacian noise) [23]. Comparing the case for both $p = 2$ and $q = 2$ with equation (2.3), one can notice that it is equivalent of the formulation of SR-LS by [5].

There is an additional issue posed when $q = 1$, since in such case the cost function defined by (2.4) is non-differentiable. This is surpassed transforming this cost function into a weighted sum version of the Gaussian log-likelihood. The performance of these methods was assessed using the Root-Mean-Square Error (RMSE) and comparing it to benchmark algorithms. The resultant RMSE of both the proposed algorithms under Gaussian noise (SLNN) and Laplacian noise ($SL-l_1-MD$ and SD) were lower than

all the existing methods, the latter specially in the presence of outliers.
The convexity and tightness of SLCP is also evaluated, and it is found out, by simulation results, that the convexity as well as the tightness of SLCP is superior to the previous SDR. Results seem to indicate that this convexity behaviour is mostly because SLCP has a higher probability of having a solution with rank near 1.

2.1.1.2 Cooperative Localization Paradigm - optimization based methods

A complete self-calibration is necessary to know, precisely, the positions of all the elements of a network (sensors and sources), since for large networks it is often impossible to place the sensors in predefined positions, with precision (sensors may be randomly distributed or its position may vary in time). Hence, in contrast to the previous approaches that were single-source, in this Subsection, the main methods proposed in literature under the collaborative paradigm are discussed.
Similarly to the single-source case, centralized methods which estimate all the network positions based on the minimization of the LS criterion, and also on its weighted version were proposed in [40] and [16], respectively. Later, methods relaxing the original nonconvex problem applying a convex SDP relaxation were also developed. Regarding such methods, a special focus is made on the work in [32], under the localization methods which use EDMs, in Section 2.1.1.4. However, it has been shown that these SDP relaxations generate large optimization problems, which, for large scale networks, makes the problem intractable.
Therefore, with the prosperity of large sensor networks, the adoption of a distributed approach (in which every node performs a similar processing contributing equally to the final result), contrasting with the former centralized algorithms, becomes extremely important. Such approach solves, in addition, limitations related to privacy or safety issues, enabling, at the same time, the method scalability and failure resilience [43]. Despite these facts, the majority of the methods found in literature still follow a centralized architecture, yet distributed algorithms have been emerging.
The distributed methods can be further divided into two subsets: initialization dependent and initialization independent. The first set comprises methods which directly solve the original nonconvex problem, as the work in [12]. In the latter, there is an intermediate convex relaxation step, with a solution that must approximate the global problem solution, regardless of the given initialization. Such approach separated from the initialization includes both sequential, as in [41], and parallel methods, as in [42] and [43], showing convergence guarantees.
The parallel method proposed in [43] outperforms previous works [32, 42, 22] not just in accuracy but also in efficiency, since there is a substantially lower complexity owing to a smaller number of inter-node communications, introducing a faster and more scalable algorithm to solve the distributed sensor network localization. The sensor network, in [43], is represented as an undirected graph with both sensors (unknown positions) and anchors (known positions) as nodes. The edges of the graph are the noisy range measurements available between sensors and between each sensor and a subset of the anchors. This new method minimizes the convex underestimator of the maximum likelihood nonconvex cost function for the sensor network localization problem (in (2.5)), resorting to the convex envelopes of its parcels.

$$\underset{x}{\text{minimize}} \quad \sum_{i\ j} \frac{1}{2}(\|x_i - x_j\| - d_{ij})^2 + \sum_i \sum_{k \in A_i} \frac{1}{2}(\|x_i - a_k\| - r_{ik})^2 \tag{2.5}$$

Where, in (2.5) x denotes the sensors positions and a the anchors positions, similarly to the single source scenario. As for d_{ij} and r_{ik}, they denote the noisy range measurements performed between pairs of two sensors and pairs of one sensor and one anchor, respectively, passed by collaborative messages through

neighbouring nodes.

2.1.1.3 Euclidean Geometry and Euclidean Distance Matrices

To better tackle the hereinafter presented works, which resort to EDMs to solve localization problems, one should have a solid knowledge base of convex optimization.
Consider a set of vectors $a_1, ..., a_n$ in $\mathbb{R}^n$, such that their known Euclidean lengths are: $l_1 = \|a_1\|_2, ..., l_n = \|a_n\|_2$. The Gram matrix associated to the vectors $a_1, ..., a_n$ is given by, according to the literature [9], $\mathbf{G} = \mathbf{A}^T\mathbf{A}$, with $\mathbf{A} = [a_1, ..., a_n]$ and it expresses lengths, distances and angles (that are invariant under orthogonal transformations). Furthermore, $\mathbf{G}_{ij} = a_i^T a_j$ and the diagonal entries of $\mathbf{G}$ are equal to l_i^2, $i = 1, ..., n$. The Gram matrix is, thus, symmetric and positive semidefinite. These characteristics can be used to express several geometric problems as convex optimization problems, with $\mathbf{G}$ as the optimization variable. The realizability of such problems imposes $\mathbf{G} \succeq 0$ and $\mathbf{G}_{ii} = l_i^2$.
The distance between two vectors a_i and a_j is $d_{ij} = \|a_i - a_j\|_2 = (l_1^2 + l_j^2 - 2a_i^T a_j)^{1/2} = (l_1^2 + l_j^2 - \mathbf{G}_{ij})^{1/2}$. In Euclidean Distance (ED) problems the only concern are the distances, d_{ij} between the vectors, while the lengths and the angles are ignored [9]. The distances are invariant not only under orthogonal transformations but also translations. A matrix $\mathbf{D}$ is an EDM if and only if its entries are the squares of the Euclidean distances between any vectors of a configuration. Moreover, an EDM $\mathbf{D}$ is characterized by being negative semidefinite on $\mathbf{1}^T$, i.e., $\mathbf{u}^T\mathbf{D}\mathbf{u} \preceq 0$ for all $\mathbf{u}$ with $\mathbf{1}^T\mathbf{u} = 0$, and verify $\mathbf{D}_{ij} \geq 0$ and $\mathbf{D}_{ii} = 0$. The previous conditions form a set of linear equalities, inequalities and a matrix inequality in $\mathbf{D}$. Therefore, any ED problem that is convex in the squared distances can be formulated as a convex problem with variable $\mathbf{D}$ [9].

2.1.1.4 Euclidean Distance Matrices in Localization Problems

The following works [15, 23, 31, 32] adopt the cooperative paradigm exploring EDMs. The first takes a LS approach while the latter resort to the EDM completion problem.
An approach resorting to Euclidean Distance Matrices (matrix $\mathbf{D}$ defined in (2.7) is an EDM), that shows that exists an approximate solution to the network self calibration problem by transforming the original nonlinear LS cost function minimization in (2.3) in a bilinear matrix formulation, is proposed in [15]. This algorithm performs automatic 3D localization of a set of sensors in an unknown environment and can deal, additionally, with the highly probable case of missing data. To the knowledge of the authors in [15], the previous calibration methods did not present any procedure to handle with missing data. The 3D positions of both sources and anchors are jointly estimated given a set of time of flight (TOF) of the transmissions generated at the sources. Any kind of assumption is made on the sensor localization. The only assumption is that the time of the emission events at the sources must be known (sensors are synchronized). The events considered may be both acoustic or electromagnetic, since they are signal propagating in the space as a spherical wave originated in the sources, thus such approach regard various applications.
In a network with M sources and N sensors, the measured TOF (between emission time of event j and arrival time at sensor i of the same event) can be expressed according to (2.6):

$$t_{i,j} = c^{-1}\|x_i - a_j\| + n_{i,j}, \tag{2.6}$$

where c is the signal velocity propagation and $n_{i,j}$ is an i.i.d. Gaussian random variable representing the error in the measurements. The estimated distance between i and j is, consequently, $d_{i,j} = ct_{i,j}$. These

estimated distances ($d_{i,j}$) can be organized in a $N \times M$ matrix as follows (2.7):

$$\mathbf{D} = \begin{bmatrix} d_{1,1} & d_{1,2} & ... & d_{1,M} \\ d_{2,1} & d_{2,2} & ... & d_{2,M} \\ ... & & & ... \\ d_{N,1} & d_{N,2} & ... & d_{N,M} \end{bmatrix}. \tag{2.7}$$

Following this notation, the maximum likelihood estimation of $\mathbf{x}$ (which is the sensor location matrix, with the coordinates of the ith sensor denoted as (x_i, y_i, z_i)) and $\mathbf{a}$ (the event source location matrix, where (a_j, b_j, c_j) are the unknown coordinates of the jth event) is given by the nonlinear LS problem [15] in (2.8):

$$\mathbf{x}^*, \mathbf{a}^* = \underset{\mathbf{x},\mathbf{a}}{\operatorname{argmin}} \sum_{i=1}^{N} \sum_{j=1}^{N} (\|x_i - a_j\| - d_{ij})^2. \tag{2.8}$$

Since (2.8) has the unwanted local minima, an algorithm in [15] is devised to find the correct solution. This algorithm has two steps and estimates the relative 3D position of the sensors providing that, for the 3D case, at least 4 sensors and 4 event sources are spread through the space, in a non coplanar way. In the first step, the number of unknowns is reduced from $3 \times (M + N)$ (M being the number of sources and N the sensors) to just 9 parameters, regardless the total number of sensors in the network; to achieve this, a Singular Value Decomposition (SVD) is employed. The second step finds those 9 parameters by solving the nonlinear least square problems (2.8) resorting to the simpler bilinear approach.
Assuming the TOF measurements with no errors, using the previous notation, in [15] a set of $M \times N$ equations are defined (2.13)

$$d_{i,j}^2 = x_i^2 + y_i^2 + z_i^2 + a_j^2 + b_j^2 + c_j^2 - 2x_i a_j - 2y_i b_j - 2z_i c_j. \tag{2.9}$$

Operations performed over this set of equations (to remove the quadratic terms) lead to a set of $(N - 1) \times (M - 1)$ equations. This resulting set can be rewritten in matrix form as in (2.10):

$$-2\hat{\mathbf{x}}\hat{\mathbf{a}}^T = \hat{\mathbf{D}}, \tag{2.10}$$

which is the sought bilinear form in the sensors and events coordinate vectors.
The $(N - 1) \times (M - 1)$ $\hat{\mathbf{D}}$ matrix verifies rank=3, for the no noise case; in the presence of noise $\hat{\mathbf{D}}$ will likely have rank higher than 3. In the latter case, only the three higher singular values will be considered in the SVD of matrix $\hat{\mathbf{D}}$. Hence,

$$\begin{aligned} &\hat{\mathbf{D}} = \mathbf{UVW} \quad \textit{and} \\ &\hat{\mathbf{x}} = \mathbf{UC}, \qquad -2\hat{\mathbf{a}}^T = \mathbf{C}^{-1}\mathbf{VW}. \end{aligned} \tag{2.11}$$

The relationships in (2.11) are verified for a matrix $\mathbf{C}$, named "mixing matrix" as it mixes the components obtained by the SVD to form the correct solution, that is 3×3 and invertible. Determining the 9 unknowns of $\mathbf{C}$ can be casted in a nonlinear LS minimization problem. Although this is a reduced and simpler problem when comparing to the original, the iterative minimization method can still get trapped into local minima. Assuming an additional constraint on the geometry of the problem (that an event position coincides with a sensor position), a closed form solution to the estimation of the components of $\mathbf{C}$ can be derived.
For the more complex case of missing data, matrices $\mathbf{D}$ and $\hat{\mathbf{D}}$ contain unknown entries, which impedes the used of SVD. This is solved using a rank constraint matrix imputation problem. Factorizing $\hat{\mathbf{D}} =$

FG and defining the variable m for the missing entries, the following optimization problem is obtained:

$$\text{minimize} \quad \|\hat{\mathbf{D}}(m) - \mathbf{FG}\|^2. \tag{2.12}$$

An iterative algorithm to solve (2.12) is presented. This algorithm is said to be general to the numerous contexts in which rank constraint matrix imputation problems are involved. Moreover, this procedure shows similarities with the *power factorization* method used for the Structure from Motion topic (as is discussed in Subsection 2.1.2 Vision). Once matrices **F** and **G** are determined, the coefficients of the mixing matrix **C** are found with the same procedure explained in the non missing data situation.

A Wireless Sensor Network Localization (WSNL) is said to be solvable if it has a unique valid realization, also known as global rigidity [26]. The network anchors play a pivotal role, as they ensure the global rigidity in $\mathbb{R}^K$, thus the set of anchors should have $K+1$ or more elements. These facts are the basis for the exploration of SDP to solve EDM completion problems, addressed in the next articles [23, 31, 32], since such relaxation is fundamental as low-dimensional EDM completion is NP-hard [24].
The established methods for determining the positions of sensor nodes and a moving target in a network (SLAT problem) are based on ML function, as described in the previous Subsection. However, such methods require an initialization with an approximate solution in order to avoid getting trapped in the local minima. Since in sensor network localization (SNL) it is frequent that the positions of several nodes have to be simultaneously computed from pairwise range measurements, Euclidean Distance Matrices provide a suitable and elegant framework for this topic [23]. Oğuz-Ekim et al. in [32] propose a centralized localization method that uses incomplete and noisy (both Gaussian and Laplacian) range measurements, and is based on a modified EDM completion problem and SDP, and not in a Bayesian framework as [20] (in which bearing information is estimated from camera images - see Subsection 2.1.2 *Vision*). Once solved the completion problem for a block of range measurements, thus finding the initial sensor and target positions, the likelihood function is iteratively improved through Majorization-Minimization (MM). Also, a scheme, in which the current target/node position is estimated from previous locations, is developed to avoid solving the larger EDMs with time. More specifically, it is proposed a two stage algorithm composed by a *startup* and *update* phases. *Startup* phase main goal is to obtain an outline of the network configuration from a batch of measurements, whereas in the *update* phase new target observations are incrementally treated as they become available, improving also all previous determined locations. Both phases include, in its turn, initialization and refinement steps. The first calculates approximate locations (in *startup* it is solved an EDM completion problem while in *update* it is solved a source localization problem) whilst the second is an iterative step using MM to refine the likelihood function. Besides solving the SLAT problem with modest prior knowledge on target/sensor positions with moderate complexity, [32] uses cost functions with plain ranges (non-squared) which improves robustness in the case of high noise. Suitable cost functions are derived for Gaussian and Laplacian noise, being more robust, in both cases, than standard EDM methods, due to the use of plain ranges. Under Gaussian noise in the *update* phase it is used the SLCP time-recursive method, under Laplacian noise the $SL-l_1$ is applied.
Focusing on the EDM problem formulation with squared ranges, a **D** matrix (denominated pre-distance matrix) with elements $\mathbf{D}_{ij} = d_{ij}^2$ and diagonal entries zero is used to find an EDM **E** (satisfying $\mathbf{E}_{ij} = \|y_i - y_j\|^2$) that is nearest to **D**, in the LS sense. Thus, for 2D, this problem is formulated

according to equation (2.13) [32]:

$$\begin{aligned} \underset{\mathbf{E}}{\text{minimize}} \quad & \sum_{i,j\in\mathcal{O}} (\mathbf{E}_{ij} - d_{ij})^2 \\ \text{subject to} \quad & \mathbf{E} \in \text{EDM}, \quad \mathbf{E}(\mathcal{A}) = \mathbf{A} \\ & rank(\mathbf{JEJ}) = 2, \end{aligned} \tag{2.13}$$

where $\mathbf{J} = (\mathbf{I}_\rho - \frac{1}{\rho}\mathbf{1}_\rho\mathbf{1}_\rho^T)$ is a centering operator and $\rho = m + n + l$ (m is the number of unknown target positions, n of unknown sensors and l of anchors) and $\mathcal{O}$ is the index set for which range measurements are available. Matrix $\mathbf{E}$ should satisfy the properties listed in the previous Subsection. And the rank constraint guarantees that the solution is compatible with a configuration of sensor/anchor/targets in $\mathbb{R}^2$. Furthermore, $\mathcal{A}$ is the index set of anchor distances and $\mathbf{A}_{ij} = \|a_i - a_j\|^2$ is the corresponding EDM submatrix that enforces the *a priori* information. A convex SDP compact relaxed formulation (since the rank constraint is ignored), in matrix notation, is presented in (2.14) and is called EDM with squared ranges

$$\begin{aligned} \underset{\mathbf{E}}{\text{minimize}} \quad & \|\mathbf{W} \odot (\mathbf{E} - \mathbf{D})\|_F^2 \\ \text{subject to} \quad & \mathbf{E} \in \text{EDM}, \quad \mathbf{E}(\mathcal{A}) = \mathbf{A}, \end{aligned} \tag{2.14}$$

where $\mathbf{W}$ is a mask matrix having zero entries in the free elements of $\mathbf{D}$.
A similar process is performed for the plain ranges case, obtaining the EDM completion formulation in (2.15)

$$\begin{aligned} \underset{\mathbf{E}}{\text{minimize}} \quad & \sum_{i,j} (\sqrt{\mathbf{E}_{ij}} - d_{ij})^2 \\ \text{subject to} \quad & \mathbf{E} \in \text{EDM}, \quad \mathbf{E}(\mathcal{A}) = \mathbf{A} \\ & rank(\mathbf{JEJ}) = 2, \end{aligned} \tag{2.15}$$

and the relaxed SDP (obtained using the epigraph technique, introducing variable $\mathbf{T}$ and dropping, again, the rank constraint) in (2.16)

$$\begin{aligned} \underset{\mathbf{E},\ \mathbf{T}}{\text{minimize}} \quad & \sum_{i,j} (\mathbf{E}_{ij} - 2\mathbf{T}_{ij} d_{ij}) \\ \text{subject to} \quad & \mathbf{T}_{ij}^2 \leq \mathbf{E}_{ij}, \\ & \mathbf{E} \in \text{EDM}, \quad \mathbf{E}(\mathcal{A}) = \mathbf{A}. \end{aligned} \tag{2.16}$$

Since the solution of these problems is a distance matrix, the Gram matrix should be obtained (factorizing **-JEJ**) to, afterwards, extract the node coordinates by SVD [24] and solving a Procrustes problem.
Simulation results show that these methods are more robust to outliers than existing ones and are run reasonably fast for small networks (about 30 unknown positions). Even further, for small Gaussian noise, the Cramér-Rao lower bound is almost achieved.
Convex formulations (2.15) and (2.16) are equally applied in the underwater scenario, by Gomes et al. in [23], for which a solution in simulation, for a network with approximately 50 nodes, is reliably found. In the experiments of [23], it was verified that the formulation (2.16) has better numerical properties than (2.14), allowing the use of a faster solver.
More recently, however, Soares et. al, in [43], propose an optimal distributed algorithm that outperforms previous methods based on SDP and on edge-based relaxations. Furthermore the algorithm in [43] also has superior performance (both in accuracy and in communication volume) than the method based in EDM completion of [32].

2.1.2 Vision

This Section includes some relevant work on localization systems based on vision, namely using a camera network that follows a target in [20], mobile cameras that only recognize markers, whose locations are known, in [29] or cameras recognizing tags in order to measure ranges and perform mapping in [38]. Feature descriptors, such as the Scale-Invariant Feature Transform (SIFT) method from [27], have been also used in mapping and localization systems without targets or markers, thus recognizing objects or natural features, in [18].
Moreover, in parallel to localization based on video images, some important topics from the field of image processing are discussed since they can be helpful tools to perform such image aided localization. Some literature on these topics will be addressed afterwards, targeting on image recognition and reconstruction in [28] and image classification in [11].

Camera networks are perhaps the most common type of sensor network nowadays. However, a key problem in deploying such networks is their calibration, i.e., determining the location and orientation of each network sensor, so that the observations in an image can be mapped into locations in the real world [20]. Furthermore, a camera pose can be represented by 6 parameters: 3 position parameters (X, Y, Z) and 3 angles (for example: roll, pitch, yaw), therefore the calibration procedure must recover these parameters. The notation (X, Y, θ), where θ is the rotation around the z-axis, is known as the absolute parametrization of a camera pose.
For a network with a single type of sensor, Funiak et al. [20] proposed an approach for camera network calibration addressed by solving a localization and tracking (SLAT) problem, where both the trajectory of a moving target and all the camera poses are estimated. In such problem, the cameras collaborate to track a moving object through the environment, reasoning probabilistically (to handle the non-linearities of the projective transformation and the dense correlations arisen between distant sensors) about which camera poses are consistent with the observed images of this environment. Therefore, with cameras placed throughout the environment at unknown locations, with an object moving accordingly to an unknown (smooth) trajectory, the network can automatically calibrate itself. The main contributions of such work are a novel representation (named relative over-parametrization - ROP) and hybrid conditional linearisation techniques that enable the use of a single Gaussian variable to a Kalman Filter solution to the SLAT problem and a stable and online algorithm that converges always to the same solution as the centralized approach (contrary to the existing similar approaches that are off-line and centralized, such as [37]). The authors claim that the performance of their solution is very good, even when the communication is lossy, as it was experimented on a real camera network with 25 nodes, with relatively low link quality between nodes (due to packet loss or interference), generating pose estimates within 90% confidence intervals. The proposed solution also obtains the respective uncertainty of each estimate.
The approach used in [20] (SLAT) is related to SLAM, since SLAT can be understood as a SLAM problem where the cameras are the landmarks (anchors) and the moving object plays the role of the robot. Some aspects of SLAT are related also to computer vision topic Structure from Motion (SFM), which consists in recovering the 3D geometry of a scene as well as the trajectory of the camera motion, making correspondences between feature points. This is in the line with the 3D scene reconstruction from thousands of images, in [1], accomplished in an very efficient way, due to the use of novel parallel distributed matching and reconstruction algorithms.
More recently, triggered by the widespread and popularity of RGB-D cameras (a sensor which provides both color images and dense depth maps), such as the Microsoft Kinect or Asus Xtion Pro, several new approaches to SLAM algorithms have been proposed. In fact, in [18], a system that simultaneously generates a dense 3D model of the surrounding environment and estimates the RGB-D camera pose (and

trajectory) with respect to this model is presented. Endres et al. claim that the approach in [18] is suitable for online operation and achieves an accuracy of 9.7cm, on average, in an RGB-D benchmarked dataset.
Another localization system based on vision is presented in [29]. This solution is designed to work indoors and is based on a low-cost system that relies only on the use of off-the-shelf camera phones to determine user location in real time, by detecting fiduciary marks, thus helping navigation. This approach opposes to other existing indoor positioning sensing technologies (ultrasonic, infrared or radio sensors), as it does not require any permanent electronic infrastructure; it only needs 2D markers spread across the environment providing a unique ID.
The usage of fiduciary markers is a well-established mechanism in mobile applications, its unnatural visual appearance makes them noticeable, easing the process of identifying information hot spots. Moreover, there are numerous marker-tracking software libraries that can estimate a camera's position and orientation with respect to markers. Yet, these markers raise questions concerning the visual clutter.
The scanning of the environment is continuous and real-time, allowing for a continuous navigation based on real-time marker tracking, that became available shortly before the publication of the article [29]. In this work [29], a controlled user study was performed to compare the proposed solution with a GPS-like localization method and the location awareness of the users was also evaluated. Such experiments revealed that such system (considered the first marker-based navigation running on the user's own mobile phones successfully deployed) works well and fits the real-world requirements, providing a substantial increase of location awareness and satisfactory tracking accuracy for the purpose (around 50cm).
The solution developed by Mulloni et al. [29] was turned into a commercial product for conference guide indoor navigation system. Such tracking based on computer vision was proven to be cost-efficient, since it only requires placing some posters with markers at the site, rather than deploying a whole active beacon system infrastructure.
In the thesis work [38], static tags are used, as well, to measure range and to perform mapping. The limitation present in this work is that the camera needs to see all the targets and, in particular, observe two tags simultaneously. Figure 2.2 shows an example of some tags spread in an indoor environment.

Figure 2.2: Tags spread in an environment to perform localization and mapping (from [38])

Beyond the use of markers, [29] reports indoor navigation systems based on the SIFT method [27]. Such systems rely on a server to outsource the actual pose estimation, providing limited scalability and

long latency (with a processing time approximately 10 seconds per image), thus making the system unsuitable for large-scale continuous navigation deployment.

Furthermore, image recognition and reconstruction could be an extremely useful tool to devise a localization method based on objects or natural features rather than unnatural tags or markers.
In this subject, Marques et al. in [28] proposed a solution to the Structure from Motion (SFM) problem for the case with degenerate data (due to rank deficiency in the measurements) using a new factorization algorithm enabling the estimation of a unique 3D shape, regardless the data degeneracies. Hence, this solution can reconstruct a 3D scene captured by a moving camera where the points are not known in every captured image and when the "full rank" requirement of the feature track matrix, to estimate missing data, is not satisfied. Prior methods would obtain, in the presence of degenerate data (one single degenerate frame) which occurs quite often in reality, reconstructed shapes with high errors [28].
This missing data problem has a similar formulation to the one presented in (2.14). In this case, though, matrix $\mathbf{E}$ has rank 4, due to the orthographic camera model assumed, and this is used to estimate the unknown data. The rigid factorization introduced in [28], which computes shapes and motion under scaled orthography in one optimization step, is similar to the power factorization, but imposes constraints on the rank and on the rigid motion, in each step. The latter constraint is essential to make the estimation of the missing data possible, as the rank condition is not enough. This algorithm is correctly solved in closed-form and the object's shape is estimated, in its last step, given by a simple LS solution.

Image classification is another issue that can be of significant importance in localization systems based on objects and natural features, since correct labels should be assigned to the regions of an image. In this line of work, Cabral et al. [11] introduce a weakly supervised system for multi-label image classification, standing out from the standard fully supervised (with manual bounding boxes or pixelwise segmentations). Identifying correctly the regions of an image and hence the objects present allows to determine the position of the camera, using the translation obtained between the camera and the known objects.
Weakly supervised algorithms learn using labels without any localization information. This article, [11], makes use of the additive property of histograms to bypass the combinatorial nature of the existing Multiple Instance Learning (MIL) methods. According to the additive property, the histogram of a whole image is a weighted sum of the histograms of all it subparts - Bag of Words. Therefore, the algorithm proposed in [11] aims to factorize the histogram of an entire image, obtaining the features plus the background. This approach enables posing this problem as a rank minimization, filling the unknown entries (in the test set) and correcting known features and labels, thus casting the weakly supervised image classification under a matrix completion framework (that copes with missing data as well as outliers). The matrix with the observations is formed by three sub-matrices including the training set, the test set and the histograms representative of each class found. Each column of this matrix has the image histogram and the respective label assignment and the entries to be estimated are the labels of the test set and the histograms in the set of the representative class histograms. Under such framework, two convex rank minimization algorithms are proposed: MC-Pos and MC-Simplex, for which the convergence is proven.
This work ([11]) stands on the important development (in 2008) of using the nuclear norm as a surrogate of rank minimization, under broad conditions. Such advance enables the relaxation of a highly nonconvex and non-differentiable function as rank using its convex envelope, estimating subspace ranks automatically. With this minimization problem a joint classification (assigning labels to features) of all the test images can be achieved, by completing the entries of the rank deficient matrix. Additionally, to avoid trivial solutions and add robustness to the matrix completion, the objective function is penalized with a loss function.
Experiments validate the low rank assumption (that histograms of objects of the same class have a com-

mon low dimensional subspace) and compare the methods presented in [11], MC-Pos and MC-Simplex, with a Support Vector Machine (SVM) baseline and diverse state-of-the-art MIL methods. The results of such experiments show that both MC-Pos and MC-Simplex perform comparably or better than the existing MIL methods, due mainly to the loss functions and feature error correction. Hence, the class representative histograms allow for class localization in a multi-label image problem. One of the limitations pointed out is that only one representative histogram is captured per class.

