UNIVERSITÀ
DEGLI STUDI
DI PADOVA

Sede Amministrativa: Centro Interdipartimentale di ricerca di Studi ed Attività Spaziali “G.Colombo”
Sede Consorziata: Università degli Studi di Trento

Dipartimento di Ingegneria civile, ambientale e meccanica

SCUOLA DI DOTTORATO DI RICERCA IN : Scienze Tecnologie e Misure Spaziali (STMS)
INDIRIZZO: Misure Meccaniche per l’Ingegneria e lo Spazio (MMIS)
CICLO XXV°

Studio e applicazione di metodi di misura del moto mediante sistemi opto-elettronici

Study and application of motion measurement methods by means of opto-electronic systems

Direttore della Scuola : Ch.mo Prof. Giampiero Naletto
Coordinatore d’indirizzo: Ch.mo Prof. Stefano Debei
Supervisore :Ch.mo Prof. Mariolino De Cecco

Dottorando : Mattia Tavernini

To my family

Acknowledgments

There is an important number of people to thank. First and foremost my advisor, prof. Mariolino De Cecco for his advice, guidance, support during these three years and to give me the opportunity to attain this result.

I would like also to thank Luca Baglivo whose support was also very important, and my lab mates Francesco, Nicolò and Alberto, always willing to help me, not only in my studies.

I am very gratefully also to Alberto Vale, João and Nuno: they give me opportunity to work with them and moreover in a marvellous city.

A special thanks also go to all the mates of the Centrale, especially Silvio and Daniele whom remember me that there is a world outside the university.

Finally, but most importantly, I owe a special thank to my family that support and bear me in all these years and to give me the oppportunity “to be studied“, as we say in my town.

Abstract

This thesis addresses the problem of localizing a vehicle in unstructured environments through on-board instrumentation that does not require infrastructure modifications. Two widely used opto-electronic systems which allow for non-contact measurements have been chosen: camera and laser range finder. Particular attention is paid to the definition of a set of procedures for processing the environment information acquired with the instruments in order to provide both accuracy and robustness to measurement noise. An important contribute of this work is the development of a robust and reliable algorithm for associating data that has been integrated in a graph based SLAM framework also taking into account uncertainty thus leading to an optimal vehicle motion estimation. Moreover, the localization of the vehicle can be achieved in a generic environment since the developed global localization solution does not necessarily require the identification of landmarks in the environment, neither natural nor artificial. Part of the work is dedicated to a thorough comparative analysis of the state-of-the-art scan matching methods in order to choose the best one to be employed in the solution pipeline. In particular this investigation has highlighted that a dense scan matching approach can ensure good performances in many typical environments. Several experiments in different environments, also with large scales, denote the effectiveness of the global localization system developed. While the laser range data have been exploited for the global localization, a robust visual odometry has been investigated. The results suggest that the use of camera can overcome the situations in which the solution achieved by the laser scanner has a low accuracy. In particular the global localization framework can be applied also to the camera sensor, in order to perform a sensor fusion between two complementary instrumentations and so obtain a more reliable localization system. The algorithms have been tested for 2D indoor environments, nevertheless it is expected that they are well suited also for 3D and outdoors.

Sommario

La tesi affronta il problema della localizzazione di veicoli in ambienti non strutturati mediante sistemi di misura che, montati a bordo del veicolo, non richiedono modifiche dell'ambiente di navigazione. La scelta è ricaduta su due strumenti opto-elettronici largamente utilizzati, camera e Laser Range Finder (LRF), i quali consentono di effettuare misure senza contatto e quindi non intervenire sull'ambiente. Particolare attenzione è stata posta alla definizione di una serie di procedure per l'elaborazione dei dati acquisiti da questa strumentazione al fine di ottenere delle informazioni affidabili e robuste alle sorgenti di rumore ambientali. Un importante contributo di questo lavoro è lo sviluppo di un procedura di associazione robusta ed affidabile che consente di tener conto di tutti gli aspetti probabilistici in maniera tale da poter essere utilizzata in un algoritmo di localizzazione globale SLAM basato sulla teoria dei grafi e fornire una stima ottimale del moto del veicolo. Inoltre, la localizzazione del veicolo può essere eseguita in un ambiente generico dato che questo metodo di localizzazione globale non richiede l'identificazione di caratteristiche particolari nell'ambiente. Parte del lavoro è stata dedicata ad un'analisi esaustiva dei metodi di stima del moto fra scansioni laser, allo scopo di identificare il metodo con prestazioni migliori da impiegare nel metodo di localizzazione. Questo ha consentito di evidenziare come un metodo di comparazione denso permetta di ottenere buone prestazioni in diverse tipologie di ambiente. L'efficacia del metodo di localizzazione globale implementato è supportata da una serie di valutazioni sperimentali in diversi ambienti, anche di elevati dimensioni. Riguardo alla camera, è stato sviluppato un metodo robusto di visual odometry, il quale ha evidenziato come tale strumento permetta di affrontare delle situazioni nelle quali le informazioni del laser non sono sufficienti per stimare la posa del veicolo. In particolare, data la generalità del metodo di localizzazione globale, questo può essere facilmente applicato anche alla camera, al fine di ottenere la fusione di informazioni fra due strumentazioni complementari e quindi ottenere un sistema di localizzazione più affidabile. Gli algoritmi sono stati testati in un ambiente indoor bidimensionale, ma si prevede che possano essere utilizzati anche in ambienti tridimensionali e outdoor.

Contents

List of Figures

Chapter 1

Introduction

The mobile robotics has been for several years confined in industrial and few cutting-edge applications. The main reason and limitation was the working scenario: to ensure the complete governance of the robot, the environment had to be controlled and perfectly known. Only in some really particular applications, such as unmanned planetary rover this limitations have been dealt using also the human intervention.

Nowadays, thanks to the large amount of research results in the lasts two decades, the mobile robotics is no more confined in the industrial and space applications but now includes also a new field, the service robots. The service robots field, in addition to being a new market in continuous expansion [1], gives also the opportunity to test new methods and algorithms that in future can be applied also to more critical fields, like the industrial and spatial ones. Though this application field may seem simple, it must deal with a general scenario, human interactions and less restrictions on the environment knowledge. These arguments become relevant in the mobile robotics field: the robots can navigate in the environment without rigid connections while instead a classic manipulator is strictly constrained to the environment.

Mobile robots must deal with a large variety of environments. Furthermore, the environment can also change with time and usually it is not possible to place sensors in it. This leads to a mobile robot that must be able to collect information from the environment and so navigate by using only the data from on-board sensors. Additionally it must be able to cope with variations in the environment. The variations covered may be not of any type, their time scale must be related also to their importance. While is reasonable that a man or another vehicle might moving in the environment, more important mutations like the environmental framework must have a slight dynamic.

On the one hand this means that during the navigation task the robot has to explore the environment or using information previously achieved by

taking care of the environment variations. On the other hand, once that the robot is capable to perform these operations, it means that it would be able to navigate in a general scenario. The importance of this aspect has been also emphasized in the *Strategic Agenda for Robotics in Europe* [2] in which, as depicted in Figure 1.1, the exploration and inspection task has been chosen as one of the macro-applications. It must be aware that this classification is very restricted to single applications: for instance also the logistic robotics needs a navigation with similar requirements than the exploration. In the same image is also noticeable that in all the sectors there is the need of mobile robots able to localize themselves in an unknown environment.

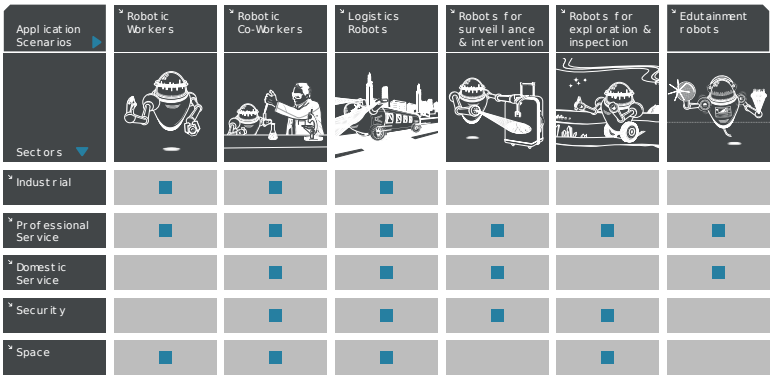


Figure 1.1: Robotics application scenarios in the different service sectors [2].

The generic considerations reported so far are useful to introduce the topic of the thesis, which is the motion measurement by opto-electronic systems. Being more specific, the motion measurement here is meant as the localization of an unmanned vehicle, but also the detection of objects in the environment and their localization with respect to the vehicle: even though they seems to be very different, it must be highlighted that the basic framework is very similar.

Strictly connected to the requirements of navigation in an unknown environment is the choice of opto-electronic sensors. Their main advantage is that they permits non-contact measurements and they can be easily fitted also in small vehicles. In addition, by selecting the appropriate sensors, they can return information on the surrounding environment, without requiring its modification. It should be pointed from the outset that they can provide only partial information, no matter how accurate or reliable they are. Moreover, the motion measurement must be inferred from these limited information. This is probably one of the limiting problems of these instrumentation.

With reference to the vehicle localization, when are used sensors able to collect information on the surrounding environment, and these information

are recursively used to estimate the vehicle position, the reference problem is called *Simultaneous Localization And Mapping (SLAM)*. This is a well known problem that theoretically allows to achieve the best estimation of the pose vehicle in a scenario without a priori knowledge of the environment. As evidence of its importance there is the huge amount of scientific works produced on this topic in the last decade. Against of all the works done in this field and the results obtained, as will be explained further, there are still some key points that require to be investigated, in order to consider the SLAM problem as solved. One of these is the data association: indeed, also the SLAM is based on recursive matching of environmental measurements. Due to the fact that these one are partial information and additionally the environment might undergo mutations, a reliable data association procedure is needful.

The work done for the thesis is based on these premises. The sensors used are a 2D laser range finder (LRF) and a camera: also if very different, both are opto-electronic sensors, capable to collect information from the environment. Particular attention was paid in modelling the uncertainties involved in the phase of matching the environmental measurements. The measurement uncertainty is propagated in the overall procedure up to the final motion estimation.

In addition, a more detailed investigation has been performed on the localization with laser range finder, in order to understand in which manner use the data acquired from the sensors and to build a reliable localization system used in the benchmark of the camera localization results. In this way it is possible to obtain a weighted estimation using the information available in a optimal sense. The present work is organized in three parts, starting from the theory and the literature review on mobile robot localizations, passing trough the proposed algorithm and ending with the experiments investigation and validation of the algorithms.

In the first part (Chapter 2) is reported a more detailed descriptions of the thesis topics briefly reported in this introduction. In addition it is possible to find a review of the related scientific works in literature.

In the second part (Chapter 3) are defined formally the approaches and algorithms implemented for the vehicle localization. In particular it is described the data organization and the data processing.

In the third and last part are briefly introduced the instrumentation used and the results (Chapter 4), and finally (Chapter 5) the conclusions.

Chapter 2

Review of the Literature

In this chapter is introduced a more detailed description on the three main topics of the thesis: the processing of the information acquired from the environment, the data association and the incremental and global localization approaches. For each point is also reported a literature review on the most relevant scientific works. The aim of this survey is to understand deeply the problems, the evolution of the methods used in its solution and to verify the possibility to improve or design alternative methods with a more suitable and accurate formulation.

2.1 Environment Information Processing

The aim of the thesis is to measure the motions of an agent, in this particular case an *Automated Guided Vehicle (AGV)*, through an unknown environment by processing the information collected from the environment. As reported in the introduction, to achieve the localization in a generic environment, there are strong limitations on the sensors. The instrumentation used have to fulfil two simple rules:

- all the components of the sensors must be fitted on the vehicle.
- the sensors operating principle has not to require the modification of the environment.

Although these constraint appear simple to be met, only few categories of instrumentations can be used and just a small number of sensors for each category are suitable. As already mentioned, the opto-electronic systems like laser range finder and camera are the most indicate. In the last two decades several scientific works in literature have proved their worth. There are also other sensors that meet the requirement, one for all the odometric reconstruction with encoder on wheels. The main drawback in this case is that they allow only a dead reckoning localization with well known problems

in terms of error drift in the vehicle pose estimation, due to the not perfect modelling of the wheel and its contact with the ground. In addition the odometric reconstruction can be done only on wheeled vehicles: currently it is not a big deal but probably in the future it will be a strong limitation.

Opto-electronic systems instead can retrieve more information on the surrounding environment that can be used in the localization; also regards on the vehicle kinematic they have few limitations. Their main drawback is instead in the measurement, since they did not return a direct information of the motion but this one has to be inferred from the retrieved data. This is the first problem to solve: how to extract from the environment, or rather from its sensor representation, useful information for the vehicle localization. This choice is not an end in itself, because also the implications on the two other main aspects, the data association and the localization, should be considered. For this reasons, before to disclose the methods implemented, it is better to give an overview on the literature works done until now.

Prior to talk about specifically on the sensors it is also better to define the kind of scenario used. The scenario can be defined using different criterion:

- structured or unstructured environment
- a priori knowledge of the environment
- environment geometry
- outdoor or indoor environment

As already mentioned in the introduction, the aim to localize the vehicle in the more general kind of scenario. This means that the scenario will be unstructured, or rather an environment with a not well defined structure, which may be subject to mutations. The second criterion in case of unstructured environment is in some way already defined: it is not possible to have a priori knowledge of the environment when it is not structured. Another important criterion is the environment geometry. If the environment can be represented with a set of basic geometric shapes, it can also be described in terms of this geometrical characteristics: on the contrary, for its definition it must be used a more general procedure. The last classification is based on outdoor or indoor environment. In this thesis it has been chosen to work in an indoor scenario, mainly because it is more easy to control and check the results. However it can be said that if the outdoor scenario allows to use in an efficiently manner the opto-electronic systems, the algorithms and approaches developed here can be used also in this situation.

With this premise is possible now to proceed in the analysis of the state of the art of the sensors used, the laser scanner and the camera.

2.1.1 Laser scanner

There are several nomenclatures to identify this type of sensors, laser scanner, Laser Range Finder (LRF), Light Detection And Ranging (LIDAR): all the nomenclatures try to define in few words its working principle. The key concept is the estimation of the *Time of Flight (TOF)* of the laser beam path from the emitter to the illuminated target and back to the detector. Because of in common laser scanner, for the time measurement, is used a pulsed light emission, the method is classified as *Pulsed Modulation* (Figure 2.1), that is in contrast with the *Continuous Wave Modulation* used instead in other instruments (like the TOF Cameras). The configuration with one emitter

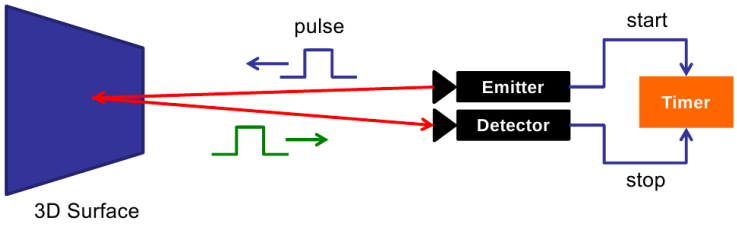


Figure 2.1: Graphical representation of a TOF sensor with a Pulsed Modulation scheme.

and one detector returns the measure of distance of a single point: in order to collect information from the the environment it is not enough. For these reason in the early 90s has been introduced *Cyclone* [3], the first laser scanner capable to recover a bi-dimensional profile of the environment suitable for the use in vehicle localization. The main component is again a couple of laser emitter and detector, but thanks to a rotating mirror, the laser beam can be pointed in different direction. The novelty of this approach was the high rotational velocity of the mirror: in this way the effects of the time-delay between the measurements is minimized. Moreover the effectiveness of this solution is confirmed by the fact that almost all the current laser scanners use the same principles (Figure 2.2(a)). Further improvements has led to laser scanners capable to acquire three-dimensional information from the environment, as the one depicted in Figure 2.2(b).

In the algorithms implemented for the thesis it has been chosen to use a bi-dimensional laser scanner, though, has previously reported, nowadays already exist instrumentation that can return three-dimensional data. This choice requires some justifications, because is mandatory to ask why do not use sensors that can return more information on the surrounding environments. The answers to this question are several. By figuring out the possible targets for the localization procedure proposed, in most of them the vehicle motion can be considered as planar. In this case there is not the need to use more data in addition to those provided by bi-dimensional sensors.

(a) Sick Laser scanner (2D)

(b) Velodyne Laser scanner (3D)

Figure 2.2: Laser scanner commonly used nowadays: a bi-dimensional industrial-grade laser scanner in which is visible the spinning mirror for a 180°*deg* field of view (a) and a three-dimensional outdoor laser scanner in which there is an array of basic sensors that rotate together for a 360°*deg* field of view (b).

On the other hand, by using a bi-dimensional laser scanner, there is a lack of information outside the plane defined by the laser beams: it is possible to take care of it by using an additional camera, as reported in the Section 2.1.3. Furthermore, as reported in the introduction, the SLAM is not yet a completely closed problem in its bi-dimensional formulation and there is still work to do in order to define the SLAM as a fully reliable solution, also in the bi-dimensional case. In addition the algorithms and approaches used in the bi-dimensional formulation can also be used, with the necessary modifications, in the three-dimensional field.

2.1.1.1 Motion estimation with laser scanner

The motion estimation using a laser scanner is the key-point to achieve to the localization of a mobile robot, as will be explained in the Section 2.3. The localization can be defined as a main procedure that requires subroutines: one of the most important is the motion estimation from the sensors, the laser scanner in this case. The aim of the vehicle localization is to estimate, for a planar motion, its *pose* $\mathbf{X}$, composed by position and attitude $\mathbf{X} = [x, y, \theta]^\top$ referred to the environment reference frame Σ_{env}. Using laser scanner, at each vehicle pose is possible to acquire a measure of the environment, commonly defined as *scan* q, which consists in a set of ordered points. The scan acquired from the laser in two different vehicle poses $\mathbf{X}_1$ and $\mathbf{X}_2$, can be used to estimate the laser motion by matching the two observations (Figure 2.3). For this reason the operation is called *scan matching*. Because of the laser is rigidly connected to the vehicle, also the vehicle pose variation is known.

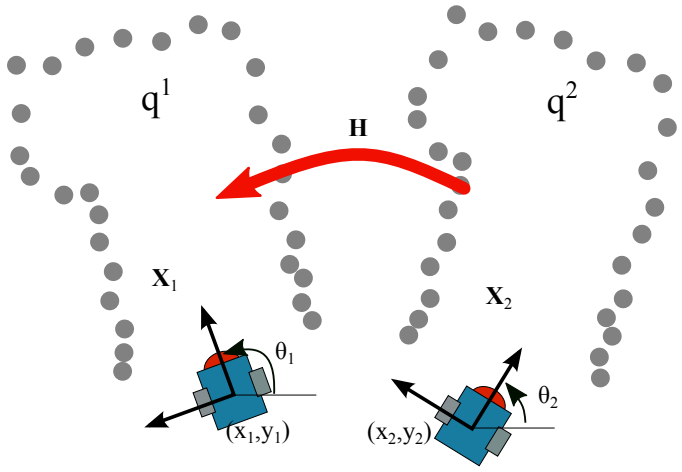


Figure 2.3: Graphical representation of the motion estimation **H** with laser scanner between two vehicle poses $\mathbf{X}_1$ and $\mathbf{X}_2$ by observing the q^1 and q^2 scans acquired.

More in general the scan matching can be defined as the procedure that allows to obtain the estimation of the planar transformation **H** between two vehicle poses or the vehicle pose and a map. A map is defined as synthetic representation of a known environment: the type of data stored are strictly related to the approach used in the scan matching. As it is possible to figure out, there are several approaches to match the scan, and the key for a fast reliable and accurate localization is hidden behind too this selection.

Analysing the most important works in literature, it can be discerned two main categories:

- Feature based scan matching.
- Dense scan matching.

This classification is based on the type of data used in the planar transformation solution. A *feature based approach* actually require two steps. The first one is the identification in the environment of natural landmarks, defined *features*, and their parametrization, in both the scans. In the second step, the features identified in the two scans are associated in order to perform the scan matching. Unlike the feature based approaches, *dense matching approaches* do not need to identify and describe features, since the whole scan is used to provide the solution at the scan matching problem.

These categories are well known from the very early literature works on vehicle localization by using laser scanner. One of the first implementation of a scan matching approach was the Cox [4] application to an experimental autonomous cart in the 90s, in which he used a dense matching approach. Few times later, a similar approach was also applied by Gonzalez [5] in a

more formal manner. On the same year Shaffer proposed a feature based method for underground mine vehicle navigation [6] and also a first comparison between the two categories of approaches [7]. After this first introduction of the features, the successive publications go ahead in parallel by keeping this classification. In all these applications the scan matching uses a priori known map in order to perform a global localization (Section 2.3): the further step was the matching between two laser scans in order to avoid the map knowledge limitation. From the side of the feature based approach, the works of Arras [8] and Einsele [9] start to deal with the problem of the localization and unique description of features. In parallel, also the dense matching approaches were applied between two laser scans: a remarkable work was done by Lu [10], in which was explicitly introduced a closed form solution for the scan matching. As prove of the importance of this new methods, Gutmann in [11, 12] performed again a comparison of these two main approaches, also by studying the influence of the noise in the laser measurements.

At the same time, new procedures to associate data between two scans has been introduced, but more or less this new approaches were the evolution of the *Iterative Closest Point (ICP)* used in the Besl [13] and Zhang [14] articles. The aim of these implementations was to improve the robustness of the ICP approach, like in the approach implemented by Montesano [15].

In the further years there were released essentially refinements of all these techniques, by introducing probabilistic aspects in the feature identification and in the data associations for both the approaches and by improving their performances for real time applications. Relevant results on probabilistic improvement are described in the Jensfelt article [16] and in the Censi works: in [17] he reported a very accurate closed form uncertainty estimation of a generic ICP approach, while in [18] is described how to estimate the uncertainty lower bound from the map of a known environment.

Since the features used until nowadays are based on segments identification in the scan (for a detailed description see Section 3.5.3), there are a lot of references also on this topic. While Diosi [19] perform a complete discussion on line uncertainty in Hesse Plane and a complete analysis of the laser scanner sources of noise, the work of Nguyen [20] and Martinez-Cantin [21] deal with the accurate segments identification. In the former it is possible to find an useful investigation on the performances of the most used algorithms to extract segments, while on the latter it is proposed a new procedure for the segments identification in a fully parametric way.

With regards to the dense matching approaches, the improvements were few. A remarkable work has been done by Censi [22] which, instead of using a classical scan matching with points, he introduced a scan matching based on the environment shape. In addition, in the last decade, another category of dense matching approaches, based on the polar representation of the scan, has been in introduced. A relevant work in this field is the Diosi one [23].

2.1.2 Camera

Cameras are nowadays largely used as measurement system in several applications, mainly to measure or reconstruct the three-dimensional shape of objects in the camera field of view. These kinds of applications are usually carried out using artificial markers or by working in controlled scenarios and, additionally, using more than a camera. In the last decade, thanks to some remarkable scientific works, this limitations has been overcome. It is possible to find applications capable to reconstruct the surrounding environment without the use of artificial landmark and without a priori knowledge of camera positions.

The most important motivation to use the camera as a measurement system is due to the large amount of information that this sensor can return. With a single shot a camera is capable to return a representation of all the visible surrounding with information on colour, textures and shadings. Its main drawback is that in order to represent the three-dimensional world in the bi-dimensional domain of the image it is required a projection. As will be explained in the further sections this is an irreversible transformation. The system of equations that this projection defines is undetermined: it can return only an infinite set of solutions known up to the *scale factor* (Section 2.1.2.2). The problem is resolvable by adding constraints to the problem: the most used one is the addition of another camera in order to resolve the scale factor ambiguity by using the cameras in a stereo configuration. A noticeable applications of this approach is the *Mars Exploration Rovers* navigation systems, based on the *Navcam* stereo-camera (Figure 2.4). In this thesis it has been decided to not use the solution of

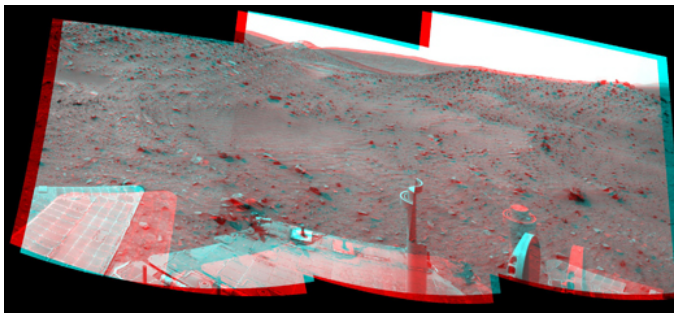


Figure 2.4: Picture mosaic from the Curiosity Navcam.

stereo-camera. For several reasons, reported in the Section 3.5.1, it is more advisable the use of only one camera.

2.1.2.1 Motion estimation with camera

The motion estimation using only a camera has been really improved in the last few years. This aspect is particular interesting: indeed the most im-

portant formulations to model the camera and its transformations are well known from more than 30 years. The main problem in the application of these approaches was the computational cost to process the large amount of information in each single image, and the lack of reliable image features detectors. Like in the laser scanner data, also in this case there is a duality on the approaches used to infer information from the images. As already mentioned the first one is based on the *features detection* and their successive association; the other approach is the *dense reconstruction*. Differently from the laser scanner these approaches, especially for the motion estimation, were not developed in parallel. Most of the works was focused on the features: indeed the dense approach is still called *reconstruction*, that is related to the classical procedure to use camera in the reconstruction of three-dimensional shapes.

In this section, are reported the most relevant scientific works, not only in the specific application of mobile robot localization, but in a more general scenario. A useful and complete survey on this approaches can be found in the Scaramuzza and Fraundorfer tutorials [24, 25]. Actually the approaches are categorized in *Visual Odometry* and *Visual SLAM*. Really interesting works in this last category were presented by Davison [26] and Chli [27]. In their publications is possible to find a complete argumentation of the visual SLAM with a camera, starting from the feature extraction to the map building: all these procedures are carried out by using a full covariance approach to estimate the motion uncertainty. Another interesting work is [28], in which is explained how it is possible to reconstruct the structure and the motion of a camera across multiple views by identifying the edges in the images.

Scientific works more focused on vehicle localization are [29], that is one of the early work applied on vehicles motion estimation, Mouragnon [30] and Konolige [31] one: they are relevant because they work in a complete unknown and very large environment, reporting good results. In conclusion, is possible to mention also the works proposed by Ortin [32] and Scaramuzza [33], in which is introduced the robust estimation in the association phase (for a detailed description see Section 3.6).

2.1.2.2 Epipolar geometry

The estimation of the motion of a camera, as reported previously, can be performed in several ways. In any case all the methods that use a camera have to deal with the mathematical and geometrical description of the image formation and the formulation of the multiple view geometry. For this reason in the further paragraphs it will be introduced the basic geometry to describe the camera. Furthermore, by studying these basic notions,s it is possible to understand better the limitations of using cameras to estimate the motion. The most used formulation in the computer vision is the one described in

Zisserman book [34].

From a mathematical point of view, a camera can be described as a mapping between the three-dimensional space of the world and the bi-dimensional space of the image; there are several geometric formulations to describe this phenomenon. The most used and studied camera model is the one base on the *central projection*, which is a good compromise between formulation simplicity and approximation of the real principle. The central projection can be also progressively modified by modelling additional characteristics, in order to give a better representation of the real situation. This basic camera model is the *pinhole camera*: it is based on the perspective projection of 3D points on the camera sensor. To define its relations, it must be taken as the centre of projection the origin of the camera coordinate frame **C** and consider the plane $Z = f$ as the *image plane* (Figure 2.5). With the

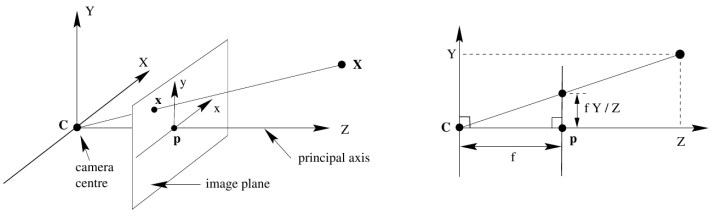


Figure 2.5: Camera perspective model used to map 3D points into the image plane [34].

pinhole camera model a point in the three-dimensional space $\mathbf{X} = [X, Y, Z]^T$ is mapped into a bi-dimensional point $\mathbf{x}$ in the image plane. This point in the image plane can be easily computed using triangles similarity. In addition it is possible to denote the centre of projection **C** as the *camera centre*, while the line from camera centre and perpendicular to image plane is called *principal axis*.

Representing the world and image points with homogeneous vectors, the central projection can be easily expressed in term of matrix multiplication:

$$\mathbf{x} = \mathbf{P}\mathbf{X} \tag{2.1}$$

where **P** is defined as the *camera projection matrix*. The geometric relationship and the camera projection matrix can be improved considering the principal point offset of Figure 2.6: the origin of the image coordinate frame is not on the the *principal point p* defined by the optical axis. So the matrix multiplication expressed in homogeneous coordinates becomes:

$$\begin{pmatrix} X \\ Y \\ Z \\ 1 \end{pmatrix} \mapsto \begin{pmatrix} fX + Zp_x \\ fY + Zp_y \\ Z \end{pmatrix} = \begin{bmatrix} f & 0 & p_x & 0 \\ 0 & f & p_y & 0 \\ 0 & 0 & 1 & 0 \end{bmatrix} \begin{pmatrix} X \\ Y \\ Z \\ 1 \end{pmatrix} \tag{2.2}$$

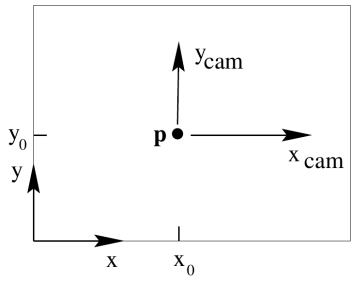


Figure 2.6: Coordinate frames in camera image plane [34]. The perspective relationship project a 3D point in the $\Sigma(x_{cam}, y_{cam})$ coordinate frame, while images points are expressed in the $\Sigma(x, y)$ coordinate frame.

where f is the focal length of the camera and $\mathbf{p} = [p_x, p_y]^\top$ is the principal point offset on the image plane. This formulation can be expressed conveniently in a more concise form:

$$\mathbf{x} = \mathbf{K}\,[I \mid 0]\,\mathbf{X}_{cam} \tag{2.3}$$

where $\mathbf{K}$ is the *camera calibration matrix*, and $\mathbf{X}_{cam}$ is the position of the three-dimensional point expressed in the coordinate frame of the camera, which has the origin in the camera centre $\mathbf{C}$ and it is called *camera coordinate frame*.

$$\mathbf{K} = \begin{bmatrix} f & & p_x \\ & f & p_y \\ & & 1 \end{bmatrix} \tag{2.4}$$

In addition, for a general finite camera, other parameters can be added, as the number of pixel per unit distance in image coordinate (considering unequal scale factors in each direction) and the *skew* parameter s. The updated camera calibration matrix has the following form:

$$\mathbf{K} = \begin{bmatrix} \alpha_x & s & x_0 \\ & \alpha_y & y_0 \\ & & 1 \end{bmatrix} \tag{2.5}$$

where α_x and α_y are the representation of the camera focal length f in terms of pixel dimension respectively in the x and y direction. The same transformation must be applied also to the parameter $\mathbf{p} = [p_x, p_y]^\top$, which in pixel dimension is denoted as $\mathbf{x}_0 = [x_0, y_0]^\top$ (Figure 2.6).

In general, points in the three-dimensional space are expressed in a coordinate frame different from the camera one. Assuming that the camera pose is known, the transformation that must be applied to transform a point in a generic three-dimensional coordinate frame to obtain its position in the camera coordinate frame is:

$$\mathbf{X}_{cam} = \begin{bmatrix} \mathbf{R} & \mathbf{t} \\ 0 & 1 \end{bmatrix} \mathbf{X} \tag{2.6}$$

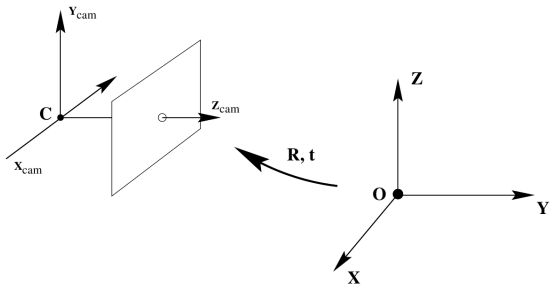


Figure 2.7: Camera 3D coordinate frame $\Sigma(X_{cam}, Y_{cam}, Z_{cam})$ and geometrical transformation R, t to the global coordinate frame $\Sigma(X, Y, Z)$ [34].

The matrix representation con be easily updated with this information:

$$\mathbf{P} = \mathbf{K}\,[\mathbf{R} \mid \mathbf{t}] \tag{2.7}$$

Usually, the parameters that define the camera calibration matrix $\mathbf{K}$ are called *intrinsic parameters*. Their definition is possible only after a calibration procedure of the camera.

The rotation $\mathbf{R}$ and translation $\mathbf{t}$ between the camera coordinate frame and the global coordinate frame are defined *extrinsic parameters*. Even these can be defined during the calibration, and remain valid until the position and orientation of the camera coordinate frame are fixed.

An important consideration must be done on the perspective model: because it is essentially a mapping operation from a three-dimensional space to a bi-dimensional space, there is a loss of information. This mapping led to an undetermined system without a unique solution: a point on the image plane produces infinite solution in the three-dimensional space. This is noticeable by observing the Figure 2.5: all the infinite solutions lie on the line that pass through the camera centre $\mathbf{C}$ and the point $\mathbf{x}$ on the image plane.

This means that to retrieve information from the camera, at least two view is needed, in order to obtain two incident projection lines Figure 2.8. With two view the reference geometry is the *epipolar geometry*. It is essentially the geometry that describe the intersection of a plane which contain the three-dimensional point and its projections on the two camera image planes.

Given a 3-D point in space $\mathbf{X}$ and his projection in first view $\mathbf{x} = \mathbf{PX}$ and second view $\mathbf{x}' = \mathbf{P}'\mathbf{X}$, epipolar geometry allows to compute one these three problems at a time:

Correspondence geometry use of the constraint between $\mathbf{x}$ and $\mathbf{x}'$ image points in order to find their correct correspondences.

Camera geometry with a set of corresponding image points, it is possible to estimate the camera matrices $\mathbf{P}$ and $\mathbf{P}'$.

Scene geometry with two corresponding image points $\mathbf{x}$ and $\mathbf{x}'$ and the camera matrices $\mathbf{P}$ and $\mathbf{P}'$, computes the position of three-dimensional point $\mathbf{X}$.

2.1.2.3 Fundamental and Essential Matrix

Classical epipolar geometry deals with epipolar plane, epipolar lines and their intersections. This geometrical representation is not the more suitable in modern application where it is preferred a mathematical representation with matrices. The *Fundamental Matrix* is the algebraic representation of epipolar geometry. In particular it represents the epipolar constraint: the projection of a point that belongs to the first camera image plane $\mathbf{x}$ must lie on the epipolar line l' on the second camera image plane (Figure 2.8).

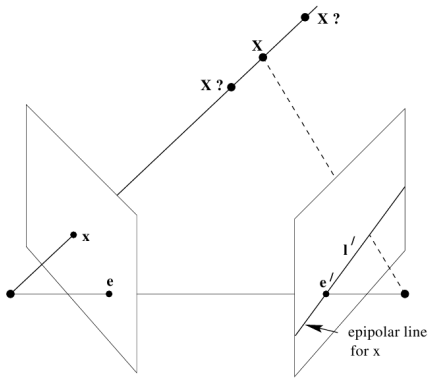


Figure 2.8: Epipolar constraint graphical representation. The projection of the point $\mathbf{x}$ on the second image must lie on the epipolar line l'.

