University of Trento

Alberto Fornaser

Data fusion of images and 3D range data

Prof. Mariolino De Cecco

Prof. Daniele Bortoluzzi

Dott. Luca Baglivo

2014

University of Trento

Data fusion of images and 3D range data

Final Examination 07 / 04 / 2014

Board of Examiners

Prof. Oreste Salvatore Bursi, Università degli Studi di Trento

Prof. Dionisio P. Bernal, Northeastern University, Boston

Prof. Michel Destrade, National University of Ireland Galway

Dott. Andrea Giovanni Calogero, Università Milano-Bicocca

Dott. Paola Falugi, Imperial College, London

Sommario

Un robot è una macchina che incorpora decenni di ricerca e sviluppo. Nate come semplici dispositivi meccanici, queste macchine si sono evolute insieme alla nostra tecnologia e conoscenza, raggiungendo un grado di automazione mai immaginato prima.

Il sogno moderno è rappresentato dalle robotica cooperativa, in cui i robot non lavorano solo per le persone, ma con le persone. Tale risultato può essere raggiunto solo se queste macchine sono in grado di acquisire conoscenze attraverso la percezione, in altre parole, in grado di raccogliere misure da sensori ed estrarre informazioni significative riguardo all'ambiente al fine di adeguare il proprio comportamento.

Questa tesi tratta il tema del riconoscimento e presa automatici di oggetti da parte di Veicoli a Guida Automatica, AGV, robot utilizzati oggigiorno negli impianti di logistica automatica.

Lo sviluppo di una tecnologia in grado di assolvere a tale compito rappresenterebbe un notevole salto tecnologico rispetto alla struttura attualmente utilizzata in questo settore: rigida, fortemente vincolata e con bassissima interazione uomo macchina. Automatizzare il processo di presa rendendo i veicoli più intelligenti aprirebbe molte possibilità, sia in termini di organizzazione degli impianti, sia per i notevoli risvolti economici derivanti dall'abbattimento di molti dei costi fissi associati. La logistica è infatti un settore di nicchia, in cui i costi della tecnologia rappresentano il vero limite per la sua diffusione, costi dovuti in principal modo ai limiti tecnologici attuali. Il lavoro è quindi finalizzato a realizzare una tecnologia autonoma, utilizzabile direttamente a bordo degli AGV moderni, con modifiche minime in termini di hardware e software.

Gli elementi che hanno consentito di sviluppare tale dispositivo sono l'approccio multi sensore e data-fusion.

La tesi inizia con l'analisi dello stato dell'arte realativo al settore della logistica automatica, concentrandosi principalmente sulle applicazioni e ricerche più innovative legate all'automatizzazione delle fasi di carico/scarico merci nei moderni impianti logistici. Ciò che emerge dall'analisi è l'esistenza di un divario tecnologico tra il mondo della ricerca e la realtà industriale: i risultati e le soluzioni proposte dal primo sembrano non soddisfare i requisiti e le specifiche della seconda.

La seconda parte della tesi è dedicata ai sensori utilizzati: telecamere industriali, scanner laser di sicurezza planari 2D e telecamere 3D a tempo di volo (TOF). Per ciascuno di questi dispositivi un processo specifico (ed indipendente) è sviluppato al fine di riconoscere e localizzare Euro-pallet: le informazioni che gli AGV necessitano per eseguire la presa di un oggetto sono le tre coordinate che ne definiscono la posa all'interno dello spazio 2D, $[x, y, \theta]$*, la posizione ed assetto. L'attenzione è indirizzata sia alla massimizzazione dell'affidabilità degli algoritmi sia alla capacità di fornire una corretta stima dell'incertezza dei risultati. Il contenuto informativo associato all'incertezza rappresenta infatti un aspetto fondamentale per questo lavoro, in cui la caratterizzazione probabilistica dei risultati e l'adozione delle linee guida del settore della metrologia sono la base per un nuovo approccio al problema. Ciò ha permesso sia la modifica di algoritmi dello stato dell'arte che lo sviluppo di nuovi, realizzando un sistema che, nell'implementazione e testing conlusivi, ha dimostrato una affidabilità nel processo di identificazione sufficientemente elevata da conformarsi agli standard industriali, 99% dei casi positivamente identificati.*

La terza parte tratta la taratura del sistema. Al fine di garantire un processo affidabile di identificazione e presa è infatti fondamentale valutare le relazioni che intercorrono fra i diversi dispositivi di misura, taratura sensore-sensore, ma anche mettere in relazione i risultati ottenuti con la macchina, taratura sensore-robot. Queste tarature sono passaggi critici al fine di caratterizzare la catena di misura tra oggetto di riferimento ed il controller del robot. Da questa catena dipende infatti l'accuratezza nello svolgimento della procedura presa e, più importante, la sicurezza dei tale operazione.

La quarta parte rappresenta l'elemento centrale della tesi, la fusione delle identificazioni ottenute dai diversi sensori. L'approccio multi sensore è una strategia che

permette di superare eventuali limiti operativi dovuti alle capacità metriche dei singoli sensori, prendendo il meglio dai diversi dispositivi e migliorando in questo modo le prestazioni dell'intero sistema. Ciò è particolarmente vero nel caso in cui vi siano sorgenti di informazione indipendenti, le quali, una volta fuse, forniscono risultati molto più affidabili rispetto alla semplice comparazione dei dati. A causa della diversa tipologia dei sensori utilizzati, cartesiani come il laser e la TOF, e prospettici come la camera, una specifica strategia di fusione è stata sviluppata. Il principale vantaggio derivante da tale processo è l'affidabile reiezione dei possibili falsi positivi, i quali potrebbero causare situazioni molto pericolose come l'impatto con oggetti o peggio.

Un ulteriore contributo di questa tesi è la previsione del rischio nella manovra di presa: conoscendo l'incertezza nel processo di identificazione, in taratura e nel moto del veicolo, è possibile valutare l'intervallo di confidenza associato a una inforcata sicura, quella che avviene senza impatto fra forche e pallet. Ciò è essenziale per la logica decisionale del AGV al fine di garantire una macchina sicura durante tutte le fasi di lavoro.

L'ultima parte della tesi presenta i risultati sperimentali. Le tematiche elencate sono stati implementate su un robot reale, testando il comportamento degli algoritmi sviluppati in diverse condizioni operative.

Abstract

A robot is a machine that embodies decades of research and development. Born as a simple mechanical devices, these machines evolved together with our technology and knowledge, reaching levels of automation never imagined before.

The modern dream is represented by the cooperative robotics, where the robots do not just work for the people, but together with the people. Such result can be achieved only if these machines are able to acquire knowledge through perception, in other words they need to collect sensor measurements from which they extract meaningful information of the environment in order to adapt their behavior.

This thesis speaks about the topic of the autonomous object recognition and picking for Automated Guided Vehicles, AGVs, robots employed nowadays in the automatic logistic plants.

The development of a technology capable of achieving such task would be a significant technological improvement compared to the structure currently used in this field: rigid, strongly constrained and with a very limited human machine interaction. Automating the process of picking by making such vehicles more smart would open to many possibilities, both in terms of organization of the plants, both for the remarkable economic implications deriving from the abatement of many of the associated fixed costs. The logistics field is indeed a niche, in which the costs of the technology represent the true limit to its spread, costs due mainly to the limitations of the current technology. The work is therefore aimed at creating a stand-alone technology, usable directly on board of the modern AGVs, with minimal modifications in terms of hardware and software.

The elements that made possible such development are the multi-sensor approach and data-fusion.

The thesis starts with the analysis of the state of the art related ot the field of the automated logistic, focusing mostly on the most innovative applications and researches on the automatization of the load/unload of the goods in the modern logistic plants. What emerges form the analysis it is that there is a technological gap between the world of the research and the industrial reality: the results and solutions proposed by the first seem not match the requirements and specification of the second.

The second part of the thesis is dedicated to the sensors used: industrial cameras, planar 2D safety laser scanners and 3D time of flight cameras (TOF). For every device a specific (and independent) process is developed in order to recognize and localize Euro pallets: the information that AGVs require in order to perform the picking of an object are the three coordinates that define its pose in the 2D space, $[x, y, \theta]$*, position and attitude. The focus is addressed both on the maximization of the reliability of the algorithms and both on the capability in providing a correct estimation of uncertainty of the results. The information content that comes from the uncertainty represents a key aspect for this work, in which the probabilistic characterization of the results and the adoption of the guidelines of the measurement field are the basis for a new approach to the problem. That allowed both the modification of state of the art algorithms both the development of new ones, developing a system that in the final implementation and tests has shown a reliability in the identification process sufficiently high to fulfill the industrial standards, 99% of positive identifications.*

The third part is devoted to the calibration of system. In order to ensure a reliable process of identification and picking it is indeed fundamental to evaluate the relations between the sensing devices, sensor-sensor calibration, but also to relate the results obtained with the machine, sensor-robot calibration. These calibrations are critical steps that characterize the measurement chain between the target object and the robot controller. From that chain depends the overall accuracy in performing the forking procedure and, more important, the safety of such operation.

The fourth part represent the core element of the thesis, the fusion of the identifications obtained from the different sensors. The multi-sensor approach is a strategy that allows the overcome of possible operational limits due to the measurement capabilities of the individual sensors, taking the best from the different devices and thus

improving the performance of the entire system. This is particularly true in the case in which there are independent information sources, these, once fused, provide results way more reliable than the simple comparison of the data. Because of the different typology of the sensors involved, Cartesian ones like the laser and the TOF, and perspective ones like the camera, a specific fusion strategy is developed. The main benefit that the fusion provides is a reliable rejection of the possible false positives, which could cause very dangerous situations like the impact with objects or worst.

A further contribution of this thesis is the risk prediction for the maneuver of picking. Knowing the uncertainty in the identification process, in calibration and in the motion of the vehicle it is possible to evaluate the confidence interval associated to a safe forking, the one that occurs without impact between the tines and the pallet. That is critical for the decision making logic of the AGV in order to ensure a safe functionality of the machine during all daily operations. Last part of the thesis presents the experimental results. The aforementioned topics have been implemented on a real robot, testing the behavior of the developed algorithms in various operative conditions.

Keywords: [Robotics, Sensor fusion, Calibration, Object recognition, Localization, Logistic]

A tutti i miei mentori.
Passati, presenti e futuri.
Primi fra tutti i miei genitori.

Acknowledgements

I would like to express my thanks to my tutors, especially to Prof. Mariolino De Cecco, the head of the Measurement Instrumentation and Robotics (MIRo) Laboratory, without whom I would never have visited and have become a member of MIRo. He deserves my deepest gratitude for adopting me into the group, for providing excellent research and social facilities. I feel very privileged to have had the opportunity to study and work in this excellent research group.

I am greatly indebted to my co-advisor Dr. Luca Baglivo. He has skillfully guided me throughout my research with his supportive attitude, vast scientific knowledge, capacity, and sympathy. His continuous support, patience, willingness, considerable help, and guidance have been of outmost importance for finishing my PhD study.

A special thanks goes to Ing. Antonio Selmo, whose support was for me more than fundamental for what has been achieved in this thesis.

The work would have been impossible or at least much less enjoyable to complete without the support, assistance, and discussions with my colleagues and friends: Dr. Francesco Setti, Dr. Mattia Tavernini, Nicolò Biasi, Michele Confalonieri and Enrico Zappia.

Alberto

CONTENTS

LIST OF FIGURES

CHAPTER 1

INTRODUCTION

"Any sufficiently advanced technology is indistinguishable from magic." Arthur C. Clarke

1.1 Preface

AUTOMATION can be defined as one of the most pervasive and influencing field for our life. In the last three decades the improvements and the developments that Research achieved in robotic, mechatronics and automation in general had in most cases a direct impact on us. It is sufficient to think to the improvements in automotive, the birth of specific fields like the domotics, innovative medical tools etc. Many of these applications bring with them not just a technological improvement but also a deep change in the way of thinking, and imaging, our life.

Robotics is maybe the most important and well known representative of the automation among the different technological fields, from the fiction to real life application. In every robotic device a critical element is the perceptual system. From it depends the entire behavior of the machine and the capability in interacting with the environment. As humans, a robot needs to collect information to accomplish tasks. From the complexity of the tasks derives the *quantity* and/or the *quality* of information to be acquired.

For the older machines, in the past decades, that element was satisfied with relatively simple devices and strategies: the tasks of a robot were in most cases well

defined, no explicit interaction or versatility were required. It is indeed not unusual to see a robot of the 90's locked inside a cage in order to keep the people outside a potentially dangerous area. Those robots were thought to be isolated from humans and from the environment.

Such structure, however, is became today disadvantageous. The new economical policies caused a deep change in the production system, marking as new objective the *lean* organization. The modern edge technology is no more sufficient, and it must be improved toward an higher level of automation. What people need now are robots capable to operate and cooperate together, machine-machine and machine-human, to modify their operative behavior depending on the situation, robots that are fully autonomous in taking choices, requiring less supervision.

In order to develop such technology an answer must be provided to some specific questions:

- *how can the perceptive level of robotic systems be improved?*
- *how does a robot deal with situations that cannot be controlled or planned a priori?*
- *what is the grade of automation required?*
- *how far does the human interaction go?*

These questions represent the motivation of the modern Research in every robotic field, from the manipulators to the mobile robots.

1.2 Robot perception

Ordinary operations can be achieved by humans thanks to our capably in understanding an manipulating the environment around us. The same cannot be said for a machine. In order to positively achieve a given task a robot must be developed and organized specifically for that purpose: more general is the task more complex becomes the development.

A key element in every automatic system is the capability in collecting information. The *quantity* and the *quality* of the incoming data are related to the perceptive performances of the machine: the link between the environment and the logic of the robot.

Perception is achieved through sensors. A sensor is a device that measures or detects internal robot conditions (proprioceptive sensors) or external environmental conditions (exteroceptive sensors). Proprioceptive sensors are, for instance, wheel encoders, compass, inclinometers, accelerometers, and gyroscopes. Exteroceptive sensors are, for instance, cameras, laser range finders, and sonar. There are a wide variety of sensors used in mobile robots, this thesis focuses on three: **cameras, 2D laser scanners, 3D Time Of Flight cameras (TOF)**.

1.3 Motivation and objectives

The objective of this work is the development of the autonomous picking of objects by an Autonomous Guided Vehicle using no priori information. In this thesis are analyzed the topics of objects identification, localization and picking inside a generic, not structured, environment. These are related to transversal subjects necessary to make the robot operative and able to fulfill the task, like the multi-sensor calibration and data fusion, or to ensure the safety of the operation, like the uncertainty analysis or the safety check. The primary goal remains however the development of a functional device that ensures a reliable source of information for real industrial machines.

The main benefits that such technology could offer are an higher level of versatility in the ordinary operations and the freedom to work in less constrained environments. These features could have a relevant impact on the modern technology, improving it but also radically lowering its costs, now prohibitive.

Here are listed and briefly described the objectives of the thesis.

Object identification and localization

The grade of versatility and autonomy of a robot comes from the quantity of constrains required to make it works properly: less they are more the robot is considered *autonomous*. For this research, the sought grade of autonomy is represented by the capability of an AGV to recognize the target object(defined a priori) without external informations, or the help of the constraints that nowadays can be found in the plants.

The main objective is the object localization, from that derives the motion of the AGV and the picking of the object. The localization is however dependent on the capability of the system to identify the presence, or not, of the object. Since the involved field is the industrial one, the main characteristic that the device must ensure is the reliability of the results: for the purposes of the application, *the better solution is the one that provides no answer rather than a wrong one*. It is indeed acceptable to miss some identifications, it is instead not the occurrence of any dangerous action derived from a wrong identification.

Such result can be achieved by different design choices: increasing *number* of sen-

sors, increasing the *quality* of the sensors (so as the costs), or both. From that choice depends the overall affordability, and so the spread, of the developed technology.

3D data analysis

There are technologies and sensors that till the last decade had prohibitive costs, these are now reaching an affordable cost for the industrial use. An example is the time of flight camera, a sensor that provides both a 3D depth map of the environment and a 2D grayscale image.

One objective of this thesis is to compare the performances of the 2D laser scanner and the 3D TOF in order identify the best solution for the application. The criteria of the choice rely on the reliability, robustness and repeatability of the results.

Despite the TOF is an innovative technology, it is very advantageous because it embeds in a single device both Cartesian and perspective data, resulting more compact and versatile compared to a multi-sensor solution like the laser and the camera.

It must be however underlined that this technology can not substitute to the planar lasers on the AGVs because of safety reasons (TOF camera are not yet qualified as safety devices). This part of the work is therefore finalized to a prototypal development rather than an industrial implementation.

Data Fusion

The multi-sensor approach covers from the will to combine the positive characteristics of different sensors and at the same time to surpass their respective limitations or drawbacks.

That is a common practice for sensors that share the same measurement typology, for example accelerometers and GPS (both are metric sensors), but it is quite uncommon for those that do not share it, a camera (perspective sensor) and a laser (Cartesian). For this second case a specific fusion strategy must be developed in order to match the information provided by the two devices, defining a new data structure shareable among them.

For the purposes of this thesis the main objective is to fuse the identifications ob-

tained by the sensors. Since the identifications involve measurements it is suitable that the fusion includes the influence of the uncertainties associated to input results.

Vehicle to sensors calibration

The goal of the work is to develop a solution that enables the AGV to autonomously locate and *pick* pallets placed inside a cargo area of a warehouse. In order to make this task possible it is necessary to evaluate the extrinsic parameters that relate the position of the sensors on board with the reference system of the forklift. These parameters can be evaluated by means of a dedicated calibration, operation that is critic because from it depends the overall accuracy of the forking procedure.

Such calibration is not common and there are limited number of similar cases in literature. Usually the nominal values are kept as reference because sufficiently accurate for the purposes of the application. For the current task, the accuracy in determining these parameters finds a direct dependency not only on the performances but also on the safety of the process. An high accuracy in estimating where the sensors are on board means a low probability in the occurrence of an impact between the forks of the AGV and the pallet.

Safety check

Given an identification, no informations are provided about the safety of the maneuver that the AGV will perform in order to pick the object.

One important feature that the device must include is the evaluation of potentially dangerous situations, in which the uncertainty associated to the calibration, the identification process, the vehicle parameters and the measurement chain in general could cause an impact between the forks and the pallet feet. Such analysis must consider the geometry of the system and be able to provide a confidence interval, a threshold or a score usable by the control logic of the AGV for the decision making process, evaluating if the task is suitable for the accomplishment or not.

1.4 Main Contribution

From state of the art it can be noticed that there is a very strong interest in this topic. Many examples of similar applications can be found till the early 90's. Nevertheless it must be pointed out that none of these works has found a real technological transfer from the research field to the industrial world. No traces of innovative products can be found on the sites of the main AGV manufacturers, the same that financed, and are still financing, most of the aforesaid works/projects. That leads to two possible analysis of the facts:

- there is no real interest in this application
- the solutions proposed till now do not satisfy the requirements of the industrial world

Considering that there are still works and project ongoing about this topic, the second hypothesis can be assumed as the correct one.

This thesis analyzes the problem, introducing a new solution based on the use and combination of different methodologies. Many of the works related to autonomous pallet picking are the result of researches from the fields of computer science and robotics, mostly focused on the codification of algorithms rather than a proper industrial development. The contribution of this thesis comes instead from the measurement field, focusing on the analysis of the parameters of influence over the different processes and the probabilistic characterization of the results.

First element to underline is the development of a strategy that uses or two complementary sensor like a 2D planar laser and a camera, or a single time of flight camera, applying innovative modalities of data fusion entirely based on the probabilistic characterization of the identifications obtained independently from the different information sources/devices.

Another contribution is the calibration between the Cartesian sensor, the laser scanner or the TOF, and the vehicle. It a fully automatic process that evaluates the extrinsic parameters of a couple, or more, of sensors from a their synchronous motion. For the purposes of the current application, and the sensor used, this calibration

has been structured in order to use the environment as source of information.

A further important topic it is the estimation of the risk for a given maneuver. Thanks to the analysis of uncertainties involved in the measurement chain, combined with the path planner of the AGV, it is possible to evaluate if the maneuver of picking is safe or not. Such predictive model is based on a probabilistic characterization of the final displacement error between target and reached pose after a maneuver. That is useful to avoid the impact with the pallet, but it can also applied to those applications that involve the motion of the vehicle in regions with limited maneuvering space.

The result is a fully working device, ready to be installed and used on most of the modern AGVs.

1.5 Outline of the thesis

The outline of the thesis is the following.

Chapter 2

It is presented the state of the art related to the automatic logistic. Different works are presented specifying the influences or the points of interest associated to the actual research, classifying them about the technology used or developed. In this chapter are also presented the industrial cases that show technological improvements related to the topic of this thesis: the development of a more advanced and *smart* AGV.

Chapter 3

The sensors are the perceptive part of an automatic machine. In this chapter are described the sensors used and the algorithms developed in order to identify the pallet. For each device are analyzed the performances of the identification process in terms of accuracy, repeatability and efficiency. In every process in addition to the identification of the object is performed the analysis of the uncertainty of the results, a key element for the data fusion and the risk evaluation.

Chapter 4

The multi-sensor approach needs the knowledge of the geometric parameters that relate different devices of the system. To achieve the picking of the pallet the measurement chain between object, sensor and AGV must be known. These informations can be evaluated through dedicates calibration processes. In this chapter are presented the motivations and the modalities that ensure the required grade of accuracy in determining such geometric relations between sensors and AGV.

Chapter 5

In order to ensure reliable results from the identification process the multi-sensors strategy is adopted, such architecture implies the fusion of the data in order to be

effective. The chapter initially focuses on how the fusion can be achieved between Cartesian and perspective devices, then are described the choices that make the developed strategy, and technology, suitable for the industrial requirements in terms of reliability and robustness.

Chapter 6

In order to analyze the entire task of identification and picking of a pallet from a perspective of operative safety, a specific analysis, based on the influence of the uncertainties factors and error budget, is performed to evaluate safety of the forking maneuver. The logic behind this operation involves the geometry of the pallet, the geometry of the vehicle and it is aimed to evaluate the risk of having an impact between the forks and the pallet feet.

As final part of the research the experimental results are presented in this chapter. In order to test and verify the performance of the algorithms developed an experimental robot was configured and used to simulate the behavior of a real AGV.

CHAPTER 2

STATE OF THE ART

The state of the art of the automated logistic is very rich from the point of view of the research but very poor in terms of actual technological transfer and on-field implementation.

2.1 Introduction

ONLY few among the many examples of success of the automation are known by people. Most of them, and probably the most important ones, took place in the industrial field, with limited visibility, and the appreciation only by the experts of the field.

The production system was and is the economic motor of our society. Our lifestyle is deeply dependent on its efficiency and the capability in creating goods from which follow richness and wealth. The technological growth related to this field finds its birth with the fist industrial revolution during the second half of the 1700, and till now it has not yet stopped. Every year a huge amount of resources are spent to strengthen our technologies, to increase the production rate, to improve the versatility of our factories. In some cases, these bring a direct benefit to the entire system.

There are however some particular niches, usually not well known, in which there is not the strength or the foresight in financing a development capable to cause a real improvement of the modern technology. One of these is the logistic and the related automation.

During the 90's a new way to organize and manage the warehouses of the factories

took place. That caused the birth of a new mobile robot, the so called AGV, Automatic Guided Vehicle, fig. 2.1. This machine was developed in order to substitute the fork-lifts guided by humans, they were designed to move autonomously the materials and the goods between the manufacturing lines or stock areas of a factory. That can be defined as the beginning of a new technological field: the automatic logistic.

Figure 2.1: Examples of modern AGVs

During the last decade this sector has grown thanks to important elements that have been highlighted from the spread of such technology. The main advantages that an automatic plant grants are:

- an AGV is safer than an human operator (it cannot be distracted, it does not get tired etc)
- an automatic plant in which are used AGVs is less expensive compared to one in which there are involved human workers, fig. 2.2
- the production rate is higher compared to the not automatized implants
- the AGVs replace a *low value added* jobs with higher level professional positions (technicians and programmers)

But there are also some drawbacks:

- an automatic plant is very expensive, at least hundreds of thousands of Euro
- an AGV is less versatile than an human operator
- once programmed, the automatic warehouse is difficult to be reorganized
- the setup of the system is critic and it is very expensive both in economic terms and time

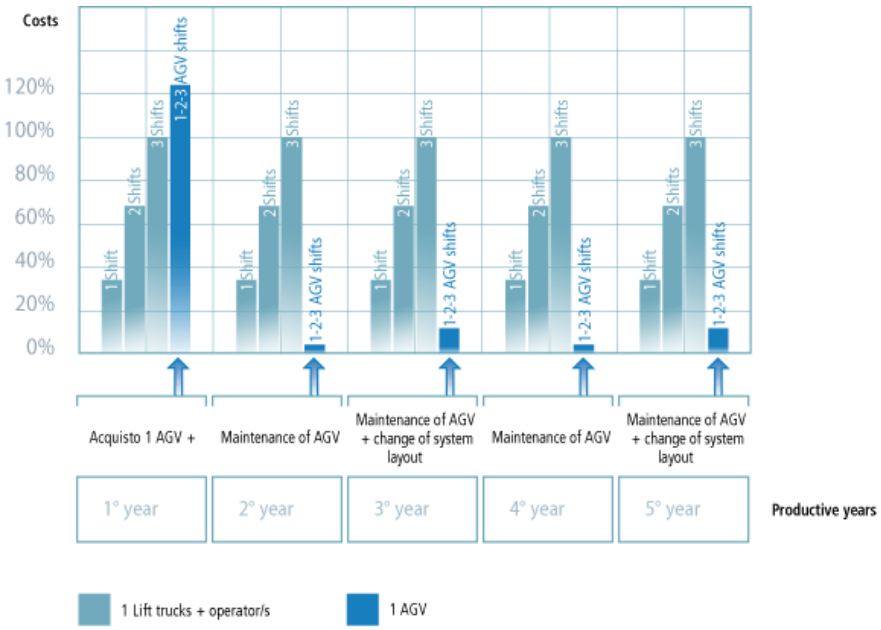


Figure 2.2: System AGV: return of investment analysis

For these reasons this technology is spreading, but slowly and only among those (big) enterprises that can afford the initial investment.

2.2 Automatic logistic

Nowadays the AGVs are used in many production environments such as the automotive industry, logistics, container harbors and so on. This thesis speaks of the automatic industrial logistic, field in which the main task of the AGVs is to load/unload pallets placed in the automatic plants in order to feed the manufacturing lines.

In these systems, the path of an AGV is usually strictly defined and can not be changed on demand. The environment in which the AGVs operates must be well constrained with pallets stored in structured stations with high repeatability, fig. 2.3. AGVs must also have a limited interaction with pedestrians/workers moving around. The installation of the system needs frequently a lot of reconstructions in the infrastructure or a brand new design of the production environment that foresee the AGV employment. That is due to many reasons. First is due to the installation of the localization system by distributing artificial landmarks (for example reflectors in known positions for laser triangulation) all over the motion area. Second to the safe path planning and control of the vehicles. Third to the installation of very accurate pallet stations able to cope with low-flexibility automatic systems. Therefore, AGVs are not often used in medium sized enterprises. Especially in non-structured warehouses the usage of AGVs, which can only follow a predefined path and where artificial landmarks cannot be always visible, is not possible.

The focus of the modern Research is the development of a small and flexible AGV for semi- or non-structured warehouses, capable in identifying pallet and performing the pinking in a more unconstrained way. This means that the pallet location inside the warehouse shall be simply defined by stock areas, fig. 2.4. The pallet location will no more be accurate and the AGV path will no more strictly defined being autonomously computed according to the actual pallet location.

For this aim the AGV shall be able to navigate from the unload till the load area also using natural landmarks, than identify the pallet (by exteroceptive sensors), compute the relative position with respect to it and to the environment, plan a feasible and safe trajectory to load the pallet by its fork. Then the vehicle shall be able to find the way back to the unload station.

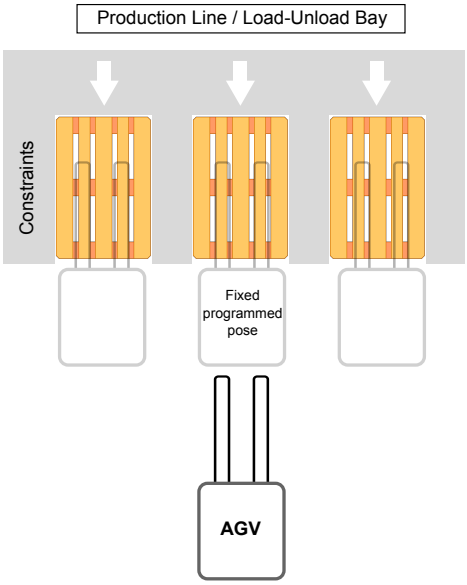


Figure 2.3: Warehouse: modern structure, constrained

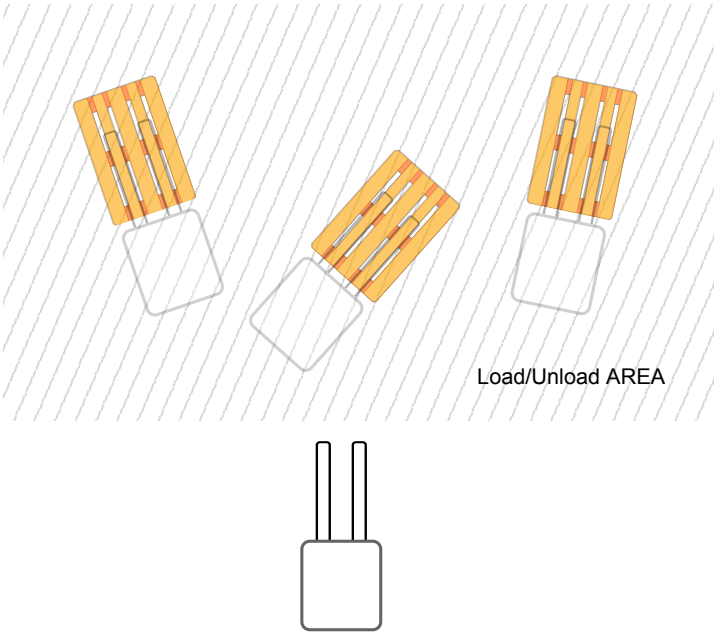


Figure 2.4: Warehouse: unconstrained load/unload bay

The system should need only a minimum of changes in the infrastructure of the warehouse. Such AGV must be able to communicate with a central control station that manages the overall logistics. The control station submits only the pallet identifier for the pallet which has to be picked up and unload the station. Then the system must be able to find the exact position and orientation of the pallet, navigating also using natural landmarks, calculate a flexible path to pick up the pallet, control accurately the path and unload the pallet autonomously.

Many non-automatic warehouses foresee pallet in not accurate locations, pedestrian workers moving around, while trucks arriving at parking zone being unloaded by man-driven transpallet. This is a typical case of non-structured environment where a fully autonomous and *smart* AGV could find the best utilization.

On the contrary, actual automatic warehouses are typically well structured for pallet locations (typical repeatability in the order of some millimeters), path strictly planned a priori, limited interaction AGV/workers. In these scenarios the versatility of a robotic vehicle is seen as a negative feature for the plant, in which the management and control of the fleet is base on the knowledge in every instant of the position of the AGVs along their defined paths. The overall structure in such a configuration is much more expensive than the non-structured case.

2.3 State of the art in logistics

There are two levels of state of the art. First is the academic level were many studies can be found about AGV, with publications also regarding the current research. The second is the industrial one were there is almost nothing of commercially available to recognize a pallet, plan and control a flexible path to reach and load it. Different reasons lie behind this situation, mainly related to the technological requirements that such application needs.

The power of processors in a recent past was not enough to solve in real-time optimization problems within reasonable time constraints. Those capabilities are increasing over time. The same about the accuracy and resolution of the instruments, it was not enough to guarantee an accurate estimation of the object location. Laser scanners had a resolution of 1°, while now they reach less than 0.25°. Laser angular resolution is directly proportional to the number of points collected on the pallet. For example, at a distance of 5 meters with the older resolution only 10 points of the pallet can be collected, with recent instruments about 40 and more. It is easy to estimate that the ratio of accuracy obtained is 4 times higher compared to a configuration in which nor recognition neither localization could be obtained. Similarly, in the past, cameras had a number of pixels that could be inadequate to cover the whole field of view in front of the vehicle. The number of pixels of industrial cameras are also increasing leading to a situation very similar to the laser capabilities.

There are other reasons than technology, related to the methodology used for the development. Several methods were defined in order to localize the pallet, but none of them copes with the **99.9**% of reliability required by the industrial applications. Each method has one or several weak points that makes it a valuable algorithm/system for research but not an effective industrial application. A low reliability in the identification process implies a high level of maintenance and frequent monitoring, with an higher probability of failure and danger for the plant.

But, if these different methods are compared, it is possible to discover that they are in many cases complementary to each other. This means that, if different approaches are simultaneously taken into account, and there is a capability to estimate the sta-

tus/accuracy of each method, it is also possible to combine them in order to have a system with a higher level of reliability. The basis of the sensor-fusion methods.

2.3.1 Academic research

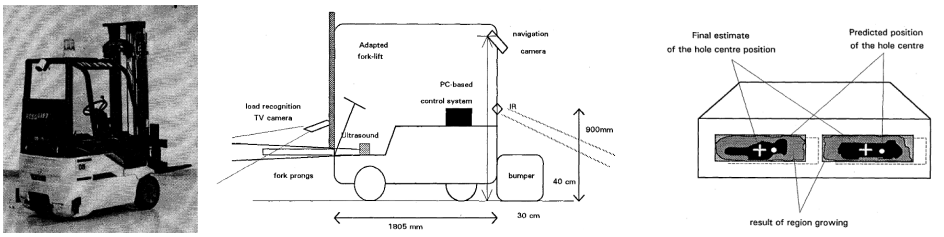
There is a huge production of publications in the field of AGVs in every aspect of its design and control. For the objectives of this thesis the interest focus on the analysis of those concerning objects/pallet identification and localization. As far as it is impossible to detail all the work done only the most relevant references will be provided. These works differs among them mostly for the sensor used to perform the identification of the objects. The devices are 2D laser range finders or cameras, but alternative cases can be found.

2.3.1.1 Camera

The earliest works are all based on the use of industrial cameras, and the most famous example is the ROBOLIFT, Garibotto et al. (1996, 1997), fig. 2.5. ROBOLIFT was the first real attempt in developing a fully autonomous vehicle based on the perception on the environment through artificial vision. The techniques described in these works are almost the same that can be found in the subsequents: the identification of the pallet is achieved by recognizing the dark spots generated over the image by the gaps between the pallet feet. This approach works properly only in those situations in which a priori knowledge of the position of the pallet is given, allowing the algorithm to isolate and process only specific areas of the input image, Regions of Interest (ROI). This suggests that the logic implemented with ROBOLIFT was mostly aimed to check the presence of the pallet in a known position instead of performing a real recognition and localization.

Another remarkable work is Seelinger and Yoder (2005), which differs from the previous for the use of markers on the pallet and on the tines of the AGV. This approach seems to be more robust but it totally misses of generality: it is indeed unthinkable to customize all the pallets of a plant to achieve the autonomous picking.

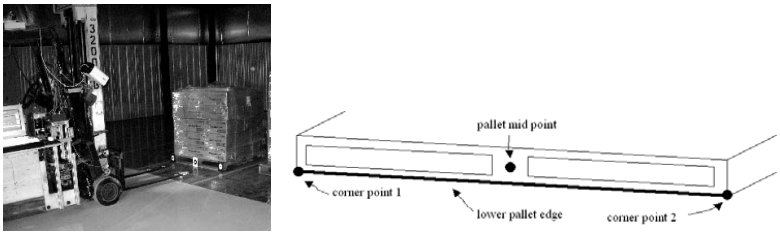
The authors present an improvement of the system, Seelinger and Yoder (2006),



(a) ROBOLIFT (b) Schematic (c) Region growing method

Figure 2.5: ROBOLIFT by Garibotto et al.

which avoids the use of markers and instead analyzes the lines created by the edges of the pallet on the ground. The identifications of the edges in an image is a technique commonly used in computer vision applications to identify objects or ROIs. The light is a critical parameter for this technique, which uses the gradients evaluated from the intensity levels of pixels of the image, usually monochromatic, to determine the points in which there is a sharp transition between light and dark, and probably an edge of an object. This approach is a good example of academic research that does not fulfill the industrial requirements: the identification of a line in an image is a task that is strongly dependent on the illumination of the scene, the correct evaluation of the gradients, the noise removal etc. Those are all elements that are difficult to be controlled in an automatic plant unless to constrain the environment.



(a) AGV prototype (b) Pallet detection by edge extraction

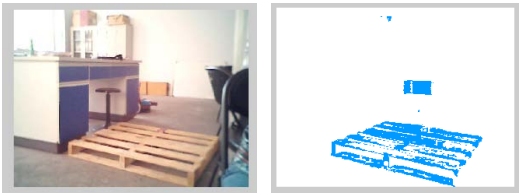
Figure 2.6: Seelinger et al.

An alternative approach that avoids the use of the gradients it is the segmentation of by means of colors. The logic is to characterize the color of the target object in terms of hue, saturation and brightness(HSV) and then define a set appropriate

tolerances, a model, useful to isolate inside the image only those pixels that match to such color values. This methodology has been used in two works, Pagés et al. (2001); Cui et al. (2012), showing good experimental results,fig. 2.7 fig. 2.8.

(a) Input image (b) Color segmentation (c) Line extraction

Figure 2.7: Pages et al.



(a) Input image (b) Color segmentation

(c) Gradient (d) Line extraction

Figure 2.8: Guang-zhao et al.

The chromatic segmentation is however related to the environmental conditions too, like in the previous cases. The development of a model of color suitable for a repeatable chromatic segmentation it is very complex and often needs a trade-offs between completeness of the silhouette of the object selected and inclusion of wrong pixels with similar colors close to the object itself: strict thresholds cause fractioned segmentations, slack thresholds cause the selection of pixels that do not belong to the object.

There is also further critical element to underline, the most common pallets, the wooden ones, have a high color variability, intra-class(a pallet could be brighter or

darker depending on the wood used, humidity etc), and inter-model(the wood could have stains, cracks, etc.). In some cases, the manufacturer color their pallets, but this is not a general solution to the problem of identification.

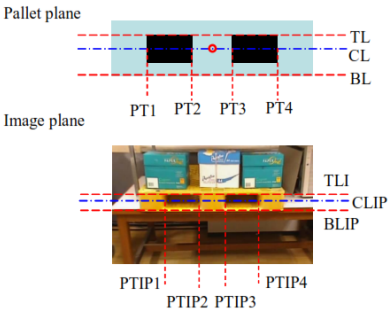
More recent works are instead based on more sophisticated analysis of the images, like the line extraction combined with an associated shape reconstruction and matching, Sungmin and Minhwan (2008). The results are convincing, but one element must be pointed out. In this work the authors use plastic pallet. That is indeed a critical aspect because the pallet made with such material have a clean and regular surface. The main benefit is a more stable and continuous chromatic distribution of the silhouette on the image, with more exalted gradients and so an *easier* detection of the edges of the pallet. Such configuration involves a more stable identification and localization process compared to the previous examples. However the same performances cannot be obtained if applied on a common wooden pallet: the cracks and the colors of the wood cause inhomogeneous surfaces, and so a line detection that is not stable and reliable. As shown in fig. 2.10, the gradient operator generates many lines around the pallet, but also inside of it, an input image very different from the one presented in the aforementioned paper.

A different approach can be seen in the work of Cucchiara et al. (2000). In this case the identification is achieved using a template matching based on the position of the corners around the frontal gaps of the pallet, the aim is to evaluate if the camera is pointing the right object. Such technique is interesting, allowing a more general strategy for object recognition: as for the humans, a detector should *learn* by it own the features of the objects, without referring to externally provided geometrical properties. The *learning* allows a higher level of the versatility by ensuring a faster adaptation of the software to new target objects.

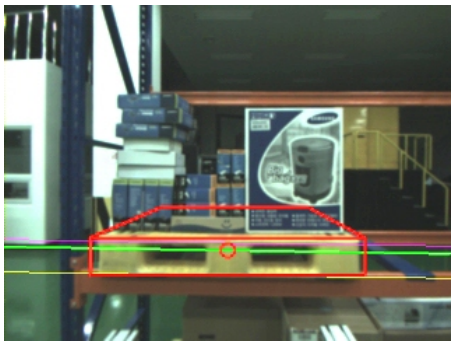
The main drawback is however the usage of a very tight ROI to trim part of the image, both to speed up the algorithm both to make the processing local: the pallet must be placed in a fixed position and at a given distance in front of the vehicle in order to have a silhouette of the frontal face of the pallet with the right dimensions for the ROI. This method is therefore functional only in those cases in which the pose of

(a) Gradient and color segmetnation

(b) Line extraction



(c) Perspective lines matching



(d) 3D pose

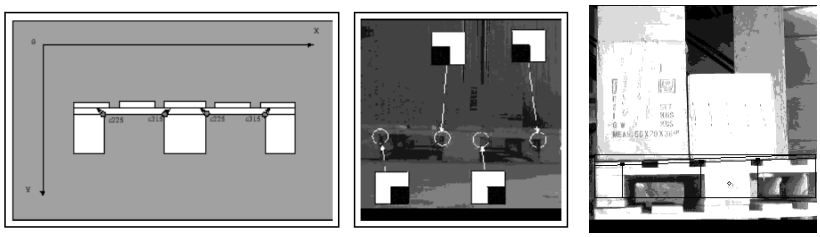
Figure 2.9: Byun et al.

(a) Original image

(b) Gradients

Figure 2.10: Wooden pallet example

the pallet is known a priori, in which it is just necessary a check before the picking.



(a) Features from pallet (b) ROI

Figure 2.11: Cucchiara et al.

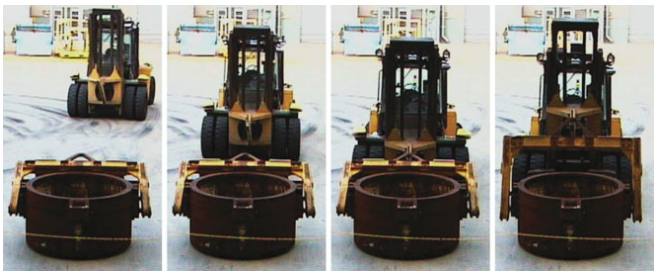
In order to automate the process of picking a more general approach is required: no narrow ROIs should be defined, the recognition of the object must be independent from the position of the pallet inside the image.

Other than pallets

A interesting example of a successful industrial application that takes advantage of the computer vision to achieve the autonomous object localization and picking is the work of Pradalier et al. (2008). In this paper is presented the strategy used to automate an Hot Metal Carriers(HMCs), a large forklift-type vehicle used to move molten metal in aluminum smelters. The automation is performed recognizing and tracking two markers placed on the crucible, using these references the vehicle plans a trajectory and perform the picking. Admitting that the use of the markers placed on the target object is not a valid solution for the objective of this thesis, this paper remains a remarkable example for understanding how, and why, the automation can play an important role for those jobs potentially dangerous or fatiguing for the workers.

(a) HMC

(b) targets



(c) Autnomous picking

Figure 2.12: Automatic Hot Metal Carrier by Pradalier

2.3.1.2 Laser

Among the publications that deal with the topic of pallet localization through Cartesian sensors, like laser scanners or similar, a division can be made: some use 2D sensors, other 3D.

2D Laser Range Finder

Early works related to the use of laser scanner as sensor for object detection in logistic are the papers of Lecking et al. (2005, 2006). Lecking and Wulf present a work accomplished in collaboration with Still, a German manufacturer of forklift. The objective is to automate a manual vehicle, including among the various operations the autonomous identification and picking of Euro Pallets. The sensors used are two: a SICK S3000 (safety sensor) to localize pallet laying on the ground and a LMS200 (measurement sensor), mounted on the forks, to localize pallet at variable heights.

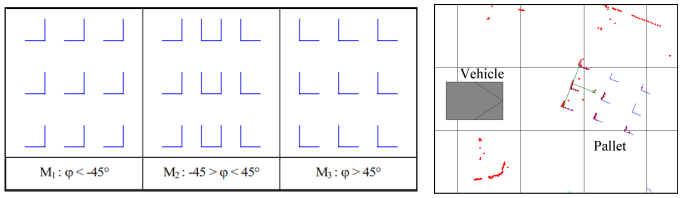
Two identification algorithms are presented. The first is aimed to a preliminary assessment, using the information of reflectivity associated to the laser scansion in order to identify metal strips attached to the external blocks of the pallet, fig. 2.13(b)(c).

The second, fig. 2.13(d)(e), aimed instead at the operating phase of the machine, using a point-to-point ICP, Iterative Closest Point, and three models of the pallet distinguishing partial representations of it. The proposed algorithm tries to minimize the displacement between the model of the pallet and the laser points belonging to the real one, performing in practice a 2D matching.

One important element must be underlined: the ICP algorithm used in this paper converges to the solution only if a *correct* initialization is given to the algorithm. The model of the pallet must be placed close to its real position in space, at least closer than $\pm$150 millimeters along the longitudinal and transversal directions and less of $\pm 15^{\circ}$ in relative orientation. The priori knowledge of the environment and the possible position of the pallet are therefore needed for usage of this method.

The results of this work are very convincing but there are no references of any technological transfer to an industrial machine as result of the research project.
A similar works is the one proposed by Zhendong et al. (2010), where a feature to

(a) Still AGV (b) Modified pallet (c) Metal plate detection



(d) Patrial models (e) ICP results

Figure 2.13: Lecking et al.

feature matching is used. The article is not well organized and it is not possible to really understand the final performance of the process. This works must be however considered as a further example of research applied to autonomous pallet picking.

More interesting is instead the work of Baglivo et al. (2008), where a more advanced algorithm is presented. Scan matching algorithms can be divided in two different classes: ICP-like (Iterative Closest Point) and algorithms that directly minimize an energy function.

The authors propose a method that belong to the second class, which doesn't require explicit correspondences as ICP and it is instead based on the numerical minimization of a function. This function is directly related to the fitting between measures and object model. Unlike the ICP that needs the scan and the model to be spatially near to each other (a few centimeters and a few degrees), the searching field for the unknown position of object that match current measured scan can be in this second case very large (some meters and tens of degrees).

The proposed energy-based method initially performs a feature extraction and then a direct search minimization aimed to solve the problem of undesired local minima

associate the *cost* functional employed. A ray casting sensor simulation is lastly used to build a convergence criterion that employs an ICP search and increase th accuracy of the localization process.

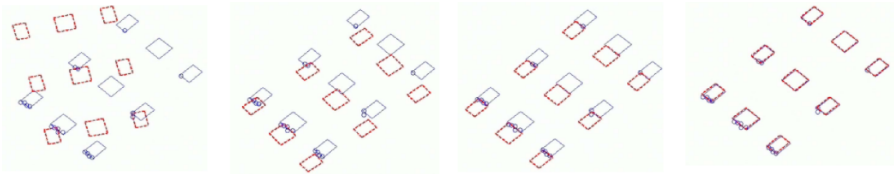
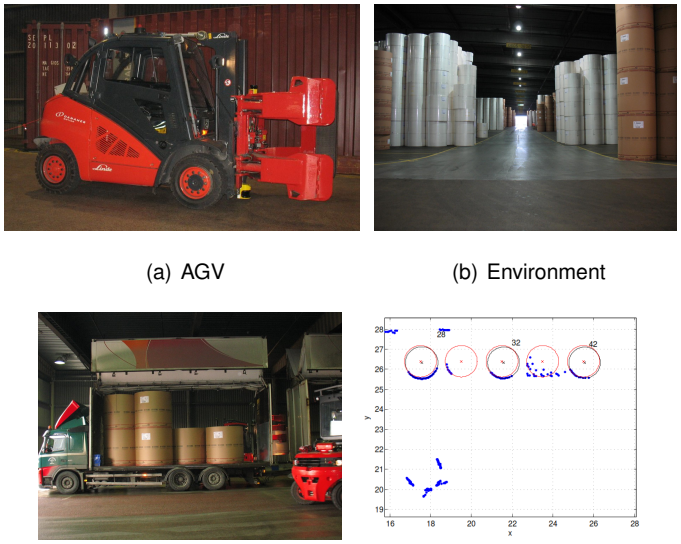


Figure 2.14: Matching result by Baglivo et al.