2.1.3 Fusing Range and Vision

Coupling different sensors can compensate for the deficiencies of a single modality alone. Hence, in [7], Biswas et. al use a combination of distance and angle information (Angle of Arrival (AoA)) to solve the problem of localization in Wireless Sensor Network (WSN). The solution presented in [7] is under the framework of Euclidean distance geometry and applies, for the first time, the technique of SDP relaxation to a mixture of sensed variables. Furthermore, this localization method is specially useful in a scenario integrating heterogeneous sensors, more specifically considering any ranging mechanism (RSSI, TOF or ultrasonic) as well as image sensor nodes.
The extension of the SDP model for the case where two types of measurements are used is achieved through trigonometric relations (only valid in $\mathbb{R}^2$) that define an optimization problem with quadratic constraints, which, in turn, can be relaxed into an SDP. Assuming m anchors a_k $(k = 1, ..., m)$ and n targets x_j $(j = 1, ..., n)$, with both a_k, $x_j \in \mathbb{R}^2$ and also let d_{kj} be the Euclidean distance measure between anchor a_k and target x_j and d_{ij} the distance between the points x_i and x_j known. Considering the case when the distance measures may be corrupted with noise, according to [7], the quadratic problem can be formulated as in (2.17)

$$\begin{aligned} \text{minimize} \quad & \sum_{i,j} |\alpha_{ij}| + \sum_{k,j} |\alpha_{kj}| \\ \text{subject to} \quad & \|x_i - x_j\|^2 = (d_{ij})^2 + \alpha_{ij} \\ & \|a_k - x_j\|^2 = (d_{kj})^2 + \alpha_{kj}, \end{aligned} \tag{2.17}$$

where the alpha factors are the added noise. Note that, so far, only distance measurements are considered, the angle integration is achieved with (2.22).
Moreover, if $\mathbf{x} = [x_1\ x_2\ ...\ x_n]$ is the $2 \times n$ matrix that is to be determined, the constraints of the latter formulation can be then defined as in (2.18)

$$\begin{aligned} & \|x_i - x_j\|^2 = \mathbf{e}_{ij}^T \mathbf{x}^T \mathbf{x} \mathbf{e}_{ij} \\ & \|a_k - x_j\|^2 = (a_k; \mathbf{e}_j)^T \begin{bmatrix} \mathbf{I} & \mathbf{x} \end{bmatrix}^T \begin{bmatrix} \mathbf{I} & \mathbf{x} \end{bmatrix} (a_k; \mathbf{e}_j), \end{aligned} \tag{2.18}$$

where $\mathbf{e}_{ij}$ is a vector with the ith position equal to 1, the jth position equal to -1 and zero everywhere else; and $\mathbf{e}_j$ is a vector of zeros except in the jth position, where it takes the value -1.
Thus, (2.17) can be set in the matrix form, in (2.20). As intermediate steps of this approach, described in [8], the matrix $\mathbf{Y}$ is defined as $\mathbf{Y} = \mathbf{x}^T\mathbf{x}$ and a relaxation to $\mathbf{Y} \succeq \mathbf{x}^T\mathbf{x}$ is performed, since, once again, the problem in nonconvex. As previously addressed, this is a rank constraint relaxation which is a key aspect in the solution of NP-hard problems under SDP. Also, a matrix linear inequality is achieved in (2.19),

$$\mathbf{Z} = \begin{bmatrix} \mathbf{I} & \mathbf{x} \\ \mathbf{x}^T & \mathbf{Y} \end{bmatrix} \succeq 0. \tag{2.19}$$

The final SDP formulation is (2.20):

$$\begin{aligned} \text{minimize} \quad & \sum_{i,j} |\alpha_{ij}| + \sum_{k,j} |\alpha_{kj}| \\ \text{subject to} \quad & (1;0;\mathbf{0})^T \mathbf{Z}(1;0;\mathbf{0}) = 1 \\ & (0;1;\mathbf{0})^T \mathbf{Z}(0;1;\mathbf{0}) = 1 \\ & (1;1;\mathbf{0})^T \mathbf{Z}(1;1;\mathbf{0}) = 2 \\ & (\mathbf{0},\mathbf{e}_{ij})^T \mathbf{Z}(\mathbf{0};\mathbf{e}_{ij}) = (d_{ij})^2 + \alpha_{ij} \\ & (a_k,\mathbf{e}_{ij})^T \mathbf{Z}(a_k;\mathbf{e}_{ij}) = (d_{kj})^2 + \alpha_{kj}, \\ & \mathbf{Z} \succeq 0. \end{aligned} \tag{2.20}$$

The AoA integration is made considering that all angle measurements from each sensor are relative to its own axis and that the orientation of each axis in the global/world coordinate system is unknown. Naturally, the field of view must be regarded and specified, depending on the type of sensing technology used. In [7] it is also considered that the measurement of the angle between 2 sensors (both anchors or unknown nodes) as seen from another sensor exists.
This work of Biswas et. al [7] presents algorithms for two cases: when both angle and range information is available and when purely the angle is available.
Regarding the first case, the absolute angle θ_{kj} between the pair anchor a_k and unknown node x_j can be expressed (assuming that all anchors know their absolute orientation in respect to the world referential) by:

$$\frac{a_k(2) - x_j(2)}{a_k(1) - x_j(1)} = \tan\theta_{kj}, \tag{2.21}$$

where $a_k(1)$ and $a_k(2)$ are the coordinates x and y, respectively, of the position of anchor a_k and $x_j(1)$ and $x_j(2)$, likewise, of the position of x_j. The previous linear constraint (2.21) can be written as in (2.22) and therefore is in an appropriate form to be reduced to an SDP formulation.

$$\mathbf{A}_{kj}\mathbf{x} = \tan\theta_{kj} \cdot a_k(1) - a_k(2) \tag{2.22}$$

In (2.22) $\mathbf{A}_{kj}$ is a $2 \times n$ matrix of all zero except in $\mathbf{A}_{jk}(1,j)$ that is equal to $\tan\theta_{kj}$ and $\mathbf{A}_{jk}(2,j)$ equal to -1; $\mathbf{x}$ is, again, the matrix of the sensor nodes positions that needs to be determined.
Additionally, combined with the distance information, the AoA between an anchor and a sensor (a_k and x_i) or 2 sensors (say x_i and x_l) as sensed by a third node (x_j) can be found, respectively, θ_{kji} and θ_{ijl}. These two previous angles can be expressed as in (2.23)

$$\begin{aligned} & \frac{(x_j - a_k)^T (x_j - x_i)}{d_{jk} d_{ji}} = \cos\theta_{kji} \\ & \Leftrightarrow \|x_i - a_k\|^2 = d_{jk}^2 + d_{ji}^2 - 2 d_{jk} d_{ji} \cdot \cos\theta_{kji}, \quad \text{and} \\ & \frac{(x_j - x_i)^T (x_j - x_l)}{d_{ji} d_{jl}} = \cos\theta_{ijl} \\ & \Leftrightarrow \|x_l - x_i\|^2 = d_{ji}^2 + d_{jl}^2 - 2 d_{ji} d_{jl} \cdot \cos\theta_{ijl}, \end{aligned} \tag{2.23}$$

where d_{ij} (d_{jl}) are the distances measured between sensors i and j (j and l). Similarly to the constraint in (2.22), the quadratic constraints in (2.23) can be relaxed and expressed in an SDP form.
When only AoA information is available, the followed approach maintain an Euclidean distance geometry 'feel', as claimed by the authors, since the quadratic term in the formulation is relaxed into an SDP. Based on a geometric property concerning circles and angles, it is presented a set of equations in the unknowns

x_i, x_j, x_k and r_{ijk}^2 (radius of the circle) for all sets of points that share angle information. Again, the devised constraints are quadratic in $\mathbf{x} = [x_1\ x_2\ ...\ x_n]$ and the SDP relaxation applied in the previous case can be equally used (that is $\mathbf{Y} \succeq \mathbf{x}^T\mathbf{x}$). Now, the problem has the unknowns $\mathbf{Y}$, $\mathbf{x}$, r_{ijk}^2 and the linear constraints of the anchors angle measurements (presented in (2.22)) can also be included. In [7] it is also shown that, despite the introduction of the unknowns r_{ijk}^2, a unique feasible solution exists for a set of equations, if an inequality depending on the number of unknowns and target sensors is verified.
Biswas et. al also raise some practical questions such as how to define the objective function when the measurements are noisy. Softer SDP models which introduce some slack are a possible alternative, and the function to be minimized in the resulting SDP can be, for instance, the absolute sum of the errors. Another objective function, defined as the sum of the squared errors is specially interesting, since the SDP formulation becomes a relaxation of a nonlinear LS problem, equivalent to SR-LS of [5] and in [15].
Other practical consideration made concerns the anchor placement when the angle information is noisy, this issue takes on critical importance. More specifically, if the anchors are placed in the interior of the network, the estimation of points outside the convex hull formed by the anchors are erroneous, as the estimated positions tend to lie in the interior. Hence, it is suggested that the anchors are placed on the network's perimeter and that such placement is a feasible assumption.
The simulations in [7] are performed for 40 nodes, with measurement noise modelled as additive, both for distances and angles and 4 anchors are placed in the corners of the network region. In these simulations the combined distance-angle and the pure AoA approaches are compared. The main drawn conclusions, further explored in [6], are that both methods can provide an accurate localization. Particularly, it is better to use both range and AoA information when the radio range and measurement noise are low (since this method uses more information) whereas the pure AoA approach performs better when the radio range and noise are high. Furthermore, it is addressed the interest and the need to study what an optimum radio range should be. Another remark done, based on the simulation results, regards the number of constraints, which grows significantly with the number of nodes and so the computational solution time. A localization method that is distributed with a constraint selection method is the given suggestion to make large networks tractable in this context.

Additionally, in networks with more than one type of sensor, and in which the simplifying assumption of [7] is not made, an accurate separate calibration for each type is normally required. Even with different set-ups, these calibrations result in the same self-calibration problem, whose goal is to simultaneously localize the sensors and the targets. In the following, the considered state-of-the-art work developed in self-calibration with different sensor fusion is presented.
A new closed-form solution to jointly calibrate heterogeneous sensors (video and range sensors) in a network is proposed by Crocco et al. in [14]. In this fusion approach (that has available range distances as well as visual information on the target position), it is used a rank constraint for both range and image data. Specifically, this rank condition allows to find an initial affine solution via bilinear factorization, as in [15], presented previously in Subsection 2.1.1.1. In this framework, both the positions of the range sensors, cameras and the target are unknown, apart from the knowledge of the position of some sensor anchors. Yet, if these anchor positions are not available, this algorithm is still able to provide a solution, using only the metric constraints from the cameras.
The range sensor calibration follows, in short, the procedure in [15], obtaining, in the end, a bilinear formulation. As each sensor is supposed to estimate the distance between itself and the target (for example, using RSSI), a matrix whose entries are the squared of the estimated distances can be built (matrix $\tilde{\mathbf{D}}$). Identically to what is introduced in [15], eliminating the quadratic terms of the equations, one can reach a set of $(M-1) \times (N-1)$ linear equations, that in matrix form is $-2\tilde{\mathbf{S}}\tilde{\mathbf{t}} = \tilde{\mathbf{D}}$. Then, a SVD is applied to this matrix bilinear formulation, once again, similarly to what is performed in [15].

The elements of a mixing matrix can be found by solving a LS problem, which gives the correct solution to the range calibration. Furthermore, if no anchor positions are available, the relative positions can still be found for a less than a 3D translation and rotation.
On the other hand, the video camera calibration consists in computing the parameters of an affine camera and the 3D position of a target, only based in image measurements. The target points ($\mathbf{t}$) are projected in the image, forming the the image measurement matrix $\mathbf{G}$ ($2 \times n$):

$$\mathbf{G}_k = \begin{bmatrix} u_{k1} & ... & u_{kn} \\ v_{k1} & ... & v_{kn} \end{bmatrix} = \begin{bmatrix} \mathbf{R}_k | \mathbf{z}_k \end{bmatrix} \begin{bmatrix} t_{11} & ... & t_{n1} \\ t_{12} & ... & t_{n2} \\ t_{13} & ... & t_{n3} \\ 1 & ... & 1 \end{bmatrix} = \begin{bmatrix} \mathbf{R}_k | \mathbf{z}_k \end{bmatrix} \begin{bmatrix} \mathbf{t} \\ \mathbf{1}^T. \end{bmatrix} \tag{2.24}$$

For a single target moving, observed by c cameras, the bilinear form equation is derived:

$$\mathbf{G} = \begin{bmatrix} \mathbf{G}_1 \\ ... \\ \mathbf{G}_c \end{bmatrix} = \begin{bmatrix} [\mathbf{R}_1 | \mathbf{z}_1] \\ ... \\ [\mathbf{R}_c | \mathbf{z}_c] \end{bmatrix} \begin{bmatrix} \mathbf{t} \\ \mathbf{1}^T \end{bmatrix} = \mathbf{C} \begin{bmatrix} \mathbf{t} \\ \mathbf{1}^T \end{bmatrix}, \tag{2.25}$$

such form (2.25) implies a rank-4 constraint on matrix $\mathbf{G}$, since the matrix $\mathbf{C}$ (that contains the c camera matrices) has dimensions $2c \times 4$. In the case of affine cameras, an alignment to the image centroid can be made such that:

$$\tilde{\mathbf{G}} = \tilde{\mathbf{C}}\tilde{\mathbf{t}}. \tag{2.26}$$

The matrix form in (2.26) is a classical SFM problem formulation, known in Computer Vision since it leads to efficient closed-form solutions (for both 3D reconstruction and camera calibration). Such method is based on applying SVD to $\tilde{\mathbf{G}}$, similarly to what is made for the range sensors calibration step. The specificity in this case is the exploration of the constraints imposed by the camera models. The final solution for the image calibration can be found via a LS problem.
Computing the joint closed solution using the range and visual constraints of the heterogeneous sensors is possible because both measurement data share a common subspace, defined by the target position $\mathbf{t}$. Hence, both types of measurement are combined, using (2.26) and the equivalent expression for range, to form the measurement matrix $\mathbf{Y}$ of size $(m + 2c - 1) \times (n + 1)$ in (2.27).

$$\mathbf{Y} = \begin{bmatrix} \tilde{\mathbf{D}} \\ \tilde{\mathbf{G}} \end{bmatrix} = \begin{bmatrix} -2\tilde{\mathbf{S}} \\ \tilde{\mathbf{C}} \end{bmatrix} \tilde{\mathbf{t}} \tag{2.27}$$

A single SVD can be applied to this matrix in order to obtain the first factorization, which can be then upgraded to metric. It is noted that it might be necessary to perform a normalization of the data to balance the SVD results. Furthermore, the constraints of each of the c cameras are recast to define a LS problem (according to [14] it is formulated as $\hat{\mathbf{U}}_i vech(\mathbf{H}) = \mathbf{0}$, where the operator $vech(\mathbf{H})$ is defined as the half-vectorization, obtaining a column vector by vectorizing only the lower triangular part of $\mathbf{H}$). Also, the range constraints should be embedded into a LS form, so that it is possible to jointly solve both constraints. In [14], the anchor constraints can be expressed as $\hat{\mathbf{A}} vech(\mathbf{H}) = \check{\mathbf{a}}$ (where $\hat{\mathbf{A}}$ has the known anchor coordinates). The resulting final equation embodying the camera and range sensor constraints is:

$$\begin{bmatrix} \hat{\mathbf{A}} \\ \hat{\mathbf{U}} \end{bmatrix} = vech(\mathbf{H}) = \begin{bmatrix} \check{\mathbf{a}} \\ \mathbf{0} \end{bmatrix}, \tag{2.28}$$

in which $\mathbf{H}$ is a matrix defined as $\mathbf{Q}_j \mathbf{Q}_j^T$ ($\mathbf{Q}_j$ is a mixing matrix found by using the anchor positions

and camera models constraints together). To compute $\mathbf{H}$, equation in (2.29) can be solved simply with a pseudo-inverse:

$$vech(\mathbf{H}) = (\mathbf{F}^T\mathbf{F})^{-1}\mathbf{F}^T\mathbf{r}, \quad \text{where} \qquad \mathbf{F} = \begin{bmatrix} \hat{\mathbf{A}} \\ \hat{\mathbf{U}} \end{bmatrix} \qquad \text{and} \qquad \mathbf{r} = \begin{bmatrix} \check{\mathbf{a}} \\ \mathbf{0} \end{bmatrix}. \tag{2.29}$$

A real test was performed with a RGB-D device (that generates a high number of range measurements coupled with standard images). In this test, the data matrix $\mathbf{Y}$ allowed to obtain the target 3D position, for each instant, together with the 3D location of the depth sensor and the camera parameters.
As future work, Crocco et al., in [14], mention the extension to the missing data case in order to allow more complex and larger networks to be also calibrated according to this method.

Discussion of existing hybrid schemes

Although the clear relevance of the work by Biswas et al. in [7] and Crocco et al. in [14], the approach proposed in this dissertation differs from the previous and goes beyond, since it surpasses the limitations present in [7, 14]. More specifically, [7] shows a space dimensionality constraint, as the proposed method can be applied exclusively in 2D and in [14] there is a limitation of the observations, since observations of the same target points must be performed with the different sensors (range and video). In this dissertation work, however, two distinct sets of target points (in two different networks) in the 3D space will be sensed by the heterogeneous sensors: a set of wireless sensor nodes and a set of objects; thus the vector to be estimated is larger than the one in [14], as the position estimation of both types of targets, which produce different measurements, is sought. Moreover, a seamless and natural integration of both types of information is pursued. Range information (either from acoustic or radio signals as Wi-Fi and from image properties) and pose/shape (from time varying range or pose estimation from vision) are harmoniously integrated in the same formulation. The fact that range information will depend on the translation between the source and the sensor while pose will depend not only on the translation but also on the rotation matrix leads to a formulation with two minimization problems (one related to the observed objects and the other to the estimated distances) in which the translation vector is the same.

Chapter 3

Methodologies

This Chapter addresses the approach followed throughout the work in order to solve the proposed problem for this thesis, describing the new methodologies devised for such purpose. The single-source hybrid localization was the first issue to be tackled, thus a new method that fuses distance and angle information to perform a target localization is proposed in Section 3.1, including a novel optimization based formulation for such hybrid scenarios. Such formulation is presented in detail in Subsection 3.1.1 and a procedure to solve the self-calibration problem of the different sensor networks, which emerged when implementing the hybrid localization algorithm in a real scenario, is described in Subsection 3.1.2. An extension of the hybrid localization to a multi-source node scenario, aiming a cooperative paradigm, is proposed in Section 3.2. The optimization criterion is similar to the one for the single-source case, but the actual solution algorithm is entirely different and aimed at large problem sizes that often arise in the collaborative paradigm, even for a relatively modest number of network nodes.

3.1 Single Source Hybrid Localization algorithm

3.1.1 Localization based on ranges and vision - FLORIS

In order to perform the fusion between range measurements from sensor anchor nodes and visual information gathered from video, a joint formulation was developed and is presented in this Section.
This formulation, named FLORIS from Fusion Localization using Ranges and Incident Streaks, exhibits some similarities to the SLCP algorithm [34] (described in Subsection 2.1.1) and is extended to 3D space using a variation of the technique adopted in the SLNN algorithm, also from [34].
Assuming an environment with two sensor networks comprising different types of anchors, namely, (1) sensor nodes that can retrieve the time of flight thus returning range information, and (2) camera-equipped nodes that can retrieve visual features, it is possible to estimate the current localization of a target, based on measurements from the hybrid network.
An illustrative example of the assumed configuration/scenario is depicted in Figure 3.1, where a target node moves in an environment including a wireless sensor network and visual features.

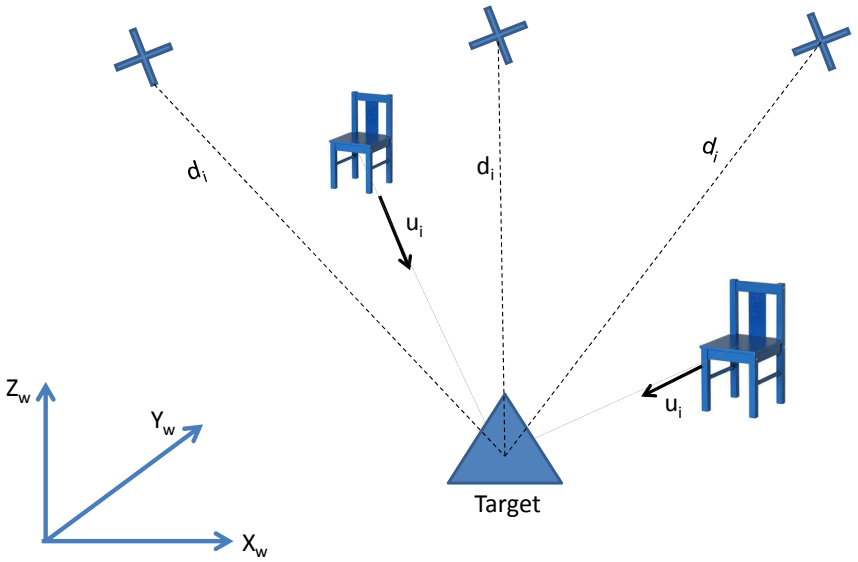

Figure 3.1: Localizaion scenario assumed for the hybrid single-source formulation

The posed problem is a single-source localization problem, in which ranges can be obtained with the wireless sensor network and the visual features provide angle information. Therefore, both range and angle information are encoded in the cost function presented in (3.1)

$$f(x) = \sum_{i \in \mathcal{R}} D^2_{C_i}(x) + \sum_{i \in \mathcal{T}} D^2_{L_i}(x), \tag{3.1}$$

where x is the target position to be determined. Furthermore, D_{C_i} represents the euclidean distance operator to C_i which is a circumference centered at each wireless anchor a_i and with radius d_i, and D_{L_i} denotes the same euclidean distance operator but to the straight line L_i originated at the visual anchor a_i and with orientation u_i. The intuitive idea behind (3.1) is that this formulation minimizes, on the one hand, the distance of the target position estimate relative to the circumferences centered at the wireless anchors with a radius d_i and, on the other hand, the distance between the target position estimation and the line, originating at the visual anchor, with orientation u_i (see Figure 3.2 with the geometric representation of the terms in the joint cost function (3.1)). The positions of the m reference points (anchors) $a_i \in \mathbb{R}^n, i = 1, ..., m$ are known *a priori*, and of these, the ones whose indices belong to set $\mathcal{R}$ provide range measurement to the source, d_i, whereas those with indices in the set $\mathcal{T}$ measure bearings represented by the unit direction vector u_i in which the target "sees" each visual anchor, obtained using a visual sensor. In (3.1), while the squared range $D^2_{C_i}(x)$ is given by $(\|x - a_i\| - d_i)^2$, $D^2_{L_i}(x)$ is given by $(x - a_i)^T(\mathbf{I}_n - u_i u_i^T)(x - a_i)$ according to [4]. It follows that (3.1) can be equivalently expressed as in (3.2),

$$f(x) = \sum_{i \in \mathcal{R}} (\|x - a_i\| - d_i)^2 + \sum_{i \in \mathcal{T}} (x - a_i)^T(\mathbf{I}_n - u_i u_i^T)(x - a_i), \tag{3.2}$$

for a considered space with dimension n (space of interest is $\mathbb{R}^n$).
The formulation related to the ranges (the left component of (3.2)), similarly to [34], is the ML estimate under i.i.d. Gaussian noise, however, no such interpretation can be given for the component associated with the angle information (the right term of (3.2)).

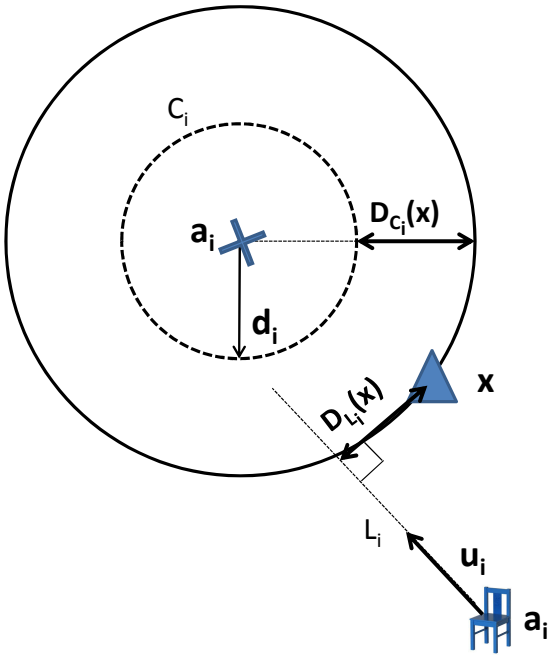

Figure 3.2: FLORIS geometric scheme (depicting target x, one acoustic anchor a_i, one visual anchor a_i and the distances $D_{C_i}(x)$ and $D_{L_i}(x)$ to be minimized)

3.1.1.1 2D formulation - representation in the complex plane

The range parcel of the cost function (3.2) is shown, in [32], to be equivalent to the usual simple (not squared) range approach (2.2). The approach followed here to minimize (3.2) for the 2D space (n = 2) is reminiscent from SLCP [34], yet, to include both types of information there are two constraints (one from range and the other from vision), as in formulation (3.3), where the total number of anchors (m) is equal to the sum of the sensor nodes and the visual features ($m = R + T$).

$$\begin{aligned} \underset{x,y_i}{\text{minimize}} \quad & \sum_{i=1}^{m} \|x - y_i\|^2 \\ \text{subject to} \quad & y_i = a_i + d_i e^{j\phi_i}, \quad i \in \mathcal{R} \\ & y_i = a_i + e^{j\varphi_i} t_i, \quad i \in \mathcal{T} \end{aligned} \tag{3.3}$$

In formulation (3.3), the first condition represents the range constraint, since the position of the source can be interpreted as being on a certain direction, encoded by an angle ($e^{j\phi_i}$) [34], multiplied by the distance measured (d_i) and summed to the position of anchor a_i represented in the complex plane. Regarding the second condition, it represents the visual constraint, as, similarly to the range case, we can encode the observation of an object using its orientation ($u_i = e^{j\varphi_i}$) times the length (t_i). Furthermore, these two constraints can be presented together in a compact matrix form, with equation (3.4)

$$\mathbf{y} = \mathbf{a} + \mathbf{R} \begin{bmatrix} \boldsymbol{\theta} \\ \mathbf{t} \end{bmatrix}, \tag{3.4}$$

where $\mathbf{a}$ stacks vertically both acoustic and visual anchors coordinates and for the 2D case $\mathbf{R}$ is the diagonal matrix

$$\mathbf{R} = \begin{bmatrix} d_1 & \cdots & 0 & 0 & \cdots & 0 \\ 0 & \ddots & 0 & 0 & \cdots & 0 \\ 0 & \cdots & d_R & 0 & \cdots & 0 \\ 0 & \cdots & 0 & e^{j\varphi_1} & \cdots & 0 \\ 0 & \cdots & 0 & 0 & \ddots & 0 \\ 0 & \cdots & 0 & 0 & \cdots & e^{j\varphi_T} \end{bmatrix}, \tag{3.5}$$

and

$$\boldsymbol{\theta} = \begin{bmatrix} e^{j\phi_1} \\ \vdots \\ e^{j\phi_R} \end{bmatrix}, \mathbf{t} = \begin{bmatrix} t_1 \\ \vdots \\ t_T \end{bmatrix}. \tag{3.6}$$

The solution of the problem in (3.3) with respect to x, if y_i is fixed, is simply the center of mass of the configuration ($x = \frac{1}{m}\sum_{i=1}^{m} y_i$). Hence for 2D, in the complex plane, the cost function of the formulation in (3.3) can be written as $\mathbf{y}^*\mathbf{J}\mathbf{y}$[1], where $\mathbf{J}$ is the projection matrix ($\mathbf{J} = \mathbf{I} - \frac{1}{m}\mathbf{1}\mathbf{1}^T$). Using the latter formulation, and replacing $\mathbf{y}$ by the expression in (3.4), the following formulation is derived:

$$\begin{aligned} \underset{\boldsymbol{\theta},\mathbf{t}}{\text{minimize}} \quad & \left(\mathbf{a} + \mathbf{R}\begin{bmatrix}\boldsymbol{\theta} \\ \mathbf{t}\end{bmatrix}\right)^* \mathbf{J} \left(\mathbf{a} + \mathbf{R}\begin{bmatrix}\boldsymbol{\theta} \\ \mathbf{t}\end{bmatrix}\right) \\ \text{subject to} \quad & |\theta_i| = 1, \quad i \in \mathcal{R} \\ & t_i \in \mathbb{R}_+, \quad i \in \mathcal{T}. \end{aligned} \tag{3.7}$$

Expanding the cost function in (3.7) yields the following matrix form, a quadratic formulation in $\begin{bmatrix}\boldsymbol{\theta} & \mathbf{t} & 1\end{bmatrix}^T$ (3.8)

$$\begin{aligned} \underset{\boldsymbol{\theta},\mathbf{t}}{\text{minimize}} \quad & \begin{bmatrix}\boldsymbol{\theta}^* & \mathbf{t}^* & 1\end{bmatrix} \begin{bmatrix}\mathbf{R}^*\mathbf{J}\mathbf{R} & \mathbf{R}^*\mathbf{J}\mathbf{a} \\ \mathbf{a}^*\mathbf{J}\mathbf{R} & \mathbf{a}^*\mathbf{J}\mathbf{a}\end{bmatrix} \begin{bmatrix}\boldsymbol{\theta} \\ \mathbf{t} \\ 1\end{bmatrix} \\ \text{subject to} \quad & |\theta_i| = 1, \quad i \in \mathcal{R} \\ & t_i \in \mathbb{R}_+, \quad i \in \mathcal{T}. \end{aligned} \tag{3.8}$$

Applying the trace of a matrix, whose cyclic nature lets the order of the matrices to be swapped, hence considering $\mathbf{W} = \begin{bmatrix}\boldsymbol{\theta} & \mathbf{t} & 1\end{bmatrix}^T \begin{bmatrix}\boldsymbol{\theta}^* & \mathbf{t}^* & 1\end{bmatrix}$, the final formulation can be defined as in (3.9)

$$\begin{aligned} \underset{\mathbf{W}}{\text{minimize}} \quad & tr\left(\begin{bmatrix}\mathbf{R}^*\mathbf{J}\mathbf{R} & \mathbf{R}^*\mathbf{J}\mathbf{a} \\ \mathbf{a}^*\mathbf{J}\mathbf{R} & \mathbf{a}^*\mathbf{J}\mathbf{a}\end{bmatrix}\mathbf{W}\right) \\ \text{subject to} \quad & rank(\mathbf{W}) = 1 \\ & \mathbf{W} \succeq 0 \\ & tr(\mathbf{W}_{i,i}) = 1, \quad i \in \mathcal{R} \\ & \mathbf{W}_{i,nm+1} \geq 0, \quad i \in \mathcal{T} \\ & \mathbf{W}_{nm+1,nm+1} = 1. \end{aligned} \tag{3.9}$$

Neglecting the rank-1 constraint and performing a simplifying expansion of the sub-matrices of the cost function in (3.9) in real and imaginary parts, a relaxed SDP is obtained which can be efficiently solved

[1] $\mathbf{X}^*$ denotes the conjugate transpose of $\mathbf{X}$.

using generic solvers. In this way, the goals of finding the phases vector $\boldsymbol{\theta}$ as well as the translations vector $\mathbf{t}$ are jointly achieved, through matrix $\mathbf{W}$, provided that its rank is (at least approximately) one. Finally, the source position x is estimated as the average of the m points of y_i, defined according to (3.4), which are directly obtained from the rightmost column (or bottom row) of $\mathbf{W}$.
Alternatively, it is possible to obtain, in a first step, only the phase vector $\boldsymbol{\theta}$, from the matrix $\mathbf{W}$ in the SDP problem (3.9) and, afterwards, use this vector to minimize the cost function relative to $\mathbf{t}$.
It was verified that the resulting estimated source positions are similar for both approaches.