The epipolar line l' can also be seen as the projection in the second image of the point $\mathbf{x}$ in the first camera, because of the three-dimensional point $\mathbf{X}$ is known up to the scale factor. Hence the relation between $\mathbf{x}$ and l' is a projective mapping from point to line:

$$\mathbf{x} \mapsto l' \tag{2.8}$$

The Fundamental Matrix formulation is actually based on this mapping, and it can be defined as follow:

$$l' = \mathbf{F}\mathbf{x} \tag{2.9}$$

where $\mathbf{F}$ is the Fundamental Matrix. Since this mapping project a point in the image plane from bi-dimensional space to a line, a one-dimensional

space, the Fundamental Matrix is not a full rank matrix. $\mathbf{F}$ is a 3×3 matrix of rank 2.

The Fundamental Matrix has become largely known since by using it is easy to check the *correspondence condition*.

$$\mathbf{x}'^{T}\mathbf{F}\mathbf{x} = 0 \tag{2.10}$$

The Fundamental Matrix permits to satisfy that condition for any pair of corresponding points $\mathbf{x} \mapsto \mathbf{x}'$ in the two images. This is an important relation, because it gives a way to characterizing the Fundamental Matrix without reference to the camera matrices, only in terms of corresponding points in two different views.

Additionally, the Fundamental Matrix has some important properties, that can be used to solve several problems:

Transpose if $\mathbf{F}$ is the matrix of the pair of cameras $(\mathbf{P}, \mathbf{P}')$, then $\mathbf{F}^{\top}$ is the matrix of the pair in opposite order.

Rank equal to 2 because it is not a full-rank matrix, it also satisfies null determinant $det(\mathbf{F}) = 0$.

Seven degree of freedom $\mathbf{F}$ is a 3×3 homogeneous matrix, with seven degrees of freedom. The number of independent variables is reduced by the null-determinant constraint and by the common scaling of mapping which remove two degrees of freedom.

Another important matrix in the epipolar geometry is the *Essential Matrix*. It is the specialization of the *Fundamental Matrix* to the case of normalized image coordinates. In other words it could be used under the assumption of calibrated cameras (for the intrinsic parameters). Furthermore it could be useful since it has fewer degrees of freedom and additional properties if compared to the Fundamental Matrix.

With the aim to understand the relationship between Fundamental and Essential Matrix, it is better to introduce the normalize image coordinates. The camera matrix can be decomposed as $\mathbf{P} = \mathbf{K}[\mathbf{R} \mid \mathbf{t}]$, and used in the mapping relation 2.1. If the calibration matrix $\mathbf{K}$ is known, it is possible to obtain the point $\widehat{\mathbf{x}} = \mathbf{K}^{-1}\mathbf{x}$ and then $\widehat{\mathbf{x}} = [\mathbf{R} \mid \mathbf{t}]\,\mathbf{X}$. The geometrical entity $\widehat{\mathbf{x}}$ is the image point expressed in *normalized coordinates*: it is possible to express a point in this coordinate frame only with a camera calibrated in the intrinsic parameters.

To define the Essential matrix, it must be considered a pair of normalized camera matrices $\mathbf{P} = [\mathbf{I} \mid \mathbf{0}]$ and $\mathbf{P}' = [\mathbf{R} \mid \mathbf{t}]$. The corresponding fundamental matrix is called *Essential matrix*, and has the form:

$$\mathbf{E} = [\mathbf{t}]_x\,\mathbf{R} = \mathbf{R}[\mathbf{R}^{\top}\mathbf{t}]_x \tag{2.11}$$

Also the essential matrix can represent the correspondence condition 2.10, but in terms of normalized coordinates:

$$\widehat{\mathbf{x}}'\mathbf{E}\widehat{\mathbf{x}} = 0 \tag{2.12}$$

By substituting $\widehat{\mathbf{x}}$ and $\widehat{\mathbf{x}}'$, and comparing this relation with the one for fundamental matrix 2.10, it follows the relation:

$$\mathbf{E} = \mathbf{K}^\top \mathbf{F} \mathbf{K} \tag{2.13}$$

This equation represents the relation between the Fundamental and the Essential Matrix. Like the Fundamental Matrix, also the Essential Matrix has some important properties:

- the Essential Matrix $\mathbf{E} = [\mathbf{t}]_x \mathbf{R}$ has only five degrees of freedom (*dof*). Rotation and translation have both three *dof*, but one in the translation is lost due the scale factor ambiguity.
- the Essential Matrix is a 3×3 matrix: two of its singular values must be equal and the third must be zero.

In conclusion, the Fundamental and the Essential Matrix allows to solve the correspondence equation between two set of points acquired in two different views. If the camera is calibrated in its intrinsic parameters, is possible to recover the variation of the extrinsic parameters between the two views. Unfortunately, the camera positions, and so the motion estimation, is known up to a scale factor in the translation.

2.1.3 Laser scanner and camera

In the previous Sections 2.1.1 and 2.1.2 the laser scanner and camera has been introduced as measurement systems to estimate the vehicle motion between two poses. Their main characteristics are reported in Table 2.1, and as presumable these instrumentations return quite different information. Looking closer an comparing these characteristics is possible to notice that laser scanner and camera are complementary instruments: on can supply the weakness of the other.

While the laser scanner returns a bi-dimensional environment representation on the plane defined by the laser beams, the camera achieve to collect a large amount of informations also out from this plane. On the other hand, all the camera information are known up to a scale factor, whereas the laser return already a metric information. For this reason also the uncertainties in the laser pose estimation are usually limited, because of the camera requires the knowledge of the scale factor. Lastly there is the reflection problems: in the laser scanner when the emitted laser beam hit some particular kind of surfaces the measures is not reliable. The camera instead use the environment light, so if there are some problems of reflection they can be solved for instance with optical filters.

Table 2.1: Laser Scanner and Camera measure characteristics comparison.

	Laser Scanner	**Camera**
Environment representation	2D shape	2D representation of 3D space, with colour and texture information
Metric measurements	Return directly a distance measurement	Distances and dimensions determined up to a scale factor
Uncertainties	Low uncertainty in frontal direction	Scale factor increase the translational uncertainty
Reflection problems	Problems on some type of surfaces	Can be solved with optical filters

2.1.3.1 Motion estimation with laser scanner and camera

By taking care of all the considerations reported until now on camera and laser scanner, the next step is how to get these sensors working together. In the previous literature review, has been seen how they each sensor can estimate the motion independently. Indeed in most of the scientific works they are used as exclusive instruments, while they are complementary and they can share information to improve the motion estimation and the environment knowledge.

During the development of the thesis work, two categories of approaches to share the sensors information were been identified:

- by estimating the motion independently with each instruments and exchanging the informations on the motion measured. Thereafter it is possible to merge the information with sensor fusion, in order to achieve a better solution.

- by merging the measurements collected from the instruments in order to increase the amount of information and in a later stage compute the motion based on this fused information.

While there are few relevant works on the first category, it is possible to figure out a framework in which laser scanner and camera exchange information on the pose estimation, not only in the final step, but also in an intermediate step in order to solve the camera scale factor ambiguity. The Naikal work [35] is very closed to this idea ant it proved the effectiveness of this idea in a vehicle localization framework.

Concerning the second type of approach, it is possible to find more interesting scientific works. One of the first implementation of this approach

was done by Arras [36]: he proposed an approach based on the identification of features in the laser scanner, and vertical edges in the camera. These entities are representative of the same natural landmark in the environment, and so by fusing these information it is possible to achieve a robust motion estimation based on this redundancy. Later, Andersen [37] used a similar approach to detect the road width by fusing laser and camera information.

2.2 Data Association

Data association is the second topic of the thesis. It is really important since it is required in all the scan matching approaches and also in the data association from two different views in the motion estimation with the camera. How it will be explained in the further sections, it has also to deal with uncertainty, outliers and similar problems.

2.2.1 What means data association?

In many applications, not only in mobile robots localization, when there are more measurements of the same phenomenon, in order to obtain useful information it is needful to associate the data of different observations. In some cases, the number of measurements for each observation is very limited and the data association could be an unnecessary step. Actually, also in this cases, there can be important problems in the processing of information due to the presence of *outliers* in the measurement.

In the case of mobile robot localization with sensors that are capable to acquire information from the environment, since the motion estimation has to be performed on this data, it is required a reliable data association. In this way it is possible to infer the vehicle motion. Because of the data association is a general procedure, it is possible to identify two general sets of points q^1 and q^2 and use them to introduce the approach (Figure 2.9). What makes the data association a difficult problem is the identification of the correspondences between these sets of points.

A popular approach to solving the problem is the class of algorithms based on the *Iterative Closest Point (ICP)*, introduced by Besl [13]. ICP is widely used because it ensure a very simple implementation and good performances. Usually it need also an initial estimation and if it is reasonably good the algorithm converges quickly to the correct solution, as demonstrated in [13].

The ICP can be identified nowadays as a class of algorithms, because of from its original concept several improvements have been developed. The main reason is the presence of missing data or *outliers* in the association: it means that only by using the ICP it is not possible to recover all the right correspondences in an reliable way. Relevant works that deal with this kind of problem was reported by Wang [38] and Trucco [39]. While the former

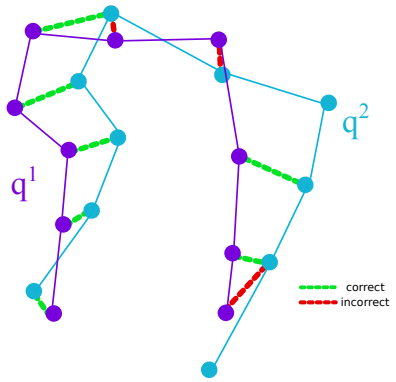


Figure 2.9: Data association between two set of points q^1 and q^2: the ICP algorithm can create incorrect associations.

introduced this problem and proposed a solution, in the scientific work of Trucco is mentioned for the first time the word *Robust ICP (RICP)*. In [40] is possible to find a comparison of several ICP approaches: in particular are highlighted the benefits achieved by a robust implementation. In conclusion it is possible also to report a recent work [41], in which is investigated the importance of the unique correspondences. Actually, the ICP can be performed using only information on the points position: the points selected are the more closest. But it is also possible to add more information on the points besides their positions. In this way the correspondences are performed not only by taking care of the distances, but also on the other aspects able to identify uniquely the same point in two sets.

2.2.2 Detection of outliers: the LMedS method

As remarked in the short introduction, on of the most relevant problem in the data association procedure it the presence of outliers. Actually this is a real problem in many other applications when there is the need to model a phenomenon with a mathematical interpretation. Also the data association required for the vehicle pose estimation can be seen as a procedure to model the transformation between two poses. This aspect will be explained better in Section 3.4.1, where is reported the procedure implemented in order to achieve a robust data association.

In this section the detection of outliers is presented with reference to a generic model interpolation. Let f be a generic model described by the following relation:

$$\mathbf{y} = f(\mathbf{x}), \tag{2.14}$$

where $\mathbf{x} \in \mathbb{R}^n$ is the vector of the n coefficient parameters used in the model, which must be inferred from the m measures $y_i, i = 1, ..., m, m > n$. Assuming, to simplify the explanation here, that the function $f(\mathbf{x})$ is a linear

function on $\mathbf{x}$ which can be wrote as:

$$f(\mathbf{x}) = \sum_{j=1}^{n} x_j \theta_j, \tag{2.15}$$

where $\theta = [\theta_1, ..., \theta_n]^\top$ is the vector of independent variables used to describe the problem. The solution of the problem can be achieve with a multiple regression in order to estimate the vector θ, by using the measures y_i. If the problem is well posed, this means the resolution of an overdetermined system of equation.

There are several methods to solve the system, but the Gold Standard in this situation is the *Total Least Squares (TLS)* method. It is easy to demonstrate that in presence of measurements with normal distribution, the TLS is the best estimator and it solves the problem:

$$\min_{\theta} \sum_{i=1}^{n} r_i^2, \tag{2.16}$$

where r_i are the residual for each measurement y_i and is defined as:

$$r_i = y_i - \sum_{j=1}^{n} x_j \theta_j. \tag{2.17}$$

The simplicity of application of this method is in contrast with its lack of robustness: one single outlier can have a relevant effect on the data interpolation with a model (Figure 2.10). For this reason it is useful to introduce

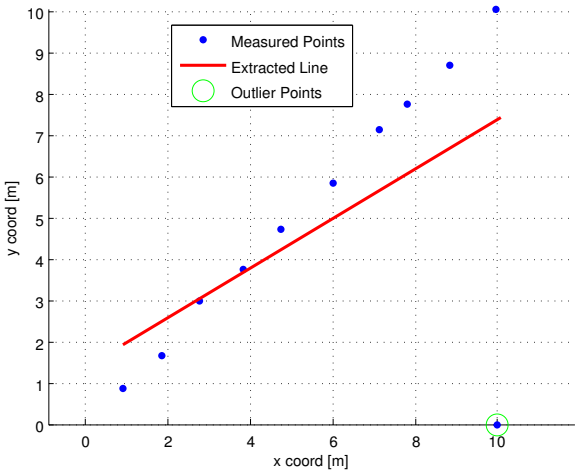


Figure 2.10: Least Squares line interpolation in presence of outliers.

the *breakdown point* ϵ^*, that is defined as the smallest percentage of outlier

data that cause the failure of the estimation of the model parameters. In the TLS method this value is $\epsilon^* = 0$: it means that the TLS can not deal with outliers.

In order to increase the breakdown point, a well known method is the *Chauvenet's criterion*. It is based on the statical analysis of the residuals: the outliers are identified as the residuals with less probability. The main drawback of the problem is that the residuals are computed after the interpolation of all the data, hence also including the outliers.

To increase the breakdown point, Siegel [42] and then Rousseeuw [43] identify the limiting factor of the TLS in the sum used in the functional (Equation 2.16). Using instead the median, it is possible to achieve an higher breakdown point of $\epsilon^* = 0.5$: intuitively if there is an outlier percentage of the 50%, by choosing the median of the residuals computed with the ideal model, is possible to select the last inlier value in the set of data. The method implemented in this way is called *Least Median of Square (LMedS)*.

In [43] are also reported and proved two important characteristics of the LMedS:

- there always exist a solution to the LMedS problem.
- the breakdown point ϵ^* for the LMedS approach is equal to the 50%: it is the maximum value achievable using the median.

If on the one hand it permits to achieve an high percentage of outlier rejection, on the other hand its implementation has a very low efficiency. Rousseeuw estimate that also in a good implementation, the time required for the LMedS is an order of magnitude higher than the same interpolation in the TLS. In addition its implementation is not simple. Because of the median must be evaluated on the residuals computed on a model without inlier, the LMedS algorithm requires multiple model evaluations on a set of sub-samples selected from all the measurements $\mathbf{y}$. The fact that it requires multiple evaluation on small samples is to ensure that at least one of this samples does not contain outliers. Lets define the following variables:

- the vector $\mathbf{x}$ of the n parameters of the model.
- the vector $\mathbf{y}$ of the m model observations. To ensure a unique solution it must be respected the constraint $m \geq n$.
- the vector of residuals $\mathbf{r}_{\mathbf{x},\mathbf{y}}$ computed on the model using the parameters $\mathbf{x}$ and the observations $\mathbf{y}$.
- the number minimum of observations s to estimate a vector of parameters $\mathbf{x}$.
- the number minimum of samples N to ensure that at least one sample is composed only by inlier.

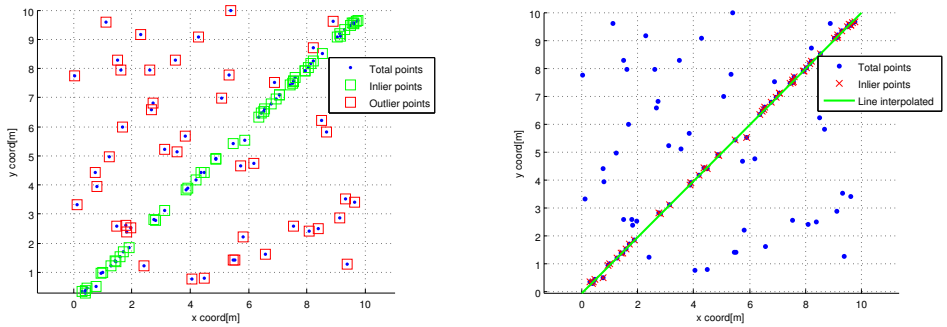
The process can be represent schematically in five steps:

1. Random selection of N samples $\Gamma_k, k = 1...N$, each one of size equal to the minimum number of observation s.
2. Estimation of the solution $\mathbf{x}_k$ on the samples $\Gamma_k, k = 1...N$.
3. For each solution $\mathbf{x}_k$ must be evaluated the medians $m_k, k = 1...N$ of the squared residuals $\mathbf{r}_{\mathbf{x}_k,\mathbf{y}}$ computed for all the observations $\mathbf{y}$:
$$m_k = median\left(\mathbf{r}^2_{\mathbf{x}_k,\mathbf{y}}\right). \tag{2.18}$$
The medians of all the samples are stored in the vector $\mathbf{M}$.
4. Selection of the least median square from the vector $\mathbf{M} = [m_1...m_N]^\top$ and the respective set of parameters $\mathbf{x}_k$.
5. Identification of the outliers based on the residuals $\mathbf{r}_{\mathbf{x}_k,\mathbf{y}}$ of the sample with the least median square and evaluation of the solution by using only the inlier observations.

The number N of samples must ensure the presence at least one of the samples is without outlier. The evaluation of this number requires an initial knowledge of the outlier percentage in the observations. With this assumption, the probability P to create randomly a sample without outliers is:
$$P = 1 - (1 - \omega^s)^N, \tag{2.19}$$
where ω is the fraction of outlier. From this equation is easy to compute the number of samples N:
$$N = \frac{log(1-P)}{log(1-\omega^s)}. \tag{2.20}$$
It must be specified that the fraction of outlier ω used in this formulation is a guess value, while for the probability is usually used $P = 0.99$.

Recalling the last step of the method, actually the solution $\mathbf{x}_k$ corresponding to the least median squares is evaluated on the minimum number of observations. In order to achieve a more reliable result, it is possible to identify the inlier observations and so refine the solution (Figure 2.11). First of all, to identify the outliers it is required to compute a threshold that must be applied to the residual of the samples computed in the least median square set:
$$\hat{\sigma} = 1.4826\alpha\left(1 + \frac{5}{n-s}\right)\sqrt{min(\mathbf{M})}, \tag{2.21}$$
where α is a parameter that can be adjusted to select correctly the outliers: in most of the application can be used $\alpha = 2.5$.

A very similar approach to the LMedS is the *Random Samples Consensus (RANSAC)*. The main difference is in the threshold definition: while in the LMedS the threshold $\hat{\sigma}$ is based on a statistical evaluation of the residuals, in the RANSAC approach, it is a fix parameter selected by the user.



(a) Outliers detection (b) Model interpolation

Figure 2.11: LMedS results with high outlier percentage ($\omega = 0.5$) on a line interpolation example.

2.2.3 Statistical approach to data association

The LMedS procedure can be applied also in the data association step: the model to be interpolated is the vehicle pose variation between two environment observations. In this case the outlier identification is performed on the correspondences between the two set of data processed from the information acquired by the instrumentation in the two vehicle pose. In order to evaluate these correspondences, it is possible to apply the LMedS algorithm to the ICP framework: in this way the residuals of the model interpolation are the distances of the matched point in the two set.

Moreover the LMedS algorithm is based on the assumption of random variables with a Gaussian probability distribution, since only in this way the outliers can be identified and removed. This assumption when applied to the data association requires that the data used in the association must be represented in this terms. Rather, the association residuals must have a Gaussian distribution which implies that also the observations of the environment should be represented under these conditions.

To ensure a robust data association and so a reliable solution, often the euclidean distance between points is not enough. Actually, when is possible to associate to these points also information on uncertainties, there are algorithms that can weigh the correspondences by taking care also of the uncertainty. The most known algorithm used for this purpose is the *Mahalanobis Distance* [44]: to be applied it requires that the point uncertainty must be described with by a multivariate Gaussian distribution. By defining two points $\mathbf{q}^1 = [q_1^1 ... q_p^1]^\top$ and $\mathbf{q}^2 = [q_1^2 ... q_p^2]^\top$, the Mahalanobis distance is defined as:

$$d(\mathbf{q}^1, \mathbf{q}^2) = \sqrt{(\mathbf{q}^1 - \mathbf{q}^2)^\top \mathbf{C}^{-1} (\mathbf{q}^1 - \mathbf{q}^2)} \tag{2.22}$$

where $\mathbf{C}$ is the sum of the covariance matrix $\mathbf{C}_{\mathbf{q}^1}$ and $\mathbf{C}_{\mathbf{q}^2}$ respectively of the two points $\mathbf{q}^1$ and $\mathbf{q}^2$. An important feature of this algorithm is that the

distance evaluated is the confidence level [45] between the distributions of the two points. Hence it allows to evaluate the probabilistic distances: when applied to a set of points, it is possible to give more weight to the points with less uncertainty. Because of the Mahalanobis distance evaluation can be used also in the LMedS method, in which the residuals must be evaluated several times, the time required by its evaluation can be important. In Montiel work [46] is possible to find useful consideration on Mahalanobis distance algorithm simplification.

In addition, also the minimization of residuals in the data association procedure can be performed by taking care of this probabilistic distance. Avoiding numerical solution, closed form solutions can be achieved by using *Singular Value Decomposition (SVD)* techniques. In Irani work [47] is reported a generalization of the SVD factorization with uncertainty. Without resorting to the SVD, Censi in [22] has developed a closed form solution that can also take care of the uncertainties.

2.2.4 Joint Compatibility test

In Section 2.2.1, with reference to the data association methods, it has been introduced the ICP, because it is the most known and used method. Though the ICP performances in terms of reliability can be improved with robust implementations, it still have an important limitation: it is based on a *nearest neighbour approach*. This means that its performances depend mainly from the initial estimation of the data association solution, no matter as robust is the algorithm used to define the correspondences.

Neira, in his work [48], performed an accurate analysis of the nearest neighbour approach, by illustrating its limitations. The limiting characteristic is on the association predictions: by using nearest neighbour approach the prediction is evaluated independently for each point. The drawback of this kind of prediction is on the estimated uncertainty that is extremely high if compared with the one that can be achieved using the correlation between all the predictions. Using a statistical approach in the data association, a prediction with high uncertainty implies the possibility that more correspondences can be within the confidence level. This means that there is an high probability to select wrong associations.

By correlating, instead, all the predictions is possible to obtain a joint probability with a lower uncertainty: for this reason the new method proposed is called *Joint Probability Test*. In order to evaluate the joint probability by avoiding individual correspondences, it is required an algorithm capable to analyse all the space of possible correspondences in search of the hypothesis with the largest number of probabilistic associations. The *Branch and Bound* algorithm permits to perform this search in an optimal way.

2.3 Incremental and Global Localization

In this section is introduced the last main topic of the thesis, the localization. Localization can be considered as the final step to perform after the elaboration of the information acquired from the environment and the data association. This does not mean that the localization is the most important step, because if the previous operations are not performed in the correct way, the localization fails.

One of the key points to obtain a full autonomous mobile robot is the navigation. As described by Siegwart [49] this process requires four important building blocks:

perception how the robot use its sensors to obtain useful data from the surrounding.

localization how the robot solve its position and orientation in the environment.

cognition how the robot decide the action to achieve its goals.

motion control how the robot plan and control the motion to follow a desired trajectory.

The most studied and still open problem of these four blocks is the *localization*. Several approaches has been proposed to localize vehicles, starting by the use of encoder on wheels until the localization based on monocamera.

Localization is the operation used to estimate the position and orientation of an agent in the reference frame of the environment. Position and orientation in general can be defined in a three-dimensional space, with six degrees of freedom. Obviously to estimate the three-dimensional position and orientation is also needed instrumentation capable to collect three-dimensional data information, like 3D laser scanner, stereo-camera or mono-camera (with some expedients). Though the vehicle localization in a bi-dimensional space may seem easier than the three-dimensional one, it is still an open problem. In this thesis most of the work has been applied to bi-dimensional laser scanners, but the methods proposed can also be applied to other range finder instrumentation and cameras.

In Section 2.1, the indoor environment has been identified as the target scenario. This is not a limitation, indeed most of the existent automated vehicles use classical techniques with global sensors (laser triangulation, wire or magnetic guidance) that requires the human manipulation of the environment. Current works in literature are focused in the study of methods that can work in this environment scenario without theses constraints. From this point of view, as seen in Section 2.1, a perfect sensor has to fit completely on the vehicle, and must return a measurement of the environment without

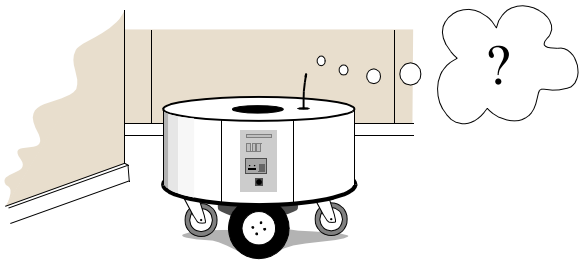


Figure 2.12: The localization problem: "Where am I"? [49].

any aid from other sensors. Opto-electronic instruments are closed to this definition, but still have to deal with some issues:

- Environment discretization: the sensor returns a environment representation that depends on its pose. The information in different acquisitions does not have explicit connection: for instance, the beams of a laser scanner can acquire a set of points on a wall. By changing the laser pose, the same wall is represented with different points.

- Measurement noise: the uncertainty of measure in range finders is due to the sensor noise but also from the reflective surface (determined by material, incidence angle, etc.).

However, by means of this instrumentation, especially with laser scanners, it is possible to reconstruct the robot pose in most of the cases better than with the odometric approach. Obviously the precision and accuracy on pose estimation depend on method used to process data.

The most important manner to classify the vehicle localization is on the method used to reference the robot position in the environment coordinate frame. The simplest approach is the matching of sequential observation in order to reconstruct the robot pose by concatenating all the transformations. This is defined *incremental localization*. Instead, if each new measurement is obtained by matching the observations acquired in the vehicle coordinate frame with information stated in the environment coordinate frame, the operation is defined *global localization*. Techniques that use maps, or approaches based on SLAM are classified as global localization. The main advantage of the global localization is the reduction of the drift in the pose estimation error. Ideally, in this case the error on the pose evaluation is due only by the error in the matching approach of the local observation with the global information. Differently, the main problem of the incremental localization is the error drift in the pose estimation. Indeed, because of the information matching is performed incrementally, the error on the pose evaluation is not only due to the last matching, but is also connected to the errors in the previous vehicle pose estimation.

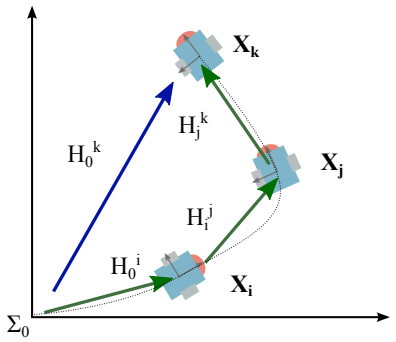


Figure 2.13: Incremental (concatenation of green vectors) and global localization (blue vector) comparison.

In order to care of these differences and to develop reliable methods, it is indispensable to evaluate the uncertainty of the localization. In this thesis the uncertainty of the vehicle state is described by a *multivariate normal distribution*, since it is a good approximation of the real noise, and in addition it allows a parametric representation with a covariance matrix.

Defining $\mathbf{X} = [x, y, \theta]^\top$ the vehicle pose, the correspondent covariance matrix is described by:

$$\mathbf{C} = \begin{bmatrix} c_{xx} & c_{xy} & c_{x\theta} \\ c_{xy} & c_{yy} & c_{y\theta} \\ c_{x\theta} & c_{y\theta} & c_{\theta\theta} \end{bmatrix}. \tag{2.23}$$

2.3.1 Incremental Localization

When there is not a-priori knowledge of the environment in which the vehicle is moving, the incremental localization is the simplest method to implement. It is based on the concatenation of incremental motion estimations: in few words is a dead reckoning method. The development of this method requires 'only' to develop a method to process the environment information and to associate these data. On the other hand it has a limitation in the drift of the pose estimation error, no matter how accurate or reliable are the algorithm. For this reason it is not used as a standalone localization method. However the incremental localization can return good results on short movement and for this reason it is useful to compute an estimation of the vehicle poses that can be uses as an initial guess for a global localization method.

Because of these problems, few works have discussed its scientific characterization: in [50] are reported some useful information on this topic. From a probabilistic point of view, it must be considered the uncertainty propagation in order to take care of the error drift. To introduce this concepts, let $\mathbf{X}_1 = [x_1, y_1, \theta_1]^\top$ and $\mathbf{X}_2 = [x_2, y_2, \theta_2]^\top$ be two vehicle poses in which are acquired the environment information, as depicted in Figure 2.14. In

terms of incremental localization, the pose $\mathbf{X}_1$ is the previous pose known, and $\mathbf{X}_2$ is the current vehicle pose to estimate. By supposing to be able

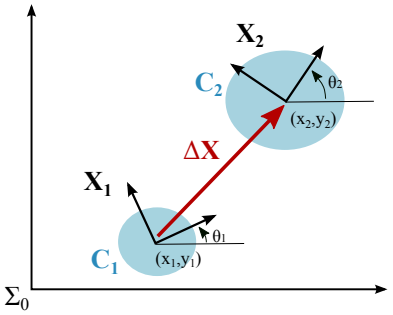


Figure 2.14: Incremental localization procedure and uncertainty propagation from pose $\mathbf{X}_1$ to $\mathbf{X}_2$.

to process the information from the environment and to perform the data association, it is possible to estimate the motion between the two poses. The result is a vector that describes the variations of the pose, denote with $\Delta\mathbf{X} = [\Delta x, \Delta y, \Delta\theta]^\top$. In this way the current vehicle pose $\mathbf{X}_2$ can be estimated by updating the pose $\mathbf{X}_1$ with the measure $\Delta\mathbf{X}$:

$$\mathbf{X}_2 = \begin{bmatrix} x_2 \\ y_2 \\ \theta_2 \end{bmatrix} = \begin{bmatrix} x_1 + \Delta x \cdot cos(\theta_1) - \Delta y \cdot sin(\theta_1) \\ y_1 + \Delta x \cdot sin(\theta_1) + \Delta y \cdot cos(\theta_1) \\ \theta_1 + \Delta\theta \end{bmatrix}. \tag{2.24}$$

In order to take care of the error propagation, the measurements uncertainties must be combined. The general solution to this problem is to uses the *propagation of uncertainty* method:

$$\mathbf{C}_{out} = \mathbf{J}\mathbf{C}_{in}\mathbf{J}^\top, \tag{2.25}$$

where $\mathbf{C}_{out}$ is the output covariance matrix, $\mathbf{C}_{in}$ is the covariance matrix associated to the inputs and $\mathbf{J}$ is the Jacobian of the system of equation that describe the relation between input and output.

In the incremental localization, the covariance matrix to estimate is $\mathbf{C}_2$, the one associated to the vehicle pose $\mathbf{X}_2$. The system of equation on which compute the Jacobian is the one reported in Equation 2.24. The input in this case are the pose variation $\Delta\mathbf{X}$ and also the previous pose $\mathbf{X}_1$. The Jacobian formulation is:

$$\mathbf{J} = \frac{\partial \mathbf{X}_2}{\partial \left[x_1, y_1, \theta_1, \Delta x, \Delta y, \Delta\theta\right]^\top}, \tag{2.26}$$

and by applying the derivative the explicit formulation becomes:

$$\mathbf{J} = \begin{bmatrix} 1 & 0 & -(\Delta y \cdot cos(\theta) + \Delta x \cdot sin(\theta)) & cos(\theta) & -sin(\theta) & 0 \\ 0 & 1 & \Delta x \cdot cos(\theta) - \Delta y \cdot sin(\theta)) & sin(\theta) & cos(\theta) & 0 \\ 0 & 0 & 1 & 0 & 0 & 1 \end{bmatrix}. \tag{2.27}$$

Also the associated input covariance matrix has to be composed in terms of previous pose uncertainty and the estimated variation. Since this two entities can be considered uncorellated, the input covariance matrix can be defined as:

$$\mathbf{C}_{in} = \begin{bmatrix} \mathbf{C}_1 & \mathbf{0}_{3\times3} \\ \mathbf{0}_{3\times3} & \mathbf{C}_{\Delta\mathbf{X}} \end{bmatrix}. \tag{2.28}$$

Thank to this formulation, the main limitation of the incremental localization becomes more evident: the covariance $\mathbf{C}_2$ of each new pose $\mathbf{X}_2$ concatenated to the previous vehicle pose $\mathbf{X}_1$ will increase by accumulating the uncertainty of the pose variation $\mathbf{C}_{\Delta\mathbf{X}}$.

2.3.2 Global Localization

The global localization has been developed in order to avoid the main problem of the incremental estimation: the error propagation. To bypass the error accumulation the only option is to match the information acquired in each new vehicle pose directly with the information defined in the global coordinate frame of the environment. In this way, the uncertainty of each pose is due only to the matching process and is completely independent from the previous vehicle pose estimation. In addition is possible also to deal with the *kidnapped robot* problem.

As evidence of these considerations, there is the fact that almost all the localization systems widely used are based on global localization. Important examples are the approaches based on laser triangulation (used in industrial applications) and the GPS (used for outdoor localization). As working principle they are the same method: both they can be though of as the application respectively for indoor and outdoor environment. Unfortunately they have also important limitations. First of all they require a network of additional sensors widely distributed in the navigation environment in well known positions. In addition, by using this sensors is usually only possible to localize the robot. Information on surrounding environment can not be achieved and operations like obstacle avoidance can not be performed. It is clear that this methods require heavy constraints on environment, and they are suitable for structured scenario.

With the evolution of methods the constraints required on navigation environment became few. In particular, with the introduction of new instrumentations able to recognize the environment (like sonar, laser and camera) and so obtain local information, a new technique was introduced: the *Simultaneous Localization and Mapping (SLAM)*. Early works were done by Chatila and Smith [51, 52]: while the former introduced the problem of the consistent world modelling, the latter established the basis for the description of the uncertainties of the vehicle pose and map.

An intermediate step between the methods with distributed global sensors and SLAM was the navigation in a-priori defined environment map.

The robot can compare its measurements with the map and so estimate its pose in the global coordinate frame defined by the map without the drift in pose estimation error. This means that a priori knowledge of the map is needed: the robot can navigate only in a well known environment. Furthermore the map represent the environment state in a precise time; this is not necessarily consistent with the environment state at the time the map is used.

These are relevant limitations and the SLAM technique has been propose to avoid them. SLAM, as its acronym says, does not require a-priori knowledge of the environment because of the map is built up while at the same time that the robot localize itself. Since they are simultaneous processes, the localization can also use the current map estimation to perform a more accurate evaluation of the robot position.

Obviously the problem solution with SLAM has an higher complexity, but it can be applied widely. It is well suitable for example in environment exploration or navigation in non static and well structured scenario. Like all the algorithms, SLAM too is in a continuous evolution: nowadays the mapping part of the procedure is often skipped and replaced with similar approaches really closed to it. In fact map can change in time and it requires a lot of memory to store information of large environment.

In this work the attention is focused in an alternative procedure to avoid the mapping task, though maintaining the benefit off the global localization.

2.3.3 Mapping, Filtering and Keyframe-based optimization

The acronym SLAM describes a generic problem which solutions are the vehicle localization and the map definition. Only in its early releases it described a method to build a probabilistic map by using a Bayesian fusion off whole the information [52]. Nowadays it defines a research field constantly evolving and several probabilistic approaches has been proposed besides the Bayesian fusion.