This paper represents the starting point of the current research.

Other than pallets

An interesting work is the project MALTA, Bouguerra et al. (2009). It presents a project aimed to automate the loading operations of rolls of paper in a paper mill. Although the operating conditions are simplified by the shape of the reference object, a cylinder of known sizes, this case can be considered success in terms of technology transfer from research to industry.



(a) AGV

(b) Environment

(c) rolls of paper to be loaded/unloaded

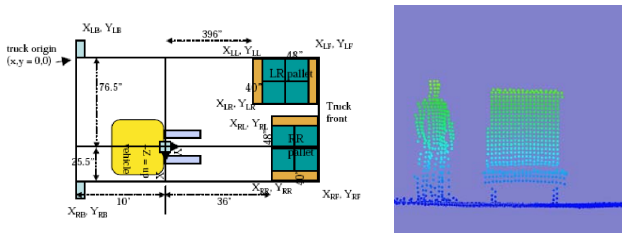
(d) Laser readings

Figure 2.15: MALTA project by Bouguerra

3D Laser Range Finder

About the 3D sensors there are the papers of Bostelman et al. (2006) and Karaman et al. (2010); Teller et al. (2010). The subject treated is the automatic unloading of pallets from a truck. The differences between these works are due to the operative conditions, from which were derived very different solutions.

In the work of Bostelman it is analyzed an industrial scenario: the truck reaches the cargo area of the plant and the rear part of the trailer becomes a gate in which the AGV should physically enters to unload the pallets. The aim is to automate the initial stage of the automatic manufacturing line: the unload of the material/goods that must be fed to the plant. The controlled operative condition allows the imposition of very strong constraints like the definition of the volumes of the loads or the position and the sizes of the trailer directly inside the reference system of the AGV. That simplifies the analysis of the data and the identification of the objects by just segmenting the 3D point and matching the volumes.

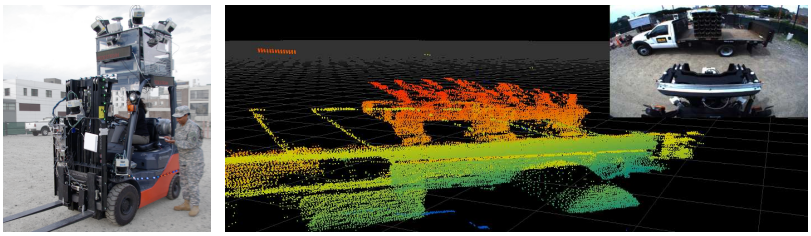


(a) Operative scenario (b) 3D data

Figure 2.16: Bostelman et al.

The second example is instead a project financed by DARPA, Defense Advanced Research Projects Agency, fig. 2.17(a). The objective is to develop an autonomous outdoor forklift capable to unload pallet from the side of a truck under the remote supervision of a not specialized operator. A planar laser scanner is fixed on the forks of the forklift and the 3D point cloud is generated by lifting the fork while recording

the data stream from the sensor. The identification of the pallet is achieved by recognizing the frontal face of the pallet from the 3D point cloud, vertical analysis, and the distribution of the points between the rear feet, depth analysis, fig. 2.17(b). The system presented is very interesting from the point of view of research but the type and the *number* sensors involved imply a cost that is not suitable for the industrial field.



(a) AGV (b) 3D data

Figure 2.17: Karaman et al.

An important element must be underlined: both the works use 3D data collected through 2D sensors. The strategy in collecting data is almost the same in both cases: moving a planar laser (rotating it in the first case, and shifting in the second) and building the 3D cloud according to the pose of the sensor. This choice comes from a lack of a suitable 3D technology: there are few 3D sensors, even less that can work outside, and almost all of them are very expensive.

2.3.1.3 Hybrid

The current research finds its birth in the works of De Cecco e Baglivo and the Project AGILE. Related to this projects are the papers Baglivo et al. (2009, 2011). Compared to the previous work of Baglivo, these present a new strategy based on the combined use of a camera and a 2D laser scanner.

These works describe two different stages of the research project. The first paper can be considered an evolution of the work of 2008. An improved version of the laser matching algorithm is presented. The innovation in this case is the first trial in combining the laser with a camera. The strategy adopted by the authors is a color segmentation associated to an adaptive model of color. Once evaluated the extrinsic parameters between laser and camera, the range data associated to the pallet are projected on the image, the points identified are then used as seeds for a region growing algorithm based on chromatic homogeneity. From that region are defined both the color and the thresholds that better model the chromatic variability of the pixels associate to the silhouette of the pallet in the image.

This algorithm is less dependent on the light condition compared to the others that use a fixed model of color. A further important aspect is that this algorithm is dynamic and so the model used for color segmentation can be adapted on request depending on the conditions of the environment, high/low illumination conditions, or the color of the pallets, colored or not. Must be however said that such segmentation is presented as a *visual filter*. No real identification is performed by the camera.

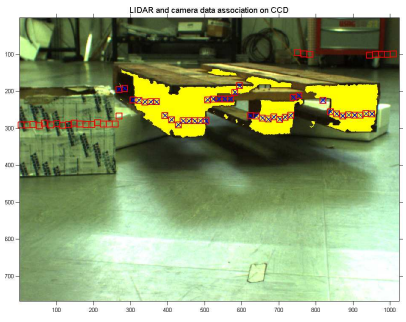
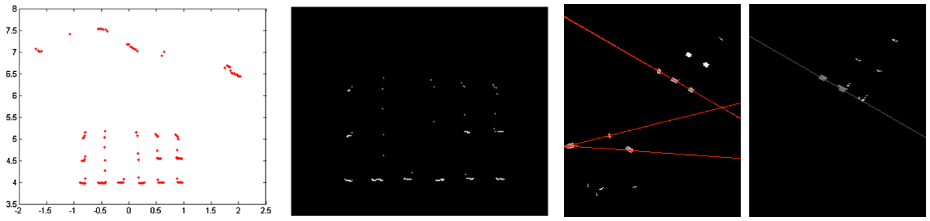


Figure 2.18: Dynamic color segmentation by Baglivo et al.

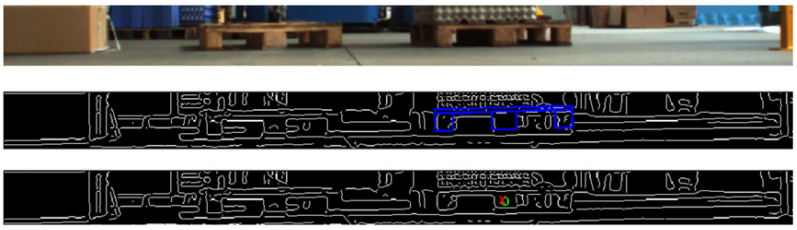
The work of 2011 presents a more advanced stage of the project with totally new strategies. About the laser side, a new algorithm based on the rasterization of the scansion is used: the range data are converted to a binary image and by means of image processing techniques, like convolution and line extraction(base Hough transform), the frontal pallet feet are identified, fig. 2.19(a)(b). The main advantage in this approach, compared to the previously used method (based on an energetic optimization), is the speed of elaboration, faster, and the computational cost, lower.

The second innovative element is the image processing algorithm achieved with the camera: a template matching based on the Chamfer/Hausdorff distance (more reference in chapter 3). The processing starts with the identification of the edges of the input image as a new binary image. Such image, which contains feature and non-feature pixels, is transformed by means of a distance function into an image in which each pixel value denotes the distance to the nearest feature pixel. After computing this new image, a template is convolved with it and the result is normalized by the number of edge pixels in the template. In practice, the template acts as a mask convolved with the image, which only selects the *distance* values of pixels corresponding to the edge pixels of the template; then the mean of those pixels is computed obtaining the position of the object in the original image fig. 2.19(d).

The two identifications are then compared performing the so called *camera consensus* in which the camera acts as control device for the information obtained from the laser. The results are convincing, but some negative elements can be highlighted: the elaboration is almost entirely based on the laser (asymmetry of the process), the video processing of the camera is dependent on a laser initialization. More details will be provided in chapter 3 and 5: processing and data fusion.



(a) 2D laser data rasterization (b) Laser identification



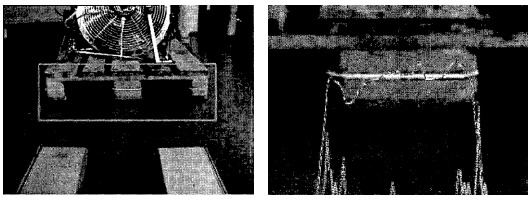
(c) Camera identification

Figure 2.19: Multi-Sensor strategy by Baglivo et al.

2.3.1.4 Other

Lastly there are some papers focusing on different techniques and devices.

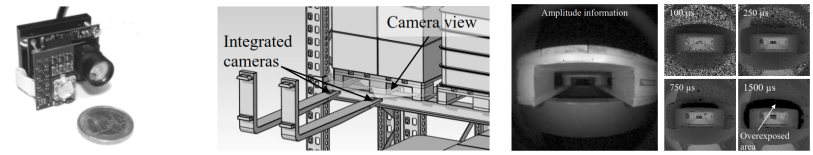
One examples is the work of Nygards et al. (2000), where a Sheet-of-Light range camera IS used to trace the profile of the frontal face of the pallet and achieve the matching.



(a) Pallet identification (b) Central foot identification

Figure 2.20: Nygards et al.

Kleinert and Overmeyer (2012) is paper in which a *time of flight camera* is used to check the presence of the pallet on a rack. The sensor has a limited field of view and it is used as safety control to avoid impact between the forks an the pallets.



(a) Senro used and configuration (b) Depth readings

Figure 2.21: Forking control system by Kleinert

2.3.2 Industrial systems

The automated logistics is one of those fields in which technology plays a key role. The differences in the development of an AGV rather than an forklift are many and radical, requiring a deep knowledge and expertise in mechatronics, robotics and automation. An AGV is indeed more similar to a robot than a forklift. That explains why the number of manufacturers is very limited.

Although there is standardized technology in the manufacturing of the AGVs, it is still possible to differentiate and characterize the various producers according to specific functionalities that the different machines and models offer.

RMT robotics is a Canadian enterprise that develops AGV for autonomous transportation and delivery in manufacturing and warehousing facilities. They declare to sell autonomous vehicles that are able to develop a map of the surrounding being guided by an operator during its initialization, than they are able to localize within this non structured environment by comparing a local map acquired via a laser scanner and the global map acquired with the human aid. They declare also to provide obstacle avoidance capabilities. No references about autonomous pallet picking or similar devices.

`www.rmtrobotics.com, www.adamsgv.com`

Mobilerobots inc is a partner of RMT robotics that offers more or less the same solutions.

`www.mobilerobots.com/ARCS.html`

Swisslog offers the TransCar LTC AGV operating inside hospitals environments. The vehicle uses laser guidance to localize the robot without artificial landmarks but in the presence of clear corridors. Even in this case the autonomous part is only related to the path planning and control.

`www.swisslog.com`

In the year 2000 **NDC Australia** announced the *Pallet Finder*. According to NDC magazine it consisted in a tool that permits an AGV to identify a pallet and engage its forks even when it is improperly placed. The system made use of a SICK laser scanner and a proprietary software. If one takes a look at the NDC site no Pallet Finder is enclosed in the products list. Neither using the search tool on the site there is an outcome. About the path planning no details are available therefore it is not clear if it was a flexible planning method embodied in the routine. Besides the previous information, the product seems not to exist. An hypothesis is that the developed system was not reliable for industry for the reasons explained before.

`http://www.ndcta.com.au`

The United States Patent n° 6952488, 4 October 2005, *System and method for object localization* describes a system based upon a camera and a method based upon an algorithm of pallet border lines extraction.

Bluebotics is a spin-off of the ETHZ. This company produces an high level technology that ensures the localization of the AGV using natural landmarks and a self constructed map. This reality represent the first real industrial case of state of the art technology applied in real products. Even in this case no autonomous localization is performed.

`http://www.bluebotics.com/`

Finally there are many industrial vehicles exploiting wire guidance, laser triangulation guidance with retroreflectors and that rely on well placed pallets for automation. In the first case its navigation consists of a wire buried inside the floor on which an AC current is fed while an antenna on the lower part of the vehicle senses the wire following its path. Flexibility is something not foreseen in this system although a high level of reliability must be underlined. In the case of laser guidance with retroreflectors the vehicle needs a structured environment but has the flexibility to redesign its path via PC therefore without deeply changing the environment as for the previous case.

From what has been said it is clear how the manufacturers of AGVs have focused on how maximize the performance of the plants sacrificing the versatility. This solution is due to the lack of a technology able to improve the level of artificial intelligence of vehicles, decreasing, or (even better) removing, the constraints currently needed in these structures. It is, however, remarkable that many initiatives, financed by the manufacturers themselves, were stared in the past, and still keep starting, with the aim to cause a technological evolution, today more and more necessary given the limits of the market and the no longer economic sustainability of the current automatic systems.

CHAPTER 3

SENSORS

The usage of more sensor is a good practice in order to achieve a high level of reliability in a measurement process. The problem of objects identification is a topic that could find in the adoption of this practice a solution.

3.1 Introduction

OBJECT RECOGNITION is a well known and studied topic: given a set of data, for example an image or any data source, the objectives are the identification of the presence of a target object and, if found, evaluate its position inside a chosen/defined reference system.

Provide a solution to such problems means to drastically increase the perceptive capabilities of any automatic machine, enabling the execution of more complex tasks, with a more advanced, and *smart*, interaction with the environment. That is however not simple to be achieved, and in most cases the solution depends both on the requirements of the application both on the operating conditions in which the robot will work.

Humans solve similar tasks daily, with little effort, mostly thanks to the experience acquired during their life and the innate capably in adaptation to the different situations. Methods and techniques that try to mimic such behavior represent the actual cutting edge research, with the wish of developing the so called *Artificial Intelligence*.

Given the objective of this work and the specificity of the task, the current research is focused on the development of a functional structure rather than a general method or framework. The AGVs need indeed to identify a single *class* of objects: *pallets*. The most common in Europe is the Euro Pallet, fig. 3.2(b), model taken as reference. The techniques and algorithms developed are however general for that class of object and usable with minimum modification with the other pallet formats.

As underlined in the state of the art, chapter 2, there are many proposed solutions to the identification of pallets for the autonomous picking, but none of them was able to produce a real technological transfer from the Research level to the industrial world.

The objective of the current research is then the development of a *functional* solution able to provide reliable results and a technology that can used directly on the field. That is achieved by employing multiple sensors and dedicated identification routines. The independence of the algorithms is indeed a key feature to increase the reliability of the process: less information are shared between the elaboration processes, more reliable is the conclusive comparison, *fusion*, of the results.

The sensors used are a camera, a 2D laser scanner, a 3D TOF camera, all industrial qualified devices. The objectives for each identification process are:

- to *identify* the presence of pallets
- to *localize* them

The target information is the position of the pallet: the 3D vector $[x, y, \theta]$. These are indeed the information required to plan a trajectory that the AGV will follow to fork the pallet.

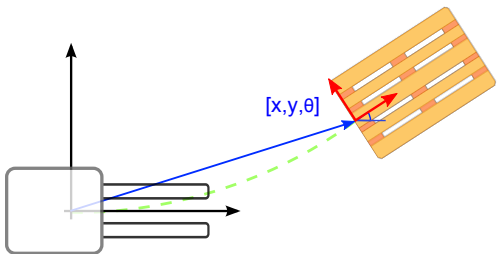
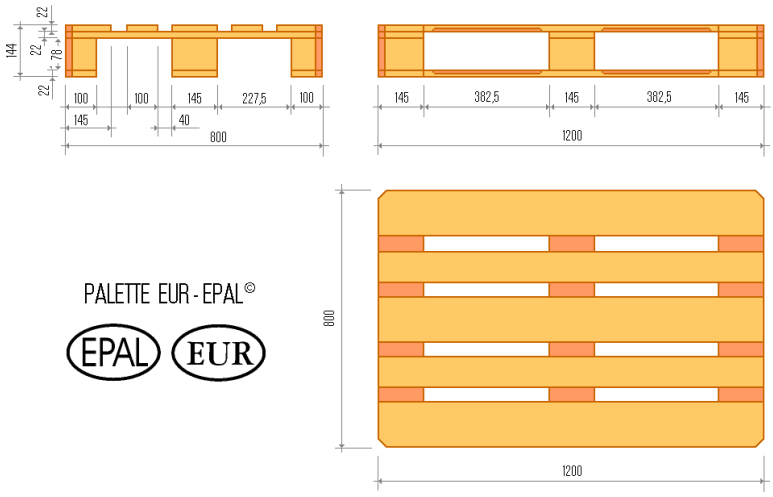


Figure 3.1: AGV and pallet

(a) Most common pallet formats



(b) Euro-Palllet

Figure 3.2: Pallet formats

3.2 Camera

The camera is the technological counterpart of the human eye. It is therefore clear why many computer vision works try to mimic it in order to improve the level of the artificial intelligence based on visual perception.

Among the various aspects related to the usage of a camera in an application, the most important one is the acquisition of information from a wide field of view, with data that characterize the objects not only in geometric terms, like the shape, but also many other information like textures, colors, light etc.

As drawback, however, such huge amount of information implies a high computational cost and therefore a slow processing.

A further element to consider is the loss of the 3D information about the volume of the objects, which are reduced to 2D projection over a plane, the image. This device offers therefore great opportunities for the development of an identification strategy based on the recognition of the *silhouette* of a objects.

3.2.1 AGILE

AGILE project (Baglivo et al. (2011)) represents the starting point for the current search. The algorithms developed in that context highlighted good potentials but also severe limitations. The complete description of the graphic elaboration is presented in the thesis of Biasi, Biasi (2010).

In summary, the pallet identification is achieved starting from the calculation of the gradients of the input image by applying the Canny 2D derivative operator. The result is a binary image in which all the pixels have value zero except the ones associated to the regions of the image in which there is a sharp transition between light and dark, a gradient. These pixels are used as seeds for a morphological dilatation, generating a map of distances between pixel coordinates of the image and points of the gradients, fig. 3.3. That map is used for a dual identification process based on two standard computer vision techniques from the class of algorithms called *Distance Transform*: Hausdorff distance, Huttenlocher et al. (1993), and Chamfer distance, Barrow et al. (1977). Both of them perform a template matching generating a voting map in which

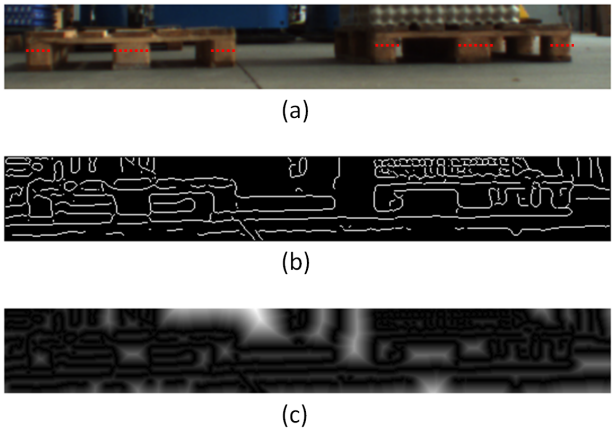


Figure 3.3: Biasi: Distance transform

the pixel coordinate with the highest value represent the solution, the identification. These maps are similar, but different characteristics can be highlighted: Hausdorff, fig. 3.4(a), evaluates the minimum distance between points and model, generating a map of sharp-cornered results (all the points close to the object in the image have the same score/value); Chamfer,fig. 3.4(b), performs instead a convolution, from that derive smoother distributions of scores (mean based operation).

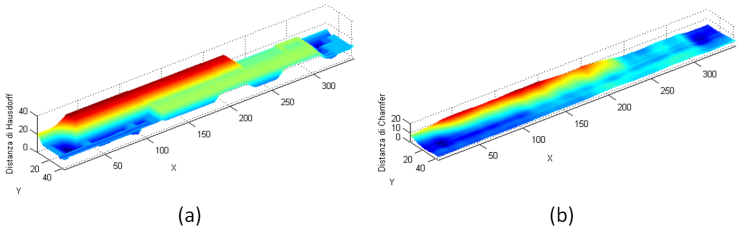


Figure 3.4: Biasi: Hausdorff vs Chamfer

The results are visible in fig. 3.5, blue Hausdorff and red Chamfer. The identifications are overlapped, meaning that the two algorithms evaluate approximately the same solution.

These two identifications are *compared* by calculating the relative displacement of

Figure 3.5: Biasi: Identifications

the central point of the pallet face: if the displacement is lower than a threshold the identification is considered valid, otherwise both are rejected.

The limitations of this identification process are two:

- the dependence from the laser: in order to correctly build the frontal model of the pallet it is necessary to know its distance from the camera, information provided by the laser
- the evaluation of the gradients is critic, this operation is not stable and robust because dependent on the environmental conditions: small light variations condition the performances of the algorithm

Such elements motivated the use of ROIs (Region Of Interest) inside the image to perform a local elaboration and strengthen the image processing. These ROIs are defined starting from the laser identification, establishing a connection of direct dependence between the two devices: the laser controls the camera.

In fig. 3.4 are presented the results of the process (in green the laser initialization).

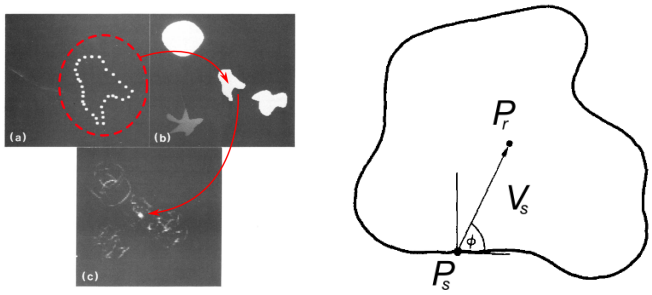
Figure 3.6: Biasi: Results

As the author writes, such image processing is more similar to a checking procedure of the laser information rather than an identification by itself.

3.2.2 Generalized Hough Transform

After AGILE, the initial objective was the development of a modified version of the previous algorithm, focusing on how make independent the camera from the initialization of the laser.

The first attempt was an algorithm derived from another standard computer vision technique: *the Generalised Hough Transform*, a voting algorithm that employs a more advanced model compared to the techniques described before, BALLARD (1981); Okada (2009).



(a) Voting logic (b) Referenfe point of the model

Figure 3.7: Ballard: Generalization of the Hough Transform

The model is built starting from the silhouette of the target object: a reference coordinate P_r must be defined in order to express the identification as a point in the image. A set of points $P_S(i)$ are sampled with a given spatial resolution from the silhouette of the object (according to the size and complexity of the model). From each $P_S(i)$ is calculated a vector $V_S(i)$ pointing to P_r. The model is constituted by a table in which the V_Ss are organized and associated to the directions of the segments from which each point $P_S(i)$ is sampled. Fig. 3.8 present the construction of the model for the pallet.

The Generalised Hough transform uses such data structure to run a more advance template matching.

Given the binary image of the gradients, the algorithms initially estimates for each

Figure 3.8: Hough: pallet model

pixel which is the direction of the gradient it belongs to. The next step is the identification of all those vector of the model that have a compatible directions with the ones found in the image, these are used to assign a vote (+1) to all the pixel coordinates (of a voting map initially set to zero) pointed by the vectors identified in that way. It must be highlighted that each direction can have even more than one compatible vector, a point of the gradient can then cause the voting of many pixel coordinates.

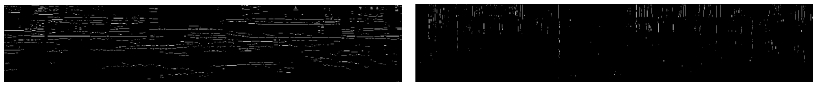
The result of the processing is a map in which are accumulated all the votes. The identification is given by determining the coordinates on the map with the highest score, that are associated to the reference point of the model and so the identification on the image.

The silhouette of the pallet and the operative condition allows a modification of the algorithm: the pallet is symmetric and usually lies on the floor; the directions of segments can therefore be subdivided and organized in just two main groups: vertical and horizontal.

To simplify the model and speed up the algorithm, only the gradients oriented along those directions, identifies using the Sobel operator, are used in the voting step, fig. 3.9. Fig. 3.8 shows the multiple vectors assigned to the vertical and horizontal directions, these build a cross, its center is the point with the theoretical highest accumulation rate.

The voting map resulting from the process is the one in fig. 3.10(a): it is not a continuous surface because the gradients are represented as a set of pixel coordinates (paths of maximum variation in the image), causing the accumulation of votes in isolated points. This simplify the identification of the maximum value in the map but also makes the manipulation of the results very complex.

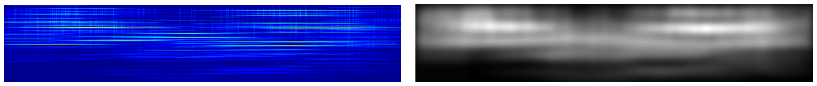
(a) Image (b) Canny operator



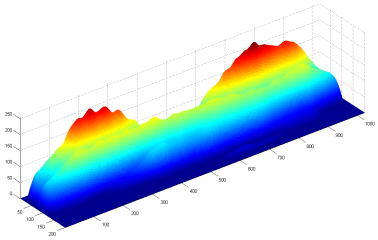
(c) Sobel horizontal gradient (d) Sobel vertical gradient

Figure 3.9: Hough: oriented gradients

The map is therefore convolved with a Gaussian filter in order to smooth the surface, fig. 3.10(b)(c).



(a) Score map (b) Score map: smoothed



(c) 3D distribution of the scores

Figure 3.10: Hough: voting map

In fig. 3.11 the map is superposed to the original image.

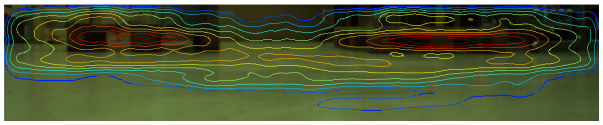


Figure 3.11: Input image vs Hough scores

The regions with the highest score are colored in red. As can be seen these are centered on the central block of the two pallets visible in the image. The result is convincing, but is still suffer of the limitation: the dependence between model and distance of the pallet.

The algorithm can be modified in order to use an parametric model based on an hypothetical pose of the pallet, morphing the silhouette by adopting a scale factor and a distortion. To identify the pallet without a priori information it is necessary first to run a brute force test over a set of possible hypothetical position in space in front of the camera, second to compare all the voting maps obtained and identify the highest score.

Such configuration is extremely computational expensive, but, more important, a key element for a reliable identification process is still missing: *finding the coordinate with the highest score does not imply the identification of the object*, it means only that in the image there is an area which produces more votes than others. The elaboration misses of a valued/threshold to use as *absolute reference* useful to assert that the identification achieved is due precisely to a pallet and not from the differential responses of the image.

3.2.3 HOG

The research moved towards more advanced computer vision techniques, from the fields of Pattern Recognition and Machine Learning.

Object recognition is one of the modern challenges for the computer vision. It studies the problem of identifying and localizing objects from specific categories, such as people or cars, inside static images. This is a difficult problem because objects in such categories can vary greatly in appearance, variations arise not only from environmental conditions, like illumination and viewpoint, but also due to nonrigid deformations and intraclass variability in shape. For example, people wear different clothes and take a variety of poses, while cars come in various shapes and colors.

The state of the art presents many papers, with various solutions and methods, usually very different among them. One algorithm however recurs more: the so called HOG.

HOG, Histogram of Oriented Gradient, is an acronym that embodies different techniques and works, all based on the identification of the objects using a model built from a map of points and a set of associated gradient directions, fig. 3.12. The HOG can be intended as an evolution of the Generalized Hough Transform.

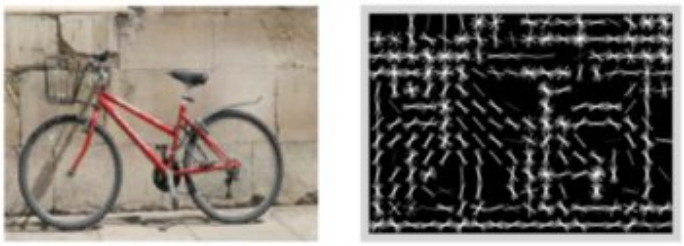


Figure 3.12: Example of HOG model

Among the different authors that proposed a solution based on HOG, Felzenszwalb and his work Felzenszwalb et al. (2010) is taken as reference. The identification process proposed achieved state of the art results on the PASCAL VOC benchmarks and the INRIA Person data set, winning the *Lifetime Achievement* Prize in 2010.

The works is also very interesting for many reasons: first it presents an exhaustive

analysis on the problem of object recognition, second the paper details very well the assumptions and choices taken for the development, lastly the author provides the entire algorithm in the form of Matlab/Mex code usable for testing. This recognition process is used in the current development as initial step for a custom pallet identificator.

As Felzenszwalb explains, his object detection system represents highly variable objects using mixtures of multi-scale deformable part models. These models are trained using a discriminative procedure that only requires bounding boxes for the objects in a set of images.

The approach builds on the *pictorial structures framework*, which represent objects by a collection of parts arranged in a deformable configuration. Each part captures local *appearance properties* of the object while the deformable configuration is characterized by spring-like connections between certain pairs of parts. Deformable part models, such as pictorial structures, provide an elegant framework for object detection.

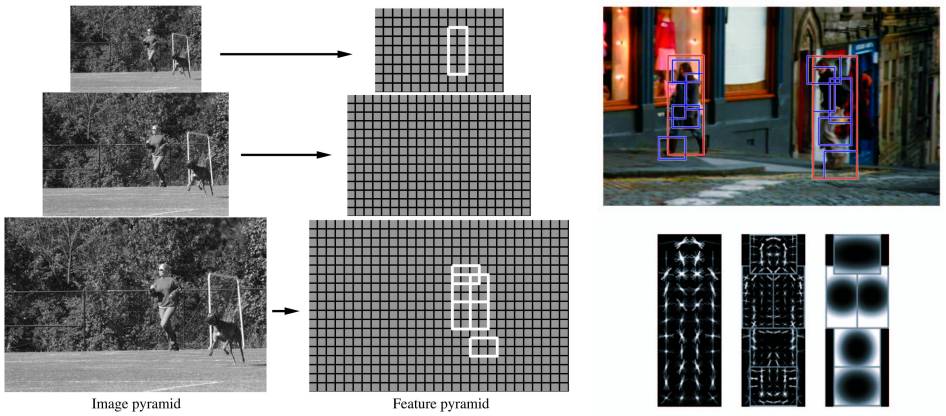
Improving the performances of the identification by enriched models is however very difficult. Simple models have historically outperformed sophisticated models in computer vision, speech recognition, machine translation, and information retrieval. One reason is that rich models often suffer from difficulties in training.

For object detection, rigid templates and bag-of-features models can be easily trained using discriminative methods such as Support Vector Machines (SVM). Richer models are more difficult to train, in particular because they often make use of latent information. Consider the problem of training a part-based model from images labeled only with bounding boxes around the objects of interest. Since the part locations are not labeled, they must be treated as latent (hidden) variables during training.

The Dalal-Triggs detector (Dalal and Triggs (2005)), which won the 2006 PASCAL object detection challenge, used a single filter on histogram of oriented gradients (HOG) features to represent an object category.

This detector uses a sliding window approach, where a filter is applied at all positions and scales of an image. The detector can be thought as a classifier which takes as input an image, a position within that image, and a scale. The classifier determines whether or not there is an instance of the target category at the given position and scale. Since the model is a simple filter, the score can be computed as $\beta \cdot \Phi(x)$, where β is the filter, x is an image with a specified position and scale, and $\Phi(x)$ is a feature vector. A major innovation of the Dalal-Triggs detector was the construction of particularly effective features.

Felzenszwalb enriches the Dalal-Triggs model using a *star-structured part-based model* defined by a *root* filter (analogous to the Dalal-Triggs filter) plus a set of *part filters* and *deformation models*. The score of one of this star models at a particular *position* and *scale* within an image is the score of the root filter at the given location plus the sum over parts of the maximum, over placements of that part, of the part filter score at its location minus a deformation cost measuring the deviation of the part from its ideal location relative to the root. Both root and part filter scores are defined by the dot product between a filter (a set of weights) and a subwindow of a feature pyramid computed from the input image. Figure 3.13(b) shows a star model for the person category.



(a) Pyramidal structure used in the identification process (b) Trained model: person

Figure 3.13: Felzenszwalb: HOG structure

To train models using partially labeled data, it is used a latent variable formulation of SVM, Andrews et al. (2002), call by the authors *latent* SVM (LSVM). In a latent SVM, each example x is scored by a function of the following form:

$$f_\beta(x) = \max_{z \in Z(x)} \beta \cdot \Phi(x, z)$$

β is a vector of model parameters, z are latent values, and $\Phi(x, z)$ is a feature vector. In Felzenszwalb's star models, β is the concatenation of the root filter, the part filters, and deformation cost weights, z is a specification of the object configuration, and $\Phi(x, z)$ is a concatenation of subwindows from a feature pyramid and part deformation features. The model of an object category is built with a mixture of star models. The score of a mixture model at a particular position and scale is the maximum over components of the score of that component model at the given location. In this case the latent information, z, specifies a component label and a *configuration* for that component. Fig, 3.14 shows a mixture model for the bicycle category.

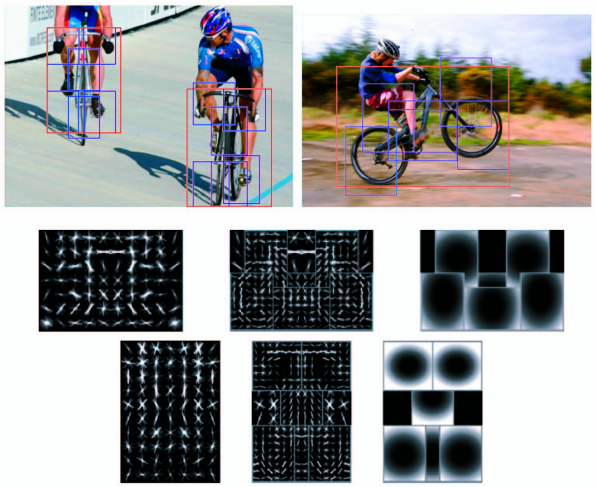


Figure 3.14: Felzenszwalb: bicycle

To obtain high performance using discriminative training it is important to use large training sets. In the case of object detection, the training problem is highly unbalanced because there is vastly more background than objects. This motivates a process of searching through the background data to find a relatively small number of potential false positives, or hard negative examples. In the paper are analyzed data-mining al-

gorithms for SVM and LSVM training, proving that data-mining methods can be made to converge to the optimal model defined in terms of the entire training set.

In the paper it is also shown how the locations of parts in an object hypothesis can be used to predict a *bounding box* for the object. This is done by training a model specific predictor using least-squares regression. Felzenszwalb uses bounding boxes derived from root filter locations. The system uses the complete configuration of an object hypothesis,z, to predict a bounding box for the object. This is implemented using functions that map a feature vector $g(z)$, to the upper left, (x_1, y_1), and lower right, (x_2, y_2), corners of the bounding box. For a model with n parts, $g(z)$ is a $2n + 3$ dimensional vector containing the width of the root filter in image pixels (this provides scale information) and the location of the upper left corner of each filter in the image. After training a model, the output of the detector is used on each instance to learn four linear functions for predicting x_1, y_1, x_2, y_2 from $g(z)$. This is done via linear least-squares regression, independently for each component of a mixture model. Fig. 3.15 illustrates an example of bounding prediction for a car detection.

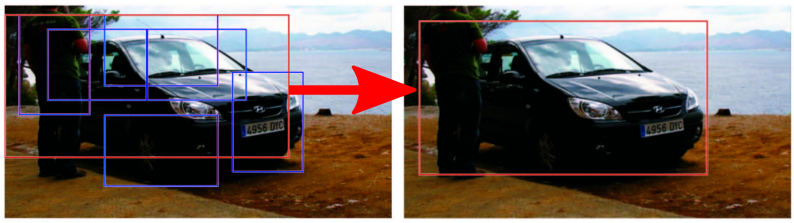


Figure 3.15: Felzenszwalb: identification

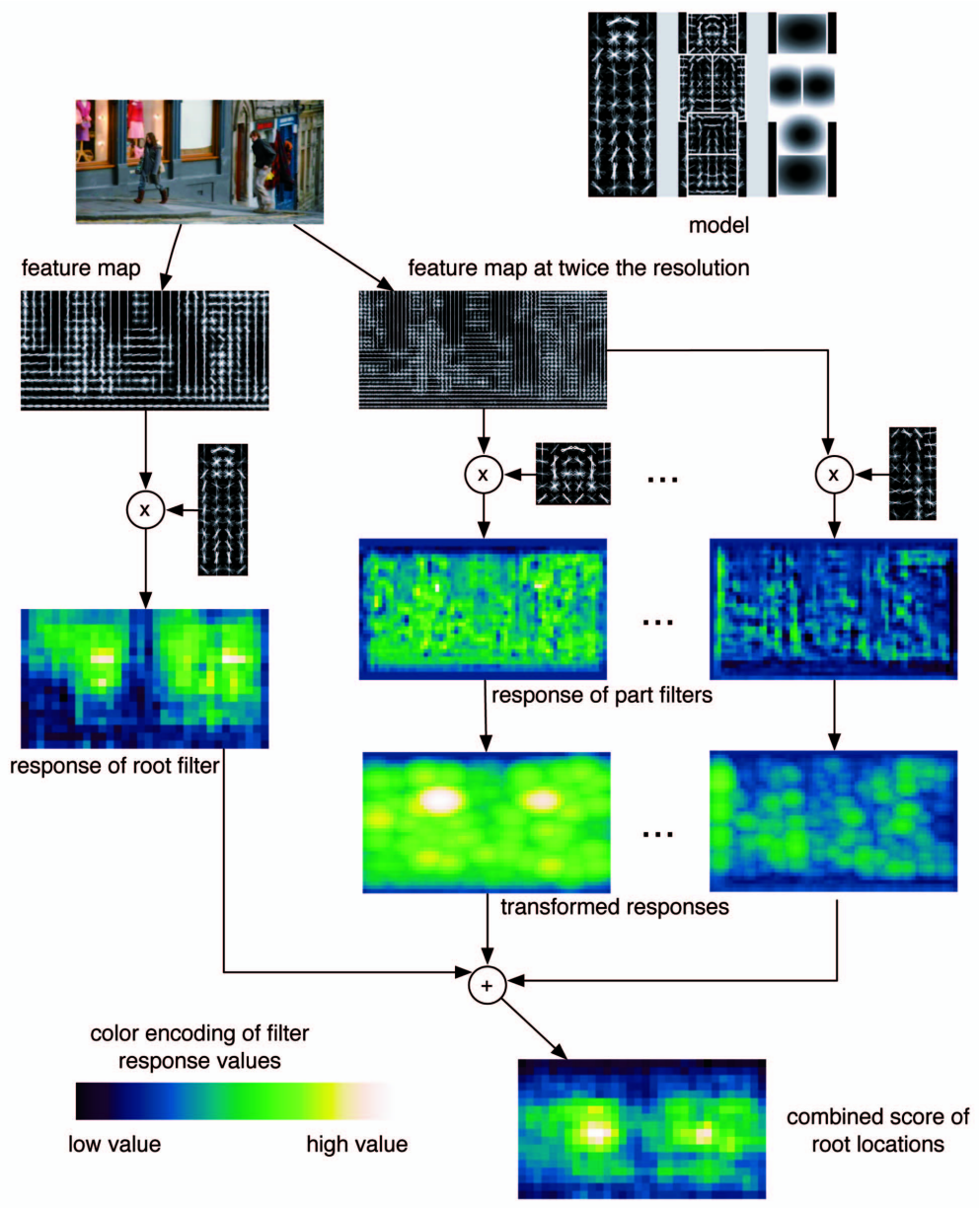


Figure 3.16: Felzenszwalb: elaboration logic

Usually multiple overlapping detections are obtained for each instance of an object. A greedy procedure eliminates the repeated detections via nonmaximum suppression. After applying the bounding box prediction method described above, set of detections *D* for a particular object category in an image is obtained. Each detection is defined by a bounding box and a score. The detections in *D* are sorted by score, greedily selecting the highest scoring ones while skipping detections with bounding boxes that are at least 50 percent covered by a bounding box of a previously selected detection.

The code provided with the paper was tested in order to verify its efficiency. A training set of images of pallets was collected and used for the training, fig. 3.17.

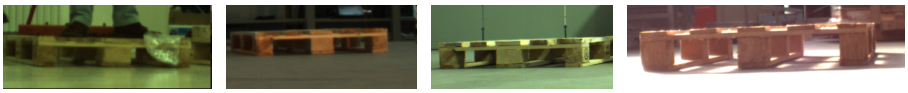


Figure 3.17: HOG: example of training set for the pallet

The resulting model is visible in fig. 3.18;

Figure 3.18: HOG: model of the pallet

In fig. 3.19 are presented the results obtained from generic images acquired in a warehouse without any control of the environment.

Figure 3.19: HOG: pallet identifications

The identification process proved great potentialities. Most cases in which the pallet is actually seen by the camera the image processing provides a correct identification.

The algorithm was also tested in limit situations: generic objects are placed deliberately in order to create a silhouette comparable to a pallet, fig. 3.20. In these cases no identifications occur.

Figure 3.20: HOG: false samples

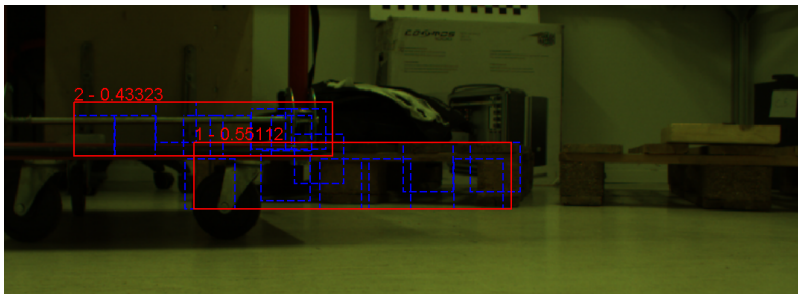


Figure 3.21: HOG: wrong identification

Despite these performances, some limitations and issues were highlighted during the tests.

First of all the elaboration time: the process is computational expensive, at least 10 seconds for elaborating of an image of resolution 1024x768 pixels.

Second, the system does not always return correct identifications, fig. 3.21.

The definition of a bounding box is not a suitable result for the purposes of the application: the reference must be position of the central socket of the pallet, the most meaningful geometric feature usable for the forking.

The same for the score assigned to the identification: such value indicates how strong was the response of the convolution over the image. An absolute method to determine if the solution is reliable is still missing.

3.2.3.1 Pallet Identificator

The code of Felzenszwalb was modified in order to:

- speed up the process
- increase the reliability of the identification
- avoid false positive identifications
- define a more advanced parameter of *quality*(not just a score)

The speed of the process is related to two elements: code efficiency and number of pixel to be processed.

About the code, the first step was the complete translation of the algorithms from Matlab to C/C++, developing a dedicated library. The increment in speed is around 1 order of magnitude.

About the number of pixel two different action were taken: that value depends indeed both on the size of the image both on its resolution.

A ROI is defined in the image in order to crop the image, fig. 3.22(a)(b). Defined the position of the camera on board the vehicle, the silhouette of the pallets (placed on the floor) is always inside a portion on the image. The more the camera is parallel to the floor the thinner is this region. The full width of the image is kept, the height and position of the ROI are instead tuned manually once placed the camera on board the AGV/robot. Elaborate only a slice of the image drastically increases the speed the identification. The use of the entire image it is instead useful only in order to identify pallets lifted from the ground (at the cost of time).

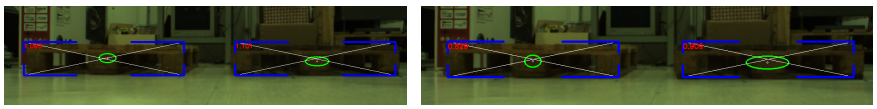
The resolution of an image is another important parameter to be consider in order to speed up the process, despite it decreases the number of pixel when lowered, the influence on the process is way more deep. This parameter is directly related to the information content: the reduction of the number of pixel through the a rescaling is comparable to the decimation or the sub-sampling (depending on the method used for the rescaling) of a set of sampled points in a signal. The modality by which the

(a) Input image (b) Cropped region (c) Scaled region

Figure 3.22: Image preprocessing

decimation occurs causes a different response from the analysis of the new data; if the decimation is too high there is even the loss of the information content from of the signal. The algorithm used for the rescaling therefore influences the results. For the image, the one that produces the best performances is the cubic interpolation (sub-sampling).

Must be highlighted that there is anyway a loss of information with the rescaling, element that can be verified using the same image with different scales: the lower is the resolution, the faster is the process, but also less identifications occur. Fig 3.23 show 2 different results achieved with different scaling factors, more details about the ellipses are provided later in this chapter (bigger is the ellipses less accurate is the identification). The elaboration with scale 1.0 (original size) is performed in 0.9 seconds, the one at scale 0.6 in 0.3 seconds.



(a) Scale 1.0 (b) Scale 0.6

Figure 3.23: Influence of the resolution

A lower limit in the rescaling exists: the minimum scale factor usable with an image of 1024x250 pixel (dimension of the ROI) is 0.5. Lower values cause the failure of the process.

The algorithm builds a pyramidal structure from the input image, this structure is convolved with the model in order to evaluate the scores from which derive the iden-

tifications: if the starting resolution is too low ($0.5 \times [1024 \times 250] \rightarrow [612 \times 125]$*pixel*, 100 pixel is the lower limit) the pyramidal rescaling produces levels that are too small to achieve any identification, making in practice useless the subsequent elaboration. Despite this limitation the scale can be changed freely in the admissible interval $0.5 \rightarrow 1.0$.

Given this lower limit, the choice of this parameter is related to the operative conditions.

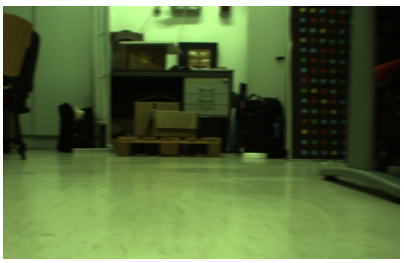
In order to identify pallets far from the camera, 3 to 5 meters away, a value close to 1.0 should be used, keeping in practice the full resolution of the input image. A far object produces a silhouette on the image made of few pixels, it is then fundamental to not lose or approximate any information at the cost of the speed, otherwise any rescale could cause the loss of the identification.

On the opposite, with pallets close to the camera, 2-3 meters far, a scale factor of 0.5-0.6 can be used without losing performances: the silhouette is well defined on the image and the loss of some pixel does not cause the failure of the process.

The modification that however radically improved the performances of the algorithm is a routine of post-processing based on the elaboration of the pyramidal maps, output of the convolution between model and the pyramidal representation of the input image.

As seen for the Generalized Hough transform, the voting procedures produces the accumulation of scores in limited regions of the image. The same happens in the HOG. In fig. 3.24 are represented some of the different pyramid levels obtained after the convolution between the input image and the model, (left$\rightarrow$ pyramid head; right$\rightarrow$ pyramid base)

The areas of the maps corresponding to pallet position on the image have an higher score. That is due to the training of the model, which only enhances specific graphical features assigning a more votes in a very localized spot. This element is used to develop a more advanced characterization of the *quality* of the identification, increasing the rejection of the false positive cases.