3.1.1.2 3D formulation

To solve this problem in 3D, only minor changes are applied. More specifically, considering the 3D space, the variables are no longer represented using the complex plane, therefore the formulation for $\mathbb{R}^3$ is in (3.10).

$$\begin{aligned} \underset{x, y_i \in \mathbb{R}^3}{\text{minimize}} \quad & \sum_{i=1}^{m} \|x - y_i\|^2 \\ \text{subject to} \quad & y_i = a_i + d_i \theta_i, \quad \|\theta_i\| = 1, \quad i \in \mathcal{R} \\ & y_i = a_i + u_i t_i, \quad t_i \in \mathbb{R}_+, \quad i \in \mathcal{T} \end{aligned} \tag{3.10}$$

Comparing the latter formulation in (3.10) with (3.3), it is possible to notice that the performed 2D geometric interpretation is now extended to the 3D case, in which θ_i plays the same role as $e^{j\phi_i}$ and u_i takes the place of $e^{j\varphi_i}$. Similarly to the 2D case, the cost function in (3.10) can be expressed as $\mathbf{y^T J y}$, which allows to follow the same rationale.
Furthermore, matrix $\mathbf{R}$ as well as the projection matrix $\mathbf{J}$ are now defined as follows:

$$\mathbf{R} = \begin{bmatrix} d_1 & 0 & 0 & 0 & 0 & 0 & \cdots & 0 & 0 & 0 & 0 & 0 & \cdots & 0 \\ 0 & d_1 & 0 & 0 & 0 & 0 & \cdots & 0 & 0 & 0 & 0 & 0 & \cdots & 0 \\ 0 & 0 & d_1 & 0 & 0 & 0 & \cdots & 0 & 0 & 0 & 0 & 0 & \cdots & 0 \\ 0 & 0 & 0 & d_2 & 0 & 0 & \cdots & 0 & 0 & 0 & 0 & 0 & \cdots & 0 \\ 0 & 0 & 0 & 0 & d_2 & 0 & \cdots & 0 & 0 & 0 & 0 & 0 & \cdots & 0 \\ 0 & 0 & 0 & 0 & 0 & d_2 & \cdots & 0 & 0 & 0 & 0 & 0 & \cdots & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & \ddots & 0 & 0 & 0 & 0 & 0 & \cdots & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & \cdots & d_R & 0 & 0 & 0 & 0 & \cdots & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & \cdots & 0 & d_R & 0 & 0 & 0 & \cdots & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & \cdots & 0 & 0 & d_R & 0 & 0 & \cdots & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & \cdots & 0 & 0 & 0 & u_{1x} & 0 & \cdots & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & \cdots & 0 & 0 & 0 & u_{1y} & 0 & \cdots & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & \cdots & 0 & 0 & 0 & u_{1z} & 0 & \cdots & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & \cdots & 0 & 0 & 0 & 0 & u_{2x} & \cdots & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & \cdots & 0 & 0 & 0 & 0 & u_{2y} & \cdots & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & \cdots & 0 & 0 & 0 & 0 & u_{2z} & \cdots & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & \cdots & 0 & 0 & 0 & 0 & 0 & \ddots & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & \cdots & 0 & 0 & 0 & 0 & 0 & \cdots & u_{Tx} \\ 0 & 0 & 0 & 0 & 0 & 0 & \cdots & 0 & 0 & 0 & 0 & 0 & \cdots & u_{Ty} \\ 0 & 0 & 0 & 0 & 0 & 0 & \cdots & 0 & 0 & 0 & 0 & 0 & \cdots & u_{Tz} \end{bmatrix}, \tag{3.11}$$

$$\mathbf{J} = \mathbf{I}_{mn} - \frac{1}{R+T} \cdot \mathbf{1}_m \mathbf{1}_m^T \otimes (\mathbf{I}_n \cdot \mathbf{I}_n)\,. \tag{3.12}$$

Since the cost function (3.2) used as the basis for the previously proposed algorithm is not the ML function (it would be the ML function for the ranges terms only, under Gaussian noise), a study regarding the behaviour of this function becomes both necessary and relevant. Therefore, multiple simulations comprising randomly generated hybrid networks (with both acoustic and visual anchors) and with no additive noise were performed, using *MATLAB* and CVX/SDPT3[2] as a general-purpose solver. The contour lines of the cost function in (3.2) were represented, for a grid of 2D coordinates (x and y), in order to visualize the shape of the resultant function. Figures 3.3 and 3.4 depict the cost function error obtained for two runs.

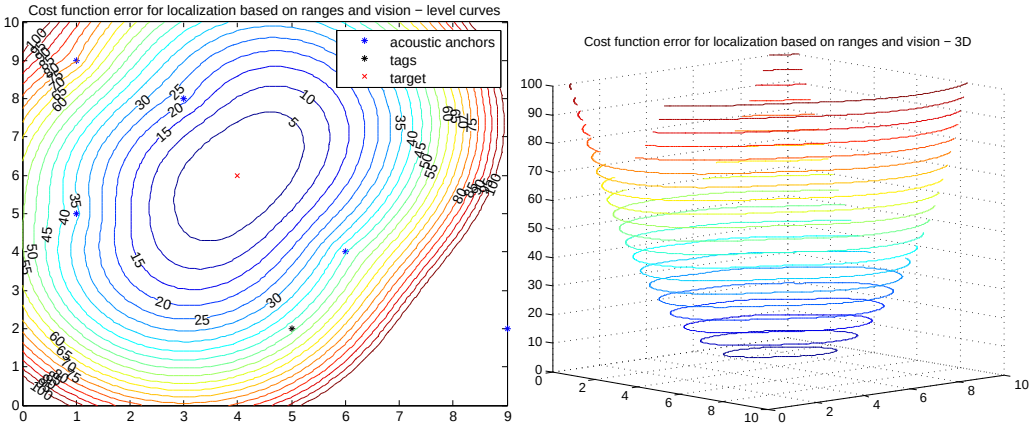

Figure 3.3: Cost function error (in distance units) of FLORIS formulation, for a network of 5 acoustic anchors and 2 visual tags - 2D and 3D visualizations

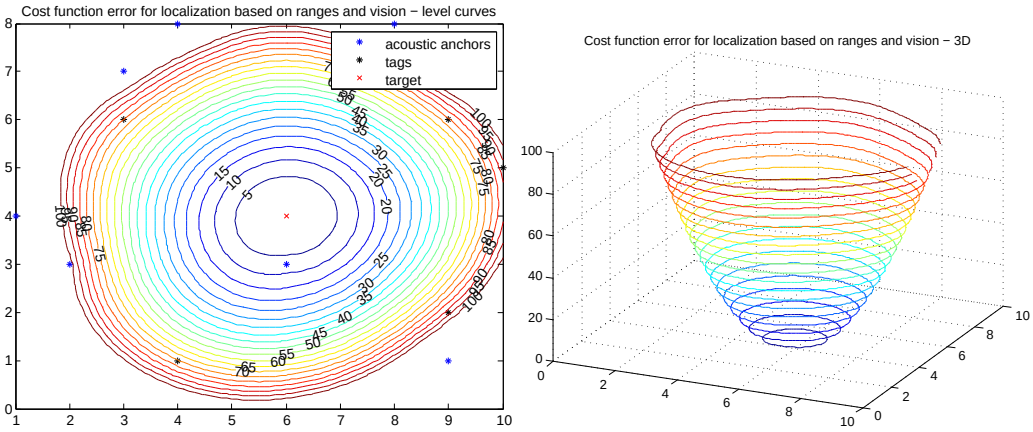

Figure 3.4: Cost function error (in distance units) of FLORIS formulation, for a network of 8 acoustic anchors and 5 visual tags - 2D and 3D visualizations

Analysing the plots obtained, it is possible to conclude that, despite not being the ML estimate, the obtained functions are said to be "well-behaved", as they exhibit a near convex shape, with the estimated target position at the minimum and without the undesired local minima. Moreover, it was observed that, as the number of anchors increase, the function becomes even more "well-behaved" (steeper cone) and similar to a convex shape. Under these conditions, the convex relaxation used in FLORIS is expected to

[2]http://www.math.nus.edu.sg/ mattohkc/sdpt3.html (accessed in October 2014)

yield an (approximately) rank-1 solution with high probability.
A thorough set of simulations were performed on randomly generated networks in order to characterize and test the performance of the previously proposed algorithm. Chapter 4 presents the results of such simulations. A performance comparison with other two state-of-the-art single-source methods indicates a leading performance of FLORIS for high noise cases, due to the particular robustness of the estimates accuracy in such scenarios.
The first experimental set-up developed to implement this hybrid localization algorithm consisted of a video camera mounted in a solidary way with a Cricket listener node. The most relevant tests performed and results obtained in this set-up are, also, presented and discussed in Chapter 4. In this case, the practical feasibility of FLORIS is completely attested, as the results attained surpass in accuracy the ones provided by the other methods, for the same experimental dataset.

It should be highlighted that the above proposed hybrid localization algorithm differs from the methods presented in Chapter 2, particularly, [7] and [14], thus making a novel contribution to this field. In the proposed scenario for this thesis there are two separate networks of heterogeneous sensors which detect and measure different variables, namely range and angle from vision. However, in the work found in the state-of-the-art, only one network of sensors that are able to *simultaneously* measure different types of variables is explored. Additionally, FLORIS, adopts the same strategy for relaxation used in SLCP in [34], and thus provides a very tight approximation to the original nonconvex problem, conferring a better performance when compared to the R-LS method in [5]. Besides the aforementioned aspects, in such previous works only simulation results are presented yielding neither deployment nor experimental tests in a real scenario.

3.1.2 Self-calibration process for the two separate sensor networks

From the existence of the two detached sensor networks, which contrasts with the related work found (more specifically in [7, 14]), emerged the need to calibrate both networks, as a precondition to perform localization. In fact, in [7] it is assumed that the obtained measurements are the absolute angles, since all the anchor nodes are aware of their orientation relative to the global coordinate system. This differs from our scenario, in which the network of sensors measuring angles is detached from the network of ranging sensors. Hence, such assumption cannot be made and a method to relate and determine the transformation between the two is required. The present Subsection details the process developed to achieve such calibration. We highlight that this is another major original contribution of this dissertation.
To fuse the information conveyed by both sensors, which belong to two different sensor networks, both measurements have to be referenced in a common coordinate system. However, as shown in Figure 3.5, the acoustic anchors form a network detached from the network of the visual features. As each network has its own separate coordinate system (coordinate system denoted as c for the sensor network and a for the visual anchors), it becomes necessary to define a global referential. Furthermore, Figure 3.5 depicts the target (a device comprising a video camera that can measure transformations to the visual markers attached to an acoustic sensor that can gather range measurements to the wireless sensor nodes) performing a trajectory in the localization scenario which includes both types of anchor nodes and all the translations and rotations that can be measured or should be determined through the calibration process.
In this case, the range measurements obtained acoustically will lead to a target position estimation in

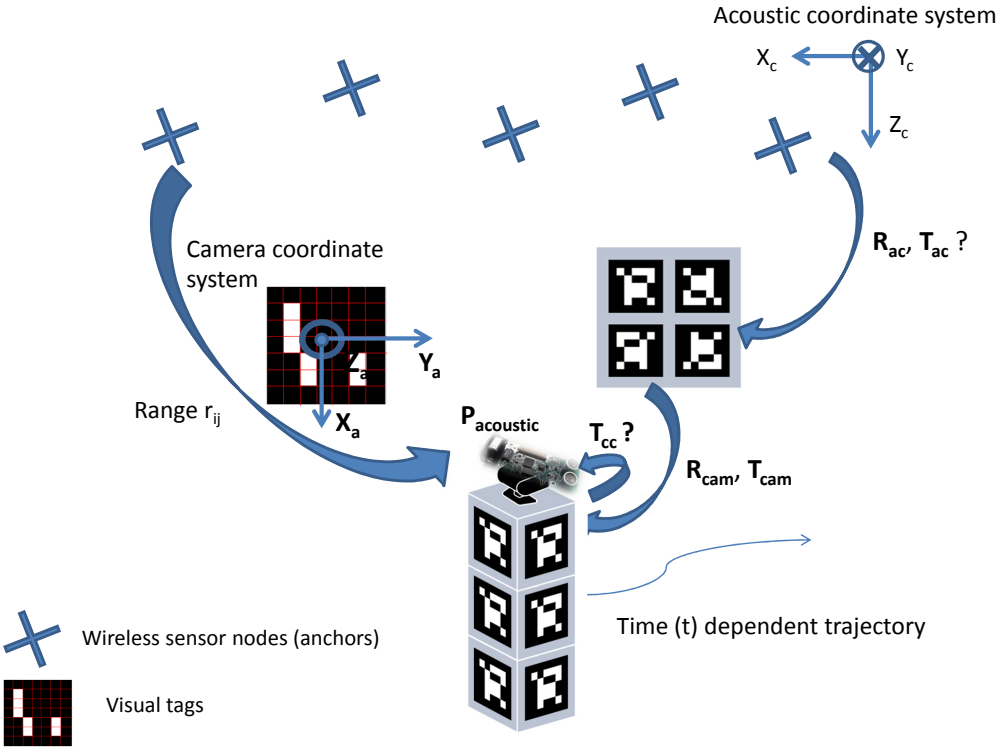

Figure 3.5: Scheme for the self-calibration of the two detached networks (acoustic and visual)

reference to the coordinate system defined by the acoustic anchors. On the other hand, the orientation of the camera, relative to the identified visual features, is determined in a coordinate system defined by such features.

The method that is explained hereinafter succeeds in determining the transformation (rotation matrix and translation vector) between the two sensor networks, without any prior information, only by using pairs of estimated acoustic target positions and camera positions. Hence, this process is called self-calibration, as there is no need for prior information. The massive flexibility and convenience provided by such assumptions-free process is definitely a decisive advantage regarding the experimental deployment when compared to previously proposed methods.

Considering a set-up similar to the one presented in Figure 3.5 and a target formed by an acoustic sensor attached to a camera, an algebraic method for the auto-calibration is derived. Such method derivation will use the notation of Figure 3.5 throughout.

The unknowns of this process are $\mathbf{R}_{\text{ac}}$, which denotes the rotation matrix between the camera referential and the acoustic referential system, $\mathbf{T}_{\text{ac}}$, which represents the translation vector between the same two networks, and, since the range and orientation are not measured by the same sensor, it is also necessary to determine $\mathbf{T}_{\text{cc}}$, which is the translation between the focal point of the camera and the center of the Ultrasound (US) receiver of the acoustic sensor. Additionally, let $\mathbf{R}_{\text{cam}}$ and $\mathbf{T}_{\text{cam}}$ be the rotation and translation, respectively, measured by the camera relative to the visual features and $\mathbf{P}_{\text{acoustic}}$ be the acoustic sensor localization computed using the ranges, in the referential defined by the acoustic network. In the envisaged procedure, sets of range measurements, between the acoustic anchors and the target, as well as pairs of $\mathbf{R}_{\text{cam}}$ and $\mathbf{T}_{\text{cam}}$, from the camera detecting a visual feature, can be collected. With the first set of data, the position of the acoustic sensor in the acoustic reference system ($\mathbf{P}_{\text{acoustic}}$) can be estimated using an algorithm based on a Least-Squares methodology such as the SR-LS [5] or on an SDR approach, such as the SLCP [33]. On the other hand, it is possible to determine the rotation and translation of a camera relative to a visual feature ($\mathbf{R}_{\text{cam}}$ and $\mathbf{T}_{\text{cam}}$), and therefore the camera position

in the visual features coordinate system is straightforwardly computed.
Using the matrix form, combining the rotation matrices and translation vectors, in accordance to the scheme in Figure 3.5, the relation in (3.13) follows

$$\begin{aligned} & \mathbf{M}_{\mathrm{AC}}\ \mathbf{M}_{\mathrm{CAM}}\ \mathbf{M}_{\mathrm{CC}} = \mathbf{M}_{\mathrm{ACOUSTIC}} \\ \Leftrightarrow & \begin{bmatrix} \mathbf{R}_{\mathrm{ac}} & \mathbf{T}_{\mathrm{ac}} \end{bmatrix} \begin{bmatrix} \mathbf{R}_{\mathrm{cam}}^{(\mathrm{t})} & \mathbf{T}_{\mathrm{cam}}^{(\mathrm{t})} \\ \mathbf{0} & 1 \end{bmatrix} \begin{bmatrix} \mathbf{R}_{\mathrm{cc}} & \mathbf{T}_{\mathrm{cc}} \\ \mathbf{0} & 1 \end{bmatrix} = \begin{bmatrix} \mathbf{I} & \mathbf{P}_{\mathrm{acoustic}}^{(\mathrm{t})} \end{bmatrix} \\ \Leftrightarrow & \begin{bmatrix} \mathbf{R}_{\mathrm{ac}}\mathbf{R}_{\mathrm{cam}}^{(\mathrm{t})}\mathbf{R}_{\mathrm{cc}} & \mathbf{R}_{\mathrm{ac}}(\mathbf{R}_{\mathrm{cam}}^{(\mathrm{t})}\mathbf{T}_{\mathrm{cc}} + \mathbf{T}_{\mathrm{cam}}^{(\mathrm{t})}) + \mathbf{T}_{\mathrm{ac}} \end{bmatrix} = \begin{bmatrix} \mathbf{I} & \mathbf{P}_{\mathrm{acoustic}}^{(\mathrm{t})} \end{bmatrix}, \end{aligned} \tag{3.13}$$

where the notation $\mathbf{P}_{\mathrm{acoustic}}^{(\mathrm{t})}$ is used in (3.13) to represent a time-dependent variable.
From the last step of (3.13), it follows that

$$\begin{cases} \mathbf{R}_{\mathrm{ac}}\mathbf{R}_{\mathrm{cam}}^{(\mathrm{t})}\mathbf{R}_{\mathrm{cc}} = \mathbf{I} \\ \mathbf{R}_{\mathrm{ac}}(\mathbf{R}_{\mathrm{cam}}^{(\mathrm{t})}\mathbf{T}_{\mathrm{cc}} + \mathbf{T}_{\mathrm{cam}}^{(\mathrm{t})}) + \mathbf{T}_{\mathrm{ac}} = \mathbf{P}_{\mathrm{acoustic}}^{(\mathrm{t})}. \end{cases} \tag{3.14}$$

Taking the difference between two generic time instants, t_1 and t_2, for the second equation of the system (3.14), one can obtain the expression (3.15):

$$\left[\mathbf{R}_{\mathrm{ac}}(\mathbf{R}_{\mathrm{cam}}^{(\mathrm{t}_2)}\mathbf{T}_{\mathrm{cc}} + \mathbf{T}_{\mathrm{cam}}^{(\mathrm{t}_2)}) + \mathbf{T}_{\mathrm{ac}}\right] - \left[\mathbf{R}_{\mathrm{ac}}(\mathbf{R}_{\mathrm{cam}}^{(\mathrm{t}_1)}\mathbf{T}_{\mathrm{cc}} + \mathbf{T}_{\mathrm{cam}}^{(\mathrm{t}_1)}) + \mathbf{T}_{\mathrm{ac}}\right] = \mathbf{P}_{\mathrm{acoustic}}^{(\mathrm{t}_2)} - \mathbf{P}_{\mathrm{acoustic}}^{(\mathrm{t}_1)}. \tag{3.15}$$

The expression (3.15) can be simplified into (3.16), according to the auxiliary scheme in Figure 3.6, which has the same notation of the algebraic derivation performed below.

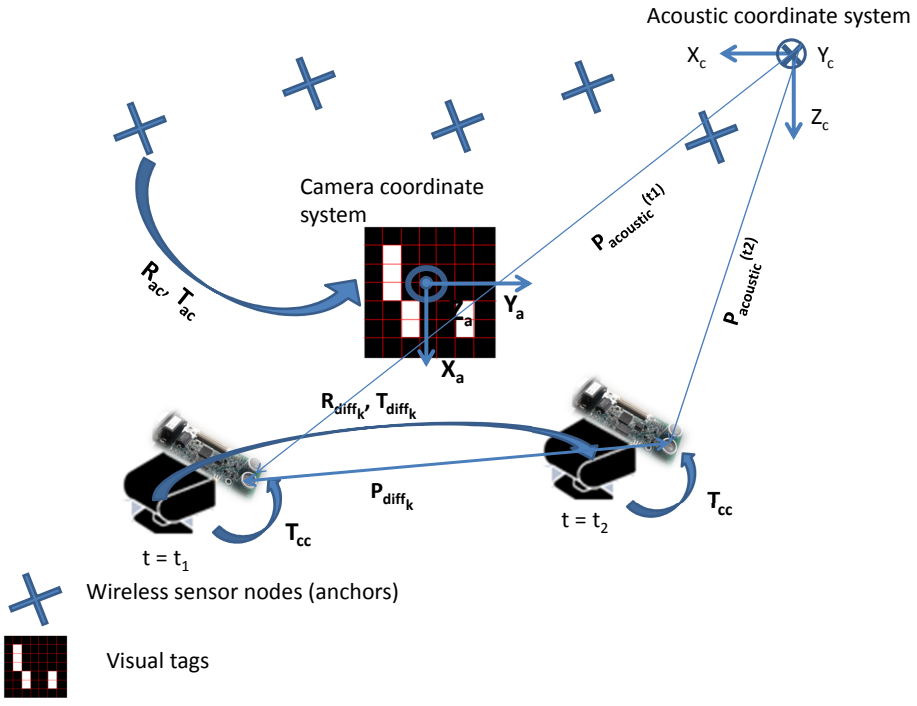

Figure 3.6: Auxiliary scheme for the derivation of equation (3.16)

$$\underbrace{\mathbf{P}_{\mathrm{acoustic}}^{(\mathrm{t}_2)} - \mathbf{P}_{\mathrm{acoustic}}^{(\mathrm{t}_1)}}_{\mathbf{P}_{\mathrm{diff_k}}} = \mathbf{R}_{\mathrm{ac}} \left[\underbrace{(\mathbf{R}_{\mathrm{cam}}^{(\mathrm{t}_2)} - \mathbf{R}_{\mathrm{cam}}^{(\mathrm{t}_1)})}_{\mathbf{R}_{\mathrm{diff_k}}} \mathbf{T}_{\mathrm{cc}} + \underbrace{(\mathbf{T}_{\mathrm{cam}}^{(\mathrm{t}_2)} - \mathbf{T}_{\mathrm{cam}}^{(\mathrm{t}_1)})}_{\mathbf{T}_{\mathrm{diff_k}}} \right] \tag{3.16}$$

In (3.16) the translation vector between the acoustic and the visual network ($\mathbf{T}_{\mathrm{ac}}$) is cancelled, leaving just two unknown variables $\mathbf{R}_{\mathrm{ac}}$ and $\mathbf{T}_{\mathrm{cc}}$.