In a recent work [53] Durrant-Whyte gives a survey of all the methods proposed until 2005. This date is really important since up to this year all the scientific works were concentrated in the implementations of the SLAM using probabilistic approaches based on the *Extended Kalman Filter (EKF)*. Relevant works in this field were made from Duckett in [54, 55], where he already studied methods to obtain fast map learning, and Konolige in [56, 57]. While in the former he studied the SLAM in large scale map, in the latter he dealt with the reduction of map information to increase the performance of the approach. More focused on the *Visual SLAM*, Davison in [26] proposed also a EKF approach to estimate the camera pose.

In 2005 Thrun with the scientific work [58] and the book [59] introduced a new probabilistic approach to the SLAM, the *GraphSLAM*. The novelty of the SLAM implemented in this new approach is that it does not use a EKF

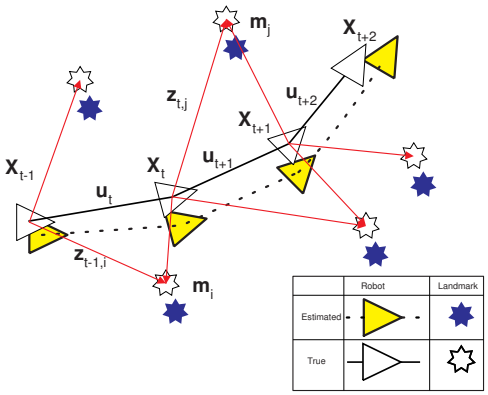


Figure 2.15: SLAM problem graphic representation [53]: the method estimate simultaneaously the robot pose $\mathbf{X}$ and the map $\mathbf{y}$ landmark location $\mathbf{m}$ using information on motion $\mathbf{u}$ and environment measurements $\mathbf{z}$.

to *filter* the information. The GraphSLAM also change the probabilistic description of the SLAM problem. From a probabilistic point of view there are two main forms of SLAM problem that are classified by taking account the temporal reference in which it is performed [59].

The former is known as *online SLAM problem* and it estimates the posterior using the current pose and map. Defining with $\mathbf{X}_t$ the pose at the time t, with $\mathbf{y}$ the map, $\mathbf{z}_{1:t}$ the measurements and $\mathbf{u}_{1:t}$ the controls the posterior is estimate by using only the information on the current pose:

$$P\left(\mathbf{X}_t, \mathbf{y} | \mathbf{z}_{1:t}, \mathbf{u}_{1:t}\right). \tag{2.29}$$

The algorithms for the online SLAM can work well in incremental: the past poses are already discarded and also the past controls and measurements can be discarded after the posterior evaluation.

The latter SLAM approach is defined *full SLAM problem* and it uses instead all the past information and all the path $\mathbf{X}_{1:t}$ to estimate current posterior:

$$P\left(\mathbf{X}_{1:t}, \mathbf{y} | \mathbf{z}_{1:t}, \mathbf{u}_{1:t}\right). \tag{2.30}$$

It could seems a slight difference but it has important implications: the most important one is the structure of the dependency and the connections between all the vehicle poses. However they are both important in the estimation of the posterior that is the gold standard in SLAM. From a more practical point of view, this classification remarks the division from classical approaches based on iterative Bayesian fusion and a new kind of approaches that use whole the past information instead filtering.

The most important example of online SLAM algorithm is the *EKF-SLAM*, that applies the Extended Kalmann Filter to filter the vehicle poses using a maximum likelihood data association. To implement this approach

are required some simplification; first of all is *feature based*. As consequence it requires a good features detector and descriptor in order to perform a correct data association in every situation: it can not work properly in scenario with ambiguous features. Another important limitation is that it can use only positive information on feature. The miss-detection of a features could not be registered with the EKF approach.

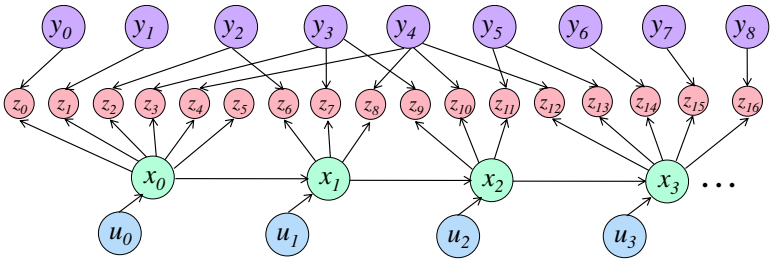


Figure 2.16: Localization and mapping with EKF approach. Using control information $\mathbf{u}$ and measurement $\mathbf{z}$ it allows to reconstruct the vehicle poses $\mathbf{X}$ and build the map $\mathbf{y}$.

As said previously, full SLAM problem methods can be defined as a new type of approaches. The main reason that has led to their definitions is to avoid the limitations of filter based methods. The filter operation of the EKF-SLAM must be replaced with methods capable to infer the posterior using all the previous information. Approaches widely used for this purpose are the *Graph Theory* and the *Bundle Adjustment*. Although the Bundle Adjustment [60, 61] is more used in Computer Vision thanks to their capabilities to work in non-linear problem solution, it can be used also in SLAM. A significant drawback of the Bundle Adjustment is due to the numerical optimization procedure that must be used.

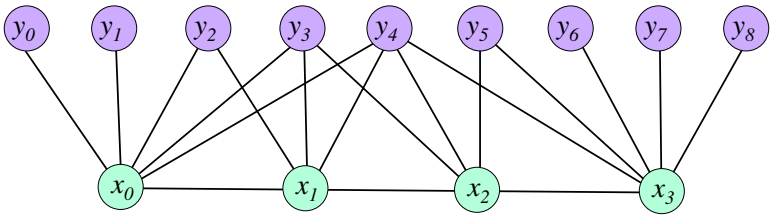


Figure 2.17: Localization and mapping in graph based SLAM approach. The vehicle poses $\mathbf{X}$ are reconstructed directly from the measurements $\mathbf{z}$ of the landmark in the environment $\mathbf{y}$, without filtering operation.

In order to avoid the numerical optimization it possible to use methods based on the Graph Theory. Graph Theory allows to solve full SLAM problems easily then Bundle Adjustment in closed form solution. It is so important that a new branch of SLAM has been defined: the *GraphSLAM*.

Early works on graph based SLAM was presented in the late 90s by Lu [62] and Gutmann [63], but important results has been achieved only few years ago, when it was understood the importance of the full SLAM. As matter of fact the posterior of full SLAM naturally forms a sparse graph. Furthermore without explicit use of minimization resolver, is possible to achieve to least square problem solution.

Thus classical SLAM like the EKF implementation and graph based SLAM can be seen as the extremes of the current SLAM knowledge. The choice of which one is better depends on the application, but currently cutting-edge applications use the graph based SLAM. This is due to some advantages, not limited to computation and memory storage performance. The first advantage is the map: with some graph based approaches the map can be update while in other it is possible to skip the map generation process. In fact in classical SLAM methods the map is the way in which is represented the most updated information, while in graph based ones the same information is stored in a graph with soft constraints. Further advantages are given on precision because of the EKF-SLAM filter all the previous informations in a map: in a given time t all the measurements are inferred and stored in the map. New measurements can not access to previous data, because they are "hidden" by the filter operation (Figure 2.18).

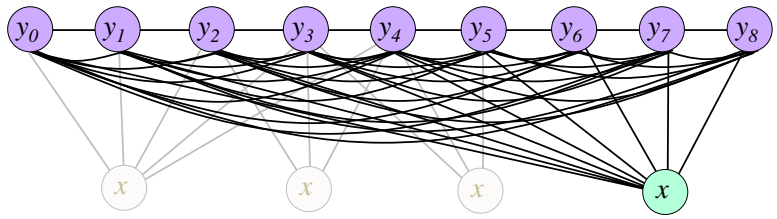


Figure 2.18: Filtering operation in online SLAM, all the information are summarized with respect to the last pose state vector $\mathbf{X}$.

On the other hand the graph based approach can use all the data collected along the path and it can also correct the previous information in terms of vehicle pose and map with new measurements. This means that by this way it is possible to have a better estimation of both map and vehicle poses since all the achievable information are used. With these premises it could be said that graph-based method can produce maps with higher accuracy with respect to the map computed by classical EKF-SLAM approaches.

Obviously also graph-based slam has limitations: actually the main problem is that the size of graph grows with time. This drawback is probably the most important, and actually the research is aiming to solve it.

A relevant work on the differences of this approaches has been presented by Strasdat in [64]. The article perform a rigorous analysis of relative advantages of online and full SLAM. The methods are applied to the monocular

SLAM problem, with a series of experiments in simulation and real applications. The most important result achieved is the performance comparison between two important aspects in the visual SLAM: the number of features and the number of frames used in the camera motion estimation. In all the cases studied the results proved that it is more important to have an high number of features instead of an high number of frames. The work main conclusion is that filter based SLAM algorithms might be convenient only when small processing budget is available, but the bundle adjustment optimization, also on key-frame, is superior elsewhere.

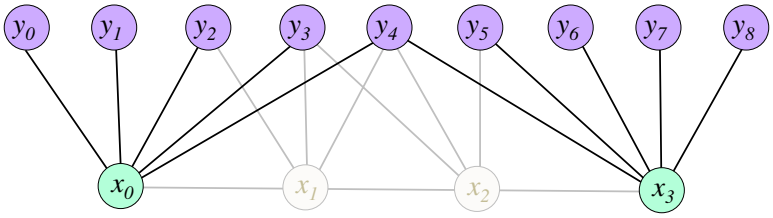


Figure 2.19: Graph optimization using key-frame: by using only the most representative vehicle poses the computational cost of the graph solution can be decreased.

By applying these considerations to the work here presented it is possible to do some consideration. First of all it must be noted that nowadays the processing budget is not a limitation, so the bundle adjustment optimization or graph based framework can have high performances. Furthermore, the comparison between number of features and frames is clearly in favour of the full SLAM approaches. Online SLAM requires highly distinctive features in order to perform best matching operations and to update the map as better as possible. In the generic scenario of navigation hypothesized in this thesis it can be difficult to find an elevate number of unique features stable on several vehicle poses. Hence the use of the online SLAM is not the best choice since it works with few features and many observation. On the other hand, as it will be shown in Section 2.3.4, the implementation of the full SLAM with graphs allows to perform the global optimization using only the most relevant vehicle poses (Figure 2.19) without affecting the performance in terms of accuracy and reliability of the localization. Furthermore, by selecting the right algorithm it is also possible to avoid the mapping: thus there are no limitations on the type of information used to describe the environment.

2.3.4 Graph Theory

The graph theory may seem a new method since most of the cutting-edge methods on SLAM use it: on the contrary it is a very old concept. It was introduced by Euler and the original idea is dated 1736: few publications

were produced after its discovery. Only thanks to the growth of the computer science there was the need to formalize the theory: an early work is the Coxeter book [65]. In [66] is also possible to find a more recent discussion that take in consideration the innovation introduced by the new approaches in the computer science. Graph theory has been used mainly to represent complex networks of communication and data organization, in several and different fields (communications, computer science, chemistry, human relations, etc.). Here are reported only the principal notions in order to make easy to explain how it can be used in the solution of the full SLAM problem.

A graph can be defined as a pair of sets $G = (\mathbf{V}, \mathbf{E})$, where $\mathbf{V} = [v_1, ...v_m]^\top$ is the vector of *vertexes* of the graph G and $\mathbf{E} = [e_1, ...e_n]^\top$ is the vector of the *edges*. In literature vertexes are also called *nodes* or *points* and edges are called *lines* or *links*. Actually the graph is defined by the connections of the vertexes with the edges, as depicted in Figure 2.20.

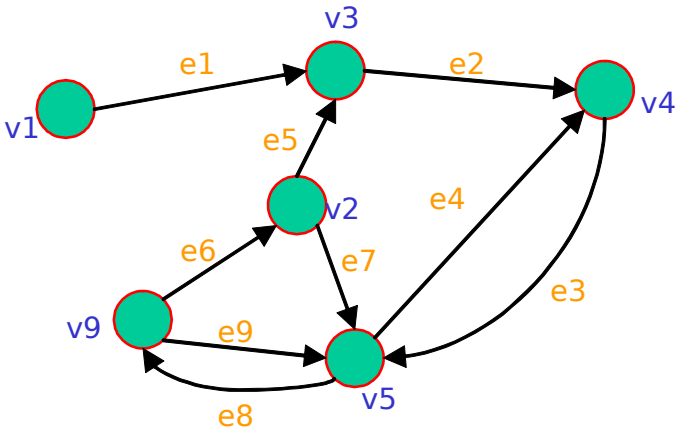


Figure 2.20: Representation of a directed graph.

Graphs can be generalized allowing loops, or rather a link on the same vertex, and also with multiple parallel links on the same vertexes as reported in Figure 2.20. On the side of connections, if it is possible to assign an information on the direction of a link, the graph can be defined as *directed graph*. In the previous figure, the links represented in the graph are of this kind. In this way edges have a direction, that is needful to solve the graph when applied to the full SLAM problem. Actually, in order to solve the full SLAM problem, the graphs considered are without loops and only with directed links.

Last important theoretic concept of the Graph Theory are the *walks*. A walk can be defined as a sequence of vertexes and edges, with an initial and a final node, and all the intermediate nodes. Each link can be covered more than one time: it is possible to define that:

- a walk is **closed** if initial and final vertex are the same one.

- a walk is a **path** if each vertex is covered only one times.
- a walk is a **cycle** when it can be defined closed and it is a path except for the first and last nodes.

This definition seems to be not useful, but when in the following sections it will be introduced the similarity between the graphs and the data fusion these notations will become important.

2.3.4.1 Graph Theory as linear system solver

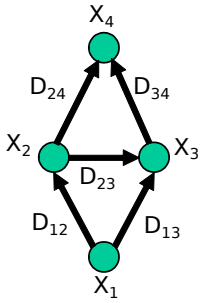
The extension of the full SLAM problem representation of the vehicle pose estimation with the Graph Theory is really simple. Indeed, how expressed also in Section 2.3.3 the posterior of the full SLAM naturally forms a sparse graph. In this case each node is a vehicle pose and the directed edges between vertexes are the estimation of the transformations between these poses. In order to describe better the problem, since the nodes are representing the vehicle poses, they can be defined $\mathbf{X}$ and the links, that represent the pose variations between the poses can be denoted with $\mathbf{D}$.

To localize the vehicle the Graph Theory must be used to solve the nodes $\mathbf{X}$ giving a set of pose transformation estimation $\mathbf{D}$. If it is possible to draw only a *path* the graph solution converge to an incremental estimation: each node is connected only with the previous and following one. In this case is possible to demonstrate that the number of equations created by the graph is equal to the number of variables and so it exists a single unique solution.

Otherwise, if a generic node of the graph has more than two links, it is possible to have more than one observation to reconstruct the pose associate to the node. An example is reported in Figure 2.21(a), in which is depicted a situation with 4 different vehicle poses $\mathbf{X} = \{\mathbf{X_1}, \mathbf{X_2}, \mathbf{X_3}, \mathbf{X_4}\}$. At each one of these poses is associated the data processed from the environment observation: between all these measurements is possible to perform a data association operation and so the estimation of the five pose transformation reported in Figure 2.21(a). The evaluation of the $\mathbf{X}$ poses is not trivial, because of the number of equations in this case is higher than the number of variables. Taking for example the pose $\mathbf{X}_3$, starting from the first pose $\mathbf{X}_1$, it can be estimated several walks. At least 3 path can be drawn:

- by a direct estimation from $\mathbf{X}_1$ to $\mathbf{X}_3$ with $\mathbf{D}_{13}$.
- by concatenating $\mathbf{D}_{12}$ and $\mathbf{D}_{23}$.
- by concatenating $\mathbf{D}_{12}$, $\mathbf{D}_{24}$ and $(\mathbf{D}_{34})^{-1}$.

This means that the system to solve is overdetermined: the gold standard in this case is the least square optimization. Further it will be demonstrated that the solution of this system of equations performed by solving the graph



(a) Graph representation

Pose \ Link	12	13	23	24	34
1	-1	-1	0	0	0
2	1	0	-1	-1	0
3	0	1	1	0	-1
4	0	0	0	1	1

(b) Graph incidence matrix

Figure 2.21: Graphical representation of a directed graph and its incidence matrix

is exactly the same of the least square minimization. This minimization include all the others poses: so the graph theory allows to solve the vehicle poses taking account of all the possible links, like a filter.

The simplest way to solve a system with a least square minimization is to use the *pseudo-inverse method*. For this reason it is required a matrix representation of the graph and system. This representation is possible thanks to the *Incidence Matrix* $\mathbf{M}$ in which are stored the connections and their directions. It is composed of $\{0, -1, 1\}$ values and has $m \times n$ dimension, where m is the number of nodes and n is the number of all the possible estimations of pose transformation. For a graph with the characteristics previously reported, without loops and with directed links, this number can be evaluated with the follow relation:

$$n = \frac{m(m-1)}{2}. \tag{2.31}$$

The values in the incidence matrix $\mathbf{M}$ represent the positive or negative link directions or the absence of a link when the value is 0. For a simple graph in which there are 2 nodes u and v, with a link uv directed from u to v, the incidence Matrix $\mathbf{M}$ has a -1 to the u row and uv column, a $+1$ to the v row and uv column and 0 in all the other components. In Figure 2.21 is reported a more elaborate example and it is represented also the incidence matrix for that case.

The definition of the incidence matrix allows to write the general formulation of the graph theory:

$$\hat{\mathbf{D}} = \mathbf{M} \cdot \hat{\mathbf{X}}, \tag{2.32}$$

where $\hat{\mathbf{X}}$ is the $md-$dimensional vector which is the concatenation of all the poses $\mathbf{X}_1, \mathbf{X}_2, ..., \mathbf{X}_m$ each one of size d: for a planar motion the pose has three dof and so d is equal to 3. The term $\hat{\mathbf{D}}$ is a $nd-$dimensional vector obtained by the concatenation of all the link measurement $\mathbf{D}_{ij}$ where i and j are the indexes of the two nodes linked.

The most interesting case is when to the measure $\mathbf{D}_{ij}$ is associated an uncertainty. By assuming that this uncertainty can be described with a Gaussian distribution it can be denoted with the covariance matrix $\mathbf{C}_{ij}$. The least square criterion for optimal estimation is to minimize the following sum of squared Mahalanobis distances:

$$W = \sum_{0 \leq i < j \leq m} (\mathbf{X}_i - \mathbf{X}_j - \mathbf{D}_{ij})^\top \mathbf{C}_{ij}^{-1} (\mathbf{X}_i - \mathbf{X}_j - \mathbf{D}_{ij}). \tag{2.33}$$

In this formulation it must be remarked that the relations between the poses are considered linear. Actually the relations between two poses are not linear: the solution to this problem are reported in Section 2.3.4.3. The Equation 2.33 can also be represented in matrix form:

$$W = (\hat{\mathbf{D}} - \mathbf{M}\hat{\mathbf{X}})^\top {\mathbf{C}_{\hat{\mathbf{D}}}}^{-1} (\hat{\mathbf{D}} - \mathbf{M}\hat{\mathbf{X}}), \tag{2.34}$$

where the term $\mathbf{C}_{\hat{\mathbf{D}}}$ is the covariance matrix associated to $\hat{\mathbf{D}}$: since all the measurements are independent, it is a $nd \times nd$ diagonal block matrix. The solution of $\hat{\mathbf{X}}$ can be computed using the weighted pseudo-inverse method:

$$\hat{\mathbf{X}} = (\mathbf{M}^\top {\mathbf{C}_{\hat{\mathbf{D}}}}^{-1} \mathbf{M})^{-1} \mathbf{M}^\top {\mathbf{C}_{\hat{\mathbf{D}}}}^{-1} \hat{\mathbf{D}}. \tag{2.35}$$

From the same weighted pseudo-inverse algorithm is possible also to compute the covariance of the the concatenated pose vector $\hat{\mathbf{X}}$ is:

$$\hat{\mathbf{C}} = \left(\mathbf{M}^\top {\mathbf{C}_{\hat{\mathbf{D}}}}^{-1} \mathbf{M} \right)^{-1}. \tag{2.36}$$

By analysing the previous relations some important observations can be done. First of all, the graph theory minimization criterion has a big drawback: it can deal only with *linear relation* between nodes, and so it is able to solve only system of linear equations. This aspect is really restrictive and it forces to use some tricks like linearisation of the equation as reported in [62].

Another important observation is on the covariance matrix $\mathbf{C}_{\hat{\mathbf{D}}}$: it has $nd \times nd$ dimensions, but assuming that the measurements are uncorrelated it becomes a diagonal block matrix, so the solution can be simplified [62]. In addition, in the graph based SLAM the matrix $\mathbf{G} = \mathbf{M}^\top {\mathbf{C}_{\hat{\mathbf{D}}}}^{-1} \mathbf{M}$ is called *information matrix*, and is really closed to the one reported in [58] except for the information of the map.

2.3.4.2 Fusing information with Graph Theory

As demonstrated, Graph Theory can solve optimisation problems also by weighting the links with the uncertainty of the correspondent measurements. In addition, the solution is also minimized in order to obtain the best estimation of the posterior. Indeed, in simple cases the graph theory solve problem by using sensor fusion based on Bayes Theorem.

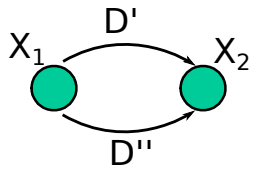


Figure 2.22: Graph with parallel connections.

Let $\mathbf{X}_1$ and $\mathbf{X}_2$ be two vehicle poses with two different estimation of pose variation $\mathbf{D}'$ and $\mathbf{D}''$. In graph theory the problem can be represented using a network with two parallel links (Figure 2.22) for the measurements. This is also a typical situation in which apply sensor fusion: defining with $\mathbf{C}'$ and $\mathbf{C}''$ the covariance matrix associated to the links and applying the graph solution reported in Equation 2.35 it can be achieved these relations:

$$\begin{array}{c} \mathbf{X}_1 = (\mathbf{C}'^{-1} + \mathbf{C}''^{-1})^{-1}(\mathbf{C}'^{-1}\mathbf{D}' + \mathbf{C}''^{-1}\mathbf{D}'') \\ \mathbf{C} = (\mathbf{C}'^{-1} + \mathbf{C}''^{-1})^{-1} \end{array}. \tag{2.37}$$

The result is equivalent to the one found using the data fusion and reported in Smith and Cheeseman [52]. Smith and Cheeseman's algorithm has however an important limitation: it can be applied only to networks formed by serial and parallel connections. In very simple cases like the one reported in Figure 2.21 classical EKF-SLAM approaches that are based on Bayesian data fusion can not be applied, or rather can be applied only by filtering all the past poses. This simple example is very useful to understand the great advantage of the graph theory with respect to the classical filter: the new information can also modify the previous evaluations, and so the current estimation of the measure.

2.3.4.3 Solving bidimensional non linear systems

The main drawback of the approach based on Graph Theory is that it requires linear relations between nodes. In many cases this aspect could be a relevant problem: actually the relations used in the vehicle localization are non linear. By taking account only of the vehicle position, the problem can be solved with linear equations, but introducing also the vehicle attitude the problem becomes non linear since the attitude introduces trigonometric function. However the Graph Theory has such significant benefits that justify its implementation in SLAM problem.

Since the initial works, the main problem was the use of graphs in non linear cases. A seminal paper is the Lu and Milios one [62], they were the first to describe the SLAM as a set of links between poses, and to formulate a global optimization algorithm. In the same work it is also reported a possible solution to the Graph Theory limitation to linear problem with a linearisation of the pose relations in the neighbourhood of an initial guess

estimation of the vehicle poses. Many subsequent works were based on this paper. The Lu and Milios algorithm was successfully implemented by Gutmann [67], who reported numerical instability due to the extensive use of matrix inversion required by the linearisation process. An additional problem emphasized is that the initial estimation has to be really close to the real solution, otherwise the method converges to local minimum. This problem can be significant in large environment in which the vehicle perform a closed loop: there is an high probability that the scans used to close the loop can have high error displacement and so the loop closure may be performed in wrong way. Other relevant works are [68] and [54] in which they provided efficient techniques for solving such instability problems. The relation between covariances and the information matrix is discussed in Frese work [69].

The *GraphSLAM* proposed by Thrun [58] is based on the *Grapichal-SLAM* of Folkesson [70] and it is probably the most advanced implementation of the Graph Theory. It has currently a drawback: since its implementation is based on the identification of features in the environment it is possible that in some scenario this approach can not work properly.

An interesting work as been presented by Martinez [71]. From the theoretic point of view it assumes some limiting hypothesis and in addition it does not take account of uncertainties. Though the method is very simple if compared with the Lu and Milios one and with some improvement it could be a valid alternative.

2.3.5 Closed form solution and uncertainty estimation

The aim of the localization is to estimate, for a planar motion, the vehicle pose $\mathbf{X} = [x, y, \theta]^\top$ at each time. As seen in the previous sections, incremental and global localization differ on the method used to reference the measurements to the environment coordinate frame. However both the methods have to use the same basic procedure to compare the information extracted from the environment in order to perform the data association.

When the data processed is the one from the laser scanner, the procedure described above is defined as *scan matching* and it allows to identify the vehicle pose variation between two measurements, while using the camera the same approach is defined *image registration*.

Without loss in generality and independently from the instrumentation, it is required that the methods used to evaluate the pose variation must be formulated with a *closed form*. In other words, after the data association step the evaluation of the planar transformation has to be carried out without using numerical optimization. The motivations for this request are many: first of all it must be considered that by using numerical optimizations the solution can have local minimum and convergence problems. In addition there are reasons strictly connected to the time required. Numeri-

cal methods have not always the same number of iterations and usually are slower than a closed form solution. This aspects connected to the time are really important in the robust implementation with the view of a robust implementation of the ICP. This will be explained better in the Section 3.4: with the aim to ensure the right data association it is used a robust implementation of the ICP which requires multiple evaluation on different samples of the original set. The use of a numerical optimization inside this procedure is not feasible.

In addition it must be also recalled that in order to obtain a more reliable localization, no matter if by an incremental or by a global procedure, it must be take care also of the uncertainty of each association. Since the association are made on set of points the most used methods are based on the *Singular Value Decomposition (SVD)* that thanks to efficient implementations allows to solve this kind of problem. The SVD is used to carry out the solution by means of a *Principal Component Analysis (PCA)*: in the Irani work [47] is explained a method to use in the SVD factorization by taking care the uncertainty. Another general method to solve the pose variation by weighting the association with their uncertainty is reported in [22].

The covariance estimation is an important process to qualify every measurement; moreover it is indispensable in case of sensor fusion. In addition,

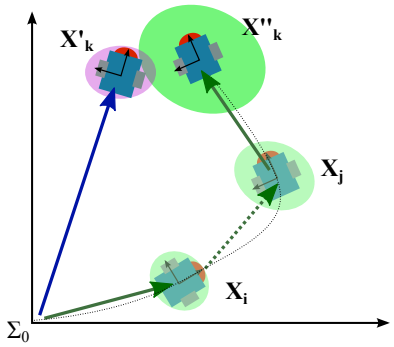


Figure 2.23: Representation of the vehicle pose uncertainties in the graph.

in the proposed graph based SLAM the link weighting with uncertainty is a truly important aspect: without considering it, all the possible walks to a node have the same importance. This means that a direct connection between two nodes has equal weight of any other walk passing through several nodes and end on the same node. A graphical interpretation of this aspect is reported in Figure 2.23. Let $\mathbf{X}'_k$ and $\mathbf{X}''_k$ be two estimations of the the vehicle pose $\mathbf{X}_k$ returned from two different graph paths. These estimations can be different not only in terms of position and attitude, but also in the uncertainty: since the $\mathbf{X}''_k$ has been evaluated by concatenating different measurements its uncertainty is higher. The optimal solution is to fuse the poses by taking care if their uncertainty. All these operations are included

in the graph solution with uncertainty weighting.

As reported in Censi paper [17], covariance estimation is not a trivial problem. This is true especially when there is not a direct correlation between the covariance of the inputs and the one of outputs. In such situation is difficult to propagate the uncertainty; an example is exactly the data association with ICP. The only relation between input and output is the error function to minimize, since the minimization procedure can be not easy to analyse (for example it can be a numerical optimization procedure). In these cases the existing methods to estimate the covariance are inaccurate or computational too expensive to be used in real-time.

The method proposed in [17] allows to achieve a good accuracy in covariance estimation, with a low computational cost. Differently from other methods that can be too much optimistic or can overestimate the covariance, the one presented is more accurate. It is based on the known method of uncertainty propagation that use a first order approximation. Let $\hat{\mathbf{x}}$ be the solution of a minimization algorithm A with a error function F that depends on the measurements $\hat{\mathbf{z}}$ such that $\hat{\mathbf{x}} = argmin_{\mathbf{x}} F(\hat{\mathbf{z}}, \mathbf{x})$. Then the covariance on the solution can be estimated using the uncertainty propagation:

$$cov(\hat{\mathbf{x}}) \simeq \frac{\partial A}{\partial \mathbf{z}} cov(\mathbf{z}) \frac{\partial A}{\partial \mathbf{z}}^{\top}. \tag{2.38}$$

Since A is a minimization algorithm, it does not have a close form to derive. Because of in the solution $\hat{\mathbf{x}}$ the error function gradient is null it is possible to obtain:

$$\frac{\partial A(\mathbf{z})}{\partial \mathbf{z}}\Big|_{\mathbf{z}=\hat{\mathbf{z}}} = -\left(\frac{\partial^2 F}{\partial \mathbf{x}^2}\right)^{-1} \frac{\partial^2 F}{\partial \mathbf{z} \partial \mathbf{x}}\Big|_{\mathbf{x}=A(\hat{\mathbf{z}})}. \tag{2.39}$$

Using this relation it could be obtained the final formulation of the uncertainty propagation from a measurement $\hat{\mathbf{z}}$ to the solution $\hat{\mathbf{x}}$ also when the minimization algorithm A has not a derivative, by only knowing the error function F:

$$cov(\hat{\mathbf{x}}) \simeq \left(\frac{\partial^2 F}{\partial \mathbf{x}^2}\right)^{-1} \frac{\partial^2 F}{\partial \mathbf{z} \partial \mathbf{x}} cov(\mathbf{z}) \frac{\partial^2 F}{\partial \mathbf{z} \partial \mathbf{x}}^{\top} \left(\frac{\partial^2 F}{\partial \mathbf{x}^2}\right)^{-1}. \tag{2.40}$$

Analysing the covariance formulation it is evident that the only source of error considered is the sensor noise on the measure $\mathbf{z}$. Other sources of error are not considered: for example it is assumed that the algorithm can not be trapped in a local minimum. It has been tested that with wrong converge the covariance estimation increase quickly, so it is possible to deal with this kind of problem, and also check the right convergence of the solution for example in the ICP framework.

Moreover the uncertainty estimation in the data association is very useful also in situations in which there are not information in the environment suitable for the data association. In this case the all the matching for an

undetermined system of equation that does not allow to evaluate a unique solution. Undetermined situations can be also considered as source of error: classical examples in a bi-dimensional pose estimation can be corridor or a circular environment without any other information in terms of texture or colour. In this case also the data association methods can fail: anyway it is possible to detect these situations because in this case the value of one of the covariance matrix eigenvector becomes very high.

Chapter 3

Methods and Algorithms

In this chapter are introduced the methods and the algorithms developed for the research topics of the thesis and used to solve the problem of the vehicle motion estimation in a general scenario. The review of literature reported previously is useful in order to emphasize the problems on these topics and to suggest the development directions to be taken.

In particular it has been noticed the lack of robust methods suitable for a generic scenario. If on the one hand in the state-of-the-art works there are rigorous studies that deals with robust and reliable localizations, on the other hand these applications are limited to some categories of environment. Since the aim is to introduce a general method that can work in every condition and environment, it has been evaluated also the use of more than one instrument in order to overcome the situations in which one of the sensor can not return reliable information from the environment.

To achieve the target of this kind of localization it has been shown as the SLAM is if not the only one, at least the best method. In regard of the SLAM method there are still some questions proving that this problem is not yet completely resolved. Indeed, in order to achieve a solution there are some strong assumptions in terms of non linearity modelling and in the descriptions of the uncertainties with a Gaussian distribution.

3.1 Working Procedure

The aim of the thesis, as already explained in the previous chapters, is the motion estimation of a mobile robots. It has been also shown that it is a field in continuous evolution: for this reason it was really important to follow a well defined working procedure.

To achieve to the final objective of vehicle localization, the work has been divided in the three main topics already used to classify the literature review: the processing of information from the environment, the data association and finally the incremental and global localization. This work

division is useful in the explanation of the arguments but it is also a formal classification. In fact the topics can not be developed independently since they are closely connected: the choice of a particular solution in one of them can cause strong limitations or completely prevent the final objective. While the processing of the environment information from the measurements is the most studied topic, the main issues are in the two other points. Especially the data association must be very robust and reliable, in both the incremental and global estimation. Indeed, in the incremental estimation are required data association approaches with high reliability in order to limit the drift error, while in the global are needed also robust approaches since the associations can involve the comparison of information collected from very far observations.

The first part of the work was focused on the use of the measurements from the laser scanner mainly because of the sensor returns directly metric information of the surrounding information. The laser scanner is an instrument widely used in the research field of mobile robotics: it is possible to find a huge amount of test and datasets to compare the results in different application. Since the final target is to implement a SLAM approach to localize the vehicle in a general scenario, as seen in the Section 2.3.3 classical implementations based on the EKF-SLAM have strong limitations on the mapping accuracy and in the fact that the filtering approaches require a map described by a set of features. Using a SLAM approach based on the Graph Theory is possible to avoid the filtering operation and increasing the map accuracy. The problem in this case is that the methods implemented in literature are still based on feature: in order to remove this limitation has been developed a SLAM approach for a general scenario using a dense scan matching approach.

To achieve this final objective is important to have a robust and reliable data association approach: part of the thesis work has been spent also on this topic. Preliminary results on the data association approach developed for the thesis show that it can be also used not only in the SLAM method but also in other different applications. An example was the implementation of these methods in the *International Thermonuclear Experimental Reactor (ITER) Remote Handling (RH)* project [72] .

Concerning the motion estimation by means of the camera, the working procedure followed has been quite similar to the one used in the laser. It has been chosen to develop the vehicle localization based on camera after the one based on the laser in order to have an accurate method to test the results. The intention of the work here reported using as sensor the camera was to provide a framework on which apply the data association and global localization approaches already used with the laser.

3.2 Sensor Data Processing

The elaboration of the data collected from the sensors is the first operation to perform in order to estimate the vehicle motion. As explained in Section 2.1 it can be classified according to the amount of processing required. For the laser scanner and camera sensors this classification is the same and it defines two categories: feature based and dense method. However, the implications on the typology of environment that involves the choice of a method rather than the other are different for the two sensors.

Because of the amount of information collected by the laser scanner sensor is limited to a bi-dimensional set of points, the feature can only be defined on the geometric interpolation of the environment. This is a strong limitation on the type of environment, that can be overcame by limiting the data processing and using the raw information.

With respect to the camera, the limitations on using feature are less than in the case of the laser scanner data. This is due to the large content of information in a image, which it is not limited to a geometric representation but also includes information on colour, texture and so on. Hence the features identification in the images is not a limitation on the environment type.

These considerations are really important in the selection of the data fusion method to apply on laser and camera. Approaches already used in literature in which the camera is used to increase the information of the laser features (Section 2.1.3.1) do not use all the achievable information in the camera images and they limit the vehicle navigation in environment that can be described with simple geometries.

3.3 Laser Scans Processing

The LRF sensor returns a set of bi-dimensional points that represent its discretization of the environment in a polar coordinate system since its working principle is based on a rotating mirror. These points can be simple transformed in a Cartesian coordinate system and defined as *laser scan* or simply *scan* here identified with $\mathbf{q}$. A laser scan is an ordered set of points which represents a 2D measurement of the environment profile in the local reference frame of the laser instrument. At each point of the scan is associated an uncertainty due to the noise of the sensor in the beam range and angular measurements. The uncertainty expressed in the polar coordinate system can be mapped in the Cartesian coordinate system using the uncertainty propagation formulation.