(a) Input image

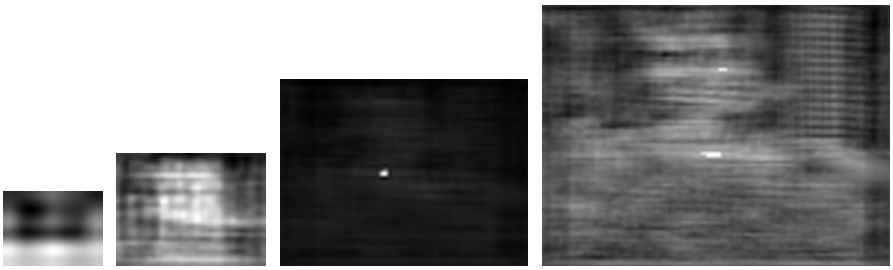


Figure 3.24: Pyramidal levels from standard HOG

If an object is clearly visible inside the image, it is logical to think that it will be correctly identified inside different levels of the pyramid. This is a form of redundancy, and so an element that ensures reliability in defining the correctness of an identification. The multi level check, however, is not sufficient to perform a direct evaluation of the quality.

The pyramidal levels are indeed not comparable with each other due to the difference in their sizes; the first step is then the resize of all the levels in order to create an homogeneous data structure with the same sizes of the input image. The main advantages are:

- the direct comparison between the pyramidal rescaled distributions and original image
- the resized maps keep their information content, making it compatible between the different levels

A low resolution map, from the top of the pyramidal structure, once rescaled to the maximum size, base of the pyramid, presents a smoother distribution of the votes

compared to the ones coming from an higher resolution level. That derives from the upscaling operation, in which the definition of the interpolation options influences the morphology of the map. This effect correctly represents the fact that an identification occurred at low resolution can't be more *accurate* than one achieved at higher resolution, the scores are less localized in a single spot.

Figure 3.25: Homogenized pyramidal levels

What it is required at this point of the elaboration is the evaluation of a descriptor or a data structure that includes in a single object all the information shared among the rescaled pyramidal maps.

Such element was developed starting from a set of concepts typical of the probabilistic field. Given n probabilistic distributions, their multiplication is the most direct and efficient method to combine them. The different trends are in this way merged, exalting the most frequent ones while reducing the less influentials (noise, random elements, etc.). The same strategy is used with the rescaled pyramidal maps.

The first step is the transformation of each map from a scoring distribution to a *pseudo-probabilistic* one: the process of identification is not probabilistic, so the transformation represent only a redefinition of the values of the map with the aim of approximate a probabilistic distribution.

The values of each map M_L, where L is the level of the pyramid, are shifted in order to have minimum value equal to 0 (M_{L_0}); then it is applied a normalization: the sum of all values of the map must be equal to 1, M_{L_N}.

$$\begin{aligned} M_{L_0} &= M_L - min(M_L) \\ M_{L_N} &= \frac{M_{L_0}}{\sum(M_{L_0})} \end{aligned} \tag{3.1}$$

The multiplication of normalized maps M_{L_0} requires a reference map as initialization: the *confusion map*. The confusion map represents the equal probabilistic distribution over the defined working space, and it is built assigning the same score to all the bins of the map: $1.0/num(bins)$.

The process of multiplication is then performed iteratively: the reference map M_R, initially equal to the confusion map, is multiplied with the first of the normalized maps. The resulting map is again normalized and substituted to M_R. The process is repeated for all the input maps.

$$M_R = \frac{M_R \cdot M_{L_N}(i)}{\sum(M_R \cdot M_{L_N}(i))} \tag{3.2}$$

The result is a map that presents a very localized spot, fig. 3.26.

(a) Input image (b) Map Combination (c) Superposition of the map over the image

Figure 3.26: Input image vs resulting combined map

The described method is functional only if only one pallet is seen by the camera. If not so, the multiplication fails, producing a map with a homogeneous distribution: no spots can be detected.

That comes from the distributions associated to multiple objects: the maps of the scores don't present the same distribution for both the objects inside a given map, neither share the same identifications among the levels (the accumulation of scores), fig. 3.27. This causes the multiplication of maps that have *null* (low score) regions for a pallet and an accumulation for the other, lowering or nullifying the resulting distribution: it is indeed sufficient a single null to lose most of the accumulation and so the identification of the object.

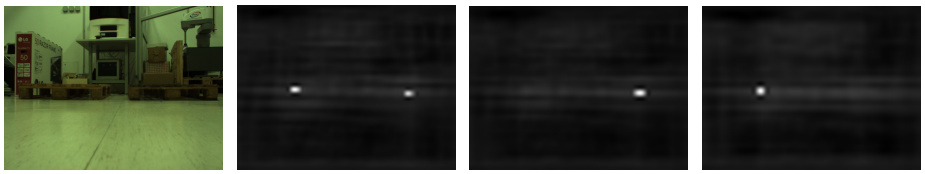


Figure 3.27: Multiple pallets

This issue is solved using a sub window of the resized (normalized) maps.

Each object identified by the Felzenszwalb routine is associated to a bounding box, used as reference for the identification, fig 3.15, and in which level of the pyramid such identification occurred. That information is used to trim the maps in those levels in which the *ith* object is identified, creating a structure of sub regions associated to each identification (the operation is simplified thanks to the homogeneous sizes of image and maps), fig. 3.28. The multiplication process it is then performed not on the entire

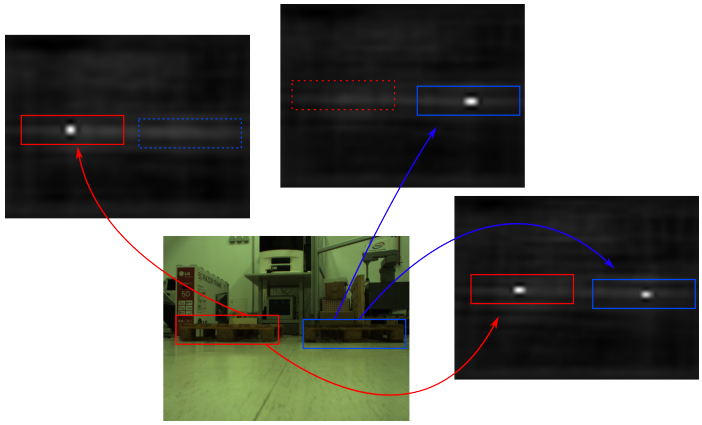
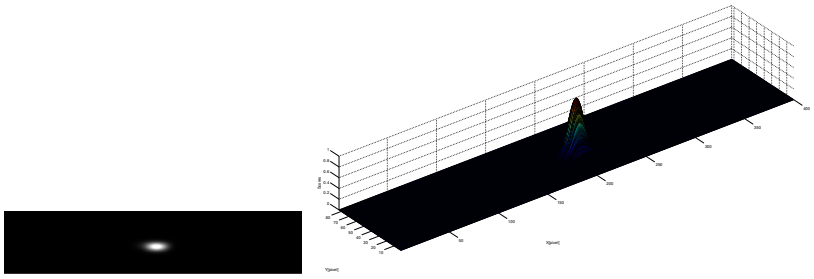


Figure 3.28: Sub-maps combination

map, but on the specific region, avoiding the occurrence of null cases. I this way the two identifications do not influence each other, obtaining as result two localized spots.

Given a identification, or more, and the resulting map of scores, it is then necessary develop a descriptor that analyzes such distributions and provides an evaluation of the *quality* of the identification.

The experimental evidences highlighted that the distribution obtained from the combination of the maps is similar to a 2D Gaussian aligned with the principal direction of the image.



(a) Local distruibution (b) 3D representation

Figure 3.29: Gaussianity of the distributions of scores

Such geometry can be modeled evaluating the coordinates of the peak and a diagonal covariance matrix (x and y are considered not correlated). It must be underlined that the actual choice is specific for the application, due mostly to the silhouette of the pallet: symmetrical and regular. With a more complex object a more advanced analysis could be required, characterizing in the most appropriate way the resulting distributions.

From such distribution two vectors V are built with the maxima values of the map along the vertical and horizontal directions, fig. 3.30.

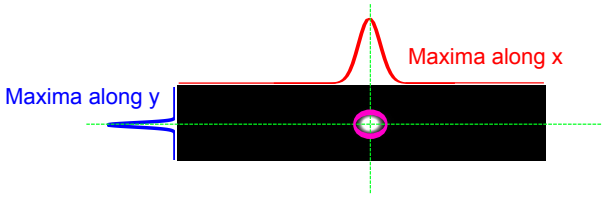


Figure 3.30: Distribution decomposition

From each vector of maxima scores V, the position the peak V_{peak} and the standard deviation V_{std} of the curve are evaluated with a single read of the data.

$$V_{peak} = \sum \left((idx + 1) \frac{V[idx] - min(V)}{\sum(V) - numel(V) \cdot min(V)} \right)$$
$$V_{std} = \sqrt{\sum \left((idx + 1 - V_{peak})^2 \frac{V[idx] - min(V)}{\sum(V) - numel(V) \cdot min(V)} \right)} \tag{3.3}$$

The combination of the results along the two direction represent the approximation of the 2D Gaussian distribution.

An identification now includes: a bounding box, a reference point associated to the peak of the map, a covariance matrix. From that structure a new representation is developed.

A covariance matrix can be represented as an ellipse in space, which is a more intuitive and meaningful way to represent a distribution: wider is the ellipse greater is the covariance and so uncertainty of the data. Each identification is then represented on the image as a bounding box and an ellipses centered on the reference point, peak of the scores, fig. 3.31.

Such representation was fundamental for the understanding of the occurrence of wrong identifications. They are due to the position of the peak and/or the size of the distribution inside the boundary box.

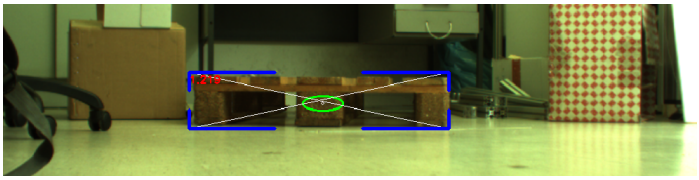


Figure 3.31: Identification: boundary box and ellipse

About the position of the maximum, fig. 3.32 shows how the wrong identification is associated to a distribution of scores shifted from the center of the box.

Such result is caused by a not correct reorganization of the parts of the model: the distribution are correctly generated by a pallet, which is identified (the one on the right), associating however a wrong boundary box.

Figure 3.32: Identification failure: peak position

Also the dimensions of the ellipse, and so the covariance matrix, plays an important role in the rejection of the false positive.

In fig 3.33 is reported a sample in which it is possible to see the difference in dimension between a correct and a false identification. The dimensions of the ellipse is directly correlated to the distribution of the scores in each bounding box. In the false positives the distribution is not a 2D Gaussian surface, it is instead more similar to a mixture of Gaussian distributions. The standard deviation, evaluated by means of single Gaussian model, results in those cases very high, with an associated wide ellipse.

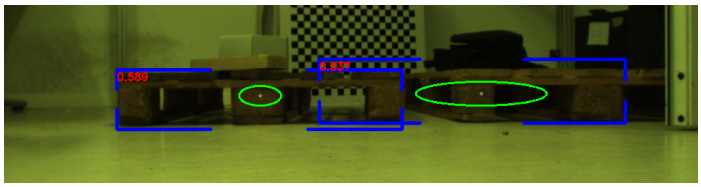


Figure 3.33: Identification failure: covariance dimensions

Such effect is related to the probabilistic approximation of the distributions: all the bins of the maps must sum to 1, if the distribution is not localized it means that all the bins share an higher mean value. This *homogeneous* distribution is then modeled as a Gaussian, with a consequent uncertain position of the peak and an high covariance. The choice of modeling the distribution as a single 2D Gaussian surface finds in this effect its strength, offering a further control over the results.

The final step of the post processing is the rejection of false positives cases. Two thresholds are defined about the position of the peak inside the boundary box and the dimensions of the covariance ellipse.

The values of these thresholds were tuned during the entire test phase of the algorithm, defining in the end the following values:

- **displacement from the center** : valid if less than 20% of the width/height of the bounding box
- **dimensions of the covariance ellipse** : valid if less than 50% of the width/height of the bounding box

It is important to underline that this structure is only finalized to check the identifications in order to avoid the false positive cases, one of the main requirement for the current application. Lower values are more preventive, rejecting even correct identifications; higher values are less, with the occurrence of false detections. The values proposed are the ones that denoted the best performances in terms of minimum rejection of correct cases and maximum rejection of false cases.

In fig. 3.34 shows rejection of the false positive case by means of the threshold verification: in blue are the bounding boxes of all the identifications, in green the covariance ellipses of only those identifications evaluated as valid.

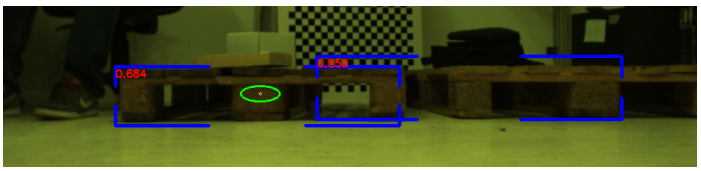
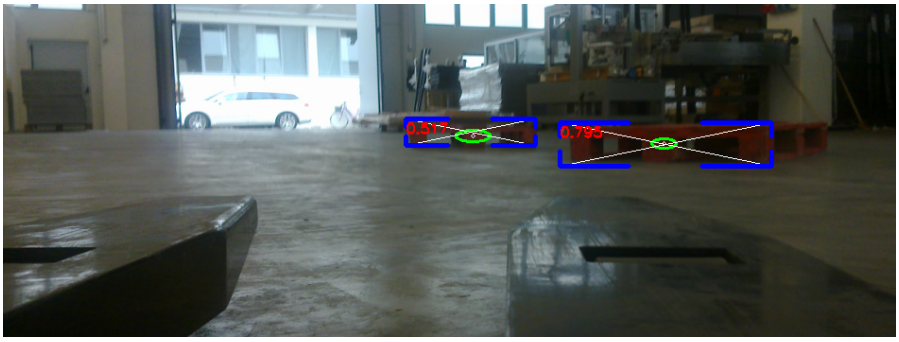


Figure 3.34: False positive rejection

The paper of Kanazawa and Kanatani (2001) rises an interesting question with its title: ***Do we really have to consider covariance matrices for image features***. Despite the topic analyzed is different form the current one, the answer to that question is the same provided in the end by the author: **yes**, the covariance is definitely a very useful information for image processing.

In fig. 3.35 are shown two different examples of successful identification of pallets with the described procedure.

(a) Image from a phone camera



(b) Image from AGV

Figure 3.35: Examples of positive Identifications of pallets

3.2.4 Results

The proposed algorithm proved good operative performances, both in terms of processing speed and reliability. Since the introduction of the discrimination based on covariance no false positives have yet occurred.

The software developed runs at 3Hz. The parameters of the process, the scaling factor and the position of the ROI, can be changed directly on line. In fig. 3.36 are presented some frames of a continuous process of identification. As can be seen in the frames 6-7-8, the identification is correctly influenced by the position of the pallet: it is close to the edge of the image, losing part of its silhouette, making such identification more *uncertain*. Interesting is also the capability of the software to identify a partial view of the pallet, frame 3, providing the bounding box as reference of the possible identification, but not the ellipse: the conditions on the position and size of the distribution of the scores are not fulfilled. The ellipses is represented only when the check on the distribution is surpassed, defining the identification as valid.

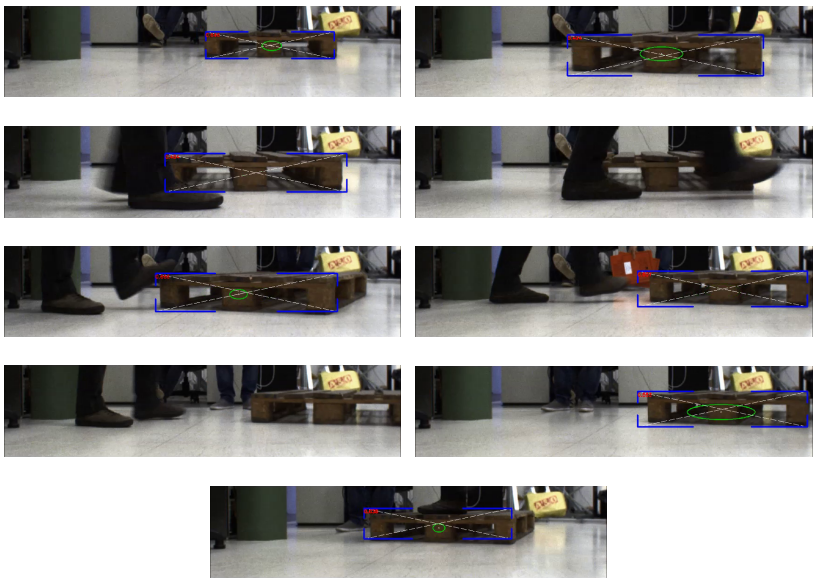
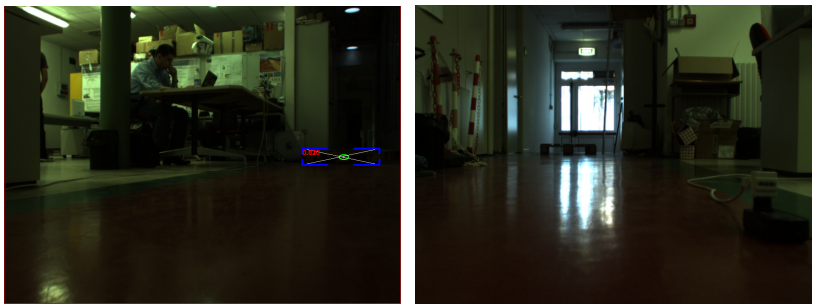


Figure 3.36: Continuous image identification

The system was tested in critical condition. In fig. 3.37 is shown how the identification of the pallet is successfully achieved even in low-light conditions.

Some limitations still remain. The main issue highlighted is the missed identification when a light source lights the pallet from behind, fig. 3.37(b). The resulting image presents dark foreground objects, shadowed, that causes the reversal of light and dark areas, dark feet and bright gaps, producing the failure of the identification.



(a) Low light conditions (b) Backlight

Figure 3.37: Testing cases

Despite this last element the overall performances of the process are convincing, allowing a continuous identification (no tracking) of the pallet in standard operative (not controlled) conditions.

3.3 Laser

The modern AGVs, due to safety reasons, must be equipped with sensors able of identify a person lying on the ground in front, or in general close to, the vehicle. That in order to not harm the people and avoid the collision with objects while the AGVs are moving inside the plants. These devices are safety 2D laser scanner.

The manufacturers of AGVs use those sensors merely for purposes of safety, not taking advantage of the rich source of information that they are capable to provide: the 2D outline of the environment, fig. 3.38.

This technology is now established at the industrial level and reached a grade of performance such that not exploiting the sensor for its totality appears to be a waste. It is sufficient to think of the fact that between the range of possible sensors offered, the cheaper ensure a scanning angular range of 250°, an angular resolution of 0.25° and a depth of field well above 20 meters (sensors dedicated to the measurements have the best performance, higher costs but can not be used for security purposes).

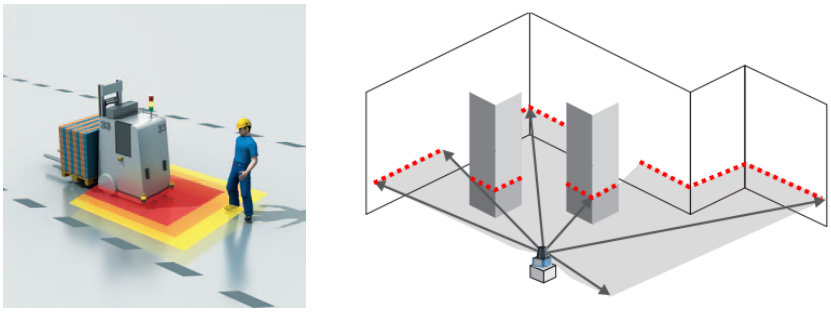


Figure 3.38: Safety laser scanners

This sensor should be employed in the identification of pallets for two reasons, one economical and one technical: the sensor is already on the AGV, the information content is sufficiently accurate in terms of Signal Noise Ration and resolution to ensure a proper identification, chapter 2. From that derives that only minimum modifications of the AGV are needed, adding no further Cartesian sensors on board and so without any additional costs. The only constrain to fulfill is the height of the sensor from the floor, lower as possible.

The identification process presented in this section represents the evolution of the work done by the research group in the AGILE project, Baglivo et al. (2008, 2009, 2011); Biasi (2010). The experimental results achieved in such context proved that the usage of the laser is both an efficient and effective method to identify a pallet, fig. 3.39. Efficient because the matching is not computationally expensive, effective because the identification of the pallet is correctly achieved with various operating and environmental conditions.

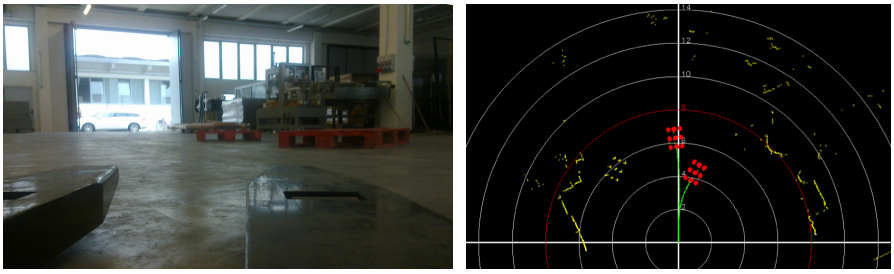


Figure 3.39: AGILE: laser identification

Minor bugs were highlighted during the test campaign of AGILE, but a more important limitation was noticed: the solution proposed lacks of any characterization of the uncertainty. For each identification just a *quality score* is computed, derived from the *percentage* of the pallet model matched with the points of the laser scansion. That value expresses only partially and in an approximately way how *accurate* the matching was. A more structured analysis is necessary.

As for the camera, once verified that the process of identification is accurate and robust, the development of a structure that monitors the uncertainty is useful to preventively filter the outputs of the process and reject most of the false positives. At the same time such information is a key element of the data fusion strategy performed as conclusive step of the overall identification process.

In order to fulfill to such requirements the laser part was totally revised and re-coded from scratch, developing a new identification method.

The development was conducted in collaboration with Mattia Tavernini, using some of the methods presented in his PhD Thesis, Tavernini (2013), a research on the topic of 2D laser scan matching for robotic self-localization.

The localization of a robot is achieved by matching a set of laser scansions acquired during an unknown motion inside an unknown environment (a simplified case is the one in which the map is instead known). The logic is to incrementally build the map of the environment by *matching* the laser scansions each other. This allows to calculate the robot motion and its absolute position inside that map. The solution and implementation are however not trivial since the construction of the map requires the knowledge of the actual pose of the robot, and at the same time the pose of the robot derives from the knowledge of the map: it is a problem of recursive optimization. The geometric parameters related to robot and scansions are evaluated incrementally, minimizing the relative displacement of the scansions evaluating a cost function based on a ICP-like routine, Iterative Closest Point. Depending on the methods, the choices of the developer, the operative condition and the environment, the method converges to a solution. In the state of the art literature can be found several works related to this topic, many different solutions and methods are proposed. That underlines how a robust and univocal solution is still missing and that the research is still ongoing.

Many of the techniques developed in such research field can be employed in the current application, these two tasks can be indeed considered similar. In order to identify and object it is necessary to recognize peculiar elements that distinguish it among the data, *features*. Once identified those, it is possible to segment the scansion and isolate candidate regions in which to perform a local matching, aligning the theoretical shape of the object, a model, with the selected laser points. The same operations involved in the alignment of laser scansions.

3.3.1 Algorithm

The matching process can be subdivided in 4 main steps:

- initialization
- candidates evaluations
- local optimization
- results organization

The steps must be accomplished in sequence and each is dependent from the previous.

The most important element to underline is that the research of the pallet is double: the first is global and aimed to find possible candidates solutions among the entire laser scansion, the second is local and aimed to check if the input candidate solutions is valid, refining the estimation of the pose $[x, y, \theta]$ of the pallet.

The algorithm start with the initialization of the data.

The input is a laser scansion expressed in polar coordinate $[\rho, \theta]$: ρs are the distances measured by the laser beam, θs are the angular values associated to the ρs. If more that one scansion is used the algorithm uses the mean of the ρs vectors. The points far more than 7 meters are removed.

The reliability of any identification process is always related to the amount of the data involved in the matching, the more far is the pallet less are the laser beams that intercept it, the less accurate and robust will be the elaboration. Given the maximum admissible distance of 5 meters, due to the minimum density of the data over the object (1 point every 2.2 centimeters at 5 meters), the theoretical maximum distance in which a laser points could be potentially related to a pallet is 5 + 1.2m (1.2 meters is the longitudinal size of the pallet), 0.8 meters of tolerance is left.

The initial step of the matching is a routine that searches potential candidate solutions by identifying the central block of the pallet inside the laser scansion. The

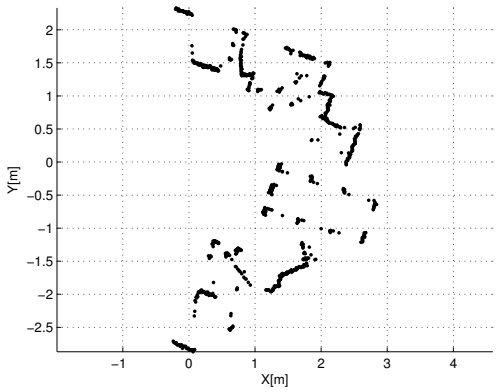


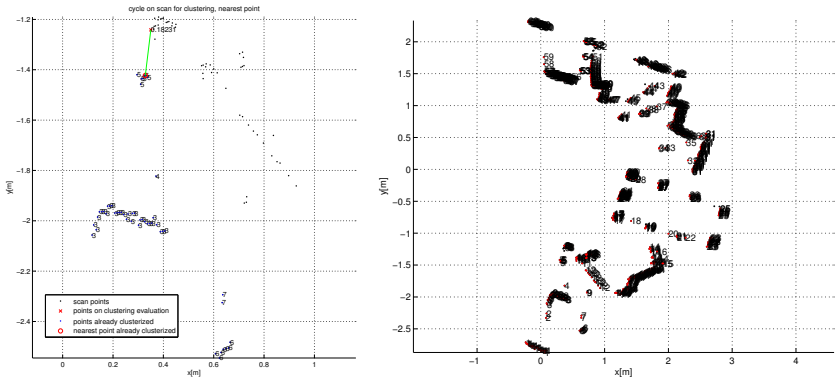
Figure 3.40: Laser input data

identification is achieved by recognizing segments that have a compatible length with such geometrical element, 14.5cm, plus the dimensional tolerances due to the manufacturing process: $\pm$1cm along the longitudinal direction, $\pm$2cm along the transversal one.

The data are segmented by generating cluster of near points. Assigned the first laser point to the first cluster, the clusterization is performed by calculating the displacement between the last point added to the current cluster and the closest not clustered point: if the distance is lower than a defined threshold the point is added to the cluster, if not, the cluster is closed and a new one is created. The maximum admissible distance between two consecutive points must be lower than 11.4cm, equal to half of the gap between the feet of the pallet. That threshold allows to group the data in clusters sufficiently continuous to isolate entire pallet feet but at the same time to create different clusters with the three segments of the frontal face, fig. 3.41.

Once the clusters are defined, it must be computed if these can be modeled as segment or not. The fitting of the segments is achieved with a custom routine that uses the probabilistic method Least Median Square, LMEDS, by Rousseeuw (1984). The algorithm was developed by Tavernini in his thesis.

In the context of autonomous localization, a good strategy is the synthesis of the environment by means of geometrical primitives and features, like the segments. The main benefit of that method, compared to a least square fitting, is a more robust



(a) Distace evaluation (b) Clusters

Figure 3.41: Clusterization

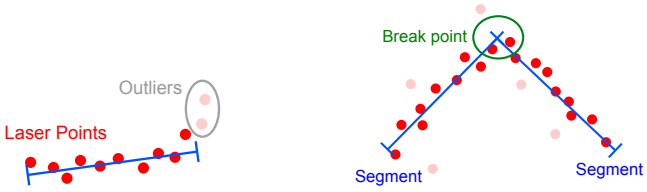
rejection of the outliers, fig. 3.42(a), a key feature in order to achieve a robust and reliable segmentation with the data involved in this process: small clusters, even shorter than 10cm (4-5 points), in which even the distributions made of theoretically aligned data can vary of $\pm$2cm due to the laser accuracy.

A further important element implemented by Tavernini is the robust evaluation of the end points of the segments from a cluster. Two sides of an edge could be merged into a single cluster due to the continuity of the points; the definition of a dedicated feature, that analyzes the alignment of the points, evaluates the best coordinates in which the break of the cluster produces two consecutive, linear, segments, fig. 3.42(b).

Such method is used in the current identification process to model each cluster as a segment.

The identified segments must then surpass a double check:

- the length of the segment must be compatible with the dimension of the central block of the pallet (and tolerances)
- the number of points of the cluster must be at least equal to 3, the minimum condition to verify the linearity of given the noise level of the data (from the laser technical specifications)



(a) Segment from cluster

(b) Break of the cluster

Figure 3.42: Segmentation

In fig. 3.43 are reported the possible candidate segments resulting from the described process.

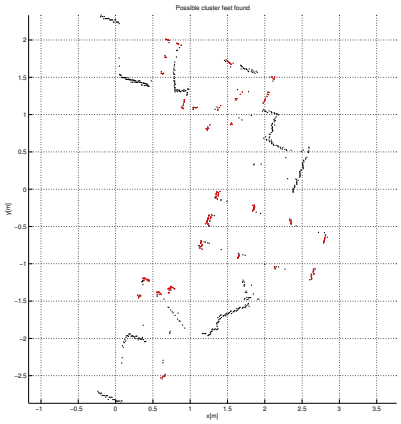


Figure 3.43: Candidates segments

The central block is larger than the others on the sides, so the previous operation rejects all the segments that do not belong to the shape of the pallet. The walls and the bigger objects are therefore removed and not considered in the successive steps of the elaboration.

What distinguishes a pallet from other objects is its shape in terms of the geometric relations between the segments. A brute force processing analyzes all the possible combinations of 3 segments calculating the relative displacement and alignment.

A combination of segments is considered valid and potentially belonging to the pallet only if two conditions are fulfilled:

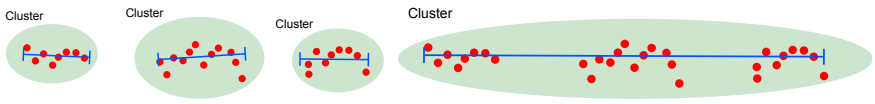
- the distance between the central points of the segments must be compatible to the distance of the frontal pallet feet and tolerances, $35cm \pm 2cm$
- the segments must be aligned, the vectors between the central points of the segments must have an angular displacement lower than 8°, this values is derived from the noise of the laser points and the dimensional tolerances of the pallet

The segments that don't fulfill these conditions are rejected.

The final step of the global localization is then the identification of the candidate poses of the pallets.

The evaluation of the pose from the central of the three clusters is an operation that does not ensure a reliable and repeatable measurement: the noise on the data, given the limited dimension of segment, strongly influences the calculation of the pose, especially for the attitude.

The structure of the pallet allows however to take advantage of the leverage effect associated to the planarity of the frontal face: the three segments are merged in a single cluster, which is used again as input in the same routine that computes the segments. Such operation evaluates a new longer segment, that better fits the frontal face of the pallet, achieving a more stable and accurate identification of the pose, less influenced by the noise superposed to each of the three sub-segments, fig. 3.44.



(a) Input segments (b) Frontal segment

Figure 3.44: Candidate refinement

Once evaluated such candidate positions a last check is performed.

The geometry of the pallet includes 3 repetitions of the *frontal* block along the longitudinal direction. That can lead to the identification of more candidates belonging to the same pallet. In order to avoid such occurrence the area in front to each candidate is analyzed checking if it is free from other identifications. If not so, the identifications are compared: if these are aligned and compatible with the geometry of the pallet the ones that do not belong to the frontal face are rejected, fig. 3.45.

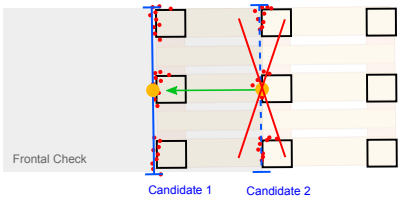


Figure 3.45: Frontal check

The result of this part is a matrix of 3D vectors $V = [x, y, \theta]$ of the candidate poses. The input scansion is then segmented according to such matrix: from each candidate solution V_i a ROI of dimension 1.0×1.4 meters (wider than the pallet) is defined within the scansion, fig. 3.46.

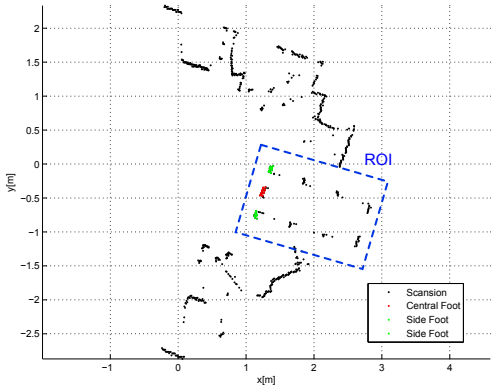


Figure 3.46: Central segment candidates and ROI

The successive step is a local minimization between the points inside the ROI and the model of the pallet. In this part the modeling of the uncertainty represented the

key factor for a robust process of matching.

The input structure, a set of laser points, segments and an initial candidate pose, is extended including the uncertainty associated to the laser points in the form of covariance matrices $C_{LP_{\rho\theta}}$.

From the datasheet of the device are retrieved the information regarding the *error* associated to the laser beam:

err_ρ **= 0.02cm** : the laser beam provides measurements with and accuracy of $\pm$2cm

err_θ **= 0.05**$^\circ$: the laser beam has a conical shape so there is an uncertainty associated to the angular displacement of the measure

These two elements are considered not correlated: the accuracy of the linear measurement is connected to the electronics of the device (the evaluation of the time of flight of the laser beam by means of a phase shift), the angular uncertainty instead to the technology of the laser beam. The mutual influence can be considered negligible. The covariance matrix is then expressed as

$$C_{LP_{\rho\theta}} = \begin{pmatrix} err_\rho^2 & 0 \\ 0 & err_\theta^2 \end{pmatrix} \tag{3.4}$$

$C_{LP_{\rho\theta}}$ is constant for the polar representation. It must be adapted to all the points of the scansion according to their position in the Cartesian space, the one used in the entire process. That is achieved by applying the formulation of covariance propagation

$$C^* = JCJ' \tag{3.5}$$

where J, Jacobian matrix, is the derivative of the transformation T that evaluates $P^* = T(P)$. In this case P is the polar laser point $P_i = [\rho_{P_i}, \theta_{P_i}]$, *transformed* into the Cartesian $P_i^* = [x_{P_i}, y_{P_i}]$.

$$T_{P_i} = \begin{pmatrix} \rho_{P_i} cos(\theta_{P_i}) \\ \rho_{P_i} sin(\theta_{P_i}) \end{pmatrix} \tag{3.6}$$

Applying the partial derivatives on ρ and θ follows

$$J_{P_i} = \begin{pmatrix} cos\left(\theta_{P_i}\right) & -\rho_{P_i} sin\left(\theta_{P_i}\right) \\ sin\left(\theta_{P_i}\right) & \rho_{P_i} cos\left(\theta_{P_i}\right) \end{pmatrix} \tag{3.7}$$

and then the covariance of each Cartesian point P_i^* of the laser scansion

$$C_{P_i}^* = J_{P_i} C_{LP_{\rho\theta}} J'_{P_i} \tag{3.8}$$

Once computed the covariance matrix for all the points inside the ROI, the minimization process starts positioning the model of the pallet according to the candidate pose.

As for all the minimization problems, in order to ensure the convergence of the minimization, the initial solution should be close to the final result. That helps the algorithm to avoid local minima associated to the cost function used. This condition is achieved thanks to the robust estimation of the candidates and the removal of most of the possible outliers that could influence the minimization.

Each point of the laser scansion is then associated to the closes segment of the pallet model. If the distance between a point and the closest segment is greater than 20cm the point is rejected. Such threshold was defined from the experimental results achieved with the clusterization. The candidate position, thanks to the final refinement based on the entire width of the pallet frontal face, proved to be stable, accurate and repeatable, with an error in attitude compared to the *ground truth* solution lower than 10°, value from which is derived the maximum admissible distance between points and model.

Given a laser point P_L, the first step of the matching evaluates projection of P_L over the closest segment of the model.

Because of the dimensional tolerances due to the manufacturing process, the model could be different form the real pallet acquired in the scansion. This is modeled by including in the model two covariance matrices associated to the end points of each segment, fig. 3.47(a). For this reason the position of the intersection point is *uncertain*, the evaluation of a point is not enough. Such effect in included in the matching by calculating the covariance matrix associated the intersection point.

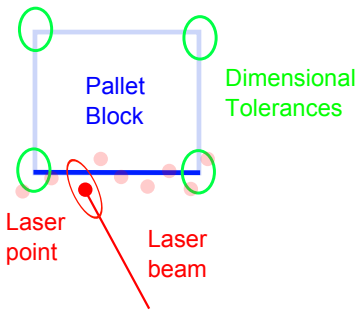


Figure 3.47: Uncertainty factors in the laser projection

The formulation is the same used for the evaluation of the covariance of the laser points, 3.5.

The input covariance in this case is more complex because dependent on three sources of uncertainty:

- uncertainty of the position of the laser point
- uncertainty of the position of the endpoints of the segment ($2\times$)

From that follows the covariance matrix

$$C = \begin{bmatrix} C_{Laser} & 0 & 0 \\ 0 & C_{Seg_{Init}} & 0 \\ 0 & 0 & C_{Seg_{end}} \end{bmatrix} \tag{3.9}$$

where C_{laser} is the 2×2 covariance matrix of the laser point ($C_{P(i)}$), $C_{Seg_{Init}}$ and $C_{Seg_{End}}$ are the 2×2 covariance matrices associated the ending points of the segment.

The point of intersection is calculated using the vectorial notation. From a couple of points, the generic line s that passes thought them can be written as

$$s(\eta) = \eta\vec{v} + q \tag{3.10}$$

where η is a scaling factor for the versor $\vec{v} = \Delta P/|\Delta P|$, q is one of the two points of the segment used as reference in space. Using the same notation both for *ith* laser point

(and the origin the laser beam, $[0;0]$), $L(\gamma)$,and for the associated closest segment, $S(\lambda)$, the intersection is calculated by solving

$$S(\lambda) = \lambda \begin{bmatrix} s_x \\ s_y \end{bmatrix} + \begin{bmatrix} s_{x_0} \\ s_{y_0} \end{bmatrix} = \gamma \begin{bmatrix} l_x \\ l_y \end{bmatrix} + \begin{bmatrix} l_{x_0} \\ l_{y_0} \end{bmatrix} = L(\gamma) \tag{3.11}$$

in which the all the parameters are known expect for λ and γ.

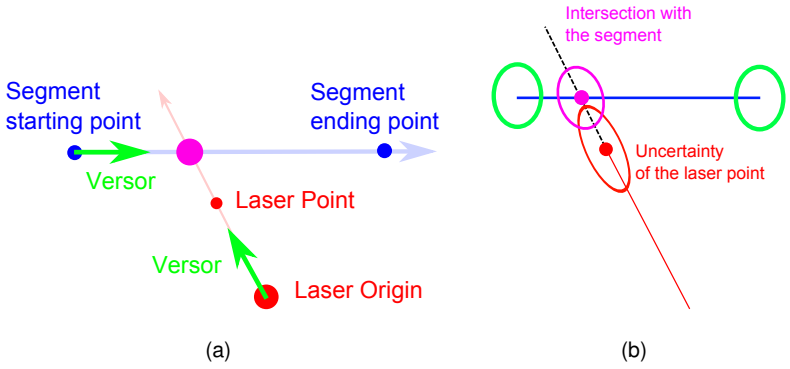


Figure 3.48: Uncertainty of the Intersection point

The solution is the following:

$$\begin{aligned} P_{intersection} &= S(\lambda^*) = \lambda^* \begin{bmatrix} s_x \\ s_y \end{bmatrix} + \begin{bmatrix} s_{x_0} \\ s_{y_0} \end{bmatrix} \\ \gamma &= \frac{\lambda s_y + s_{y_0} - l_{y_0}}{l_y} \\ \lambda^* &= \frac{l_y(l_{x_0} - s_{x_0}) + s_{y_0} - l_{y_0}}{l_y s_x - s_y} \end{aligned} \tag{3.12}$$

Substituting λ^* in 3.11 and differentiating along $[x_{laser}, y_{laser}, x_{s_0}, y_{s_0}, x_{s_1}, y_{s_1}]$ (coordinates of the laser point and the endpoints of the segment) the Jacobian matrix can be defined. Applying then 3.5 the covariance of the intersection point is computed.

The data structure includes 2 points and 2 covariance matrices, fig. 3.48(b). This is used to structure a cost function that calculates the distance between the laser points and their intersection with the model.

The function chosen is the Mahalanobis distance, Mahalanobis (1936).

$$D = \sqrt{(\Delta)' C^{-1} (\Delta)} \tag{3.13}$$

That is a probabilistic definition of distance, in which the displacement Δ of two distribution is computed and weighted on the uncertainty of the data, the covariance matrix C.

In the current configuration this notation finds its best usage: the couples of points that have an high relative displacement or an high associated uncertainty (or both) are the ones with the highest Mahalanobis distance (these points are not *compatible* from a more standard definition of the Mahalanobis distance). That can be used in an optimization process minimizing such values by correcting the position $[x, y, \theta]$ of the model inside the ROI. In this way the points that are more far from the model are the one that have the higher influence in the minimization, that makes the convergence of the matching faster by correcting in few iterations the reative angular orientation between data and model. This cost function is used in a non linear Levenberg Marquardt optimization.

The process of minimization is iterative (IPC like matching), a routine developed by Tavernini in order to minimize the relative displacement of successive laser scansions for the localization of robots, Tavernini (2013): the intersection of laser points and segments must be performed inside the cost function, updating the model and so the associations.

The result of the process is a set of *optimized* positions $[x, y, \theta]$, identifications of pallets.

Once localized the position of the pallet it is important to evaluate the uncertainty of the matching in order to understand how accurate and reliable such information is: the less accurate is the correspondence between the model and the points selected, the higher is the uncertainty.

Evaluate the uncertainty of a result obtained by an optimization process is however not a simple task; usually it is necessary to characterize the convergence of the

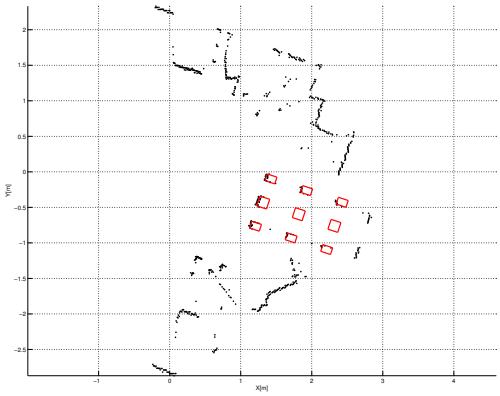


Figure 3.49: Laser identification

optimizer and the space associated to the cost function used. An approximate form is however presented in the work of Censi, Censi (2007).

The paper presents an analysis on the estimation of the uncertainty of a process of incremental localization and mapping in which a set of laser scansion must be aligned by means of an ICP: an algorithm that involves an optimization. The proposed solution evaluates the covariance matrix of the result by means of a simple matrix product.

Let $\widehat{x}$ be the result of an algorithm A minimizing an error function J, which depends on the measurements $\check{z} : \widehat{x} = A(z) = argmin_x J(\check{z}, x)$. The covariance of $\widehat{x}$ can be approximated as:

$$cov(x) \simeq \left(\frac{\partial^2 J}{\partial x^2}\right)^{-1} \left(\frac{\partial^2 J}{\partial z \partial x}\right) cov(z) \left(\frac{\partial^2 J}{\partial x^2}\right)^{T} \left(\frac{\partial^2 J}{\partial z \partial x}\right)^{-1} \tag{3.14}$$

where everything is computed at $\widehat{x}$,$\check{z}$.

As the authors explains, 3.14, is an extended formulation derived from the first order approximation of the covariance

$$cov(\widehat{x}) \simeq \left(\frac{\partial A}{\partial z}\right) cov(z) \left(\frac{\partial A}{\partial z}\right)^{T} \tag{3.15}$$

Since A is not in closed-form, it is not easy to compute $\partial A/\partial z$. However, $A(z)$ and z are bound by an implicit function. In fact $\widehat{x}$ is a stationary point of J; a necessary condition is that the gradient is null at $\widehat{x} : \partial J(\check{z}, \widehat{x})/\partial x = 0^T$. In this case, the implicit

function theorem gives an expression for $\frac{\partial A}{\partial z}$:

$$\frac{\partial A(z)}{\partial x^2} = -\left(\frac{\partial^2 J}{\partial x^2}\right)^{-1}\left(\frac{\partial^2 J}{\partial z \partial x}\right) \tag{3.16}$$

with $z = \check{z}$ and $x = A(\check{z})$.

In the current matching algorithm the terms $\left(\partial^2 J/\partial x^2\right)^{-1}$ and $\left(\partial^2 J/\partial z \partial x\right)$ are computed from the formulation of distance between the laser points and the their intersections with the segments: the first is a 3×3 matrix ($[x, y, \theta]$); the second is a $3 \times n$ matrix (n is the number of point matched with the model). The covariance matrix $cov(z)$ includes all the covariances of all the associations of segments and laser points, $n \times n$ matrix. The result is a 3×3 covariance matrix of the position $[x, y, \theta]$ of the identified pallet.

X[m]	Y[m]	θ[rad]
1.248	-0.408	-0.317

Table 3.1: Identification: result

		1E-04[m]
0.052	-0.0002	-0.021
-0.0002	0.0535	-0.0697
-0.021	-0.0697	0.2234

Table 3.2: Identification: covariance

The algorithm was ported to low level code, C/C++, boosting the speed of the process. Using a single scansion as input the elaboration takes less than 80 millisecond, achieving the real time processing with the laser SICK S3000 (sampling time of 120 milliseconds).

Increasing the number of scansions, in order to lower the noise on the data, the process becomes delayed depending on the amount of the input data: the best operative performances are achieved with 5 scansions, processing time around 660-690ms ($5 \times 120 \rightarrow 600ms$ of acquisition time).

The dedicated interface for the laser processing is visible in fig. 3.50.

The identification process involves additional elements like the fusion of the data, the calibration of the sensors, the path planning, obstacle avoidance etc.

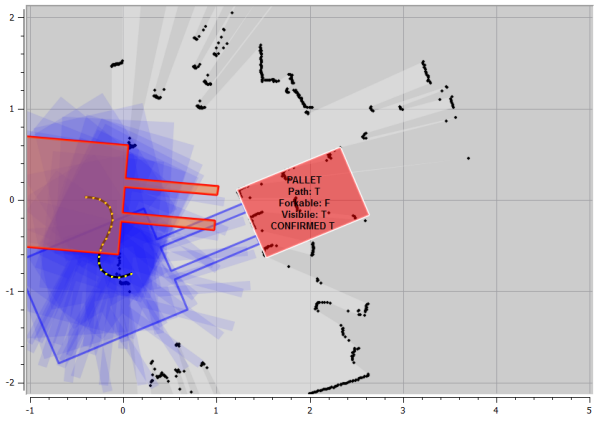


Figure 3.50: Laser GUI interface

Some of these elements are visible in the image as the parameters written over the pallet:

Pallet : The object identified is a pallet

Forkable : The pallet can be reached in safety, no impacts (path planning and obstacle avoidance)

Visible : The pallet is placed inside the field of view of the camera (laser-camera calibration)

Confirmed : Both laser and camera confirm the identification of the pallet (sensor fusion)

In this case the pallet is recognized, visible, confirmed but not reachable: the shape and size of the AGV would cause an impact with obstacles along the path (lower left corner of the image).