Appending/stacking horizontally all the N (number of measurements -1) pairwise differences measured, equation (3.16) can be expressed as in (3.17)

$$\underbrace{[\mathbf{P}_{\text{diff}_1} \cdots \mathbf{P}_{\text{diff}_N}]}_{\mathbf{P}_{\text{diff}}} = \mathbf{R}_{\text{ac}} \left(\underbrace{[\mathbf{R}_{\text{diff}_1}\mathbf{T}_{\text{cc}} \cdots \mathbf{R}_{\text{diff}_N}\mathbf{T}_{\text{cc}}]}_{\mathbf{R}_{\text{diff}}(\mathbf{I}_N \otimes \mathbf{T}_{\text{cc}})} + \underbrace{[\mathbf{T}_{\text{diff}_1} \cdots \mathbf{T}_{\text{diff}_N}]}_{\mathbf{T}_{\text{diff}}} \right), \tag{3.17}$$

where $\mathbf{I}_N$ denotes an identity matrix of $N \times N$, $\mathbf{R}_{\text{diff}}$ is formed by horizontally stacked differences of rotations matrices (which no longer hold the properties of a rotation matrix) and $\mathbf{T}_{\text{diff}}$ includes the pairwise differences of translation vectors.
According to (3.17), the error function can be defined as

$$\mathbf{E} = [\mathbf{P}_{\text{diff}} - \mathbf{R}_{\text{ac}} \left(\mathbf{R}_{\text{diff}}(\mathbf{I}_N \otimes \mathbf{T}_{cc}) + \mathbf{T}_{\text{diff}}\right)]. \tag{3.18}$$

With this formulation, $\mathbf{R}_{\text{ac}}$ can be estimated by solving an orthogonal Procrustes problem[3], which determines the closest orthogonal rotation matrix that best fits two sets of points. Here, one set of points is given by $\mathbf{P}_{\text{diff}}$ (which are the pairwise differences of the position estimations obtained through the acoustic sensor network, alone) and the other by $\mathbf{R}_{\text{diff}}(\mathbf{I}_N \otimes \mathbf{T}_{\text{cc}}) + \mathbf{T}_{\text{diff}}$, hence the solution of the Procrustes problem will give an estimate of $\mathbf{R}_{\text{ac}}$.
In order to reduce the complexity of the algebraic calculations in such cases, the Frobenius norm is usually applied. Since one of the ways in which this norm can be defined is $\|A\|_F^2 = tr(A^*A)$, we apply the trace operator to $\mathbf{E}^*\mathbf{E}$, thus $tr\{\mathbf{E}^*\mathbf{E}\}$. Taking the trace truly simplifies the calculations, since the expression is transformed in an inner-product (element-wise computation), while no information is lost, as the eigenvalues hold all the matrix information. This can also be regarded as computing the average of the error covariance matrix.
After some algebraic manipulation, a new expression for the error function is obtained in (3.19)[4].

$$\begin{aligned}\mathbf{E}(\mathbf{T}_{\text{cc}}, \mathbf{R}_{\text{ac}}) = tr\{&\mathbf{P}_{\text{diff}}^T\mathbf{P}_{\text{diff}} - 2\mathbf{P}_{\text{diff}}^T\mathbf{R}_{\text{ac}}\left(\mathbf{R}_{\text{diff}}\mathbf{I}_N \otimes \mathbf{T}_{\text{cc}} + \mathbf{T}_{\text{cam}}\right) + (\mathbf{I}_N \otimes \mathbf{T}_{\text{cc}})^T\mathbf{R}_{\text{diff}}^T\mathbf{R}_{\text{diff}}(\mathbf{I}_N \otimes \mathbf{T}_{\text{cc}}) + \\ &2\mathbf{T}_{\text{cam}}^T\mathbf{R}_{\text{diff}}\mathbf{I}_N \otimes \mathbf{T}_{\text{cc}} + \mathbf{T}_{\text{cam}}^T\mathbf{T}_{\text{cam}}\}\end{aligned} \tag{3.19}$$

To determine the unknown $\mathbf{T}_{\text{cc}}$, we should minimize the error, thereby, using the function $\mathbf{E}(\mathbf{T}_{\text{cc}}, \mathbf{R}_{\text{ac}})$, defined in (3.19) to form the cost function, yields the following optimization problem in (3.20).

$$\begin{aligned}\underset{\mathbf{R}_{\text{ac}}, \mathbf{T}_{\text{cc}}}{\text{minimize}} \quad & tr\{\mathbf{P}_{\text{diff}}^T\mathbf{P}_{\text{diff}} - 2\mathbf{P}_{\text{diff}}^T\mathbf{R}_{\text{ac}}\left(\mathbf{R}_{\text{diff}}\mathbf{I}_N \otimes \mathbf{T}_{\text{cc}} + \mathbf{T}_{\text{cam}}\right) + (\mathbf{I}_N \otimes \mathbf{T}_{\text{cc}})^T\mathbf{R}_{\text{diff}}^T\mathbf{R}_{\text{diff}}(\mathbf{I}_N \otimes \mathbf{T}_{\text{cc}}) \\ & + 2\mathbf{T}_{\text{cam}}^T\mathbf{R}_{\text{diff}}\mathbf{I}_N \otimes \mathbf{T}_{\text{cc}} + \mathbf{T}_{\text{cam}}^T\mathbf{T}_{\text{cam}}\} \\ \text{subject to} \quad & \mathbf{R}_{\text{ac}}^T\mathbf{R}_{\text{ac}} = \mathbf{I}_3\end{aligned} \tag{3.20}$$

The problem in (3.20) is a constrained optimization problem, since the matrix $\mathbf{R}_{\text{ac}}$ has to belong to the space of rotation matrices. Consequently, the method to minimize this objective function consists in using the Lagrange multipliers so that this problem is converted into an unconstrained optimization problem (the restriction is incorporated in the same equation as the cost function). The auxiliary function given by the Lagrangian method is shown in (3.21)

$$\mathcal{L}(\mathbf{T}_{\text{cc}}, \mathbf{R}_{\text{ac}}, \lambda) = \mathbf{E}(\mathbf{T}_{\text{cc}}, \mathbf{R}_{\text{ac}}) + tr(\lambda(\mathbf{R}_{\text{ac}}^T\mathbf{R}_{\text{ac}} - \mathbf{I}_3)), \tag{3.21}$$

[3] http://en.wikipedia.org/wiki/Orthogonal_Procrustes_problem (accessed in October 2014)
[4] As all the values are real (positions in the 3D space and rotation matrices), the conjugate transpose notation (X^*) is replaced, simply, by the transpose (X^T).

where λ is a constant called the Lagrange multiplier.
In accordance with this method, to estimate $\mathbf{T}_{\text{cc}}$ we need to solve the constraint-free problem in (3.22)

$$\underset{\mathbf{R}_{\text{ac}},\lambda}{\text{maximize}} \quad \left(\underset{\mathbf{T}_{\text{cc}}}{\text{maximize}} \quad \mathcal{L}(\mathbf{T}_{\text{cc}}, \mathbf{R}_{\text{ac}}, \lambda) \right). \tag{3.22}$$

To determine the extreme of the previous objective function, its partial derivative with respect to $\mathbf{T}_{\text{cc}}$ (the unknown variable) is computed and set to zero, yielding

$$\frac{\partial \mathcal{L}(\mathbf{T}_{\text{cc}}, \mathbf{R}_{\text{ac}}, \lambda)}{\partial \mathbf{T}_{\text{cc}}} = -2 \sum_i \mathbf{R}_{\text{diff}_i}^T \mathbf{R}_{\text{ac}}^T \mathbf{P}_{\text{diff}_i} + 2 \left(\sum_i \mathbf{R}_{\text{diff}_i}^T \mathbf{R}_{\text{diff}_i} \right) \mathbf{T}_{\text{cc}} + 2 \sum_i \mathbf{R}_{\text{diff}_i}^T \mathbf{T}_{\text{cam}_i} = 0. \tag{3.23}$$

Algebraically manipulating the expression in (3.23), results in the the following expression (3.24) to compute $\mathbf{T}_{\text{cc}}$:

$$\mathbf{T}_{\text{cc}} = \left(\sum_i \mathbf{R}_{\text{diff}_i}^T \mathbf{R}_{\text{diff}_i} \right)^{-1} \left(\sum_i \mathbf{R}_{\text{diff}_i}^T \mathbf{R}_{\text{ac}}^T \mathbf{P}_{\text{diff}_i} - \sum_i \mathbf{R}_{\text{diff}_i}^T \mathbf{T}_{\text{cam}_i} \right). \tag{3.24}$$

Additionally, following the computation of $\mathbf{R}_{\text{ac}}$ and $\mathbf{T}_{\text{cc}}$, the translation vector between both networks' referential systems $\mathbf{T}_{\text{ac}}$ is computed as:

$$\mathbf{T}_{\text{ac}} = \overline{\mathbf{P}_{\text{diff}}} - \mathbf{R}_{\text{ac}} \left(\overline{\mathbf{R}_{\text{cam}}} \mathbf{T}_{\text{cc}} + \overline{\mathbf{T}_{\text{cam}}} \right). \tag{3.25}$$

In light of the above algebraic derivation of the formulation to self-calibrate two separate sensor networks, an iterative method can be devised to estimate both $\mathbf{R}_{\text{ac}}$ and $\mathbf{T}_{\text{cc}}$.
A workflow summarizing the preceding newly proposed self-calibration method can be found in Figure 3.7. Since $\mathbf{T}_{\text{cc}}$ is considerably smaller when compared with $\mathbf{T}_{\text{cam}}$, a first estimate for $\mathbf{R}_{\text{ac}}$ ($\mathbf{R}_{\text{ac}}(k)$) can be

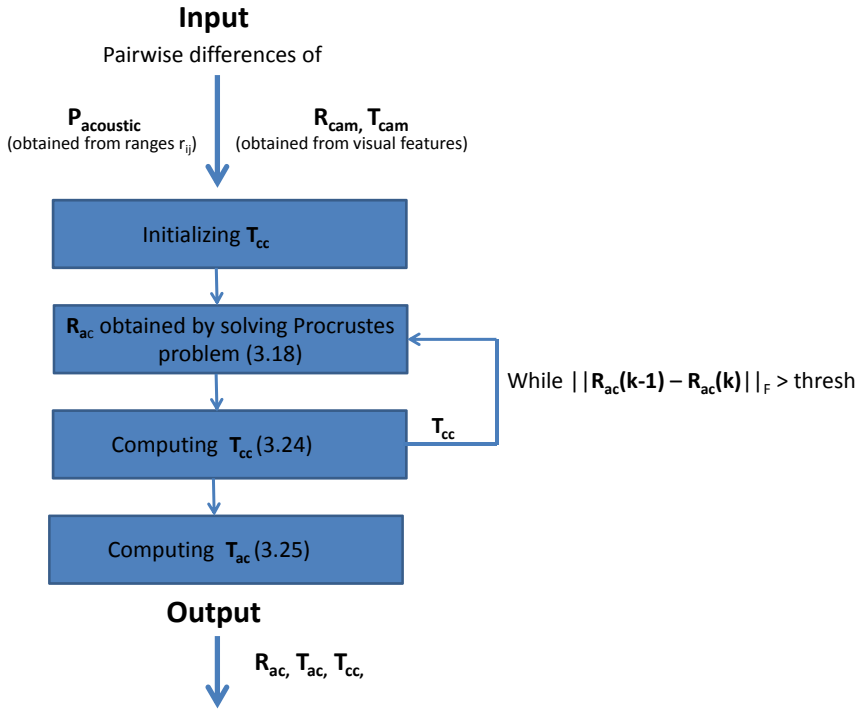

Figure 3.7: Outline of the proposed self-calibration procedure workflow

obtained neglecting $\mathbf{T}_{\text{cc}}$ ($\mathbf{T}_{\text{cc}} = 0$) or, alternatively, with an initialization of the value of $\mathbf{T}_{\text{cc}}$, if available.

3.2 Sensor Network Hybrid Localization algorithm

3.2.1 Collaborative Localization based on ranges and vision - CLORIS

In respect to non-cooperative techniques, the cooperative localization paradigm can notably improve the performance of localization, not just in terms of accuracy but also in increased coverage [45]. In fact, with numerous sensors exchanging position information it is possible to achieve higher accuracy in location estimates for a larger fraction of the network nodes.
In such collaborative paradigm, the envisaged scenario comprises, as illustrated in Figure 3.8, several targets roaming in the same environment communicating among them, transmitting the ranges measured (to the beacons and to the other targets) and incident streaks (again, to the visual anchors and to the other targets), in this way enhancing localization.

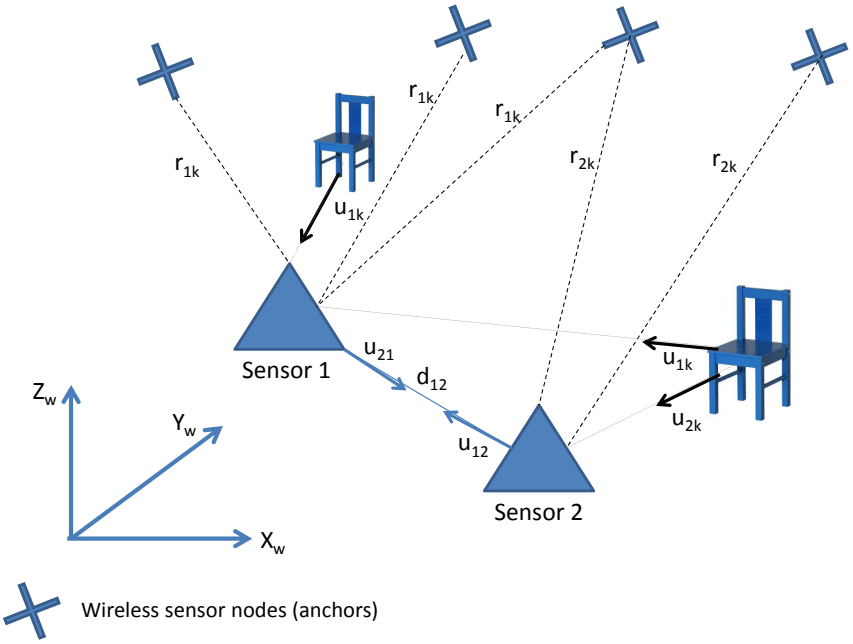

Figure 3.8: Localizaion scenario assumed for fusion sensor network localization approach

The methodology used in this collaborative approach is an extension of FLORIS scheme to this new, more complex, paradigm. More specifically, the cost function formulation is in line with our previous formulation for the single-source case. However, the algorithm relaxation to obtain the cooperative solution follows a distinctive strategy. Effectively, it has not a SLNN flavour, instead it follows the approach presented by Soares et. al in [43], introducing the necessary modifications to accommodate the component regarding the orientation measured between nodes, here represented by the presence of straight line sets. Thus, this approach constitutes the newly proposed CLORIS method.
In order to ease the understanding of the upcoming problem formulation notation, we reformulate the single-source problem of FLORIS in (3.1) as

$$\underset{x}{\text{minimize}} \quad f(x) = \sum_{k\in\mathcal{A}} D^2_{S_k}(x-a_k) + \sum_{k\in\mathcal{A}} D^2_{L_k}(x-a_k)\,, \tag{3.26}$$

where $D^2_{S_k}(x-a_k)$ is the squared euclidean distance operator (defined as $D^2_W(x) = \inf_{y\in W}\|x-y\|^2$) of point $x-a_k$ to the set formed by a sphere centered at the origin and with radius r_k and $D^2_{L_k}(x-a_k)$ is the squared euclidean distance of point $x-a_k$ to the straight line passing the origin with orientation u_k.

The sensor network localization problem differs from the previous problem since range and angles between sensors are also measured. Thus, in the cooperative paradigm two new terms, associated with the measurements between pairs of sensors, are added to the cost function, which can be written as

$$f(\mathbf{x}) = \sum_{i\sim j} D^2_{S_{ij}}(x_i - x_j) + \sum_i \sum_{k\in\mathcal{A}_i} D^2_{S_{ik}}(x_i - a_{ik}) + \sum_{i\sim j} D^2_{L_{ij}}(x_i - x_j) + \sum_i \sum_{k\in\mathcal{A}_i} D^2_{L_{ik}}(x_i - a_{ik}), \tag{3.27}$$

where, similarly to (3.26), S_{ij} is the sphere generated by noisy range measurement (d_{ij}) between sensors i and j and L_{ij} the set formed by the straight line passing in sensor i and j with measured unit vector u_{ij}. Moreover, for each sensor i, $\mathcal{A}_i$ is the subset of anchors whose distance to i is available as a noisy range measurement and $i \sim j$ denotes an edge ($\in \mathcal{E}$, the set of edges) between the sensors i and j, if there is a noisy range measurement between them and both nodes can communicate with each other.
Further, to obtain a convex function, the components of (3.27) which represent distances to a sphere ($D^2_{S_{ij}}$ and $D^2_{S_{ik}}$) can be relaxed by distances to a ball, according to (3.28):

$$D^2_{S_{ij}}(z) = \inf_{\|y\|=d_{ij}} \|z-y\|^2 \qquad \text{is replaced by} \qquad D^2_{B_{ij}}(z) = \inf_{\|y\|\leq d_{ij}} \|z-y\|^2. \tag{3.28}$$

Such relaxation is illustrated by Figure 3.9.

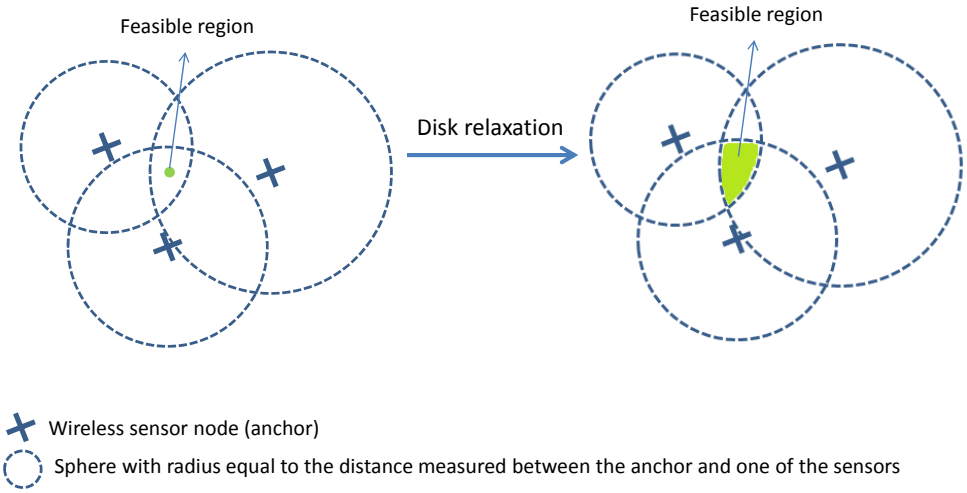

Figure 3.9: Scheme illustrating the disk relaxation performed, showing the feasible region after the relaxation

And an identical replacement is performed for the terms associated with anchor measurements. The obtained balls B_{ij} and B_{ik} are the convex hulls of the respective spheres. Furthermore, the distances $D^2_{B_{ij}}$ and $D^2_{B_{ik}}$ are the convex envelopes of $D^2_{S_{ij}}$ and $D^2_{S_{ik}}$, respectively.
Replacing the terms in (3.27) by the relaxation performed in (3.28) and the respective for the anchor dependent measurements yields the final convex cost function for the fusion approach of the disk relaxation based method CLORIS.

$$f(\mathbf{x}) = \sum_{i\sim j} D^2_{B_{ij}}(x_i - x_j) + \sum_i \sum_{k\in\mathcal{A}_i} D^2_{B_{ik}}(x_i - a_{ik}) + \sum_{i\sim j} D^2_{L_{ij}}(x_i - x_j) + \sum_i \sum_{k\in\mathcal{A}_i} D^2_{L_{ik}}(x_i - a_{ik}), \tag{3.29}$$

This resultant cost function (3.29), inspired in [43] is not the convex envelope of (3.27), indeed it is an underestimator of the primary problem. Yet, Soares et. al [43] present a discussion, for range-only collaborative localization, in which is concluded that this type of underestimator can provide very good approximations with quantifiable sub-optimality.

At this point it is important to make the following remark regarding a comparison between the relaxation strategy followed in CLORIS and the EDM approaches discussed in the state-of-the-art. Although cooperative only range EDM based methods can dispense, up to a certain point, the *a priori* knowledge of anchors positions, that is not the case for the present disk relaxation. Indeed, the knowledge of the anchor coordinates is crucial for this approach. Nevertheless, and as a positive feature of this disk relaxation, while in the SDP relaxations carried out in EDM based methods a factorization, without any rank guarantees, has to be necessarily performed, the present approach circumvents such inconvenient complexity. Thus, while in the EDM case we have to optimistically expect, with no guarantees, a rank-1 solution, in the CLORIS case we work directly with the anchors coordinates, as discussed bellow.
Now that a suitable cost function was found, the method addressed in this dissertation to obtain the solution follows the distributed sensor network localization algorithm via the accelerated Nesterov's gradient descent, introduced in [30].
Following [43] and adding the terms related to the angle information, the cost function (3.29) is expressed as

$$\underset{\mathbf{x}}{\text{minimize}} \quad f(\mathbf{x}) = \underbrace{\sum_{i\sim j} D^2_{B_{ij}}\left(\mathbf{A}_d\mathbf{x}\right) + \sum_{i\sim j} D^2_{L_{ij}}\left(\mathbf{A}_l\mathbf{x}\right)}_{g(\boldsymbol{x})} + \underbrace{\sum_{i}\sum_{k\in\mathcal{A}_i} D^2_{B_{ik}}\left(x_i - a_{ik}\right) + \sum_{i}\sum_{k\in\mathcal{A}_i} D^2_{L_{ik}}\left(x_i - a_{ik}\right)}_{h(\boldsymbol{x})}, \tag{3.30}$$

in which matrices $\mathbf{A}_d$ and $\mathbf{A}_l$ perform the coordinate differences for the acoustic and the visual anchors, respectively. These matrices are computed as $\mathbf{A}_d = \mathbf{C}_d \otimes \mathbf{I}_p$ and $\mathbf{A}_l = \mathbf{C}_l \otimes \mathbf{I}_p$ (p is the space dimension and this method is valid for both practical cases of $p = 2$ and $p = 3$), where $\mathbf{C}_d$ is the arc-node incidence matrix with the distances of the network and $\mathbf{C}_l$ is the arc-node incidence matrix relative to the orientations measured. The division into the sum of two functions $g(\mathbf{x})$ and $h(\mathbf{x})$ is useful for computing the gradient of the cost function.
The function $D^2_{B_{ij}}(z)$ is convex and differentiable, with gradient

$$\nabla D^2_{B_{ij}}(z) = z - P_{B_{ij}}(z), \tag{3.31}$$

where $P_{B_{ij}}(z)$ denotes de orthogonal projection of point z onto the convex set (ball) B_{ij}. Likewise, the gradient of $D^2_{L_{ij}}(z)$ is

$$\nabla D^2_{L_{ij}}(z) = z - Pl_{L_{ij}}(z), \tag{3.32}$$

$Pl_{L_{ij}}(z)$ is the orthogonal projection of point z onto the set formed by the straight line L_{ij}.
Hence, by (3.31), (3.32) and applying the gradient operator rules, the gradient of $g(\mathbf{x})$ is obtained as in (3.33).

$$\begin{aligned}
\nabla g(\mathbf{x}) &= \nabla D^2_B(z) + \nabla D^2_L(z) \\
&= \mathbf{A}_d^T \nabla D^2_B\left(\mathbf{A}_d\mathbf{x}\right) + \mathbf{A}_l^T \nabla D^2_L\left(\mathbf{A}_l\mathbf{x}\right) \\
&= \mathbf{A}_d^T\left(\mathbf{A}_d\mathbf{x} - P_B(\mathbf{A}_d\mathbf{x})\right) + \mathbf{A}_l^T\left(\mathbf{A}_l\mathbf{x} - Pl_L(\mathbf{A}_l\mathbf{x})\right) \\
&= \mathcal{L}_d\mathbf{x} - \mathbf{A}_d^T P_B(\mathbf{A}_d\mathbf{x}) + \mathcal{L}_l\mathbf{x} - \mathbf{A}_l^T Pl_L(\mathbf{A}_l\mathbf{x})
\end{aligned} \tag{3.33}$$

In expression (3.33) B and L denote the Cartesian product of the balls B_{ij} and the straight lines L_{ij}, corresponding to the edges with range or bearing information, respectively. Furthermore, $\mathcal{L}_{d(l)}$ is defined as $\mathbf{L}_{d(l)} \otimes \mathbf{I}_p$, being $\mathbf{L}_{d(l)}$ the Laplacian matrix related to the distance (angles) graph.

Regarding $h(\mathbf{x})$, let $\nabla h(\mathbf{x}) = (\nabla h_1(x_1), ..., \nabla h_n(x_n))$, where each term is:

$$\begin{aligned} \nabla h_i\left(x_i\right) &= \sum_{k \in \mathcal{A}_i} \nabla D^2_{B_{ik}}\left(x_i - a_{ik}\right) + \nabla D^2_{Lik}\left(x_i - a_{ik}\right) \\ &= \sum_{k \in \mathcal{A}_i} \left(x_i - a_{ik}\right) - P_B(x_i - a_{ik}) + \sum_{k \in \mathcal{A}_i} \left(x_i - a_{ik}\right) - Pl_L(x_i - a_{ik}). \end{aligned} \tag{3.34}$$

From (3.33) and (3.34), the overall gradient of cost function $f(\mathbf{x})$ can be computed according to expression (3.35).

$$\begin{aligned} \nabla f\left(\mathbf{x}\right) &= \mathcal{L}_d\mathbf{x} - \mathbf{A}_d^T P_B(\mathbf{A}_d\mathbf{x}) + \mathcal{L}_l\mathbf{x} - \mathbf{A}_l^T Pl_L(\mathbf{A}_l\mathbf{x}) \\ &+ \begin{bmatrix} \sum\limits_{k \in \mathcal{A}_1} \left(x_1 - a_{1k}\right) - P_B(x_1 - a_{1k}) + \sum\limits_{k \in \mathcal{A}_1} \left(x_1 - a_{1k}\right) - Pl_L(x_1 - a_{1k}) \\ \vdots \\ \sum\limits_{k \in \mathcal{A}_n} \left(x_n - a_{nk}\right) - P_B(x_n - a_{nk}) + \sum\limits_{k \in \mathcal{A}_n} \left(x_n - a_{nk}\right) - Pl_L(x_n - a_{nk}) \end{bmatrix} \end{aligned} \tag{3.35}$$

As both $\nabla g(\mathbf{x})$ and $\nabla h(\mathbf{x})$ are Lipschitz continuous functions, it is possible to compute the Lipschitz constants, which according to [43][5] are $L_g = 2\delta_{\max}$ ($\delta_{\max}$ is the maximum sensor node degree) and $L_h = \max\left|\mathcal{A}_i\right| : i \in \mathcal{V}$ (which denotes the maximum degree of the anchors nodes); for this fusion case the graph includes both range and orientation anchors of the network. Such constants, resulting in the Lipschitz constant $L_{\hat{f}} = 2\delta_{\max} + \max\left|\mathcal{A}_i\right| : i \in \mathcal{V}$, together with the gradient of $\hat{f}$ enables the use of the parallel algorithm to minimize this cost function described below. Parallelizability is a crucial concern in the development of distributed inference algorithms for sensor networks, but is beyond the scope of this work. Therefore, no advantage is taken here of the specific form of (3.35).

To find the minimum of $\hat{f}$ the followed approach is the Nesterov's optimal method [30], as in [43], since the cost function has a Lipschitz continuous gradient in (3.35). The choice of Nesterov's over conventional descent gradient methods allows for a tremendous speed increase at almost no additional computational complexity cost. At a high level, in this approach each node estimate is given an arbitrary initial value and its position estimate ($x_i(k)$) is updated resorting to a combination (w_i) of the two previous iterations estimates ($x_i(k-1)$ and $x_i(k-2)$). The previous loop can be performed up to a predefined maximum number of iterations or it can stop when the gradient is smaller than a preset threshold. The pseudo-code of this algorithm is presented in detail in Algorithm 1.

[5]The expressions for the Lipschitz constants in [43] were originally derived under the assumption that all sets were balls, but they still hold true even when some of these sets become lines.

Algorithm 1 Parallel method based on accelerated Nesterov's optimal method

Input: $L_{\hat{f}}$, $\{d_{ij} : i \sim j \in \mathcal{E}\}$, $\{r_{ik} : i \in \mathcal{V}, k \in \mathcal{A}\}$, $\{u_{ij} : i \sim j \in \mathcal{E}\}$, $\{u_{ik} : i \in \mathcal{V}, k \in \mathcal{A}\}$;

Output: $\hat{\mathbf{x}}$;

1: $k = 0$;

2: $x_i(0)$ and $x_i(-1)$ is randomly assigned, for each node i;

3: **while** stopping criterion is not achieved, each node i **do**

4: $\quad k = k + 1$;

5: $\quad w_i = x_i(k-1) + \frac{k-2}{k+1}\left(x_i(k-1) - x_i(k-2)\right)$;

6: $\quad w_i$ is broadcast to the neighbours of node i;

7: $\quad \nabla g(w_i) = \delta_{d_i} w_i - \sum_{j \in N_i} w_j + \sum_{j \in N_i} \mathbf{C}_{d_{(i\sim j,i)}} P_{B_{ij}}\left(w_i - w_j\right) + \delta_{l_i} w_i - \sum_{j \in N_i} w_j + \sum_{j \in N_i} \mathbf{C}_{l_{(i\sim j,i)}} Pl_{L_{ij}}\left(w_i - w_j\right)$;

8: $\quad \nabla h(w_i) = \sum_{k \in \mathcal{A}_i} w_i - P_{B_{ik}}(w_i) - Pl_{L_{ik}}(w_i)$;

9: $\quad x_i(k) = w_i - \frac{1}{L_{\hat{f}}}\left(\nabla g_i(w_i) + \nabla h_i(w_i)\right)$;

10: **end while**

11: **return** $\hat{\mathbf{x}} = \mathbf{x}(k)$

In Algorithm 1 δ_{d_i} is the degree of a node i to which range is measured and δ_{l_i} is the same for a node i from which orientation can be obtained. Considering, for the illustrative range case example, $P_B(\mathbf{A}_d\mathbf{x}) = \{P_{B_{ij}}(x_i - x_j)\}_{i\sim j \in \mathcal{E}}$ the concatenated projections onto a ball, multiplying it by $\mathbf{A}_d^T$ will perform simultaneously all the projections relative to the edges of each node i. All these simultaneous operations are denoted by $\mathbf{C}_{d_{(i\sim j,i)}}$, defined as the entry $(i \sim j, i)$ of the arc-node incidence matrix $\mathbf{C}_d$. The previous notation yields the expressions in line 7 of Algorithm 1:

$$\sum_{j \in N_i} \mathbf{C}_{d_{(i\sim j,i)}} P_{B_{ij}}(x_i - x_j) = \left(\mathbf{A}_d^T P_B(\mathbf{A}_d\mathbf{x})\right)_i \text{ and } \sum_{j \in N_i} \mathbf{C}_{l_{(i\sim j,i)}} Pl_{L_{ij}}(x_i - x_j) = \left(\mathbf{A}_l^T Pl_L(\mathbf{A}_l\mathbf{x})\right)_i.$$

Having introduced the newly proposed CLORIS method, a note regarding the reasons for adopting such methodology over an EDM approach seems convenient. Though the outlined plan was to modify the EDM formulation to incorporate angles, we could not succeed in that line of work. At the same time, this disk relaxation set-up proved to be highly attractive due to the natural and seamless fit that it could provide for the extension of the single source (FLORIS) cost function, regardless the distinct relaxation performed afterwards.