As introduced in Section 2.1.1.1 the final goal of the motion estimation is to evaluate the vehicle pose $\mathbf{X}$ by computing an optimal spatial transformation between two scans acquired in two different vehicle position, or

between a scan and the environment map. A *map* is defined as a synthetic representation of a known environment: the type of data stored are strictly related to the approach used in the scan matching. The processes of associating data between the scan and the map or between two scans and of computing the spatial transformation go under the name of *scan matching*. The classification in terms of data processing is defined as:

- Feature based matching: natural landmarks are identified in a scan, and the pose estimation is a result of their matching either with the landmarks of another scan or of the map.
- Dense matching: all the points in a scan are used to perform the matching procedure and so to return a pose estimation.

It must be also recalled that a fundamental step, in order to be able to associate an uncertainty to the vehicle position, is the evaluation of the uncertainty of the matching. With the knowledge of the map, path and laser parameters, using the method reported in [18] the Fisher Information Matrix is computed and thus the minimum achievable uncertainty on pose estimation as well. In this way it is possible to compute the *uncertainty lower bound* of the vehicle pose. In addition, by observing the Fisher Information Matrix, it is possible to detect the *undetermined situations*. In these cases, like corridor or circular environment, is not possible to obtain a solution from the scan matching with any approach. The navigation in an unknown environment assumes that the map is unknown: the method previously described can not be applied in this kind of navigation, but it provides some useful suggestions to check undetermined situations during the scan matching.

3.3.1 Feature based scan matching

A feature can be defined as natural landmark of the environment that is invariant with the observation point of view and locally unique: these are key requirements to perform a reliable matching. They have been introduced to avoid some typical problems of the laser scanner measurements: different scans of the same scenario are strongly dependent on the observation pose; measurement noise and discretization can bring to wrong data association and feature extraction can mitigate this problem.

When the environment is structured and defined by a set of basic geometric entities, the easiest features to extract from the scan are *segments* and their intersection points called *corners*. The extension of the feature definition to geometrical entities of higher order, like curves, is possible but in most of the applications the highest order employed is the linear one. Even with the simplest features, the identification must deal with partial occlusions or discontinuities in the landmark surfaces. To ensure a reliable

pose estimation, it is required a complete uncertainty evaluation. Each feature must be assigned a position and an uncertainty, and the minimization problem in the matching framework should be a weighted least squares problem which takes into account features uncertainties. This has been done in the present work.

In order to achieve a feature unique local description, in [8] were defined several types of features. From a practical point of view, the definitions of many kind of natural landmarks can create some difficulties, because of changing in point of view can involve also a feature transformation (Figure 3.1). Hence it is possible to define only a feature type, corner. Every corner is defined by the intersection of two segments and because of the laser acquisition has a clockwise or counter-clockwise rotation, these segments can be sorted with respect to this rotation. In some particular situations some features are defined as the end point of a segment, like the one depicted in Figure 3.1 in position $\mathbf{m}_{i+1}$. This is due to the lack of information on the hidden segment. By describing it as a corner it is possible to take care also of the changes in the feature information.

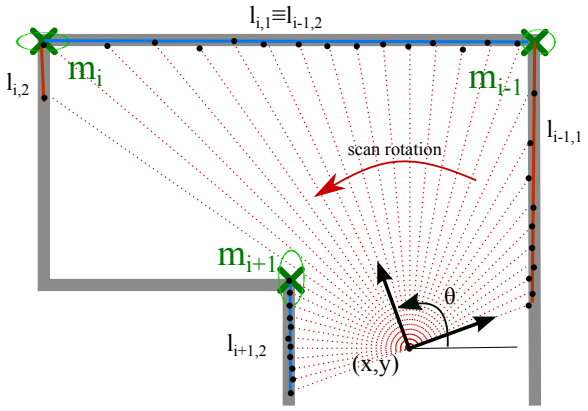


Figure 3.1: Schematic representation of features. While M_{i-1} and M_i are defined by two segments each, M_{i+1} is defined by only one segment: by the vehicle motion it will be detected as corner.

A feature can be described as a set:

$$M = \{\mathbf{m}, \mathbf{C}_m, \mathbf{l}_1, \mathbf{C}_{l_1}, \mathbf{l}_2, \mathbf{C}_{l_2}\} \tag{3.1}$$

where M is the set of feature information, $\mathbf{m} = [m_x, m_y]^\top$ is the feature position in the scan reference frame with covariance matrix $\mathbf{C}_m$, $\mathbf{l}_1$ and $\mathbf{l}_2$ are the segments connected with $\mathbf{m}$, sorted according to scan rotation direction (Figure 3.1), $\mathbf{C}_{l_1}$ and $\mathbf{C}_{l_2}$ are their uncertainties in the Hesse plane. The covariance matrix of a feature position, defined in the scan reference frame, can be expressed as:

$$\mathbf{C}_m = \begin{bmatrix} c_{xx} & c_{xy} \\ c_{xy} & c_{yy} \end{bmatrix}. \tag{3.2}$$

The generic segment is defined in the Hesse plane by the vector $\mathbf{l} = [l_\delta, l_r]^\top$ and its covariance matrix $\mathbf{C}_l$ is:

$$\mathbf{C}_l = \begin{bmatrix} c_{\delta\delta} & c_{\delta r} \\ c_{\delta r} & c_{rr} \end{bmatrix}. \tag{3.3}$$

A cartesian point which lies on the line segment must satisfy the constraint:

$$q_x cos(l_\delta) + q_y sin(l_\delta) = l_r, \tag{3.4}$$

where $\mathbf{q} = [q_x, q_y]^\top$ is the point in Cartesian coordinates. Using this features description, instead of using clusters of noisy scan points, allows to achieve a better local unique description, and so to perform a more reliable matching.

3.3.1.1 Segment identification

Segments can be extracted with a classical split and merge approach: as reported by Nyugen [20] it is the best choice for a geometrical environment. A graphical representation of the split and merge iterative approach is depicted in Figure 3.2. The performance of this approach can be improved by pre-processing the set of points with a clusterization. The main problem

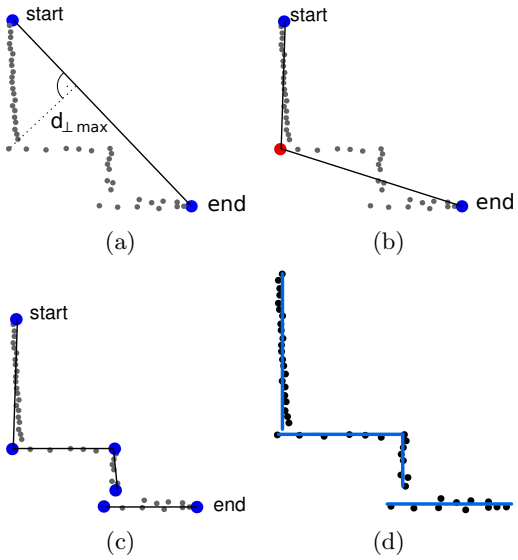


Figure 3.2: Split and merge segments identification on cluster of point, identification of the maximum distance point (a) and segment splitting (b). The process stops when the maximum distance is under a threshold (c) and is possible to interpolate the segments (d).

of this algorithm is the tuning of the constant parameters: by supposing to

know the laser scanner measurement noise it is possible to obtain a non-parametric method.

Otherwise, for cluttered environment, a method that can ensure high performances and a good segment extraction, without constant parameters, is the *Adaptative Scale Sample Consensus ASSC* [21], which also does not require the knowledge of the measurement noise. In this method is introduced the concept of the *scale* that in classical method as the split and merge is related only to the noise standard deviation of the sensors. Actually there are several factors that have influence on the scale, like the angle of the beam, the kind of surface and so on: the scale is a property of each segment of the scene. The method is based on the maximization of an utility function $U(\theta)$ defined as:

$$U(\theta) = \frac{n_k}{\sigma_k}, \tag{3.5}$$

where θ is the vector of the line parameters, n_k is the number of points in the segment and σ_k is the standard deviation of the segment interpolation residuals. If on the one hand this method works fine in cluttered environment, on the other hand when the environment is composed by segments of different size connected each other it has some limitations. These are due to the fact that the utility function promotes the ratio between the number of points and the residual standard deviation.

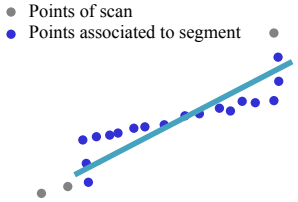


Figure 3.3: ASSC segment identification scale problem on connected segments.

The problem is reported in Figure 3.3: to the segment with high scale are associated also the points of the adjacent segments. It is possible to limit these problem by interpolating the segments with a method that allows the outlier identification, like the LMedS (Section 2.2.2).

Though with these methods is possible to obtain good results, they require a post-processing procedure to fix problems due to occlusions and surface discontinuities. Concerning the segments interpolation, given the measured points it is possible to follow the method described in [19] which also allows to reliably compute the segments uncertainties in the Hesse coordinate systems.

3.3.1.2 Feature identification

The feature identification and their uncertainty estimation are based respectively on segments parameters and their uncertainties. When a feature can be identified as the intersection of two segments, like a corner, its position is defined by the segment intersection. Writing the two segments in the form of equation 3.4 is possible to define a system in which the unknowns q_x and q_y are the coordinate of the feature $\mathbf{m}$. The feature uncertainty can be expressed by propagating the input uncertainties of the connected segments:

$$\mathbf{C}_m = \mathbf{J}\mathbf{C}_{l_{1,2}}\mathbf{J}^\top, \tag{3.6}$$

where $\mathbf{J}$ is the Jacobian of the system of equations used to compute the intersection point, and $\mathbf{C}_{l_{1,2}}$ is the joint covariance matrix of the intersecting segments $\mathbf{l}_1$ and $\mathbf{l}_2$. The Jacobian is defined as:

$$\mathbf{J} = \begin{bmatrix} \frac{\partial m_x}{\partial \delta_1} & \frac{\partial m_x}{\partial r_1} & \frac{\partial m_x}{\partial \delta_2} & \frac{\partial m_x}{\partial r_2} \\ \frac{\partial m_y}{\partial \theta_1} & \frac{\partial m_y}{\partial r_1} & \frac{\partial m_y}{\partial \delta_2} & \frac{\partial m_{y,i}}{\partial r_2} \end{bmatrix}. \tag{3.7}$$

The composed covariance matrix is defined as:

$$\mathbf{C}_{l_{1,2}} = \begin{bmatrix} \mathbf{C}_{l_1} & \mathbf{0}_{2\times 2} \\ \mathbf{0}_{2\times 2} & \mathbf{C}_{l_2} \end{bmatrix}, \tag{3.8}$$

where the segments $\mathbf{l}_1$ and $\mathbf{l}_2$ are considered uncorrelated.

In case of features defined only by one segment, that is an endpoint of a segment, it can be used the same approach by replacing the missing segment with the next laser beam closest to the endpoint. The parameters of the ray in Hesse plane are assumed uncorrelated and the standard deviation of the angular parameter δ of the missing segment is considered proportional to the laser angular resolution. In addition, when a segment end point is detected it must be checked also that the feature is in the foreground. On the contrary the detected point is not a stable natural landmark.

3.3.1.3 Feature matching

Once two set of features $\left\{M^1, M^2\right\}$ are available, the matching procedure can be implemented. To ensure a reliable solution the matches are selected using a probabilistic gate where the Mahalanobis distance between the features of the candidate pair is evaluated. When global localization with respect to the map is performed, the set M^2 corresponds to the map features while M^1 corresponds to the scan. The set M^1 is transformed to M^{1*} according to the initial guess for the pose estimation. This practically means that the feature positions, their covariances and the segments information are transformed in the reference system of the second scan or of the map according to the tentative pose transformation. Then is applied the

association filter. To strengthen the association check, also the probabilistic distances of the intersecting segments in the Hesse plane are evaluated for each feature by computing the following equations:

$$\begin{cases} {D_m}^2 = \left(\mathbf{m}_i^{1*} - \mathbf{m}_j^2\right)^\top \left(\mathbf{C}^*_{m_i^1} + \mathbf{C}_{m_j^2}\right)^{-1} \left(\mathbf{m}_i^{1*} - \mathbf{m}_j^2\right) \\ {D_{l_1}}^2 = \left(\mathbf{l}_{i,1}^{1*} - \mathbf{l}_{j,1}^2\right)^\top \left(\mathbf{C}^*_{l_{i,1}^1} + \mathbf{C}_{l_{j,1}^2}\right)^{-1} \left(\mathbf{l}_{i,1}^{1*} - \mathbf{l}_{j,1}^2\right) \\ {D_{l_2}}^2 = \left(\mathbf{l}_{i,2}^{1*} - \mathbf{l}_{j,2}^2\right)^\top \left(\mathbf{C}^*_{l_{i,2}^1} + \mathbf{C}_{l_{j,2}^2}\right)^{-1} \left(\mathbf{l}_{i,2}^{1*} - \mathbf{l}_{j,2}^2\right) \end{cases} \tag{3.9}$$

where D_m is the Mahalanobis distance between the j^{th} feature position in M^2 and the i^{th} feature position in M^{1*}, with a transformed covariance matrix $\mathbf{C}^*_{m_j^1}$. D_{l_1} is the probabilistic distance between the first segment $l_{i,1}^2$, with uncertainty $\mathbf{C}_{l_{i,1}^2}$, of feature M_i^2 and the first segment $l_{j,1}^{1*}$, with uncertainty $\mathbf{C}^*_{l_{j,1}^1}$, of feature M_j^{1*}, all expressed in the reference frame of the set M^2. Likewise, D_{l_2} is the probabilistic distance of the second segments connected to each features.

Using also the information on the connected segments is possible to create features locally unique. In case of comparison between a corner and an end point is possible to neglect the evaluation on the missing segment.

The solution kernel for the computation of the planar transformation matrix between the two feature sets is obtained in a closed form by weighting with uncertainties, in practical terms by minimizing the sum of Mahalanobis distances of the associated features:

$$\min_{\mathbf{H}} \sum_i \left\| \mathbf{H} m_i^1 - m_i^2 \right\|^2_{\mathbf{C}_{ij}^{-1}}, \tag{3.10}$$

where m_i^1 and m_j^2 are the associated features respectively in the first and in the second vehicle pose and the norm used is defined as $\|\mathbf{v}\|^2_{\mathbf{C}} = \mathbf{v}\mathbf{C}\mathbf{v}^\top$ and so corresponds to the Mahalanobis distance. For this purpose valid methods are the factorization with uncertainties [47] or the optimization in closed form reported in [22]. With regard to the solution uncertainty, in Section 2.3.5 is described a general method to estimate the uncertainty of optimization problem.

Lastly there is the problem of the detection of undetermined situations. In the feature based method is quite simple to detect these kind of situations, since the matching requires the presence of at least two features to solve the planar motion. When the number of feature is not sufficient a solution can not be achieved.

3.3.2 Dense scan matching

The dense scan matching approaches have been introduced in order to overcome the limitations of the feature in terms of geometrical description of the environment. In the dense matching category more than one method can be identified. The most used method matches pairs of points from two different scans or from one scan and a suitable point-based representation of the map, it is called *Point to Point (PP)* matching. An alternative version of this method matches the points of a scan with a line interpolation of the other, instead of matching pairs of points. It is called *Point to Line (PL)* matching.

3.3.2.1 Point to Point matching

Point to Point (PP) matching is the simplest implementation to perform scan matching. It must be recalled that it works on the sensor representation of the environment, with measurement noise. Furthermore, differently from the features, each point in the scan is not univocally distinguishable. For this reason is important to use a robust implementation of the ICP framework to achieve an accurate solution. These characteristics limit the performance of the approach. To ensure a good accuracy and precision, the approach must be applied to a moderate pose variation between the two scans. It can also work with large pose variation but in this case its precision decreases. In each ICP iteration it could be found the best estimation of the planar

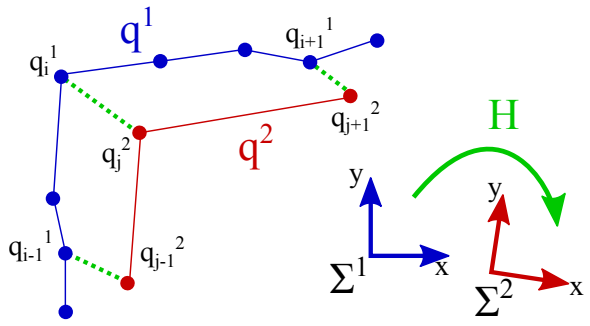


Figure 3.4: Schematic representation of the PP procedure between two set of points q^1 and q^2, which are respectively espressed in their reference frame Σ^1 and Σ^2, that differ for the transformation **H**.

transformation H between the two set of points q^1 and q^2, by minimizing the functional:

$$\min_H \sum_i \left\| H q_i^1 - q_j^2 \right\|^2, \tag{3.11}$$

where q_i^1 and q_j^2 indicate respectively the i^{th} point in the first set associated with the j^{th} point on the second set and **H** is the planar transformation

between the two point sets. Like the other dense matching approaches for laser scanner, there are some critical situations in which the pose estimation fails. In corridor with all parallel walls or in circular environment the PP cannot return correct solutions, and unlikely the PL approach it is not possible to detect these situations. Also in this case the covariance can be computed in closed form with the method reported in Section 2.3.5.

Against all these drawbacks, the PP has also an important advantage, its robustness to large initial displacement errors. If the laser scans are closer enough to each other in terms of displacement and rotation this method can work also without initial guess of pose estimation.

Concerning the pose estimation, omitting numerical optimization, almost all the other methods are based on Principal Component Analysis evaluation. A classical approach is to use the Singular Value Decomposition (SVD), that requires some improvement [73] in order to avoid the typical problem of this method. Otherwise a simpler method has been proposed by Lu [10].

3.3.2.2 Point to Line matching

The Point to Line PL approach has been proposed to obtain a better modelling of the dense scan matching. The rationale is that the ICP has to solve a *surface matching problem* even though the laser scans are quite rough discretizations of the environment surfaces.

The new formulation can be stated as follow: given a reference surface S^{ref} and a set of points q, the problem is to find the planar transformation $\mathbf{H}$ that minimizes the distances of the points q to their transformed projections on the reference surface S^{ref} using the transformation $\mathbf{H}$:

$$\min_{\mathbf{H}} \sum_i \left\| \mathbf{H}q_i - \prod \left\{ S^{ref}, Hq_i \right\} \right\|^2, \tag{3.12}$$

where the symbol $\prod \left\{ S^{ref}, \cdot \right\}$ denotes the projection on S^{ref} of the i^{th} point of the q set. This formulation needs a representation of the surface S^{ref} that is known if the scan is compared to a map, and so the solution is easy [5]. Otherwise, if the method is used to match scans, the surface formulation must be estimated. In this case is anyway possible to interpolate with a surface one of the two scans, in order to evaluate the normals on which compute the point to surface distances. The surface representation can be done in several ways, but a method quite simple, without any problem of interpolation stability, is the linear interpolation and for this reason the approach is defined *Point to Line*. The line is the one that, given a point of a scan, contains the two closest points in the reference scan (Figure 3.5); the goal is to minimize the distance of each point to its associated line.

The ICP in this approach is used on q_j^2, the j^{th} point of the second scan, to find the two closest points in the first scan, denoted by $q_{i,1}^1$ and

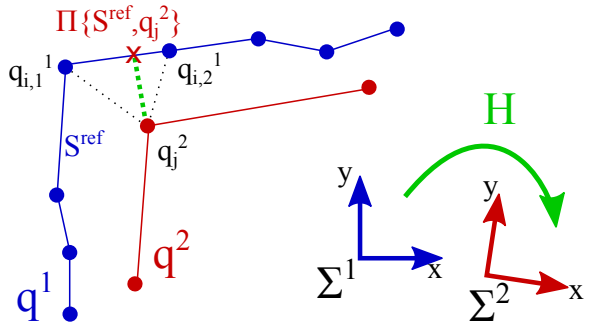


Figure 3.5: Schematic representation of the PL procedure: for each point q_j^2 in the second set, two points $q_{i,1}^1$ and $q_{i,2}^1$ are selected in the first one. With a linear interpolation it is possible to approximate the reference surface S^{ref} and find the point reprojection on it.

$q_{i,2}^1$. These two points define a line, with normal $\mathbf{n}_i = [n_{x,i}, n_{y,i}]^\top$, which is used to solve the surface matching problem. The functional to minimize is defined by equation (3.12):

$$\min_{\mathbf{H}} \sum_j \left\| \mathbf{n}_i \left[\mathbf{H} q_j^2 - q_{i,1}^1 \right] \right\|^2 . \tag{3.13}$$

Solely with this simple scan interpolation is possible to increase the performance of the dense matching approaches. In addition, since the approach does not work on a single point of the scan, the solution can be accurate also with large position variation. The main drawback of this approach is the reliability: it is very sensitive to the presence of outliers in the matching. For this reason it requires a robust implementation of the data association and a good initial pose estimation.

Also in the PL scan matching minimization can be performed in closed form [22] like its uncertainty (Section 2.3.5). Unlike the PP approach, with the PL is possible to detect the undetermined situations in which the solution is not well defined. The evaluation of the Fisher Information Matrix in this case is not feasible, but observing the covariance matrix associated to the solution is noticeable that in case of undetermined situations one of its eigenvectors tends to the infinite. This means that uncertainty and so the solution is not defined in one of the pose parameters.

3.4 Laser Data Association

A critical task in the scan matching process is the data association. In this work the different implementation pipelines of each approach share the same data association algorithm both for the robust association of features or for PP and PL matching. The algorithm is based on a robust improvement of

the *Iterative Closest Point* (ICP) used in [14] and [13]. More specifically, the robustness is achieved by applying the *Least Median of Squares* (LMedS) [43] method in order to detect the outlier associations. The pose estimation for all the approaches has a closed form solution thus reaching a good efficiency of the LMedS implementation.

Although this robust implementation of the ICP, it must be take account of some important considerations. The first one is the difference between the feature based and dense matching: while to a feature is associated a descriptor that ensure its local uniqueness, it is not possible to build a descriptor for the single scan point. Secondly, the robust implementation with LMedS on the association can work until the 50% of outliers. This limit is adequate when the pose variation between the two laser measurements is limited, but because of the data association must be used also in the global localization whit large pose variations, it must put in practice other preventive measure to limit this percentage.

The first one is a sifting process based on the *sensor field of view*. It requires an initial estimation of the pose variation that is used to detect the laser scans common area and so to limit the number of points or features that can create outlier associations.

A second preventive operation is to use *bi-unique correspondences* in the definition of the associations. In classical ICP algorithm implementation the correspondences are defined by searching the closest points between a first set of points to the second one. Thus a point of a set can have multiple correspondences with points in the other set: most of these correspondences can be outliers. To avoid this situation it is possible to carry out the ICP by evaluating the correspondences from the first set to the second one and vice versa. This double correspondences can be processed to find the connections available in both the directions and so defining the bi-unique correspondences.

3.4.1 Detection of outlier associations

The implementation of the ICP algorithm with the sensor field of view filter and the bi-unique correspondences does not ensure the correctness of all the associations. The wrong correspondences between points or features of the two sets are defined as outlier associations. Because of all the scan matching approaches are based on a least square minimization the presence of this outliers prevents the settlement of a reliable solution. For this reasons there is the need of a robust ICP capables to detect the wrong associations. In Section 2.2.2 has been presented a method that can detect the presence of the outliers, the LMedS algorithm. This algorithm can be used in the detection of outlier associations by considering the ICP correspondences as measurements and the planar transformation that describes the pose variation as the model parameters. By evaluating the planar transformation on several

samples of correspondences is possible to detect the outlier associations.

Let q^1 and q^2 be two sets of points or features on which perform the scan matching and $\hat{\mathbf{H}}$ be is the initial estimation of the planar transformation between the two vehicle poses. The LMedS general procedure can be summarized in the following steps:

1. Transformation of the first set of points q^1 in the reference frame of the second set using the initial estimation $\hat{\mathbf{H}}$: the new set of point thus obtained is denoted with q^{1*}.

2. Filtering of the q^{1*} and q^2 sets by taking care of the sensors field of view.

3. Definition of the bi-unique correspondences $\left\{q_i^1, q_j^2\right\}$ using the ICP algorithm between the two set of points q^{1*} and q^2.

4. Selection of N random samples $\Gamma_k, k = 1...N$ on the bi-unique correspondences $\left\{q_i^1, q_j^2\right\}$. The number of correspondences in each sample must be equal to the minimum number of associations needed to evaluate the planar transformation between the two set of points: to estimate the planar transformation the minimum number is $s = 2$. The number of samples N must be evaluated using the Equation 2.20.

5. Estimation of the planar transformation $\mathbf{H}_k$ using the association stored in the samples $\Gamma_k, k = 1...N$.

6. Computation of the association distances for all the correspondences found by the ICP algorithm. The distances must be evaluated following the procedures illustrated for each scan matching approach. The distances are the residual of the model interpolation and are stored in the vectors $\mathbf{r}_{\mathbf{H}_k}, k = 1...N$.

7. Evaluation of the vector of the medians $\mathbf{M} = [m_1, ..., m_k, ...m_N]^\top$ associated to the squared residuals $\mathbf{r}_{\mathbf{H}_k}, k = 1...N$:

$$m_k = median\left(\mathbf{r}^2_{\mathbf{H}_k}\right). \tag{3.14}$$

8. Selection of the least median square m'_k from the vector $\mathbf{M} = [m_1...m_N]^\top$ and its residuals vector $\mathbf{r}'_{\mathbf{H}_k}$:

$$m'_k = min(\mathbf{M}). \tag{3.15}$$

9. Computation of the residuals threshold $\hat{\sigma}$:

$$\hat{\sigma} = 1.4826\alpha\left(1 + \frac{5}{n-s}\right)\sqrt{m'_k}. \tag{3.16}$$

In the feature based scan matching the threshold is set to the desired confidence level, since the residuals in this case are Mahalanobis distance.

10. Identification of the outliers association by comparing the residuals $\mathbf{r}'_{\mathbf{H}_k}$ of the sample with the least median square with the threshold $\hat{\sigma}$.

11. Evaluation of the final solution by using only the inlier associations.

Actually using the LMedS method there would be no need to use an iterative algorithm like the ICP because of the sampling procedure consents to estimate the planar transformation on a set of points without outliers. Indeed ICP uses the iterations to overcome the presence of outliers. However if the initial motion estimation is distant from the final solution quite a few associations are incorrect: the LMedS classifies them as outliers and computes the final solution on a small number of association. Instead, by performing another iteration of the proposed framework, the ICP can work on a better initial motion estimation and so it can return more correct associations that LMedS recognizes as inlier. The final solution is more accurate since it is evaluated on a large number of associations.

In terms of computational cost the three scan matching approaches are similar. If on one hand the time required from dense matching approaches in the association phase is higher than in the feature-based approach due to the higher number of matches to evaluate, on the other hand the feature approach has an additional step that is the natural landmark identification.

3.4.2 Consistent association with Joint Compatibility

The consistent data association with *Joint Compatibility (JC)* permits to avoid the ICP procedure by exploring all the possible associations between two set of points. This is particularly useful when there are a limited number of associations to evaluate and so the wrong estimation of a correspondence can cause the failure of the scan matching approach. It must be notated that the use of the LMedS procedure in the ICP framework is very similar to this algorithm.

The combination of LMedS and ICP defines probably a method with a slower convergence since it requires more than one iteration while using the JC only one iteration is needed. On the other hand the LMedS evaluates the association by considering limited number of minimal correspondences combinations: the JC instead perform an exhaustive search on all the possible combinations. Increasing the number of points in the sets the LMedS uses always the same number of evaluations while the number of combinations in the JC, also using the Branch and Bound algorithm, increase exponentially. From this point of view the JC is more like a brute force research algorithm.

For this reason in this thesis work has been preferred the implementation of the LMedS algorithm in the ICP procedure than the Joint Compatibility data association.

3.5 Camera Images Processing

The decision to estimate the vehicle motion in a generic environment prevents to fuse the information in the environment data processing as seen in Section 3.2: camera and laser scanner have to provide the vehicle motion independently. While in the laser scanner the basic operation is the scan matching, in the camera images processing the analogous procedure is called *Visual Odometry (VO)*.

The Visual Odometry is the process of estimating the ego-motion of an agent (e.g., vehicle, human, and robot) using only the input of a single or multiple cameras attached to it. The term was coined in 2006 by Nister [74] in his paper: the name was chosen for its similarity to the wheel odometry. Likewise, using the VO is possible to estimate incrementally the vehicle poses by examining the changes that motion induces on the images acquired by the on board cameras.

The most used technique in this field is the classical odometry and almost every vehicle has rotary encoders to measure wheels rotation. Unfortunately it has some drawbacks:

- Error drift on pose estimation due to the slippery and slide of wheels on the floor.
- It can not be applied to mobile robots with non-standard locomotion methods.
- It assumes that the vehicle moves in a plane.

Using instead a measurement system that can estimate its position observing the environment, it is possible to overcome the odometry drawbacks. With three-dimensional information the vehicle could be able to localize itself also in non-planar environment. Moreover, the scenario observation is indispensable to planning the vehicle path. In this order of things VO could be an important improvement to vehicle localization. Its capabilities makes VO an interesting supplement to wheel odometry or to other navigation systems.

The problem of recovering relative camera poses and the three-dimensional structures from a set of camera images (calibrated or non calibrated) is known in the computer vision as *structure from motion (SFM)*. It is a field studied since the early 1980s, but real-time working systems has been deployed only in the last years.

3.5.1 Monocular vs Stereo Cameras

Though in the work here reported has been used only one camera it is interesting to illustrate the differences with multi-camera approaches. In this context it should be emphasized that the state of the art works use the monocular pose estimation.

Visual Odometry algorithms can be classified in several way. One of the most important classification is based on the number of camera used:

1 camera only one camera it is used, the pose between two frames is unknown. It is defined as *Monocular VO*.

2 or more cameras it is used stereo or trinocular configurations for the cameras. The cameras relative position must be well known. Though it is possible to have more than two cameras is defined *Stereo VO*.

This classification is really important not only for the implementation of the method, but also for the results that can be achieved. Most of the research done in VO was produced using stereo cameras: indeed the VO is a particular case of the SFM, which was developed starting from multi-camera setups. The main characteristic of the stereo VO algorithms is that the three-dimensional position of features is evaluated by using a triangulation procedure in each vehicle pose. Successively by comparing these set of points computed in different vehicle poses it is possible to estimate the camera motion.

The alternative to the stereo vision is to use a single camera. In this case there are no information on camera poses, since by using only a camera they are the unknowns of the problem. The difference from the stereo scheme is that both the relative motion and the three-dimensional structure must be inferred from the bi-dimensional information in the image plane.

The main drawback of using a monocular approach is that the motion can only be recovered up to a scale factor. The reason is that in the epipolar geometry there are three main variables: the correspondences, the camera and the scene geometries. By knowing two of them it is possible to achieve a complete information on the third one. In this case already two of them are unknown: the cameras extrinsic parameters, since they are the ego-motion to estimated and the scene geometry, because it has been assumed a navigation in an unknown scenario. It means that a complete information can not be achieved, in fact the scale factor can not be computed with only these information. The absolute scale can be determined from:

- Direct measurements of a known object in the environment .
- Motion constraints.
- Integration with other sensors.

Anyway also the stereo VO has some drawbacks:

- The camera extrinsic parameters must be computed with a calibration.
- The camera extrinsic parameters degradation over the time.
- The stereo VO can degenerate to the monocular case when the distance to the scene is much larger than the stereo baseline.

3.5.2 Visual Odometry

Once that the basic idea of the VO has been introduced, it is possible to go further in the formulation of the problem. The scenario is the following: a vehicle is moving through the environment, with a rigidly attached camera system. At each discrete time t the camera takes a picture I_t: by using a monocular system at the time n is possible to obtain a set of images $I_{0:n} = \{I_0, ..., I_n\}$. Because of the VO is an incremental procedure, at each time t it allows to compute the rigid body transformation from the previous camera position and orientation to the actual one:

$$\mathbf{H}_{t-1}^{t} = \begin{bmatrix} \mathbf{R}_{t-1}^{t} & \mathbf{t}_{t-1}^{t} \\ 0 & 1 \end{bmatrix}, \tag{3.17}$$

where $\mathbf{R}_{t-1}^{t}$ is the rotation matrix and $\mathbf{t}_{t-1}^{t}$ the translation vector of the camera pose at the time t respect to its pose at the time $t-1$. Using

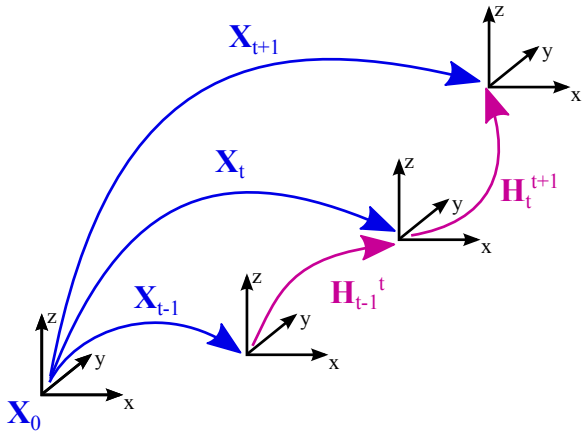


Figure 3.6: Visual Odometry incremental pose estimation.

a recursive estimation it is possible to construct the set of all subsequent motions $\mathbf{H}_{1:n} = \{\mathbf{H}_0^1, ..., \mathbf{H}_{n-1}^n\}$ and then the set of camera poses $\mathbf{X}_{0:n} = \{\mathbf{X}_0, ..., \mathbf{X}_n\}$ that describes the camera motion with respect to the initial coordinate frame. Each camera pose is obtained by concatenating all the previous transformations and so at this stage the vehicle is localized by an incremental localization.

The relations presented previously describe the geometry of the problem, while the complete VO pipeline is reported in Figure 3.7. It is noticeable as the main tasks are essentially the same defined in the laser scanner approach. For every new image acquired I_t the first two steps consist in the

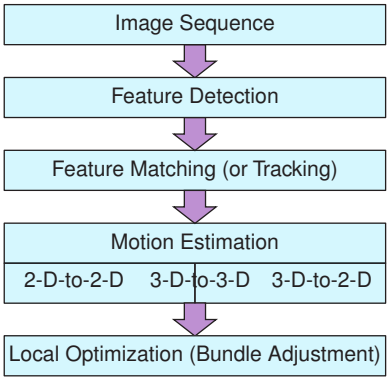


Figure 3.7: Visual Odometry framework for general purpose.

detection and matching of the bi-dimensional features in the current image with those from the previous frame: they must be the reprojection on image plane of the same three-dimensional feature. Like in the laser scanner data association, the robust data association is an important task in order to estimate correctly the motion. There are several ways to perform motion estimation that depend on the number of camera used in the VO. In the case of monocular VO, since with a single view it is not possible to achieve three-dimensional information due to the unknown scale factor, the comparison is between the bi-dimensional features information in two consecutive views. Thus the camera poses are recovered incrementally. It is possible to perform a refinement work over the last m images and this technique is the *windowed bundle adjustment*.

Following the information flow given by the reported flow chart, the first topics must be the feature detection and matching. For some reasons that will be explained in the sequent Section 3.6, it is better to explain first the motion estimation. Indeed there are some basic aspects that must be known in order to apply a robust matching.