Comments

The algorithm is functional but some critic elements must be highlighted.

If there are more pallets placed close each other the clusterization could fail: parts from different pallets could indeed be merged in a single cluster. It must however pointed out that in real plants it is unlikely to find two pallets attached or closer than 10-20 centimeters: such configurations are dangerous because from an error of the AGV could derive the damage of more that one load, a minimum of spatial tolerance is required in order to ensure the safety of the picking. The common practice suggests a gap between consecutive pallets of at least 30 centimeters, optimal for the current method.

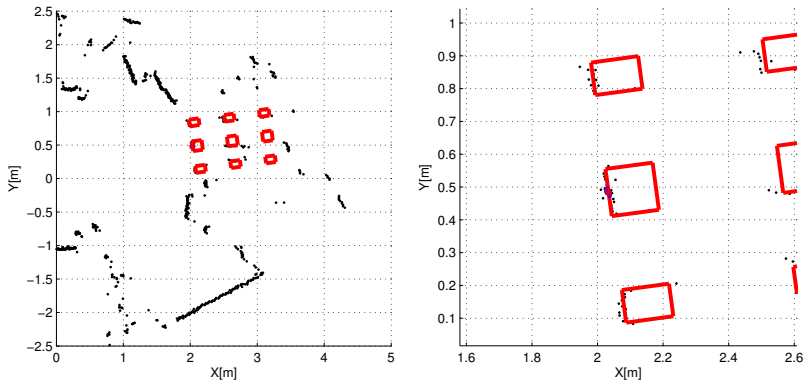
A further element is the height of the sensor form the floor. The laser plane is commonly placed in the middle part of the device, in the SICK S3000 is about 4cm far from the bottom side of the sensor. That makes difficult to place the sensor in a position in which the laser plane intersects the pallet at half of the height of the feet. The sensor can not be too low for safety reasons, not too high in order to not decrease the monitored area (changing the aim of the laser). For these reasons the position of the sensor must be carefully planned and tested on the field with every AGV model.

3.3.2 Results

In this and in the next paragraph are reported two different tests performed in order to characterize the metric performances of the algorithm, evaluating not only the accuracy of the results, but also the repeatability and the efficiency of the overall process of identification.

In fig. 3.51 are reported the results obtained from 900 successive (continuous) identifications achieved in stationary conditions: the position of pallet and sensor is fixed for the entire duration of the test, the pallet was placed in a generic unknown position in front of the sensor. With this configuration was tested the repeatability of the process, analyzing the variability of the identifications of the same object over time. The element that influences the results is the noise on the laser data, which causes variations in the clusterization phase, slightly changing the segments, but also the optimization, because of the different distribution of the point around the model.

	X[m]	Y[m]	θ[rad]
Mean	2.032	0.484	0.128
Sigma	0.004	0.002	0.005



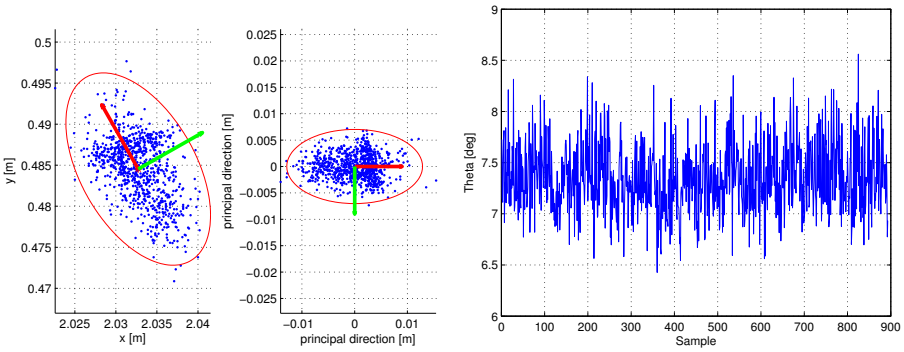
(a) Pallet (b) Distribution of the identifications

Figure 3.51: Repeatability of the laser identification

Fig. 3.52 presents a more detailed representation. In fig. 3.52(a) the red ellipse represent the 95% confidence interval of the distribution of the identifications, underling how the variability of the results remains within the value of $\pm$ 2 centimeters along the transversal direction, the one critical for the correctness of the picking of the pallet (chapter 6).

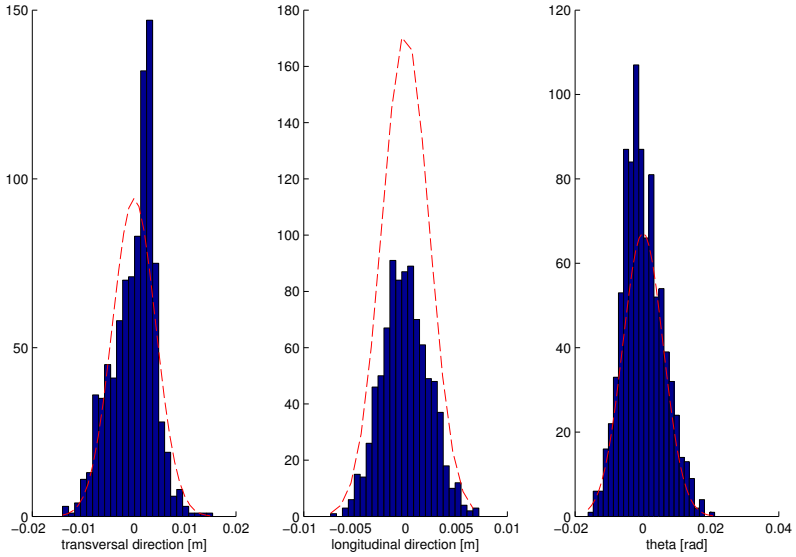
This value must not be confused with the covariance associated to each identification, which expresses the uncertainty of the matching between laser data and model. The covariance of the distribution must be therefore interpreted as the uncertainty of the process as influenced by the noise over time, while the covariance of the identification as uncertainty of the elaboration as influenced by the noise in the single sample. These two are however correlated: by increasing the number of input scansions, the noise of the data is lowered (mean of the input ρs), so as the dispersion of the results.

The histograms in fig. 3.52(c) show that the distributions of the result is Gaussian, due to the randomness of the noise over the laser samples.



(a) Position

(b) Attitude



(c) Histograms

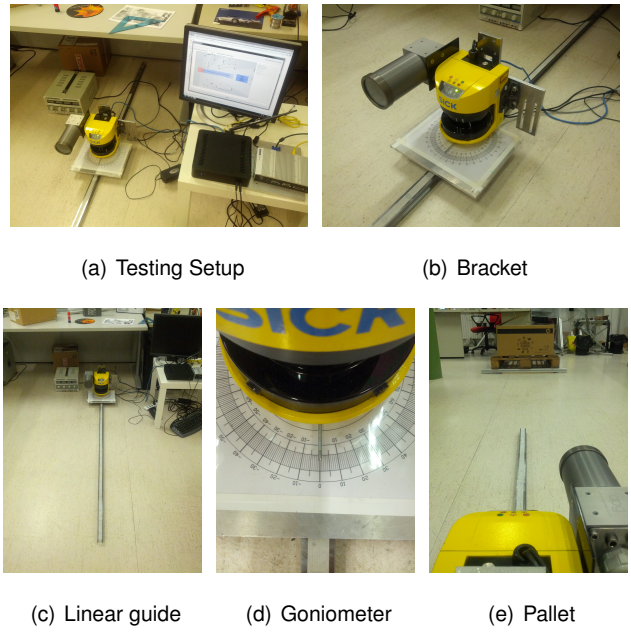
Figure 3.52: Distributions of the laser identifications

3.3.3 Metric analysis

Once completed the development of the algorithm, a test campaign was run in order to characterize its performances under different operative conditions.

Since the camera is capable to provide only the recognition of the pallet but not its localization, the Cartesian sensors are the ones designated to provide the measurement of the pose. That information is indeed the one used by the AGV to plan a trajectory and perform the picking. A measurement that is not accurate, reliable and repeatable is not suitable for the purposes of the application.

A dedicated testing setup was used: the laser is constrained on a bracket made of a goniometer placed on a linear guide, fig. 3.53.



(a) Testing Setup (b) Bracket

(c) Linear guide (d) Goniometer (e) Pallet

Figure 3.53: Metric analysis: setup

A pallet is placed in a fixed position, the linear guide is instead moved and oriented in three different configurations: **frontal, transversal and diagonal**. In each configuration **6 positions** are defined, each step with a relative displacement of 30

centimeter from the previous. For each position **5 angular orientations** are considered, steps of 15°: $[-30^\circ, -15^\circ, 0^\circ, +15^\circ, +30^\circ] + \eta$, where η is the angular offset required to approximately align the laser to the pallet longitudinal direction. For each configuration, spatial and angular, **3 different setup** of the process are used: single, 3 and 5 input laser scansions.

Each elaboration is considered complete once **100 identifications**, correct or not, are achieved.

The overall number of tested configurations is 270, with 27000 identifications.

During the tests the system is commanded by an external PC, connected thought TCP/IP, that acts as asynchronous controller, fig. 3.53(a). The number of requests sent to achieve the 100 identifications is recorded.

Such setup is useful to simulate the operative conditions in which the system will work on the field: an AGV that uses this identification system should stop close to a cargo area and send a request of identification asking if there are pallets around. It is therefore critical to understand how much *efficient* is the process of identification: because of the variability of the environmental condition or the influence of the noise on the data, it is possible that a pallet is actually inside the area, potentially identifiable, but however not recognized. *How should the AVG continue in this cases?*

For the current purposes the *efficiency* is defined as the capability of the elaboration to answer to a request with a correct identification. This parameter is useful to understand how the different influences, related to random conditions (noise) or not (environment, position of the pallet etc), limit the functionality of the device.

In order to group and to organically report the data acquired, the following structure is used for each test modality.

Setup of the test

Positions and orientation of the laser compared to the fixed position of the pallet.

Analysis of the Covariances

The data are presented grouping all the identifications achieved in each config-

uration in two modalities:

poses vs scansions used, merging the distributions associated to each angle

poses vs angular displacement, merging the distributions obtained using a different number of scansions

These two representations are aimed to highlight unexpected trends in the data sets.

In order to provide a more meaningful representation of the distributions the 95% confident interval Ellipses are used instead of the point clouds. The distribution are translated and rotated in order to align them to the reference system visible in fig 3.54.

The color used in the plots is the same for all the data sets: red 1 scansion, green 3, blue 5.

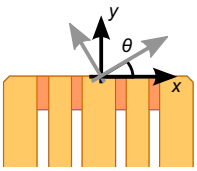


Figure 3.54: Metric analysis: reference coordinates

Analysis of the Results

The data acquired are organized in 3 tables, one for each setup (number of scansions). This parameter is indeed the only one that the user can modify in order to strengthen the measurement (at the cost of processing speed). The single scansion is the fastest configuration, but also the most vulnerable to the noise over the data. On the opposite 5 scansions have a better signal nose ratio (mean of the value), but the vehicle must be stopped for a longer time(600ms).

The structure of each table is the following:

Number of Requests The total number of requests from the external PC in order to obtain 100 identifications, correct or not.

Correct Identifications Number of identification that are consistent with the operating conditions.

An identification is considered *correct* if it differs from the mean of the distribution less than ± 5 cm and $\pm 3^{\circ}$. The experimental evidences show that the process of identification is stable, the distribution of the results has an uncertainty usually lover than the thresholds chosen. If the uncertainty of the distribution is higher than the thresholds, the identifications more far from the mean of the distribution are re-classified as non correct.

The wrong identifications are usually related to random noise that generates of *ghost* shapes in the scansion, these identifications are then usually located far away from the real position of the pallet, so clearly detectable. If more than once identification is generated from each process only the not correct one is kept (the worst case).

Missed Identifications Number of processes that produces no identifications.

Wrong Identifications Number of identification considered not correct.

Null Identifications During the test campaign a bug was identified. Some identifications were equal to $[x = 0, y = 0, \theta = 0]$, defined as *Null*. In order to keep unchanged the structure of the tests, such identifications are classified as missed: subtracted from the number of correct identification but not added to the not correct ones; a missed identification lowers the efficiency of the process. The bug was subsequently successfully solved.

Correct Identifications Rate Rate between the number of correct identifications and the reference number of 100 identifications.

Wrong Identifications Rate Rate between the number of wrong identifications and 100. This parameter is critical for the current application because from it derives the reliability of the entire system: for the purposes of the device a missed identification is preferable to a false one.

It must be however pointed out that during these tests no other sensors were used together with the laser, the aim was indeed to verify the performance of the laser elaboration. The multi-sensors fusion strategy, chapter 5, solves the problem of the false identifications.

Efficiency of the Process The ratio between the number of correct identification and the number of total requests. A value close to 100% means an high probability in having an immediate, correct, identification from a given request/query; lower values instead suggest that successive elaborations could be necessary in order to identify the pallet. Such parameter is also fundamental for the decision making process: if no identifications occur the AGV must chose if there is a pallet that is not identified or there is no pallet at all. A high efficiency means to have an higher probability in choosing the right option stating that if no identifications occur then there aren't pallets in front of the vehicle.

Identified Paths

In the final part are shown five figures, one for each nominal orientation, in which are reported the paths and covariance for each identification. The objective is to verify if there are critic poses/configurations in which the variability of the identifications produces relevant differences in the path planning or even its failure.

The entire data set is reported in Appendix A.

A more synthetic representation of the results is presented here merging the data in three overview tables.

All the identification ratios are satisfying, more than **99.00%**: the precess of identification can therefore be considered reliable, accurate and repeatable.

About the false positives, it must be underlined that the use of just one scansion turned out to be dangerous, with 4 wrong identifications. The same for the elaboration efficiency: with more than 10% of unsuccessful elaborations is difficult to define a strategy able to evaluate if the pallet is not recognized or rather missing.

On the opposite the use of 5 scansion is clearly the best choice, with optimal performances both for accuracy and for efficiency. A longer delay, around half second, seems to be a price worth paying for.

Table 3.3: 1 scansion, overall

Theoretical Id. Numb.	**9000**		
Requests	9913		
True Id. Numb.	8978	**Identification Ratio**	**99.76%**
False Id. Numb.	4	**False Detection Ratio**	**0.04%**
No Id.Numb.	913	**Unsuccessful Elaboration**	**10.14%**
Zero Id.Numb.	18		

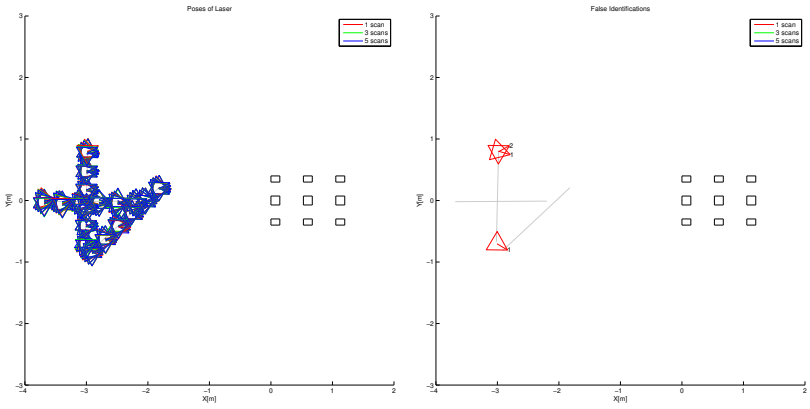
Table 3.4: 3 scansions, overall

Theoretical Id. Numb.	**9000**		
Requests	9222		
True Id. Numb.	8994	**Identification Ratio**	**99.93%**
False Id. Numb.	0	**False Detection Ratio**	**0.00%**
No Id.Numb.	222	**Unsuccessful Elaboration**	**2.47%**
Zero Id.Numb.	6		

Table 3.5: 5 scansions, overall

Theoretical Id. Numb.	**9000**		
Requests	9139		
True Id. Numb.	8996	**Identification Ratio**	**99.96%**
False Id. Numb.	0	**False Detection Ratio**	**0.00%**
No Id.Numb.	139	**Unsuccessful Elaboration**	**1.54%**
Zero Id.Numb.	4		

Fig. 3.55(a) presents all the poses of the laser in the tests. This scheme must be take as reference for the successive maps.



(a) Overall laser poses (b) False identifications

Figure 3.55: Laser poses

First map, fig. 3.55(b), the most important one, shows the poses in which wrong detections occurred. For each pose it is reported the number of identifications among the 100 acquired. The worst case is the one with two false positives. The identifications occur only when a single scansion is used and only when the relative orientation between laser and pallet is elevate. The noise of the data and the exposed geometry of the pallet mark such cases as complex and difficult to be correctly solved. It must also be underlined that a similar configuration is unlikely to happen: usually the AGVs are programmed to reach the pallet frontally, the same for a cargo area. The problem underlined is not manifested when more scansions are used.

Fig. 3.56 shows the poses of the laser and the associated success rate. Lower rates are due to the occurrence of the wrong identifications, the ones shown in fig. 3.55(b), but also the occurrence of the null cases, bug of the software. The second case is more frequent than the fist, meaning that a wrong identification is a rare event.

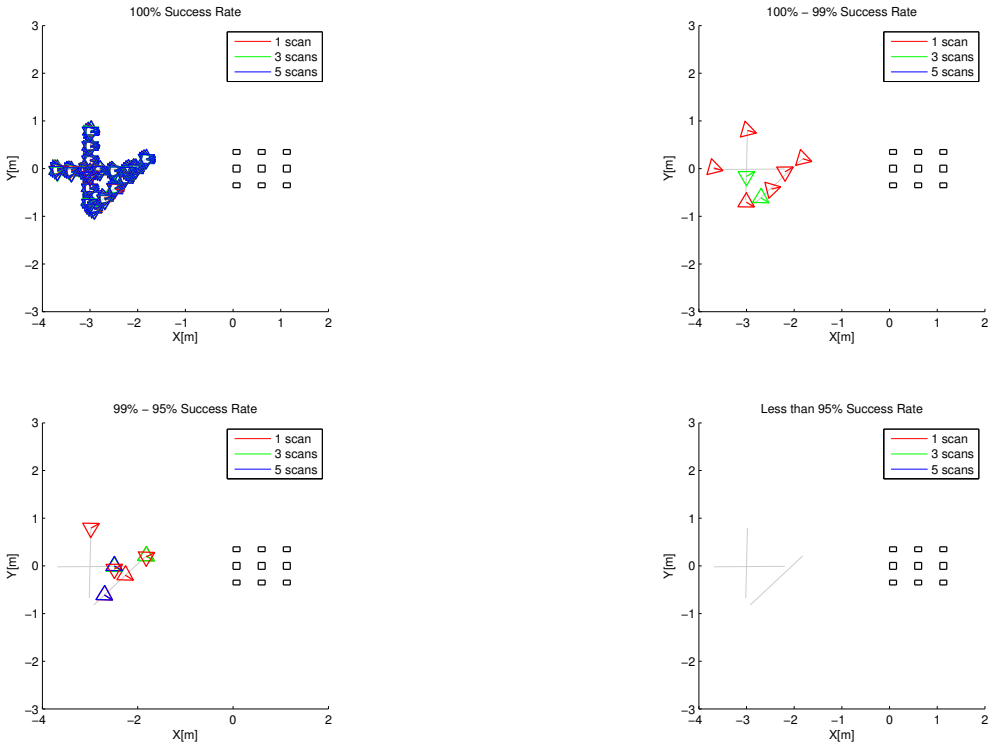


Figure 3.56: Success ratio of the identification

Lastly the efficiency. Fig. 3.57 presents the poses of the laser compared to 4 different level of efficiency. The configuration with the better efficiency is the one with 5 scansion. It is interesting to notice that there is a pose in which even with 5 scansions there is an situation of inefficiency: the laser is far from the pallet, with a relative angle close to the limit of $\pm 15^{\circ}$, the shape acquired and the noise superposed cause a set missed identifications.

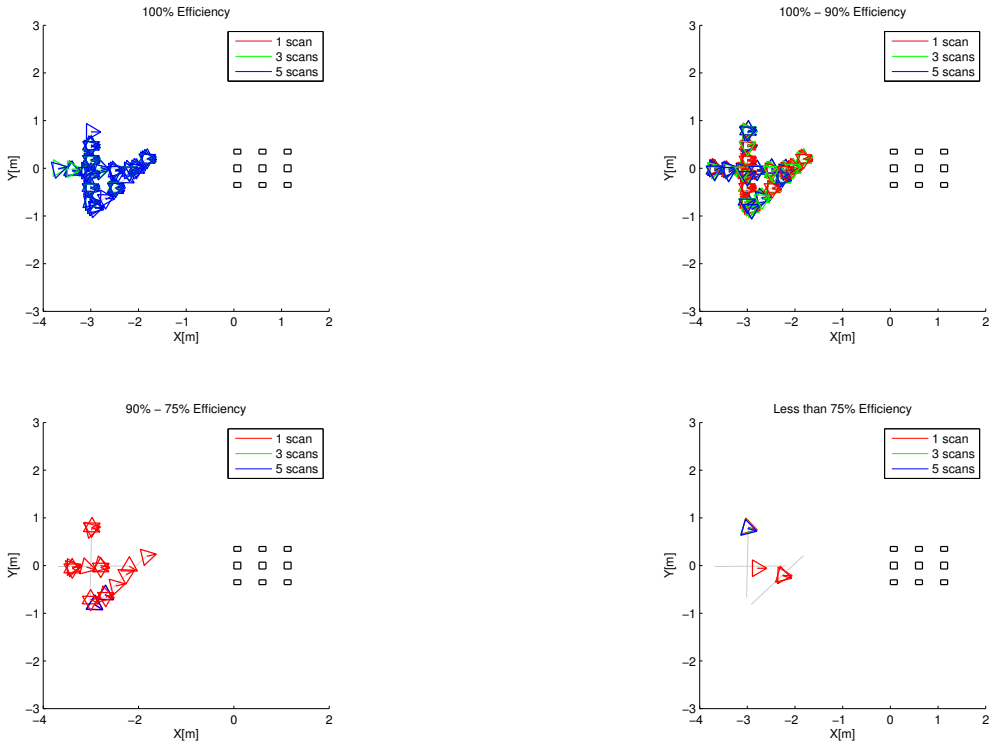


Figure 3.57: Efficiency of the identification

3.4 TOF

The second Cartesian sensor used is the TOF camera. This technology have spread during the last five years, thanks mostly to a more and more affordable price.

(a) MESA RS4000 (b) PMD CamCube 2.0 (c) FOTONIC E70P

Figure 3.58: TOF camera models

In literature there are many works focused on localization and navigation inside unknowns environments based on the use of such technology, Wang et al. (2009); Biswas and Veloso (2011), but there are also some papers related to the logistic field, examples are the works of Kleinert and Overmeyer (2012); Weichert et al. (2013). Even if the titles recall the actual application, in those works the TOF is used only for the autonomous palletization and depalletization phase of the production line.

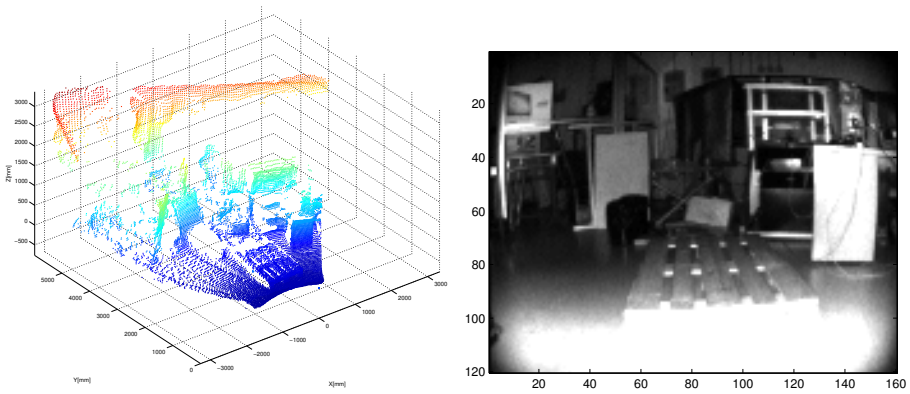
Some interesting references can instead be found on the sites of the manufactures of the TOF cameras and the datasheets of their products. These documents highlight and foresee the possible use of the 3D technology for *Forklifts*, fig. 3.59.

(a) MESA, datasheet (b) FOTONIC, video

Figure 3.59: TOF technology vs logistic

Compared to a laser scanner, a TOF camera offers a more advanced data structure, fig. 3.60:

- a 2.5D measurement of the environment (2.5 because it is projective, the objects are not entirely seen), fig. 3.60(a)
- a gray scale image, fig. 3.60(b)



(a) 3D depth map (b) Grayscale image

Figure 3.60: TOF data

The double nature of the output can be used to develop two different, parallel and independent identification processes, combining in a single device both the functionalists of the laser and the camera. A further advantage of the 3D information is the possible extension of the identification to pallets lifted from the ground, task not achievable with the planar laser scanner.

The device used is a FOTONIC E70P, fig. 3.58(c), a TOF camera of a resolution of 160x120 pixel, maximum depth of 7 meters, industrial class.

3.4.1 Depth

The object recognition from range data is a well studied topic, examples among the various papers that represent the state of the art are the ones of Mian et al. (2006); Schnabel et al. (2007); Biegelbauer et al. (2010). The structure of such solutions and algorithms can be substantially divided in 2, not necessarily consecutive, steps:

- the recognition of the shape of the target object inside an input 2.5D/3D point cloud
- the matching of the model of the object with the range data in order to estimate its position

Starting form the second, the matching between the model, the theoretical shape of the object, and the point cloud is usually achieved using an ICP. There are many versions of such algorithm, but all of them share the same elaboration logic: the minimization of the relative displacement between two input datasets. The minimization is carried out iterativelly as shown in the diagram 1. After each iteration the model should be closer to its counterpart in the dataset, asymptotically converging to a final position. The exit condition of the loop, together with the criteria on the selection of the inliers, are the elements that diversify the works and the results of the matching.

An exit condition usually adopted is the mean distance between points and model, lower is this value, more accurate (probably) is the correspondence between the point cloud and the model.

Such logic however works only if important conditions are met:

- the initial position of the model, the initial guess, must be close to the final solution (in order to ensure the convergence of the algorithm)
- a low noise level on the data
- the percentage of points from the point cloud that are associated to the target object must be elevate, few outliers
- the more dense is the point cloud, the more accurate is the matching

Data: *Model*: Sythetic Geomety, CAD file or Reference Point Cloud

Data: *Input*: Point Cloud

Result: H: transformation matrix that aligns the model with the point cloud

H = Initial Guess;

Score = Inf ;

while *!Matched or Iterations* $<$ *Iterations limit* **do**

  Update of the pose of the model $\rightarrow$ $Model_H = H \cdot Model$;

  Evaluation of the closest points of the input clout to $Model_H$;

  $Inlier = S(Model_H, Cloud) \rightarrow$ Selection criteria ;

  $Outlier = Cloud - Inlier$;

  Calculation of a scoring factor based on the inliers associated to the model:

  $Score = F(Inlier)$;

  if $|Score_{now} - Score_{previous}| >$ *Convergence threshold/criteria* **then**

    Update H from the Inlier ;

  else

    Matched ;

  end

end

Algorithm 1: 3D ICP matching logic

The *recognition* is instead usually achieved *after* the ICP: if the algorithm does not converge, it means that the matching conditions are not fulfilled, so the object is considered not recognized. In most cases the recognition is achieved by a trial and error procedure in which different objects are tested in the matching process, keeping the ones that obtain the best results in terms of mean distance between points and the identified position of the model.

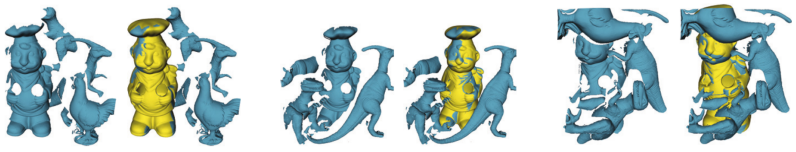
These conditions strongly influences the usability of such algorithms. The initial guess is an information that not always is known.

A further drawback is related to the amount of data: more dense is the point cloud to be processed, slower is the processing.

Some papers describes methods to overcome such limitations using descriptors to strengthen the matching, Lo and Siebert (2009); Pratikakis et al. (2010); Papazov and Burschka (2011). These identify and use local features in order to work with big point clouds, in which the target object is placed inside a generic unstructured environment. For the purpose of the current work, the ones with the higher relevance are those in which is presented a global converging algorithm. Papazov et al. (2012) is the one that better matches with the current application.

The topic is the autonomous robotic grasping of known objects, a task similar to the picking of a pallet: a robot that has to grasp an object must initially recognize and localize it and then plan a movement to achieve the grasping, fig. 3.61(a). From the

(a) Robot used for the grawsping



(a) 62.3% object occlusion (b) 70.4% object occlusion (c) 86.2% object occlusion

(b) Object identification in clutterd point clouds

Figure 3.61: Papazov et al.

experimental results proposed the method seems to achieve excellent performances, identifying and properly matching the model even in conditions of cluttered data or object occlusion, fig. 3.61(b).

The most interesting element of this work is however the use of a set of geometric *invariants* that simplifies and accelerates the matching process. The elaboration is based on the class of algorithms called *surface registration*.

Rigid surface registration consists of computing a rigid transform which aligns two surfaces.

Assume that S is a surface represented by a set of oriented points. According to Winkelbach et al. (2006), for a pair of oriented points $(u, s) = ((p_u, n_u), (p_v, u_v)) \in S \times S$ a descriptor $f : S \times S \to \Re^4$ is computed as:

$$f(u,v) = \begin{pmatrix} f_1(u,v) \\ f_2(u,v) \\ f_3(u,v) \\ f_4(u,v) \end{pmatrix} = \begin{pmatrix} \|p_u - p_v\| \\ \angle(n_u, v_v) \\ \angle(n_u, p_v - p_u) \\ \angle(n_v, p_u - p_v) \end{pmatrix} \tag{3.17}$$

where $\angle(a, b)$ is the angle between the vectors a and b.

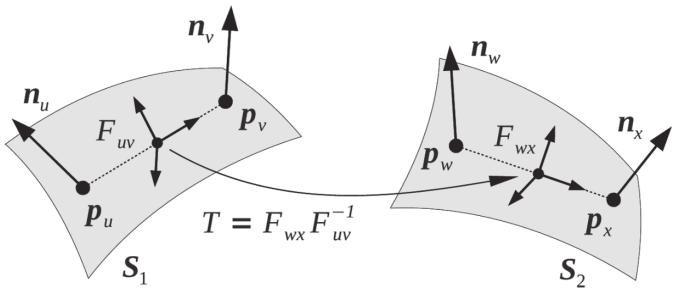


Figure 3.62: Papazov: descriptors

In order to register two surfaces S_1 and S_2, each one represented by a set of oriented points, the method proceeds as follows. It samples uniformly oriented point pairs $(u, v) \in S_1 \times S_1$ and $(w, x) \in S_2 \times S_2$ and computes ans stores their descriptors $f(u, v)$ and $f(w, x)$ in a four-dimensional hash table. This process continues until a collision occurs, i.e. until $f(u, v)$ and $f(w, x)$ end up in the same hash table cell. Computing the rigid transform T which aligns (u, v) to (w, x) gives a transformation hypothesis which registers S_1 to S_2.

$$T = F_{wx} F_{uv}^{-1} \tag{3.18}$$

is computed based on the pairs' local coordinate systems, each one represented by a 4×4 matrix (homogeneous coordinates) F_{uv} respectively F_{wx}.

$$F_{uv} = \begin{pmatrix} \frac{p_{uv} \times n_{uv}}{\|p_{uv} \times n_{uv}\|} & \frac{p_{uv}}{\|p_{uv}\|} & \frac{p_{uv} \times n_{uv} \times p_{uv}}{\|p_{uv} \times n_{uv} \times p_{uv}\|} & \frac{p_u + p_v}{2} \\ 0 & 0 & 0 & 1 \end{pmatrix} \tag{3.19}$$

where $P_{uv} = P_v - P_u$ and $n_{uv} = n_u + n_v$. Here F_{wx} is defined analogously by replacing the induces u and v in 3.19 with w and x, respectively. The transformation hypothesis T generated in this way is evaluated by transforming the points of S_1, i.e. $p_i' = Tp_i, \forall p_i \in S_1$ and counting those p_i' which fall within a certain ϵ-band of S_2.

According to Winkelbach, this process of generating and evaluating hypotheses is repeated until any of the following stopping criteria are met:

- a hypothesis is good enough
- a predefined time limit is reached
- all combinations are tested

Such approach is extended to all the surface of the model, building an hash table for the matching process. This is based on a *RANSAC*, Fischler and Bolles (1981) matching, in which a subsequent set of correspondences between model and data are iteratively evaluated. In 2 is reported the elaboration stricture of the algorithm proposed by Papazov.

Data: Object model M

Result: A list of transformations T, solutions of the identification

1)initialization:

compute an octree for the scene S to produce a modified scene $S^* \rightarrow$ discretization of the input cloud ;

T is set to empty ;

2)number of iterations:

compute the number N of iterations ;

forall the *N times* **do**

3)sampling:

sample a p_u uniformly from S^* ;

$L = x \in S^* : \|x - p_u\| \in [d - \delta_d, d + \delta_d] \rightarrow$ distance sub sampling ;

sample a p_v unifromly from L ;

4)normal estimation:

estimate normals n_u at p_u and n_v at p_v ;

$(u, v) = ((p_u, n_u), (p_v, n_v)$;

5)hash table access:

$f_{uv} = (f_2(u, v), ..., f_4(u, v))$;

access the model hash table cell at f_{uv} and get its oriented model point pairs (u_j, v_j) ;

6)generate and test:

forall the (u_j, v_j) **do**

get the model M of (u_j, v_j) ;

compute the rigid transform T that aligns (u_j, v_j) to (u, v) ;

if Acceptance function*(M,T)* **then**

T is added to the solutions ;

end

end

end

7)removing conflicting hypotheses:

remove conflicting hypotheses from T ;

Algorithm 2: Papazof: elaboration scheme

From the guidelines of the paper was developed a Matlab routine. The model used is the one shown in fig. 3.63, built in parametric form in order to vary the density of the points on demands, instead of the octree discretization, and so the hash table of the features. It is used only the frontal face of the pallet, choice derived from an

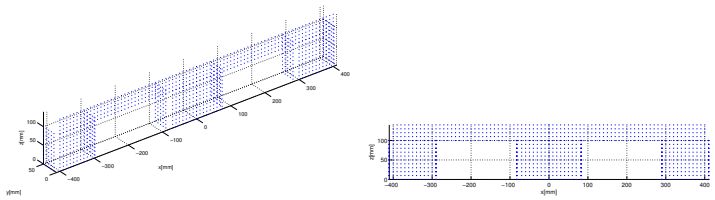


Figure 3.63: Discrete 3D model of the pallet

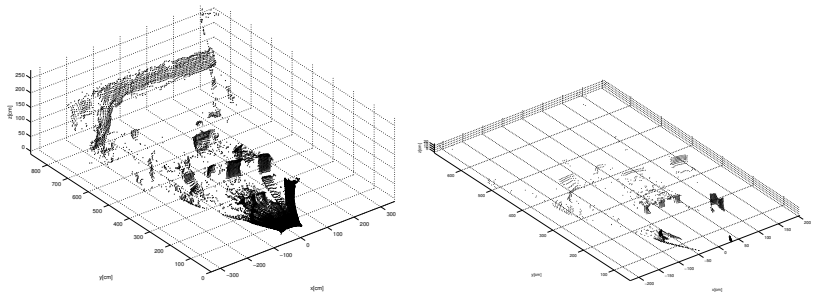
observation: the pallets can be both loaded or unloaded, the upper part can then be or not seen by sensors. The only invariant part is therefore the frontal face.

The models contains also part of the sides of the feet. The symmetrical shape of the face can lead to backward solutions, in which the model is oriented pointing outside the pallet. The implementation of these parts strengthen the process of matching by forcing the right orientation of the model over the data.

A further element that improves the performances of the algorithm is the preprocessing of the depth map: it is oriented in order to have the z normal to the ground. After this transformation it is possible to define a threshold in the admissible height z of the points: the ones higher than the pallet, placed on the floor ($z = 0$), are removed (a tolerance of 10% is kept), fig. 3.64. This decimation implies that the pallets lifted from the ground cannot be identified, but the main benefit is a more robust and fast elaboration.

A critical element is the evaluation of the right density of the mesh of the model. A value too high or too low, with an admissible variability of millimeters, cause the failure of the matching. This problem is not highlighted in the paper for three two reasons:

- the sensor is placed in a fixed position
- the objects are placed in limited, known, area



(a) 3D map (b) Decimated map

Figure 3.64: TOF: depth map preprocessing

- the sensor is not a TOF but a Microsoft Kinect

This setup allows the acquisition of *dense*, *undistorted*, point clouds, in which the density of the points over the surface of the objects is approximately constant, parameter used in the octree discretization of the model.

The same it is not true for the current application: the position of the pallet is not known a priori and can vary of meters. To that it must also be added the distortion of the lenses of the sensor and the limited resolution of the chip, these cause a high variability of the number of points over the surfaces depending on the position of the pallet in front of the TOF. The same object, seen from 2 different point of view, can generate different features depending on the density of the points.

The identification of the right density depends on the position of the pallet, the position of the pallet is found using the right density: this is a chicken-egg problem.

Such issue is solved by running a brute force processing, accomplishing the matching with different densities in the model.

Each identification resulting from the *ith* process, performed with the *ith* density, is stored in a vector as a 3D pose in space, $[x, y, z, \alpha, \theta, \gamma]$, position and attitude. Once completed the brute force analysis for the defined range of densities, these poses are used to move the model inside the 3D environment and isolate subspaces of the input point cloud: given the shifted model, only the points of the 3D cloud closer than 10cm are kept for the successive analysis, fig. 3.65. The regions with less points than

the model are rejected.

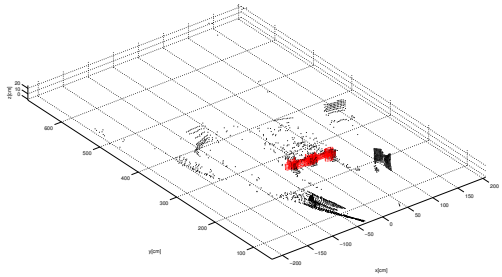
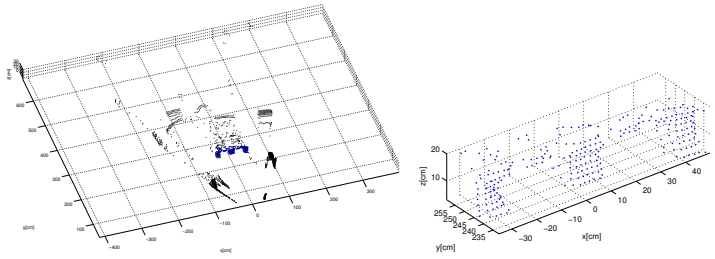


Figure 3.65: TOF: candidate solutions

From each subspace is evaluated a score factor, equal to the mean distance between the points of isolated region and the ones of the model. Formally

$$s_i = \frac{\sum_1^N min \, \|P(p) - M(m)\|}{N} \tag{3.20}$$

where N is the number of point of the subregion, $min \, \|P(p) - M(m)\|$ the minimum distance of a $P(p)$ point of the subregion from the $M(m)$ point of the model. The pose associated to the minimum score is the one that represent the best identification. This, and the associated isolated points, are then used in a refinement process based a standard ICP, so the result of the elaboration.



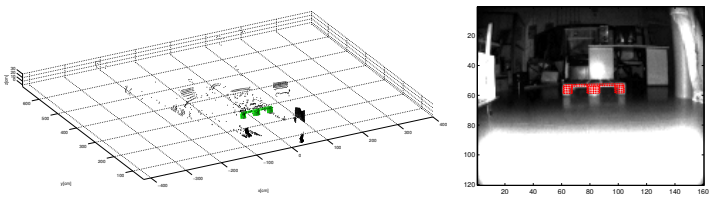
(a) Sub region identified from the best candidate

(b) Subregion, detail

Figure 3.66: TOF: sub region selection

The result of the matching is represented in a dual form: as the model superposed to the 3D point cloud, fig. 3.67(a), or as the model over the image, fig. 3.67(b).

Since the camera and the depth sensor are the same, the Bouguet calibration can be used to achieve a first approximation of the intrinsic parameters of the camera (for the image, not the depth), thanks to that it is possible to project the result obtained in the 3D on the image. Such step has the main advantage to express the result directly inside the image, a step broadly used in the fusion phase in order to merge in a single result the identifications of the different devices (laser and camera).



(a) Model inside the 3D map (b) Madel over the image

Figure 3.67: TOF: identification

In fig. 3.68 are shown the identifications from a sequence of samples acquired mounting the TOG on a moving robot.

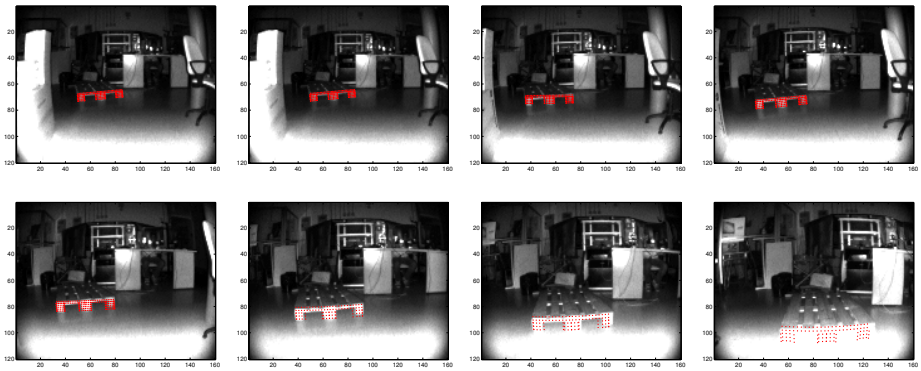


Figure 3.68: TOF: sequence

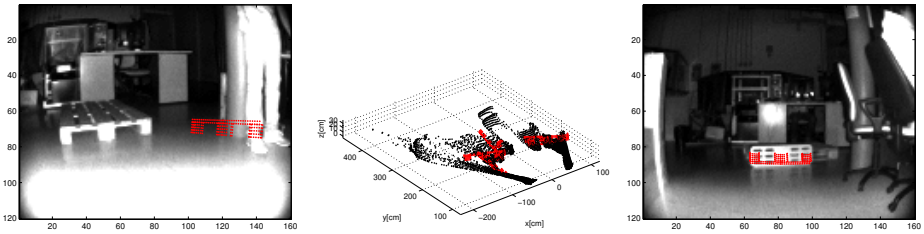
The main issue in this process is the time consumption. The brute force must be tuned in order to optimize the admissible densities to apply to the model. The range depends on the operating conditions: for a pallet far more than 3 meters the density should be higher than 2cm (1 point each 3 centimeters, a grid), closer pallet are

correctly identified if the model is instead more dense, 1cm.

From the experimental results the best performances in terms of processing time and correctness of the matching are obtained using a range of densities from 1.5 to 2.5cm.

Despite some positive results, achieving the correct identification of the pallet in generic environment, the algorithm resulted to be nor stable nor repeatable. The control of the mean distance between points and model is not a robust method. The great variability of the scene and the limited number of points of the target object limit the capability of the algorithm to achieve a matching. The missed identifications are not isolated cases..

A further important limitation of this algorithm is the missing of an absolute reference/method that evaluates if the identified object is indeed a pallet: some false identification occur, the worst condition for the purposes of the work.



(a) Wrong identification (b) Wrong candidates (c) False positive

Figure 3.69: TOF: wrong identifications

The Papazof algorithm was subsequently tested in the form of C/C++ code using the dedicated library *ObjRecRANSAC* of the Point Cloud Library, PCL, Rusu and Cousins (2011).

`[http://docs.pointclouds.org/1.7.0/`
`classpcl_1_1recognition_1_1_obj_rec_r_a_n_s_a_c.html]}`

No improvements were highlighted.

3.4.2 Perspective

One of the most interesting aspects of the TOF cameras is the synchronous acquisition of a grayscale image together with the 3D point cloud. Such data can be directly used in order to get a second identification, this time based on the silhouette of the object rather than on its geometrical shape. The strategy adopted is the same used for the RGB camera, a HOG processing is run as identificator.

The main issue in using such algorithm with this device is the low resolution of the outputs. The HOG requires the evaluation of a model, built by mean of a supervised training. That operation is achieved without problems with the RGB camera (images of 1024x786 pixels), the images of 160x120 pixel from the TOF cause instead the crash of the procedure.

The motivation is related to the small sizes of the regions that mark the pallet inside the images used for the training, these are used to build the pyramidal structure of the model. The training needs at least a minimum size (and resolution) for the input samples, condition not fulfilled with the image format of the TOF.

A possible solution is to acquire the samples maximizing the size of the pallet n the image, that is however not possible because the emission of the infrared lamps saturates all the objects closer than 1 meter far from the camera, making them not visible.

A functional solution is instead the resize of the images, doubling their size through a cubic interpolation. With such configuration the training converges to a solution, fig. 3.70(a).

In fig. 3.70(b) is shown an example of correct identification using the model obtained doubling the size of the training set.

Compared to the RGB camera the performances are however very poor. The low resolution of the input images and the high distortion caused by the lenses limit the success ration of the process. Even in a controlled environment with stable conditions many identifications are missed.

(a) Model of the pallet

(b) Identification

Figure 3.70: TOF: HOG identification

(a) Image form TOF (b) Synthetic depth map

Figure 3.71: Image vs depth map

A further strategy has been tested. Instead of using directly the images acquired by the sensors, a synthetic depth maps are generated from the point cloud, fig. 3.71, and used as input for the identification process. Even in this case the maps must be rescaled to complete the training. Such approach didn't achieve better result in terms of identification rate.

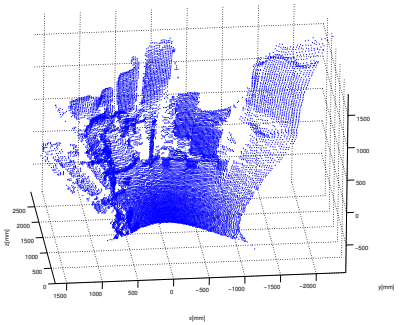
The strategy used for the camera seems in this case not convincing. The limit of the resolution deeply influences the processing, causing the HOG to be not so effective as for the camera and so not suitable for an industrial usage.

3.4.2.1 Segmentation

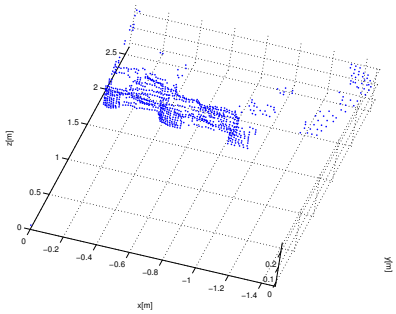
Interesting it is the use of the identification obtained from image as initialization of the 3D matching, inverting in practice the logic implemented in AGILE.