Chapter 4

Results

In this Chapter the main simulation and experimental results attained for both FLORIS and CLORIS are presented, in Sections 4.1 and 4.2, respectively.
Extensive tests were run, in simulation, in order to analyse the performance of the hybrid localization methods proposed in this dissertation. All the conditions in which such tests were run, as well as the outcomes, are detailed in Subsections 4.1.1 and 4.2.1.
As an experimental testbed for indoor localization was also developed in the scope of this thesis, findings from applying the new hybrid algorithms on a real scenario are addressed in Subsections 4.1.2 and 4.2.2. Moreover, is included a quantitative performance analysis, benchmarking FLORIS and CLORIS against methods found in the literature, aiming to demonstrate the improvements introduced by the contributions given in this work.
Regarding the experimental implementation, some hardware details as well as physical limitations (some overcome, while others not) are also discussed.

4.1 Single Source Localization - FLORIS

4.1.1 Simulation

4.1.1.1 Preliminaries

In order to assess the performance of the hybrid localization algorithm proposed in this thesis, several simulation tests were performed. Both the position estimation error and the rank of matrix $\mathbf{W}$ were studied. The analysis of the rank of $\mathbf{W}$ allows to assess the formulation in (3.9) and its relaxation, since when $\text{rank}(\mathbf{W}) = 1$, the relaxed solution coincides with the optimal one for the non-relaxed and original problem.
The following experiments were run using MATLAB R2013a and the general-purpose SDP solver CVX/SDPT3[1]. Networks of acoustic and visual anchors were randomly generated in a $[0, 5] \times [0, 5] \times [0, 5]$ cube with white Gaussian noise added both to the distances obtained for the ranging part and to compute the orientations in the visual part. Such measurements (of range and orientation) were modeled to be noisy according to the expression $d = \hat{d}(1 + w)$, where d and $\hat{d}$ are, respectively, the corrupted by noise

[1] http://www.math.nus.edu.sg/ mattohkc/sdpt3.html (accessed in October 2014)

and the actual range (or orientation) measurements and $w \sim \mathcal{N}(0, \eta^2)$ denotes a normally distributed random variable with zero mean and variance equal to the square of the noise factor. This noise model intends to make the simulations as similar as possible to the real scenario and is supported by the studies (discussed in Section 4.1.2) analysing the distance and the angle measurements distribution, in the real set-up. Additionally, we can find in the literature authors also claiming that the position estimates noise increases with the magnitude of the real measured distances.
The accuracy is evaluated computing the RMSE, for every set of MC Monte Carlo runs, which is defined as in (4.1)

$$RMSE = \sqrt{\frac{1}{MC}\sum_{i=1}^{MC} \|x_i - \hat{x_i}\|^2}, \tag{4.1}$$

where x_i and $\hat{x_i}$ are the true and the estimated source positions for the i^{th} run.
Furthermore, regarding analysis of execution times, the following simulations were conducted on a machine equipped with an Intel i5 1.7 GHz CPU and 6 GB of RAM.

4.1.1.2 Results

Firstly, the position estimation error and the rank of $\mathbf{W}$ were analysed for different noise factors. The considered noise factor values were $\eta = (0.001, 0.005, 0.01, 0.05, 0.1, 0.5, 1)$. $MC = 500$ Monte Carlo runs were executed for each value of noise factor and for the same anchor network configuration. Thus, with these tests it is possible to assess the performance of the algorithm relative to the noise factor.
Results for two different network configurations are presented. The first configuration comprises 6 acoustic anchors and 4 visual anchors, these values match the real installed set-up and, for that reason, the study of such configuration was considered relevant. Additionally, the minimal configuration necessary to perform localization in $\mathbb{R}^3$, totalizing 4 anchors (3 acoustic and 1 visual), is also addressed. In this way, the algorithm is characterized for the extreme minimum nodes case. It should be stressed that the latter case refers to a situation in which resorting to the hybrid algorithm it is still possible to estimate a position, even though the number of acoustic anchors is insufficient for pure range-based localization (3 or less for 3D and 2 or less for 2D). To that purpose is enough that one single additional pattern is detected. This is relevant since it might be a common situation in real application scenarios.
Tables 4.1a and 4.1b show that the percentage of resultant matrices $\mathbf{W}$ with rank-1 property (matrix $\mathbf{W}$ was considered rank-1 when performing SVD the first singular value was 16 times greater than the second) is substantially high, thus the relaxation performed (according to what is discussed in Subsection 3.1.1) is valid as the optimal value for the original nonconvex problem is very frequently found.
In Figure 4.1 the RMSE obtained for 500 runs for two fixed networks (with 6 acoustic and 4 visual

Noise Factor	Rank($\mathbf{W}$)=1 (%)
0.001	100
0.005	83.0
0.01	72.4
0.05	62.6
0.1	72.6
0.5	88.6
1	88.4

(a) Network with 6 acoustic and 4 visual anchors

Noise Factor	Rank($\mathbf{W}$)=1 (%)
0.001	100
0.005	100
0.01	97.0
0.05	72.6
0.1	75.4
0.5	94.6
1	98.2

(b) Network with 3 acoustic anchors and 1 visual anchor

Table 4.1: Rank-1 percentage solutions of FLORIS algorithm for fixed configuration networks, for different noise factors

anchors and 3 acoustic anchors and 1 visual anchor), in the considered range of noise factors, is presented. As expected, when the number of anchors is higher the position estimates are more accurate, since more

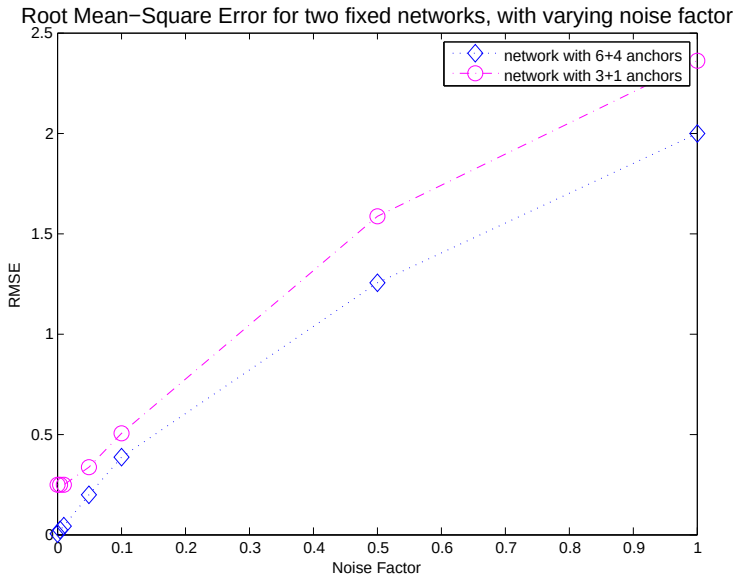

Figure 4.1: Performance of FLORIS algorithm, for two configuration networks, for several noise factors

information is used. Nevertheless, this result shows that the FLORIS is able to perform reasonably accurate localization with 4 anchors (when only 3 acoustic nodes are available).

For comparison with the fixed configuration case (shown in Figure 4.1), 500 Monte Carlo runs were executed for each value of noise factor, yet, this time for varying network configurations in every run, keeping the number of both acoustic and visual anchors constant. With these tests we aim to observe the algorithm performance under configuration diversity.

Tables 4.2a and 4.2b report the relative frequency for rank($\mathbf{W}$)=1 and Figure 4.2 depicts the RMSE for the same range of noise factors, but now for varying network configurations.

Noise Factor	Rank($\mathbf{W}$)=1 (%)
0.001	88.2
0.005	75
0.01	75
0.05	65.4
0.1	71.2
0.5	82
1	67.8

(a) Network with 6 acoustic and 4 visual anchors

Noise Factor	Rank($\mathbf{W}$)=1 (%)
0.001	45.6
0.005	39.2
0.01	42.4
0.05	53.2
0.1	64.4
0.5	88.8
1	91.8

(b) Network with 3 acoustic anchors and 1 visual anchor

Table 4.2: Rank-1 percentage solutions of FLORIS algorithm for different network configurations, for different noise factors

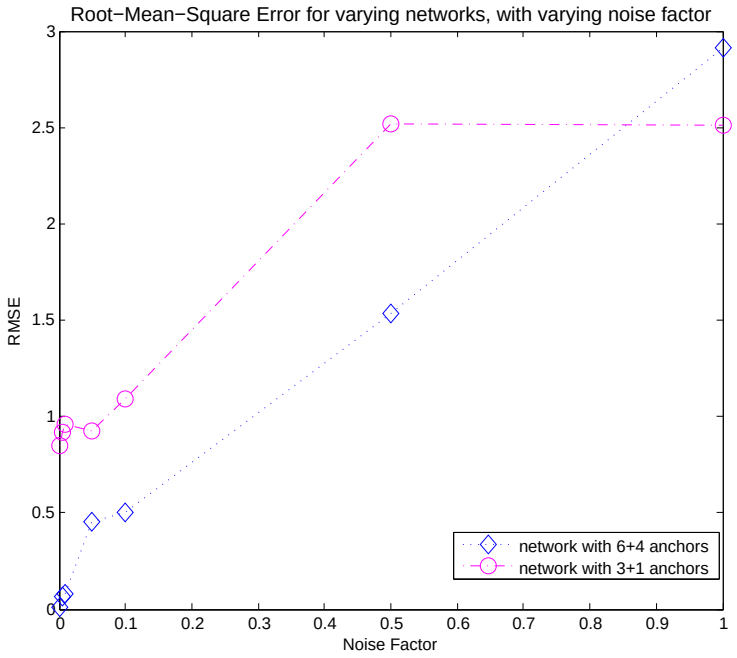

Figure 4.2: Performance of FLORIS algorithm, for different configurations networks of 6+4 and 3+1 anchors, for several noise factors

From Tables 4.2a and 4.2b and Figures 4.1 and 4.2 it is possible to conclude that the fixed configuration networks seem to be favourable to the hybrid algorithm as the rank-1 percentages are higher and the RMSE lower than for the varying case. Nonetheless, for the variable configurations with 6+4 anchors, the percentage of rank and the RMSE are not far from the values obtained for the fixed case. Considering the 3+1 anchors configuration, the values show larger discrepancy, being, less frequently, the achieved relaxed solution equal to the optimal solution. This can be explained as such configuration is the minimal one in which 3D localization is plausible, and, being a corner case, it can generate quite adverse situations for the method.

Aiming at a more detailed analysis of behaviour in individual runs, Figures 4.3 and 4.4 present, respectively, the histograms of the rank of $\mathbf{W}$ and the norm of the error for each Monte Carlo run, for a configuration with 6 acoustic and 4 visual anchors, and a noise factor values fixed at $\eta = (0.01, 0.1)$.

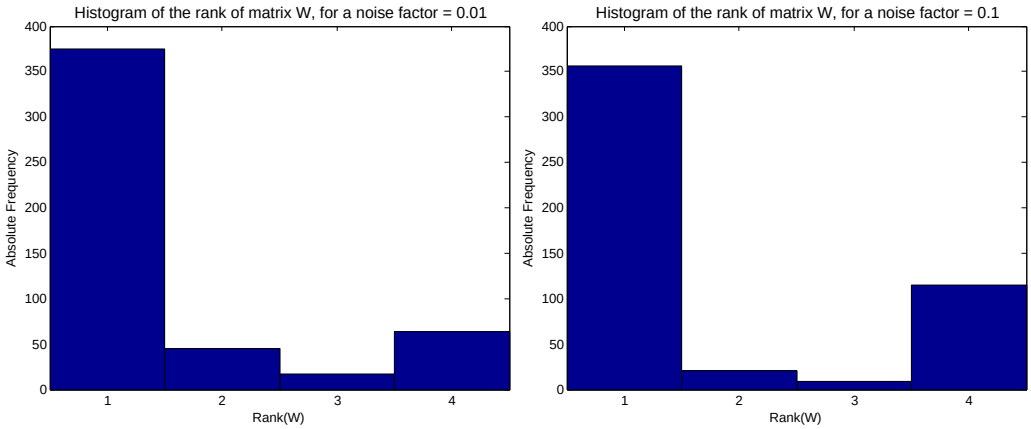

Figure 4.3: Histogram of the rank of matrix $\mathbf{W}$ for a configuration with 6 acoustic and 4 visual anchors, for noise factors 0.01 and 0.1

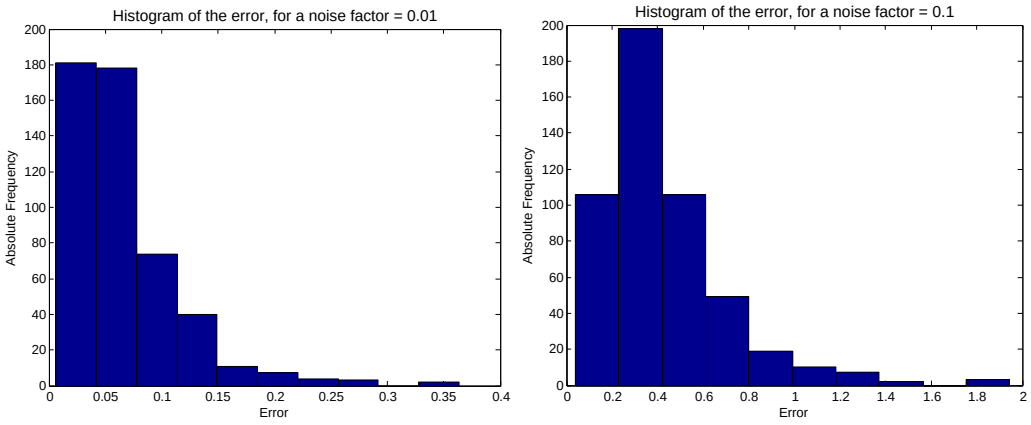

Figure 4.4: Histogram of error norm for a configuration with 6 acoustic and 4 visual anchors, for noise factors 0.01 and 0.1

From Figures 4.3 and 4.4 it is possible to conclude that the error norm is clustered around the low values (up to 0.1 for $\eta = 0.01$ and 0.6 for $\eta = 0.1$), which is associated with the tightness of the relaxation depicted in the first histograms.

With a constant noise factor (of 0.1) and varying the number of anchors of the configuration it is possible to analyse the evolution of the estimation error depending on the total number of anchors and also on their type. Figure 4.5 shows the RMSE obtained for $500MC$ trials for each network formed by the combination of 5 to 15 acoustic and visual anchors. From this simulation we can state that the ideal configuration within this node range, i.e. the network for which the RMSE reaches the minimum, was found to be 15 acoustic and 13 visual anchors.

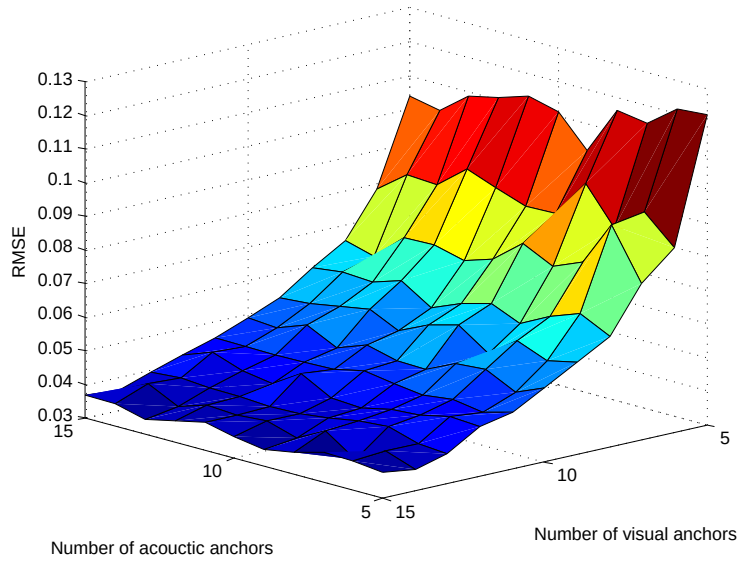

Figure 4.5: RMSE for network configurations combining different number (between 5 and 15) of acoustic and visual anchors

Regarding the running time of FLORIS, it was observed that a 3D position estimate is, usually, obtained in approximately 1 second (for networks of approximately 10 nodes), using a machine with the characteristics described in 4.1.1.1. Furthermore, the running time slightly increases with higher noise factors and the same trend is observed for higher number of anchors, as shown in Figures 4.6 and 4.7.

These behaviours are consistent to what was expected, as higher noise factors make the convergence to the solution harder and higher number of network nodes increase the number of variables of the optimization problem, thus augmenting the complexity of the algorithm.

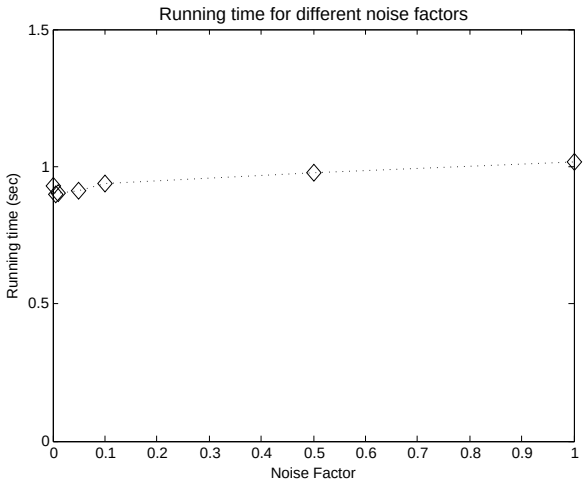

Figure 4.6: Average running time, for different noise factors, for a network of 3 acoustic anchors and 1 visual anchor

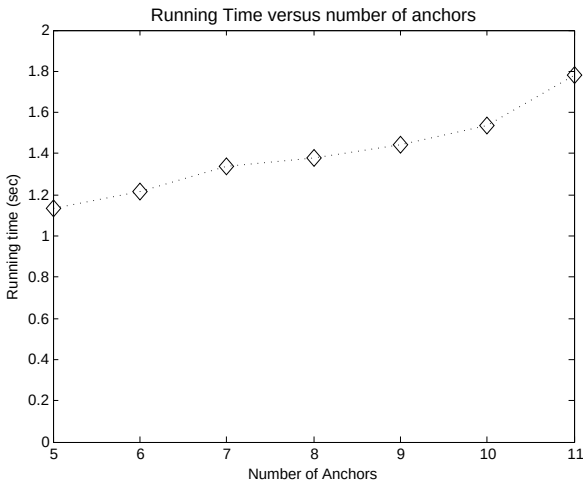

Figure 4.7: Average running time versus number of anchors (with 2 visual anchors from the total)

In order to assess the newly proposed FLORIS, a comparison in terms of accuracy is performed regarding other state-of-the-art algorithms, namely SR-LS from [5] and SLNN from [34]. The latter methods were chosen due to the performance analysis made in [34], which shows that the SLNN method has (for 3D space) higher accuracy than the previous proposed methods with similar formulations, while SR-LS provides less accurate but much faster results. Therefore, if an accurate estimate, under uniform Gaussian noise, is pursued then SLNN is claimed to be the preferred, yet if the time constraints prevail, the chosen method should be SR-LS.

Figure 4.8 depicts RMSE values, for several values of noise factor $\eta \in [0.001, 1]$, and for 500 randomly generated network configurations, comprised by 6 acoustic and 4 visual anchors, for each factor.

Naturally, for the 3+1 network configurations such comparison shall not be conducted, as, due to the ambiguity in 3D space in such instances, the non-hybrid algorithms fail to produce a correct position estimate.

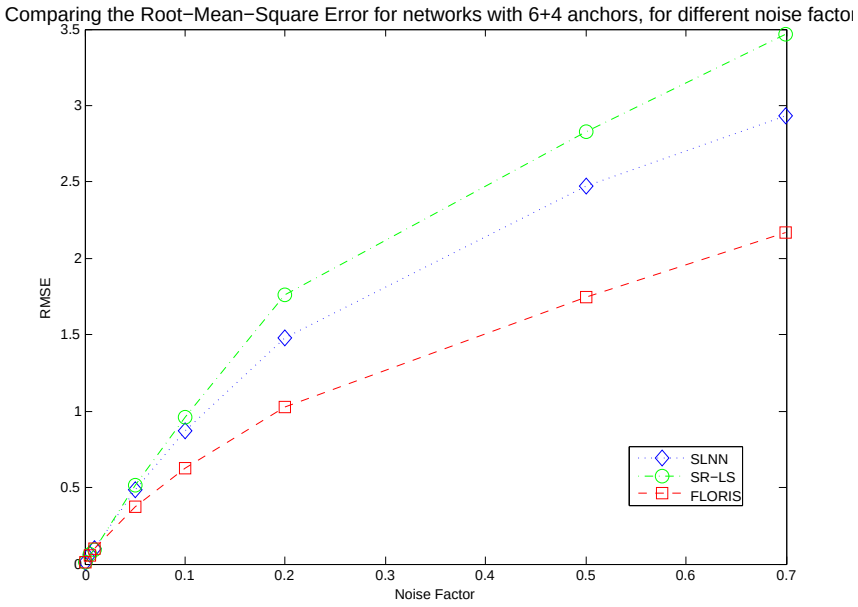

Figure 4.8: 3D source localization performance; comparison among the RMSE of the three methods, for various networks of 6 acoustic and 4 visual anchors and for several Gaussian noise factors

In this case there is a significant improvement in the position estimation accuracy induced by the addition of the bearing to the range information, through the proposed formulation. Therefore, we can state that, under the usual random conditions, FLORIS outperforms, for all noise factors, the other two state-of-the-art methods, found to be the most accurate for this scenario.
Nonetheless, it can be claimed that the latter comparison might be somewhat unfair, as SLNN and SR-LS take into account less information (the 6 acoustic anchors alone) than the hybrid method (which adds information from other 4 anchors). Thus, Figure 4.9 depicts the simulation results of comparing the RMSE of SLNN and SR-LS with FLORIS for an equal number of anchors. In this case, a network of 6 acoustic anchors was generated to test SLNN and SR-LS and a network in which 3 of the latter acoustic anchors were converted into visual ones is used to test the hybrid approach.

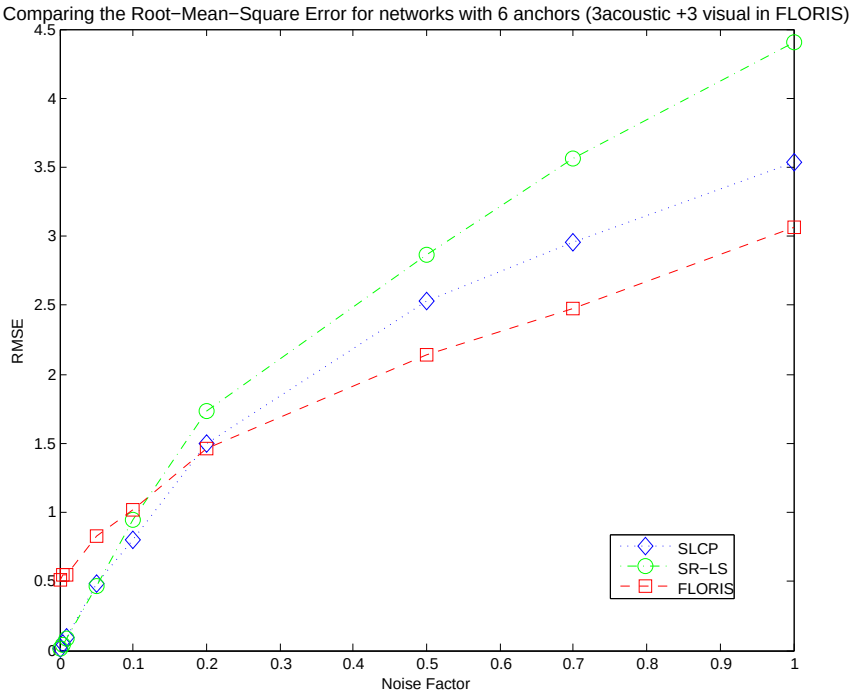

Figure 4.9: 3D source localization performance; comparison of the RMSE among three methods, for various networks of 6 acoustic or 3 acoustic + 3 visual anchors, for several Gaussian noise factors

Under these conditions, the performance of FLORIS is closer to the other methods. In fact, it can be noticed that for very low noise factors, the accuracy of FLORIS is slightly lower than for SLNN and SR-LS. On the other hand, the proposed method seems to be consistently more robust for noisier measurements. This is a very interesting and important finding, specially since this work aims at experimentally implementing the hybrid localization method in real cases scenario, which inherently corresponds to a situation with high noise factor.

Furthermore, an analysis of the running times for this group of algorithms is executed, in order to demonstrate the practical feasibility of the newly proposed localization algorithm.

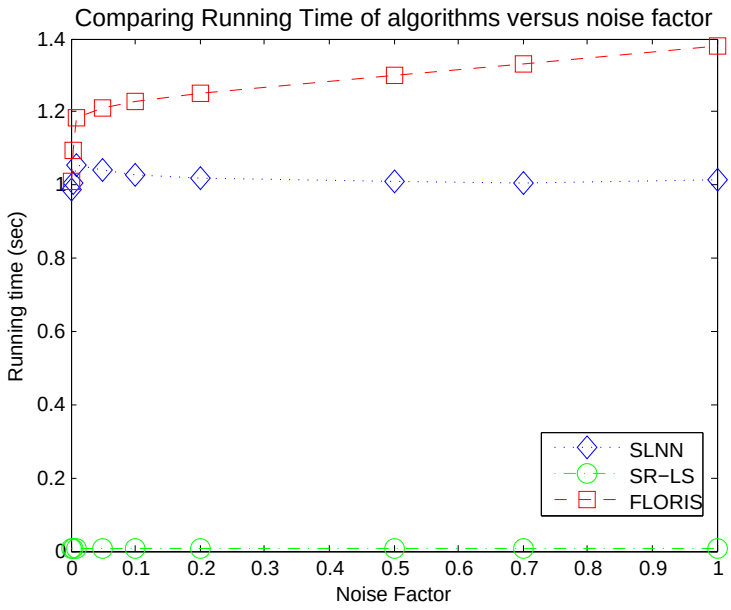

Figure 4.10: Running times comparison, for networks of 6 acoustic or 3 acoustic + 3 visual anchors for FLORIS, for varying noise factor values

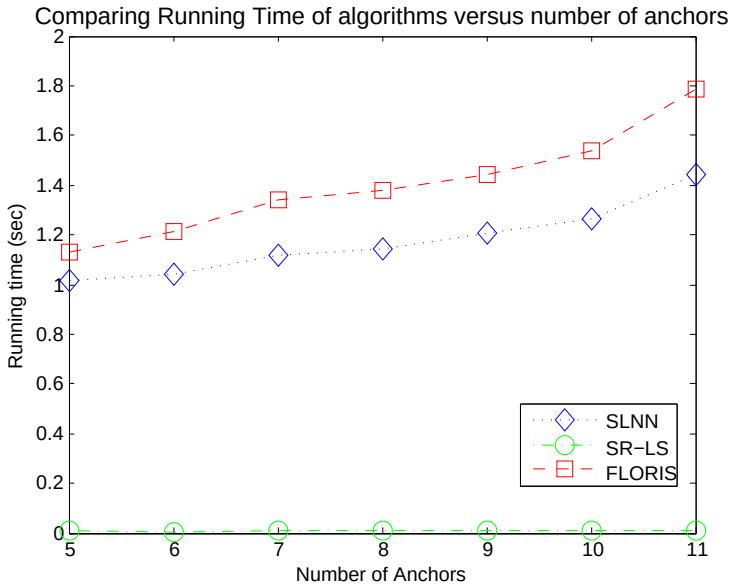

Figure 4.11: Running times comparison, for networks of different number of anchors (for FLORIS 2 of the acoustic anchors were converted into visual)

It can be observed in Figures 4.10 and 4.11 that the complexity of FLORIS seems to be comparable to the SLNN's, since the running times are similar (showing a fairly constant behaviour). As for the SR-LS, it shows a very high speed when compared with the former, which corroborates the discussion in [34] about the tradeoff between speed and accuracy for this method. In summary, given its RMSE, SR-LS should be chosen only when time or computational resources are constrained.
It should be noted that the discrepancy between the obtained values and the ones presented in [34] stems from the different CPUs in which the experiments were conducted.
In Figure 4.10 there is a small increment of the hybrid method running times for higher noise factors however, for such cases the new algorithm provides better position estimates than the other methods. Thereby, if resources are not severely constrained, FLORIS should be preferred in such cases. Such

conclusion can be inferred by looking at Figure 4.12, which compares the achieved RMSE with the execution speed of the algorithms, for a noise factor of 0.2.