The motion estimation for a monocular camera is the 2-D to 2-D case: the only information achievable are the features identified and matched on the camera image plane in two different poses. As shown in Section 2.1.2.3, using the Fundamental and the Essential matrices it is possible to describe the relations between points on camera in different poses. The properties of Fundamental and Essential matrices can be used to solve these relations. In particular the Essential matrix has less unknown variables and more constraint: it means that by using its formulation it could be easier to estimate the camera motion. The Essential matrix requires the employ of

the normalized coordinates. To use them the camera calibration matrix $\mathbf{K}$ must be known: hence the camera must be calibrated in the intrinsic parameters. With these premises, by applying the correspondence condition defined in Equation 2.12 it is determined a system of equations:

$$\widehat{\mathbf{x}}^t \mathbf{E}_{t-1}^k \widehat{\mathbf{x}}^{t-1} = 0, \tag{3.18}$$

where $\widehat{\mathbf{x}}^t$ and $\widehat{\mathbf{x}}^{t-1}$ are the sets of features found respectively in the images I_t and I_{t-1} and $\mathbf{E}_{t-1}^t$ is the Essential matrix which contains the motion estimation information:

$$\mathbf{E}_{t-1}^t = \widehat{\mathbf{t}}_{t-1}^t \mathbf{R}_{t-1}^t, \tag{3.19}$$

in which $\mathbf{R}_{t-1}^t$ is the rotation matrix between the two camera poses, $\mathbf{t}_{t-1}^t = [t_x, t_y, t_z]^\top$ is the translation and:

$$\widehat{\mathbf{t}}_{t-1}^t = \begin{bmatrix} 0 & -t_z & t_y \\ t_z & 0 & -t_x \\ -t_y & t_x & 0 \end{bmatrix}. \tag{3.20}$$

To estimate the motion in the context reported there is a set of standard algorithms, most of them are build for Fundamental matrix but they can also work with the Essential matrix. The most used algorithms are:

eight-point algorithm is the first solution proposed. It allows to solve the problem with a system of linear equation using a singular value decomposition.

five-point algorithm is the minimal case for 2-D to 2-D correspondences. It has became the standard in presence of outliers.

The simplest method to compute the essential matrix is the eight-point algorithm. Rearranging the matrix in Equation 3.18 and stacking the correspondence condition from eight points gives a linear equation system:

$$\mathbf{A}\mathbf{E}^* = 0 \tag{3.21}$$

where $\mathbf{A}$ and $\mathbf{E}^*$ are matrices built by stacking the information of the matrix product in Equation 3.19. The parameters of $\mathbf{E}^*$ are defined by solving this system. Using more than eight points, the overdetermined system can be solved using the SVD factorization that leads to the equation $\mathbf{A} = \mathbf{U}\mathbf{S}\mathbf{V}^\top$. The estimation of $\mathbf{E}^*$ is defined by the last column of $\mathbf{V}$. A valid essential matrix after the SVD is $\mathbf{E}^* = \mathbf{U}\mathbf{S}\mathbf{V}^\top$ with $\mathbf{S} = diag\{s, s, 0\}$, in order to fulfil the Essential Matrix constraint on its singular values (Section 2.1.2.3). To satisfy all the constraint, the correct Essential matrix is $\overline{\mathbf{E}}^* = \mathbf{U} diag\{1, 1, 0\} \mathbf{V}^\top$.

The eight-point algorithm has some drawbacks. First of all its solution can degenerated when the three-dimensional points associated to the features identified are coplanar. Moreover it is not to robust to the presence of

outlier associations. To avoid the co-planarity problem it is possible to use the five-point algorithm, but also this one has some drawbacks:

- the system to solve is not linear and it generates a tenth degree polynomial.
- eight-point works with calibrated and uncalibrated cameras, five-point deals only with calibrated cameras.
- eight-point approach can solve overdetermined systems of equation, while the five-point works only with exactly five points.

Taking care of these considerations, since both the methods are based on a SVD decomposition, the returned values are not the rotation and the translation but the SVD matrices $\mathbf{U}, \mathbf{S}, \mathbf{V}$. In order to infer the camera motion between the two frame it is easier to assume that:

- the 1st camera matrix is $\mathbf{P} = [\mathbf{I} \mid 0]$.
- the 2nd camera matrix is $\mathbf{P} = [\mathbf{R} \mid \mathbf{t}]$.

This means that the solution is referred to the first camera reference frame. Using the properties of essential matrix, it is possible to find *four solution of motion*, plus the scale factor ambiguity:

$$\begin{aligned} \mathbf{R} &= \mathbf{U}\left(\pm\mathbf{W}^T\right)\mathbf{V}^\top \\ \mathbf{t} &= \mathbf{U}\left(\pm\mathbf{W}\right)\mathbf{S}\mathbf{U}^\top, \end{aligned} \tag{3.22}$$

where the matrix $\mathbf{W}$ is defined as:

$$\mathbf{W} = \begin{bmatrix} 0 & \pm 1 & 0 \\ \mp 1 & 0 & 0 \\ 0 & 0 & 1 \end{bmatrix}. \tag{3.23}$$

The scheme of the four solutions is depicted in Figure 3.8 and it is noticeable that they are essentially the combination of two opposite rotations and two opposite translations of the second camera. The right solution can be chosen by taking care that a reconstructed three-dimensional point $\mathbf{X}$, in order to be a real point, must be in front of both camera views. Thus, to select the right solution is sufficient to check the *cheirality* [75].

In conclusion, using these formulations, it is possible to reconstruct the camera pose. Ideally by using this formulation there is no restriction on the type of motion: it must be recalled that the rotation matrix $\mathbf{R}$ and the translation vector $\mathbf{t}$ describe a three-dimensional motion. For some applications, like unmanned vehicle that navigate in a planar surface, it is enough to reconstruct the motion in a plane. If the camera has a *planar motion* the geometric relations can be simplified. A very recent work on this topic is the *1-Point RANSAC* presented in [33]. Though it is a remarkable

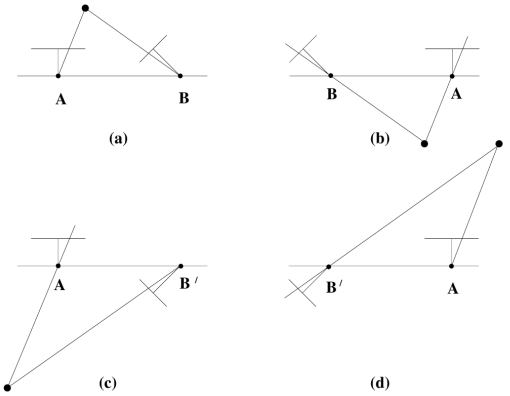


Figure 3.8: The four possible solutions associated to the Essential Matrix [34]: the right solution is the one (a) with the triangulate point in front of both cameras.

work in the vehicle motion estimation field, it requires strong restrictions on the motion, as reported in the same article by the comparing with a five-point algorithm.

The procedure implemented to estimate the motion between two vehicle poses is based on the Ortin work [32], in which it is also assumed a planar motion but with less constraint than the previous approach. As reported in

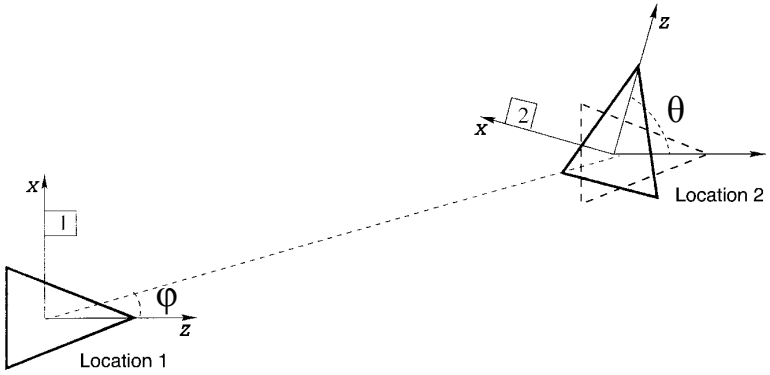


Figure 3.9: Schematic representation of the camera planar motion [32].

Figure 3.9 the camera has to move inside a plane with also the optical axis z parallel to the motion plane. Respecting these constraints the rotation matrix $\mathbf{R}$ and the translation vector $\mathbf{t}$ can be defined in terms of the camera planar pose parameters. In addition, since the translation is defined up to a scale factor it can be represented only by the direction of a unit vector. So the rotation and translation of the second camera with respect to the first

camera coordinate frame are:

$$\mathbf{R} = \begin{bmatrix} cos(\theta) & 0 & sin(\theta) \\ 0 & 1 & 0 \\ -sin(\theta) & 0 & cos(\theta) \end{bmatrix} \quad \mathbf{t} = \begin{bmatrix} sin(\varphi) \\ 0 \\ cos(\varphi) \end{bmatrix}. \tag{3.24}$$

This notation can be used in the definition of the Essential matrix reported in the Equation 3.19 obtaining its representation in terms of the two angular values that define the planar motion:

$$\mathbf{E} = \begin{bmatrix} 0 & -cos(\varphi) & 0 \\ cos(\theta - \varphi) & 0 & cos(\theta - \varphi) \\ 0 & sin(\varphi) & 0 \end{bmatrix}. \tag{3.25}$$

At this point it is possible to define the new variables:

$$\begin{array}{ll} C_\varphi = cos(\varphi) & C_{(\theta-\varphi)} = cos(\theta - \varphi) \\ S_\varphi = sin(\varphi) & S_{(\theta-\varphi)} = sin(\theta - \varphi). \end{array} \tag{3.26}$$

By substituting this variables in $\mathbf{E}$ and its stacked version $\mathbf{E}^*$ in the Equation 3.21, it could be found a linear system solvable using SVD and at least three points, being $\left[C_\varphi, S_\varphi, C_{(\theta-\varphi)}, S_{(\theta-\varphi)}\right]^\top$ the unknown parameters vector.

In conclusion, using the previous methods it is possible to compute the pose transformation between two frames. The motion estimation can be done in a three-dimensional space using the five-point and eight-point algorithms or in a bi-dimensional space by supposing a camera planar motion. In both the cases there is a drawback: the translation is known up to a scale factor. To recover the trajectory $\mathbf{X}_{0:n}$ of the camera in the image sequence, the transformations $\mathbf{H}_{0:n}$ have to be concatenated (Figure 3.6) To do this, the proper relatives scaled for subsequent transformations must be computed. A way of doing this it is to triangulate the three-dimensional points from the features matched in two subsequent image pairs and compute the distance ratio from couples of matched points: this one is the method implemented in this work. A better procedure to recover the scale is to use the trifocal tensor. In both the solution all the transformations have the same global scale factor, but it is still unknown.

In order to estimate the real scale factor is possible to use one of the method listed in Section 3.5.1. It must be remembered that the aim is the vehicle localization in a generic scenario, so the most feasible one is the fusion of information on the motion with another instrument.

3.5.3 Feature identification

In order to estimate the motion of the camera using one of the method proposed, it is necessary to carry out the two previous operation of the VO

pipeline of Figure 3.7: the identification and the matching of the features. The selection of the features must take care of some important aspects. The first one is that the navigation occurs in an unknown scenario, features based on the identification of geometrical landmark are questionable. Secondly, the feature must be detected and matched in several frames in which the camera may undergo significant pose variations. Hence it is needed an algorithm capable to identify highly distinctive and local invariant features. Only in this manner it is possible to match correctly the features with different camera poses and environment conditions.

An interesting work in which can be found a comprehensive analysis on the local invariant features is [76]. In this survey several types of feature detectors have been tested and evaluated by taking into account the most required characteristics:

- *Interest Point, Region or Local Feature*: a feature can be defined as a point, a uniform region or a point with the information of the neighbourhood.
- *Rotation invariant*: the feature descriptor changes slightly with the rotation of the camera.
- *Scale invariant*: the feature descriptor changes slightly with the distance of the camera.
- *Repeatability*: given two images of the same scene under different conditions, an high percentage of features must be detected in both the images.
- *Localization accuracy*: the feature should be accurately localized in the images.
- *Robustness*: the feature should be little sensitive to the scene conditions.
- *Efficiency*: the feature detection must be suitable for real-time applications.

Analysing the results, one of the best methods is the *Scale Invariant Feature Transform (SIFT)* [77]. It is based on a *Difference-of-Gaussians (DoG)* detector which means that it can work also in non structured environments. The features of the SIFT are image points with a descriptor of the neighbourhood that permits to obtain good performances in almost all the properties listed previously. As can be easily understood from the SIFT acronym its main characteristic is the scale invariance, however the SIFT detector performs extremely well in matching and image retrieval.

In conclusion the SIFT is a very efficient method, but it has also some drawbacks. The first one is the it requires the definition of thresholds to

remove features in low contrast regions and edges. Indeed SIFT creates a lot of key points but the ones that do not have enough contrast or lie on edges are not locally invariant features. The thresholds must be tuned on the type of environment and on the camera used. Another drawback is the lack of an uncertainty estimation of the single feature: the SIFT includes also a matching operation that returns a score based on the similarity of the matched features, but there is not information on the uncertainty associated to the definition of the feature position in the image.

3.6 Camera Data Association

The processing of the information acquired from the camera with the modalities reported above permits to obtain a set of highly distinctive features for each image. Using the SIFT detector this distinctiveness is reached by a descriptor of the neighbourhood: in addition this algorithm offers also the possibility of a matching process capable to deal with these descriptors and to return the appropriate associations.

Unfortunately the SIFT is not faultless. It has a good localization accuracy but is possible that some features can be detected with less accuracy of others. Furthermore the features are locally invariant and distinctive, so it is possible that in a repetitive environment the matching operation can create incorrect correspondences. As shown previously the methods used in the VO to solve the Essential Matrix and so compute the camera motion are sensitive to the presence of outliers in the feature associations. Thus, also in the VO is required a data association with outliers identification.

In Section 3.4 it has been introduced the data association for the laser scanner data based on a ICP algorithm with a LMedS robust implementation. The same concept can be applied also in this case with the exception of certain steps. Since the common SIFT implementations come with the matching procedure already implemented, the initial ICP steps can be avoided. After that, the steps to follow are exactly the same.

Defining also in this case as q^1 the set of features in the first image in comparison and with q^2 the ones in the second image, with reference to the framework listed in Section 3.4.1 the operations to perform in the camera data associations are:

1. Selection of N random samples $\Gamma_k, k = 1...N$ on the matched features $\left\{q_i^1, q_j^2\right\}$.
2. Estimation of the camera motion $\mathbf{H}_k$, described by the three-dimensional rotation matrix $\mathbf{R}_k$ an translation vector $\mathbf{t}_k$, from the Essential matrix evaluated using the association stored in the samples $\Gamma_k, k = 1...N$.
3. Computation of the association distances for all the matched features. The distances are stored in the vectors $\mathbf{r}_{\mathbf{H}_k}, k = 1...N$.

4. Evaluation of the vector of the medians $\mathbf{M} = [m_1, ..., m_k, ...m_N]^\top$ associated to the squared residuals $\mathbf{r}_{\mathbf{H}_k}, k = 1...N$:

$$m_k = median\left(\mathbf{r}^2_{\mathbf{H}_k}\right). \tag{3.27}$$

5. Selection of the least median square m'_k from the vector $\mathbf{M} = [m_1...m_N]^\top$ and its residuals vector $\mathbf{r}'_{\mathbf{H}_k}$:

$$m'_k = min(\mathbf{M}). \tag{3.28}$$

6. Computation of the residuals threshold $\hat{\sigma}$:

$$\hat{\sigma} = 1.4826\alpha\left(1 + \frac{5}{n-s}\right)\sqrt{m'_k}. \tag{3.29}$$

7. Identification of the outliers association by comparing the residuals $\mathbf{r}'_{\mathbf{H}_k}$ of the sample with the least median square with the threshold $\hat{\sigma}$.

8. Evaluation of the final solution by using the inlier associations.

In the implementation of the data associations it must be take into account two important aspects. The first on is regarding the evaluation of the distance of the residuals $\mathbf{r}_{\mathbf{H}_k}, k = 1...N$ in each step. Features identified by the SIFT are point in the image plane and the distance defined by the use of the Essential matrix in the correspondence condition (Equation 3.18) is a reprojection error. Since the image formation is based on a projective transformations, the algebraic distance is not the same of the geometric distance [34]. The algebraic distance is a particular case of the geometric distance. Furthermore the geometric error permits to evaluate more precisely the reprojection error, but its formulation is quite difficult; hence usually it is used an its first-order approximation defined as *Sampson error*. In addition the Sampson error, or distance, in the case of Essential matrix has a formulation even simpler. Let $\widehat{\mathbf{x}}'$ and $\widehat{\mathbf{x}}$ be two set of image features points expressed in normalized coordinates and $\mathbf{E}$ the Essential matrix that describe the correspondence condition between these set of points:

$$\widehat{\mathbf{x}}'\mathbf{E}\widehat{\mathbf{x}} = 0. \tag{3.30}$$

The Sampson distance associated to the i^{th} feature correspondence between $\widehat{x}'_i$ and $\widehat{x}_i$ is:

$$d_{Sampson} = \frac{(\widehat{x}'_i\mathbf{E}\widehat{x}_i)^2}{(\mathbf{E}\widehat{x}_i)^2_1 + (\mathbf{E}\widehat{x}_i)^2_2 + (\mathbf{E}^\top\widehat{x}'_i)^2_1 + (\mathbf{E}^\top\widehat{x}'_i)^2_2}, \tag{3.31}$$

where $(\mathbf{E}\widehat{x}_i)^2_j$ denotes the square of the j^{th} value of the vector $\mathbf{E}\widehat{x}_i$.

Table 3.1: Number N of samples required by the LMedS by assuming the 50% of outliers and the 0.99 probability to define a set without outliers.

Min. set of points	8	5	3	2	1
No. of samples	1177	145	35	16	7

The second aspect is the problem dimensionality. LMedS algorithm has to select a certain number N of samples to ensure that at least one of them is composed only by correct associations. The Equation 2.20 permits to compute this number and in Table 3.1 are represented the values associated to the typical number used in the VO algorithms. It is noticeable the rapid increase of the required samples with respect to the cardinality of the problem: between the eight-point and five-point algorithms there is almost an order of magnitude.

Using the planar motion assumption the cardinality of the problem is equal to 3 and so the number of evaluation is limited to 35. In addition the minimization can work on few parameters and so the final solution can be more accurate. For this reason it has been implemented the VO with the planar motion assumptions described in the previous section.

3.7 Incremental Localization

The incremental localization is the simplest approach that can be applied having a method capable to estimate the motion between two subsequent observations. The vehicle pose at the time t is defined with $\mathbf{X}_t$ and it is computed by updating the previous pose $\mathbf{X}_{t-1}$ with the estimation of the motion $\mathbf{H}_{t-1}^{t}$, as represented in Figure 3.10. By performing this operation recursively at each time t it is possible to obtain the incremental localization. The set of the past poses $\mathbf{X}_{0:t} = \{\mathbf{X}_0, ..., \mathbf{X}_t\}$ define the *path* followed by the vehicle. In the case of planar motion the estimation $\mathbf{H}_{t-1}^{t}$ is a trans-

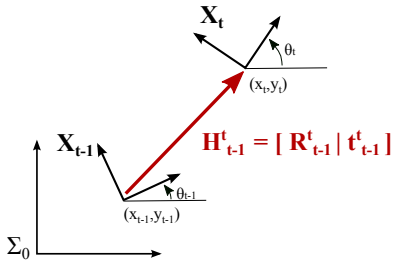


Figure 3.10: Represenation of the vehicle poses nomenclature in the incremental localization.

formation matrix composed by the bi-dimensional rotation matrix $\mathbf{R}_{t-1}^{t}$ and

translation vector $\mathbf{t}_{t-1}^{t}$:

$$\mathbf{R}_{t-1}^{t} = \begin{bmatrix} cos(\Delta\theta_{t-1}^{t}) & -sin(\Delta\theta_{t-1}^{t}) \\ sin(\Delta\theta_{t-1}^{t}) & cos(\Delta\theta_{t-1}^{t}) \end{bmatrix} \quad \mathbf{t}_{t-1}^{t} = \begin{bmatrix} \Delta x_{t-1}^{t} \\ \Delta y_{t-1}^{t} \end{bmatrix}, \tag{3.32}$$

in which $\Delta\mathbf{X}_{t-1}^{t} = [\Delta x_{t-1}^{t}, \Delta y_{t-1}^{t}, \Delta\theta_{t-1}^{t}]^{\top}$ is the pose variation described by the transformation matrix $\mathbf{H}_{t-1}^{t}$.

The pose updating is described by the Equation 2.24, in this case the formulation can be stated as:

$$\mathbf{X}_t = \begin{bmatrix} x_t \\ y_t \\ \theta_t \end{bmatrix} = \begin{bmatrix} x_{t-1} + \Delta x_{t-1}^{t} \cdot cos(\theta_{t-1}) - \Delta y_{t-1}^{t} \cdot sin(\theta_{t-1}) \\ y_{t-1} + \Delta x_{t-1}^{t} \cdot sin(\theta_{t-1}) + \Delta y_{t-1}^{t} \cdot cos(\theta_{t-1}) \\ \theta_{t-1} + \Delta\theta_{t-1}^{t} \end{bmatrix}. \tag{3.33}$$

The incremental estimation can be applied on the camera and on the laser scanner, since both the instruments can estimate the pose variation between two subsequent environment observation.

While the concatenation of the pose variations is a simple procedure, more difficult is to obtain a reliable estimation of the motion $\mathbf{H}_{t-1}^{t}$. Indeed the major drawback of the incremental localization is the error drift in the pose estimation. This problem is well known in the dead reckoning localization with encoder mounted on wheels, especially in the vehicle attitude as it can be seen in Figure 3.11.

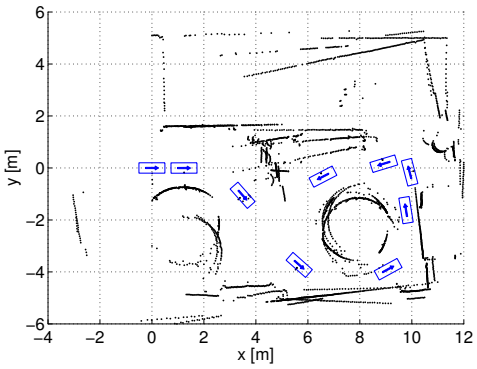


Figure 3.11: Dead reckoning localization with encoder on wheels with problems on attitude drift.

In order to limit the drift in the pose estimation error it is possible to apply at the scan matching and the visual odometry the LMedS approach robust improvements on the data association, presented respectively in Section 3.4 and Section 3.6. Previously it has been introduced the encoder odometry incremental localization: its performances are poor, but it can be very useful in the definition of the correspondences in the scan matching procedure and to estimate the scale factor in the visual odometry. By using the encoder odometry is also possible to implement a *Extended Kalman*

Filter (EKF) in order to fuse the information between this sensor and the localization based on the range sensors.

Concerning the incremental localization with the camera, the current limiting aspect is that to the features extracted from the images is not associated a position uncertainty. This aspect prevents the evaluation of the visual odometry uncertainty and so to apply the EKF. However the odometry is still important to compute the scale factor: only in this way the translation of the camera can be estimated.

On the other hand, regarding the incremental localization with the laser scanner, by using the scan matching approaches is possible to evaluate the uncertainty of the motion. This means that it is possible to apply the EKF in order to fuse the information on the vehicle pose. The implementation of the EKF depends on the vehicle kinematics, but in general it is composed by the *prediction* and *updating* steps. Let $\hat{\mathbf{x}}$ be the vehicle state vector, $\mathbf{P}$ the covariance associated to the state and $\mathbf{Q}$ the covariance of the control input $\mathbf{u}$; the prediction can be expressed by the relation:

$$\begin{aligned} \hat{\mathbf{x}}_{k|k-1} &= f(\hat{\mathbf{x}}_{k-1|k-1}, \mathbf{u}_{k-1}) \\ \mathbf{P}_{k|k-1} &= \mathbf{F}_{k-1}\mathbf{P}_{k-1|k-1}\mathbf{F}_{k-1}^{\top} + \mathbf{Q}_{k-1}, \end{aligned} \tag{3.34}$$

where $\hat{\mathbf{x}}_{k|k-1}$ and $\mathbf{P}_{k|k-1}$ are the predictions respectively of state and its uncertainty and $\mathbf{F}_{k-1}$ is the Jacobian of the function f used to compute the prediction. The encoder are associated to the input $\mathbf{u}_{k-1}$ with the uncertainty $\mathbf{Q}_{k-1}$. Furthermore, by using the encoder the vehicle state $\hat{\mathbf{x}}$ corresponds to the vehicle pose $\mathbf{X}$ and its covariance matrix $\mathbf{C}$ is equal to the state uncertainty $\mathbf{P}$. This prediction can be used to compute an initial pose estimation for the scan matching approaches which return a measurement $\mathbf{z}_k$ of the state with uncertainty $\mathbf{R}_k$. This measurement is used to update the vehicle state:

$$\begin{aligned} \hat{\mathbf{x}}_{k|k} &= \hat{\mathbf{x}}_{k|k-1} + \mathbf{K}_k\mathbf{y}_k \\ \mathbf{P}_{k|k} &= (\mathbf{I} - \mathbf{K}_k\mathbf{G}_k)\mathbf{P}_{k|k-1}, \end{aligned} \tag{3.35}$$

in which $\mathbf{G}_k$ is the Jacobian of the function g used to map the measurement to the state, and $\mathbf{y}_k = \mathbf{z}_k - g(\hat{\mathbf{x}}_{k|k-1})$ is defined as the innovation. The matrix $\mathbf{K}_k$ is the Kalman gain determined by the relation $\mathbf{K}_k = \mathbf{P}_{k|k-1}\mathbf{H}_k^{\top}\mathbf{S}_k^{-1}$ in which is used the innovation covariance $\mathbf{S}_k = \mathbf{G}_k\mathbf{P}_{k|k-1}\mathbf{G}_k^{\top} + \mathbf{R}_k$.

It should be noted that the scan matching approaches evaluate the transformation between two vehicle poses. The current vehicle pose, and so the measurement $\mathbf{z}_k$ used in the update step of the EKF, is computed incrementally from the best evaluation of the previous vehicle pose using the Equation 3.33. Very similar argument is also valid for the uncertainty $\mathbf{R}_k$, that must be computed using the formulation reported in Section 2.3.1. The best estimation of the previous vehicle pose on which apply the incremental solution is the solution returned by the EKF in the prior routine.

On the chance that the encoder estimation is not available, it is still possible to work with the laser. Indeed the scan matching approaches can

afford to work without the initial guess poses if the pose variation between the two laser scans is not excessive. With the aim to obtain a more reliable method is possible also in this case to use an EKF. The formulation is similar to the one reported above except that the input is not defined. The prediction can be inferred from the previous vehicle states. For this purpose the vehicle state, and also its uncertainty, used in the EKF must include the pose velocities:

$$\hat{\mathbf{x}} = [x, y, \theta, \dot{x}, \dot{y}, \dot{\theta}]^\top. \tag{3.36}$$

Using instead the camera without the encoder pose estimation, it is not possible to recover the scale factor. To perform the vehicle incremental localization with the camera there must be an exchange of information with the laser in order to evaluate the translation scale factor.

3.8 Global Localization with Maps

The global localization with maps lies outside the localization in unknown environment, but it is a good test field for the scan matching approaches. It has been developed during a collaboration on the preliminary analysis of localization systems for the *International Thermonuclear Experimental Reactor (ITER) Transfer Cask System (TCS)* [78].

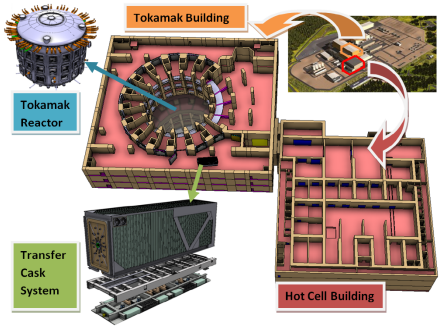


Figure 3.12: ITER scenario and TCS vehicle.

The procedure is very similar to the incremental localization: at each time t the vehicle pose $\mathbf{X}_t$ is estimated with a scan matching approach. The difference in the global localization with map is in the input used in the scan matching. Instead of comparing the current scan with the previous one, now the scan is compared with the map. Since the environment in which the TCS navigates it is structured, with a well defined geometry, the map representation is quite simple and can be derived from the CAD model of the building. The map thus defined is simply composed by a set of points connected by segments. These circumstances are perfect to apply the feature based scan matching approach: is also possible to implement

the dense scan matching approaches in order to compare the results of the different algorithms.

Concerning the feature based approach, it requires a pre-processing to identify the features and to store the information in a structure like the one proposed in Equation 3.1. Also regarding the PL approach the changes required are few since the map is already defined with a sequence of lines that can be used as reference surface S^{ref}. Few modifications are instead required from the PP. To accomplish this scan matching are necessary two set of points. The scan acquired from the laser scanner is already a set of points, the second one scan be evaluated by simulating the laser acquisition in a predicted pose by using the map. It is essential in this case to have a laser scanner simulator that can return this kind of information from the map.

Since using the map does not mean that all the correspondences in the scan matching approach are correct, also in this case has been implemented an ICP framework with a LMedS evaluation to detect the outliers in the association (Figure 3.13).

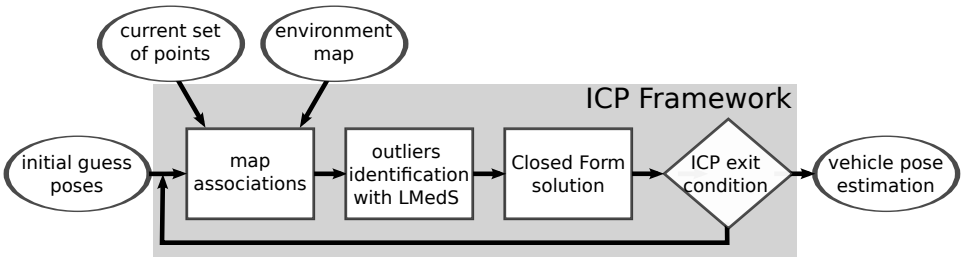


Figure 3.13: ICP framework used in the all the compared matching approaches, featured by a LMedS step for robustness improvement.

Like the incremental localization (Section 3.7) it has been implemented an EKF in order to generate the prediction of the vehicle pose and to use this information as initial pose estimation in the scan matching approaches. The implementation of the EKF, with or without the odometry, is perfectly the same of the one reported in Equation 3.34 for the prediction stage and in Equation 3.35 for the update. The main difference is in the information used in the measurement update. Since the scan matching is performed directly with information stated in the environment global reference frame, the pose estimation z_k is already the pose vehicle expressed in this reference frame. The same logic must be applied to the covariance matrix $\mathbf{R}_k$ which is the uncertainty of the global localization.

Furthermore, having a map it is also possible to evaluate the lower-bound uncertainty of the laser scan matching. Using the algorithm developed in [18] is possible to compute for each planned vehicle pose the *Fisher's information matrix* and so establish the *Cramer-Rao lower bound* of the unbiased

estimator of the vehicle pose $\mathbf{X}$.

3.9 Global Localization with Graph Theory

When there is no a-priori knowledge of the environment map the alternative for the localization is not just the incremental localization. In this cases it is possible to estimate the vehicle pose in a more reliable manner with the SLAM, which allows to localize the vehicle and build the map simultaneously. Moreover the SLAM approach meets exactly the requirements of the vehicle motion estimation in an unstructured scenario. As shown in Section 2.3.2 there are several implementations of the SLAM; the main difference is in the use of the information acquired during the navigation.

The *online SLAM* algorithm is usually based on a EKF, which is used to summarize all the previous information in the current pose and map estimation by means of a filtering operation. As reported in Section 2.3.3, this operation has limitations in the estimation accuracy of both the map and vehicle poses. In addition it requires a feature based map, since the EKF has to store and update a probabilistic map of the environment: the use of features permits to perform this operation in the easiest way.

On the contrary the *full SLAM* algorithm uses in each iteration all the previous information in terms of vehicle poses and sensor observations. Moreover it can be solved using the Graph Theory. Since the Graph Theory does not use filtering, it allows to fuse the data in a manner unattainable by the approaches based on the Bayesian formulation, as the EKF (Section 2.3.4.2). For these reasons in the thesis work has been implemented the global localization based on the Graph Theory. In addition, because of the method here proposed does not use expressly the map building, it is possible to use dense matching approaches and so carry out the localization also in a non-geometric environment.

The main drawback of the Graph Theory is that it solves only linear systems, while the global localization defines a non linear system. A lot of literature works have been produced on this topic (Section 2.3.4.3); in the following section is explained the approach here implemented.

3.9.1 Solve the non-linearity

The most relevant work in the graph-based SLAM has been proposed by Lu in [10]. In this early work the non-linearity problem was solved by linearising the vehicle motion equations in the neighbourhood of the initial vehicle pose estimation. The resulting approach was not robust, but the implemented algorithms were a fundamental step in the solution of graphs with uncertainty weighting (Section 2.3.4.1). A more recent work [71] dealt with the solution of the non-linearity problem, but using a less rigorous

probabilistic discussion. The main innovation of this algorithm is in the approach used in the linearisation: a *Two-Step* pose estimation.

In order to understand how it works it must be considered two vehicle poses $\mathbf{X}_i$ and $\mathbf{X}_j$ used in the global localization, which correspond two nodes of the graph. By performing the matching of the information acquired by the sensors in these poses is possible to estimate the pose variation $\mathbf{D}_{ij}$. The formulation that describes the relation between these entities is the one already used in the incremental localization. Defining $\mathbf{X}_i = [x_i, y_i, \theta_i]^\top$ the pose of the i^{th} node and $\mathbf{X}_j = [x_j, y_j, \theta_j]^\top$ the pose of the j^{th} node, the system of equations that describe their relation, given the pose variation $\mathbf{D}_{ij} = [\Delta x_{ij}, \Delta y_{ij}, \Delta\theta_{ij}]^\top$, can be stated as follow:

$$\mathbf{X}_i = \begin{bmatrix} x_i \\ y_i \\ \theta_i \end{bmatrix} = \begin{bmatrix} x_j + \Delta x_{ij} \cdot cos(\theta_j) - \Delta y_{ij} \cdot sin(\theta_j) \\ y_j + \Delta x_{ij} \cdot sin(\theta_j) + \Delta y_{ij} \cdot cos(\theta_j) \\ \theta_j + \Delta\theta_{ij} \end{bmatrix}. \tag{3.37}$$

In this system of equations it is noticeable that while the relations that describe the vehicle position $\mathbf{X}' = [x, y]^\top$ are non-linear, the relation between the vehicle attitudes is linear. By assuming that attitude and position of the vehicle can be solved separately, it is possible to linearise all these relations. Indeed, once that the vehicle attitude is solved, the position relations become linear. In conclusion the concept is extremely simple: the linearisation can be done by solving the problem in two step, one for the rotation and the other one for the translation. Extending this approach to the whole graph, it means that there are two system of linear equation, that can be solved easily with the Graph Theory.

In order to perform this two step evaluation, it must be assumed that the the matching errors in $\mathbf{X}'_i = [x_i, y_i]^\top$ and θ_i are not correlated. This means that the covariance matrix $\mathbf{C}_i$ associated to the generic pose $\mathbf{X}_i$ returned by the global localization has a block diagonal form:

$$\mathbf{C_i} = \begin{bmatrix} c_{xx} & c_{xy} & 0 \\ c_{yx} & c_{yy} & 0 \\ 0 & 0 & C_{\theta\theta} \end{bmatrix}_{x=x_i, y=x_i, \theta=\theta_i}. \tag{3.38}$$

This is clearly a simplification, because of the vehicle attitude and position are not independents; on the other hand the alternative solution is to linearise the equations that, how it has been reported previously, probably leads to worst results. It mus be notated that also by using classical probabilistic approaches in the SLAM there are simplifying assumptions in the uncertainties evaluation, because only serial and parallel links can be treated.