The HOG identification is associated to a boundary box that wraps the object in the image, structure used to evaluate the performances of the algorithms in PASCAL Challenge. Since the camera and the depth sensor share the same pixels, the boundary box defines a region not only for the image, but also for the point cloud. Such region can be indeed used to segment the range data, fig. 3.72(a). The 3D cloud is organized on the sensor memory as a structure of 3 matrices, one for each principal direction $[x, y, z]$; the cells of each matrix inside the boundary box correspond to the subspace to keep, fig. 3.72(c).

(a) HOG identification



(b) Depth map



(c) Segmented area

Figure 3.72: 3D segmentation from image

This strategy is functional but not optimal: the configuration that ensures the best reliability is the one in which the identifications are achieved independently and the sensors involved do not influence themselves reciprocally.

The use of the camera as input for the matching means to force a solution in a defined location, losing as consequence other possible identifications and the generality of the method. Worst it is instead the case in which the camera produces a wrong identification, a false positive: a possible, dangerous, erroneous matching could be generated form the associated depth analysis.

A further negative element to consider is the low efficiency the HOG processing with the images from the TOF. As initialization for the depth processing, the graphic part must ensure a *continuous* source of information, achieving frequent identifications to be passed to the second stage of the matching. Such performances are not achieved with the actual configuration.

Two possible solution could solve such issue: calibrate the camera in order to minimizer the distortions (still working with low resolution images, not optimal), or couple an external RGB camera as support sensor (same configuration of laser-camera). Both the solution are not optimal for the purposes of the application: both present drawbacks compared to the laser-camera solution, which is simpler, less expensive and offers more accurate and robust results.

3.4.3 Comments

Although the results are encouraging, getting a double identification of the pallet with a single sensor, there are different negative elements that mark the use of the TOF as not suitable sensor for the automatic identification of pallets in an unconstrained environment or lack of priori information.

Many limitations and issues have been indeed encountered during the development:

Deformation of the data

Both the 3D cloud and the image are very deformed, typical of the TOF technology. An entire research field is focused on how to improve the measurement performances of these devices, developing innovative calibrations based on a more refined modeling of the sensors. Kahlmann et al. (2006); Weyer et al. (2008); Fuchs and Hirzinger (2008); Chiabrando et al. (2009); Fuchs (2010); Lindner et al. (2010); Piatti (2011) are only few among the various examples of a still active research.

From the development of more accurate sensors will derive the success of many 3D application based on the recognition of objects in 3D: the models are usually built with synthetic geometries or files (CAD), these don't match with the point clouds provided by the sensor because of the deformation of the data, causing the failure even of the most advance matching techniques

The same happens with the images, the silhouette of the object is deformed, lowering the identification rate of the HOG. Compared to the results obtained with the RGB camera the ratio is about 10:1, 10 successive identifications of the camera to obtain 1 with the TOF.

Low resolution

TOF cameras use CCD chips with a limited number of pixels, usually less than 20.000 (19200 in this case). Even if that number seems elevate, compared to a standard industrial camera it is almost 2 orders of magnitude lower (1600x1200 pixels). Such characteristic represents a strong limitation for both the Cartesian and the perspective identification.

About the Cartesian part, a pallet placed 3 meters far from the sensor presents a number of 3D points over its surfaces varying from 150 to 250, these are also very noisy. Both elements deeply influence the calculation of the features used in the matching, often obtaining erroneous results between the 3D data and the model. The consequence is an unstable algorithm that must be run several time in order to optimize the model (its density) and then analyze a vector of possible results in order to isolate the most reliable ones.

Similarly, the low resolution of the image weakens the HOG processing. As explained earlier, the resolution of an image is correlated to the information content: the amount of pixels influences the accuracy and reliability of the algorithm. A low resolution input image forces the algorithm to work with few pyramidal level, the ones close to the top, lowering the overall performances.

Narrow field of view

The field of view of the TOF used is the wider found on the market: $70^{\circ} \times 53^{\circ}$. With this configuration a pallet placed 4 meters far is entirely visible only within a transversal range of about 3 meters.

Such admissible area is comparable with the field of view of the camera, but not with the one of the laser, 190°. A wide field of view is useful in order to pre-process the cargo area while the vehicle is approaching, in this way it is possible to localize hypothetical candidates before starting the real process of identification and picking. On the opposite a narrow field of view limits the operative performance of the AGV, requiring a priori knowledge of the position of the pallet, gross but necessary, in order to preventively align the vehicle with it.

Low depth

The main limit of this technology is however the measurable depth. The measurement process is based on the active illumination of the environment, so the power of the lamps defines the maximum visible depth.

Must be pointed out that there is a strong difference between an indoor environment, like a room or a laboratory, and a open one, like a warehouse. In the first case the emitted light remains inside the environment, the device indeed works with the specified measurement range. On the opposite, when working inside a warehouse, the light becomes diffuse, making some regions of the environment no visible by the sensor. In fig. 3.73 is reported a sample acquired inside a warehouse. The points are very noisy and depopulated.

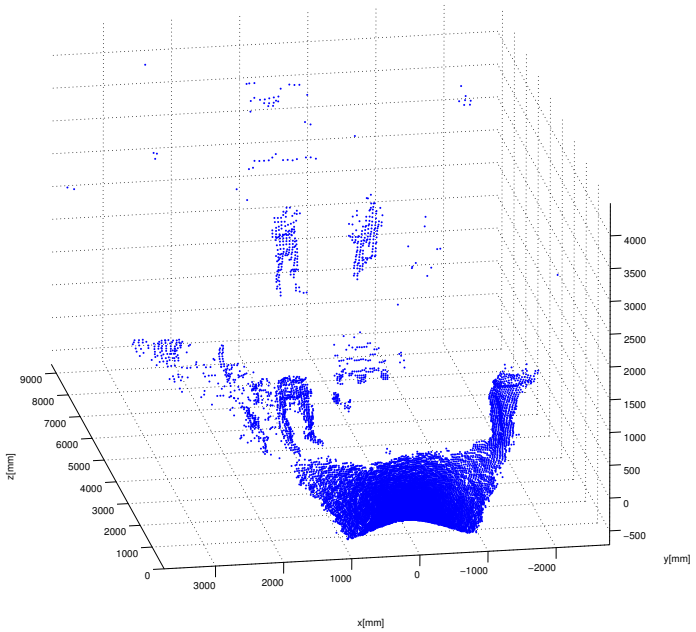


Figure 3.73: TOF: industrial environment, depth map

This 3D technology is still very limited, not ensuring the reliability required for the current application.

Further negative element is that TOF cameras are not classified as a safety sensor (do not see the bodies that absorb infrared, those colored in opaque black). For this

reason, this sensor can not replace a safety laser scanner, resulting in practice as an extra, expensive, device to be equipped on the AGV.

For what has been said so far, the TOF technology does not seem a suitable solution for the industrial usage, remaining, for now, confined to the experimental field. It must be however underlined that this technology is constantly evolving, with innovative devices every year. Just consider to the TOF used in this research (model presented in the second half of 2010), in 2014 two new models have been presented, one with an high-power emission and a depth field of view of 10 meters, and one that incorporates an RGB high resolution camera, fig. 3.74.

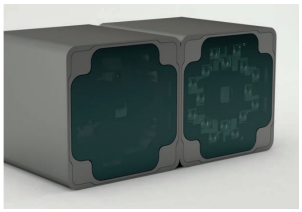


Figure 3.74: Innovative TOF: FOTONIC C-series

Seem than this technology can reach a suitable level of performances in few years, reaching the industrial level with 3D application and so an even higher level of automation compared to the one reachable with the standard 2D devices

CHAPTER 4

CALIBRATION

Perception is a process that involves more sensory organs able to detect different elements. The more complex and evolved is this process, more interesting and accurate are the information that can be obtained.
This is true for humans as for machines.

4.1 Introduction

ROBOTS, and automatic systems in general, require the evaluation of some relations in order to correctly interpret (and use) the data provided by the sensors. Typically these relations are the geometric transformations that describe the position of the devices on-board the robot, called extrinsic parameters, or characterize of the sensors themselves, the intrinsic parameters. The accuracy in evaluating such parameters always influences the operative performances of the machine. The main difference, however, that is detectable among the various applications is the dependence that exist between the tasks and such influence: there are cases in which the dependence is moderate, the data acquired are used for purposes of simple monitoring, and instead cases of high dependence, the data represent the initial step of a measurement process from which derives the entire behavior of the machine. This second case is far more critic than the first.

The actual application falls inside such category: the localization and picking of an object are indeed tasks that involve the measurement of a pose, the planning of a trajectory and the motion of a robot.

In the current application, as shown in chapter 3, there are two types of identifications: Cartesian, form the laser or the TOF, and non Cartesian, from camera. These informations must be combined, through a process of fusion (chapter 5), in order to increase the reliability of the results, checking the compatibility of the informations. The data structure used in the fusion requires both the data types, element achievable with two possible configuration of the sensors: laser-camera or TOF by itself. The main difference between the two is clearly the number of sensors, and so the number of relations, intrinsic and extrinsic parameters, that must be evaluated.

In the case of a laser-camera system, the first step is the estimation of the extrinsic parameters that relate the devices. With the knowledge of these parameters it is possible to express the data with respect to all the possible reference systems: laser to camera and camera to laser. At the same time the sensors should be characterized by identifying, if necessary, specific intrinsic parameters, like the distortions and focus of the optic mounted on the camera.

On the opposite, a TOF camera does not require any extrinsic calibration, the sensing device is same for the depth data and for the image. As underlined in the previous chapter, a fundamental step for such devices is the calibration of the intrinsic parameters in order to compensate the distortion of the depth map, 3D, and of the image, 2D. Such calibration is a state of the art topic, not treated in the current research due to the metric limitations and low performances achieved in the identification process associated to the device. Must be pointed out that, once more accurate sensors reach the industrial level and standardized calibration becomes available, the concepts and methods described in the following chapter, so as for the others, can be directly extended to this 3D technology with minimum modifications. These have been developed and tested with a laser-camera system, but they are indeed general.

A further calibration to consider is the one that evaluates the extrinsic parameters between the sensors and the machine on which they are mounted. In order to achieve tasks like the path planning, in which there is an interaction between the robot and the environment, it is fundamental to transform the information obtained by the sensors into a representation usable by the AGV. There is indeed a deep difference between

the identification of a object and the information that a robot needs to achieve the picking. This problem is also known as *Hand-Eye* calibration, which determines the relations required to move the *hand*, the robot, toward what has been seen by the *eye*, the perceptual system.

In the following paragraphs the various calibrations, used and/or developed in order to setup the vehicle for the autonomous picking, are presented:

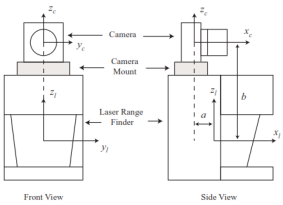
- calibration of laser and camera, intrinsic and extrinsic parameters
- calibration of laser and vehicle, extrinsic parameters

4.2 Laser and camera

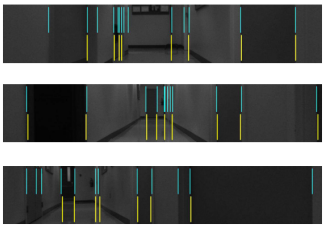
The camera-laser calibration is a well known procedure that has been studied and optimized during the last decade. It is indeed necessary for many modern robotic and computer vision applications. A non-exhaustive list of examples includes: the acquisition of ground-based city models by using the combination of a laser and a camera couple obtaining textured 3D structures, Früh and Zakhor (2004), the fusion of laser-shape features with visual appearance for object classification, Douillard et al. (2009), pedestrian detection Premebida et al. (2009), or for recognition and modeling of landmarks in outdoor self-localization and mapping, Ramos et al. (2007). Other applications are focused on 3D sensors, Scaramuzza et al. (2007), or devices derived from the motion of the entire laser-camera block, Fornaser (2009).

The fusion of the two sensor requires of the knowledge of their relative displacement, in this way it is possible to project the depth readings into the images or vice versa. This informations are usually estimated by performing a calibration between the devices, but there are also examples in which a different approach is adopted, Amarasinghe et al. (2008): the nominal geometrical values are used as initial guess for a subsequent refinement based on the planar alignment of the laser readings and graphical features extracted from the image, fig. 4.1.

(a) Sensors



(b) Scheamtic



(c) Features used for the refinement

Figure 4.1: Amarasinghe et al.

Despite the various examples of the literature in which are involved laser-camera systems, the number of published works focused precisely on the laser-camera calibration is relatively small. The most broadly used method was proposed by Zhang

and Pless (2004), a paper that describes a practical procedure where a checkerboard pattern is freely moved in front of the two sensors, fig. 4.2. The poses of the checkerboard are computed from plane-to-image homographies (Zisserman and Hartley (2000)), and the camera coordinates of the planes are related with laser depth readings for establishing a set of linear constraints in the extrinsic calibration parameters. The solution of the system of equations provides an initial estimate for the relative rotation and translation that is subsequently refined by iterative minimization of the reprojection error.

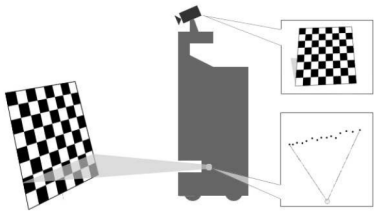


Figure 4.2: A schematic of the calibration problem by Zhang.

Among the different papers some can be underlined for the novelty in the approach and/or the modification of the aforesaid method:

- Kassir and Peynot (2010), focus on how to make the calibration process more automatic. The paper deals with the setup of a laser camera system mounted on a car. The number of samples to acquire and the time required for the calibration are not considered critic elements. The complexity of the operations and the supervision of the process are instead the elements that the author aims to simplify in order to have a tool usable in different configurations with the minimum effort. The main contribution of this works it is actually the Matlab toolbox associated to the paper, called Robust Automatic Detection in Laser Of Calibration Chessboards, RADLOCC.

  `http://www-personal.acfr.usyd.edu.au/akas9185/AutoCalib/AutoLaserCamDoc/index.html`

- Vasconcelos et al. (2012) studies and highlights the limits of the method introduced by Zhang, describing an optimized version of the standard algorithm

using the theoretical minimum number of input data for the calibration. The main contribution is in this case a more structured approach to the problem, from which is derived a quicker calibration due to the reduced number of scans and images to be acquired.

- Zhen et al. (2012) presents a new calibration procedure that uses a more complex calibrating target. Instead of a planar chessboard it is used a 3D shape: a cube with chessboard on the faces, fig. 4.3. This allows the user to acquire less samples keeping the overall accuracy of the calibration. The main drawback of this approach is the construction of the model: in order to obtain good results the quality of the model must be elevate (more complex compared to a planar chessboard).

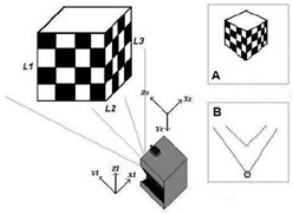


Figure 4.3: Chen calibrating target.

- Li and Nashashibi (2013), in this work it is analyzed the method of Zhang explaining and verifying the different performances that can be obtained in the calibration depending on the number of input data used. The innovative element of this work is the modifications that it is applied on the process of optimization starting from a consideration: in most cases the laser-camera system is used in automotive or robotics applications where the objective is to monitor the terrain and objects close to it, obstacles. This element usually means that the sensors aim down. Such configuration involves that the calibrating the system must be close to the floor, placing the chessboard with a side always in contact with the ground. This fact allows the introduction of additional constraints in the optimization, obtaining as a benefit an higher accuracy, compared to the original method, using the same number of input data.

Despite these methods and tools are all effective in performing the calibration pro-

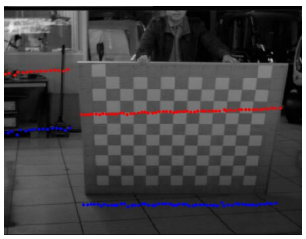


Figure 4.4: Calibrating target placed on floor by Li and Nashashibi

cedure, they still require an experienced operator that acquires the data and monitors the process in order to achieve reliable and accurate results. Since the current application is addressed to the industrial field, and the will of developing a device that requires the less possible expertise from the final users, an important requirements that the calibration procedure should ensure is the ease of use. For this reasons the method chosen and used is the one that ensures the higher level of automation: the RADLOCC.

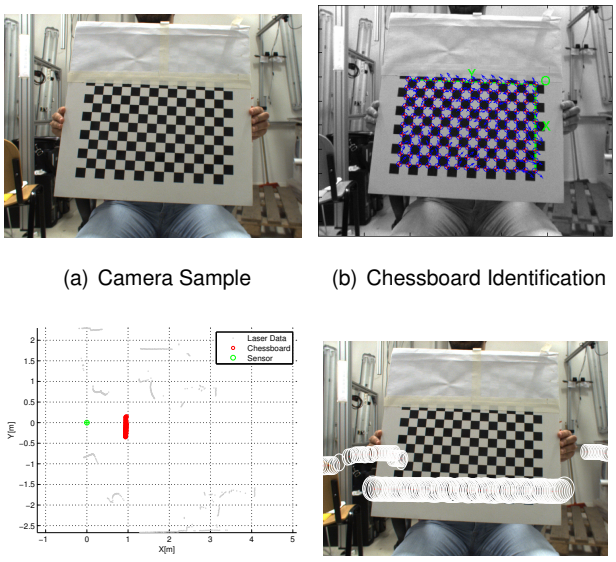
As stated by the authors, but also verified in practice form the tests performed, the RADLOCC provides an automatic and robust calibration of the laser-camera system. The goal of the procedure is to find accurate estimates of the intrinsic parameters of the camera and the rigid transformation between the camera and the laser under the assumption of known intrinsic parameters of the laser (technical specifications of the laser scansion from datasheet). The procedure can be divided in two stages:

- automatic camera calibration
- automatic extrinsic camera-laser calibration

Rather than pursue self-calibration methods, Kassir uses algorithms that automate two existing trustful calibration methods which rely on observing a calibration object and which jointly achieve complete camera-laser calibration: Bouguet's Camera Calibration, Bouguet (2008), and Zhang and Pless's extrinsic camera-laser calibration technique.

To use these techniques, the operator is required to obtain a calibration dataset, which is a set of synchronized pairs of images and laser scans, containing a chess-

board and taken from different poses. The chessboard acts as the calibration object and the size of its squares needs to be measured and provided to the algorithm. For the camera calibration, the corners of the chessboard squares need to be extracted from the images. In Bouguet's toolbox, this is attained by the outer corners being selected manually, obtaining as outputs the camera intrinsic parameters as well as the rigid transformation from the camera to the chessboard for each image. The transformation is necessary for the camera-laser calibration technique cited above, fig. 4.5(b).



(a) Camera Sample

(b) Chessboard Identification

(c) Laser scansion and chessboard selection

(d) Reprojection of the laser points over the image

Figure 4.5: Process and result fo the laser-camera calibration

The camera-laser calibration then uses the points originating from the chessboard which appear in the laser scans, figure 4.5(c), to find the camera-to-laser rigid transformation. Typically, these points have to be extracted manually. In both cases, the extraction process is the key time consuming task for the operator.

Aimed at automating the entire procedure, two algorithms are presented by Kassir: the first automatically extracts the chessboard corners from each image, Robust Automatic Detection Of Calibration Chessboards (RADOCC), and the second extracts the

chessboard points from the laser scans. With the aid of these algorithms, the required operator time is reduced to what is needed for acquiring the calibration dataset and measuring the size of the chessboard squares. The first result that can be achieved once completed the calibration procedure is the reprojection of the laser scansion over the image, fig. 4.5(d).

This automatic camera-laser calibration method demonstrates accuracy while significantly reducing operator time when compared to the other methods.

The software of Kassir has been however modified removing the automatic chessboard detection from the laser part. Such algorithm, tested under various operating conditions, fails to achieve sufficient reliability and robustness for the purposes of the current application. In order to stabilize the automatic detection of the chessboard it is necessary that the board on which the pattern is attached has dimensions not comparable with the objects in background: the identification is indeed based on the dimensional matching between clusters of continuous points from the input laser scansion and a reference length that must be set as parameter in the algorithm. Although this requirement can be fulfilled by using a large checkerboard (longer side of 1.5m or more), that solution causes issues in the calibration of the camera. The use of a large chessboard implies to move the calibrating pattern more far from the camera in order to make it entirely visible. The calibration of the camera needs that the chessboard is moved at various distances and orientations in order to cover its entire field of view, taking care to acquire some samples in which the chessboard covers the whole image (optimal for the evaluation the intrinsic parameters). It must also be pointed out that the camera calibration is heavily dependent on the number of squares of the chessboard: the more dense is the chessboard, the more accurate will be the calibration. Both the aforesaid elements are however related to the resolution of the camera: patterns that are too dense and/or too far from the camera could be acquired with a resolution too low to be correctly processed. A large checkerboard could therefore limits the procedure.

For the aforesaid reasons, a checkerboard of compact size is preferable, allowing the user to complete the procedure even inside a limited space. The pattern used

in the calibration of the current laser-camera system was a checkerboard made of squares of 2.5cm, moved at distances ranging from 0.5 to 2m (depending on the lenses).

About the laser part, the detection of the chessboard is achieved by selecting manually the laser points belonging to the board. This operation is performed inside a dedicated GUI i which the user uses a *lasso* to easily select a the points. After this step the data is passed to the calibration routine which is run without further modifications.

Despite the loss of automation in the laser part, this structure proved to be more robust, with am overall time for the calibration lower then 5 minutes.

As said in the introduction, the uncertainty plays an important role in the actual development. The aforesaid calibration provides such information. The parameters that are involved the transformation between laser and camera, and that influence the uncertainty of the process, are $[x_{lc}, y_{lc}, z_{lc}, R_{lc}, f_{lc}, c_{lc}, k_{lc}]$ where

- $[x_{lc}, y_{lc}, z_{lc}]$ is the vector that connects the reference systems of laser and camera
- R_{lc} is the rotation between laser and camera, 3 rotations
- $[f_{lc}, c_{lc}, k_{lc}]$, focal, camera canter and distribution coefficients, are the intrinsic parameter of the camera

Such parameters, and their uncertainties, are used to estimate the uncertainty of the projection of the laser data over the image, fig. 4.5.

A similar approach is used in the fusion process to project the laser identifications over the image, the formulation in that case is modified including also the contribution of the uncertainty of the *position* of the pallet, more reference in chapter 5.

Results

The result of the calibration is a structure that includes the extrinsic parameters between laser and camera,tab. 4.1 and tab. 4.2, and the intrinsic parameters of the camera, tab. 4.3.

(a) Devices mounted on an experimental AGV (b) Protoypal version (c) Final configuration of the sensors

Figure 4.6: The laser camera system used

Table 4.1: Relative displacement

X[m]	Y[m]	Z[m]
-0.1253	0.1325	-0.1459

Table 4.2: Relative orientation

-0.99975	-0.01202	-0.01878
0.012489	-0.9996	-0.02523
-0.01847	-0.02545	0.99951

Table 4.3: Intrinsic parameters of the camera

focals[x,y][pixel]	image center[x,y][pixel]	K, distorsions
1364.86	703.62	-0.10569
1359.88	477.06	0.079738
		-0.00223
		6.26E-05

Once mounted sensors on-board of the vehicle the data that can be obtained are like the ones shown in fig. 4.7, the reprojection of the laser data over the image.

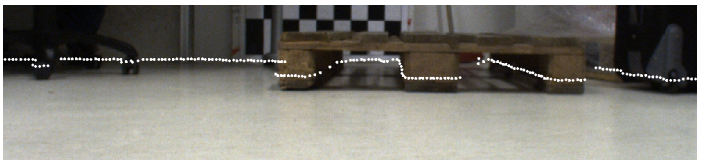


Figure 4.7: Data from laser and camera after calibration

A critical element for the application is the position of the laser: from its height from the ground depends the detectability of the blocks of the pallet. The best performances are achieved when the laser is planar to the floor, scanning the wider possible area. Such configuration not always can be adopted: usually the AGVs mount the safety laser in the middle of the forks, on the chassis of the vehicle, with an height of the scanning plane from the floor around to 10-15 centimeters. This configuration must be changed in order to lower the height of the sensors.

If the laser is instated mounted directly on the forks, like in fig. 4.6, its aim must be regulated in order to not interfere with the forks themselves.

A second element that should be highlighted is the parallax between laser and camera. This effect is independent from the calibration: the fields of view of the sensors are indeed different and shifted, causing an inconsistency in the representation of the laser points over the image and vice versa. The higher is the baseline between the sensor, the less consistent is the projection of the points. Fig. 4.8(a) shows how some laser points are projected over the image in wrong spots, that is due to the overlap inside the image of geometries that are not from the point of view of the laser. Applying a filter on the visibility of the points this effect can be solved, fig. 4.8(b).

The parallax does not represent an issue for the fusion process, chapter 5, but it is nevertheless advisable to place the two sensors in the closest possible configuration in order to have a coherent visual and geometrical representation of the environment. Sensors distant from each other cause a decrease of the possible configurations in which the pallet can be doubly identified, both by laser and the camera. Pallets with an

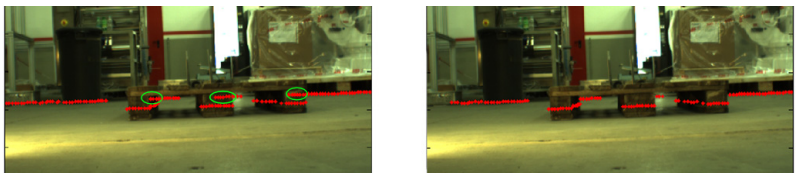


Figure 4.8: Laser points over the image, filtering effect

high relative orientation, more than 20°, are not identified due to the mismatching of the geometry exposed, laser side, or for the perspective deformation of the silhouette of the object on the image. The higher is the baseline between laser and camera, the more influencing is the angular displacement between object and sensors, reducing the probability of seeing the pallet with a suitable configuration contemporaneously by both the devices.

4.3 Vehicle to Sensors

The identification of an object a is very useful result from the point of view of the perceptual capabilities of an automatic machine, however it loses its meaning if the machine is not able to interact with the object itself. The information that the control logic of the vehicle needs is not the identified presence (or not) of the object, but where it is positioned.

The autonomous picking requires that the identification of the pallet is related to a reference position within the reference space used by the robot, the only way in which the AGV can plan a path and run a maneuver. The identification of the pallet achieved by the laser can not be directly used for the planning: the sensor works inside its own space, which is different from the one used by the AGV. The identifications must be therefore transformed in order to be used.

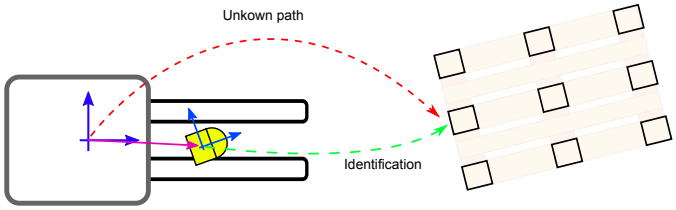


Figure 4.9: Pallet identification vs sensor position

From the academic works that deal with similar applications can be noticed that, in almost all the cases, there isn't any specific analysis on this topic, avoiding any study on how to determine these geometrical parameters and using instead the nominal ones. In most of the applications in which are involved sensors on board, an error in the nominal geometric parameters has a marginal influence on the performance of the system. In the current application this assumption falls: the evaluation of the extrinsic parameters is critical as from it directly depends the success of the picking of the pallet. A very accurate identification process associated with a poor estimation about where the sensors are placed on-board implies the planning of incorrect trajectories, with a final pose of the AGV unsuitable for the forking. It is therefore necessary to develop a calibration that estimates the pose of the sensor on-board

As for the laser-camera calibration, this process should include the estimation of the uncertainty or the covariance matrix of the parameters. The covariance is a good indicator about how much accurate the calibration process is, and it keeps trace of the uncertainty associated to the transformation evaluated. The error in the calibration has indeed a strong influence on the path planning and where to steer the AGV, from that depends the safety of the entire process, chapter 6.

To determine the pose of the sensor with respect to the vehicle the two reference systems must be somehow correlated. These, however, do not share any compatible information, 2D scansion vs a 2D map, and so it is not possible to perform a direct evaluation of the extrinsic parameters. The calibration can only be achieved by means of an indirect measurement.

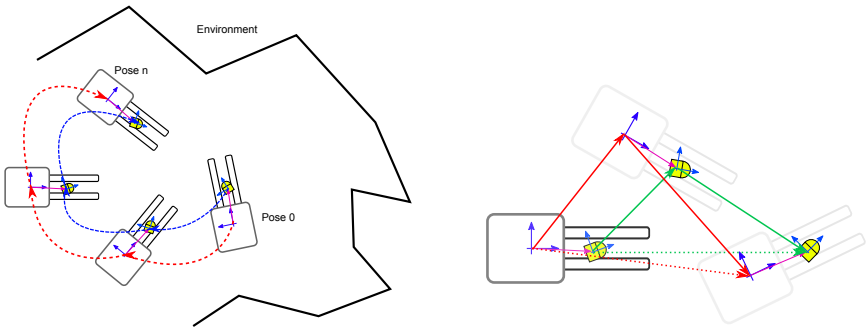
The calibration developed is based on an observation: both the systems involved provide data from which can be estimated their motion. From the readings of the encores on the wheels, applying the odometric reconstruction from the kinematic model of the robot, the motion of the AGV can be determined. From the scansions of the laser, applying localization techniques, a similar result can be achieved.

The following assumptions are taken:

- the laser and the vehicle move coplanar to the floor
- the laser and the vehicle are rigidly fixed
- the pose of the vehicle and laser scans are acquired synchronously

If the vehicle is moved in different positions in space while acquiring laser scansions, two trajectories can be defined: one for the vehicle and one for laser sensor, fig. 4.10. These differ from each other except for a constant rigid transformation that transforms one into the other and vice versa. Such transformation is precisely the position of the laser sensor expressed within the reference system of the vehicle.

Assuming to know the location of the laser sensor on board, it is possible to repre-



(a) Calibration setup (b) Relative displacements of laser and AGV

Figure 4.10: Laser-vehicle calibration setup

sent all the positions of the vehicle and laser within the AGV reference system. This originates the scheme in fig. 4.10(b), a chain of segments.

The two paths must be acquired synchronously, requirement that can not be fulfilled during the motion of vehicle due a lack of any triggering source between the robot/AGV and the laser. The system is therefore organized in order to record the samples stopping the vehicle and acquiring only in this case the laser scansion (or multiple ones). Such discretization is a not a simplification of the procedure but a forced choice due to technical limitations associated to the experimental robot used. The synchronous acquisition of laser scansions and positions of the vehicle represents the calibration set to be passed to the process as initialization.

The information content is given by the relative displacement of the reference systems of laser and vehicle along the trajectory performed. The geometry of the problem can be defined as a set of quadrilateral polygons obtained by breaking the paths into segments that connects the different poses, fig. 4.11. This geometric configuration is valid for all the possible combinations of vehicle-laser samples, all the possible permutations are then used in the calibration.

Each sub polygon is made of 4 segments:

- one that connects 2 laser poses, green
- one that connects 2 vehicle poses , orange
- one that connects the vehicle to the laser, purple, which is the same for all the sampled data ($2\times$)

The first two are evaluated with the data acquired, laser scansions and odometry, the last is the unknown of the problem.

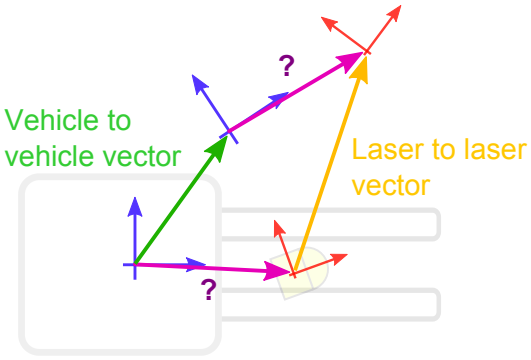


Figure 4.11: Vehicle-laser Polygon

The closure of the polygon can occur only with the right estimation of the pose of the laser, the only configuration that makes compatible the relative displacements of laser-vehicle pairs of samples. The process of calculation is then based on such geometric constraint: defined a number n of polygons, these are used inside a cost function for an optimization process aimed to close them.

The way in which the poses of vehicle and laser are estimated is critical for the purposes of calibration since from them derive the construction of the polygon. In the next two paragraphs are then described the techniques used to evaluate such poses and the associated uncertainties.

Vehicle

It is necessary to evaluate the poses in which the vehicle is stopped. Such task is facilitate in those cases in which a global localization system is employed, with a very accurate estimation of the pose of vehicle in real time. Those devices (SICK NAV) are commonly equipped on the industrial AGVs, but that is however not a general condition. The choice is to work with the source of data that any AGV has: the readings of the encoders mounted on the wheels. From this data the position of the vehicle is computed incrementally using the kinematic model of the robot.

The main difference between a global localization system and an incremental one is the uncertainty associated to the estimation of the pose: in the first case it is practically constant and always within certain range (usually close to 3-5 centimeters), in the second case the uncertainty is instead directly related to the path traveled, increasing during the maneuver. The incremental localization is instead based on a recursive formulation, in which the uncertainty associated to the geometrical parameters involved causes errors in the estimation of the pose. Since the last pose calculates is used as initialization for the successive iteration the error can only grow, *drift error*. For this reason, the incremental localization can be considered as the worst operative case. An industrial AGV, equipped with a more accurate global positioning system, will ensure more accurate results.

The lack of an absolute reference that limits the drift error implies that the pose of the vehicle becomes gradually more and more uncertain along the stops. That influences the estimation of the relative displacement of the poses of the AGV and so the segments used in construction the polygons. As previously underlined, the shape of the polygons is critical for the correctness of the calibration. Any discrepancy between the real motion and its estimation causes an incoherency in the segments of the polygon, from which derives the failure in fulfilling the constraints of the problem, the closure of the polygons: the motion of the laser in this case becomes not compatible with the one of the vehicle (and vice versa). The main consequence is the loss of the performances of the optimization, achieving a worst convergence, with higher residuals and therefore a more uncertain estimation of the position of laser on board.

Such problem is solved by including in the optimization process the influence of the errors on the measurements. Since the incremental localization suffers from the problem of drift, it is fundamental to develop a method capable of weighing the data according to their uncertainty. In this way the initial steps of the trajectory will have an higher influence on the calibration compared to the final ones (more uncertain), limiting the aforesaid effects on the optimization. In order to implement such strategy the uncertainty associated to the poses of the vehicle must be evaluated.

The available data are:

- the kinematic parameters of the vehicle (the movement is slow and therefore the dynamic effects are negligible)
- nominal geometry and uncertainty of the reference parameters (i.e. radius of the wheels, interaxis etc)
- encoder readings
- incremental localization from the controller of the robot

The work De Cecco et al. (2007a) presents a method that evaluate the uncertainty of the pose of a vehicle from the odometric reconstruction. The kinematics and navigation equations are referred on a differential drive robot, like the one used during the development: P3DX, fig. 4.12(b).

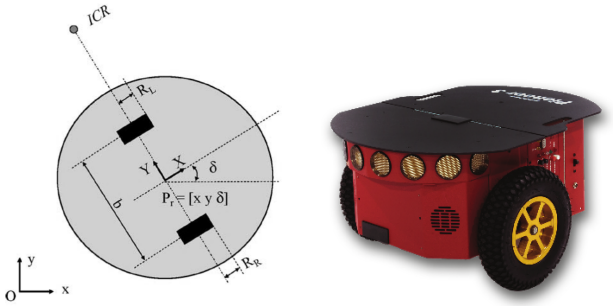
$$\begin{cases} x_{k+1} = x_k + \pi \cdot \frac{n_{Rk} R_R + n_{Lk} R_L}{n_0} \cdot cos(\delta_k) \\ y_{k+1} = y_k + \pi \cdot \frac{n_{Rk} R_R + n_{Lk} R_L}{n_0} \cdot sin(\delta_k) \\ \delta_{k+1} = \delta_k + 2\pi \cdot \frac{n_{Rk} R_R - n_{Lk} R_L}{b} \end{cases} \tag{4.1}$$

Uncertainty is expressed as the covariance matrix of the vector $X_k = [x_k, y_k, \delta_k]$.

4.1 must be rewritten by taking into account the pose increments noise

$$X_{k+1} = X_k + \Phi_{wk} + |\Phi_{wk}| \cdot \varepsilon_k \tag{4.2}$$

where Φ is a nonlinear function of $[n_{Rk}, n_{Lk}, n0, R_R, R_L, b]$ that calculates the position and attitude increments at each iteration step, n_0 is the number of the ticks of the



(a) Differential robot (b) Experimental robot: P3DX

Figure 4.12: Differential drive robot

encoders(constant). The vector ε_k is a stochastic variable that is related to kinematic model uncertainty, used to describe the error that affects pose increment. This error is proportional to increment modulus.

The variables that affect the accuracy of the estimation of the position and attitude are

$$w_k = [R_R, R_L, b, \delta_k] \tag{4.3}$$

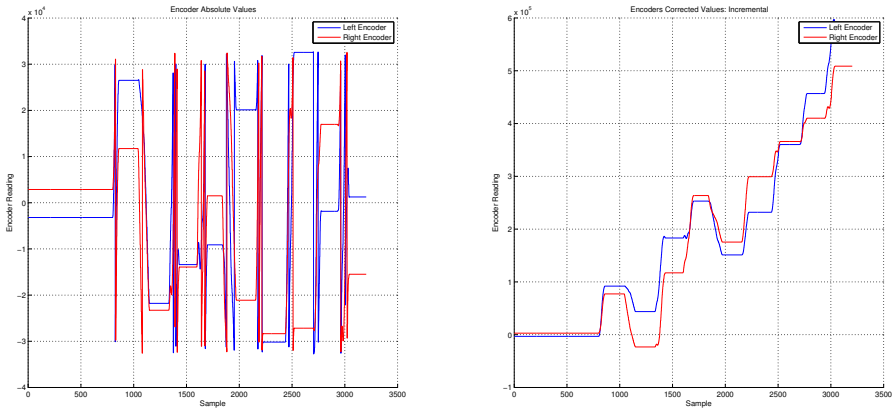
The covariance matrix of the position can then be calculated applying the following formulation

$$C_{X_{k+1}} = C_{X_k} + J_{\Phi_k} C_{wk} J_{\Phi_k}^T + J_{\Phi_k} S_{wk} I_k^T + I_k S_w k J_{\Phi_k}^T + |\Phi_k|^2 C_{\xi_k} \tag{4.4}$$

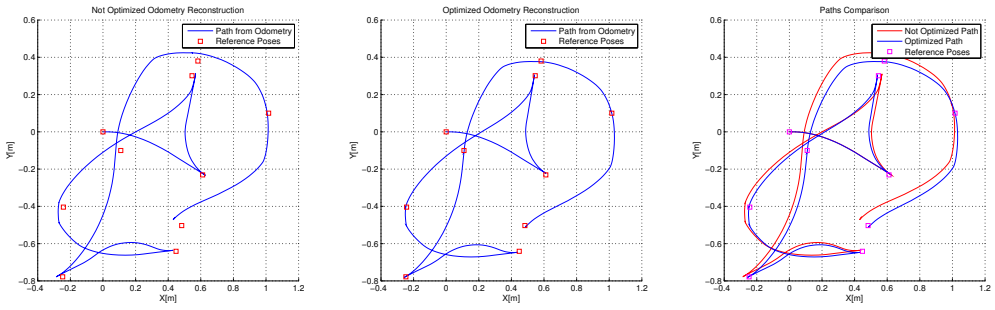
- J_{Φ_k} is the Jacobian matrix of Φ
- C_{w_k} is the diagonal matrix with the covariances of the parameters of w_k vector
- S_{w_k} is the matrix whose elements are the square root of C_{w_k} elements
- C_{ξ_k} is the diagonal matrix estimating uncorrelated model un certainty with covariances of pose increments.

This formulation is recursive and must be evaluated in each step after 4.1. Such equations are derived from the kinematic model of the robot, and must be updated if a different vehicle is used.

In fig. 4.13(b) is presented the reconstruction of the path done by the vehicle. There is a discrepancy between the odometric reconstruction of the trajectory (blue line) and the positions provided directly by the robot in the stops (red squares).



(a) Odometry



(b) Path reconstruction and optimization

Figure 4.13: Path performed by the vehicle

That is related to the kinematics parameters used by the robot controller (flashed on a chip): these differ from the ones used in the odometric reconstruction. However both the reconstructions suffer of the same drift error, the nominal parameters are different from the *real* ones, so both the sets of coordinates are not the ones actually

reached by the robot during the motion.

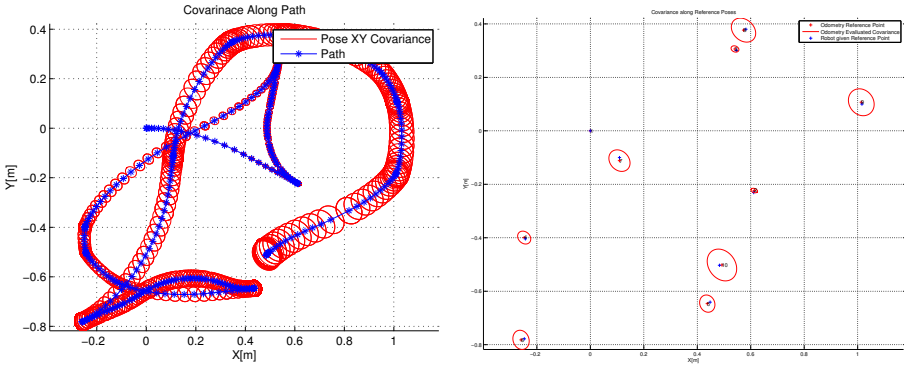
The knowledge of the geometric parameters is fundamental for the reconstruction of the trajectory, more they are accurate less will be the drift error, but they are unknown and they must be somehow evaluated. The nominal parameters from the datasheets and the technical specification are used as initial estimation.

The discrepancy between the positions is solved performing an optimization process on the nominal cinematic parameters, fitting the reconstructed trajectory on the positions provided by the robot. From the odometry are identified the samples associated to the stops of the robot. The minimization process, Levenberg–Marquardt, compute the optimal kinematic parameters that minimize the displacement in $[x, y, \delta]$ between these samples and the recorded positions. Fig. 4.13(a)(b) present the reconstruction and optimization of the kinematic parameters and the evaluated trajectory.

Table 4.4: Kinematic Parameters of the robot

	R_R, R_L [m]	b [m]	δ[deg]	n_0
Nominal	0.0977	0.333		76600
Optimized	0.095	0.330		
Uncertainty	$\pm$0.005	$\pm$0.005	$\pm$1	

The choice of optimizing the parameters, instead of the reference positions, comes from a consideration: the information that really matters in this phase is not the pose of the vehicle, but the associated uncertainty. Modeling the uncertainty allows to properly weigh the coordinates provided by the robot, minimizing the influence of the errors in the optimization. From the point of view of the final utilization of the procedure, it is also far more simple and less invasive to use the poses provided by the controller of the AGV, usually accurate, and then integrate the missing information about uncertainty. On the opposite, the modification the reference poses implies the evaluation of new the kinematic parameters, variables that must be somehow evaluated in order to reconstruct the path from the odometry. Any error in these values would produce an error in the reconstructed poses, so as for the controller of



(a) Evaluation of the uncertainly along the trajectory

(b) Input data used for the calibration, vehicle side

Figure 4.14: Path data

the AGV, adding in practice no further information to the process. Once optimized the kinematic parameters the uncertainty of the robot positions is computed form reconstructed trajectory.

Laser

The laser is not a proprioceptive device, it acquires indeed information only from the surrounding environment. The data provided, 2D maps, can be however used to estimate the motion of the sensor itself. Such operation falls in the topic of the automatic localization, one of the most studied and advanced research field of advanced (smart) robotic. This topic is still in full development, with an intensive research, and a considerable number of works that every year enrich the state of the art.Currently, diverse solutions are presented, different about strategies and finalities, with advantages but also drawbacks. There are essentially two types of automatic localization: one that uses a map of the environment as reference, and a second that uses just the data provided by the laser without any other a priori knowledge. The main difference between the two is that the first uses a reference that can be used to directly localize the robot, the second instead not. The map of the environment is built incrementally from the laser scansions acquired during the motion. That is however a chicken-egg problem: to organize the laser scansions and build a map the position of the robot must be known, the position of the robot is computed from the map of the environment. The techniques that deal with such problem are indeed called *Simultaneous Localization And Mapping*, SLAM. In order to keep the system as general as possible this second typology of localization is used. It is more complex but also less constrained, not requiring any priori knowledge of the environment.

For the objective of the calibration there is interest only in the localization of the sensor, functionality obtained with the software developed by Tavernini M. in his Ph.D. thesis, Tavernini (2013).

A synthetic representation of the process of auto-localization of the sensor is reported below.

As output the process provides the incremental reconstruction, globally optimized, of the map of the environment and the poses of the laser sensor, fig. 4.15. There are now 2 comparable chains that can be used in the optimization process.

Data: Laser Scansions

Incremental Localization:

for *all scansions* **do**

pairwise of two successive scansions, n & n+1;

Point to Point ICP, robust initialization;

Refinement:

Point to Line ICP, accurate matching of the scansions; → first raw estimation of the displacement of 2 successive positions of the laser

end

Incremental reconstruction of the path:

creation of a connection matrix;

Global Optimization: → graph theory

all the positions of the laser are optimized and weighted on the uncertainty of the matchings;

Result: Positions of the laser

Algorithm 3: Auto-localization of the laser by Tavernini

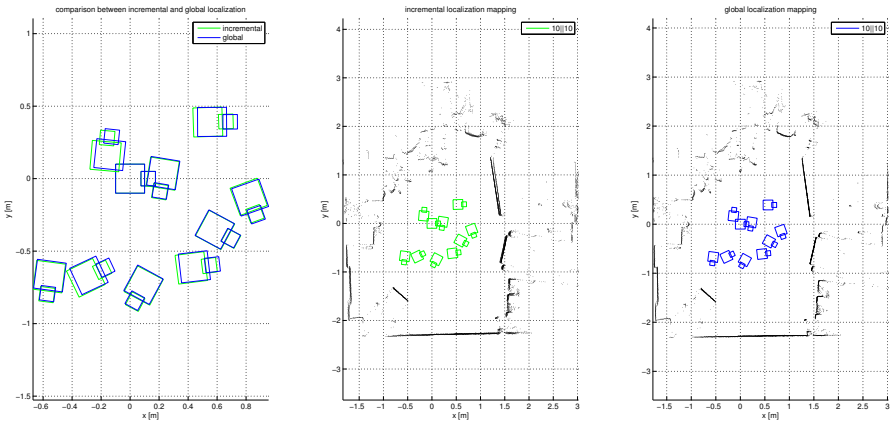
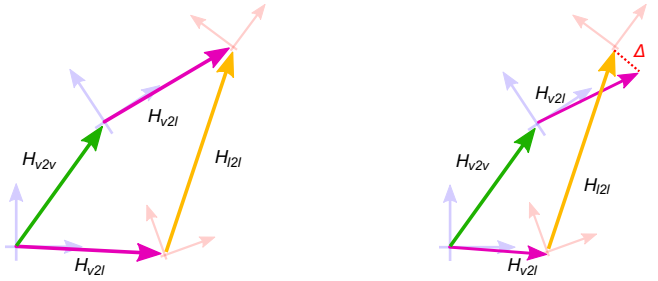


Figure 4.15: Laser poses

4.3.1 Optimization

Once evaluated the poses of laser and vehicle, the next step is the evaluation of the pose of the sensor on board by means of an optimization process. Assuming an ideal measurement system, in which the errors are limited or absent, the optimization problem can be formulated as following. The polygon is represented as two different paths made by two vectors each.