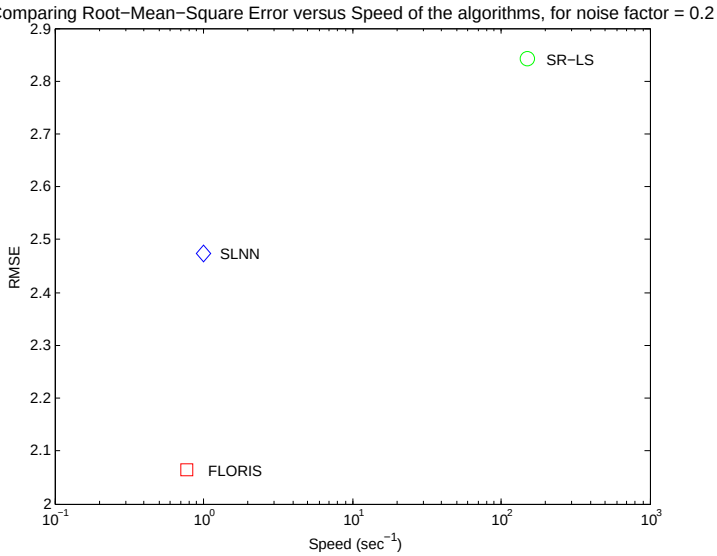

Figure 4.12: 3D single-source localization performance comparison: RMSE versus speed for the three addressed algorithms

In the same way, the running times of FLORIS, in Figure 4.11, depict an evolution similar to the SLNN case, with increasing number of anchors. According to [34] there is a fixed overhead due to software implementation which is substantial for configurations with up to around 10 anchors. Above 15 anchors the variable component becomes more meaningful and, for that reason, the study of the running time as a function of the number of anchors is done only up to 11 anchors.
Looking at Figure 4.11 we can infer that introducing visual anchors, and hence a visual constraint, carries a certain time overhead, as the running time of FLORIS appears larger than the SLNN running time for an approximate constant amount.
Additionally, the case in which the majority or even all acoustic anchors are coplanar, is addressed and studied. Although this extreme case reflects a reasonably likely real system configuration, since these sensors tend to be placed on a single plane, commonly the ceiling, the proposed algorithm, similarly to what already occurred for SLNN (3D version of SLCP), determines poor estimations of the target positions; this is due to the fact that under such circumstances the rank of matrix $\mathbf{W}$ will not be 1 so frequently. However, since it is highly likely that visual anchors will be placed in different planes, with this fusion approach, such problem can be now alleviated.
Indeed, this problem was observed in the experimental set-up, where, initially, all the acoustic beacons were placed in a single plane. In order to be able to compare the proposed method with existing ones, namely SLNN, the experimental scenario had to be adapted, moving the anchors to non-coplanar positions (as is discussed in Subsection 4.1.2).
Besides the 3+1 anchors and the coplanar cases, there is, yet, another situation for which FLORIS algorithm provides a good solution for a formerly problematic configuration. Effectively, when the target lies outside the convex hull formed by the acoustic anchors, the position estimates obtained by the previous methods present lower accuracy. If simply one visual anchor is placed outside the convex hull, FLORIS can, immediately, compute a better estimate of the target position. Again, this is a very realistic assumption, which adds more value to the proposed hybrid method.
Generating networks under such conditions (where a visual tag is placed outside the convex hull as well as the source) for 500 Monte Carlo runs with a noise factor equal to 0.1, resulted in the RMSEs presented

in Table 4.3, which substantiates the aforementioned claims.

	SR-LS [5]	SLNN [34]	FLORIS
RMSE (m)	0.0060	0.0064	0.0012

Table 4.3: 3D source localization accuracy comparison, for configurations with 6 acoustic anchors, where the target and the tag are placed outside the acoustic anchors convex hull, for noise factor equal to 0.1

In a nutshell, this Section demonstrates, supported by thorough experiments, that FLORIS is a localization method with very high potential, as it compares favourably with state-of-the-art methods. Moreover, it is verified that adding the orientation information captured through vision is a valuable asset/benefit. Experiments adding just one visual anchor to the acoustic constellation resulted in very significant improvements in the accuracy of the position estimates.
Although there are cases for which FLORIS is less accurate than SLNN and presents similar running times, other cases are identified for which FLORIS should be clearly preferred.

4.1.2 Experimental Results - Tests with the Cricket system and tags

4.1.2.1 Cricket system and ARUCO library

Cricket is a location support system for mobile and location dependent applications [36], where distances are estimated based on the difference between the time of arrival (ToA) of radio and ultrasonic signals. The radio runs at a 433MHz frequency and these devices are designed for indoor localization, as the maximum US range is approximately 10 m, when the listener and the beacons are facing each other with no obstacles in the way. In the *Cricket v2 User Manual* [2] it is claimed that Cricket provides an accuracy on the order of 1 cm for measurements up to 3.5 m and 2 cm for larger distances. Due to the simplicity of setting up and operating with such system, it was used in the prototype to prove this thesis concept. Therefore, the first prototype consisted in a listener Cricket node mounted solidary with a video camera, that could roam in the area covered by the Cricket beacons and with visual tags. In this way, both range measurements to the multiple acoustic anchors, deployed as Cricket beacons, and orientations between the camera and the tags or objects could be, simultaneously, acquired.
The first tests and experiments were performed using the prototype previously described as the moving target, as shown in Figure 4.13. The environment created for the target localization included tags (from the ARUCO library [21] - an Augmented Reality applications library based on OpenCV[3]) spread through the space and Cricket beacons. According to Figure 4.14, the beacons were, likewise, spread through the space, the majority attached to the ceiling (for practical reasons, such as permanent fixation, since it is extremely important that the anchors maintain a fixed and known position). Yet, not all beacons could be coplanar due to the relaxation limitation of SLNN for these cases.

[2]http://cricket.csail.mit.edu/v2man.pdf (accessed in October 2014)
[3]http://www.uco.es/investiga/grupos/ava/node/26 (accessed in October 2014)

Figure 4.13: Testing the first prototype, consisting of a Cricket listener mounted on a camera

Figure 4.14: Cricket beacons attached to the ceiling in the first experimental set-up

In order to obtain the rotation of the tag, so that the orientation required as input for the fusion localization algorithm can be determined, the referential and corners in the tag were considered according to Figure 4.15. The detection of the tag and the determination of its rotation matrix is performed using the ARUCO library, which is based on OpenCV [4].

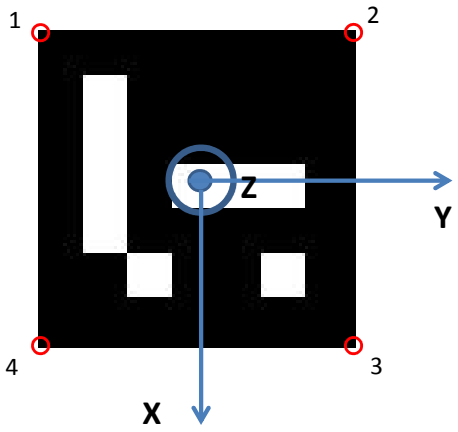

Figure 4.15: Aruco tag with referential used and corners numbered

[4]http://opencv.org/ (accessed in October 2014)

4.1.2.2 Practical setbacks and solutions

Initially, according to the plan outlined, the experimental set up only had Cricket nodes whose measurements should be sufficient to compute the camera position and tags to provide the orientation. However, during the first experiments, the Cricket system exhibited a practical limitation which jeopardized the tests. More specifically, the ultrasonic US signal was found to be highly directional and, consequently, measuring the ranges between a beacon and the listener node was almost impossible unless the US transmitter and the receiver were properly aligned.
Such limitation from the hardware proved to be a restriction to the outlined experiments. The solution found for this practical problem was the installation of a parabolic reflector, as shown in Figure 4.16, in each Cricket beacon attached to the ceiling aiming to scatter the ultrasounds. Although the reflector creates a small multi-path, this is a minor issue considering that with this alteration the listener is, contrary to what happened previously, able to receive the ultrasonic impulses from several beacons in almost every position within the limits of the set-up.

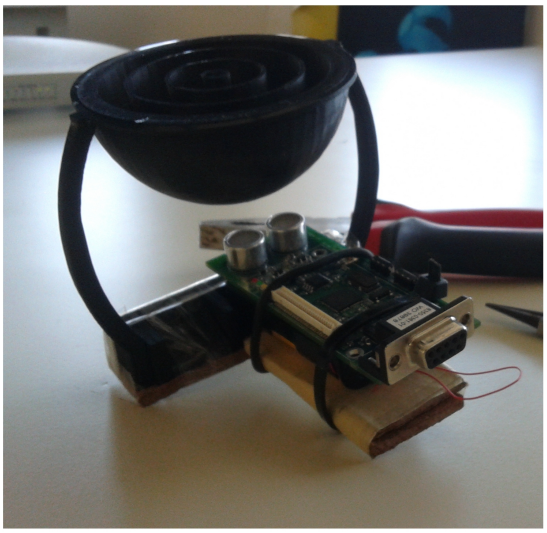

Figure 4.16: Custom-made reflector installed in the Cricket beacons

When a position estimate was, finally, obtained, another question was raised, related with the quantitative evaluation of the obtained estimate: "How can we know the true position?" To tackle this issue a new prototype was devised. This time, the set camera plus Cricket was attached to a structure covered by tags, equally from the ARUCO library. This device is showed in Figure 4.17.

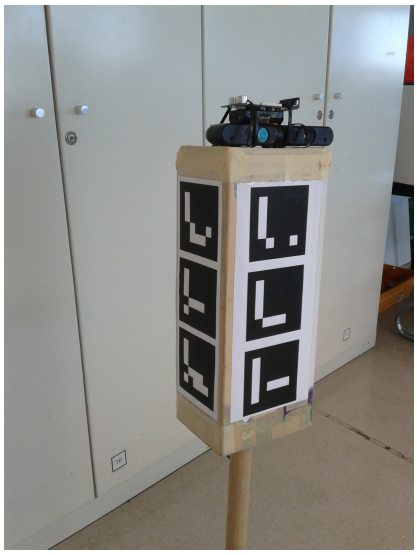

Figure 4.17: Prototype with camera and Cricket listener attached to a parallelepiped covered by tags, whose position can be determined

In this way, in addition to the prototype position determined using the ranges given by the hybrid method, it is now possible to have another estimate for this position, that can be used as the ground truth, based on the rotation and translation of the structure's tags. To do so, the box should be observed in the different positions by a static camera. Primarily, a 3D map of the layout of the tags on the box was generated using functions, which resort to RANSAC [19], developed in the work [38] and kindly provided by the author. With this map it is possible to determine the position of a camera observing the box, relative to a referential defined using one of the tags as reference, since all the rotations and translations among the tags are computed. Hence, even when this second camera cannot see the reference tag, it is still possible to determine the camera position relative to the referential (which in this case is not directly observed) using the map with the transformations of the observed tags to the reference one.
This is also extremely useful for the sensor network localization paradigm, in which a second sensor must obtain the angle to this first sensor. The procedure followed in such approach in Subsection 4.2.2 uses precisely this method, in which a second sensor can be referred in the referential system of the prototype shown in Figure 4.17.
Another improvement implemented was the replacement of single tags by boards of tags. These boards lead to better and more accurate tag detection, as failures from environmental conditions (such as lighting conditions) are less probable with multiple tags and more points (corners of the tags) are available to compute the camera extrinsic parameters. Such improvements were experimentally confirmed, comparing the angle variance, at several distances, when a single marker or a board is used. Figure 4.18 illustrates a typical distribution of angles measured to a board of 4 ARUCOS from a specific distance, all under the same environmental conditions. The resulting histogram resembles a Gaussian distribution. On the other hand, under different environmental conditions the angle variance is larger and the histogram is more different from a normal distribution, as shown in Figure 4.19.

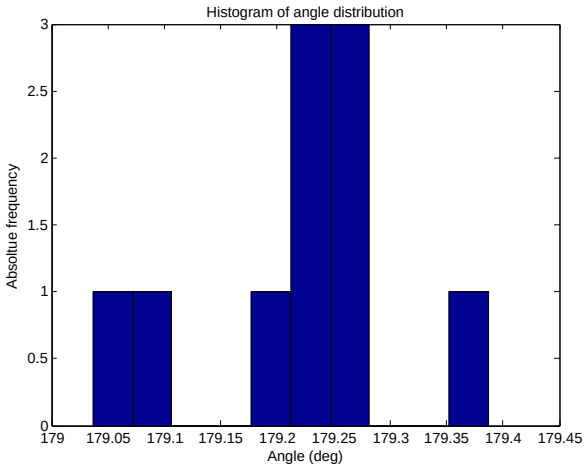

Figure 4.18: Histogram of measured angle distribution for a board 1.53 m away and under the same environmental conditions

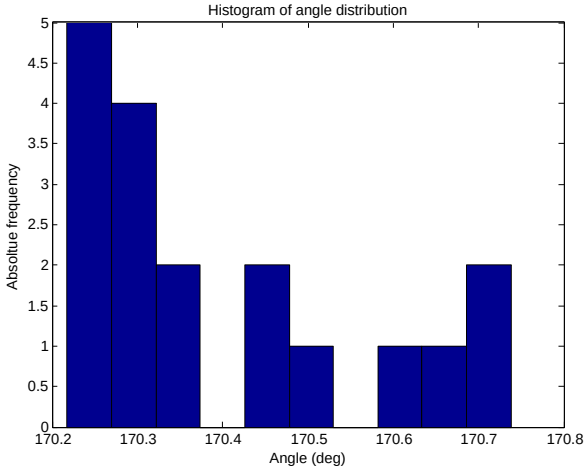

Figure 4.19: Histogram of measured angle distribution for a board 2.78 m away and under different environmental conditions

Table 4.4 summarizes the angle variance found for three different distances when only one tag and a board (with 4 tags) were detected, under uniform and non-uniform environmental conditions.

	1 tag		Board	
	Same Cond	Diff Cond	Same Cond	Diff Cond
1.53 m	0.0403	0.0669	0.0090	0.0401
2.10 m	0.1069	8.0474	0.0023	0.0238
2.78 m	31.9533	20.6532	0.0052	0.0286

Table 4.4: Variances of the measured angle (in degrees) from one tag and from a board of tags, for different distances and both with uniform and non-uniform environmental conditions

It can be observed that using a board causes a smaller dispersion of angle values, particularly when the distance increases. Additionally, the variance when the environmental conditions, such as illumination (light or shadow) vary is always larger than when the conditions are maintained, but when the board is used this effect is substantially lower.

This study also excludes the need of a possible parametric formulation of FLORIS algorithm (with a parameter affecting the angle parcel depending on the distance), considering that, in this way, the reliability of the orientation detected is not affected by the distance (up to approximately 5m of distance). Consequently, a 3D map of the board of ARUCOS on the wall was built, similarly to what is performed for obtaining the ground truth position. In this way, it is not necessary to detect the tag defined as reference conferring more flexibility to the camera movements.

Figure 4.20 shows the final experimental set-up used to acquire range and angle information to perform the tests of the new FLORIS algorithm.

Figure 4.20: Final experimental set-up, with Cricket beacon nodes, visual tags and the target comprising a Cricket listener and a camera

After some experiments with the testbed it was noted that the ranges between the anchors and the moving target provided by the Cricket system had lower accuracy than the one claimed by the manufacturer (1 cm for measurements up to 3.5 m and 2 cm for higher distances). In some cases, the measured distance would have more 10 cm than the real one, thus hindering the position estimation performance. This issue was also impacted by the reflectors mounted on the ceiling beacons, which add a small multipath, affecting the sensed range.
In order to tackle this problem a Cricket calibration was performed. This calibration consisted in collecting, for each single anchor, several pairs (25) of real and measured ranges, covering the whole set-up area, in order to be representative for all distance measurements.
A preliminary statistic study was performed, in order to characterize each beacon (since, contrary to what is claimed, the beacons show non-uniform behaviours among them concerning the range accuracy, as it was confirmed by this study). Aspects such as measurements distribution and beacon clustering (according to minimum within cluster variance) were considered and are shown in Figures 4.21 and 4.22.

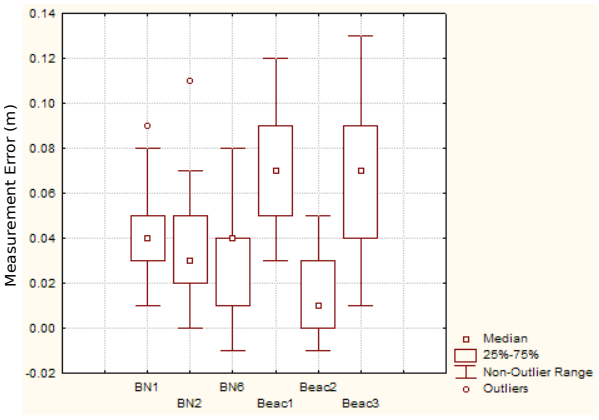

Figure 4.21: Box plots of the distance measurements collected for each of the 6 beacons of the network

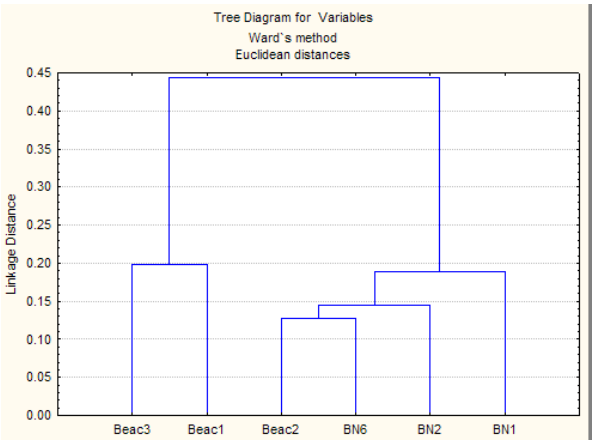

Figure 4.22: Clustering obtained for the 6 beacons, minimizing within cluster variance

It is interesting to note that the attachment of the reflector to the beacons is reflected in the similarity identified by the clusters in Figure 4.22. More precisely, *Beac2* and *BN6* do not have a reflector, since they are oriented vertically to the floor, directly to the target. That is also the case of *BN1*, however this

was mounted with a horizontal orientation, possibly causing the disparity relative to the other beacons. On the other hand, the signal from *Beac1, Beac3* and *BN2* is transmitted up to the target using the reflector. We may conclude that the introduction of the reflector somehow affects the measurement accuracy, however this can be minimized performing the beacon calibration.
Furthermore, to perform the Cricket system calibration itself, a linear regression using the data collected was performed for each beacon. This way, we can model each beacon separately and then correct/rectify the obtained range. Even though the method applied was straightforward, the localization results showed a reasonable improvement when the measurements were corrected according to the calibration performed. The quantitative analysis is presented and discussed in Subsection 4.1.2.4.
Despite all these efforts and improvements, the obtained measurements in the experimental set-up are still quite noisy. Yet, as it is discussed in Subsection 4.1.2.4, the fusion algorithm handles this type of noisy data quite well, including obtaining higher accuracy position estimation than state-of-the-art algorithms, in the same conditions.

4.1.2.3 Sensor Networks Self-Calibration

Necessarily, prior to performing tests in this environment, with both Cricket nodes and visual tags, a calibration must be conducted so that the transformation between both networks (of Crickets beacons and visual tags) is known and hence the collected data and results are meaningful and valid.
Prior to performing the self-calibration in the experimental set-up, the devised method introduced in Subsection 3.1.2 was validated, both with entirely numerical data and with numerical data for the ranges mixed with real data for the visual information (the notation used below is the same as introduced in 3.1.2).
Regarding, in the first place simulated data, a fixed translation between the two sensors ($\mathbf{T}_{\text{cc}}$) and translation vectors and rotation matrices between the acoustic and the visual networks ($\mathbf{R}_{\text{ac}}$ and $\mathbf{T}_{\text{ac}}$) were randomly generated in a $[0,\,10] \times [0,\,10] \times [0,\,10]$ cube. For the k-th observation random camera positions ($\mathbf{R}_{\text{cam}_\text{k}}$ and $\mathbf{T}_{\text{cam}_\text{k}}$) were synthesized so that acoustic node positions could be computed, following the model $\mathbf{P}_{\text{acoustic}_\text{k}} = (\mathbf{R}_{\text{ac}}\,(\mathbf{R}_{\text{cam}_\text{k}}\mathbf{T}_{\text{cc}} + \mathbf{T}_{\text{cam}_\text{k}}) + \mathbf{T}_{\text{ac}})\,(1+w)$, where $w \sim \mathcal{N}(0, \eta^2)$. Given an initialization for $\mathbf{T}_{\text{cc}}$ (which is admissible as its dimension tends to be an order of magnitude smaller than the other two unknowns), the self-calibration method iterates through estimating $\mathbf{R}_{\text{ac}}$ and $\mathbf{T}_{\text{cc}}$, until the difference between two consecutive estimates of $\mathbf{R}_{\text{ac}}$ is smaller than a predefined threshold. When the latter stopping criterion is verified, $\mathbf{T}_{\text{ac}}$ is computed, using the previously obtained $\mathbf{R}_{\text{ac}}$ and $\mathbf{T}_{\text{cc}}$ estimates. Performing 1000 Monte Carlo runs of such procedure, making $k = 100$ observations and noise factor 0.01, the errors obtained for the estimated unknowns were computed as $\mathbf{E}_{\mathbf{R}_{\text{ac}}} = \frac{1}{MC}\sum_{i=1}^{MC} \|\mathbf{R}_{\text{ac}_\text{i}} - \mathbf{R}_{\text{acGT}_\text{i}}\|_F$ for the rotation matrix and $\mathbf{E}_{\mathbf{T}} = \frac{1}{MC}\sum_{i=1}^{MC} \|\mathbf{T}_i - \mathbf{T}_{\text{GT}_\text{i}}\|$ for the translation vectors, were found to be:

$$\mathbf{E}_{\mathbf{R}_{\text{ac}}} = 0.0076 \qquad \mathbf{E}_{\mathbf{T}_{\text{ac}}} = 0.0448 \text{ m} \qquad \mathbf{E}_{\mathbf{T}_{\text{cc}}} = 0.1810 \text{ m}.$$

The previous results validate the derived self-calibration method, obtaining highly accurate estimates.
Furthermore, to better model this process to the real case, the self-calibration procedure was performed with camera positions ($\mathbf{R}_{\text{cam}_\text{k}}$ and $\mathbf{T}_{\text{cam}_\text{k}}$) acquired in the experimental set-up, using the ARUCO library. Again, 1000 Monte Carlo runs were executed with $\eta = 0.01$ for simulating with partial real data.

$$\mathbf{E}_{\mathbf{R}_{\text{ac}_{\text{half-real}}}} = 0.1027 \qquad \mathbf{E}_{\mathbf{T}_{\text{ac}_{\text{half-real}}}} = 0.3361 \text{ m} \qquad \mathbf{E}_{\mathbf{T}_{\text{cc}_{\text{half-real}}}} = 0.4359 \text{ m}.$$

The latter results show lower accuracy when compared to the totally simulated case. This may be partially explained by the total number of observations, since simulation with half experimental data comprised

28 pairs of observations and decreasing number of observations leads to a degeneration of the estimations (the Procrustes problem suffers from ill approximations for small datasets).

With the latter set up (in Figure 4.20), the procedure to determine both the rotation matrix and translation vector between the two detached networks follows, once more, the self-calibration method presented in Subsection 3.1.2. Thus, it is possible to determine the transformation between the two detached networks ($\mathbf{R}_{\text{ac}}$ and $\mathbf{T}_{\text{ac}}$) as well as the translation between the listener and the camera ($\mathbf{T}_{\text{cc}}$), as there are two different sensors assembled together, using only the data acquired regarding the listener and the camera positions. Particularly, to perform this pre-step both ranges from the acoustic anchors and rotations and translations from the visual tags detection are collected. Single-source localization methods from the state-of-the art, such as the SLNN [34] or the SR-LS [5] can be applied to estimate the Cricket listener position from the range information (in this case the SR-LS approach was used). As for the information relative to the camera, ARUCO library provides the rotation and the translation of the camera relative to the markers ($\mathbf{R}_{\text{cam}}$ and $\mathbf{T}_{\text{cam}}$), which are used to determine the camera positions. The Cricket positions estimated in the Cricket referential ($\mathbf{P}_{\text{acoustic}}$) and the camera positions estimated in the camera referential ($\mathbf{T}_{\text{cam}}$) for one self-calibration test conducted are depicted in Figures 4.24 and 4.23. Some target position estimations (namely positions number 9 and 10), seen in Figure 4.24, are clearly outliers, probably due to exceedingly noisy range measurements. Nevertheless, as it will be shown hereinafter, the outcomes of the self-calibration procedure are robust to the presence of such position estimates.
Figure 4.23 also depicts the map created for the board of ARUCOS on the wall. The origin of the visual anchors referential was defined to be the center of the tag with ID 0.

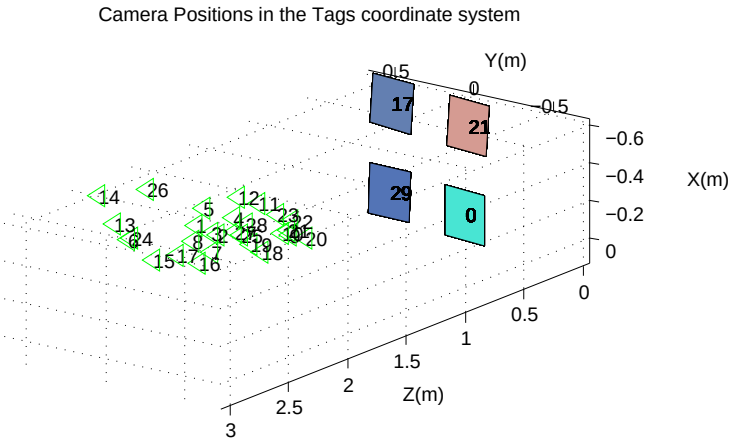

Figure 4.23: Camera positions in the tags coordinate system, collected for the self-calibration input

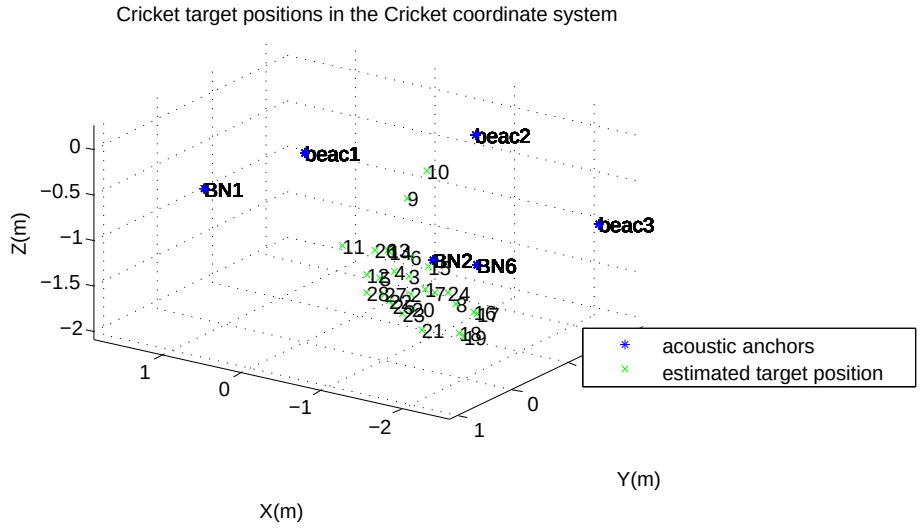

Figure 4.24: Cricket target positions in the Cricket coordinate system, collected for the self-calibration input

The Cartesian coordinates forming the acoustic anchor constellation were obtained from centering and averaging matrix $\mathbf{D}$ (an EDM consisting of all the pairwise Euclidean distances measured between the Cricket beacon nodes), resulting in $\tilde{\mathbf{D}}$ and then applying a SVD to this new projected and averaged matrix ($\tilde{\mathbf{D}} = \mathbf{U}\mathbf{\Sigma}\mathbf{V}^*$). The reconstructed beacon coordinates ($\mathbf{x}$) are computed as $\mathbf{x} = \mathbf{U}\mathbf{\Sigma}^{\frac{1}{2}}$. Later, this constellation will have to be synchronized, so that the results can be evaluated.