According to this two-step implementation, the first operation is to accomplish the global optimization of the pose attitude. Let $\mathbf{X}_{0:m} = \{\mathbf{X}_0, ..., \mathbf{X}_m\}$

be the set of m poses on which apply the global optimization. To the attitudes of these poses can be applied the system of equation defined by the graph and reported in Equation 2.32:

$$\Delta\theta = \mathbf{M} \cdot \theta, \tag{3.39}$$

where $\Delta\theta$ is the vector of the relative attitude between the graph poses, θ is the vector of the vehicle global attitudes referred to the reference frame of the pose $\mathbf{X}_0$ and $\mathbf{M}$ is the incidence matrix:

$$\begin{matrix} \Delta\theta = [\Delta\theta_{01}, ..., \Delta\theta_{ij}, ..., \Delta\theta_{(m-1)m}]^\top \\ \theta = [\theta_0, ..., \theta_m]^\top \end{matrix}, \tag{3.40}$$

in which the complete vector $\Delta\theta$ has dimension $n = \frac{m(m-1)}{2}$. The incidence matrix $\mathbf{M}$ is the connection between the measurements $\Delta\theta$ and the global estimation of attitude θ, so its dimensions are $n \times m$.

The system in Equation 3.39 can be solved by taking into account the standard deviation $\sigma_{\Delta\theta_{ij}}$ associated to the relative attitude evaluation between the i^{th} and j^{th} node. This values are stored along the diagonal of the matrix $\mathbf{C}_{\Delta\theta}$ that has dimension $n \times n$:

$$\mathbf{C}_{\Delta\theta} = \begin{bmatrix} \sigma^2_{\Delta\theta_{01}} & & & 0 \\ & \cdots & & \\ & & \cdots & \\ 0 & & & \sigma^2_{\Delta\theta_{(m-1)m}} \end{bmatrix}, \tag{3.41}$$

By applying the Equations 2.34 and 2.35 is possible to compute the global vehicle attitude:

$$\theta = (\mathbf{M}^\top {\mathbf{C}_{\Delta\theta}}^{-1} \mathbf{M})^{-1} \mathbf{M}^\top {\mathbf{C}_{\Delta\theta}}^{-1} \Delta\theta, \tag{3.42}$$

and its uncertainty:

$$\mathbf{C}_\theta = (\mathbf{M}^\top {\mathbf{C}_{\Delta\theta}}^{-1} \mathbf{M})^{-1}. \tag{3.43}$$

Once the global attitude is evaluated, it is possible to optimize the vehicle position. To obtain linear relations in the position terms, it must be applied the global attitude to all the relative position evaluations defined for each link. This means the definition of a new set of relative position evaluations from the original set of measurements $\mathbf{D}_{ij} = [\Delta x_{ij}, \Delta y_{ij}, \Delta\theta_{ij}]^\top$. The new set can be defined as:

$$\Delta\mathbf{X}'_{ij} = \begin{bmatrix} \Delta x'_{ij} \\ \Delta y'_{ij} \end{bmatrix} = \begin{bmatrix} \Delta x_{ij} cos(\theta_i) - \Delta y_{ij} sin(\theta_i) \\ \Delta x_{ij} sin(\theta_i) + \Delta y_{ij} cos(\theta_i) \end{bmatrix}. \tag{3.44}$$

The matrix $\mathbf{M}$ must be reshaped in order to be used with the position. This not due to changes of the graph links, but is to deal with the dimensionality variation: the vehicle attitude θ_i is a scalar value while the position $\mathbf{X}'_i$ is

a vector. The new incidence matrix is defined as $\mathbf{M}'$ and it has $2n \times 2m$ dimensions.

In addition it must be defined the matrix $\mathbf{C}_{\Delta\mathbf{X}'}$. It is a diagonal block matrix: each block is the covariance matrix associated to the relative position estimation between the nodes of the graph, and its dimensions are $2n \times 2n$.

$$\mathbf{C}_{\Delta\mathbf{X}'} = \begin{bmatrix} \mathbf{C}_{\Delta\mathbf{X}'_{01}} & & 0 \\ & \cdots & \\ & & \cdots \\ 0 & & \mathbf{C}_{\Delta\mathbf{X}'_{(m-1)m}} \end{bmatrix}. \tag{3.45}$$

The system of linear equation that must be solved to evaluate the vehicle global positions $\mathbf{X}'$ is:

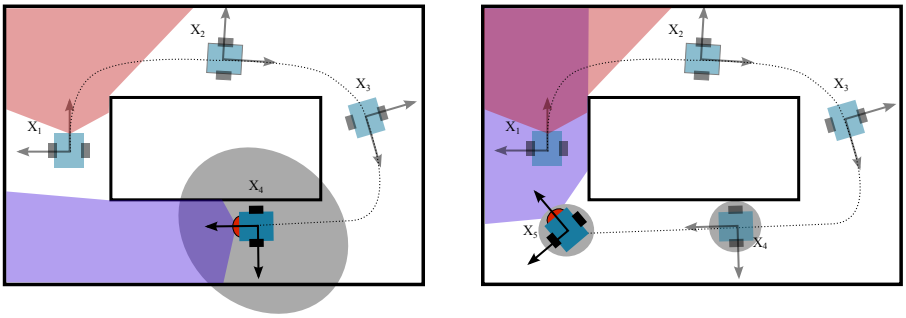
$$\mathbf{X}' = \mathbf{M}' \cdot \Delta\mathbf{X}'. \tag{3.46}$$

The global optimized vehicle position $\mathbf{X}'$ and its uncertainty $\mathbf{C}_{\mathbf{X}'}$ are achieved by solving the graph with uncertainty weighting:

$$\mathbf{X}' = ({\mathbf{M}'}^{\top}{\mathbf{C}_{\Delta\mathbf{X}'}}^{-1}\mathbf{M}')^{-1}{\mathbf{M}'}^{\top}{\mathbf{C}_{\Delta\mathbf{X}'}}^{-1}\Delta\mathbf{X}' \tag{3.47}$$

$$\mathbf{C}_{\mathbf{X}'} = ({\mathbf{M}'}^{\top}{\mathbf{C}_{\Delta\mathbf{X}'}}^{-1}\mathbf{M}')^{-1}. \tag{3.48}$$

At the end of this operations, thanks to the two-step linearisation approach, it is possible to obtain the global optimization of the vehicle pose without using numerical algorithms. Furthermore, the method does not require the building of a map.



(a) open loop uncertainties (b) close loop uncertainties

Figure 3.14: Graphical representation of vehicle uncertainties in SLAM. (a) In open loop estimation vehicle and map uncertainties grow up at at each step. When it is possible to close the loop (b), uncertainties of vehicle position and map decrease suddenly.

Summing up all the observations pointed out until this point, some important considerations on this global localization implementation can be carried out. The first one is regarding the pose connections in the graph.

The graph solution perform a Total Least Square minimization by considering all the possible walks. This means that for each node of the graph the error of all the measurements, also the indirect one, is minimized by considering also the uncertainties and their propagation along the walk. Intuitively the graph can be compared to a network with elastic connections between the nodes: low uncertainty pose estimation between two nodes correspond to a stiffer connection. The global optimization accomplished from the graph based SLAM is equivalent to find the minimum energy configuration of the network. For this reason, unlike EKF-SLAM, the proposed method can modify the previous pose estimations in order to achieve to the best solution. Maybe this feature seems to be unsafe, but instead it is a strength point: in open loop, both classical and graph based SLAM can have error drift in pose estimation. When a new observation allows to close the loop (Figure 3.14), with a graph approach all the past poses can be updated in order to estimate the more reliable vehicle path. With an EKF-SLAM instead, all the past poses are summarized with the filtering in a prior. The new observation can not fix previous poses and map: this means that the map accuracy decrease and the uncertainties of all the poses of the path are greater than the one estimated with a graph based SLAM

Another important observation is on computational cost. As seen in Section 2.3.3, in order to obtain a reliable pose estimation is more important to have an high number of landmarks than vehicle poses. It means that the full SLAM can be performed only on same key-frame without significant effects on pose uncertainty. Furthermore, using dense matching approaches is possible to increase the number of natural landmarks used in each association.

3.9.2 Pose uncertainty evaluation

By applying the proposed method for linearisation, since the estimation of the global pose has been split in two steps, the Graph Theory has to be performed twice, one for the attitudes and one for the positions of the vehicle. This means that in the solution procedure, the attitude θ and the position $\mathbf{X}' = [x, y]^\top$ are assumed independents. The independence of this parameters helps to linearise the relations. This assumption affects also the estimation of the pose uncertainty: at the conclusion of the graph based approach the vehicle pose uncertainty is described by a block diagonal covariance matrix.

Assuming that the pose of i^{th} node can be notated as $X_i = [x_i, y_i, \theta_i]^\top$, the covariance matrix after global localization must have the form:

$$\mathbf{C}_i = \begin{bmatrix} c_{x_i,x_i} & c_{x_i,y_i} & 0 \\ c_{y_i,x_i} & c_{y_i,y_i} & 0 \\ 0 & 0 & c_{\theta_i,\theta_i} \end{bmatrix}, \tag{3.49}$$

where the first and the second diagonal block are evaluated respectively from

the second and first step of the graph based SLAM linearisation procedure. The zeros out of diagonal blocks indicate that the vehicle position and attitude are uncorrelated. To compose the covariance matrix $\mathbf{C}_i$ defined above, with the assumption of independence between position and attitude, it could be performed a simply operation of matrix composition:

$$\mathbf{C}_i = \begin{bmatrix} \mathbf{C}_{\mathbf{X}'_i} & 0 \\ 0 & \mathbf{C}_{\theta_i} \end{bmatrix}. \tag{3.50}$$

The covariances $\mathbf{C}_{\mathbf{X}'_i}$ and $\mathbf{C}_{\theta_i}$ must be computed from the covariance matrices returned by the graph based SLAM. Actually the problem is that these covariance matrices are associated to the parameters of all the pose variables in the graph, while the covariances needed are only the one referred to the pose parameters of the single node. This issue can be solved with a *marginalization* of the distribution on a node.

The marginal density of a generic scalar value x on the a^{th} node is by definition the integration of all the joint densities $p(x_a, x_b)$ on the other b^{th} nodes:

$$p(x_a) = \int p(x_a, x_b) dx_b. \tag{3.51}$$

This formulation is not always easy to use and it could be unusable for a large number of variables: it is the case of the uncertainties returned from the graph solution. Assuming that the uncertainties estimated by the graph can be represented with a multivariate Gaussian, the integral formulation can be simplified. As demonstrated by Schon [79], in this case the computation of the covariance can be done very easily. Let $\mathbf{x}$ be a random vector with mean μ and covariance $\mathbf{\Sigma}$ and assume that the vector can be partitioned in $\mathbf{x}_a$ and $\mathbf{x}_b$ as follow:

$$\mathbf{x} = \begin{pmatrix} \mathbf{x}_a \\ \mathbf{x}_b \end{pmatrix}, \quad \mu = \begin{pmatrix} \mu_a \\ \mu_b \end{pmatrix}, \quad \mathbf{\Sigma} = \begin{pmatrix} \mathbf{\Sigma_{aa}} & \mathbf{\Sigma_{ab}} \\ \mathbf{\Sigma_{ba}} & \mathbf{\Sigma_{bb}} \end{pmatrix}. \tag{3.52}$$

By marginalizing the Gaussian distribution is possible to obtain the uncertainty distribution $p(\mathbf{x_a})$ of the partition $\mathbf{x}_a$:

$$p(\mathbf{x}_a) = \mathcal{N}(\mathbf{x}_a; \mu_a, \mathbf{\Sigma}_{aa}). \tag{3.53}$$

Identical approach can be applied to the covariance matrices of vehicle position $\mathbf{C}_{\mathbf{X}'}$ and attitude $\mathbf{C}_\theta$ estimated by the graph based SLAM. Assuming that these uncertainties can be represented by a multivariate Gaussian, the vectors of position and attitude can be partitioned on the values of each node. Thus is possible to marginalize the uncertainties of all the poses in the path in order to evaluate the covariance matrices $\mathbf{C}_{\mathbf{X}'_i}$ and $\mathbf{C}_{\theta'_i}$ associated to the generic $\mathbf{X}_i$ pose of the path.

Chapter 4

Experimental Results

In order to test the proposed algorithms under different environment specifications, in this work has been used several instrumentations mounted on different research vehicles and a collection of standard datasets well known in the field of vehicle localization. Since the vehicle localization is a field still in evolution, the most difficult part is the assessment of the results. Actually the problem is to achieve the information from a more accurate measurement system in such a manner to have a ground truth: for this reason the methods have been also tested on simulated data.

4.1 Instrumentation

4.1.1 Simulations and Datasets

The simulations and the datasets have been used to test and qualify the localization algorithms. For the simulation of the global localization with maps procedure has been used the simulator included in the software *RobDAC* developed by the institute IPNF of the IST of Lisbon [72]. The software has been modified in order to simulate the typical sources of noise in the laser scanner and to allow the definition of several types of laser scanner, in order to investigate on the performance of the different scan matching approaches. This laser simulator developed was also used in the creation of the simulation used in the evaluation of the incremental and global localization in unknown environment.

In order to perform a comparison of the performances between the developed algorithms and other works on the same topic has been used the MRPT dataset collection [80] of experimental measurements. Unfortunately, also in this case most of the data acquired is released without the reference of a ground truth: however it was useful to evaluate qualitatively the performance of the algorithms and to test them in several kind of environment.

4.1.2 Test vehicle

The test vehicle used, shown in Figure 4.1, is a research mobile robots MobileRobots Pioneer P-3DX. It is a differential drive vehicle for indoor appli-

Figure 4.1: Mobile Robots Pioneer P3-DX test vehicle equipped with lasers and camera.

cations, with several on-board sensors like wheel encoders, gyroscope and an array of sonar. The one used during the thesis has been also equipped with an on-board computer needful to acquire the data returned from the opto-electronic instrumentation. As depicted in Figure 4.1, to the vehicle has been added two laser scanners at different height and a camera. The laser scanners used are respectively a Sick S2000 on the bottom and a Sick LMS151 on the top: since they have different specifications it was possible to test the algorithms with different laser profiles. The camera used was industrial graded, and in particular it is a Matrix Vision BlueFox with a 1024×768 CCD sensor and an optical with $f = 8mm$ focal length.

The Sick LMS151 laser scanner is also used in the proprietary navigation software of the Pioneer P3-DX. In order to reference all the information acquired by the Sick S200 laser scanner to the vehicle pose it was used a software developed from our laboratory, based on the scan matching procedures presented on this thesis. In this way it has been also possible to calibrated the extrinsic parameters of the camera after a camera-laser calibration [81].

4.1.3 ITER vehicle prototype

The ITER vehicle prototype shown in Figure 4.2 is a model of the Cask and Plug Remote Handling System (CPRHS) used in the mock-up of the Tokamak Building of the ITER project to test different algorithms of vehicle localization, navigation and path planning. It is a rhombic like vehicle,

Figure 4.2: ITER model vehicle prototype equipped with laser and camera.

or rather a two-steering-wheels AGV which navigates inside the mock-up environment by collecting the measurement from a Hokuyo URG-04LX-UG01 laser scanner and a common webcam. The vehicle is equipped with a small micro controller in which is implemented the command of the drives In order to collect the data from the two sensors it has been used a laptop with the RobDAC software.

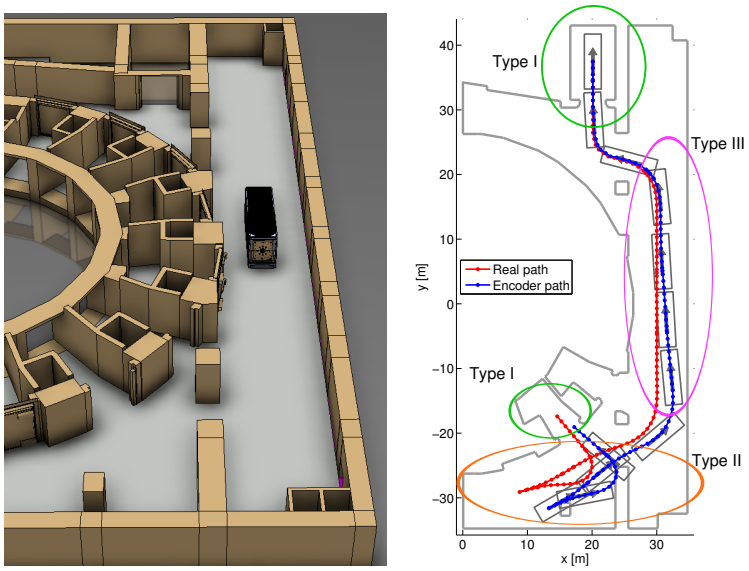
4.2 Global Localization with Maps

The first results reported is the global localization with map. Although this type of localization requires a complete knowledge of the environment, since the map is available, it allows to test the performances of the scan matching approaches when they are used in a localization framework; in this way it is possible to achieve a complete overview of their pro and cons.

The localization is performed in a representative environment that includes several type of situations commonly encountered by an indoor mobile robot, some of which can be also critical, like the one depicted in Figure 4.3. This environment has narrow areas with high symmetry and few features, wide areas with good asymmetry and number of features and areas that are very closed to critical situations.

In order to investigate on the performance of the scan matching methods in a localization framework, several laser profiles in terms of sensor performances have been tested. The considered parameters are:

- maximum distance detectable (10, 20, 100 m)
- standard deviation of noise in range (0.04, 0.1 m)



(a) 3D CAD model of the environment

(b) Map representation of the environment

Figure 4.3: CAD model of the ITER environment (a) and relative map, with real path and dead-reckoning estimation. The three type of scenarios are (I) narrow with few features, (II) wide with several features and (III) featureless and thus closed to undetermined situation. The vehicle starts from the room on the top.

- angular field of view (180° and 240° deg)
- angular resolution (0.25°, 0.5°, 1° deg)

The selection of the parameter values is based on common values of commercial lasers: while the last three parameters are more general in the scope of the present comparison analysis, the maximum measurable range must be related to the environment size. Actually the chosen values are respectively about 0.1, 0.2 and 1 times the maximum distance measurable in the environment. Since it is not feasible to change all these parameters in a real LRF and we didn't avail of an highly accurate ground-truth of the vehicle trajectory, a first evaluation has been carried out in a controlled simulation. All the combinations of these parameters define 36 different laser profiles, defined here *ID* (Table 4.1), based on which the localization procedures based on the three different scan matching approaches have been test. Two tests are provided:

- with an initial guess pose estimation given by a simulated odometry
- without any initial guess pose estimation

For both the approaches an Extended Kalman Filter (KF) is used : in the former, which is the classical approach, the odometric estimation can be used as an input control in the prediction step of the KF. In the latter the prediction must be inferred from the past vehicle states. In the case without odometry data a EKF assuming constant velocities (linear and angular) model is used. The velocity state is estimated from the observations of incremental displacements obtained by matching consecutive scans. Furthermore the localization methods have been also tested in a scale model of the real scenario.

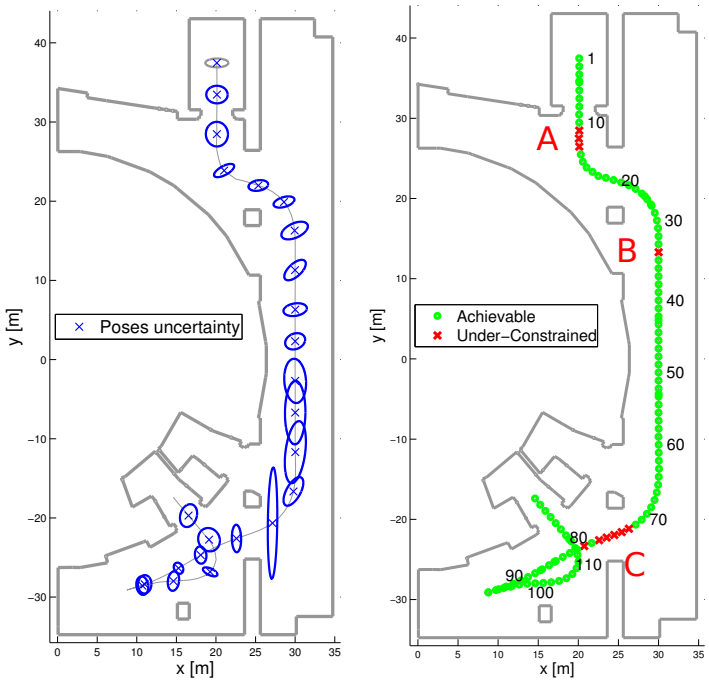
Simulation and experimental results share the same scenario (Figure 4.3), except for the 1 : 25 scale factor of the experimental mock-up . These analysis has been complemented with the analysis of the theoretical limitations of the environment, performing an *uncertainty lower bound* computation along the path. Having the best theoretical performance of the pose estimation at hand, the comparative analysis gains more generality. With the knowledge of the map, path and laser profile, using the method reported in [18] the Fisher Information Matrix is computed and thus the minimum achievable uncertainty on pose estimation as well. In addition, by observing the Fisher Information Matrix, it is possible to detect the *undetermined situations*. In these cases, like corridor or circular environment, is not possible to obtain a solution from the scan matching with any approach. In Figure 4.4(a) is depicted the analysis of the best achievable uncertainty for profile ID 29. It is observable how the uncertainty on vehicle pose increases along the corridor. In addition, by only changing the laser angular resolution (from profile ID 29 to ID 17) the poses with high uncertainty degenerate in undetermined situations, as represented in Figure 4.4(b) by the regions A, B an C.

4.2.1 Simulated results

The simulation results are reported in (Table 4.1). The best laser profile in terms of measurement noise, resolution, range and field of view, can be identified in the ID 36, while the worst one is ID 1. Since reporting the results for all the IDs can be useless, the results shown below are a selection of the whole laser profiles set, in order to emphasize the most relevant comparisons.

In Table 4.2 and 4.3 are reported the results of the simulation. For some IDs results are missing because these are cases where localization fails. This problem can occur when for several consecutive times the amount of data in the scans are not sufficient to localize the vehicle. In these situations the KF can diverge and the localization gives rise to a *kidnapped robot problem*. The solution of this problem is not the aim of this application, which is instead the evaluation of the scan matching approaches.

Among the simulations with a guess pose estimation the best results in terms of mean and maximum error correspond, as expected, to the best laser profile ID 36 (Table 4.2). By an insight into this profile (Figure 4.5) it is



(a) Path uncertainty lower bound on profile ID 29

(b) Path undetermined situations on profile ID 17

Figure 4.4: Vehicle pose uncertainty lower bound (magnified by 100) (a) and undetermined situations (A,B,C) of a lower angular resolution laser scanner (b). The numbers represent the progressive scan number along the path.

Table 4.1: Laser scanner profile IDs used in the simulation.

ID	angular FOV [deg]	angular resolution [deg]	range max [m]	range noise std [m]
1	180	1	10	0.1
5	180	1	20	0.1
9	180	1	100	0.1
10	240	1	100	0.1
12	240	1	100	0.04
17	180	0.5	20	0.1
24	240	0.5	100	0.04
27	180	0.25	10	0.04
28	240	0.25	10	0.04
29	180	0.25	20	0.1
32	240	0.25	20	0.04
34	240	0.25	100	0.1
35	180	0.25	100	0.04
36	240	0.25	100	0.04

clear that the most accurate approach is the PL, that it is capable to estimate the position with a mean error lower than 0.01 m in the best case, also when the localization is performed without the initial guess pose estimation

Table 4.2: Errors in the case of simulated vehicle pose estimation using odometric data as initial guess.

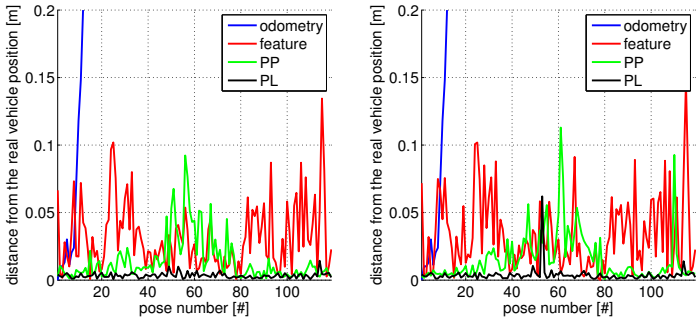
Errors in vehicle position/attitude estimation *with* initial guess pose estimation						
ID	**Feature**		**PP**		**PL**	
	mean	*max*	*mean*	*max*	*mean*	*max*
	[m]/[deg]	[m]/[deg]	[m]/[deg]	[m]/[deg]	[m]/[deg]	[m]/[deg]
1	0.24/1.19	1.06/9.03	0.07/0.30	0.39/2.85	0.05/0.33	1.38/6.12
5	0.15/0.65	0.51/5.38	0.07/0.27	0.28/0.98	0.02/0.09	0.09/0.76
9	0.16/0.54	1.10/4.54	0.08/0.31	0.56/1.19	0.02/0.06	0.11/0.68
10	0.11/0.43	0.35/6.06	0.05/0.26	0.28/0.70	0.02/0.06	0.12/1.55
12	0.05/0.21	0.16/1.63	0.06/0.31	0.27/1.13	0.01/0.02	0.03/0.23
17	0.47/1.90	2.15/7.97	0.03/0.11	0.11/0.41	0.02/0.07	0.13/0.82
24	0.04/0.10	0.30/1.00	0.03/0.16	0.14/0.57	0.01/0.01	0.02/0.09
27	0.22/1.82	1.02/18.2	0.02/0.09	0.24/0.64	0.01/0.10	0.09/2.68
28	-	-	0.02/0.07	0.13/0.81	0.01/0.04	0.05/0.94
29	0.10/0.48	0.38/2.16	0.01/0.05	0.09/0.32	0.01/0.04	0.07/0.26
32	0.04/0.08	0.18/0.41	0.01/0.05	0.04/0.25	0.01/0.03	0.03/1.29
34	0.07/0.15	0.42/1.29	0.01/0.07	0.07/0.28	0.01/0.04	0.04/1.55
35	0.04/0.12	0.17/1.44	0.02/0.09	0.17/0.33	0.01/0.02	0.05/0.98
36	0.03/0.07	0.13/0.54	0.01/0.08	0.09/0.29	0.01/0.01	0.01/0.04

Table 4.3: Errors in vehicle pose estimation in simulation with respect to the ground-truth, without initial guess pose estimation.

Errors in vehicle position/attitude estimation *without* initial guess pose estimation						
ID	**Feature**		**PP**		**PL**	
	mean	*max*	*mean*	*max*	*mean*	*max*
	[m]/[deg]	[m]/[deg]	[m]/[deg]	[m]/[deg]	[m]/[deg]	[m]/[deg]
1	-	-	-	-	-	-
5	-	-	-	-	0.06/0.08	1.95/0.57
9	0.24/0.70	1.73/4.55	0.10/0.31	0.57/1.25	0.03/0.06	0.37/0.65
10	0.18/0.59	1.71/10.9	0.06/0.25	0.25/0.70	0.02/0.05	0.31/0.33
12	0.07/0.26	0.50/2.82	0.07/0.28	0.43/0.73	0.01/0.03	0.04/0.25
17	-	-	-	-	-	-
24	0.04/0.11	0.35/0.66	0.03/0.15	0.20/0.45	0.01/0.01	0.01/0.08
27	-	-	-	-	-	-
28	-	-	-	-	-	-
29	-	-	4.17/19.5	19.7/122	0.05/0.04	1.08/0.26
32	0.06/0.10	1.92/1.11	-	-	0.01/0.01	0.87/0.07
34	0.08/0.18	0.29/1.45	0.02/0.06	0.11/0.33	0.01/0.02	0.13/0.14
35	0.05/0.15	0.35/1.51	0.05/0.11	1.01/2.08	0.01/0.01	0.03/0.12
36	0.03/0.06	0.15/0.42	0.02/0.07	0.11/0.23	0.01/0.01	0.06/0.05

(Table 4.3). The PP approach ensures good results in terms of accuracy but less in precision. Despite of the premises, the worst approach is the feature-based in terms of both mean and maximum error. Even though the features have a better statistical implementation, their limited numbers and high uncertainties lead to results worse than dense matching approaches. Theoretically the features should be the best approach, since their identification is done with a probabilistic inference on noisy data in a well structured environment. However the number of features detectable is small if compared

with the number of points in each scan. In addition, due to the environment size, the features are also far from the laser, in areas where the scan resolution is low: this entails the extraction of features with high uncertainties that consequently affect the pose estimation.



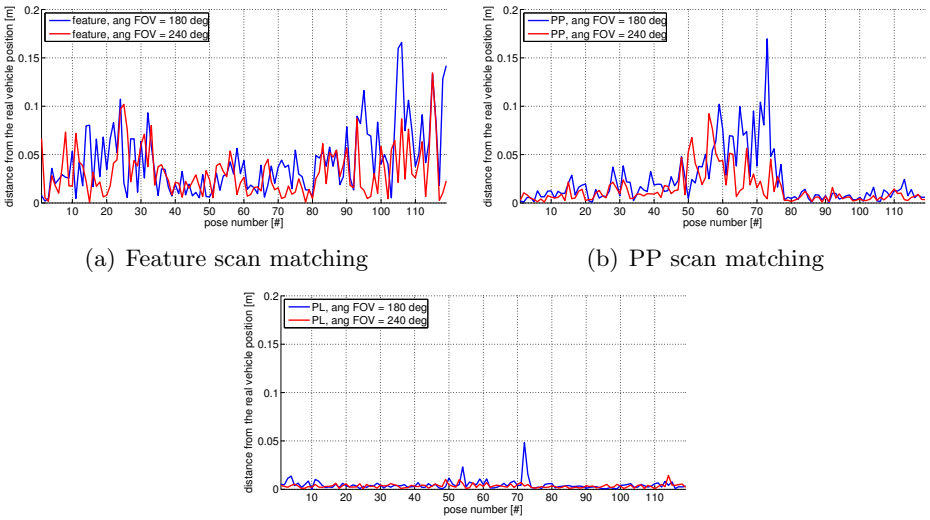
(a) with initial guess pose estimation

(b) without initial guess pose estimation

Figure 4.5: Error in vehicle position estimation along the path using the laser profile ID 36.

Next consideration is on the angular field of view: by increasing it, the error in position must decrease, especially in environments where there are several undetermined situations like corridors. This is noticeable by comparing the results on IDs (32,35,36): since the IDs 35 and 36 can avail of an high maximum range, the ID 32 with a wide 240°deg field of view but lower maximum range has comparable performances. Furthermore it can be observed in Figure 4.6 that the angular FOV is really important to avoid the undetermined in the region C, in the neighbourhood of pose number 70.

Regarding the laser maximum range, with reference to IDs (36,32,28), a degradation in accuracy by lowering the range is remarkable . Analysing the overall performances along the path, while the localization based on dense matching with initial guess pose estimation work properly, in the feature-based case the localization fails in some parts of the path. This problem is represented in Figure 4.7(a), while the comparison of the methods over all the laser maximum ranges is depicted in Figure 4.8. The feature-based approach fails on the pose 40, at the entry of corridor. This is due to a very short range (10 meters), which does not allow to identify the minimum number of features required. On the same IDs, without using the initial guess pose estimation, the results with short range laser become worst: all the approaches fail. While the feature-based and PP approach stop to work correctly on the undetermined region B, the PL continues to localize the vehicle until the region C: the PL approach can also work in region very closed to the undetermined situations. This is in accordance



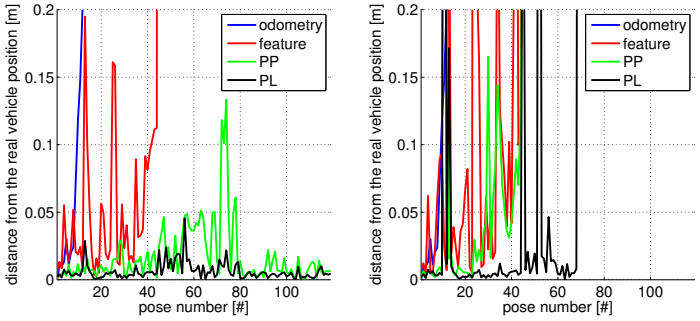
(a) Feature scan matching

(b) PP scan matching

(c) PL scan matching

Figure 4.6: Error in vehicle position estimation along the path on localizations performed with different laser *angular FOV* with initial guess pose estimation.

with a maximum error of PL in most of the case lower than in the other approaches.

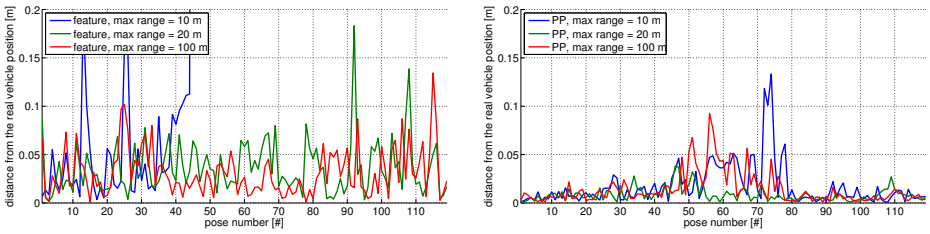


(a) with initial guess pose estimation

(b) without initial guess pose estimation

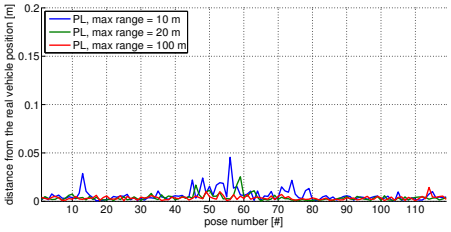
Figure 4.7: Error in vehicle position estimation along the path on localizations performed with laser profile ID 28.

Comparing the IDs (36,24,12) is possible to analyse the influence on the performances by changing the laser angular resolution: the improvements on using high resolution are poor if compared with the other profile parameters, especially for the PL approach (Figure 4.9).



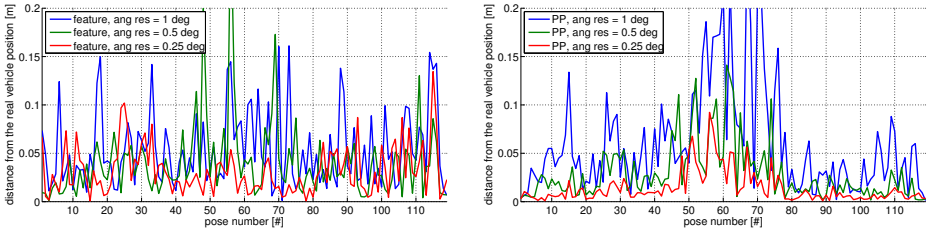
(a) Feature scan matching

(b) PP scan matching



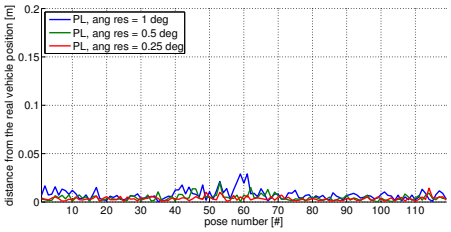
(c) PL scan matching

Figure 4.8: Error in vehicle position estimation along the path on localizations performed with different laser *maximum range* with initial guess pose estimation.