(a) Closure of the polygon (b) Error in the closure

Figure 4.16: Path data

The first path includes the laser position expressed in the vehicle reference system and two successive poses of the laser sensor.

$$Chain_1 = T_{v2l} + R_{v2l} \cdot T_{l2l} \tag{4.5}$$

The second includes the vector of two poses of the vehicle and then the vehicle-laser vector.

$$Chain_2 = T_{v2v} + R_{v2v} \cdot T_{v2l} \tag{4.6}$$

These two path have different endpoints because the vehicle-laser vector is unknown, so the polygon is opened, Δ, fig. 4.16. The simpler cost function that can be implemented involves the calculation of that opening. This function fills a vector of residuals of $n \times 2$ elements, n possible polygons, in which are stored the components of the vectors connecting the ending points of the chains, $\Delta[X, Y]$. A Levenberg-Marquardt optimization estimates the best values of $[x_{v2l}, y_{v2l}, \theta_{v2l}]$ of the vehicle-

laser transformation that minimize such residuals, closing the polygons.

$$Residual_i = \Delta[X, Y] = Chain_1 - Chain_2 = R_{v2l} \cdot T_{l2l} - R_{v2v} \cdot T_{v2l} + T_{v2l} - T_{v2v} \quad (4.7)$$

An improvement of the process includes also the alignment of the reference systems in the ending coordinates: the relative angle between the two systems can be used as third element in the vector of the residuals.

$$\begin{aligned} R_1 &= R_{v2l} R_{l2l} \rightarrow \theta_{Chain_1} \\ R_2 &= R_{v2v} R_{v2l} \rightarrow \theta_{Chain_2} \\ Residual_i &= [\Delta[X, Y], (\theta_{Chain_1} - \theta_{Chain_2})] \end{aligned} \quad (4.8)$$

As pointed out in the previous parts, the poses of laser and vehicle are data derived from conditions that are far away from the ideal ones. The error in pose estimation is not negligible, radically changing the shape of the polygons and so the convergence of the method. The cost function must be therefore modified including the uncertainty of the poses.

The uncertainty of the poses of laser and vehicle are different: the first is almost constant, the second not. The localization of the laser uses a global optimization from which follows a robust estimation of the laser transformations, in this case the amount and the type of the data provided by the sensor ensure a reliable estimation of the poses along the entire motion. On the opposite the path done by the vehicle is an information derived from an incremental measurement, in which the uncertainty can only increase. From that follows that the poses of the vehicle are accurate only in the first part of the trajectory, losing their influence on the calibration after each stop of the vehicle: if the uncertainty of a pose is elevate it is then probable that the polygons associated are deformed. For these reasons the uncertainty is used to weigh the data in the closure of the polygon, assigning more importance to the early acquisitions and less to the last ones.

The formulation of the residuals is then derived from Mahalanobis distance. To minimize the Mahalanobis distance means to reduce the difference in displacement and orientation of the distributions, increasing their overlapping. From such operation derives the closure of the polygon based not only on the coincidence of the vectors, but also the compatibility of the covariances.

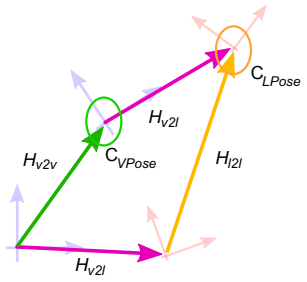


Figure 4.17: Laser poses

The updated formulation is the following. The pose of the ending coordinate of the two chains as homogeneous matrices:

$$H_{c1} = H_{v2v} \cdot H_{v2l} \qquad H_{c2} = H_{v2l} \cdot H_{l2l} \tag{4.9}$$

H_{v2l} is unknown and it is initially defined as 4×4 identity matrix. The displacement between the two chains can be calculated by multiplying the inverse of one with the other

$$\Delta H = H_{c1}^{-1} H_{c2} \tag{4.10}$$

From the matrix ΔH is built the displacement vectors X

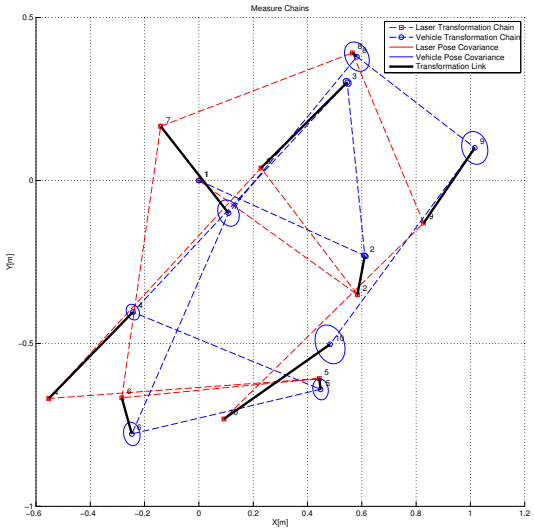
$$\begin{aligned} \Delta T &= \Delta H(1:2,3) \rightarrow [x, y] \\ \Delta R &= \Delta H(1:2,1:2) \rightarrow \theta \\ X &= [\Delta T, \Delta R] \end{aligned} \tag{4.11}$$

Such data is then weighted on the covariance of laser,C_{LPose}, and vehicle, C_{VPose}, used to build the *ith* polygon.

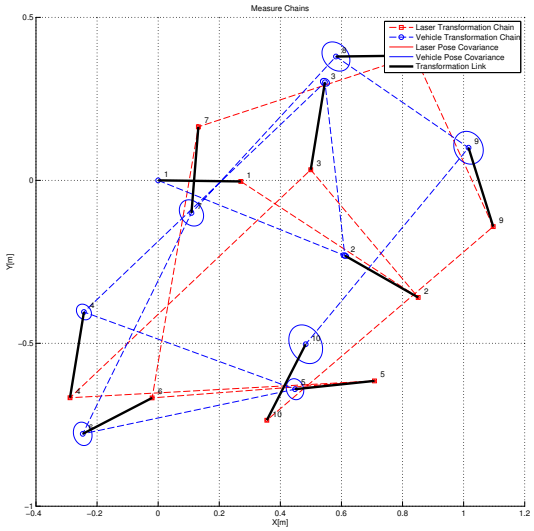
$$Residual_i = X_i'(C_{VPose} + C_{LPose})^{-1} X_i \tag{4.12}$$

In fig. 4.18 are shown the chains of laser and vehicle before and after the optimization, the segment in black represents the vehicle-laser transformation.

The use of the covariance only of the endpoints of the segments, laser and vehicle, comes from a consideration: the covariance of the different positions is expressed inside the absolute space of each sensor. Each point has therefore an uncertainty



(a) Before optimization



(b) After optimization

Figure 4.18: Positions of laser and vehicle

that is not related to the others (even more so if a global localization system is used). This element allows a simplification about how the uncertainties influence the process: starting point of the segments is considered certain and used as reference for the evaluation of the displacement vector. Include instead also the uncertainty of the

initial points means to relate the two covariances. Follow that the estimation of the displacement vector must depend on the uncertainty of the starting point: the measurement chain and the mutual influences of the covariances must then be modeled and evaluated, increasing the complexity of the problem. That involves the uncertainty propagation, obtaining even higher covariance matrices and weighting less and less the measurements of the vehicle, nullifying their influence in the calibration process. For this reason is kept the formulation with the covariance of only the endpoints of the segments, simplified structure that however have proved to be functional and enough accurate for the purposes of the application.

Result of the optimization is the position of the laser on board the AGV. Together with this data it is computed the covariance of the parameters $[x, y, \theta]$: as for the laser matching the calibration involves an optimization, so the formulation proposed by Censi(Censi (2007)) is used (Chapter 3, laser matching covariance). The results are reported in the next paragraph.

The performances of the entire procedure are strongly related to the accuracy in estimating the positions of laser and vehicle. Any increase of the drift error during the motion of the vehicle causes the loss of the information content. Each subsequent pose will have an higher uncertainty and so a lower weight in the optimization, meaning that the calibration will be based only on the first few samples acquired.

The best configuration is the one that has an uncertainty almost constant in all the positions, limited to the order of some centimeters. That implies an uniform weighting of the samples acquired during the entire motion and so a more robust optimization. Such configuration is the one that the AGVs offer thank to their global navigation systems.

4.3.2 Results

The laser-vehicle calibration has been coded and tested with an experimental robot within a generic environment, no priori information provided. In the following tables are reported the results of the calibration: position of the laser on board, $[x, y, \theta]$, and covariance.

Table 4.5: Vehicle to laser: transformation

x[m]	y[m]	θ[deg]
0.2707	-0.0036	-0.5555

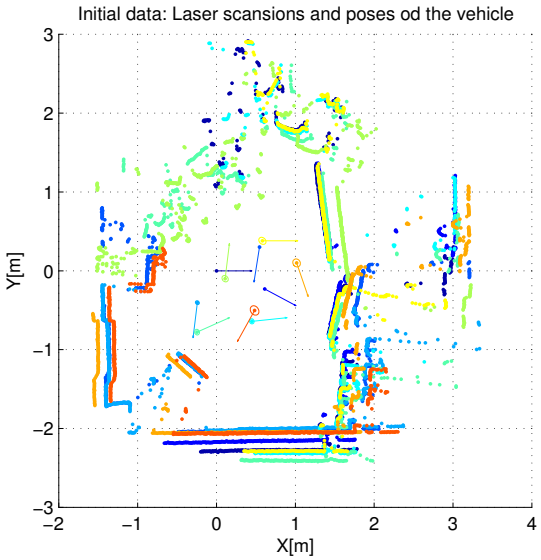
Table 4.6: Vehicle to laser: error

x_e[m]	y_e[m]	θ_e[deg]
0.0067	0.0079	0.6345

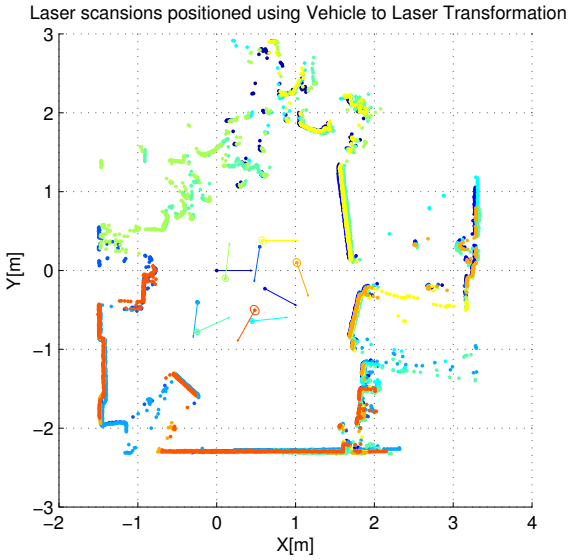
Table 4.7: Vehicle to laser: covariance matrix

		1.0E-3[m]
0.0446	-0.0242	-0.0072
-0.0242	0.0625	-0.0120
-0.0072	-0.0120	0.1226

To verify the quality of the results it is sufficient reconstruct the environment within which the calibration is performed. Fig. 4.19(a) presents such reconstruction by organizing the laser scansion according to the poses of the robot. The position of the scansions, however, depends also on the position of the sensor on board, unknown before the calibration. A null vector $[0, 0, 0]$ is used in this case.
The more accurate is the evaluation of the parameters $[x, y, \theta]$, the better is the reconstruction: the scansions match each others generating a continuous map of the environment. In Fig. 4.19(b) there is the same environment, this time reconstructed including the extrinsic parameters obtained from the calibration process.



(a) Initial Data



(b) Applying the trasformation between laser and vehicle

Figure 4.19: Result of the vehicle to laser calibration

Due to the lack of an accurate global navigation system (SICK NAV) the metric performances of this calibration results to be limited, especially for the angular displacement θ between vehicle and laser. The experimental robot can localize itself in two ways: by using the odometry (incremental localization), or using a second laser scanner together with a map of the environment (global localization). This second method should be more accurate, but it has a drawback: the position of the sensor on board the vehicle is unknown and such information it is necessary in order to correctly estimate the pose of the robot. It is a chicken-egg problem: to localize the robot must be known the pose of the sensor, to calculate the pose of the sensors must be known motion of the robot. The calibration should therefore be performed two times: a first time to estimate the pose of the laser dedicated to the localization and then again for the frontal laser using this time the localization provided by the previous sensor. Such configuration was however skipped because it doesn't add further improvements to the process itself: the industrial AGVs are already equipped with navigation systems, which provide the global localization with a very high accuracy. Such devices are calibrated and integrated in the machine by the manufacturers of the AGV, providing the continuous localization of the robot, data that can be directly used in the proposed calibration. With such configuration the calibration process will be therefore both simpler (the position is provided and so the odometric reconstruction can be skipped) and more accurate.

It must however underlined that even with the configuration presented, using the incremental localization, the accuracy of the calibration was sufficiently accurate, achieving the correct picking of the pallet with the experimental robot, chapter 6.

CHAPTER 5

SENSOR FUSION

Two sensors are way better than a single one. The two data sources can be combined taking the best from each device, achieving as main benefit a more robust and reliable knowledge of the environment.

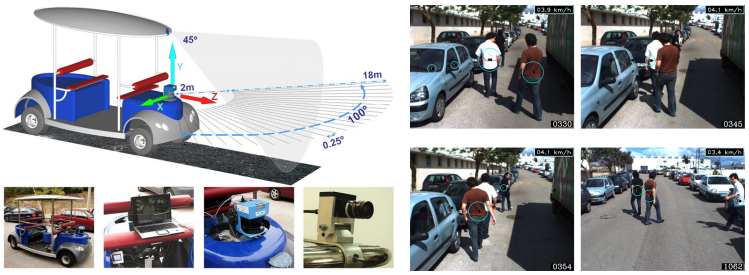
5.1 Introduction

MULTI-SENSOR strategy is a common solution when the operative condition requires very reliable results in order to properly accomplish the assigned task. The current application is one example of such conditions: the identification of the pallet is achieved by different sensors, increasing the reliability of the overall process. These information must be somehow combined in a single solution.

In this chapter is described how the identifications are fused, the methodology developed is general and usable in any system that includes a couple, or more, of Cartesian and perspective sensors. In the case of a TOF camera the process remains same, resulting even more simplified by the coexistence of both the typology of sensors in a single device.

In literature can be found several works that speak of the sensor fusion of multi-sensors systems. One of the most studied topic in which is commonly involved the laser-camera setup is the *pedestrian recognition*. Such task is fundamental for co-operative robotics, with application for the indoor, Jinshi et al. (2008), or outdoor, especially for automotive, GIDEL et al. (2008, 2009); Oliveira et al. (2010).

The objective is to improve the perceptive performances of the autonomous vehicles: in order to increase the safety of such machines it is critical to identify what is an obstacle, static in the surrounding environment, and what is potentially capable to move, a dynamic object, like people. In many of these works a classifier is used to identify the object of interest, a similar strategy compared to routine developed for the camera. The results associated to such strategy seem convincing, adding credit to the choices taken during the current development.



(a) Vehicle and laser-camera system (b) Pedestrian recognition

Figure 5.1: Oliveira et al.

Gidel et al. (2009) is an interesting paper in which are listed the possible strategies of fusion that can be implemented in such problems.

Four main architectures can be identified in the literature:

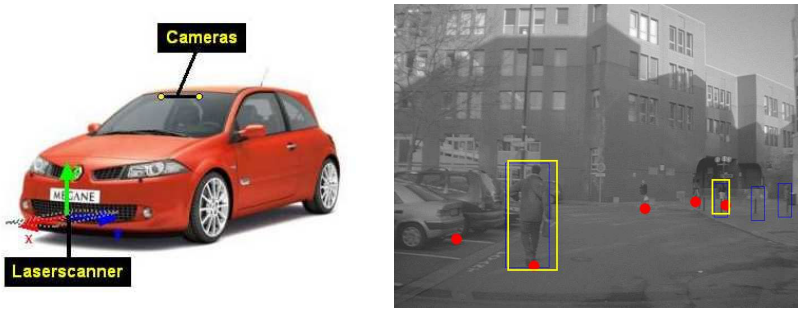
serial fusion : the laser scanner segments the scene and then provides some ROIs (Regions Of Interest), which are confirmed to match pedestrians by means of a vision based classifier, Szarvas et al. (2006)

centralized fusion : the measurements from the various sensors are merged (associated and tracked) in a same central block, Linzmeier et al. (2005)

decentralized fusion : each sensor system detects, classifies, identifies and tracks the potential pedestrians before being merged in a track-to-track fusion block, Blanc et al. (2005)

hybrid fusion : available information includes both unprocessed data from one sensor and processed data from the other one, Monteiro et al. (2006)

Gidel chooses for his application the centralized fusion architecture.



(a) Car (b) Pedestrian recognition

Figure 5.2: Gidel et al.

For the purposes of the current application, in which is achieved the independent identification of the pallet by both laser and camera, the centralized fusion seems to be the best choice as well.

In almost none of the papers listed there is any analysis of the uncertainty. Usually the data involved is processed without considering any possible influence on the results of the errors in the nominal parameters. Nygards and Wernersson (1998) presents an analysis on the covariance of a 3D laser scanner and a camera, this is one of the few examples in which there is a more structured approach of the problem of fusion.

The development and the considerations presented till here have shown how the analysis of uncertainty in the algorithms is a key element to provide a first check on the correctness of the identifications. For this reason a modified version of the centralized fusion has been developed including the influence of uncertainties.

5.2 Elaboration strategies

The configuration of the measurement system is the following: two sensors (or one that incorporates them both) able to provide independent identifications of pallets. Each identification is expressed within the reference system of the device. The extrinsic parameters are known from calibration. The data must be fused in order to provide an unique result. The main difficulty in achieving such operation comes from the different typology of the sensors, one Cartesian one not. That implies the lack of an homogeneous, and directly comparable, information content. The identifications of the pallet are indeed expressed as a position in space $[x, y, \theta]$, from the laser scansion, and pixel coordinates $[i, j]$, from the image. A notation or a reference system must be identified in order to express the data in an homogeneous form/space that can be used as base for the fusion.

An efficient data fusion process for the current application can substantially be attained with three possible modalities:

- fusing the data inside a 3D space(laser)
- fusing the data inside a 2D space(image)
- define a transformation that projects both the data to a completely new space

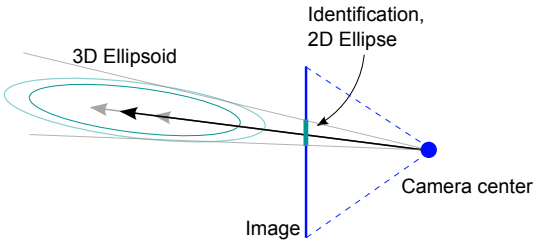
For the typology of the sensors and for the objective of the work, the third option is skipped: the use of a space that is not easy to interpret implies that the results of the fusion will be more difficult to monitor in the occurrence of errors or critical cases, with a consequential complex management of the device.

3D fusion

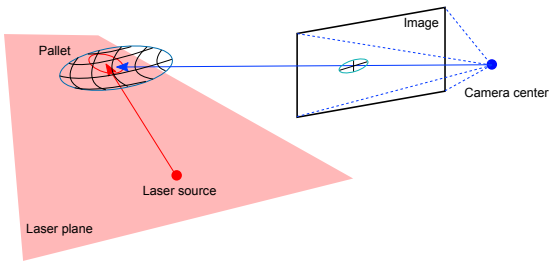
The fusion in the 3D space uses the laser as reference sensor to which associate the identifications obtained from the camera.

Although the laser scansion is planar, it is necessary to move on a three-dimensional space in order to express the information of the camera inside such plane. The 2D identifications describe the position of the silhouette of the pallet inside the image, this must be transformed into a position in space in order to be comparable with the Cartesian one from the laser.

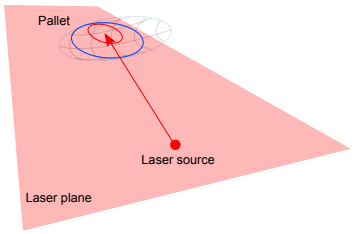
Admitting of successfully perform such transformation, the 2D ellipse associated to the identification must be defined as 3D ellipsoid to include the uncertainty from the evaluation of the pose: a depth estimation, fig. 5.3(a). If the extrinsic parameters are known, the intersection between this 3D ellipsoid and the laser plane can be calculated, blue ellipse. That generates a new 2D ellipse which can be compared with the measurement obtained from the laser, ref ellipse, fig. 5.3(b)(c).



(a) 3D depth projection



(b) Camera identification to 3D space



(c) Intersection of the 3D ellipsoid and the laser plane

Figure 5.3: Camera to world projection

The critic step in this approach is however the transformation itself from the 2D distribution of the image to a 3D one, the estimation of the position of the pallet from the camera. The reconstruction of the 3D information lost when capturing an image is indeed a topic on which an active research is still ongoing. The solution of this problem implies the recognition of the geometrical elements/features of the object (known from a model) inside the image. Given the projective pinhole model of the camera and the intrinsic parameters associated, the geometries identified in the images can be used to evaluate the position of the object along the optical rays of the camera, Zisserman and Hartley (2000). The more accurate is the identification of the geometry inside the image the better is the approximation of the position of the object in 3D space.

As shown in chapter 3, the HOG identifies a region, a rectangle, in which is contained the face of the pallet. The accuracy of this result is neither adequate nor sufficient to obtain an useful approximation of the 3D pose of the pallet in terms of position $[x, y]$ and attitude θ. That can be obtained only by performing a specific analysis on the deformation of the face of the pallet due to its pose (variation of the size) and the perspective distortion (variation of the silhouette of the face), Sungmin and Minhwan (2008). As underlined in chapter 2, such operation is strongly dependent on the accuracy of the gradients calculated from the image, operation influenced by the light condition, materials, distance of the pallet from the camera etc. The accuracy and the repeatability of the results of such algorithm are typically very poor, mainly due to the lever effect between the errors in the identified geometries of the object and its depth estimation. Even in stable conditions the estimated position can vary of centimeters, one order of magnitude greater compared to the laser. This makes the influence of the camera negligible.

In order to have a robust and reliable fusion strategy the data should have comparable uncertainties, giving in this way an equal weight on the different sources of information involved. For the reasons underlined, this approach results to be not suitable nor advantageous for the purpose of the fusion.

2D fusion

In every transformation of data there is a loss in accuracy. That derives from the uncertainty associated to the parameters involved, which cause errors in the estimation of the results. The 3D fusion involves the reconstruction of 3D Cartesian data from perspective ones, the limitation of the method comes from that operation. Between the two typologies of identification the less accurate is the one from the camera and the transformation of such identification into the laser reference system adds even more uncertainty in the results. In this way the information of the laser can not be efficiently compared to the one provided by the camera.

The implementation of a strategy in which the most accurate information, from laser, is moved to the less one, camera, generates instead two comparable distributions in terms of uncertainty. The laser identification must be projected on the image, resulting in a 2D ellipse directly comparable with the one obtained from the image processing, fig. 5.4.

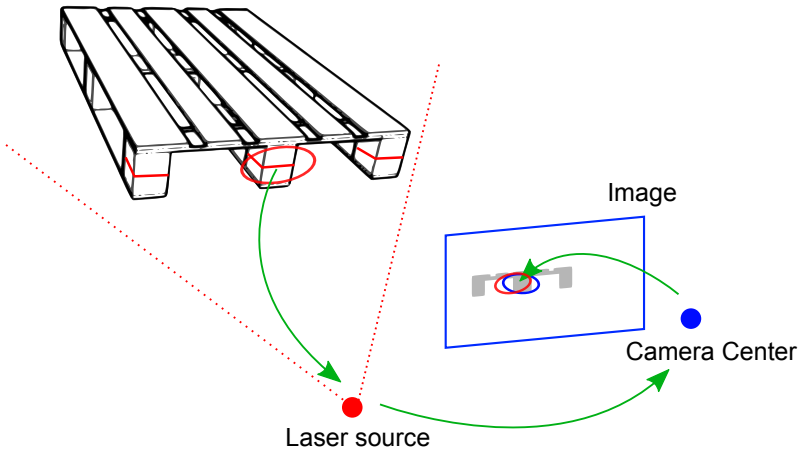


Figure 5.4: Laser to camera projection

Compared to the 3D one, it is clear how this fusion process is more finalized to the verification of the results in terms of compatibility of the identification rather than a combination of Cartesian data.

This aspect would be critic if the laser is not sufficiently accurate in determining the position of the pallet, requiring further measurements to refine such estimation. The performances of the sensor are however sufficiently elevate to ensure the required accuracy, millimeters and tenths of degrees, allowing the development of a fusion process aimed to determine if the identifications provided by the sensors are generated by false positives cases or not.

For the above reasons this 2D strategy is chosen.

5.2.0.1 AGILE

AGILE, Baglivo et al. (2008, 2009, 2011), was the starting platform during which different approaches were tested in order to combine the data from laser and camera. The resulting strategy is the one shown in the diagram at fig. 5.5. The processes of identification is totally based on the laser, structure due to the limited performances obtained from the camera (chapters 2 and 3).

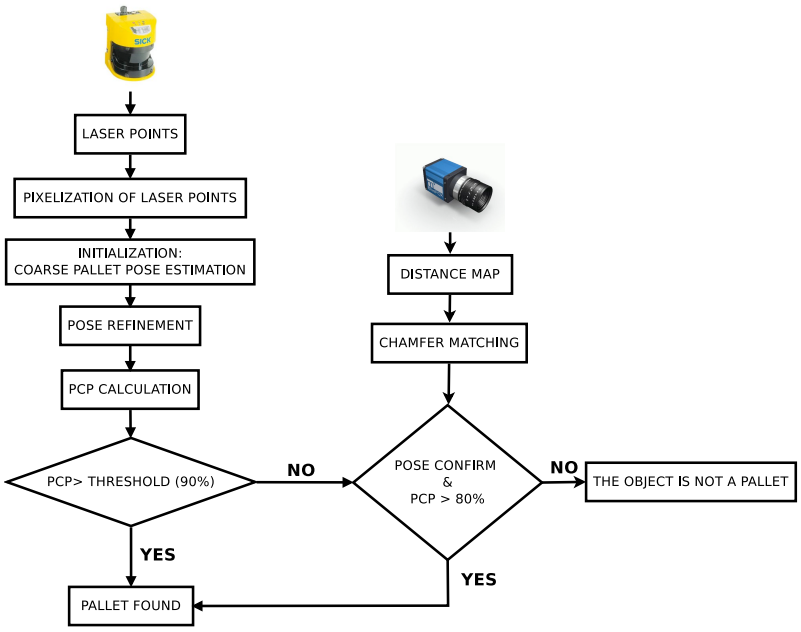


Figure 5.5: AGILE, identification structure

As more exhaustively described in the thesis of Biasi, Biasi (2010), and the paper Baglivo et al. (2011), the pallet identification is performed in two steps: in the first only the laser scansion is processed, in the second the output identifications are analyzed verifying if they are *reliable* or not. Only in this second part, and only if required, the camera is used.

In the diagram there is a strong asymmetry in the management of sensors, assigning in practice all the computational load to the laser. The identification algorithm provides as output the pose of the pallet, $[x, y, \theta]$, together with a threshold value based on the performances of the matching between laser data and model: a simu-

lated scan of the identified pallet is simulated and each obtained point is associated with its nearest from the real input scansion, PCP stands for *percentage of coupled points*. When the PCP value is less than 80% the camera processing is activated.

The camera algorithm is virtually independent from the laser, but in practice is requires an initialization (given by the laser) in order to shrink the area to be processed, and to control the dispersion of the results. Once processed the defined region of interest, it is performed a check on the *distance* between the identifications obtained from the image processing and the reprojection of the laser ones on the image, this step has been defined as a *camera consensus*. From it follows if the identification of the laser is considered valid or not.

The camera is therefore used as a support for the laser in the cases of poor matching. Such process is not properly a fusion but instead a comparison. No analysis of the uncertainty is performed.

The asymmetry of the process represents the main constrain to the fusion strategy used in AGILE. The new algorithm of image processing allows now the identification of the pallet independently from the laser and the possible development of a new method that can take advantage of both sensors, fig. 5.6.

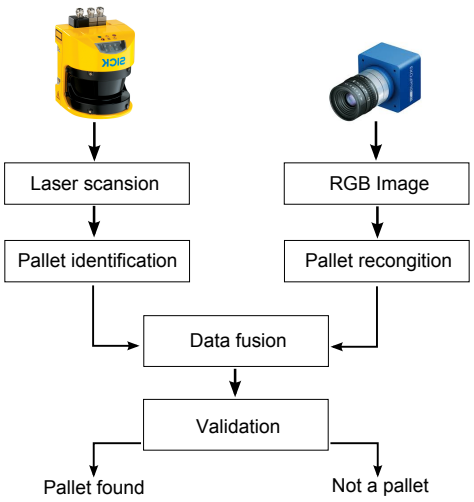


Figure 5.6: Fusion diagram: independent input

The main benefit of having two independent identification algorithms is the opportunity to delay any mutual interaction between laser and camera until the data fusion. Each identificator is capable of: acquire data, process them, analyze independently if the results are valid or not (control based on covariance and thresholds). This structure ensures more reliable data as input for the fusion process, rejecting most of the possible false positives that could influence the results.

The fusion process can be seen as a further *filter* of the results in which the input identifications are mutually checked. With this structure the only way in which a false positive can occur is that both the identifications pass the first, independent, verification of the respective identification processes and then the conclusive compatibility check of the fusion. Practically speaking, it means that both the laser and the camera identify, in the same way and in the same place, an object that is not a pallet as such.

```
acquire SCAN;
pixelize SCAN;
find LINES;
while LINES do
    create rotated model;
    convolve model on pixelized scan;
    if model match successful then
        return palletpose(i) {fill candidate poses set}
    end
end
if palletpose is not empty then
    select best pallet {user–defined criterion};
    X_pallet ← best pallet pose;
    find all points belonging to X_pallet;
else
    goto → acquire SCAN;
end
refine X_pallet {apply local minimization};
compute PCP: ;
if PCP > 90% then
    Solution found: X_pallet;
else
    if PCP > 80% then
        camera consensus;
        if camera consensus=true then
            Solution found: X_pallet;
        else
            X_pallet rejected;
        end
    else
        X_pallet rejected;
    end
end
goto → acquire SCAN;
```

Algorithm 4: Fusion procedure used in AGILE

5.3 Laser-camera data fusion

The fusion uses the informations obtained by the laser-camera calibration to project the laser identifications on the image plane. This is achieved in two steps.

Initially the laser data are transformed and expressed inside the 3D reference system of the camera. This operation involves the extrinsic parameters between the two devices:

$[x_{LC}, y_{LC}, z_{LC}, \alpha_{LC}, \beta_{LC}, \gamma_{LC}] \rightarrow H_{extrisic}$

$$P_{3D_{camera}} = H_{extrisic} \cdot P_{3D_{laser}} \tag{5.1}$$

Despite the laser data are planar,$[x, y]$, the transformation must be computed in 3D due to the relative orientation and displacement of the reference systems. The z coordinate of laser data is set to zero, reference height of the entire device.

Once computed the position of a laser point in the 3D space of the camera, the projective behavior of the sensor is applied through the pin-hole model, fig. 5.7, this transforms the 3D point P into 2D image coordinates p. In this step the parameters involved are the intrinsic ones: focal, camera center, distortions of the lens/optics.

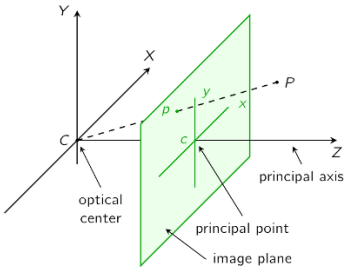


Figure 5.7: Pinhole model

In most softwares the algorithm simply projects the laser points into the image, skipping any uncertainty analysis. This is commonly due to the limited influence of such information on the process and its results. A typical example is the one in which the laser data must be associate to a color in order to texture a 3D scansion: it is sufficient to consider a region of a few pixels around the projections to identify a good approximation of the right color, any representation of the uncertainty in that case is

not necessary. In the current application the knowledge of such information is instead critical to refine the process of data fusion.

RADLOCC, chapter 4, includes the uncertainty analysis of both intrinsic and extrinsic parameters, evaluating not just the reprojected points on the image but also the ellipses of covariance, fig. 4.5(d). This result is obtained through the covariance propagation $C^* = JCJ'$, in which J is the Jacobian of the transformation used to project the data on the image and C the covariance matrix of the parameters.

Given a 3D point $P_{3Dcamera}$, $[x, y, z]$, the formulation to be applied in order to project the point on the image is derived from the pin-hole camera model, Zisserman and Hartley (2000). The point is initially projected on a plane orthogonal to z, frontal direction of the camera, distant 1 (the unit depends on the definition of the focal) from the camera center.

$$a = \frac{x}{z} \qquad\qquad b = \frac{y}{z} \tag{5.2}$$

From the 2D coordinates $[a; b]$ is computed the distance of the projection from the center of the plane (intersection with z):

$$r = \sqrt{a^2 + b^2} \tag{5.3}$$

r is used to evaluate and apply the radial distortion of the image due to the optic (k_5 is usually set to 0):

$$\begin{aligned} a_d &= a(1 + k_1 r^2 + k_2 r^4 + k_5 r^6) + 2k_3 ab + k_4(r^2 + 2a^2) \\ b_d &= b(1 + k_1 r^2 + k_2 r^4 + k_5 r^6) + k_3(r^2 + 2b^2) + 2k_4 ab \end{aligned} \tag{5.4}$$

Once computed the *distorted* 2D coordinates $[a_d; b_d]$, the final step evaluates the coordinates (in pixels) of the point on the image by including the focal length, the camera center and the geometry of the pixel (α):

$$\begin{aligned} x_{px} &= f_x(a_d + \alpha b_d) + c_x \\ y_{px} &= f_y b_d + c_y \end{aligned} \tag{5.5}$$

Given the projective formulation, the covariance propagation is achieved by differentiating x_{px} and y_{px}; the input covariance matrix C is a diagonal matrix that includes

the σ^2 of the parameters of the process (uncorrelated): focal length, camera center, distortion parameters, α. It is important to remark that the starting 3D point is the result of a previous 3D transformation between laser and camera, 5.1, so other 6 parameters must be included in the evaluation: 3 angles and a 3D vector, $[x_{LC}, y_{LC}, z_{LC}, \alpha_{LC}, \beta_{LC}, \gamma_{LC}]$.

The illustrated structure presents however a limitation: the input points are considered without uncertainty, simplifying in practice the measurement model and so the Jacobian matrix. For this reason, the formulation was modified. The result of the laser processing is the position of the pallet $[x_{plt}, y_{plt}, z_{plt}]$ with the associated uncertainty. The information related ot the uncertainty must be included in the reprojection on the image by expanding both the Jacobian and the covariance matrix. Since the camera does not provide information about the attitude of the object, the contribution of θ is not considered in this operation.

A further important element is not included in the standard notation: the laser beam is a cone and therefore it is not possible to define a priori the exact height at which the beam impacts with the objects, and so the position of the laser point on the face of the pallet. This can be included in the formulation by modeling an uncertainty in the height z, function of the distance d of the identified object and the cone opening δ_{LB} (Laser Beam) from the specification of the laser provided by the manufacturer: $error_z = d \cdot tan(\delta_{LB}/2)$.

The formulation is then modified including both the aforesaid elements

$$C^* = J_2 C_{LI} J_2' \tag{5.6}$$

J_2 : is the new Jacobian matrix in which are included the partial derivative of the laser identification. This is a 2×19 matrix:

$\partial[x_{LC}, y_{LC}, z_{LC}, \alpha_{LC}, \beta_{LC}, \gamma_{LC}, f_x, f_y, cc_x, cc_y, k_1, k_2, k_3, k_4, k_5, \alpha, x_{plt}, y_{plt}, z_{plt}]$

The focal, f, and the camera center, cc, are pairs of parameters, one for each principal direction of the image: horizontal and vertical.

C_{LI} : the diagonal covariance matrix of σ from the standard notation, C, plus the

covariance matrix of the laser (Laser Identification), C_{li}

$$C_{liU} = \begin{bmatrix} C & 0 \\ 0 & C_{li} \end{bmatrix} \tag{5.7}$$

C_{li} : 3×3 diagonal covariance matrix from laser identification,

$$[x_{plt}, y_{plt}, z_{plt}] \rightarrow C_{li} = \begin{bmatrix} C_{xy} & 0 \\ 0 & error_z^2 \end{bmatrix} \tag{5.8}$$

C^* is the output covariance matrix expressed in pixel that describes how uncertain is the projection if the laser identification on the image. This ellipse, the red one in fig. 5.8, can now be fused with the one obtained from the image processing, the blue one.

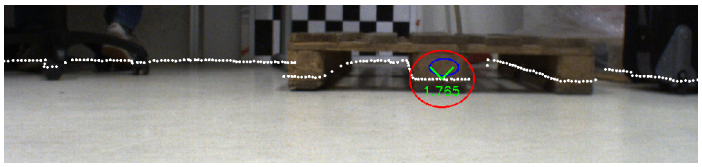
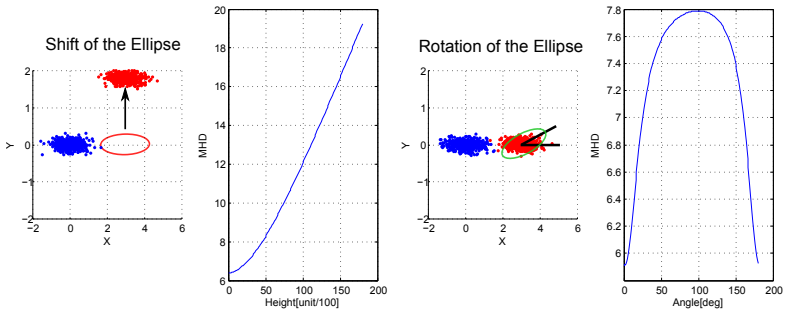


Figure 5.8: Laser projection on the image

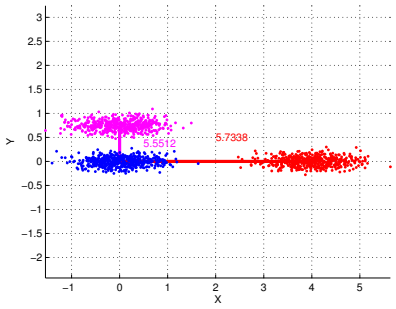
The fusion takes advantage of the knowledge of the covariances of the results. In order to verify if the two identifications, and distributions, are *compatible* it is used the Mahalanobis distance, Mahalanobis (1936), similarly as done the 2D laser matching.

$$MHD = \sqrt{(\mu_{laser} - \mu_{camera}) \left(C_{laser} + C_{camera}\right)^{-1} (\mu_{laser} - \mu_{camera})'} \tag{5.9}$$

Considering two aligned distributions at a distance l, fig. 5.9(a)(b), it can be proved that increasing both the relative angle and/or the relative displacement, the value of such distance increases. The greater is the eccentricity of the ellipses, the difference of the eigenvalues, the more exalted this effect is. If two distributions are instead overlapped the Mahalanobis distance is low, the fusion is indeed based on that: a couple of *correct* identifications of the pallet are located in the same spot on the image, close to each other; this configuration is therefore easily detectable by simply analyzing the distances between the distributions on the image and selecting the minimum ones.



(a) Shift (b) Rotation



(c) Similar values of distance

Figure 5.9: Mahalanobis distance, relative displacements

A possible drawback is instead the occurrence of a low value of distance from the alignment of the distributions along the maximum uncertainty direction, fig. 5.9(c). This second event is however prevented by the rejection of the false positives achieved directly in the identification processes. From that follows that the identifications can be aligned, but with a relative displacement on the image at least equal to the width of the face of the pallet (depending on its distance and position), distance from which derives a value not comparable to the configuration in which the ellipses are overlapped.

A further element of strength of this choice comes from the perspective effect associated both to the camera identification and to the laser projection. The distance of the object from the camera is connected to the dimensions of the ellipses: the more far is the object, smaller are the ellipses. This however has a in practice no influence on such formulation of the distance because the relative displacement of the (correct)

distributions becomes smaller as well, keeping the associated distance value almost constant. The main benefit is then a robust parameter invariant to the perspective, with the possibility to define a single value/threshold usable for all the possible configuration without further analysis on the position of the object.

Given this definition of distance, **two distribution are considered compatible, and so referring to a real pallet, if the distance is lower than 3**. Such threshold was defined from the experimental evidences and the tests run under different operative conditions: no false positives occurred since its definition.

If more identifications reach the fusion stage, a brute force elaboration compares all the possible combinations. If a distribution is evaluated compatible with more that one another (laser to camera and vice versa) all the associated identifications are rejected.

5.3.1 Optimization

The procedure described so far is functional, but a minor problem was highlighted during the development. Running control cases in which the operative conditions were planned in order to analyze the ellipses on the images, it has been noticed that the distributions of laser and camera were never concentric, but instead always partially overlapped and shifted, fig. 5.10. Such effect it is critical for the fusion: the value of distance associated to two correct, but shifted, ellipses is usually higher than 1.5. This influences the choice of the threshold on distance, requiring the definition of higher values, weakening the rejection of false positive cases. Such issue is due to the lack of any *physical* connection or relation between the laser-camera identifications and their position on the face of the pallet.

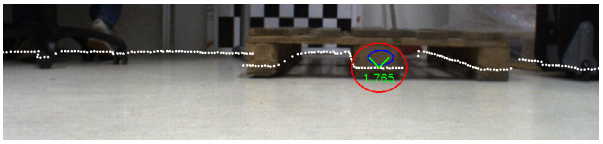


Figure 5.10: Shift of the ellipses

The projection of the laser data on the image uses the extrinsic parameters, which depend on how the devices are assembled. The position of the laser ellipse *on the image* depends therefore on a geometrical configuration of the system, which is modeled with the laser-camera calibration. The position of the ellipse *on the face of the pallet* is instead related to the height from the ground of the scanning plane. Two different relations can then be identified: *sensor to sensor*, that it is responsible to the data transfer from laser to camera, and a *sensor to environment*, from which depends the position of the laser ellipse over the object.

The identification of the camera comes instead from a different and independent process: the identification is placed at **the coordinates associated to the highest score obtained by applying the convolution of the pyramidal filters of the model (created with a training) over the image**. **The result is dependent on the model used, without a direct correlation with the geometry of the object**. For this reason, this identification can be interpreted just as the best in terms of **response to the**

model.

There is no real connection between the information obtained from laser and camera: the ellipses represent only two different results that share an homogeneous representation.

From this follows that a modification of the identifications doesn't invalidate the fusion process, this can be instead used to improve the performances of the process allowing more a efficient rejection of the false positive cases.

The identifications that are transformed are the one less constrained to configuration of the system, the ones from the camera. The laser identifications are strongly connected to the geometry of the device, any modification would cause the interruption of the *measurement chain*, influencing the meaning of the representation of the results. On the opposite the camera identification is not bound to any geometry and it is not involved in any measurement chain. The modification of the results in this case can be seen a additional (conclusive) step of the identification process.

Such operation is achieved considering the two elements that constitute the camera identification: the center of the ellipse on the image and the associated boundary box. The two are connected: both the boundary box and the position of the maximum score, isolated inside that region, come from the model trained (convolution of the pyramidal levels). This can be used to apply a corrective transformation on the coordinates of the ellipse and shift the identification to the center of the face of the pallet, superposing it with the one from the laser, fig. 5.11.

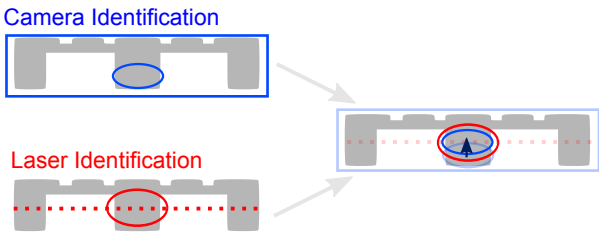


Figure 5.11: Pre-fusion optimization

Any action has to undergo to an implicit condition: its dependence on the laser-

camera calibration. Since the reprojection of the laser on the image depends on the extrinsic parameters, and so its position on the face of the pallet, the corrective action applied on the image must be related to such parameters. That can be achieved by structuring the setup of the device: after the laser-camera calibration, a further optimization routine must be run in order to minimize the shift of the ellipses and tune the corrective *pre-fusion* action.

A linear transformation is used as corrective action. The optimization procedure takes as input a set o control images and laser scansions, minimizing the output values of the fusion routine by evaluating two independent shift (no rotations are considered): one vertical and one horizontal, linearly proportional to the size of the boundary box associated to the identification.

$$C_{Optim} = \begin{bmatrix} k_w Width_{BBox} + C_x \\ k_h Height_{BBox} + C_y \end{bmatrix} \tag{5.10}$$

The results for configuration used during the development are:

- Vertical shift $\rightarrow$ **12%** Bounding Box Height
- Horizontal shift $\rightarrow$ **5%** Bounding Box Width

As stated earlier, these values depends on the relative displacement of the sensors and on the model trained. Any modification in the setup (or target object) requires a new calibration and a new optimization of the parameters.

Fig. 5.12 presents an example of the fusion with (a) and without (b) the application of the corrective action on the camera identification.

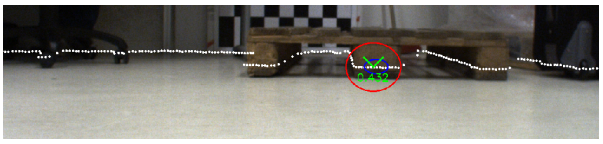


Figure 5.12: Improvement of the fusion after the optimization

As can be seen in the second image, the position of the camera identification, blue, is superposed and centered both on the laser ellipse, red, both on the laser points, white.

5.3.2 Results

The fusion process described achieved excellent operative performances, with zero false positives during the entire test phase. The elaboration is stable and ensures a proper fusion with the admissible pallet poses: $[1.5m \rightarrow 4m; \pm 15^{\circ}]$.

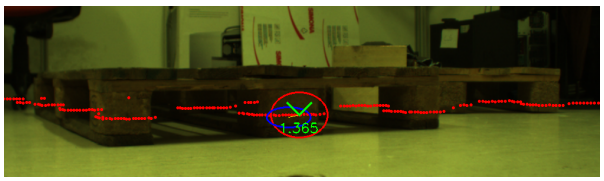


Figure 5.13: Example of fusion results

In those cases in which the attitude of the pallet is higher that 15° the fusion could fail: the deformation of the frontal face causes a shift in the HOG distribution inside the box (deformation of the model), the laser identification does not suffer of the same effect. This shift produces higher output values from the fusion, potentially out of thresholds, with the consequential discard of the identifications.

A solution to such effect could be the definition of less restrictive thresholds, increasing the admissible distance between the distributions. Remembering however one of the key aspects of this work, *better a missed identification than a not correct one*, such choice becomes not optimal because it lowers the reliability of the process, weakening the entire system. The loss of some, *limit*, identifications does not motivate such risk.

5.3.3 Critic Cases

Some critical situations can not however be solved with this system. The laser-camera device is designed to achieve the identification of the pallet positioned within the scanned plane. The structure is functional for the cost reduction that derives from using the sensors already mounted on board the vehicle (safety laser). That however implies that any interfering element placed outside such scanning plane can not be identified.

An example in which such configuration finds its limit is visible in fig. 5.14. Two pallets: one positioned on the floor, and another partially superimposed on the previous one. Between the two, only the one on the floor can be, and is, recognized.