Taking pairwise differences of both target and camera position estimates results in vectors $\mathbf{P}_{\text{diff}}$ and $\mathbf{T}_{\text{diff}}$, respectively. Additionally, the differences of rotation between the camera and the markers ($\mathbf{R}_{\text{diff}}$) are also determined from the obtained rotations of the ARUCO library. The totality of the inputs of the self-calibration process are, then, computed, and hence the transformation between the two sensor networks (consisting in $\mathbf{R}_{\text{ac}}$ and $\mathbf{T}_{\text{ac}}$) and the translation between the listener and the camera ($\mathbf{T}_{\text{cc}}$) can be readily computed for this experimental set-up.

In order to maximize the accuracy of $\mathbf{R}_{\text{ac}}$ and $\mathbf{T}_{\text{ac}}$ estimates in this practical case, and since $\mathbf{T}_{\text{cc}}$ was easily available, the value measured with a tap measure between the acoustic sensor and the focal point of the camera was assigned to $\mathbf{T}_{\text{cc}} = \begin{bmatrix} 0.0800 & -0.0435 & -0.0180 \end{bmatrix}$m.

The analysis and validation of the achieved results for $\mathbf{R}_{\text{ac}}$ and $\mathbf{T}_{\text{ac}}$ proved to be difficult, as no ground truth or intuitive reference could be obtained. Therefore, in order to facilitate the evaluation of the outcomes, the Cricket network referential was aligned with the visual network referential, by performing the synchronization of the aforementioned beacon constellation. For this purpose, one of the Cricket beacons was chosen to be the origin of the coordinate system and the coordinates of other 3 beacons, relative to the new referential (defined in order to be aligned with the origin of the visual tags coordinate system), were measured. Such measured coordinates were taken as reference and used as input, alongside the Cricket positions of the same 4 beacons in the previous non-aligned referential, to solve a Procrustes problem. In this way, rotating all the constellation with the resultant rotation matrix leads to the alignment of the Cricket coordinate system and with the Cricket positions estimates ($\mathbf{P}_{\text{acoustic}}$) obtained in this new aligned referential, the $\mathbf{R}_{\text{ac}}$ and $\mathbf{T}_{\text{ac}}$ computed can now be compared with a reference and thus evaluated.

For this case, the computed rotation and translation between the Cricket network and the visual tags coordinate systems were

$$\mathbf{R}_{\text{ac}} = \begin{bmatrix} -0.0241 & -0.9977 & 0.0640 \\ 0.0941 & -0.0660 & -0.9934 \\ 0.9953 & -0.0179 & 0.0954 \end{bmatrix}, \qquad \mathbf{T}_{\text{ac}} = \begin{bmatrix} 0.7495 & 2.0781 & 1.8151 \end{bmatrix} \text{ m}.$$

If the two coordinate systems are considered to be perfectly aligned (which is an assumption that cannot be completely confirmed), $\mathbf{R}_{\text{ac}}$ can be obtained by inspection. Moreover, $\mathbf{T}_{\text{ac}}$ was measured manually with a laser distance meter. These two quantities ground truths were found to be:

$$\mathbf{R}_{\text{acGT}} = \begin{bmatrix} 0 & -1 & 0 \\ 0 & 0 & -1 \\ 1 & 0 & 0 \end{bmatrix}, \qquad \mathbf{T}_{\text{acGT}} = \begin{bmatrix} 0.82 & 1.97 & 1.84 \end{bmatrix} \text{ m}.$$

This alignment assumption (a physical limitation) is an important issue since a small misalignment causes changes in the rotation matrix, introducing high sensitivity to the system (as the orientations used for all the position estimates are multiplied by $\mathbf{R}_{\text{ac}}$).
To evaluate the error of the first self-calibration $\mathbf{R}_{\text{ac}}$ estimate, the Frobenius norm of the difference between the experimentally obtained and the ground truth rotation matrices was computed:

$$\mathbf{E}_{\mathbf{R}_{\text{ac}}} = \|\mathbf{R}_{\text{ac}} - \mathbf{R}_{\text{acGT}}\|_F = 0.1655. \tag{4.2}$$

Moreover, the error of the resultant $\mathbf{T}_{\text{ac}}$ (computed as $\|\mathbf{T}_{\text{ac}} - \mathbf{T}_{\text{acGT}}\|$) is 0.1314 m.
Although such results can be further improved, since they derive from noisy range measurements from the Cricket system, they demonstrate that the derived self-calibration method is correct.
An improvement to this calibration results was accomplished by obtaining the Cricket position estimates ($\mathbf{P}_{\text{acoustic}}$) from directly measuring the ranges between the listener to the beacons with a laser distance meter and not synchronizing the Cricket constellation (as the aligning rotation performed can introduce additional error, due to possible errors in the reference position measurements). The results obtained with this procedure were

$$\mathbf{R}_{\text{ac}_{\text{laser}}} = \begin{bmatrix} -0.0788 & -0.9955 & 0.0517 \\ 0.0292 & -0.0541 & -0.9981 \\ 0.9965 & -0.0772 & 0.0333 \end{bmatrix}, \qquad \mathbf{T}_{\text{ac}_{\text{laser}}} = \begin{bmatrix} 0.7405 & 1.9884 & 1.8310 \end{bmatrix} \text{ m},$$

reflecting an improvement of the estimation of $\mathbf{R}_{\text{ac}}$ of 15%, as $\mathbf{E}_{\mathbf{R}\text{ac}_{\text{laser}}}$ is now 0.1406. Likewise, the error norm of the translation vector estimation using the laser is smaller by 37.5%, when compared with the first estimate, as the error norm obtained for this case is 0.0821 m.
This procedure was performed with the intention of not jeopardizing the ensuing hybrid localization performance due to a less accurate calibration. It is emphasised that this does not, in any way, undermine the calibration process as previously proposed, as a self-calibration can always be performed using the Cricket system similarly to what happens in the hybrid localization procedure; however, it is natural that the achieved accuracy is somewhat lower.

Figure 4.25 shows the final target position estimations obtained by FLORIS algorithm during a walk through the installed experimental scenario. Such positions are estimated based both on the ranges measured to the acoustic anchors and the orientations measured to the visual tags mounted on the set-up. As can be observed, the experimental scenario installed covers a volume of approximately 50 m^3, showing an appropriate considerable size/extension, given the envisaged end applications.

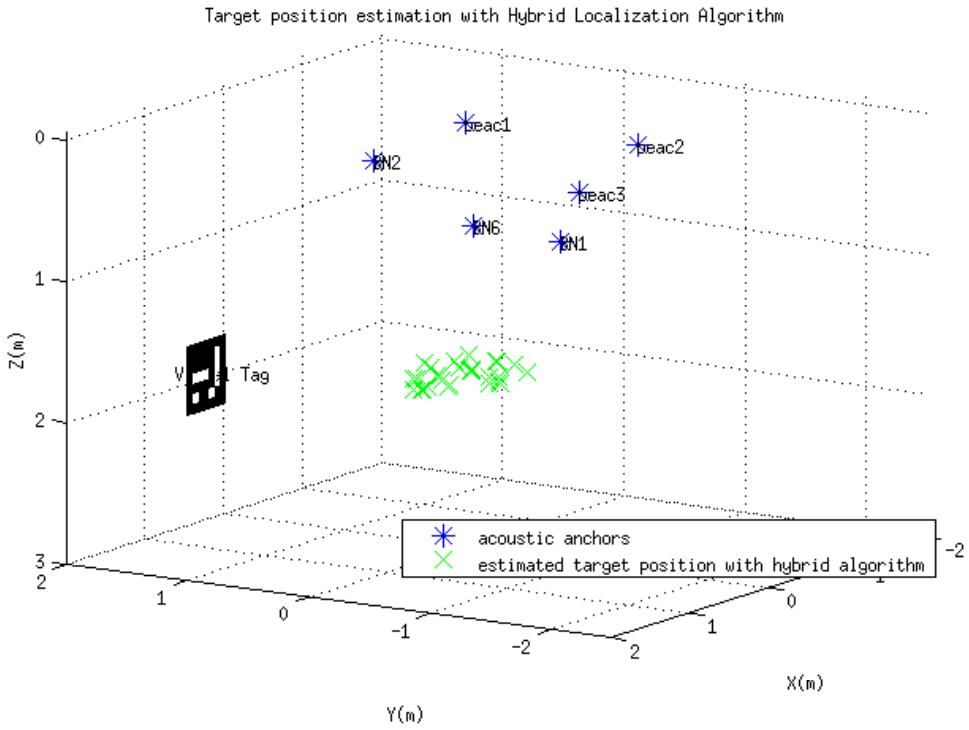

Figure 4.25: Target localization using the proposed FLORIS algorithm, during a walk through the experimental scenario

4.1.2.4 FLORIS Performance analysis

In order to support the hybrid localization approach, developed in this thesis, relative to the traditional single sensed variable approaches, comparisons between the new fusion algorithm (FLORIS) and existing state-of-the-art algorithms, which are based solely on range information are performed. In this way, it is possible to thoroughly evaluate the performance of FLORIS. Such performance analysis is presented below.

It should be stressed that the proposed hybrid localization algorithm is less stringent than other single sensed variable methods, as in this case a combination of anchors is used to perform localization. More specifically, the hybrid approach is able to successfully localize a node in certain situations where other algorithms are not. For example, in the 3D space, when information from only 3 acoustic anchors is received, it is not possible to estimate a correct position. Yet, if an object is recognized the new algorithm produces a valid position estimation. Furthermore, and contrary to previous methods, a valid position is estimated even if the target is placed outside the convex hull formed by the acoustic anchors, provided that one tag, minimum, is located outside that hull. The two previous scenarios were not only duly tested in simulation (in Subsection 4.1.1) but also experimentally verified. Hence, the method proposed in this thesis conveys further flexibility and scalability to localization systems.

Several datasets (including range and orientation measurements) were incrementally acquired in the experimental scenario shown in Figure 4.20 and used to both test the fusion approach proposed in this dissertation and perform comparisons with previous localization algorithms from the literature. Table 4.5 compares the RMSE obtained for the newly proposed fusion algorithm with SR-LS from [5] and SLNN from [34], for a set of experimental data acquired in the deployed set-up. Additionally, with the purpose

of maximizing the position estimation accuracy of these methods, a further post-processing refinement step was applied to the solutions found by the three methods. Such step was implemented using a generic minimization function (`fminseacrch` from MATLAB), which attempts to find the local minimizer of the ML functions (expressions (2.3) for the SR-LS and (2.4) for the SLNN), given the estimated position as a starting point. The RMSE results for the latter methods with the application of this minimization step are referred to as the method name with the suffix min.
As aforementioned, the experimental set-up is formed by 6 Cricket beacons, a board of ARUCO tags and the target, which comprises a Cricket listener node attached to a camera. Range (from the detected beacons) and angle measurements were collected for several positions (27) within the scenario, also the ground truth coordinates for each position were measured. The obtained bearings suffered a geometric transformation (using $\mathbf{R}_{\text{ac}}$ and $\mathbf{T}_{\text{ac}}$) in order to be described in the beacon coordinate system. This was performed following the self-calibration procedure previously introduced and described in Subsection 4.1.2.3.

	SR-LS [5]	SR-LS-min	SLNN [34]	SLNN-min	FLORIS	FLORIS-min
RMSE (m)	0.22	0.22	0.23	0.22	0.23	0.20

Table 4.5: 3D source localization performance comparison, for a set of experimental data

For this case in Table 4.5, it can be observed that FLORIS provide better accuracy estimations, when compared with the existing methods, confirming the good performance already found in simulation results. Particularly, such results reveal the quality of the proposed convex relaxation, which can provide a good starting point so that the desired minimum is reached with the post-refinement step (as in the value obtained for FLORIS-min).
This is a very fine and encouraging result, showing that although several practical limitations were found in the deployment of the experimental set-up (noisy measurements given by the Cricket System and the calibration issue) they were successfully overcome.
Globally the variance of the 3D positions computed by FLORIS for this dataset is not very different from the other methods, as $\sigma^2_{FLORIS} = 0.2490$ m, $\sigma^2_{SLNN} = 0.2106$ m and $\sigma^2_{SR-LS} = 0.2340$ m. Hence the shape of the position estimate distribution is similar for all methods. The RMSE achieved by the fusion approach reflects an improvement in localization accuracy when compared to the other two state-of-the-art algorithms, thus not only validating this newly proposed method but also pointing to a leading performance of FLORIS. These findings support that the hybrid method proposed in this thesis is a significant contribution to the area, outperforming the previous works in the case of practical noisy data. Moreover, FLORIS has proved to be a solid localization method, fully working in a real scenario and providing very good accurate estimates, including for special cases not covered by the state-of-the-art algorithms.
Applying the additional step of Cricket calibration, performed accordingly to what is described in Subsection 4.1.2.2, reduces the obtained RMSE for all methods. Furthermore, in order to maximize the possible obtained accuracy, this Cricket calibration was combined with the extra refinement step. The resulting values are presented in Table 4.6 and legitimize this calibration procedure.

	SR-LS calib	SR-LS calib-min	SLNN calib	SLNN calib-min	FLORIS calib	FLORIS calib-min
RMSE (m)	0.17	0.17	0.20	0.17	0.20	0.16

Table 4.6: 3D source localization performance comparison, for a set of experimental data with Cricket calibration

Figure 4.26 shows the histogram of the absolute error for the FLORIS estimates, from which the low error of the majority of the position estimates can be perceived.

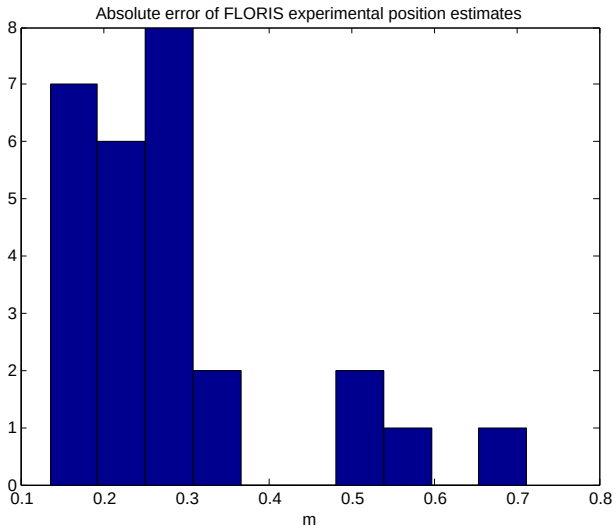

Figure 4.26: Histogram of the absolute error of target positions estimated by FLORIS during a walk through the experimental set-up

Figure 4.27 depicts the target positions estimated by FLORIS algorithm versus the ground truth positions. It can be observed that the estimated positions are, globally, very close to the ground truth, indicating a reasonable and promising accuracy.

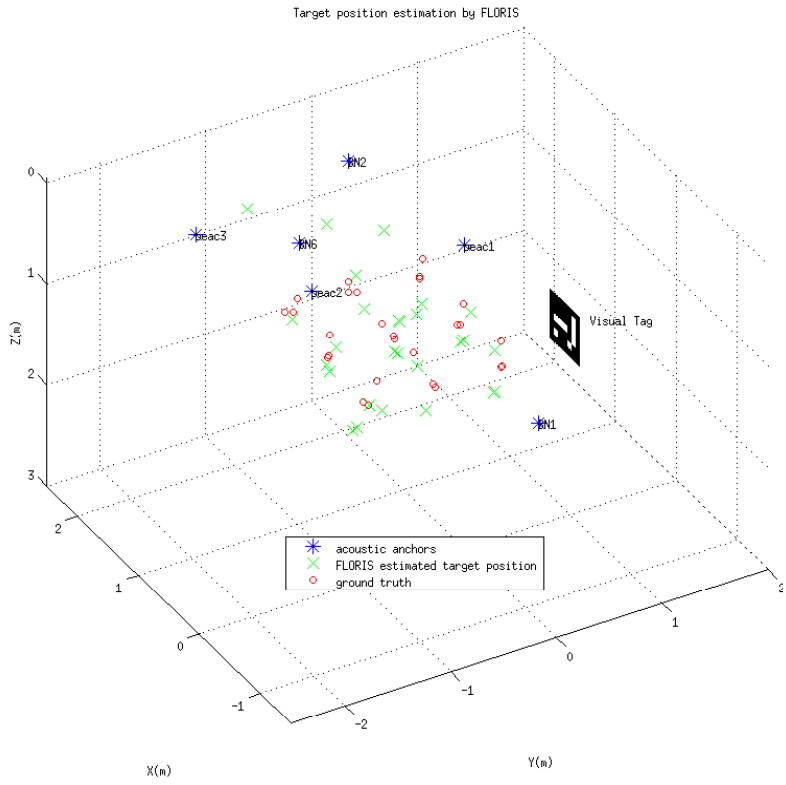

Figure 4.27: Target position estimations given by FLORIS algorithm, during a walk through the experimental scenario, versus the ground truth positions

Examining one of the cases in which the position estimate accuracy is higher than for the other methods, the improvement in localization due to the introduction of another sensed variable (specifically, the orientation information) is clearly illustrated in Figure 4.28.

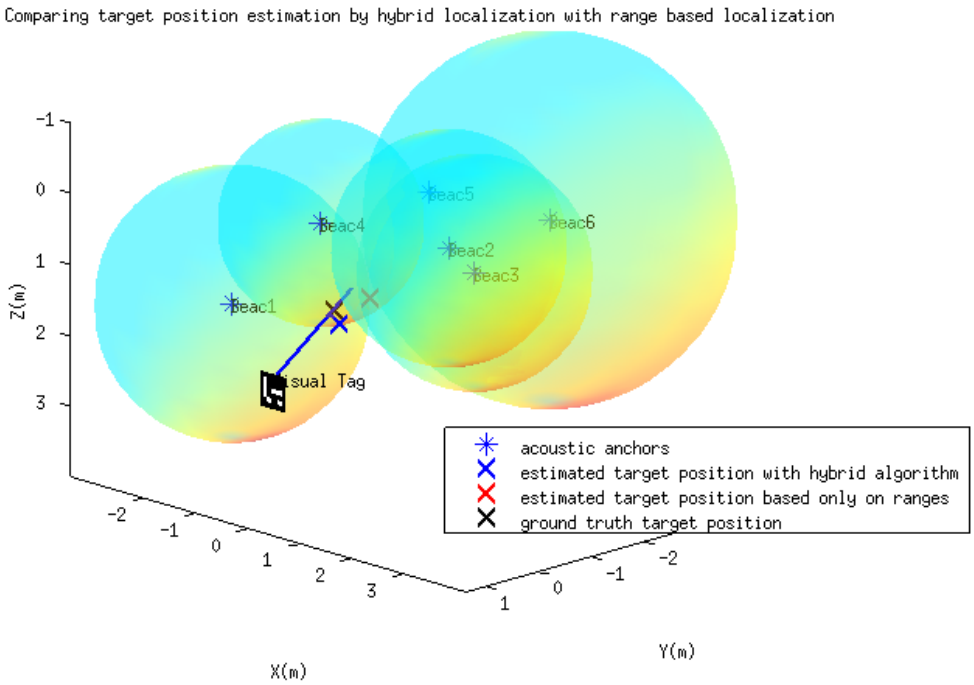

Figure 4.28: Comparing the position estimation given by FLORIS and by the SLNN with the ground truth position

Figure (4.28) depicts the two networks of the experimental set-up, spheres with radius equal to the ranges measured between each anchor and the target. Due to the highly noisy measurements acquired by the Cricket system, the five spheres (associated with the distances measured to the five anchors detected at that instant) do not intercept in a single point. Therefore, the position estimation based solely on ranges (using the SLNN algorithm and depicted with a red cross) has lower accuracy than FLORIS position estimation (depicted with a blue cross). What this picture shows is that the orientation measured between the target and the visual tag (depicted by the blue line) is quite reliable, since it intercepts the ground truth position, hence causing FLORIS position to typically have a higher accuracy than the SLNN estimate, in such noisy situations.

4.2 Sensor Network Localization - CLORIS

4.2.1 Simulation

4.2.1.1 Preliminaries

Similarly to the single source case, a set of simulation tests was conducted to evaluate the proposed sensor network fusion method CLORIS. In this case, performance will be evaluated based on the accuracy, using the RMSE (which for the sensor network localization problem is defined as in (4.3)) and the running time.

$$RMSE = \sqrt{\frac{1}{MC}\frac{1}{N}\sum_{k=1}^{MC}\sum_{i=1}^{N}\|x_i - \hat{x_i}^k\|^2} \tag{4.3}$$

In (4.3) N is the total number of nodes with unknown positions, i.e., the sensor nodes, x_i is the true position of node i and $\hat{x_i}^k$ is its position estimate in the k-th Monte Carlo run.
The following experiments were run using MATLAB R2013a and a machine powered by an Intel Xeon E5-2620 2.0GHz CPU and 64GB of RAM. The measurement noise (both for ranges and orientations) was modeled in the same way as in the single source case.
In the sensor network localization paradigm there may be networks which can not be solved, the so-called non-localizable networks, in which a unique set of sensor positions consistent with the data can not be determined [2]. Such issue can be identified in 2D applying a result from graph theory, in [2], whereby the generic global rigidity of the graph is necessary and sufficient for the network to be localizable. However, as an extension for the planar result to 3D does not exist, the proposed fusion algorithm was tested on fixed localizable networks in a [0, 1] × [0, 1] × [0, 1] cube, and not on randomly generated ones. Hence, sets of $MC = 500$ Monte Carlo trials were run for the fixed network for each measurement noise factor $(\eta) \in (0.001, 0.005, 0.01, 0.05, 0.1)$.

4.2.1.2 Results

In this Subsection the simulation results obtained for CLORIS, introduced in Section 3.2, and its comparison with the range-based method of [43], are presented. The RMSE versus the noise factor for a 3D localizable network comprising 13 acoustic anchors and 4 sensors is shown in Figure 4.29. For CLORIS case, 5 anchors were converted into visual ones, so that a more equitable comparison between the two methods can be performed.

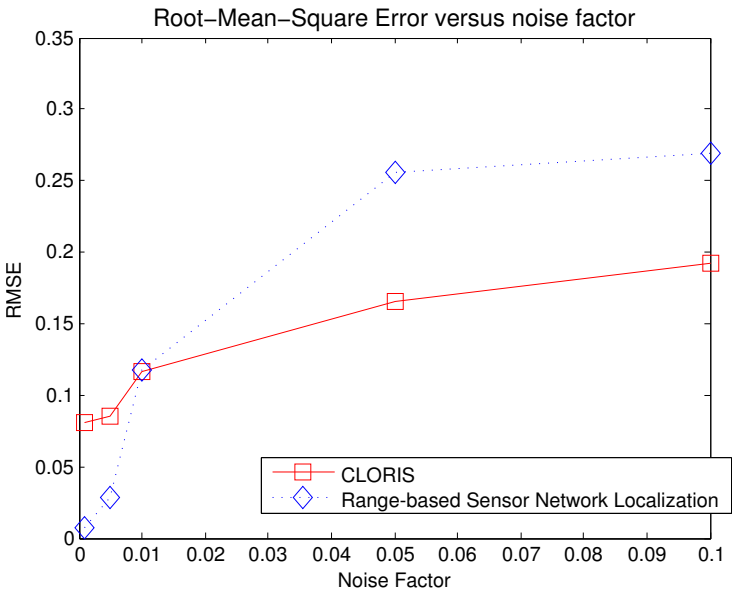

Figure 4.29: 3D Sensor Localization performance; comparison of the RMSE between CLORIS and the range-based disk relaxation method, for a fixed network of 13 acoustic or 8 acoustic + 5 visual anchors, for several Gaussian noise factors

It is interesting to note that the robustness for high noise factors demonstrated by the single-source FLORIS method in Subsection 4.1.1.2 seems to prevail in this paradigm as well with the fusion disk relaxation formulation. In fact, up to a noise factor of 0.01 the accuracy of CLORIS algorithm is surmounted by the range-based method, yet for higher noises the fusion approach achieves better results. Thus, by incorporating bearing information the network position estimations can be improved in noisy scenarios. Figure 4.30 depicts the average running times of the 500 Monte Carlo runs for each noise factor for the two methods, for the previous network.

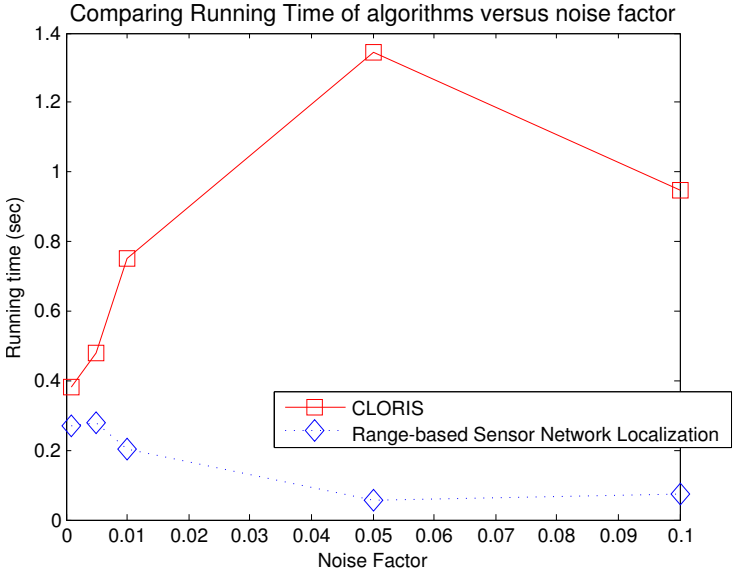

Figure 4.30: Running times comparison between CLORIS and the range-based disk relaxation method, for a fixed network of 13 acoustic or 8 acoustic + 5 visual anchors, for several Gaussian noise factors

Figure 4.30 shows that CLORIS algorithm has a larger cost in terms of time when compared with the original one. From the analysis of the simulations performed this reduction in the speed seems to be

caused by the larger number of iterations of the algorithm when including also bearing than when only considering range measurements. This gap in the number of iterations was sometimes found to be of nearly one order of magnitude. A more detailed analysis of this observed slow-down in convergence rate is deferred to future work.
Moreover, a study of the proposed algorithm performance depending on the number of anchors is presented through Figures 4.31 and 4.32.

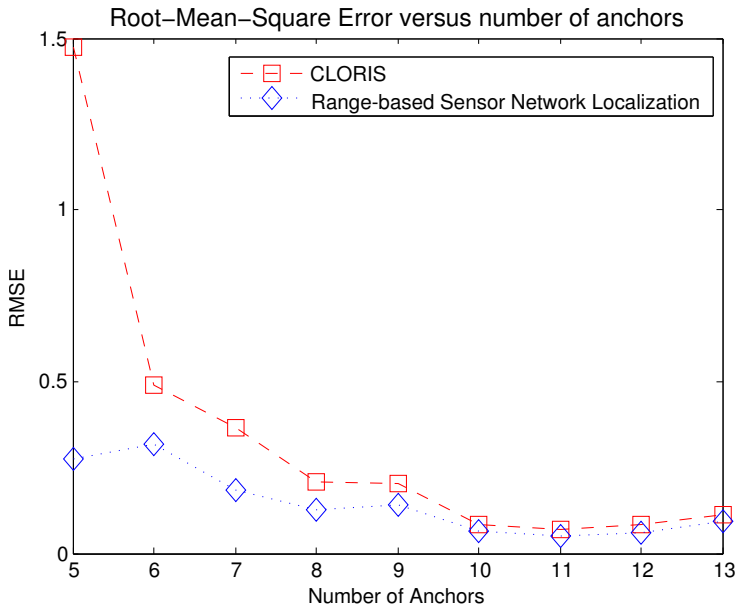

Figure 4.31: 3D Sensor Localization performance; comparison of the RMSE between CLORIS and the range-based disk relaxation method, for a fixed network with varying number of anchors (for the fusion case 2 of the acoustic anchors are converted into visual)

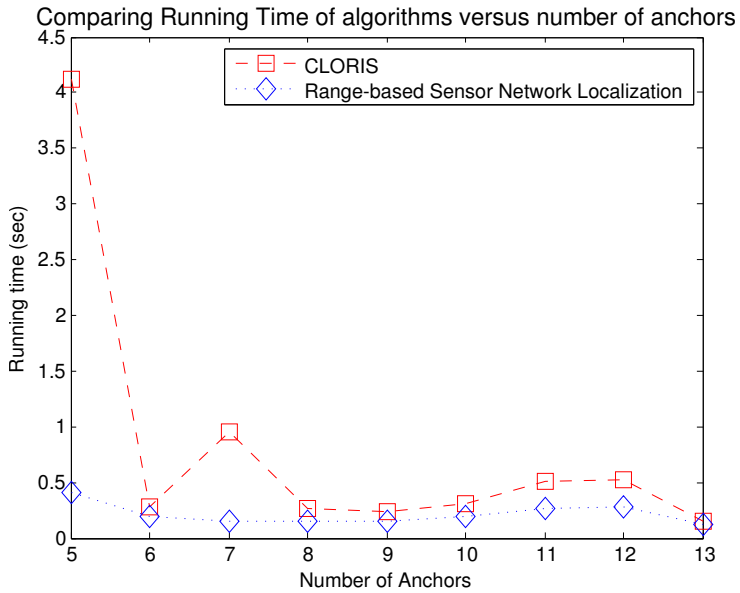

Figure 4.32: Running times comparison between the fusion and the range-based disk relaxation method, for a fixed network with varying number of anchors (for the fusion case 2 of the acoustic anchors are converted into visual)

As expected, increasing the number of anchors makes the localization algorithms estimations, gradually, more accurate. The high value found for the RMSE of CLORIS for the 5 anchors network case leads

to the conclusion that such configuration is non localizable with the visual anchors. This fact seems to be supported by the average running time depicted in Figure 4.32, which is a peak indicating an odd behaviour for this specific case.