(a) Feature scan matching

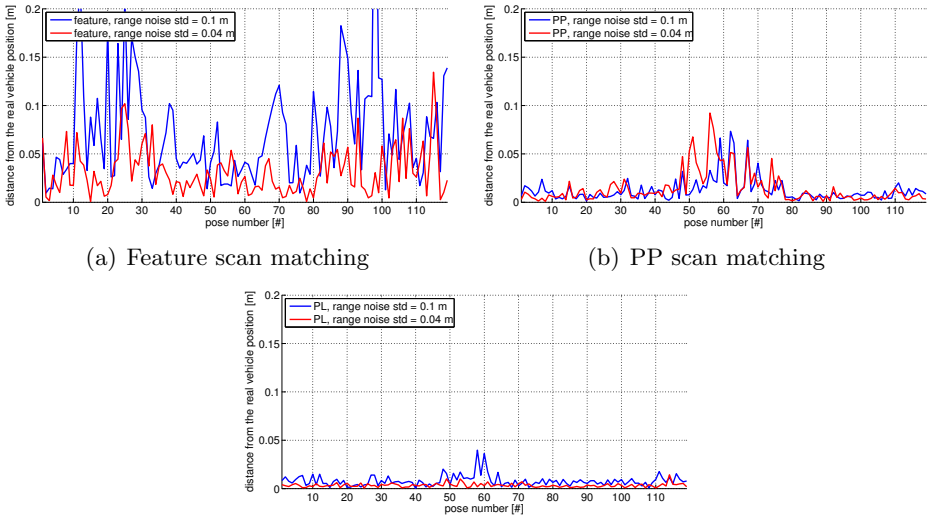
(b) PP scan matching



(c) PL scan matching

Figure 4.9: Error in vehicle position estimation along the path on localizations performed with different laser *angular resolution* with initial guess pose estimation.

Last comparison can be done on the laser range noise, by observing the IDs 36 and 34. In Figure 4.10 the same comparison is reported in a graphical manner. While the performances improvement on the dense matching approaches is limited, in the feature based approach it is worth using sensors with low noise. This is in agreement to what explained above about feature extraction uncertainty.



(a) Feature scan matching

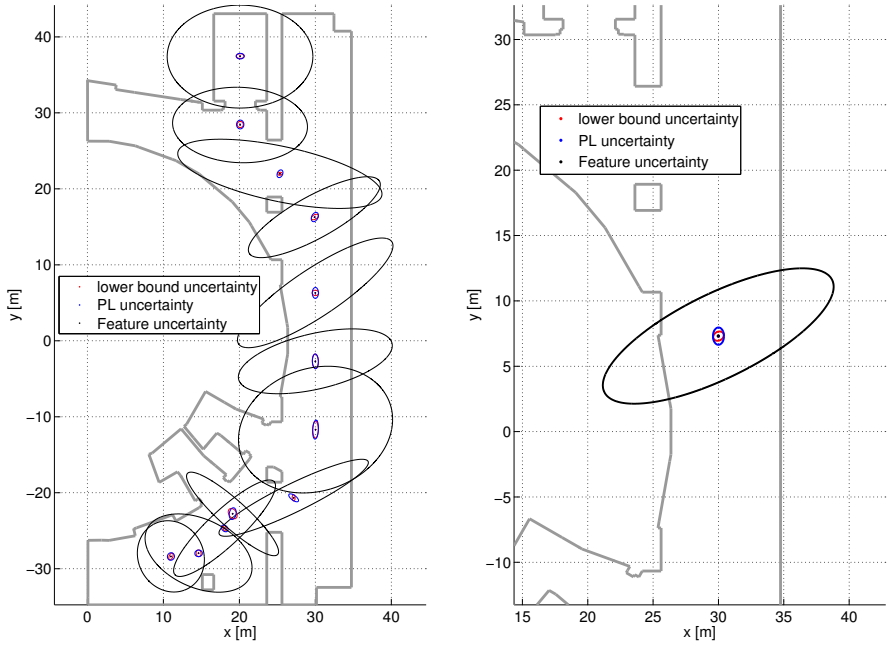
(b) PP scan matching

(c) PL scan matching

Figure 4.10: Error in vehicle position estimation along the path on localizations performed with different laser *range noise* with initial guess pose estimation.

Few comments can also be done on the worst configurations. It is important to denote that in these cases by using the odometry initial guess pose estimation to localize the vehicle is possible in almost all the profiles. Instead, without initial guess pose estimation and bad laser profiles, the number of undetermined situations (or rather the ones with the highest lower bound uncertainty) is high, and since they are located in some key-points of the path (Figure 4.4) the EKF can easily diverge.

The good performances of the PL approach are also proved by the uncertainty comparison reported in Figure 4.11. Its uncertainty estimation is close to the lower-bound along all the path. Instead the uncertainty evaluate by the feature based scan matching approach is higher than the lower-bound, and justify also the performances discussed above.



(a) Path uncertainty

(b) Uncertainty on pose 40

Figure 4.11: Uncertainty comparison (magnified by 200) between the lower-bound estimation and the scan matching approaches on profile ID 36.

4.2.2 Experimental Results

The evaluation of the experimental data have been performed in a qualitative manner, since a precise ground truth was not available. Nevertheless, importantly, it is possible to put in evidence some interesting considerations in the scope of this applications, coming from the comparison between simulations and real world tests. In the experimental results a single laser profile have been used: the *Hokuyo URG-04LX-UG01*. This laser scanner has 5.6 m maximum range, 0.352 deg angular resolution and a 240 deg field of view. Because the navigation area has a maximum length of about 3.5 m, its maximum range is enough to avoid critical situations like the one of measuring only parallel walls of a corridor and no corner. The experimental tests in this mock-up (Figure 4.12) are really challenging, for several reasons. The first one is the narrow spaces of navigation: the LRF in these situations returns biased data. In addition, in real implementations the available map can be quite different from the real environment. Hence the localization and the scan matching approaches must be very robust. Like the simulation stage, also in this case the localization has been performed using or not an initial guess pose estimation from odometry.

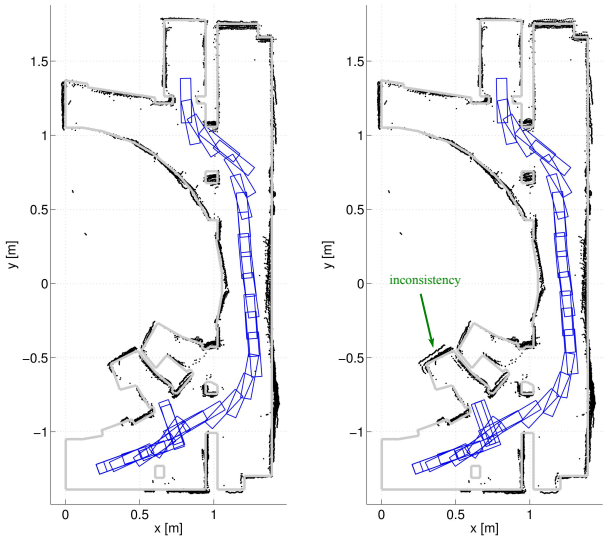
Figure 4.12: Environment and CPRHS mobile robot mock-up used to collect data for the experimental results.

Figure 4.13 reports the results of the PP scan matching approach. The differences in using (Figure 4.13(a)) or not (Figure 4.13(b)) the initial guess pose estimation are more remarkable only in the final part of the path. The reason of the scans mismatching in this zone is probably due to the laser bias in the measurements of close walls on both sides of the laser.

Also for the PL localization (Figure 4.14) the main differences are in the same part of the path. In addition this approach returns bad results in the zone A and along the main corridor: the effects are visible in the top-right walls were the points spread is higher than the PP one.

Lasts results reported are the pose estimations with the feature-based approach (Figure 4.15). In this case the differences between using or not an initial guess are almost negligible. Differently for the simulation results, in this case the results using the feature procedure are more closed to the best dense matching approach performance, and also in the final port the scans are well overlapped.

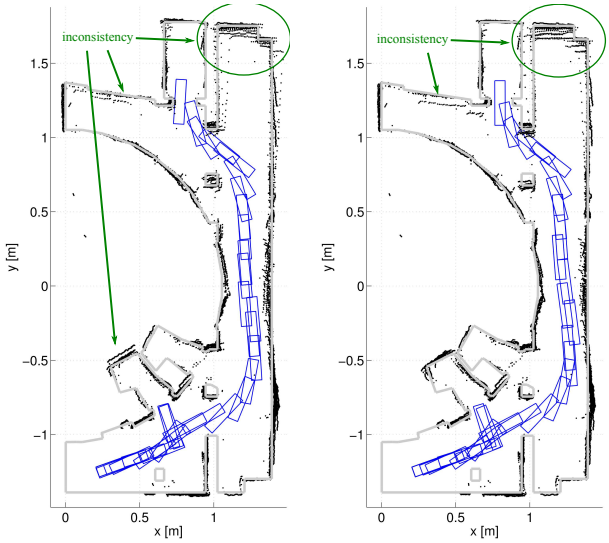
In conclusion it can be observed that in presence of a not ideal map and instrumentations with not modelled biases, the feature-based scan matching is comparable to the dense matching approaches in terms of performances. This is particularly true in terms of accuracy, while the PP approach is superior in terms of precision. Instead, the PL approach in these condition presents a low precision in the region very closed to the undetermined situations.



(a) with initial guess pose estimation

(b) without initial guess pose estimation

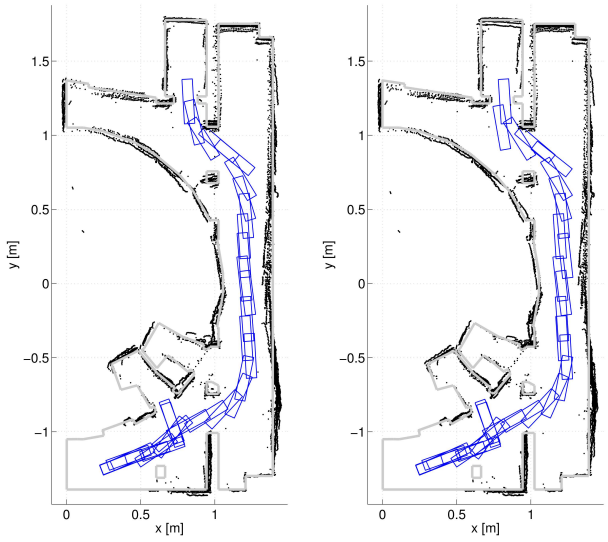
Figure 4.13: Vehicle localization in the mock-up system with *point to point* scan matching approach.



(a) with initial guess pose estimation

(b) without initial guess pose estimation

Figure 4.14: Vehicle localization in the mock-up system with *point to line* scan matching approach.



(a) with initial guess pose estimation

(b) without initial guess pose estimation

Figure 4.15: Vehicle localization in the mockup system with *feature based* scan matching approach.

4.3 Incremental Localization

In this section are reported the results achieved on incremental localization using the laser and the camera as independent sensors, in order to fulfil the requirement of navigation in a generic scenario.

4.3.1 Incremental localization with laser scanner

The analysis of the global localization with maps is very useful in order to evaluate the performances of the scan matching approaches and data association procedure. On the other hand the assessment of scan matching approaches with the global localization with map is not entirely comprehensive. There are two main limitations in this analysis, both related to the data association topic. First of all the matching is performed between a scan and the map which is an ideal measurement. In such way it is not possible to retrieve information on the matching between two sets of information processed from the measurements. Secondly, the evaluation of the error drift in the vehicle pose estimation is not possible. This prevents the understanding of the systematic errors of the data association procedure. Testing instead the algorithms in a incremental localization framework it makes possible to evaluate better the effectiveness of the data association procedure and of the scan matching approaches. Moreover the incremental

localization is indispensable to provide a reliable initial estimation of the vehicle pose for the SLAM approach.

Like the global localization with map, also in this case the analysis is performed first on the laser scan data retrieved by a simulation due to the problem of estimating a reliable ground truth and then on the experimental data of a real environment.

4.3.1.1 Simulated results

The environment simulated in this case is more general than the one used in the global localization with maps, but it is yet structured (Figure 4.16), in order to compare the performances of the feature based and the dense scan matching approaches. Taking into consideration the forthcoming use of the simulator in the SLAM approach, it also permits to draw vehicle paths on which test the loop closure. It must be specified that the aim of the incremental localization analysis is not to report an exhaustive comparison over all the laser profile parameters, but to highlight the results already reported above from a more practical point of view, otherwise this comparison and the one exposed in the global localization will be redundant. Moreover the results on incremental localization are needful to recognize the most efficient scan matching approach.

The first results presented is a comparison of the scan matching approaches over the laser parameters: from the previous investigation is already predictable a decrease of the performance by lowering the laser characteristics, but it is interesting to observe the effects on the pose estimation error drift. The selected profile laser IDs are the two reported in Table 4.4. In both the configurations the maximum range can cover all the environment; in this way it is possible to avoid undetermined situations.

Table 4.4: Laser scanner profile IDs used in the incremental localization simulation.

ID	angular FOV [deg]	angular resolution [deg]	range max [m]	range noise std [m]
2	270	0.5	100	0.04
1	180	1	100	0.1

In Figure 4.16 is reported the localization obtained with the PL approach and the best laser profile (ID2): the acquired laser scans are overlapped to the map using the incremental localization solution. A complete description of the results is reported in Figure 4.17. Immediately is noticeable the drift in the pose estimation, especially in the attitude. In particular for both the laser IDs the PP is the approach with highest error drift: this confirms the assumptions made during its presentation.

Concerning the feature based and PL approach, their results are very closed. By degrading the laser performance, the error in the localization

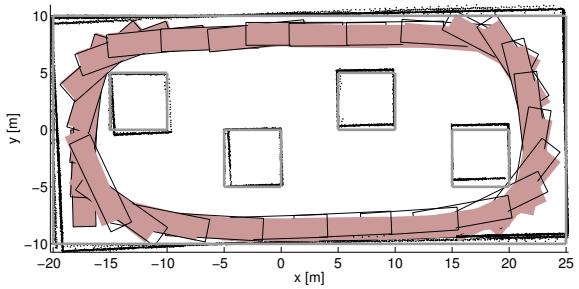
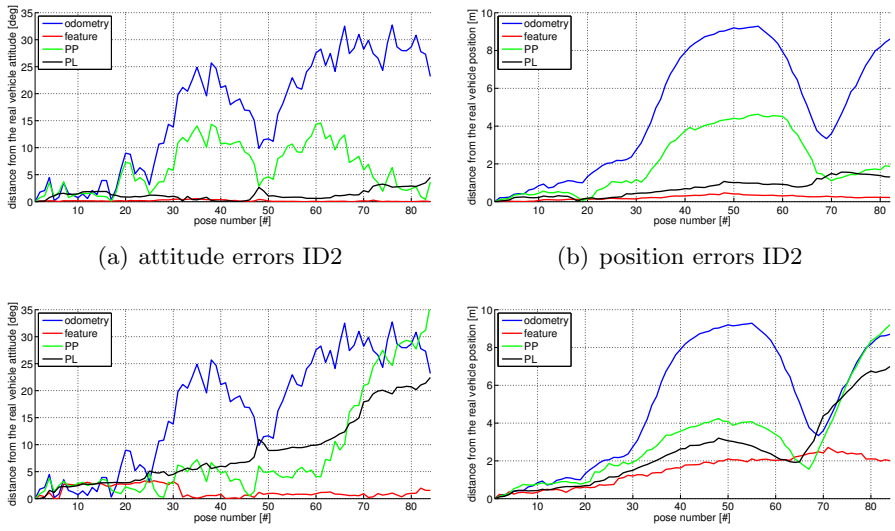


Figure 4.16: Incremental localization with PL in simulated environment using laser profile ID2.



(a) attitude errors ID2

(b) position errors ID2

(c) attitude errors ID1

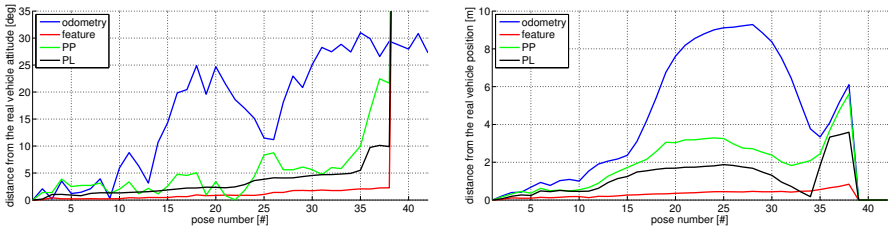
(d) position errors ID1

Figure 4.17: Vehicle incremental localization errors in position and attitude for the laser IDs of Table 4.4.

with PL increase while the performance of the features are more stable. It must be recalled that the kind of environment is the most suitable one for the features.

Another important factor to investigate is the number of laser scan acquisitions used in the incremental pose estimation. If on the one hand a high number of incremental observations can increase the error propagation, on the other hand using few acquisitions forces the scan matching to work with a lower number of points or features in common between two views. The analysis of this aspect is illustrated in Figure 4.18, where it is shown the pose estimation error on the same path used above but with halved acquisition. The performances of all the approaches decrease except for the PP

algorithm. This aspect confirms its low accuracy but also its robustness to high displacement.



(a) attitude errors ID2 (b) position errors ID2

Figure 4.18: Vehicle incremental localization errors in position and attitude with less acquisitions using the laser ID 2.

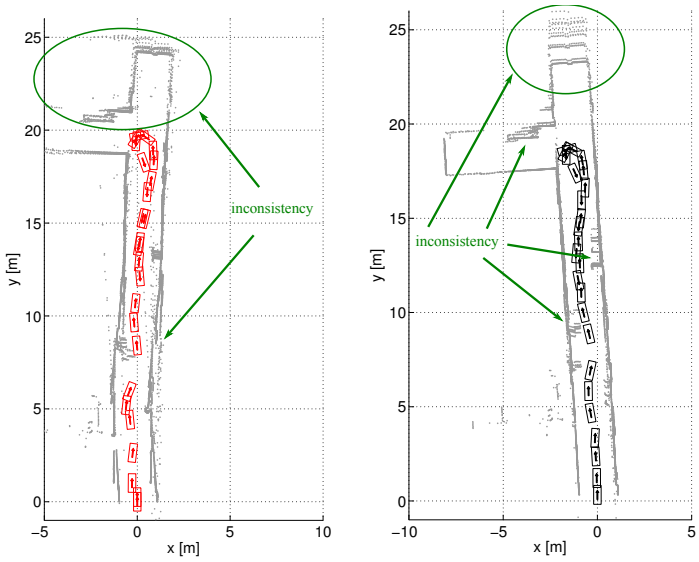
In conclusion, the experiments on simulated data shown that in a well structured environment the feature based scan matching can return good results. Also the PL approach is capable of good performance at most in an optimal environment for the features.

4.3.1.2 Experimental results

The simulation results are useful to investigate theoretic advantages and drawbacks of the scan matching approaches, but to test their actual effectiveness is essential an evaluation on a sequence of laser scans acquired in a real environment. From these evaluations is excluded the PP approach since already during the simulations has shown major limitations especially in terms of error propagation. Though it is very robust to large initial displacement and for this reason is used to estimate the initial guess for the other approaches, since the odometry does not always return good results.

By taking care that a precise ground truth is not known, the best way to represent results in this cases is to overlap all the scans referenced to the environment coordinate frame using the poses estimated with the localization. An accurate solution permits to obtain a proper definition of the navigation environment; on the contrary, with a bad pose estimation, the accuracy by which the environment is reconstructed becomes worst, with an high dispersion of points.

The first pose estimation reported is regarding a long corridor scenario (Figure 4.19) in which most of the acquisitions are closed to undetermined situations. Additionally the number of features is low and attached on the lateral walls there are small and unstable entities: the sum of these issues cause the partial failure of the feature based algorithm. Also the PL approach presents some difficulties, especially in the estimation of the frontal motion, in which instead the feature approach commits less errors.



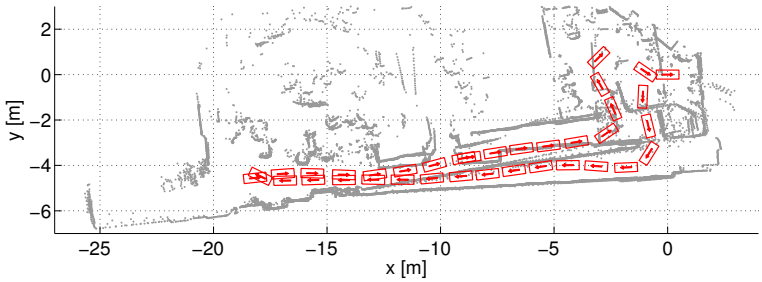
(a) feature (b) PL

Figure 4.19: Incremental localization on long corridor scenario with the feature based an PL scan matching approaches.

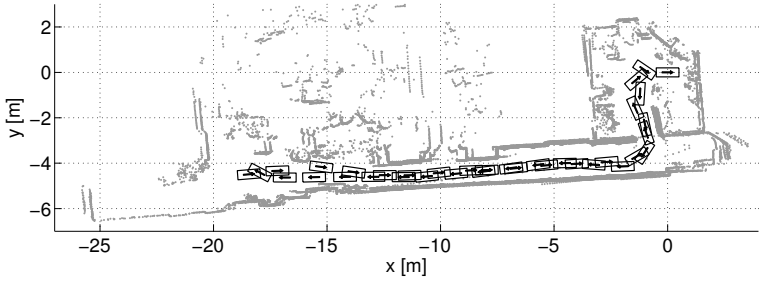
This is noticeable by observing the points dispersion on the top wall, the error is quantifiable for the PL algorithm in about 3 meters. With regard to lateral displacement and attitude instead, by using the PL approach the drift is very limited, the lateral walls are well defined and remains rectilinear also after the turn in the top part of the path.

A second analysis has been carried out in the laboratory of the university. The scenario in this case is more general, there is again a corridor that connect a small room (right side) and an open space (left side), but in this case on its walls there are more detectable features. For this reason the approach based on feature can work in a better way if compared to the previous result. Though globally, the PL approach has superior performances, but it still has problems to localize the vehicle in the corridor. How it is detectable from the wall on the far left and by the point spreading in the open space, the estimation of the motion along the corridor is not completely correct. It is instead remarkable the limited drift in the attitude. After the turn on itself the points overlapping in the corridor is fine, whereas in the feature approach this aspect is a point of weakness.

To conclude this experimental assessment it is possible to notice that, despite the simulation results, in a real environment the scans can be very cluttered. Furthermore in some kind of environment the number of detectable feature can be very low or even null. These peculiarities cause a decrease in performance of the feature based scan matching approach.



(a) feature



(b) PL

Figure 4.20: Incremental localization in the laboratory scenario with the feature based an PL scan matching approaches.

Concerning the PL algorithm, because it does not have restriction on the environment structure, it can work in almost all the situations. As see previously it has a weak point in circumstances closed to undetermined situations, like corridors.

4.3.2 Incremental localization with camera

The testing of the camera localization is quite difficult since at the moment there is not a valid simulator that can be used to generate data. Different environment conditions such as light variation, reflections camera saturation and so on are not so easy to model. For these reason the vehicle localization with camera has been tested only on experimental data. As already introduced, the aim is to obtain a localization system independent from the laser, that can be used to fuse information regarding only the motion estimation. In such way it is possible to overcome the limit of navigation in unstructured environment.

Moreover the motion estimation with camera can be useful when the laser encounters undetermined situations. This is the reason why the method has been tested in the corridor environment already used in the laser incremental localization, which results are reported in Figure 4.19. In Figure 4.21

is shown a frame of the set of images used in the VO. The images sizes

Figure 4.21: Images used in the incremental localization with camera on the long corridor environment.

are 320×240 pixels, in order to obtain a fast image processing with the SIFT detector. Scale factors of the translations are estimated by using the odometry incremental estimation between two frame acquisition.

In Figure 4.22 is depicted the incremental localization performed with the monocular VO and the comparison with the odometric and laser scan matching approach. Though the real ground truth is not known, it is possible to use as reference the best laser estimation obtained with the global localization, since the estimation has high accuracy (Section 4.4.2 for more details). The pose estimation errors in position and attitude of the localization achieved by using the VO and computed by taking as reference the global localization performed with the laser are reported in Figure 4.23. It is noticeable an elevate error drift in the position estimation, while in attitude the error are very closed to the odometry. By comparing the position error and the path estimation it can be seen that the elevated error drift in position is probably due to a wrong data association in the middle of the path.

Concerning the data association step, it has been used the LMedS algorithm described in Section 3.6 applied to the the planar motion constraint. A representation of this procedure is illustrated in Figure 4.24. The SIFT algorithm can extract highly distinctive feature, but in repetitive environment the matching based only on the feature descriptor can connect different features. In addition, also in the case in which the features matching are correct, it is possible that some features are not well defined: it is the case of feature extracted in uniform zones, as the floor or the wall in Figure 4.24. The LMedS algorithm can corrected both these kinds of wrong associations.

In conclusion the monocular VO shows discrete performances especially

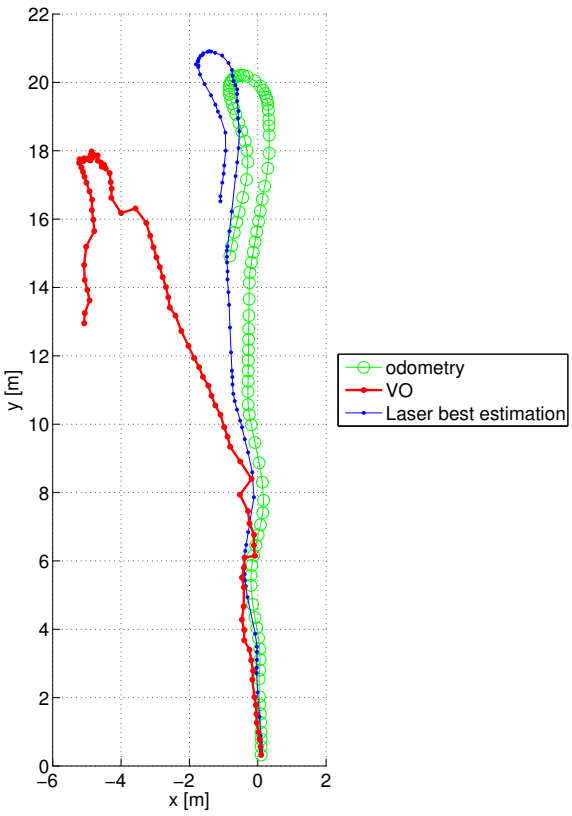
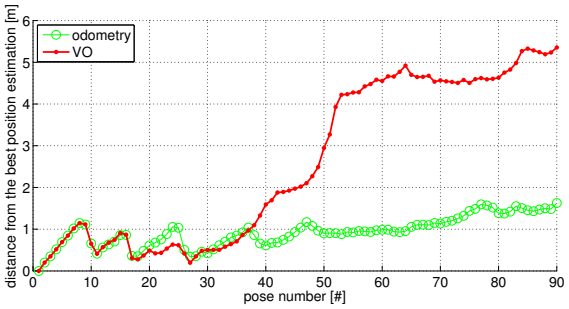


Figure 4.22: Incremental localization with monocular VO. The result is compared with the odometry and the best estimation obtained using the LRF.

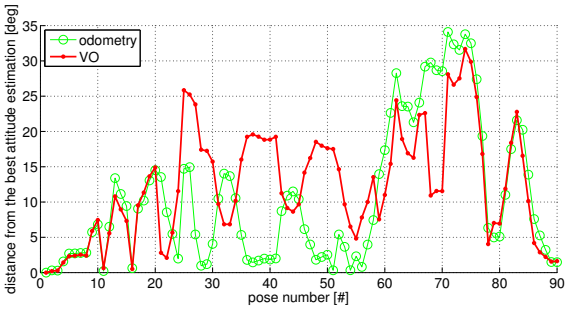
in the attitude error drift that is comparable with the one of the odometry dead reckoning. By observing the images in Figure 4.24 is noticeable that the use of a camera with a limited field of view angle requires large displacement between two image frames. For this reason to improve the VO performances must be used camera with higher angular field of view.

4.4 Global Localization with Graph Theory

The last results presented concern the global localization performed without previous knowledge of the navigation environment. In Section 3.9 it has been introduced an algorithm based on the Graph Theory that can deal with association uncertainties and non linear equation. Since the graph in the case of global localization represents the associations network between all the poses, the best results can be achieved using all the possible poses in a path. However, by applying the algorithm in this manner is a limitation,



(a) Vehicle position error



(b) Vehicle atittude error

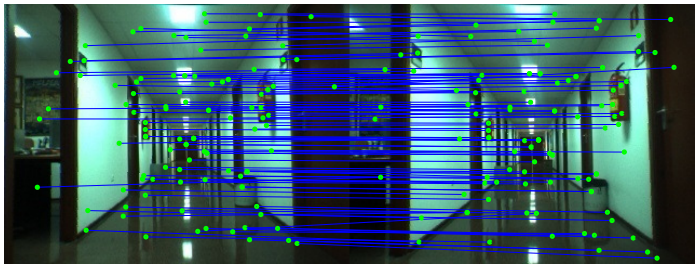
Figure 4.23: Position and attitude error in the monocular VO evaluated by taking as reference the best laser scan matching estimation.

the vehicle has to execute a complete path in order to be globally localized. Along the whole path its pose are estimated using only the incremental localization, with the problem of error drift show previously. Instead, it is possible to apply the method also while the vehicle is moving, on small sets of poses that becomes globally consistent. This procedure can be define as a *windowed* global localization: when a loop closure is detected is then possible to perform the global localization.

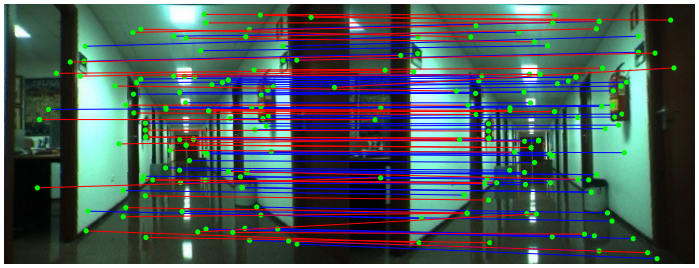
The global localization algorithm proposed is general, in this work has been applied to the laser scanner, but without modification it is possible to use it also for the camera. In particular, in order to achieve the objective of navigation in a general scenario, it has been chosen the PL algorithm which, as noticeable from the incremental localization results, can ensure good results independently from the environment.

4.4.1 Simulated results

To present the benefits in using the windowed global localization is used a an analysis on simulated data, in order to have a precise ground truth. The simulated scenario is the same of the one already used and illustrated



(a) Associations returned by SIFT algorithm



(b) Outlier associations identification with LMedS algorithm

Figure 4.24: Outliers identification in the SIFT algorithm associations. The initial association (a) are processed with an LMedS algorithm in order to detect the outlier associations (red line).

in Figure 4.16. Since the aim is to remark the importance of the global

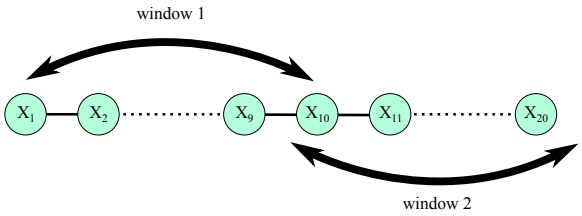
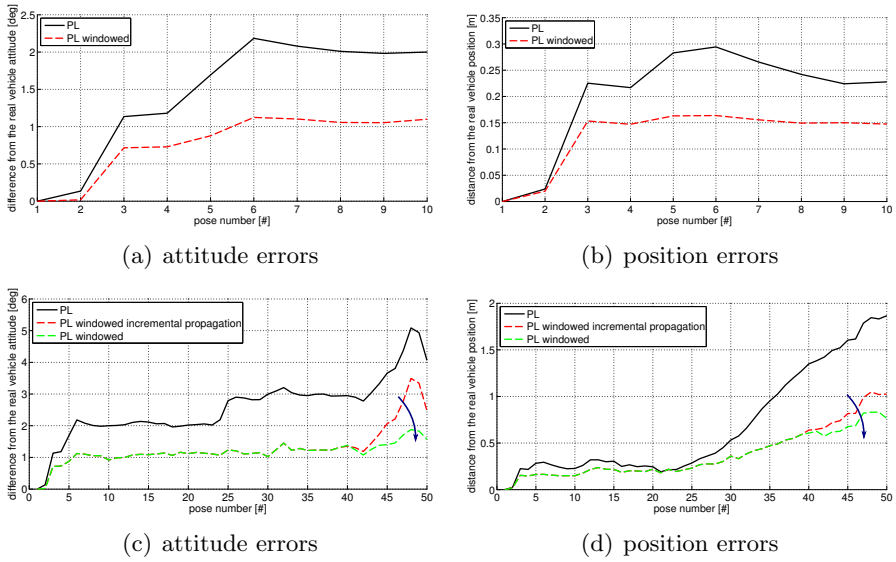


Figure 4.25: Windowed global localization each 10 vehicle poses.

localization it has been used a laser profile with low performances, and the windowed global localization is performed on sets of 10 scans each one (plus one scan to connected the sets, as represent in Figure 4.25).

In Figure 4.26 is illustrated this procedure applied in two different instant time. The vehicle starts to move and localize itself incrementally with the PL approach until it reaches a number of poses equal to 10 (the limit used in this investigation). At this point it is carried out the windowed global localization that can improve the estimation accuracy, as it is notice-

able in Figure 4.27(a) and (b). The windowed global localization creates a



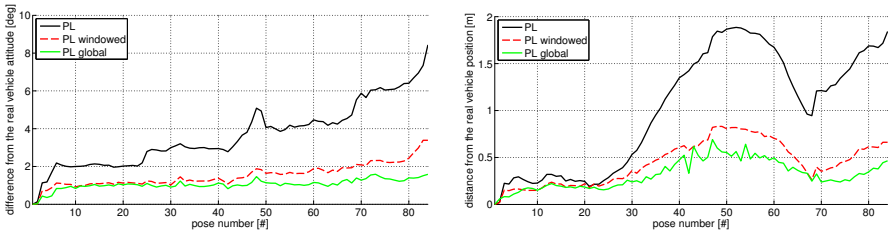
(a) attitude errors (b) position errors

(c) attitude errors (d) position errors

Figure 4.26: Windowed global localization accuracy improvement on the first 10 vehicle poses in attitude (a) and position (b). The attitude and position errors in (c) and (d) are referred to the windowed global localization on the set of poses from 40 to 50. Error evaluated on the ground truth.

small local graph on the set of vehicle poses selected, improving the localization performances. Then, starting from this optimized poses, the vehicle is localized again in incrementally until a new window of optimization can be defined (Figure 4.25). On this new window is performed again a local optimization and the update of the vehicle poses of the window. In Figure 4.26(c) and (d) are shown the errors in vehicle attitude and position at the pose 50 after the windowed global localization. In particular it can been seen the incremental localization carried out starting from the best estimation in pose 40 (red dashed line) and the improvement in terms of error after the windowed global localization (green dashed line). It is also draw the incremental localization of the vehicle (black line).

The use of this windowed global localization procedure ensure a error drift lower than the incremental localization, and so makes it possible to evaluate a better initial pose estimation for the final global optimization on the whole path, as reported in Figure 4.27.



(a) attitude errors (b) position errors

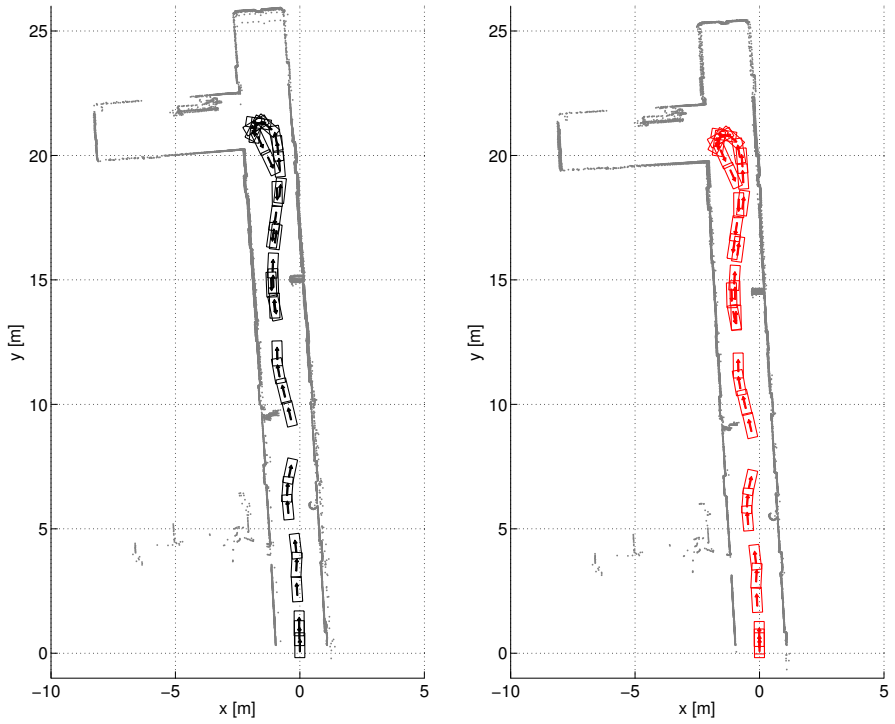
Figure 4.27: Windowed and complete global localization procedures comparison with the incremental vehicle localization.