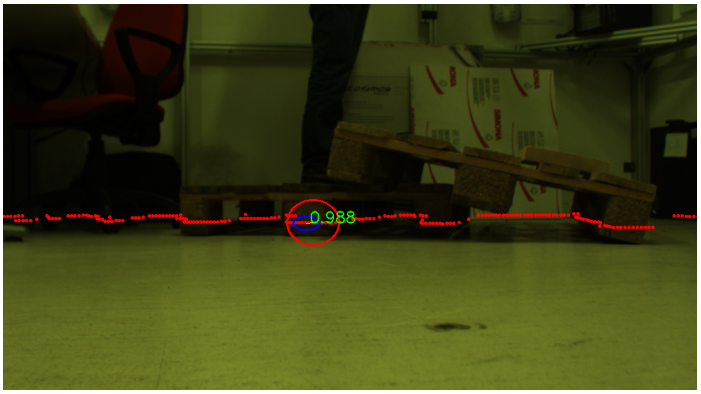


Figure 5.14: Critic case

From the practical point of view of the task, it is clear how this situation is dangerous in terms of the safety for the vehicle and for load of the pallets. Forking and lifting the identified pallet would cause the capsizing of the second, with consequences that can not be controlled. It must be pointed out that the situation illustrated can be considered borderline. Inside of an automatic warehouse it is rare the occurrence of such configuration, and even more rare that a loaded pallet it is superimposed to another.

It is however important to underline and show the limitations of the system in order to understand where to improve this device and focus the future research.

5.4 Trigger

A critic element it is the asynchronism of the sensors during the acquisition.

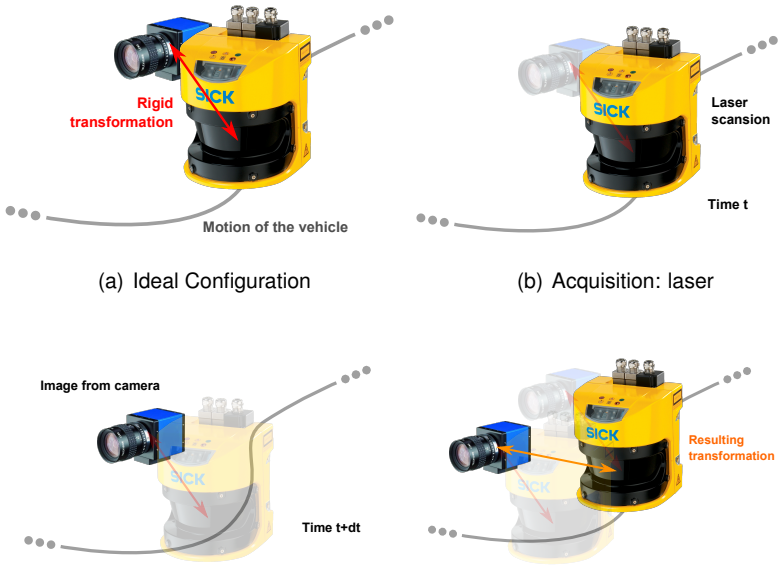
The proposed structure has a meaning only if the relative displacement of the sensors is fixed and known, information evaluated by means of the laser-camera calibration, chapter 4. However, it can also be related to the motion of the vehicle: even if the sensors are fastened and calibrated, it is still possible to have an error due to an *unknown relative displacement* of the devices.

If the vehicle is moving, an the sensor are not synchronous, any delay in the acquisitions causes a variation of the relative pose of the sensors, fig. 5.15. If the laser is acquired at the time t and the camera at the time $t + \delta t$, a displacement error related to the space traveled by the AGV during the dt interval occurs. The faster the vehicle moves the higher the error is: $e = v \cdot \delta t$ (approximatively).

From that derives an improper fusion: the projection of the laser identifications on the image is performed with nominal (calibrated) extrinsic parameters which differ from real geometric configuration of the acquisition. The fusion of the ellipses could be incongruent, with the possible rejection of correct identifications and at the same time the generation of false positive cases.

From a general point of view, multi-sensors systems are usually managed using a signal, analog or digital, that pilots a trigger port on every device, ensuring their synchronization. Such element is standard for the industrial cameras, usually combined with an illumination to control the light condition of the scene. The same, however, is not true for the laser scanners.

This second class of devices is not provided of any kind of piloting input port, the sensor has indeed to work continuously for the purposes of the safety. The only information related to the time instant in which the scansion takes place is a time stamp that is included in the data packet sent from the sensor to the PC (not all models provide it). The time stamp, however, can not be used for the synchronization with the camera: this information is sent only at the conclusion of the scansion, two entire revolutions of the laser head, too late in order to command the acquisition of an image.



(a) Ideal Configuration

(b) Acquisition: laser

(c) Acquisition: camera

(d) Motion combination and resulting transformation

Figure 5.15: Displacement of the sensor due to the asynchronism

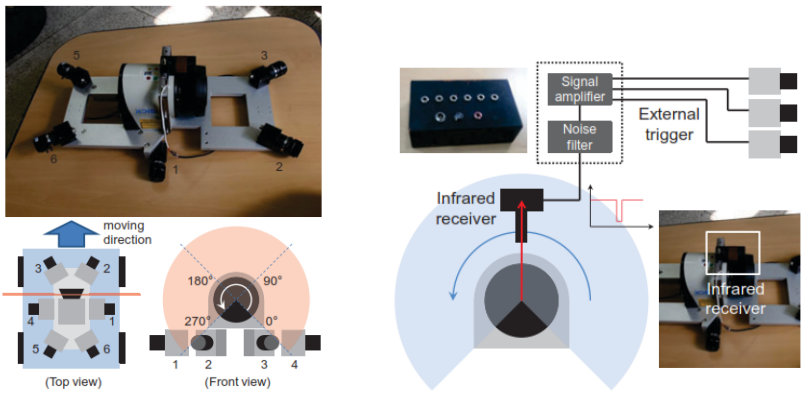
The underlined problem can be solved by adopting one of two possible strategies:

- stopping the vehicle and only then acquire the data
- using an external device that synchronizes the acquisitions

The first strategy, broadly used during development, it is advantageous in terms of implementation, not requiring any hardware modification. It has a main drawback: the pause of the motion of the vehicle. The stop should theoretically last as long as necessary to acquire the data, less than a second (depending on the number of scansions to be collected), but the real duration of the operation is related instead to the capability of the vehicle to stop the motion and start it again (dynamic performances of the machine), introducing a latency of at least 2-3 seconds. Such delay could be not acceptable as it is considered a dead time for the production.

The second strategy derives from the work of Bok et al. (2011). The paper presents

a device finalized to scan urban environments using a laser scanner and 6 cameras, fig. 5.16(a). As for the current application, the asynchronism in the acquisition is an issue. The capturing speed of the laser sensor is 50 fps, the one of the cameras is 60 fps. An infrared detector is then attached in front of the laser sensor, figure 5.16(b) and the detected laser signal is sent to the each camera after passing through a noise filter and an amplifier. In the upper left side of 5.16(b), which shows the box including the noise filter and the signal amplifier, the red terminal at the bottom receives the signal of photodetector and the upper six white terminals provide the trigger signal to all the cameras simultaneously. The trigger signal is generated 0.4 milliseconds after the infrared ray is received, time required by the hardware to process and condition the signals, delay considered negligible.



(a) The device (b) Syncronization logic

Figure 5.16: Synchronous laser-camera system by Bok et al.

An industrial device with similar functionalities was not found on the market, so a custom one was created for the current application. A dedicated electronic was developed starting from the characterization of the laser beam.

Fig. 5.17(b) shows the reading of a photodiode positioned around 60 centimeters far from the laser head. The area of the receiver, the angular resolution of the laser and the cone of the laser beams cause the identification of more consecutive peaks in a sampling interval of 0.2 milliseconds. The number of these peaks depends on

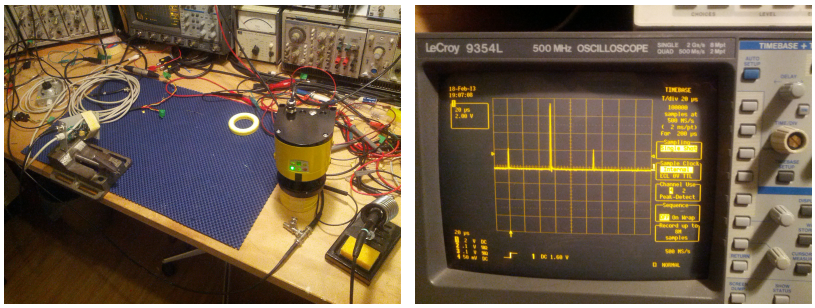
the distance between the diode and the laser head: with the configuration in which the sensor is placed at 2 centimeters the number of peaks rises to nine. With this kind of signals the generation of a synchronization pulse can not start from the crossing of a defined threshold: consecutive peaks can active the process multiple times, influencing the repeatability of the device.

A conditioning block has been developed from the analysis of the readings of the photodiode. The height of the fringes derives from the different incident power of the laser beams on the receiver, fig. 5.17(c). If the sensors position is fixed, the incident power is almost the same in every scansion (omitting the influence of vibrations etc), the integral of the signal is then constant. A low pass filter, an integrator, acts as sum, generating a curve of accumulation; comparing the value of the accumulation with a threshold value, dependent to the amplification used in the conditioning block, is emitted of the triggering signal, fig. 5.17(d). The electronics developed is placed inside the enclosure of the camera, minimizing the volume of the device, fig. 5.18(b).

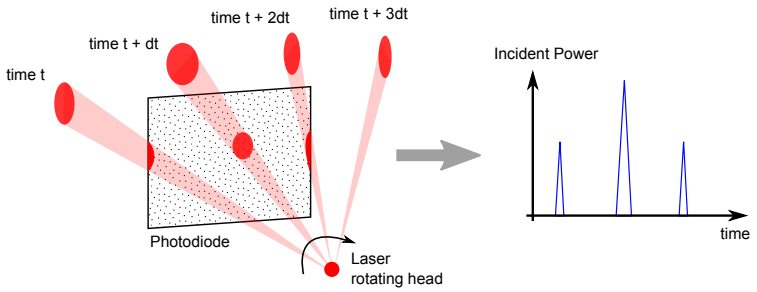
The outputs signal has a period of 40 milliseconds, interval that matches with the specifications of the laser: 1 sample each 80 milliseconds, double scansion, fig. 5.5.

The synchronous acquisition of the sensors ensures the stability of the fusion process but it is however still not sufficient to avoid the stop of the vehicle. As will be explained in chapters 6, in order to achieve the picking of the pallet an important information is required: the position of the vehicle in the instant in which the laser acquires the scansion.

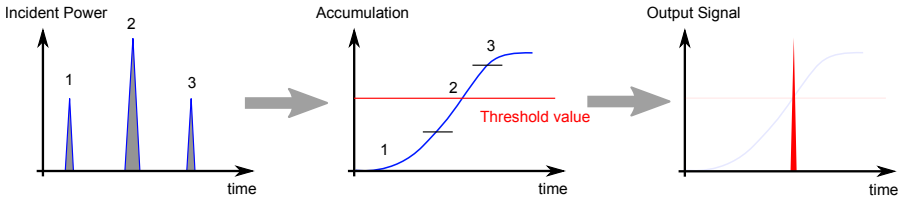
The pinking is a task that implies the motion of the AGV and so a trajectory must be planned: any asynchronism between the identification and knowledge of the pose of the vehicle causes the failure of the task, steering the AGV to a position not suitable for the forking. Such information can be obtained in two ways: by installing on board the AGV a triggerable device connected to the trigger of the laser (implying the modification of the AGV), or by reconstructing the pose of the vehicle from subsequent laser scans (SLAM: Simultaneous Localization And Mapping). The second strategy is the less invasive but it also represents another state of the art open problem.



(a) Experimental setup (b) Single laser beam detection



(c) Photodiode response

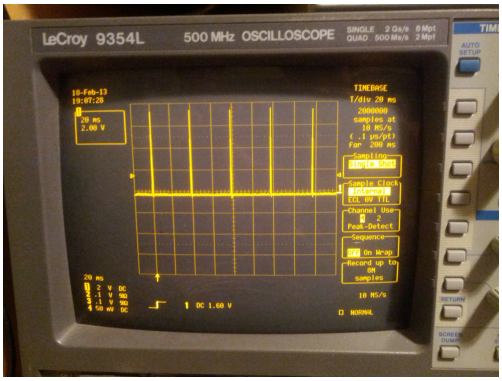


(d) Output signal generation logic

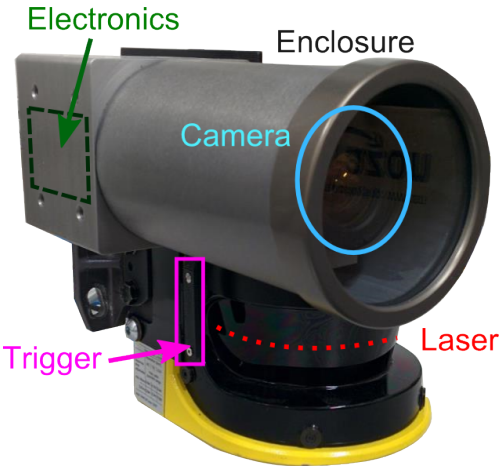
Figure 5.17: Trigger development

As said before, the less complex and invasive strategy remains the stop of the AGV.

A more robust implementation of the synchronization is achievable only after a full integration of the device on board of the AGV, designing together with manufacturer of the machine the electronics and the communications protocols required for such purpose.



(a) Trigger output



(b) Laser-camera device

Figure 5.18: Trigger Output

CHAPTER 6

EXPERIMENTAL VERIFICATION

Safety first!
The motion of an automatic vehicle inside a plant is a task that involves risks. Those must be somehow predicted and prevented.

6.1 Introduction

AN INDUSTRIAL DEVICE must ensure the proper operation in every possible operative condition, always in safety: for the people, for the objects, for the plant, for itself. The will of automating the procedure of forking has therefore to comply with that: the system must identify objects but at the same time it must verify if the picking can be performed without risks.

A peculiar aspect of the current application is the connection between measurements and risks: these are not due to the capability of identifying or not the objects (omitting the false positive cases), they are instead related to the motion of the automatic vehicle, which is the conclusive part of the autonomous picking.

The motion of a robot is always critic as it implies the management of the machine at a low level: an interaction with the controller and the motors, requiring necessarily the development of custom elements for each vehicle model. That is in contradiction with the intention of keeping the system as general as possible and usable as a *plug and play* device without modifications of the AGV, the direct interaction with the controller of the AGV must be avoided. The system was then organized in order to evaluate and generate high-level information that the vehicle can handle and use to

run autonomously its own task.

An example is a reference pose positioned frontally to the pallet. This information can be passed directly to the AGV, leaving the task of path planning and motion to its controller, or rather directly planning a trajectory and passing it as reference to be followed. This second assumption however implies that the vehicle is able to perform the maneuver planned, not obvious. The kinematics and the control system of a vehicle limits the possible trajectories that can be performed: a suitable trajectory planner is therefore necessary.

A path can be defined in several ways, like points in space, curves, but also as commands that produce the tracking of a trajectory in time. The main advantage of this second modality is that the use of control variables allows to overcome many of the kinematic constraints, generalizing the problem, and so increase the number of vehicles able to correctly follow the planned path. The main drawback is the increased complexity of the planning routine: the more general is the modeling, the greater will be the versatility in execution.

The parameters involved in this *versatile* trajectory planning can be: steer angle, curvature, torque/speed of the wheels . A possible strategy is derived from the papers Kelly and Nagy (2003); De Cecco et al. (2007b), in which are described continuous curves called *clothoids*. A custom path planner (and controller) based on the use of these curves was developed and positively used both in AGILE and in the current research during the conclusive tests with an AGV and an experimental robot.

Regardless the path chosen, every maneuver must fulfill a well defined constraint: the forks must move linearly under the pallet, so the ending part of the trajectory must be a straight line, fig. 6.1.

Given a trajectory, the risks associated to the motion derives from different possible causes: moving objects, wrong trajectories, non-feasibility of the picking, etc. The commonly adopted solution to such problem is the use of techniques of obstacle avoidance or on-line checking of the environment, Koren and Borenstein (1991); Susnea et al. (2009). These are useful for the purposes of safety of the vehicle, avoiding the impacts during the motion, but they suffer a main drawback represented by the

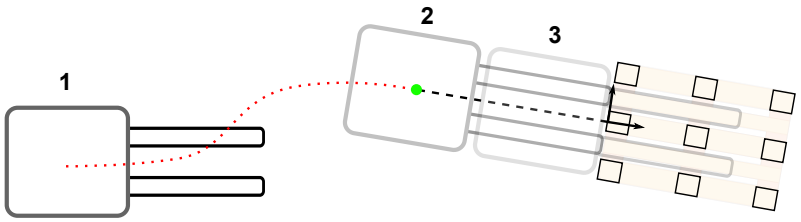


Figure 6.1: Forking procedure

occurrence of situations of operating inefficiency: avoiding an obstacle or stopping the vehicle during a maneuver are actions that modify or interrupt the task of the AGV, causing downtimes and slowdowns for the plant.

The situations just described have usually a low impact on the standard operations of the AGVs: the main task is to move the vehicle to different manufacturing/stocking areas, in which the accuracy in performing the motion is not considered a critic. Corrections of the maneuver and stops of are allowed in order to ensure the safety of the plant. The same can not be stated for the picking where the objective is to reach a target position as fast as possible with a *sufficiently hight* accuracy.

The motion of a wheeled vehicle must undergo to the holonomic constrains and for this reason there is a huge difference in planning and executing a maneuver inside a confined space or inside a wide area: the first case is very complex, the second is easier to be achieved. That derives from the limitations in mobility of which all wheeled vehicles have: a point in space can be reached with infinite maneuvers, which imply a certain number of controls and a minimum space for the execution.

Set the upper limit of 5 meters of distance, value related to the resolution of the laser and the minimum number of points required for the identification, it is mandatory that the maneuver of picking occurs within such space without further subsequent iterations.

The application of the standard techniques in this case would rise to frequent downtimes: the limited space for the maneuver does not allow the modification of the paths to avoid obstacles or to modify the motion in order to correct an error while approaching the pallet.

For these reasons, it is necessary to develop a strategy that allows the vehicle to move directly toward the pallet. Such objective can be achieved by applying state of the art techniques that combine obstacle avoidance and path planning, identifying whether there is, or not, a path that allows the vehicle to move safely within a laser mapped environment, Von Wahlde et al. (2009). That is however still not sufficient to solve the problem of the safety: even if there is a suitable path for the forking there is a marginal risk associated to the impact of the forks with the pallet.

The trajectory that points to the reference pose in front of the pallet is a clear example of a derived measurement: the coordinates are evaluated indirectly from the data of the laser and the calibration between the sensor and the AGV. The parameters involved suffer of errors that propagate in the measurement chain. Even if the identification is robust and reliable, the uncertainty about where to steer the vehicle is the most relevant element of risk of the process of picking.

From what aforesaid, it is necessary to verify whether the planned trajectory is safe or not, checking if the error associated to the measurement chain can potentially provoke an impact.

A possible strategy is a continuous tracking of the pallet during the motion of the vehicle, real-time verifying the conformity of the path traveled. This approach is advantageous in terms of control logic, with a continuous update of the information of where the pallet is placed relatively to the vehicle and so the capability to apply *minimum* corrections to the trajectory in order to compensate the errors due to the motion, the measurement or both. The main drawback of that strategy is the necessary synchronous acquisition of the poses of the vehicle and the laser scansions: to apply any corrective action the laser data must be related to the reference system of the vehicle. Path planning is a task that requires substantially two information: a starting and an ending point($[x, y, \delta]$); the first is the actual pose of the AGV, the second is the target point in front to the pallet. If the vehicle is moving, it is fundamental to know the exact position of the AGV in which the laser acquires the data, such information is indeed necessary in order to provide to the planner the proper starting and ending coordinates. Any delay in the acquisition causes an error in the estimation of the

two points: a delay in time causes a displacement in space, so as for laser-camera couple, causing the evaluation of wrong trajectories.

Such operative configuration is not easily achievable unless modifying the sensors and adding custom hardware on-board of the AGV (extending the trigger signal of laser and camera to the controller of the AGV). Such solution remains however the most attractive one because it would remove the constraint of stopping the vehicle in order to acquire the entire dataset (laser,camera,pose) required for the task.

A complementary strategy, useful in overcoming the technological limitations mentioned, is the planning of successive stops of the AGV along the main trajectory. In this way it is possible to collect the data asynchronously (the pose of the AGV is known and fixed while acquiring the laser). However, in order to maximize the operative performances of the machine, it is not advisable to stop the vehicle more than once per forking procedure. It is then essential to maximize the extraction of the information from the data acquired during the initial stop of the vehicle, planning successive stops only if strictly necessary. A method/strategy must therefore be developed in order to estimate in predictive manner the level of risk connected to the results of identification and the planned path. It must consider the parameters that influence the picking, such as the calibration between the sensors and the accuracy of the identification, evaluating the degree of uncertainty of the position of the vehicle before the linear forking of the pallet. Such information must then be compared with the geometric configuration of forks and the shape of the pallet to determine whether the errors in the measurement chain could cause an impact.

The picking of a pallet can occur in safety even if reaches the target reference pose with a displacement error, position and attitude: the forks are usually thinner than the gaps of the pallet, that must be analyzed and modeled. Such verification, called *error budget analysis*, is a key element for the decision-making process of the control logic of the AGV. That represents a direct method by means of which it is possible to evaluate if, and when, the maneuver can be performed in safety, or whether it is necessary to apply a strategy of risk minimization: approaching the pallet, stopping the AGV and performing an additional verification.

6.2 Error budget

A critical element, that it is however completely missing in the state of the art related to the autonomous picking of the pallet, is any analysis on the possible causes of failure of the proposed solutions. These works rarely speak of the procedure of forking, implying that the task is left to the control logic of the AGV and so managed by the manufacturers. Such procedure is not obvious because the uncertainty in parameters and measurements produces different effects on the final execution. Understanding and modeling them it is strategical in order to increase the level of safety of the machines.

Consider a pallet positioned 5 meters away from the AGV and an error in the nominal value of the attitude of the laser on board of just 1°: this configuration causes 8 centimeters of displacement error between the reference pose in front of the pallet and the one reached by the AGV, fig. 6.2.

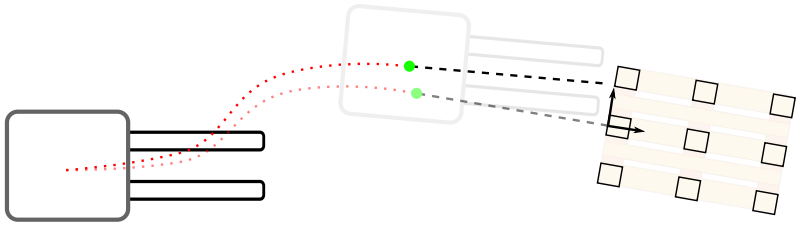


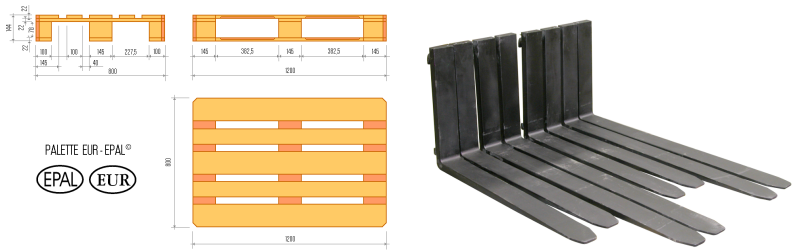
Figure 6.2: Influence of the errors in the forking

This value could seem negligible but it represents an element of failure: this error sums with the ones associated to the motion of the vehicle and the uncertainty in the identification itself, resulting ultimately in a possible impact between the forks and the pallet feet, with consequential damages and downtimes.

The discrepancy between the nominal values of the parameters and the real ones is an inevitable condition, and for this reason it is a good practice to consider such factor as influential inside the processes.

The process of forking is considered safe when the conclusive linear motion of the AGV, in which the forks pass under the pallet and between the gaps, causes

no impacts. Once stated such condition, it is possible to analyze the geometry and evaluate which errors, in reaching the reference pose, still ensures no impact. That depends on two factors: the geometry of the forks (fixed for the machine) and the position from which the vehicle starts the linear motion (variable due to the uncertainty in the identification, calibration and motion). The parameters of interest are the displacement errors between the position reached by the AGV and the reference pose, $[x_{RP}, y, _{RP}\, \theta_{RP}]$, placed in front of the pallet at a distance approximately equal to the length of the forks. Given the geometry of the pallet and forks, it is evaluated a functional that relates the displacement and the impact condition.



(a) Pallet (b) Forks

Figure 6.3: Geometries of the problem

Two different types of impact can occur: one between the end of the fork and the inner feet of the pallet, and one between the body of the fork and the feet of the front face (this second case models the impact between the forks and the face of the pallet during the initial part of the linear motion), fig. 6.4. Both these conditions derive from the linearity of conclusive part of the maneuver.

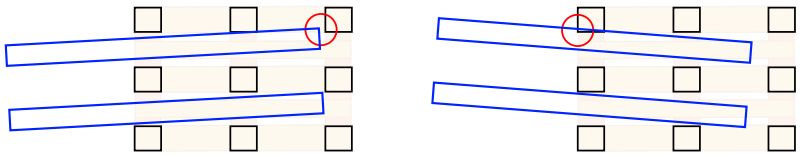


Figure 6.4: Impact cases

These configurations can be modeled using a parametric segment, dependent on the pose of the vehicle in front of the pallet, to calculate the vertical coordinate of two

reference points of the geometry in fig. 6.5: a point on the segment with longitudinal coordinate coincident with the front face of the pallet, P_1, and a point on the end the tine , P_2. Both these values must be internal to the geometric constraints imposed by the shape of the pallet: the two gaps used for the picking.

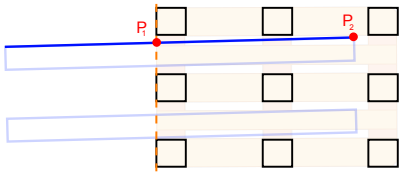


Figure 6.5: Check points

P_1 and P_2 are calculated starting from the definition of the relative displacement between pallet and forks. The pallet is placed to the coordinates $[L_{forks}, 0, 0]$ where L_{forks} is the length of the forks (the minimum distance for the linear motion), the origin forks to the coordinates $[x, y, \theta]$, variables of the problem. The target pose in this case is the origin. The value x y and θ represent the displacement and angular error in reaching such pose before the liner motion. The objective is the evaluation of those values that cause no impacts and so to define an *admissible error region*.

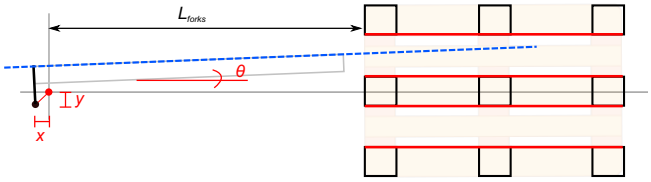


Figure 6.6: Safety Analysis: geometry and parameters involved

Defined as $h_1(x, y, \theta)$ the height of point P_1, and h_2 the one of P_2, the notations that define such values are:

$$\begin{aligned} h_1(x, y, \theta) &= y + P_{h_i} - tan(\theta)\left(P_{ff} + x\right) - \frac{\pm F_i \pm F_w}{2cos(\theta)} \\ h_2(x, y, \theta) &= y + P_{h_i} - tan(\theta)\left(cos(\theta) \cdot F_l - \frac{1}{2}sin(\theta)(\pm F_i \pm F_w)\right) - \frac{\pm F_i \pm F_w}{2cos(\theta)} \end{aligned} \tag{6.1}$$

P_{h_i} is the *ith* constrain relative to the admissible heights, the gaps between the feet of the Euro pallet [-0.300 to -0.0725, +0.0725 to +0.300]

P_{ff} is the coordinate of the frontal face of the pallet, L_{forks} (minimum)

F_l is the length of the forks

F_w is the width of a tine

F_i is the interaxis of the forks

This formulation is applied to both the tines conveniently adjusting the sign of F_i and F_w according to the side considered.

Eight nonlinear functions are derived, two for each side of the two tines. These represent the conditions in which each side of the fork is coincident with its constrain. The intersections of these functions define the 3D region of space $[x, y, \theta]$ within which an error does not cause the impact with the palled and so the failure of forking procedure, fig. 6.7.

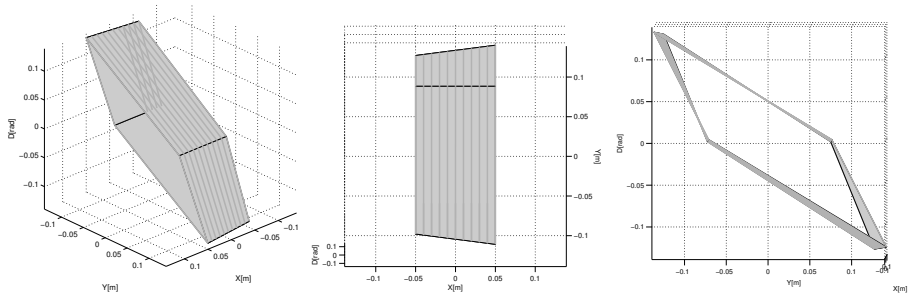


Figure 6.7: 3D admissible error volume

In order to simplify the notation, and the representation of the results, the analysis from now on will be focused on the subspace $[y, \theta]$, fig. 6.8.

This simplification is motivated by the different influence of the errors on the maneuver: from the geometry of the problem follows that the parameters with the highest influence on the impact are the error along the transversal direction of the pallet and the relative angular displacement.

An error in $[y, \theta]$ causes a shift of the forks along the direction with the lower degree of freedom, the vertical one.

On the opposite, an error along the longitudinal displacement x has in practice no influence on the impact. The x is connected to the *dimension* of the admissible area,

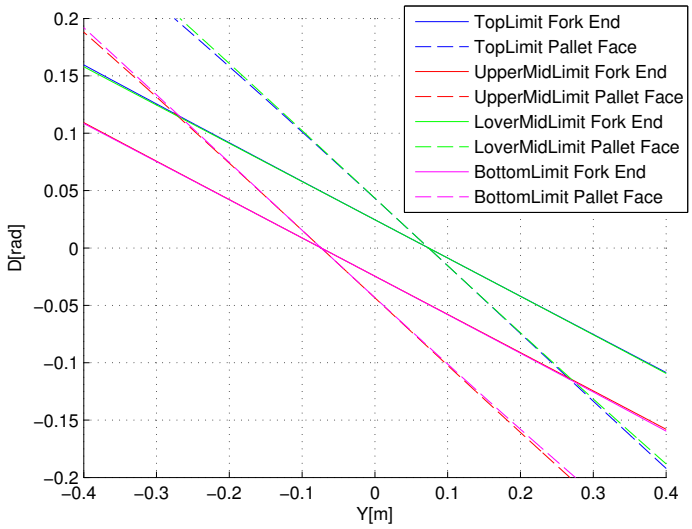


Figure 6.8: Safe region boundaries

changing the *number* of possible configurations suitable for the forking: more far from the pallet are the forks, more configurations result to be valid (higher admissible errors in $[y, \theta]$). In fig. 6.9 is presented the admissible volume highlighting the influence of a displacement along x: the blue is the configuration with an error $x = +5cm$, the red $x = -5cm$. Such variation can be considered negligible because of its minimal influence on the safety of the maneuver.

Applied the simplification, the region identified by the 4 more restrictive conditions is checked using a brute force test in which the model of the forks is moved verifying the occurrence of impacts. Fig. 6.10(a) presents the displacements of the forks that ensures no impact. Fig. 6.10(b) demonstrates that the corresponding distribution is inside the admissible error region, verifying the formulation adopted.

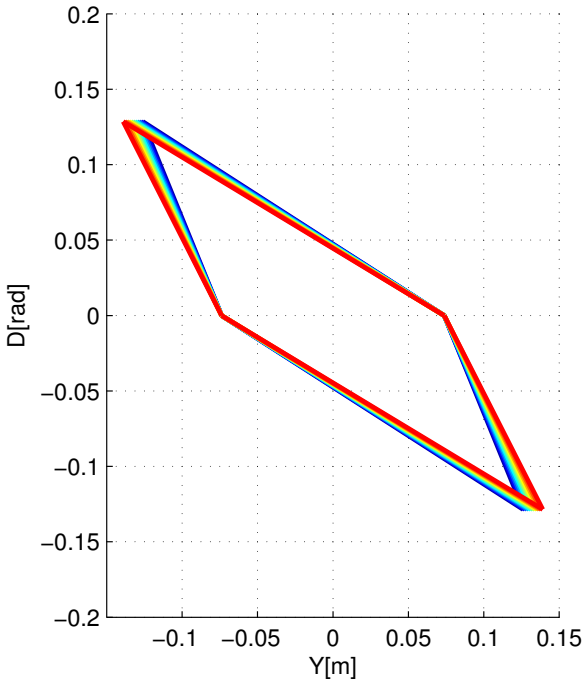
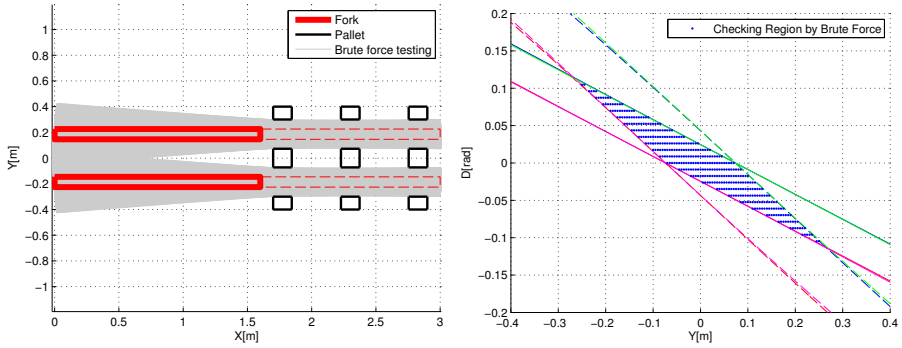


Figure 6.9: Frontal error vs the admissible area



(a) Admissible poses that ensure no impact with the pallet

(b) Brute force testing and verification

Figure 6.10: Error budget: brute force check

In order to maximize the computational efficiency, the generatrices of the region are approximated with 4 linear segments traced between the intersections of the boundaries, fig. 6.11.

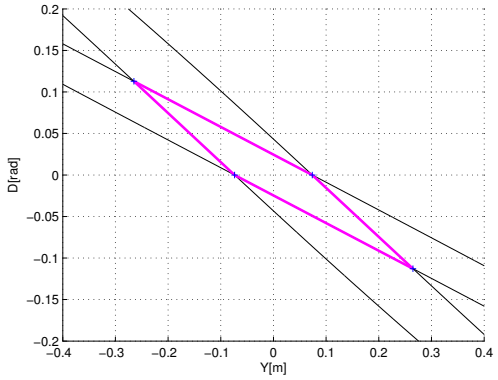


Figure 6.11: Polygonal approximation of the region

The error of such approximation can be evaluated by calculating the difference of the area built from the nonlinear functions or the polygon.

$$\Delta Area = \frac{A_{poly} - A_{Boundaries}}{A_{Boundaries}} \tag{6.2}$$

The difference between the areas is equal to the -0.0173%, negligible, verifying that the approximation occurs in regions in which the generatrices are close to linearity.

6.3 Uncertainty propagation and Safety Check

Once identified the admissible errors it is necessary to structure a procedure of verification that takes advantage of such information.

The maneuver occurs in safety only if the position reached, compared to the reference one, has a displacement error that remains inside the admissible error area. Such assumption however does not solve the problem because it represents a posteriori analysis, verifying of impact when it is already occurred or inevitable.

The question to be answered is instead: *when/how do the errors in the measurements lead to a pose that is still suitable for the forking*?

It is then necessary to structure a *predictive* method that analyzes the measurement chain and provides an estimation of the risk before the motion takes place.

As explained previously, there are two main elements than influence the measurements, and so the picking:

- laser measurement/identification accuracy
- calibration accuracy

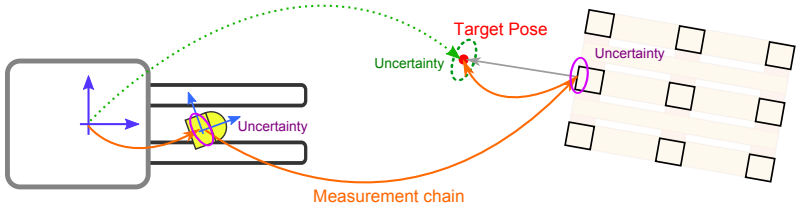


Figure 6.12: Measurement chain and uncertainties

It is therefore fundamental to model the entire measurement chain and evaluate not only the target reference pose $[x_{RP}, y_{RP}, \theta_{RP}]$, but also the associated uncertainty/covariance matrix.

From the theoretical point of view, the covariance matrix is related to the probability of occurrence of an error given a measurement: that is the information that can be correlated with the admissible error area.

The calculation of the covariance of the target reference pose can be achieved in the same way as done for the laser identification or the calibrations: $C^* = JCJ'$. In this case the covariance C is a combination of two different matrices: one for the vehicle-laser calibration, C_{V2L}, and one the identification of the laser, C_{L2P}, fig. 6.12. These are not correlated and so they can be merged in C as two independent blocks on the diagonal.

$$C = \begin{bmatrix} C_{V2L} & 0 \\ 0 & C_{L2P} \end{bmatrix} \tag{6.3}$$

Both C_{V2L} and C_{L2P} are 3×3 covariance matrices ($[x, y, \theta]$), C is a 6×6. Given the initial covariance matrices of the parameters C, it is then sufficient to evaluate the J of the transformations involved the measurement chain.

Defined H_{V2L} the homogeneous transformation matrix between the vehicle and the laser, and H_{L2P} the homogeneous transformation matrix between the laser and the pallet, the pose of the pallet expressed in the reference system of the AGV, H_{V2P}, can be computed as

$$H_{V2L} = \begin{bmatrix} cos(\theta_{V2L}) & -sin(\theta_{V2L}) & x_{V2L} \\ sin(\theta_{V2L}) & cos(\theta_{V2L}) & y_{V2L} \\ 0 & 0 & 1 \end{bmatrix} \tag{6.4}$$

$$H_{L2P} = \begin{bmatrix} cos(\theta_{L2P}) & -sin(\theta_{L2P}) & x_{L2P} \\ sin(\theta_{L2P}) & cos(\theta_{L2P}) & y_{L2P} \\ 0 & 0 & 1 \end{bmatrix} \tag{6.5}$$

$$H_{V2P} = H_{V2L} H_{L2P} \tag{6.6}$$

The coordinates of the pallet are then

$$\begin{bmatrix} x \\ y \\ \theta \end{bmatrix} = \begin{bmatrix} H_{V2P}(1,3) \\ H_{V2P}(2,3) \\ \theta_{V2L} + \theta_{L2P} \end{bmatrix} \tag{6.7}$$

From that formulation, applying the partial derivatives on
$[x_{V2L}, y_{V2L}, \theta_{V2L}, x_{L2P}, y_{L2P}, \theta_{L2P}]$
is computed the Jacobian matrix

$$J = \begin{bmatrix} 1 & 0 & -y_{L2P}cos(\theta_{V2L}) - x_{L2P}sin(\theta_{V2L}) & cos(\theta_{V2L}) & -sin(\theta_{V2L}) & 0 \\ 0 & 1 & x_{L2P}cos(\theta_{V2L}) - y_{L2P}sin(\theta_{V2L}) & sin(\theta_{V2L}) & cos(\theta_{V2L}) & 0 \\ 0 & 0 & 1 & 0 & 0 & 1 \end{bmatrix} \tag{6.8}$$

Applying covariance propagation is then computed the covariance matrix of the pose of the pallet, C_{V2P}, expressed inside the reference system of the AGV.

This is however not the information of interest: the vehicle must not be steered to the pallet face, but to the reference pose in front of it, P_{FP}. This is evaluated by adding a further transformation to the chain, H_{P2FP}, that moves back the coordinates of the pallet according to the length of the forks: $[x_{P2FP}, y_{P2FP}, \theta_{P2FP}] \leftarrow [L_{forks}, 0, \pi]$. This transformation is considered certain and error free (C does not change).

$$P_{FP} = H_{V2P}H_{P2FP} \tag{6.9}$$

That, once derived, gives the J to be multiplied with the covariance matrix C_{V2P} previously calculated. Called $\alpha = \theta_{V2P}$ and $\beta = \theta_{P2FP}$ for a more compact notation,

$$J = \begin{bmatrix} 1 & 0 & y_{P2FP}(sin(\alpha)sin(\beta) - cos(\alpha)cos(\beta)) - x_{P2FP}(cos(\alpha)sin(\beta) + cos(\beta)sin(\alpha)) \\ 0 & 1 & -x_{P2FP}(sin(\alpha)sin(\beta) - cos(\alpha) * cos(\beta)) - y_{P2FP}(cos(\alpha)sin(\beta) + cos(\beta)sin(\alpha)) \\ 0 & 0 & 1 \end{bmatrix} \tag{6.10}$$

Fig. 6.13 presents an example of real data acquired with an experimental robot, the length of the fork in this case was set to 20cm (due to the limited space in which the test was performed). In the figure all the different ellipses are expressed in the reference system of the robot and it is visible the mutual influence of the uncertainty along the measurement chain.

The ellipse on the target reference pose indicates the area in which the vehicle could be situated at the end of the trajectory (before the linear part) with a confidence interval of 95% (95 of 100 successive motions would fall inside such area). That information can be then interpreted in the following way:

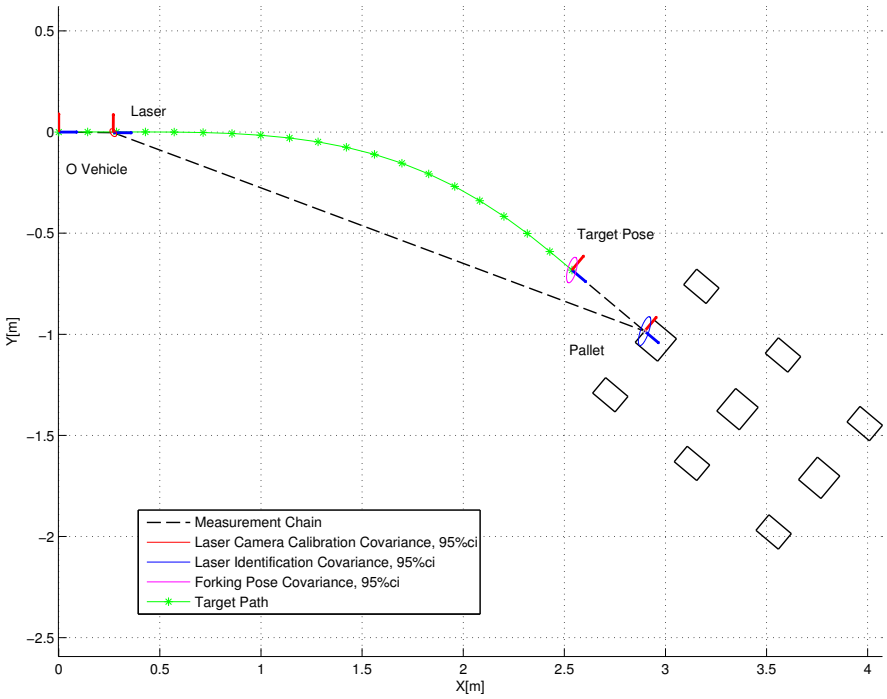


Figure 6.13: Example of measurement chain

- high uncertainties are related to an high probability of impact
- the probability of an impact is related to the covariance matrix and so to the dimension of the ellipse
- from accurate identifications and calibrations derive safe and accurate forking procedures and small ellipses

There is a direct dependence between the dimension of the ellipse, the uncertainty of the measurement and the *probability* of an impact.

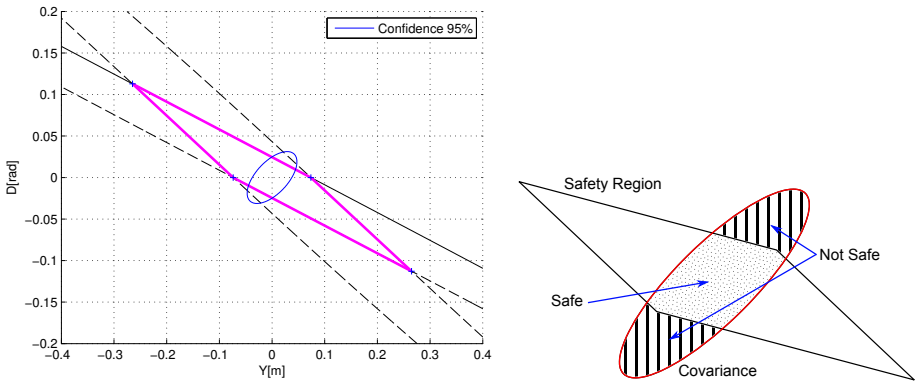
Such representation of the covariance can be compared with the admissible error area: *the first represents the possible error in reaching the target pose, the second the admissible error that causes no impact with the pallet*.

The comparison of the covariance ellipse and the admissible error region starts by positioning of the ellipse in the origin of the space $[y, \delta]$. The motivation in placing

the ellipse in the center of the axis comes from the logic used in the construction of the safe area: the information of interest is the error in reaching the position, not the position itself. The absolute coordinates lose their meaning in this analysis, focusing on the dimensions of the distributions (the amounts of the errors) rather than on their position in space.

It is however important to underline that the position of the distribution depends on the accuracy in estimating the parameters involved in the measurement chain. If these are not accurate, the covariance of the reference pose will be high, with a wider distribution, verifying in practice the higher probability of an impact with the pallet.

By superimposing the two curves, fig. 6.14, it is possible to notice that the ellipse of 95%ci (confidence interval) only partially overlaps the admissible error polygon. This indicates that there is a residual probability of an impact: the vehicle could reach a reference pose in front of the pallet with an error outside the admissible limits, with a consequent not suitable configuration for the picking.



(a) Admissible error area and 95% ellipse

(b) Regions superpositions

Figure 6.14: Reference pose uncertainty vs admissible error region

On the opposite, an ellipse that is entirely inscribed in the polygon means that the probability of an impact is close to zero: the confidence interval of 95% indicates that the target pose will be reached almost alway with an admissible error for the forking. An ellipses of 99%ci will grant the maximum level of safety (if inscribed); such configu-

ration, however, was never achieved during the test phase due to the limited accuracy of the processes involved (laser-vehicle calibration: poor robot performances in localization).

Since it is very unlikely that the entire ellipse is inscribe inside the admissible region, the information of interest becomes the *level* of safety in performing the maneuver: the AGV needs a method that provides an estimation of the risk, from which to derive suitable corrective actions, like moving closer to the pallet and perform a further check.

Such information is obtained by changing the confidence interval of ellipse associated with the reference pose. There is indeed a fundamental difference between an ellipse that is partiality inscribed and one that is instead entirely: in the first case it is difficult to evaluate which part is safe or not, in the second the whole distribution falls inside the admissible area. This second configuration can be achieved by lowering confidence interval of ellipse, obtaining a non-linear reduction of the dimension of the ellipse itself.