The previous simulations were performed for a constant noise factor of 0.01. When performing similar tests (for the same network configurations) but for a higher noise factor (0.1), both CLORIS RMSE and running times are lower (this contradicts the trend in Figure 4.30), demonstrating, again, the feasibility and robustness of CLORIS towards high noise situations.
The same simulation tests were executed for another localizable 3D network, comprising a larger number of anchors (18 acoustic anchors, from which 6 were converted into visual for testing the fusion approach) and fewer sensors (3). The global relative behaviour of the RMSE and running times are congruent with the results found for the previous network configuration. However, and as expected due to the increase of the total number of anchors, the accuracy is moderately improved (the fusion approach suffers less impact from high noise factors) while the speed deteriorates. The number of iterations CLORIS algorithm was, once more, considerably higher, noting a peak again at noise factor 0.5. Such result definitely calls for a more profound study regarding this escalation of the number of iterations when angles are included. Tables 4.7a and 4.7b display the quantitative results.

Noise Factor	Range-based	CLORIS
0.001	0.0082	0.0819
0.005	0.0695	0.1213
0.01	0.1117	0.1514
0.05	0.1976	0.1770
0.1	0.1330	0.1328
0.2	0.1800	0.1783

(a) RMSE (in distance units) versus noise factor

Noise Factor	Range-based	CLORIS
0.001	0.6269	1.3000
0.005	0.3289	2.4749
0.01	0.2100	3.6567
0.05	0.1815	4.6106
0.1	0.3942	0.8778
0.2	0.1992	0.3027

(b) Execution time (s) versus noise factor

Table 4.7: Performance comparison for a second 3D network configuration (with 2 sensors and 18 acoustic anchors, or 12 acoustic and 6 visual for the fusion case) between the CLORIS and the range-based disk relaxation approach for sensor network localization

4.2.2 Experimental Results - Fusion Sensor Network Localization

After confirming the experimental operability of FLORIS for the localization of a single target, the cooperative approach CLORIS, based on the disk relaxation, was also tested in the previously installed experimental set-up.
For that purpose, two targets, consisting, similarly to the single source scenario, on a listener cricket node mounted on a camera were deployed in the area covered by the acoustic and visual anchors. In this set-up one of the nodes whose position is unknown (called sensor 1) measures distances to a subset of available acoustic beacons (depending on the position and directionality) and the pose to the visual anchor; a second unknown node (sensor 2) measures distances to the subset of beacons available to it and also the poses to sensor 1 (which has tags enabling such measurement) and to the visual anchor, if the latter is in the field of view of the camera. Figure 4.33 illustrates such scenario, showing the image seen by sensor 2 (including sensor 1 and the tags).

Figure 4.33: View from sensor 2 camera, showing sensor 1 and the visual tag

The following Subsection 4.2.2.1 features the performance analysis of the implementation of the fusion network localization algorithm introduced in Subsection 3.2.1, comparing it with the same disk relaxation based solely on range information.

4.2.2.1 CLORIS Performance analysis

A dataset comprising, as described above, range measurements from the Cricket system and angles among the acoustic and visual anchors and the sensors in the set-up in Figure 4.20 was collected. Preliminary results showed that using both the experimental range and bearing measurements to estimate the network positions led to reasonably high error when compared to the ground truth positions. Hence, to find the source of error, datasets reproducing the experimental set-up and the instances of the networks (different sensor positions) were synthetically generated. With these exclusively synthetic datasets it was possible to reject the possibility of non-localizable networks among the instances, since the sensor position estimates obtained by CLORIS matched the ground truths. The next step was to replace the synthetic distances (with no noise) by the distances measured by the Cricket system between the sensors and the anchors as well as between the two sensors, applying the calibration accomplished for each Cricket beacon. In this case, CLORIS was still able to provide quite accurate sensor position estimates, as can be seen in Table 4.8 and Figure 4.34. This "debug" process allowed to conclude that the visual component, i.e., the computation of the angles, is a source of error. Indeed, this could be expected, since the incident directions have to be expressed in the global coordinate system and the computations to perform such operation imply several transformations ("hops") between coordinate systems, which are error prone (for instance, a small deviation in a Rotation matrix between two systems may have a significant impact on the resultant direction described in the global reference).
The incident directions on sensor 1 are computed exactly in the same way as in the single-source paradigm, resorting to the self-calibration procedure discussed in Subsection 4.1.2.3, applying equation (4.4) to translate sensor 1 to the defined global reference.

$$\mathbf{P}_{\text{cam1\_c}} = \mathbf{R}_{\text{ac}} \left(\mathbf{T}_{\text{cam1\_a}} + \mathbf{T}_{\text{cc}} \right) + \mathbf{T}_{\text{ac}} \tag{4.4}$$

The notation $\mathbf{P}_{\text{cam1\_c}}$ designates the position of camera 1 in the Cricket (global) coordinate system and $\mathbf{T}_{\text{cam1\_a}}$ the translation vector of camera 1 referred in the ARUCO coordinate system.
For sensor 2 the procedure is more complex. The incident directions in sensor 2 have to be intermediately translated into the sensor 1 tags and subsequently to sensor 1 camera coordinate systems, so that in the end they can be translated to the global coordinate system. Only when they are expressed in the global coordinate system may they be inserted in the fusion algorithm. The following transformation is performed:

$$\mathbf{P}_{\text{cam2\_c}} = \mathbf{R}_{\text{ac}} \left(\mathbf{R}_{\text{cam1\_a}} \left(\mathbf{R}_{\text{as1\_cam1}} \left(\mathbf{T}_{\text{cam2\_as1}} + \mathbf{T}_{\text{cc}} \right) + \mathbf{T}_{\text{as1\_cam1}} \right) + \mathbf{T}_{\text{cam1\_a}} \right) + \mathbf{T}_{\text{ac}} \tag{4.5}$$

Similarly to (4.4), $\mathbf{P}_{\text{cam2\_c}}$ denotes camera 2 position in the Cricket referential, $\mathbf{R}_{\text{as1\_cam1}}, \mathbf{T}_{\text{as1\_cam1}}$ are the rotation matrix and translation vector between camera 1 and the sensor 1 tags and $\mathbf{T}_{\text{cam2\_as1}}$ is the translation between camera 2 and sensor 1 tags coordinate system.
According to the previous explanation, and to circumvent the constraints imposed by the practical set-up, the results addressing the mixed case are also presented. In such case the range measurements are the ones obtained experimentally using the Cricket system, yet the incident directions are synthetically generated from the visual anchor and sensor real positions. It is believed that such assumption, in no way, diminishes the strength of this novel fusion method for sensor network localization.
The experimental results attained when two targets roamed in the set-up area are summarized in Table 4.8. These results are particularly interesting as the networks configurations obtained from the experimental dataset happen to be non-localizable when considering only acoustic anchors and range information, whereas when the bearing is added the sensor positions are estimated with good accuracy. Hence, this experimental test demonstrates the practical operability and feasibility of the new CLORIS method. It is also verified that the integration of the experimentally measured angles increases the error, nevertheless it is a quite acceptable degradation considering the practical deployment and the previous discussion.
Furthermore, from Figure 4.34 it can be observed that the sensor 1 position estimates tend to be much more accurate than sensor 2 estimates, in fact the RMSE considering only sensor 1 is 0.0204 m while the RMSE for the sensor 2 position estimates is 0.3895 m. Further analysis shall be conducted in order to understand and explain such fact.

	RMSE (m)
Range-based Disk Relaxation	1.43
CLORIS - Synthetic Directions	0.50
CLORIS - Totally Experimental	0.55

Table 4.8: 3D sensor network localization performance for a walk through the experimental scenario. Comparison among the solely range based and CLORIS algorithm with both experimental range measurements and synthetic directions and all experimental measurements

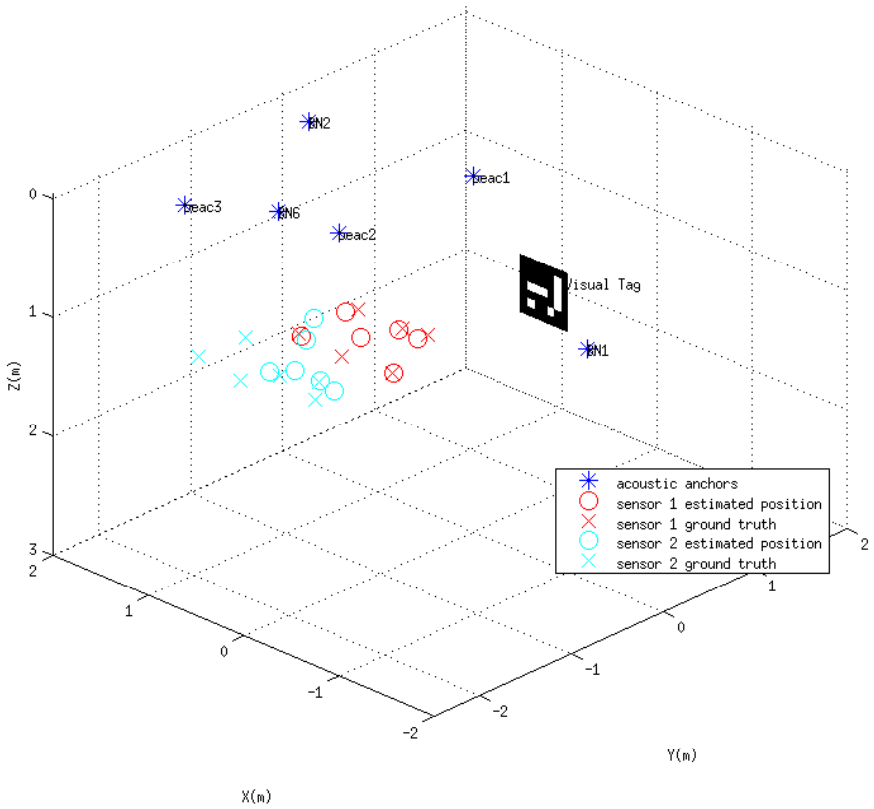

Figure 4.34: Estimated sensors positions by CLORIS (with experimental ranges and synthetic angles) versus the ground truth positions during a walk through the experimental set-up

The results achieved presented in both Figure 4.34 and Table 4.8 show a practical case in which the cooperative fusion approach developed in this dissertation provides good position estimates surpassing the previous method based on a single sensed variable. Such result is extremely encouraging regarding a further development of CLORIS method for cooperative localization.

Chapter 5

Conclusions

This Chapter presents the major contributions and attained results of this dissertation (in Section 5.1) and discusses, in further detail, some foreseen directions for future research on this topic (in Section 5.2).

5.1 Achievements

The work presented in this dissertation successfully addressed the growing need for solutions and applications for large scale heterogeneous sensor networks, taking advantage of the distinct sensing devices. More precisely, this work provides accurate indoor localization methods that can use different types of proximity information, namely ranges and incident directions estimated from video images, in a totally integrated fashion.

A new formulation based on a nonconvex LS joint formulation fusing information from ranges and angles and a tight convex relaxation for this problem were devised. The single-source method was called FLORIS, owing to the geometry of the formulation.

Additionally, a procedure to derive the observation model and its parameters - a key component for information fusion that expresses spatial data acquired from the two separate sensor networks in a common reference frame - is proposed. This calibration procedure was analytically and experimentally tested enabling integration of the localization system in the deployed experimental set-up.

In the same vein, this dissertation also targeted the collaborative scenario (i.e., the sensor network localization problem) for heterogeneous measurements. The proposed CLORIS algorithm is an extension of a previous sensor network localization method developed in-house and based on a convexification mechanism by disk relaxation.

The results obtained show that the both methods can provide accurate localization. More quantitatively, FLORIS method achieves, in numerical results, lower RMSE (thus higher accuracy) than state-of-the-art range-only methods, specially for high noise scenarios, while the running times are comparable with the former preferred method (SLNN). In this way it can be claimed that the newly proposed FLORIS has a similar performance or even outperforms, in some cases, the previous existing localization methods.

Simulation results for the collaborative scenario were extremely promising too. Again, the algorithm's most salient feature is high accuracy in the presence of strong observation noise, thus presenting a good performance trade-off when compared to the range-only methods.

Moreover, an experimental set-up was deployed leading to a validation of the feasibility and robustness of

the algorithms in real application scenarios for the two localization paradigms. Although several physical setbacks were found, most of them were surpassed, allowing for an entirely integrated implementation of FLORIS and CLORIS. Both methods showed very good behaviour in the real scenario, with better accuracy than benchmark algorithms in the conducted tests. The experimental findings clearly motivate further implementations, using higher performance technologies and devices.

5.2 Future Work

As multiple future work directions were identified during this dissertation work, this Section is divided into topics, detailing, for each, the further improvements foreseen.

5.2.1 Range

The variables sensed to obtain distance information for the proposed hybrid localization method are an important practical concern, thus alternative technologies to acoustic based sensors, such as the Cricket system which was shown to provide noisy range measurements, shall be addressed.
Received Signal Strength Indication (RSSI) is one technique that measures pseudo-ranges (since the range is computed using a function that relates the power magnitude with distance) and that can be advantageous. In fact, although the resulting range measurements can be quite noisy, they can be obtained using low complexity hardware and in an inexpensive way. Furthermore, this technology can provide large datasets, which can be most convenient to an EDM approach (discussed in 5.2.4), in which the localization problem is formulated as an EDM completion.
Another technology that should be taken into account is Wi-Fi, mainly due to its ubiquitousness in the envisaged practical application scenarios. As the majority of smart devices have Wi-Fi receptors, an hybrid system relying on such technology to obtain the distance measurements would be easy to implement with widespread acceptance.
Lastly, in challenging environments such as indoors, owing to interferences, occlusions and multi-path, [45] suggests ultra-wideband (UWB) transmission technology, albeit being currently expensive, as a promising alternative for localization. Thus, exploring (low-cost) UWB nodes is another future prospect.

5.2.2 Vision

The method developed in this work can be seen, regarding the vision component, as a first step/phase to the exploration of objects/patterns present in the environment. Effectively, an evolution from the use of (augmented reality) markers/tags as visual anchors to general known objects spread throughout the surroundings is a natural evolution for this work. Thus, object/pattern detection and recognition algorithms should be applied in order to obtain the bearing from visual anchors. This final phase would resort to the Bag of Words image representation and use methods such as SIFT (Scale-Invariant Feature Transform) [27], following the work developed by Lampert et al. [25], which allows the use of classifiers for object localization that were previously considered too slow for this task, since the convergence to the global solution is in sub-linear time.

5.2.3 Cooperative paradigm

In regard to the sensor network localization problem, although the attained results both in simulation and experimentally were positive, there is still room for improvement. Namely, one of the identified issues concerns the large number of iterations executed by the disk relaxation algorithm when angles are considered, which degrades the retrieved running times. Thus, a deeper analysis of the computations associated with the incident directions and their projections onto lines, performed during Nesterov's gradient descent method, is pertinent.

Furthermore, a potential formulation with fine tuned weights for the range and angular components shall be assessed. Effectively, exploratory tests suggest that varying the weights assigned to the heterogeneous measurements, rather than considering an equal influence in the cost function, may have a positive impact in the obtained results.

Another aspect that was proven to have impact in the experimental localization performance is the procedure to translate between the coordinate systems for the various types of nodes. Certainly this has also impact in the single-source localization paradigm, yet it has more negative consequences in the cooperative case since more poses from the multiple sensors must be expressed, via the known transformations, in the global referential system. A robust method to execute such transformation to a common coordinate system shall be devised, as such procedure has high sensitivity to small fluctuations of the entries of rotation matrices. Additionally, a flawless calibration of the cameras is mandatory for any devised procedure. In fact, the general underperformance found for the experimental results should be partly consequence of physical limitations and constraints of the available hardware.

5.2.4 Exploring other constraints

Parallel to the line of work followed in this dissertation, based on ML/LS approaches, an innovative approach exploring the low rank properties of EDM is foreseen. With the ranges available, the distance estimation should be formulated as an EDM completion problem, as found in the literature [32, 31, 23]. Furthermore, the completion problem will use the rank constraint, as the matrix rank is given by the subspace where localization is being performed (most commonly 3D). A matrix ($\mathbf{D}$) comprising all the EDM (per column, using the vectorization operation) related to each instant of the target trajectory will be built and a decomposition in a matrix with the correct "clean" distances ($\mathbf{D_r}$) added with the error ($\mathbf{E}$) is pursued. A formulation consisting in a cost function with both matrices, penalizing the error matrix and a low rank constraint imposed to the "clean" matrix will be sought. To obtain such decomposition from a matrix built from batches of measurements, dealing with the outliers, a technique called Sparse Subspace Clustering (SSC) [17] could be applied. This method [17] proposes a convex relaxation to represent data points from a high-dimensional space as a sparse linear affine combination of the remaining points from a low-dimensional subspace.

Also, a more theoretical approach may be explored regarding the use of the rank constraint in the EDM completion problem. In this topic the rank evolution in several EDMs formed by the moving target in a wireless network for each time instant will be studied. With this secondary approach we will try to devise a new formulation for the problem of localizing and tracking a single moving source (not addressed in this Thesis). In this scenario a dependence of the position estimation on prior estimates combined with the target trajectory information may be established.

Under this EDM completion approach the case in which not all distances between the sensors and anchors are available, i.e., the missing data case, is inherently addressed.

Moreover, it should be stressed that promising cues regarding a unified fusion using EDMs have been

found. In fact, low rank constraints are used not only in WSN localization (via EDMs), as aforementioned, but also, more recently, in several traditional vision methods/problems [10]. Such fact paves the way to further research in this topic, strongly suggesting that a common methodology, integrating both ranges and orientations can be devised. This approach could be applied together with 3D reconstruction maps as in [1].

References

[1] S. Agarwal, Y. Furukawa, N. Snavely, I. Simon, B. Curless, S. M. Seitz, and R. Szeliski. Building Rome in a Day. *Communications of the ACM*, 54(10):105–112, Oct 2011.

[2] B. Anderson, I. Shames, G. Mao, and B. Fidan. Formal Theory of Noisy Sensor Network Localization. *SIAM Journal on Discrete Mathematics*, 24(2):684–698, Feb. 2010.

[3] J. Bachrach and C. Taylor. *Localization in Sensor Networks, in Handbook of Sensor Networks: Algorithms and Architectures.* John Wiley & Sons, 2005.

[4] J. P. Ballantine and A. R. Jerbert. Distance from a line, or plane, to a point. *The American Mathematical Monthly*, 59(4):242–243, Apr. 1952.

[5] A. Beck and P. Stoica. Exact and Approximate Solutions of Source Localization Problems. *IEEE Transactions on Signal Processing*, 56(5):1770–1778, May 2008.

[6] P. Biswas. *Semidefinite Programming Approaches to Distance Geometry Problems.* PhD thesis, Stanford University, 2007.

[7] P. Biswas, H. Aghajan, and Y. Ye. Integration of Angle of Arrival Information for Multimodal Sensor Network Localization using Semidefinite Programming. In *Proceedings of 39th Asilomar Conference on Signals, Systems and Computers*, 2005.

[8] P. Biswas and Y. Ye. Semidefinite Programming for Ad Hoc Wireless Sensor Network Localization. In *Proceedings of the third international symposium on Information processing in sensor networks*, pages 45 – 54. ACM, 2004.

[9] S. Boyd and L. Vandenberghe. *Convex Optimization.* Cambridge University Press, 2004.

[10] R. S. Cabral, D. Fernando, J. P. Costeira, and A. Bernardino. Unifying Nuclear Norm and Bilinear Factorization Approaches for Low-rank Matrix Decomposition. In *IEEE International Conference on Computer Vision (ICCV)*, pages 2488 – 2495, 2013.

[11] R. S. Cabral, D. Fernando, J. P. Costeira, and A. Bernardino. Matrix Completion for Weakly-supervised Multi-label Image Classification. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, in press, online July 2014.

[12] G. Calafiore, C. L., and M. Wei. Distributed optimization techniques for range localization in networked systems. In *Proceedings of the 49th IEEE Conference on Decision and Control (CDC)*, pages 2221–2226, 2010.

[13] K. Cheung, W. Ma, and H. So. Accurate approximation algorithm for TOA based maximum likelihood mobile location using semidefinite programming. In *Proceedings of IEEE Int. Conf. Acoust., Speech, Signal Process (ICASSP'04)*, pages 145–148, Montreal, Canada, 2004. IEEE.

[14] M. Crocco, A. D. Bue, I. B. Barbosa, and V. Murino. A Closed Form Solution for the Self-Calibration of Heterogeneous Sensors. In *Procedings of the British Machine Vision Conference 2012*, Los Angeles, California, 2012.

[15] M. Crocco, A. D. Bue, and V. Murino. A Bilinear Approach to the Position Self-Calibration of Multiple Sensors. *IEEE Transactions on Signal Processing*, 60(2):660–673, Feb. 2012.

[16] G. Destino and G. Abreu. On the maximum likelihood approach for source and network localization. *IEEE Transactions on Signal Processing*, 59(10):4954–4970, Oct. 2011.

[17] E. Elhamifar and R. Vidal. Sparse Subspace Clustering. In *IEEE Conference on Computer Vision and Pattern Recognition, 2009. CVPR 2009*, pages 2790 – 2797. IEEE, 2009.

[18] F. Endres, J. Hess, N. Engelhard, J. Sturm, D. Cremers, and W. Burgard. An Evaluation of the RGB-D SLAM System. In *Procedings of the 2012 IEEE International Conference on Robotics and Automation*, pages 1691 – 1696, 2012.

[19] M. Fischler and R. Bolles. Random Sample Consensus: A Paradigm for Model Fitting with Applications to Image Analysis and Automated Cartography. *Communications of the ACM*, 24(6):381–395, June 1981.

[20] S. Funiak, C. Guestrin, M. Paskin, and R. Sukthankar. Distributed Localization of Networked Cameras. In *5th International Conference on Information Processing in Sensor Networks*, pages 34–42, 2006.

[21] S. Garrido-Jurado, R. Muñoz-Salinas, F. J. Madrid-Cuevas, and M. J. MaríÂŋn-Jiménez. Automatic generation and detection of highly reliable fiducial markers under occlusion. *Pattern Recognition*, 47(6):2280 – 2292, 2014.

[22] M. Gholami, L. Tetruashvili, E. Strom, and Y. Censor. Cooperative Wireless Sensor Network Positioning via Implicit Convex Feasibility. *Signal Processing, IEEE Transactions on*, 61(23):5830–5840, Dec 2013.

[23] J. Gomes, E. Zamanizadeh, J. Bioucas-Dias, J. Alves, and T. C. Furfaro. Building location awareness into acoustic communication links and networks through channel delay estimation Categories and Subject Descriptors. In *WUWNet'12 7^{t}h ACM International Conference on Underwater Networks & Systems*, Los Angeles, California, 2012.

[24] N. Krislock and H. Wolkowicz. *Handbook on Semidefinite, Conic and Polynomial Optimization.* Springer, 2012.

[25] C. H. Lampert, M. Blaschko, and T. Hofmann. Beyond sliding windows: Object localization by efficient subwindow search. In *IEEE Conference on Computer Vision and Pattern Recognition, 2008. CVPR 2008*, pages 1–8. IEEE, 2008.

[26] L. Liberti, C. Lavor, N. Maculan, and A. Mucherino. Euclidean distance geometry and applications. arXiv:1205.0349v1, 2012.

[27] D. Lowe. Object recognition from local scale-invariant features. In *Proceedings of the Seventh IEEE International Conference on Computer Vision, 1999 (Volume:2)*, pages 1150–1157. IEEE, 1999.

[28] M. Marques and J. P. Costeira. Estimating 3D shape from degenerate sequences with missing data. *Computer Vision and Image Understanding*, 13(2):261–272, Feb. 2011.

[29] A. Mulloni, D. Wagner, I. Barakonyi, and D. Schmalstieg. Indoor Positioning and Navigation with Camera Phones. *IEEE Pervasive Computing*, 8(2):2–31, Apr. 2009.

[30] Y. Nesterov. A method of solving a convex programming problem with convergence rate o (1/k2). *Soviet Mathematics Doklady*, 27(2):372–376, Feb. 1983.

[31] P. Oguz-Ekim, J. Gomes, and P. Oliveira. RSS Based Cooperative Sensor Network Localization with Unknown Transmit Power. In *Proceedings of Signal Processing and Communications Applications Conference (SIU), 2013 21st*, pages 1–4. IEEE, 2013.

[32] P. Oguz-Ekim, J. Gomes, J. Xavier, and P. Oliveira. Robust Localization of Nodes and Time-Recursive Tracking in Sensor Networks Using Noisy Range. *IEEE Transactions on Signal Processing*, 59(8):3930–3942, Aug. 2011.

[33] P. Oguz-Ekim, J. Gomes, J. Xavier, M. Stosic, and P. Oliveira. Approximate Maximum Likelihood Source Localization from Range Measurements Through Convex Relaxation. arXiv:1111.6755v1, 2011.

[34] P. Oguz-Ekim, J. Gomes, J. Xavier, M. Stosic, and P. Oliveira. An Angular Approach for Range-Based Approximate Maximum Likelihood Source Localization Through Convex Relaxation. *IEEE Transactions on Wireless Communications*, 13(7):3951–3964, July 2014.

[35] N. Patwari, J. Ash, S. Kyperountas, A. Hero, R. Moses, and N. Correal. Locating the nodes: cooperative localization in wireless sensor networks. *Signal Processing Magazine, IEEE*, 22(4):54–69, July 2005.

[36] N. B. Priyantha, A. Chakraborty, and H. Balakrishnan. The Cricket Location-Support System. In *The Proceedings of the sixth ACM International Conference on Mobile Computing and Networking (ACM Mobicom 2000)*, pages 32–43, 2000.

[37] A. Rahimi, B. Dunagan, and T. Darrell. Simultaneous Calibration and Tracking with a Network of Non-Overlapping Sensors. In *Proceedings of the 2004 IEEE Computer Society Conference on Computer Vision and Pattern Recognition, 2004*, pages 187–194. IEEE, 2004.

[38] P. F. R. C. Santos. Mapeamento e Localização em Forma Fechada. Master's thesis, Instituto Superior Técnico, 2013.

[39] D. Schneider. You Are Here. *IEEE Spectrum*, 50(12):34–39, Dec. 2013.

[40] Y. Shang, W. Rumi, Y. Zhang, and F. M. Localization from connectivity in sensor networks. *IEEE Transactions on Parallel and Distributed Systems*, 15(11):961–974, Nov. 2004.

[41] Q. Shi, C. He, H. Chen, and L. Jiang. Distributed wireless sensor network localization via sequential greedy optimization algorithm. *IEEE Transactions on Signal Processing*, 58(6):3328–3340, June 2010.

[42] A. Simonetto and G. Leus. Distributed maximum likelihood sensor network localization. *IEEE Transactions on Signal Processing*, 62(6):1424–1437, Mar. 2014.

[43] C. Soares, J. Xavier, and J. Gomes. Simple and fast cooperative localization for sensor networks. arXiv:1408.4728v1 [math.OC], 2014.

[44] S. Wang and H. Hu. Wireless Sensor Networks for Underwater Localization: A Survey. Technical Report CES-521, School of Computer Science and Electronic Engineering University of Essex, United Kingdom, May 2012.

[45] H. Wymeersch, J. Lien, and M. Win. Cooperative Localization in Wireless Networks. *Proceedings of the IEEE*, 97(2):427–450, Feb. 2009.