4.4.2 Experimental results

Analogous considerations to those expressed in the simulation results can be carried out in the experimental results: in real application the effects are clearly visible also without the ground truth. Additionally it is possible to analysis further characteristics that otherwise with simulated data are not replicable.

A first analysis is on the importance of the windowed global localization and subsequently the global localization on the whole path. In Section 4.3.1.2 during the exposure of the incremental localization on real data, with reference to the PL result of Figure 4.19, it has been noticed the problematic of the PL approach in environment similar to a corridor. These problem can be solved using a global localization algorithm, as noticeable in Figure 4.28. Only by using a windowed global localization the problem of the estimation the frontal motion in a corridor can be drastically reduced: this is observable by comparing the points spreading on the top wall and on the small object placed along the lateral walls. Using a global localization on the whole path the results are still better. The reason of this improvement is on the uncertainty weighting of the graph links: in environment like a corridor the uncertainty in its main direction is high. The links in this direction are weak, and so the global localization (and also the windowed) can move easily the poses in according to the more reliable measures, in order to achieve the best vehicle path estimation.

From the foregoing considerations is clear that the use of the uncertainty weighting is indispensable for a trustworthy method. This consideration is valid not only for corridors, but also for general environment. To demonstrate that this is a relevant aspect, in Figure 4.29 and Figure 4.30 are reported the results of global localization carried out by applying or not the uncertainty weighting to the graph solution. The aim of this confrontation is to prove that uncertainty weighting can makes the difference; though the results are quite closer, a measurement in real application must be qualified



(a) windowed global localization (b) global localization

Figure 4.28: Windowed and global localization in long corridor scenario.

with uncertainty. Differences seem to be minimal, due to the good guess trajectory returned by the incremental localization. Analysing deeply the reconstruction it is possible to notice that the global optimization performed by taking care of uncertainty Figure 4.29 is more accurate. This is clearly visible in Figure 4.30 by observing the point spreading in almost all the reconstructed walls. The error in the pose estimation is also observable in the wall between the two rooms: while using uncertainty weighting it is defined by only one segment, without taking care of the uncertainties the error entails in two different estimations of the same wall.

Another important feature to investigate is the performance enhancement fostered by the loop closure. In both the previous results, it has been granted the possibility to the global localization to close the loop, or rather to associate also the measures of the last poses to the first acquisitions. It is interesting to observe the consequences when these associations are avoided. To prevent the loop closure, it has been imposed a limit on the number of connections between subsequent nodes in the graph. The limit on forward links has been chosen in order to avoid the connections between the initial scans and the ones acquired by the laser after the round-about on the right

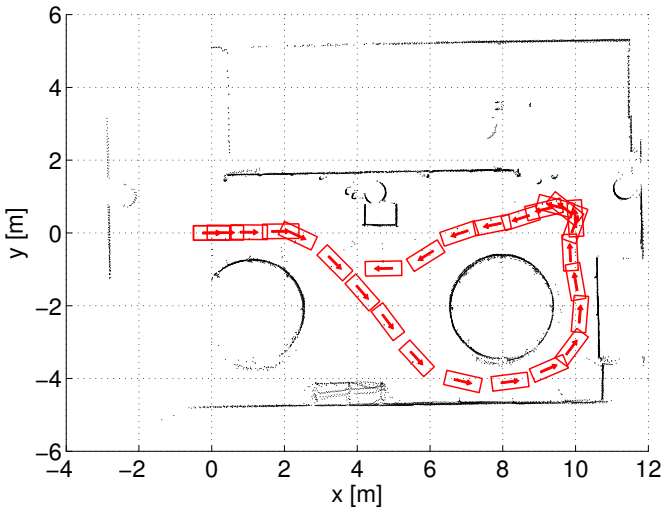


Figure 4.29: Graph based global localization on experimental data using uncertainty weighting.

circumference in the environment. The localization result is reported in Figure 4.31 and it allows to emphasize an important consideration. Since in this case it is used the uncertainty weighting, the point spreading on the wall is very limited, but are noticeable the effects of the error drift in the pose estimation. Most of the walls are defined twice: this issue is attributable precisely to the error accumulated during the round-about on the right circumference. Thus in a global localization without loop closure or with a loop closure that defines a labile network, if on the one hand the local reconstruction seems very accurate, on the other hand it is possible that the regions far from the initial acquisition can be affected by a significant error in the pose estimation.

In this context it is interesting to analyse the uncertainties in presence or not of a loop closure. This analysis is resumed in Figure 4.32 and Figure 4.33. In the first figure are depicted the uncertainties among the whole vehicle path. Here again it is proposed the comparison between the incremental, and the global optimization: the ellipses represent the position uncertainties with the probabilities of the 99.7%, while the arcs of circumference are representative of the attitude uncertainties, also related to the 99.7%. As reported in the figure note, to obtain a better representation the uncertainties were scaled of a factor 10.

Analysing the uncertainties, the considerations made previously, based on qualitative observation of the overlapped scans can be confirmed. The drift of the incremental estimation is well explained from its uncertainty, that grows at each step, while in global localization the uncertainties are limited, like the drift in the pose estimation. It can be noticed another important

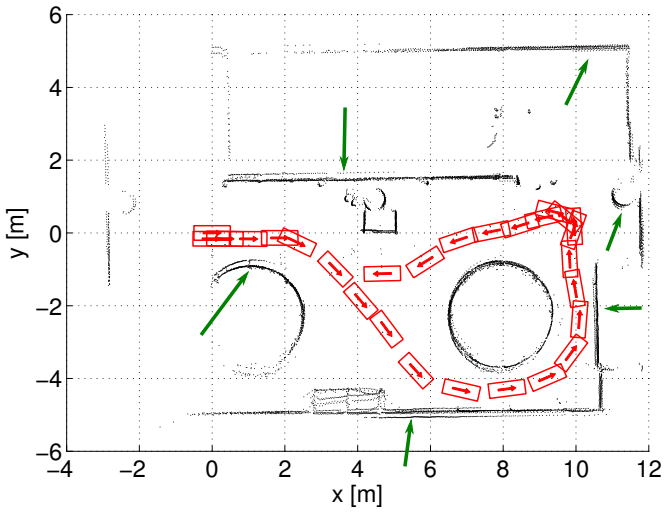


Figure 4.30: Graph based global localization on experimental data *without* the use of uncertainty weighting. Green arrows indicated the scan inconsistencies.

aspect: global localization, with or without loop closure, ensure a lower drift in vehicle attitude estimation. In Figure 4.33 are represented the pose estimation of the three procedures, in the $8th$ and $27th$ (last) scans. By this way it is possible to observe the good performances of global localization and loop closure. In the last pose the uncertainty in the closed loop global localization is at least an order of magnitude less than the incremental one. In addition the uncertainty of the global localization with loop closure is lower than the localization in open loop.

From the same figure another important feature of the graph based slam can be noticed: all the nodes are used in optimization and so all the node poses can be optimized. This is the big difference between online and full SLAM, and it is noticeable in the estimation of the pose 8. Using the graph based global localization and by forcing the loop closure, the values of position and attitude of this pose change. The same operation is unattainable by using a filtering operation, since the information of previous poses are not achievable. This means that further scan matching can not change previous poses, while using the graph based global localization all the previous poses can be changed in order to obtain the best global optimization.

Last results concern the global localization in large environments. The testing in large environment is very important to qualify the global localization algorithms. In this case the main problem is the error drift in the vehicle pose estimation used as initial guess in the global localization. If in small scale environments the drift is limited and can be easily compensated, in large scale environment it can be relevant and it can cause inconsistency as

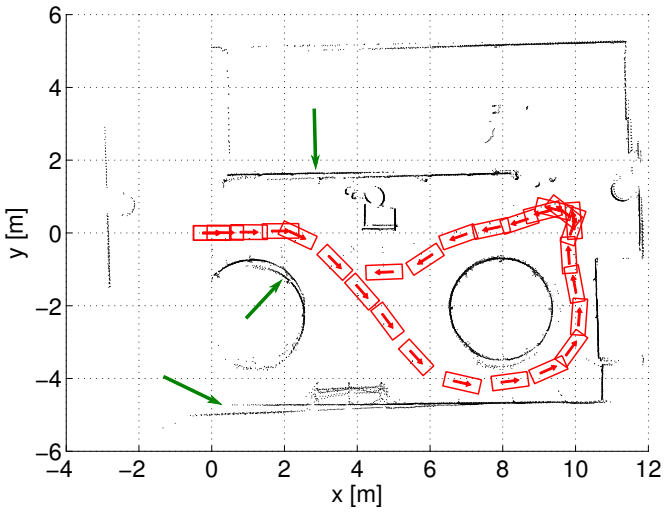


Figure 4.31: Graph based global localization on experimental data preventing the loop closure. Green arrows indicate inconsistencies due to the error in pose estimation.

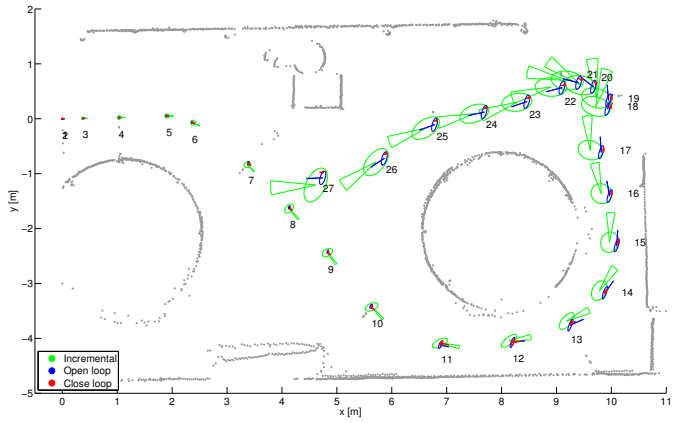


Figure 4.32: Pose uncertainties comparison between incremental and global with open and close loop pose estimation. Uncertainties are scaled by a factor 10.

represented in Figure 4.34(a). These issues can preclude the loop closure by introducing wrong associations. From this point of view is really important to estimate as good as possible the vehicle pose also when it is not possible a loop closure, using for instance the windowed global localization procedure. In Figure 4.34 are depicted the two step of the windowed global localization and the global localization performed on the whole path. It is noticeable an inconsistency in the scan overlapping in Figure 4.34(a) but is very limited considering that it has been computed in open loop. Also after the global

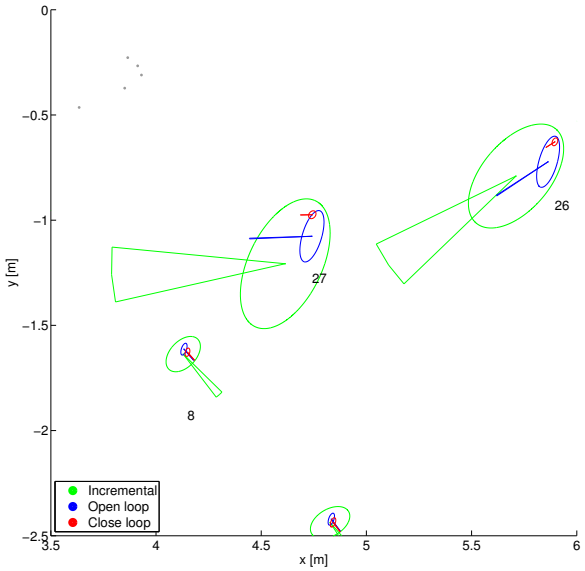
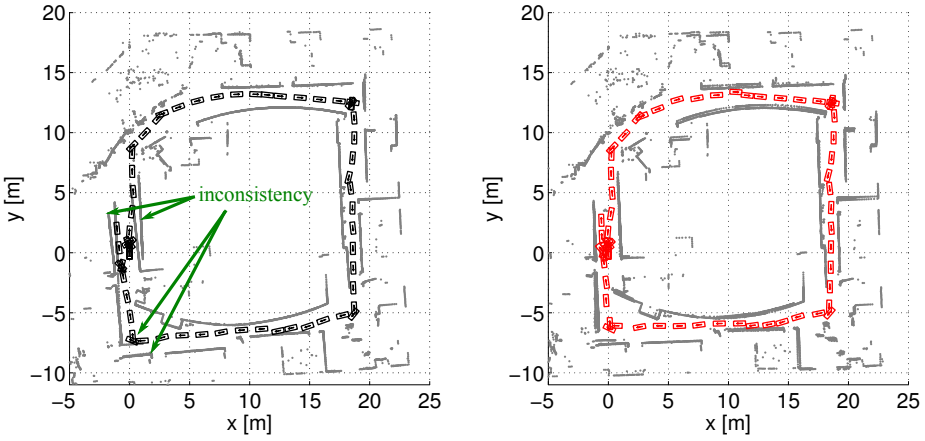


Figure 4.33: Detail of the pose uncertainties comparison between incremental an global localization.



(a) windowed global localization (b) global localization on the whole path

Figure 4.34: Graph based global localization on the Intel dataset [80] using the windowed procedure (a) and the global localization on the whole path(b).

localization (Figure 4.34(b)) are still present some inconsistencies, that are probably due to the low number of scan used in the global localization.

In addition a useful instrument that can be used to understand the state of the connections is the Information Matrix (Section 2.3.4.1) which is the inverse of the covariance matrix of the system of equation to solve. The Information Matrix associated to this global localization is illustrated in

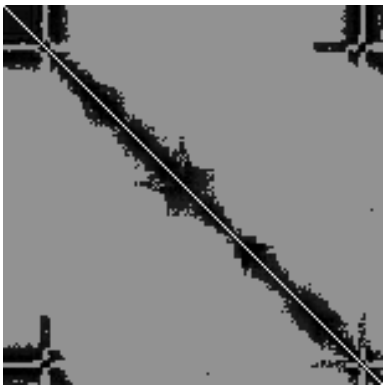


Figure 4.35: Representation of the information matrix associated to the graph based global localization on the Intel dataset [80].

Figure 4.35 as image. Darker areas represent strong connections between poses: this ones are limited in a band around the diagonal, which means that only small group of subsequent poses can be connected. This constraint is forced by the structure of the environment. It is also possible to observe the two small groups in the corners opposite to the diagonal: these areas represent the loop closure. Hence, in this case the windowed global localization is a needful instrument not only in order to estimate a reliable vehicle localization before the global localization on the whole path, but also to ensure the loop closure.

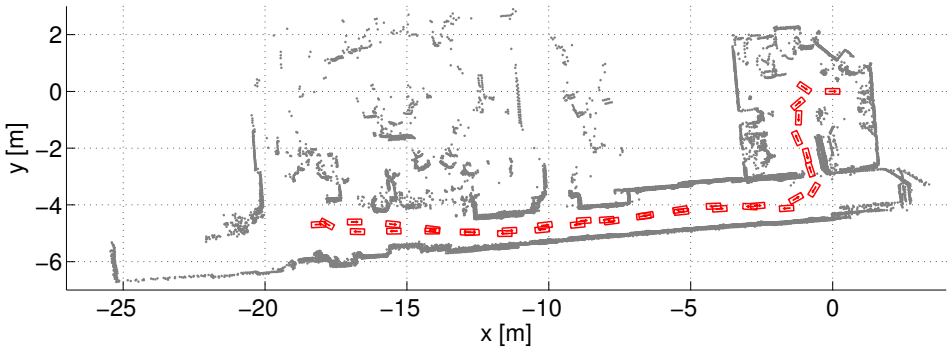


Figure 4.36: Graph based global localization in the laboratory scenario.

In Figure 4.36 is represented the result of the global localization on the same data used in the incremental localization (Figure 4.20). The improvement of the global localization are noticeable also in very cluttered environment. It is also very interesting to analyse the Information Matrix for this vehicle path, which is drawn in Figure 4.37. Like the previous information matrix there are strong connections on the diagonal, but this time there

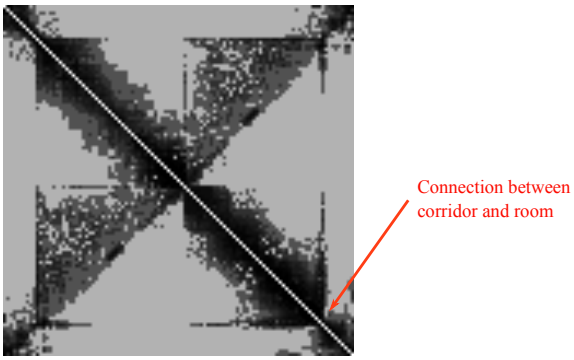


Figure 4.37: Representation of the information matrix associated to the graph based global localization in the laboratory scenario.

are also connections outside the diagonal band. This means that different portions of the vehicle path were carried out in the same environment. It is also possible to notice a loop closure between the initial and final poses and also the division of the environment in two different rooms, connected by a small set of poses.

Chapter 5

Conclusions

This thesis work presents a set of algorithms capable to estimate the motion of a vehicle in an indoor scenario, without a priori knowledge of the environment, by using the measurements acquired from a camera and a laser range finder. In particular the aim was to develop a general framework based on a probabilistic approach that can be used in both the sensors. For this reason were identified three main topics, the processing of the sensors information, the data association and the vehicle localization. With the exception of the first, which is strictly connected to the information returned from the instruments, the other procedures can be easily fitted for each type of opto-electronic measurement system.

The environment information processing was carried out by considering a generic environment. Concerning the laser data, it was highlighted how dense scan matching approaches and in particular the point to line algorithm, can be used in generic environments and geometric environments as well. This result was derived by the comparison of the three scan matching approaches in the global localization with map. Also for the camera was decided to use a data elaboration approach suitable for an unknown environment: since the information in the images are referred not only to geometric environment characteristics, in this case it was used the identification of locally invariant features with the SIFT algorithm. Furthermore it was pointed out how in some situations wherein the laser scan matching solution is not defined, the properties of the camera can be helpful to localize the vehicle.

Regard the data association was developed a robust procedure based on the LMedS algorithm, capable to deal with an high percentages of outliers. The use of this method has permitted to apply a probabilistic approach in the association of the data acquired in different poses vehicle, which is indispensable in order to obtain a reliable method.

With respect to the last topic, the vehicle localization, besides the incremental approach were developed two global localization procedures. The

first one, although not innovative since based on map, with the improvements on data processing and association, has shown results of remarkable applicative interest. In particular it was applied in a preliminary analysis for the localization with on-board laser range finder of a service vehicle used in the ITER remote handling system. A more original approach was developed in the global localization based on the Graph Theory. The solution of this SLAM problem was strengthened using a two step linearization and uncertainty weighting of the graph connections. In addition, since the algorithm does not require expressly the mapping procedure, it is possible to use also non feature based approaches. With the laser range finder it was possible to take advantage of this aspect by using the PL dense scan matching approach. The global localization with on-board laser range finder thus obtained was applied in different conditions and qualified using simulation, datasets and data acquired from a research mobile robot. Due to the generality and robustness of the algorithm developed it was possible to carried out the vehicle global localization in all these contexts without any modification of the approach or the environment. All these characteristics ensure that the same procedure can be also applied to the information elaborated from the camera, as it was already done with the data association algorithms. In such a way it would be possible to have two different motion estimation approaches which can be used in a complementary manner to guarantee the vehicle localization.

This work was useful to develop a framework for the study of the vehicle localization problem and to create a benchmark tool for the evaluation of vehicle localization algorithms. The possible further improvements in order to obtain a complete localization systems are:

- extension of the vehicle global localization approach to the camera, by qualifying the uncertainty of the image feature identification procedure.
- selection of the key-frame poses based on the information matrix returned by the Graph Theory approach, in order to solve optimally the global localization.
- representation of the vehicle pose uncertainty with a Particle Filter algorithm in order to solve the kidnapped robot problem and develop a multi robot exploration.

Bibliography

[1] Statistical Department International Federation of Robotics. World robotics 2012, august 2012.

[2] EUROP European Robotics Technology Platform. The strategic research agenda for robotics in europe, july 2009.

[3] S. Singh and J. West. Cyclone: A laser scanner for mobile robot navigation. Technical Report CMU-RI-TR-91-18, Robotics Institute, Pittsburgh, PA, September 1991.

[4] I. J. Cox. Blanche-an experiment in guidance and navigation of an autonomous robot vehicle. *Robotics and Automation, IEEE Transactions on*, 7(2):193–204, April 1991.

[5] J. Gonzalez, A. Stentz, and A. Ollero. An iconic position estimator for a 2d laser rangefinder. In *Robotics and Automation, 1992. Proceedings., 1992 IEEE International Conference on*, pages 2646 –2651 vol.3, may 1992.

[6] G.K. Shaffer, A. Stentz, W.L. Whittaker, and K.W. Fitzpatrick. Position estimator for underground mine equipment. *Industry Applications, IEEE Transactions on*, 28(5):1131 –1140, sep/oct 1992.

[7] G. Shaffer, J. Gonzalez, and A. Stentz. A comparison of two range-based pose estimators for a mobile robot. In *Proceedings of SPIE Symposium on Mobile Robots*, 1992.

[8] K.O. Arras and R. Siegwart. Feature extraction and scene interpretation for map-based navigation and map building. In *Proceedings of SPIE, Mobile Robotics XII*, volume 3210, pages 42–53. Citeseer, 1997.

[9] T. Einsele and G. Farber. Real-time self-localization in unknown indoor environments using a panorama laser range finder. In *In IEEE/RSJ International Workshop on Robots ans Systems, IROS 97*, pages 697–703. IEEE Press, 1997.

[10] F. Lu and E. Milios. Robot pose estimation in unknown environments by matching 2d range scans. *Journal of Intelligent and Robotic Systems*, 18(3):249–275, 1997.

[11] J.-S. Gutmann and C. Schlegel. Amos: comparison of scan matching approaches for self-localization in indoor environments. In *Advanced Mobile Robot, 1996., Proceedings of the First Euromicro Workshop on*, pages 61 –67, oct 1996.

[12] J.-S. Gutmann, W. Burgard, D. Fox, and K. Konolige. An experimental comparison of localization methods. In *Intelligent Robots and Systems, 1998. Proceedings., 1998 IEEE/RSJ International Conference on*, volume 2, pages 736 –743 vol.2, oct 1998.

[13] P.J. Besl and N.D. McKay. A method for registration of 3-d shapes. *IEEE Trans. Pattern Anal. Mach. Intell.*, 14(2):239–256, February 1992.

[14] Z. Zhang. Iterative point matching for registration of free-form curves and surfaces. Technical report, Institut National de Recherche en Informatique et en Automatique, 1994.

[15] L. Montesano, J. Minguez, and L. Montano. Probabilistic scan matching for motion estimation in unstructured environments. In *2005 IEEE/RSJ International Conference on Intelligent Robots and Systems, 2005. (IROS 2005)*, pages 3499 – 3504, August 2005.

[16] P. Jensfelt. Approaches to mobile robot localization in indoor environments. Technical report, PhD thesis, Signal, Sensors and Systems (S3), Royal Institute of Technology, SE-100 44, 2001.

[17] A. Censi. An accurate closed-form estimate of ICP's covariance. In *2007 IEEE International Conference on Robotics and Automation*, pages 3167 –3172, April 2007.

[18] A. Censi. On achievable accuracy for range-finder localization. In *Proceedings of the IEEE International Conference on Robotics and Automation (ICRA)*, pages 4170–4175, Rome, Italy, April 2007.

[19] A. Diosi and L. Kleeman. Uncertainty of line segments extracted from static sick pls laser scans. In *Proceedings of the Australasian conference on robotics and automation*. Citeseer, 2003.

[20] V. Nguyen, A. Martinelli, N. Tomatis, and R. Siegwart. A comparison of line extraction algorithms using 2d laser rangefinder for indoor mobile robotics. In *Intelligent Robots and Systems, 2005.(IROS 2005). 2005 IEEE/RSJ International Conference on*, pages 1929–1934. IEEE, 2005.

[21] R. Martinez-Cantin, J.A. Castellanos, J.D. Tardos, and J.M.M. Montiel. Adaptive scale robust segmentation for 2d laser scanner. In *Intelligent Robots and Systems, 2006 IEEE/RSJ International Conference on*, pages 796 –801, oct. 2006.

[22] A. Censi. An ICP variant using a point-to-line metric. In *IEEE International Conference on Robotics and Automation, 2008. ICRA 2008*, pages 19 –25, May 2008.

[23] A. Diosi and L. Kleeman. Scan matching in polar coordinates with application to slam. Technical report, Department of Electrical and Computer Systems Engineering Monash University, Clayton, 2005.

[24] D. Scaramuzza and F. Fraundorfer. Visual odometry [tutorial]. *Robotics Automation Magazine, IEEE*, 18(4):80 –92, dec. 2011.

[25] F. Fraundorfer and D. Scaramuzza. Visual odometry : Part ii: Matching, robustness, optimization, and applications. *Robotics Automation Magazine, IEEE*, 19(2):78 –90, june 2012.

[26] A.J. Davison. Real-time simultaneous localisation and mapping with a single camera. In *Computer Vision, 2003. Proceedings. Ninth IEEE International Conference on*, pages 1403–1410. Ieee, 2003.

[27] M. Chli. *Applying Information Theory to Efficient SLAM*. PhD thesis, Imperial College London, 2009.

[28] A. Bartoli and P. Sturm. Structure-from-motion using lines: Representation, triangulation, and bundle adjustment. *Computer Vision and Image Understanding*, 100(3):416–441, 2005.

[29] D. Nistér. An efficient solution to the five-point relative pose problem. *IEEE Trans. Pattern Anal. Mach. Intell.*, 26(6):756–777, June 2004.

[30] E. Mouragnon, M. Lhuillier, M. Dhome, F. Dekeyser, and P. Sayd. Monocular vision based slam for mobile robots. In *18th International Conference on Pattern Recognition*, 2006.

[31] K. Konolige and M. Agrawal. Frameslam: From bundle adjustment to real-time visual mapping. *IEEE Transactions on Robotics*, 24(5):1066–1077, 2008.

[32] J.M.M. Montiel D. Ortin. Indoor robot motion based on monocular images. *Robotica*, Vol. 19:pp. 331 – 342, 2001.

[33] D. Scaramuzza. 1-point-ransac structure from motion for vehicle-mounted cameras by exploiting non-holonomic constraints. *Int. J. Comput. Vision*, 95(1):74–85, October 2011.

[34] R. Hartley and A. Zisserman. *Multiple View Geometry in Computer Vision*. Cambridge University Press, New York, NY, USA, 2 edition, 2003.

[35] N. Naikal, J. Kua, G. Chen, and A. Zakhor. Image augmented laser scan matching for indoor dead reckoning. In *Proceedings of the 2009 IEEE/RSJ international conference on Intelligent robots and systems*, IROS'09, pages 4134–4141, Piscataway, NJ, USA, 2009. IEEE Press.

[36] K.O. Arras and N. Tomatis. Improving robustness and precision in mobile robot localization by using laser range finding and monocular vision. In *Advanced Mobile Robots, 1999.(Eurobot'99) 1999 Third European Workshop on*, pages 177–185. IEEE, 1999.

[37] J.C. Andersen, N.A. Andersen, and 0. Ravn. Vision assisted laser scanner navigation for autonomous robots. In *ISER*, pages 111–120, 2006.

[38] X. Wang, Y. Cheng, R.T. Collins, and A.R. Hanson. Determining correspondences and rigid motion of 3-d point sets with missing data. In *IEEE Computer Vision and Pattern Recognition*, pages 252–257. IEEE computer society press, 1996.

[39] E. Trucco, A. Fusiello, and V. Roberto. Robust motion and correspondence of noisy 3-d point sets with missing data. *Pattern recognition letters*, 20(9):889–898, 1999.

[40] A.W. Fitzgibbon. Robust registration of 2d and 3d point sets. *Image Vision Comput.*, 21(13-14):1145–1153, 2003.

[41] L. Zhang, S. Choi, and S. Park. Robust icp registration using biunique correspondence. In *Proceedings of the 2011 International Conference on 3D Imaging, Modeling, Processing, Visualization and Transmission*, 3DIMPVT '11, pages 80–85, Washington, DC, USA, 2011. IEEE Computer Society.

[42] A. F. Siegel. Robust regression using repeated medians. *Biometrika*, 69(1):242–244, April 1982.

[43] P.J. Rousseeuw. Least median of squares regression. *Journal of the American Statistical Association*, 79(388):871–880, December 1984. ArticleType: research-article / Full publication date: Dec., 1984 / Copyright Â© 1984 American Statistical Association.

[44] P. C. Mahalanobis. On the generalised distance in statistics. In *Proceedings National Institute of Science, India*, volume 2, pages 49–55, April 1936.

[45] B. Reiser. Confidence intervals for the mahalanobis distance. *Communications in Statistics-Simulation and Computation*, 30(1):37–45, 2001.

[46] JMM Montiel and L. Montano. Efficient validation of matching hypotheses using mahalanobis distance. *Engineering Applications of Artificial Intelligence*, 11(3):439–448, 1998.

[47] M. Irani and P. Anandan. Factorization with uncertainty. In *In European Conference on Computer Vision*, pages 539–553, 2000.

[48] Neira, J. and Tardós, J.D. Data Association in Stochastic Mapping Using the Joint Compatibility Test. *IEEE Transactions on Robotics and Automation*, 17(6):890–897, dec 2001.

[49] R. Siegwart and I.R. Nourbakhsh. *Introduction to Autonomous Mobile Robots*. MIT Press, April 2004.

[50] T. Einsele. *Localization in Indoor Environments Using a Panoramic Laser Range Finder*. PhD thesis, TU München, Lehrstuhl für Realzeit-Computersysteme, 2002.

[51] R. Chatila and J. Laumond. Position referencing and consistent world modeling for mobile robots. In *1985 IEEE International Conference on Robotics and Automation. Proceedings*, volume 2, pages 138 – 145, March 1985.

[52] R. Smith, M. Self, and P. Cheeseman. A stochastic map for uncertain spatial relationships. In *on The fourth international symposium*, pages 467–474, Univ. of California, Santa Clara, California, United States, 1988. MIT Press.

[53] H. Durrant-Whyte and T. Bailey. Simultaneous localisation and mapping (SLAM): part i the essential algorithms. *IEEE Robotics and Automation Magazine*, 2:2006, 2006.

[54] T. Duckett, S. Marsland, and J. Shapiro. Learning globally consistent maps by relaxation. In *ICRA'00*, pages 3841–3846, 2000.

[55] T. Duckett, S. Marsland, and J. Shapiro. Fast, on-line learning of globally consistent maps. *Auton. Robots*, 12(3):287–300, May 2002.

[56] K. Konolige. Large-scale map-making. In *Proceedings of the 19th national conference on Artifical intelligence*, AAAI'04, pages 457–463. AAAI Press, 2004.

[57] K. Konolige. SLAM via variable reduction from constraint maps. In *Proceedings of the 2005 IEEE International Conference on Robotics and Automation, 2005. ICRA 2005*, pages 667 – 672, April 2005.

[58] S. Thrun and M. Montemerlo. The graphslam algorithm with applications to large-scale mapping of urban structures. *International Journal on Robotics Research*, 25(5):403–430, 2006.

[59] S. Thrun, W. Burgard, and D. Fox. *Probabilistic Robotics (Intelligent Robotics and Autonomous Agents)*. The MIT Press, 2005.

[60] Bill Triggs, Philip Mclauchlan, Richard Hartley, and Andrew Fitzgibbon. Bundle adjustment - a modern synthesis. In *Vision Algorithms: Theory and Practice, LNCS*, pages 298–375. Springer Verlag, 2000.

[61] A. Bartoli and P. Sturm. Three New Algorithms for Projective Bundle Adjustment with Minimum Parameters. Research Report RR-4236, INRIA, 2001.

[62] F. Lu and E. Milios. Globally consistent range scan alignment for environment mapping. *Auton. Robots*, 4(4):333–349, October 1997.

[63] J.S. Gutmann and K. Konolige. Incremental mapping of large cyclic environments. In *1999 IEEE International Symposium on Computational Intelligence in Robotics and Automation, 1999. CIRA '99. Proceedings*, pages 318 –325, 1999.

[64] H. Strasdat, J.M.M. Montiel, and A.J. Davison. Real-time monocular SLAM: why filter? In *2010 IEEE International Conference on Robotics and Automation (ICRA)*, pages 2657 –2664, May 2010.

[65] H.S.M. Coxeter. *Regular Polytopes*. Courier Dover Publications, June 1973.

[66] T. Harju. Graph theory. Technical report, University of Turku, Department of Mathematics, 01 2011.

[67] J.S. Gutmann and B. Nebel. Navigation mobiler roboter mit laserscans. In *Autonome Mobile Systeme 1997, 13. Fachgesprach*, pages 36–47, London, UK, UK, 1997. Springer-Verlag.

[68] M. Golfarelli, D. Maio, and S. Rizzi. Elastic correction of dead-reckoning errors in map building. In *Intelligent Robots and Systems, 1998. Proceedings., 1998 IEEE/RSJ International Conference on*, volume 2, pages 905 –911 vol.2, oct 1998.

[69] U. Frese and G. Hirzinger. Simultaneous localization and mapping - a discussion. Technical report, Deutsches Zentrum fur Luft- und Raumfahrt (DLR), Institut fur Robotik und Mechatronik, 2001.

[70] J. Folkesson and H. Christensen. Graphical slam - a self-correcting map. In *In IEEE Intl. Conf. on Robotics and Automation (ICRA*, pages 383–390, 2004.

[71] J.L. Martinez, J. Morales, A. Mandow, and A. Garcia-Cerezo. Incremental closed-form solution to globally consistent 2D range scan mapping with two-step pose estimation. In *2010 11th IEEE International Workshop on Advanced Motion Control*, pages 252 –257, March 2010.

[72] Universidade Técnica de Lisboa Instituto de Plasmas e Fusão Nuclear, Instituto Superior Técnico. Iter remote handling, december 2012.

[73] D. W. Eggert, A. Lorusso, and R. B. Fisher. Estimating 3-d rigid body transformations: a comparison of four major algorithms. *Machine Vision and Applications*, 9(5):272–290, March 1997.

[74] D. Nistér, O. Naroditsky, and J. Bergen. Visual odometry for ground vehicle applications. *Journal of Field Robotics*, 23:2006, 2006.

[75] R. Hartley. Cheirality.

[76] T. Tuytelaars and K. Mikolajczyk. *Local Invariant Feature Detectors: A Survey*. Now Publishers Inc., Hanover, MA, USA, 2008.

[77] D.G. Lowe. Distinctive image features from scale-invariant keypoints. *Int. J. Comput. Vision*, 60(2):91–110, November 2004.

[78] A. Vale and I. Ribeiro. Mobile robot navigation for remote handling operations in iter. In *III Workshop de Robótica: Robótica Experimental*, 2011.

[79] T.B. Schon and F. LindSten. Manipulating the multivariate gaussian density. Technical report, Linkoping University, Division of Automatic Control, 01 2011.

[80] J.L. Blanco. The mobile robot programming toolkit, MRPT, 2012.

[81] A. Kassir and T. Peynot. Reliable automatic camera-laser calibration. In *Proceedings of the 2010 Australasian Conference on Robotics & Automation,*, Dec, 2010.