From Smith and Cheeseman (1986), given a 2×2 covariance matrix C, its representation in the form of an ellipse is achieved by defining the value of a parameter K, dependent on the cardinality of the problem (in this case 2), from which the confidence interval is evaluated as the probability P of a point to lie within that ellipse, higher is K (and P) wider is the ellipse. 6.11 is the formulation that relates K and P to the dimension of the ellipse; the commonly used values of K and P are reported in table 6.1.

$$P = 1 - e^{\frac{-K^2}{2}} \rightarrow K = 2log\left(\frac{1}{1-P}\right) \tag{6.11}$$

The axis and the orientation of the ellipse are evaluated from the eigenvalues and

Table 6.1: K and P for 2D distributions

K	10.6	9.21	7.38	5.99	4.61	2.77	1.39
P	99.50%	99%	97.50%	95%	90%	75%	50%

eigenvectors of the covariance matrix C.

$$axis_{ellipse} = \sqrt{K \cdot eigvalues(C)}; \tag{6.12}$$

From 6.11 is calculated the value of K from which derives an inscribed ellipses, and so the probability P that the displacement error of the pose reached by the AGV lies inside the admissible error area. Given a ellipse with semiaxis a and b, oriented with an angle θ, the k value of scaling that makes such ellipse tangent with a generic line $y = mx + q$ can be computed by solving

$$\left[\frac{1}{a^2} \quad \frac{1}{b^2}\right] \begin{bmatrix} (x\cos(\theta) + (mx+q)\sin(\theta))^2 \\ (-x\sin(\theta) + (mx+q)\cos(\theta))^2 \end{bmatrix} - k^2 = 0 \tag{6.13}$$

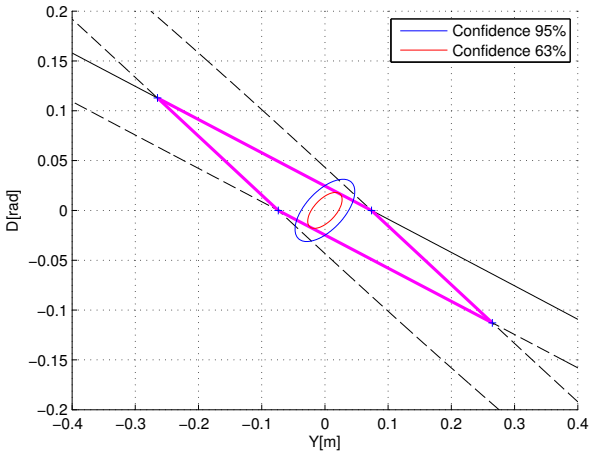


Figure 6.15: Safety check: Inscribed Ellipse

It must be emphasized that the identification of such ellipse does not guarantee the safety of the picking: *it indicates how much probable is the reaching of a suitable pose for the forking*.

The importance of such information comes from the automatization of the picking process. It is sufficient to think to pallets positioned at different distances and the possible operative choices that can be taken:

- a measurement chain of a pallet 4 meters far from the AGV suffers more of the effects of uncertainty due to the vehicle-laser calibration (lever action from

the distance between sensor and object), resulting in an inscribed ellipse with confidence interval close to 60-65%, motivating an initial approach toward the pallet and a further elaboration

- the identification of a pallet close to the vehicle has instead a lower associated uncertainty, less influenced by the uncertainty of the parameters, allowing a direct maneuver of picking

From that is possible to define a threshold value, a score or a routine that takes as input the confidence level of the inscribed ellipse and uses such information for the decision making process of the AGV.

The definition of this part is a critical for the safety of the task and must to be optimized/adjusted taking into account different elements, not all related to the geometry of the problem:

- the needs of the user
- the performances of the vehicle
- the performances of the control
- the geometry of the vehicle
- the geometry of the forks
- the geometry of the pallet
- the trajectory used for the final approach

Some of these can't be defined a priori, on field tests are required to characterize the machine and tune this last step.

Two different modalities can be achieved at the industrial level: for the AGV model, analyzing and tuning the decision making process once for all the vehicles of the line, or instead as an additional step to be included in the calibration process of every machine, testing the behavior of the machine and the interaction with the identification system. The first is more general and less accurate, the second one is slower but more safe. Once completed this last stage the AGV is fully calibrated and ready to

work, equipped with a powerful tool that improves its operative performance, the versatility and at the same time increase the safety level of the machine itself.

The described procedure can be extended to the full 3D analysis including also the error along the longitudinal direction, x. As said before the admissible error area becomes an admissible error volume and the 2D ellipse becomes a 3D ellipsoid, fig. 6.16.

The elaboration logic remains the same.

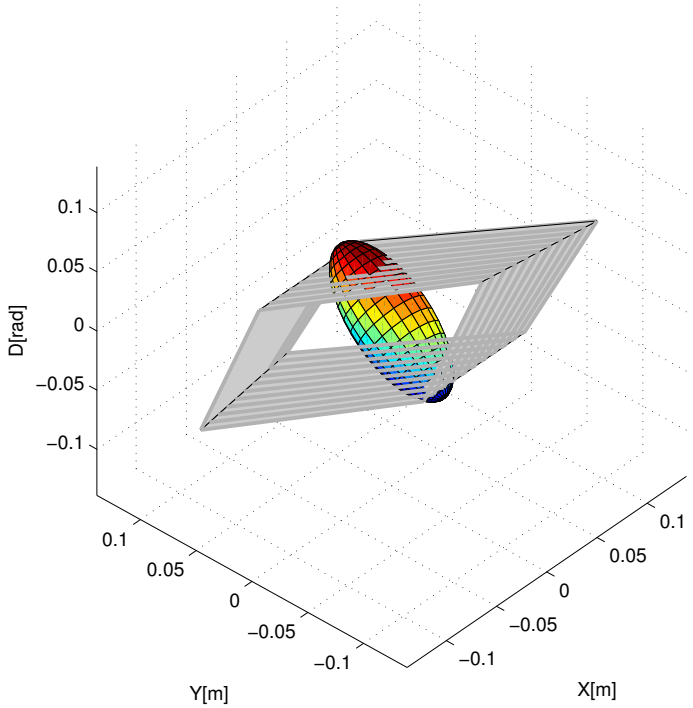


Figure 6.16: 2D regions superpositions

6.4 On field test

In the tests performed during the AGILE project, few hundred trials, the worst cases were those in which the forks barely touched the inner pallet feet, however never causing a displacement of the pallet greater than 2-3cm. The AGV used for the test was equipped with wide forks, 15 centimeters (the most common are of 8cm), with a consequent limited error margin in the picking. Such results gave credit to the development achieved, verifying the performances of the previous version of the laser identification algorithm, which has proven to be reliable and repeatable in localizing the pallets, fig. 6.17.

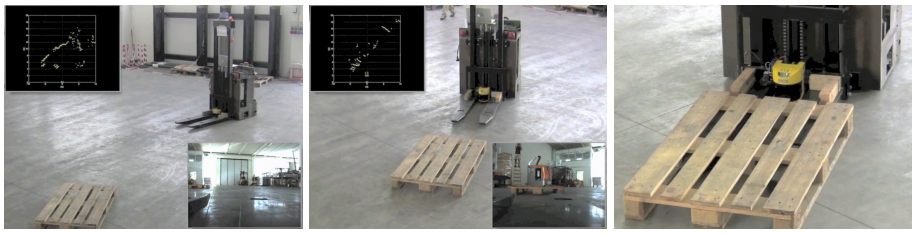


Figure 6.17: AGILE Forking

The developed system not only embeds the knowledge, the technology and the performances achieved with AGILE, it offers a new and more complete functional structure. The improvements achieved both in the identification algorithms and fusion process ensure a more robust and reliable processing with a better rejection of the possible false positive cases: **zero false positives were occurred in any testing condition from the definition of the elaboration routine**.

The functionalities have been empowered including all the necessary tools to evaluate not only whether there is, or not, a pallet close to the robot, but also if its picking is feasible and safe.

The system, in the current configuration, includes:

- **camera calibration**
- **laser-camera calibration**
- **laser-vehicle calibration**
- **laser identification**
- **camera identification**
- **laser-camera fusion**
- **path planner with continuous curvature**
- **obstacle detection**
- **AGV commands manager**

The software has been organized and installed on a dedicated industrial PC configured with an experimental robot P3DX. The robot was programmed modifying its controller in order to perform the tracking of the trajectories communicated by the identificator, fig. 6.18.

Figure 6.18: Experimental Robot: Evolution of the system

The configured system has been tested by moving the robot simulating the behavior of an industrial AGV, fig. 6.19. Despite the limitation of such a vehicle, like the torque of the motor and the poor performances in localization, the picking is accurate resulting in a suitable conclusive alignment between robot and pallet. During the testing the safety check percentage went below 60%.

Figure 6.19: Experimental Robot: Forking

CHAPTER 7

CONCLUSIONS

"It's more fun to arrive at a conclusion than to justify it."
Malcolm Forbes

7.1 Overview

THE THESIS presents the research done in order to develop an industrial device aimed at increasing the level of automation of the automatic logistic plants in which are employed AGVs.

The objective is to create a sensing system by means of which to perform the autonomous identification and forking of pallets placed in unknown positions inside of the cargo area of a warehouse.

From what shown in the initial part of the thesis, such task is well debated in the state of the art and studied for more than a decade. Much work has been carried out till now, producing many functional demonstrators and prototypes, but never completely satisfying the industrial specification related to this application. That consequently caused the stagnation of the technology used in this field, today not advantageous anymore.

What distinguishes this research from the other works does not come from the optimization of the computational aspects of the algorithms and strategies. The focus was instead aimed to the analysis and characterization of the measurements, from which was possible to maximize reliability and robustness of the identification process

and fulfill the requirements and specifications typical of every industrial device.

As shown, the autonomous picking of the pallet is not a trivial task. The development of an identification routine is only a part of the solution. The problem involves different elements, most of which represent modern challenges of the Research.

In order to fork a pallet the system must collect data, identify the object, plan a trajectory, avoid the impact with it. All these elements were analyzed, providing an operative solution.

About the identification of the pallet it was proved how the use of multiple sources of information attains the rejection of all the false positives, minimizing the occurrence of potentially dangerous situations. The combined use of laser and camera resulted to be the best configuration, achieving an higher accuracy and repeatability compared to the TOF technology.

It must however be underlined that the TOF represents the future for the logistic: the amount and type of information provided is way more rich and useful compared o the laser-camera configuration. The actual limitations are mainly due to the poor quality of the 3D data, condition that can only improve in the next years. There is indeed a strong interest from the industry, finding even now some applications based on such technology. It is sufficient to think to the evolution of these devices during the last 3 years: in 2010 just 1 TOF was classified as industrials, in 2014 more that 8, with better performances and more advanced functionalities.

The algorithms and strategies presented were coded and used inside a dedicate GUI (Graphical User Interface), software aimed at provide a simple and useful monitoring of the status of the system and the measurements achieved.

Two dedicated libraries were coded for the laser-camera configuration. For the camera was achieved a processing time of 3Hz, sufficient for the purposes of the application but also the bottleneck of the entire elaboration. For laser was instead achieved the real-time processing.

Together with the identification routines were included all the different parts required to correctly perform the picking of the pallet: intrinsic and extrinsic parameters of sensors and vehicle, obstacle detection, path planning, safety check.

About the parameters, these are estimated by two dedicated calibrations procedures, organized in order to be run as Matlab routines, generating a set of configuration files for the GUI. These information are a key element in order to evaluate the grade of the risk due to the uncertainties involved in the process.

This structure was implemented on an experimental robot verifying the correctness of the elaboration in different testing conditions.

The main contributions of this thesis are a robust identification system and a structured approach to the problem of the autonomous pallet picking, providing all tools necessary in order to setup an AGV and ensure its proper operation.

A further important element is the evaluation of the risk for the maneuver. This is studied in this case for the picking of the pallet, but it is however a general method that can be used in all those operative situation in which the shape and the dimension of the robot could cause potentially dangerous situations.

7.2 Awards

The technology born from this research received an important award in 2012, winning the funding announcement *SeedMoney* 2012 by Trentinosviluppo.

This initiative involved the submission of a business proposal based on an innovative product/device. The proposal, and so the idea and the device, was evaluated in terms of innovativeness, scientific validity, technical and economic feasibility. It was selected and financed, achieving the ninth place out of a total of 40 proposals.

7.3 Open issues and future works

The research presents a device at a more advanced stage than a prototype. At the current state of development it can be directly used on an AGV with minimum hardware modification: voltage supply and fasteners. The only element required is the design of a communication protocol with the machine, element that is however dependent to the control logic of the machine and the technical choices of the manufacturer of the AGV.

Despite the processing strategy fulfills the requirements of the industrial application, some critic elements are still pending. The system is functional, but suffers from some limitations due to some design choices taken.

The main limitation is the speed of processing: the time consumption associated to the identification limits the potentialities of the system. Comparing the two processes it is clear that there is a gap in performances between the laser and the camera: the first works in real time, the second not. The elaboration of the images is the bottle neck for the entire elaboration.

The development of a more efficient process, capable to ensure the same results in a shorter time, would allow a radical change in the interaction between the sensing block and the AGV.

A fast processing, close to the real-time, is the first fundamental step in order to design a device based not only on the identification of the pallet, but also its tracking during the maneuver of forking (synchronizing the vehicle with the sensors).

Such strategy would be optimal for two reasons:

- the vehicle does not need anymore to be stopped in order to acquire the data set: laser scansion, image, position of the AGV
- the continuous update of the data during the maneuver allows both to strengthen the forking procedure both to streamline and simplify the control of the trajectory

A frequent update of the pose of the pallet makes possible the continuous re-planning

of the trajectory for the picking, moving the AGV toward the pallet by applying minimum corrections on the maneuver. This strategy is easier and less computationally expensive compared to a control that stabilizes the motion of the AGV along a predetermined trajectory by minimizing the displacement error.

A further benefit from the tracking is the compensation of the errors during the maneuver. The uncertainty of the target pose to which the AGV must be steered it is strongly dependent on the distance and angle to the pallet: uncertainty propagation and the lever effect of the measurement. The iteration between the motion and the update of the target pose involves the gradual reduction of the influence of the calibration errors over the final measurement: after every update of the measurement the object is closer and closer, lowering the influence of the uncertainties. This assumption is true only if a proper laser-vehicle calibration is performed, necessary condition for every configuration.

A good solution to the problem of the computational speed could be the use of FPGA in which embed the code developed, task that requires specific expertise.

Another problem identified during development was the position of the laser on the vehicle. The 2D laser sensor must be positioned at an height from which the scanning plane can intersect the pallet feet at around 5cm from the ground. This specification derives from the characteristics of the laser cone emitted by the sensor, which at 5 meters of distance has a size comparable with the height of the pallet itself.

The size of the sensors used (as for models of other manufacturers) makes the fulfillment of such requirement difficult to be accomplished. The sensor should be fastened at a height of a few centimeters from the floor, configuration that is however not optimal because the dust and dirt can easily accumulate on the sensor, requiring a more frequent maintenance. There are laser models that are more compact, but none of these can be used for the current application. An example is the laser URG-04LX-UG01 from Hokuyo, the smallest on the market, which is not classified as safety laser scanner (it represents an additional device to be mounted on the AGV), or the new models SICK S300 mini, which however offers nor the accuracy nor the resolution required to identify the pallet inside the scansion (accuracy $\pm$4cm, angular resolution

higher than 0.5°).

During the development the solution to such issue was attained by testing different configurations of the sensor on board the robot, regulating the aim in order to maximize the scanning area. This configuration however underlined another problem: if the floor is not flat, condition theoretically guaranteed by the specifications of the warehouses, the the position of the scanning plane becomes influenced by the discontinuities of the floor itself, causing a vertical shift of it, sometimes aiming to the ground, sometimes higher than the pallet.

The best solution could be the adoption of an automatic tilter that continuously control and corrects the aim of the laser scanner in order to achieve the best operative conditions, such strategy must however deal with the problem of the safety: being a safety sensor, any modification of the nominal position could be not acceptable by the safety norms.

A minor problem is represented by the volume occupied by the camera. Despite the camera is rigidly attached to the laser sensor, its enclosure, required in order to certify the device, is bulky. In order to make easier the installation of the device on the AGV and minimize the space required between the forks, the video block should be redesigned using smaller sensors maybe OEM, with less sophisticated optics.

The issues described here must be then considered as new elements for a further step in the development. The main objective of this research was the development of a device able to fulfill the requirements of accuracy, reliability and safety, objective achieved thanks to an approach to the problem more focused on the characterization of the measurements rather than on the computational performances. Given this result, a new step of can be started, aimed this time on the improvement of the computational performances and an a partial redesign of the assembly: the device must be engineered. This second part is however more close to the industrial implementation rather than to the research field.
Lastly it should be emphasized as it is essential perform an extensive experimental campaign on the field. A test campaign has been carried out, both in the laboratory

and in a warehouse, but the only real way to effectively test the device is the daily use on-board of an AGV working in a real plant. *The operating conditions are the only element capable to underline any limitations or problems associated to the device.*

About the TOF technology, it does not seem, for now, a suitable solution for a reliable identification process. The results are however encouraging, motivating a further research in to verify the performances achievable by this technology compared to the progress made from the state of art.

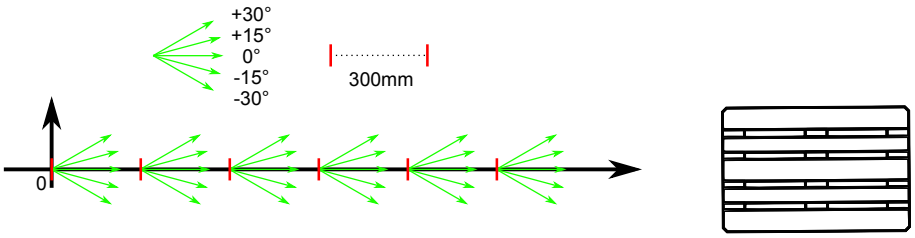
The use of this devices can be however suggested for those operative situations in which there is at least a priori rough knowledge of the position of the pallets. Compared to the laser-camera, the TOF can offer in these cases a more advanced information content: analyzing for example the volume or the integrity of the loads, the presence of obstacles not on the floor, the identification of more complex shapes. These tasks can however be performed in a robust way only if the data are accurately segmented, condition derived from the knowledge of the environment and organization of the plant.

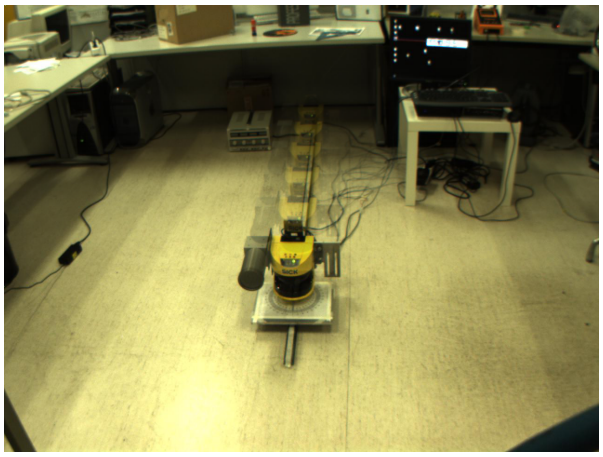
APPENDIX A

METRIC QUALIFICATION DATA

Frontal movement

Setup and covariances





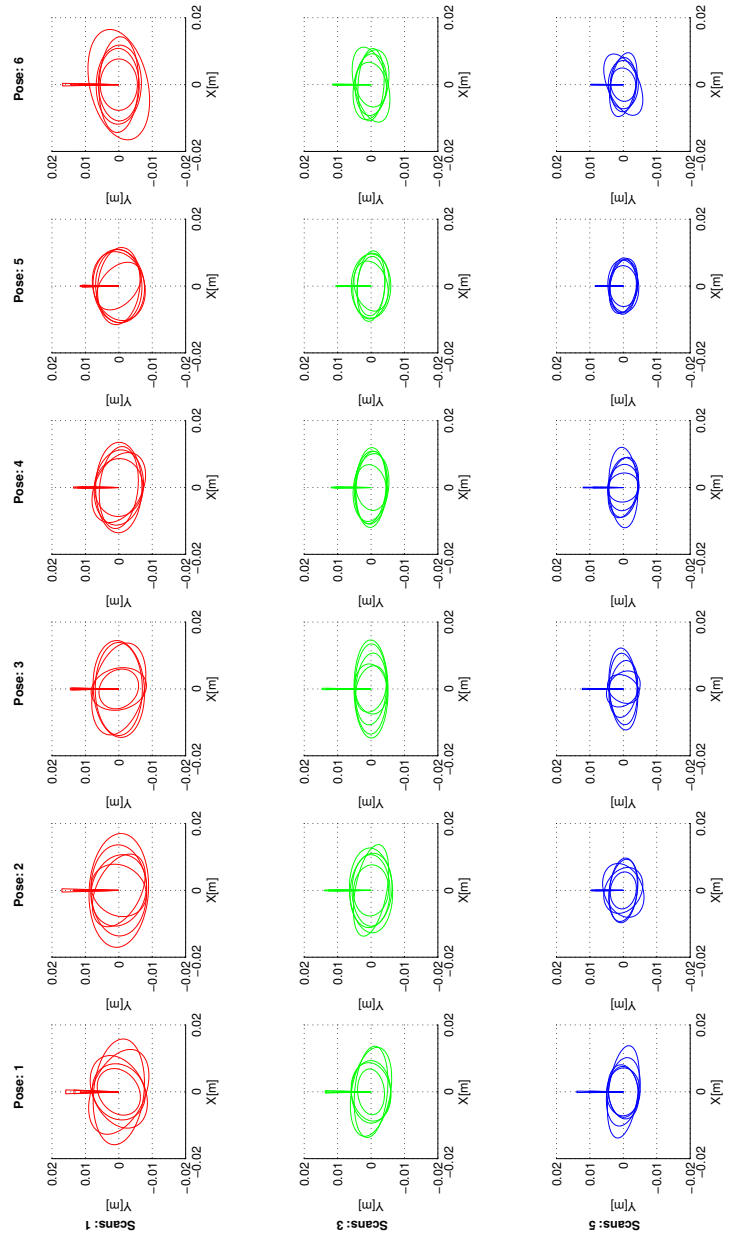


Figure A.1 : Covariances: Scans Vs Pose

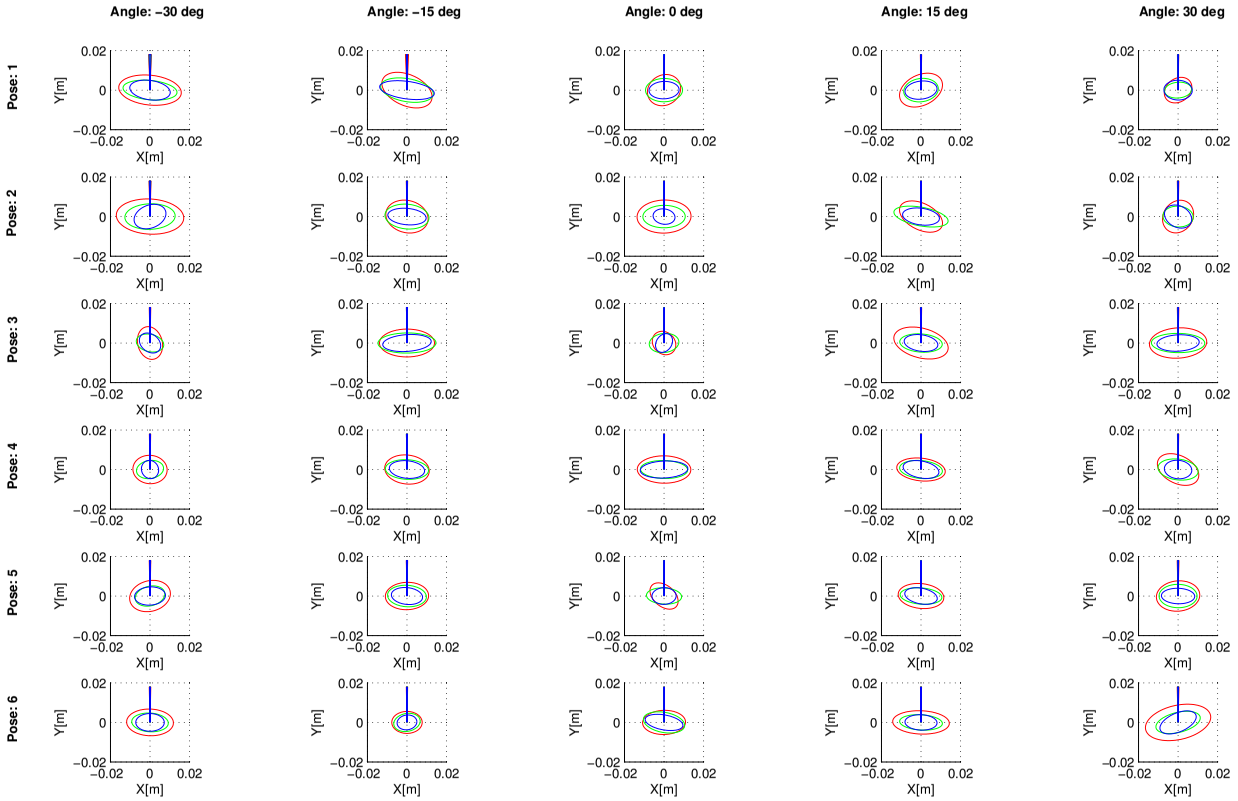


Figure A.2: Covariances: Angle Vs Pose

Results tables

Table A.1: 1 Scan Overview

Theoretical Id. Numb.	**3000**		
Requests	3315		
True Id. Numb.	2993	**Identification Ratio**	**99,77%**
False Id. Numb.	0	**False Detection Ratio**	**0,00%**
No Id.Numb.	315	**Unsuccessful Elaboration**	**10,50%**
Zero Id.Numb.	7		

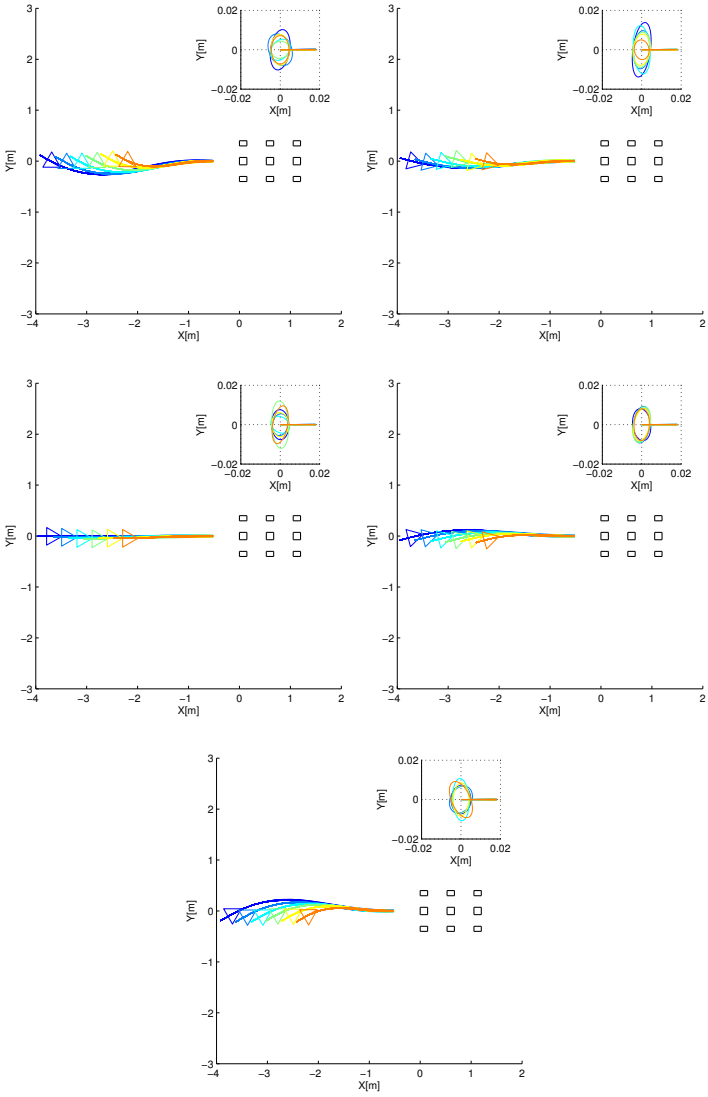
Table A.2: 3 Scan Overview

Theoretical Id. Numb.	**3000**		
Requests	3052		
True Id. Numb.	2998	**Identification Ratio**	**99,93%**
False Id. Numb.	0	**False Detection Ratio**	**0,00%**
No Id.Numb.	52	**Unsuccessful Elaboration**	**1,73%**
Zero Id.Numb.	2		

Table A.3: 5 Scan Overview

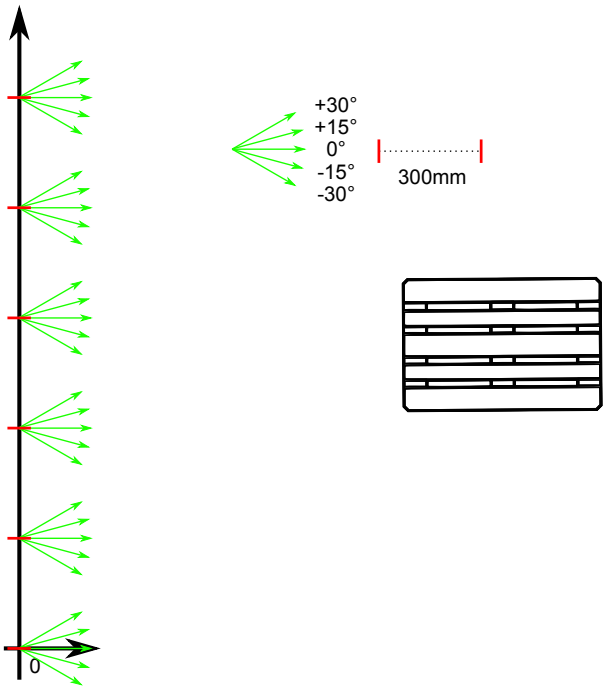
Theoretical Id. Numb.	**3000**		
Requests	3022		
True Id. Numb.	2998	**Identification Ratio**	**99,93%**
False Id. Numb.	0	**False Detection Ratio**	**0,00%**
No Id.Numb.	22	**Unsuccessful Elaboration**	**0,73%**
Zero Id.Numb.	2		

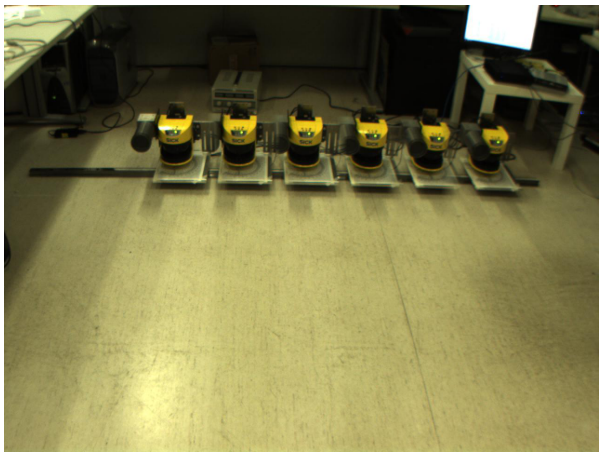
Paths



Transversal movement

Setup and covariances





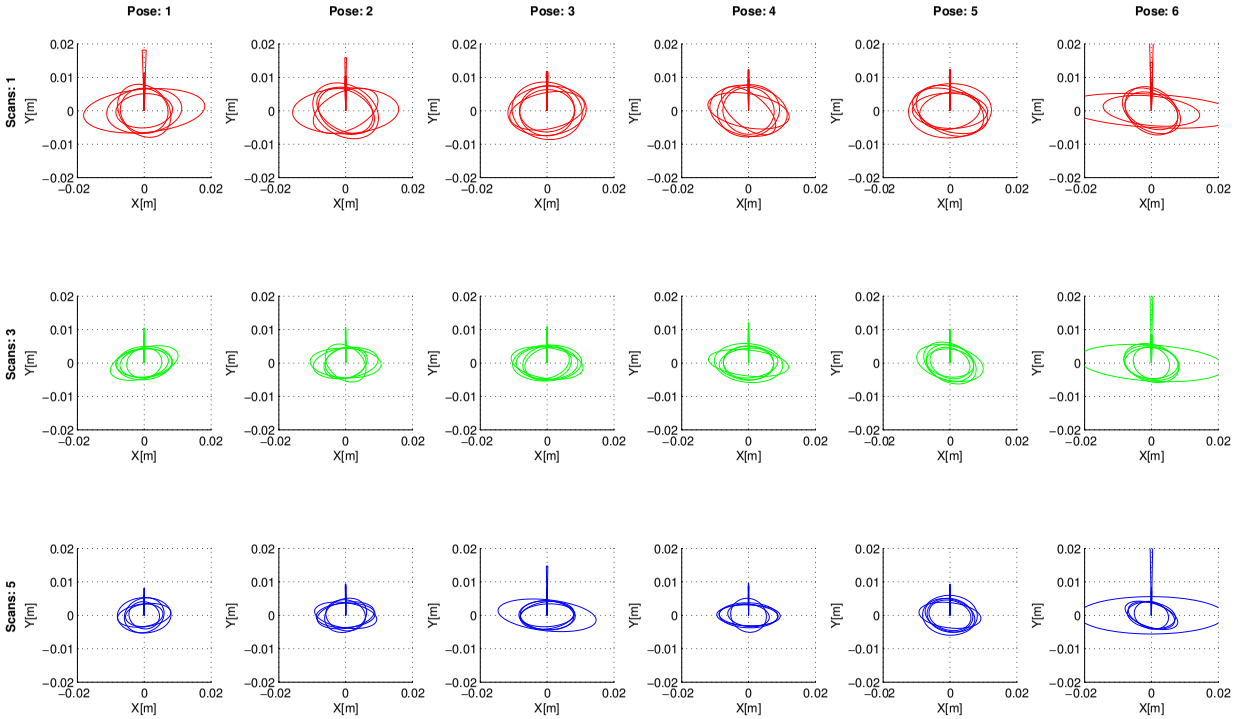


Figure A.3: Covariances: Scans Vs Pose

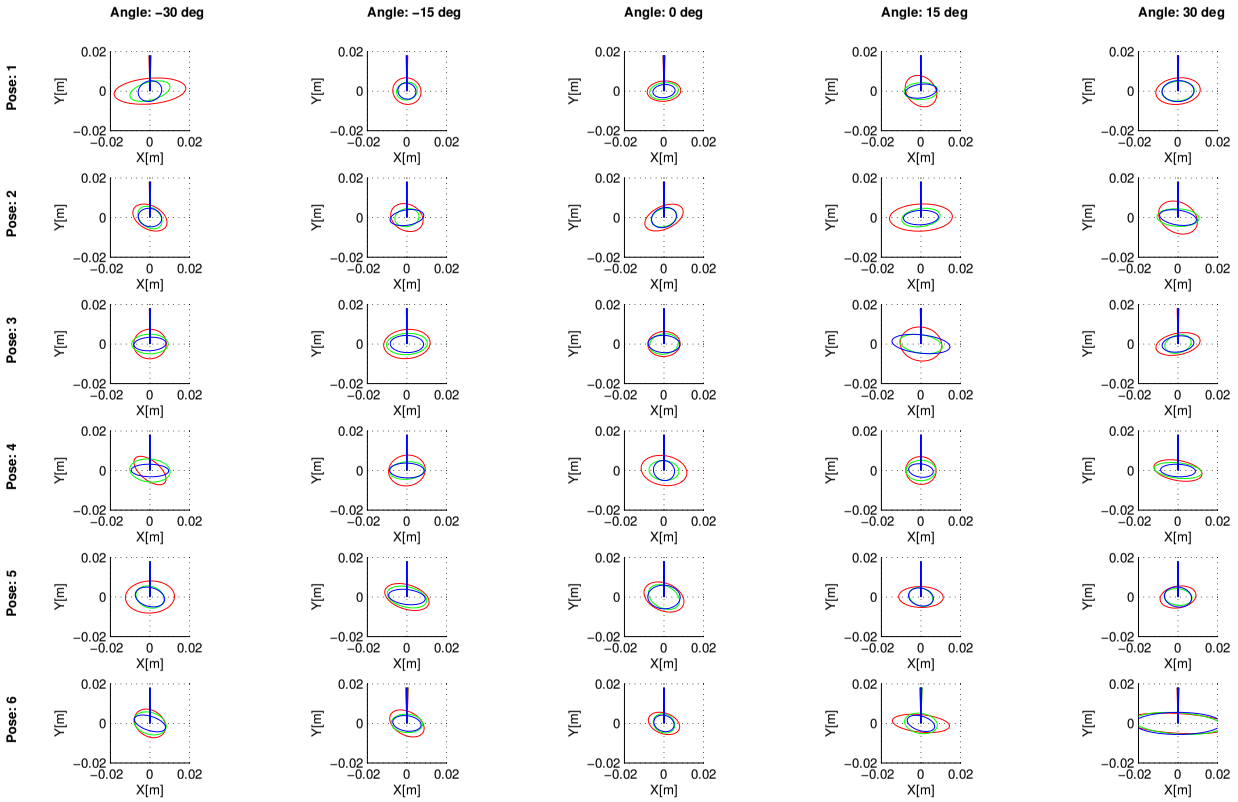


Figure A.4: Covariances: Angle Vs Pose

Results tables

Table A.4: 1 Scan Overview

Theoretical Id. Numb.	**3000**		
Requests	3291		
True Id. Numb.	2996	**Identification Ratio**	**99,87%**
False Id. Numb.	4	**False Detection Ratio**	**0,13%**
No Id.Numb.	291	**Unsuccessful Elaboration**	**9,70%**
Zero Id.Numb.	0		

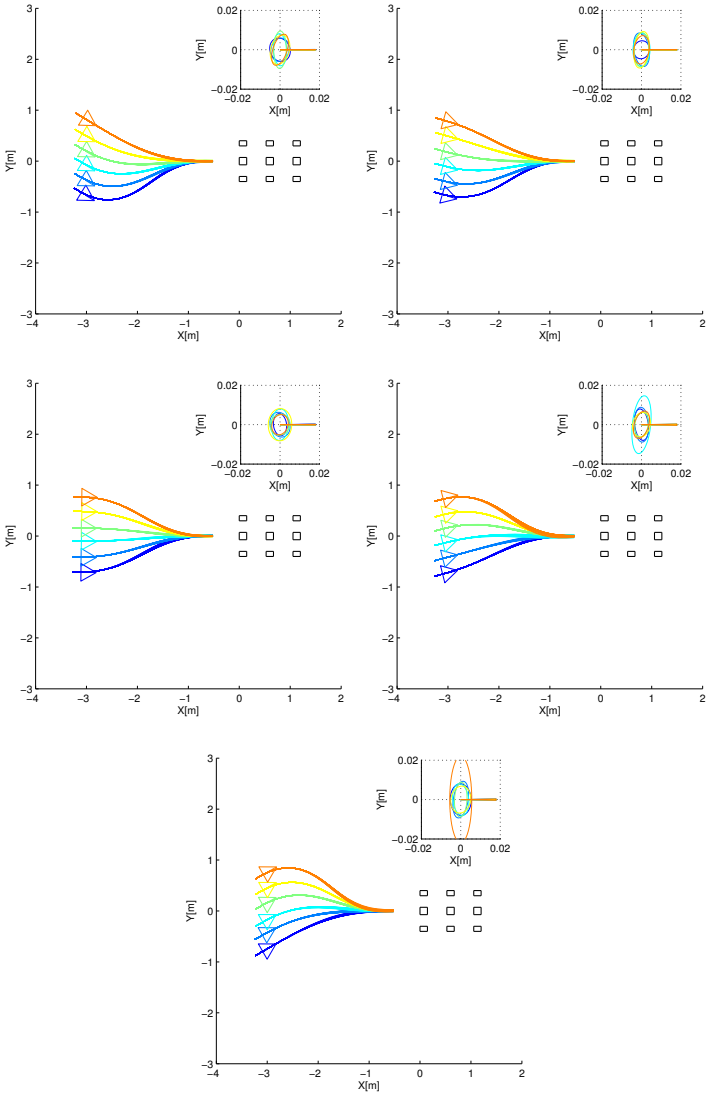
Table A.5: 3 Scan Overview

Theoretical Id. Numb.	**3000**		
Requests	3126		
True Id. Numb.	2999	**Identification Ratio**	**99,97%**
False Id. Numb.	0	**False Detection Ratio**	**0,00%**
No Id.Numb.	126	**Unsuccessful Elaboration**	**4,20%**
Zero Id.Numb.	1		

Table A.6: 5 Scan Overview

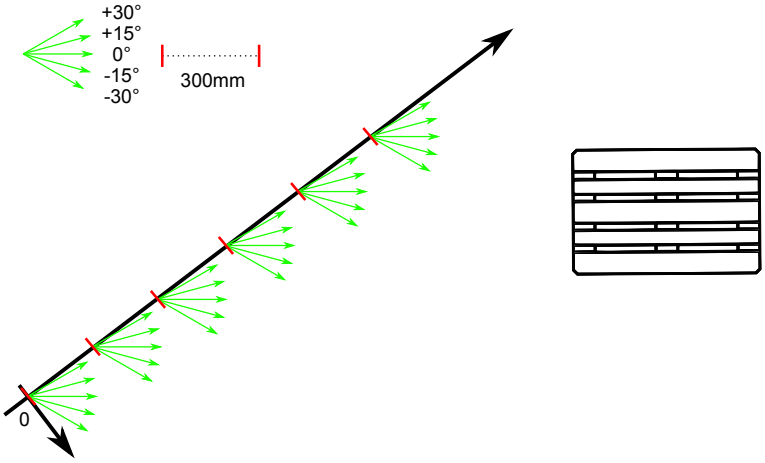
Theoretical Id. Numb.	**3000**		
Requests	3083		
True Id. Numb.	3000	**Identification Ratio**	**100,00%**
False Id. Numb.	0	**False Detection Ratio**	**0,00%**
No Id.Numb.	83	**Unsuccessful Elaboration**	**2,77%**
Zero Id.Numb.	0		

Paths



Diagonal movement

Setup and covariances



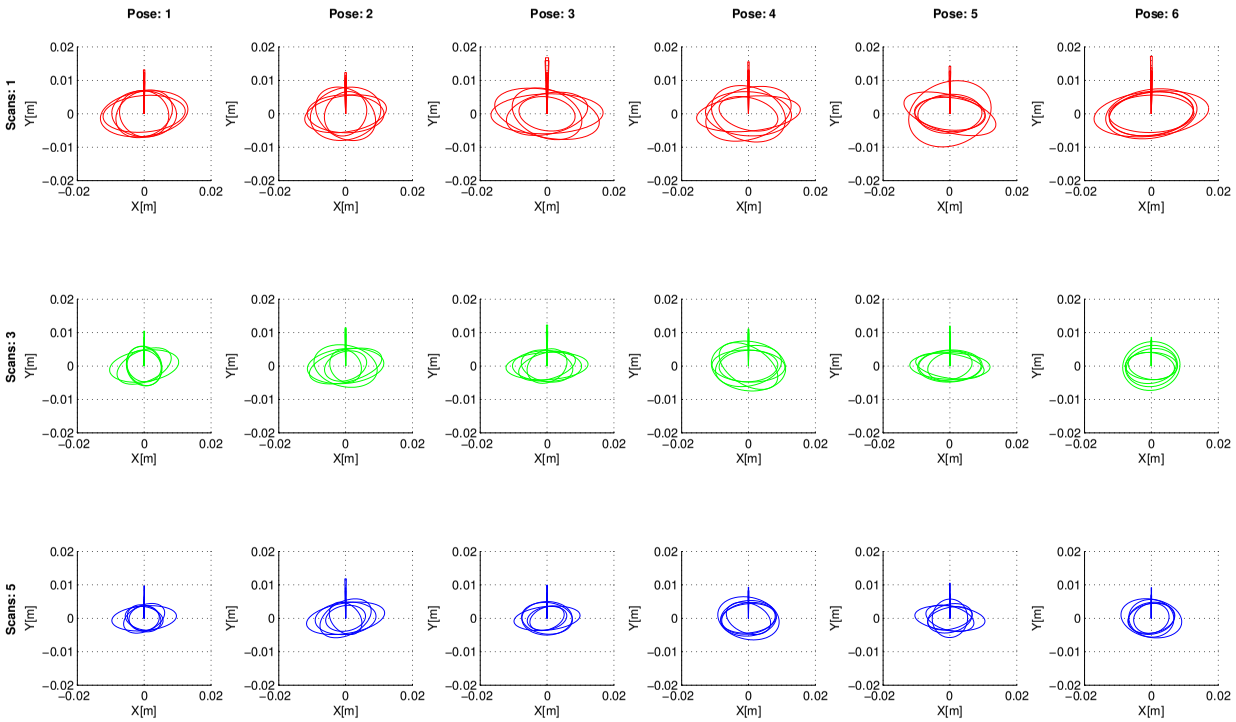


Figure A.5: Covariances: Scans Vs Pose

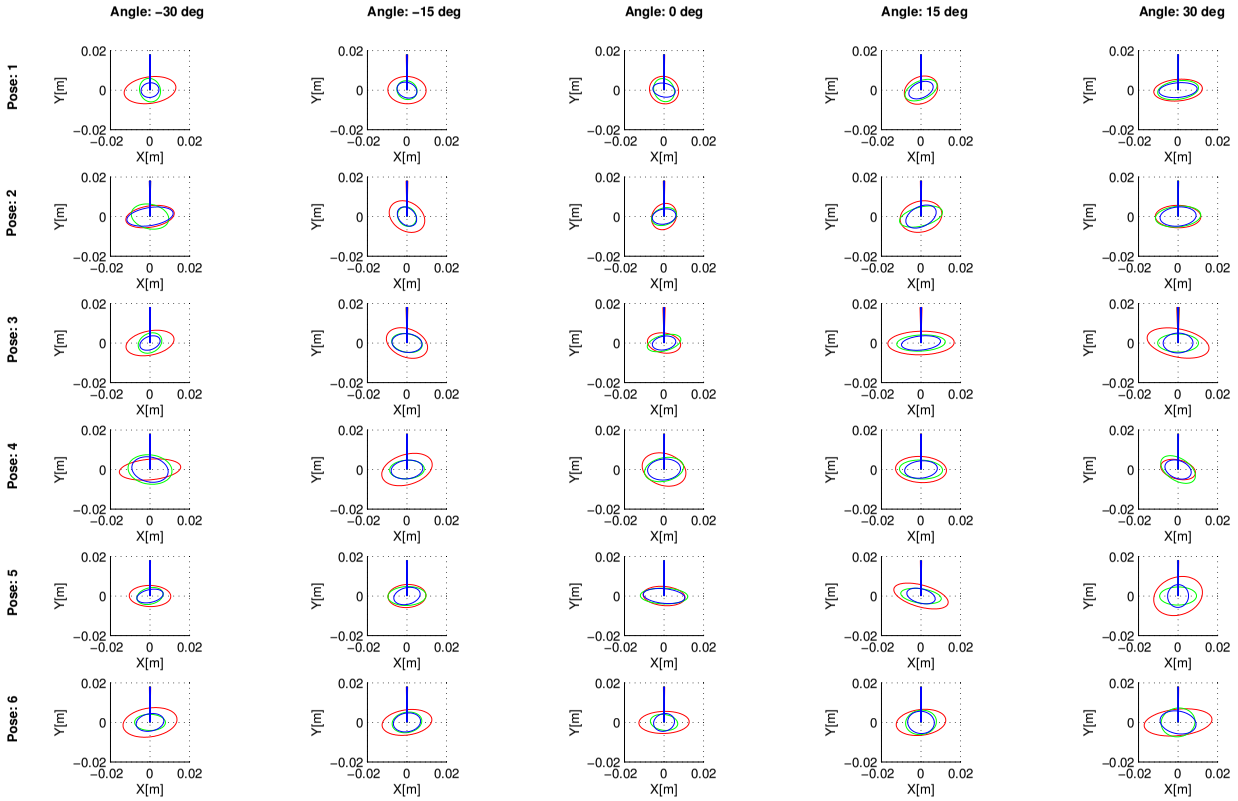


Figure A.6: Covariances: Angle Vs Pose

Results tables

Table A.7: 1 Scan Overview

Theoretical Id. Numb.	**3000**		
Requests	3307		
True Id. Numb.	2989	**Identification Ratio**	**99,63%**
False Id. Numb.	0	**False Detection Ratio**	**0,00%**
No Id.Numb.	307	**Unsuccessful Elaboration**	**10,23%**
Zero Id.Numb.	11		

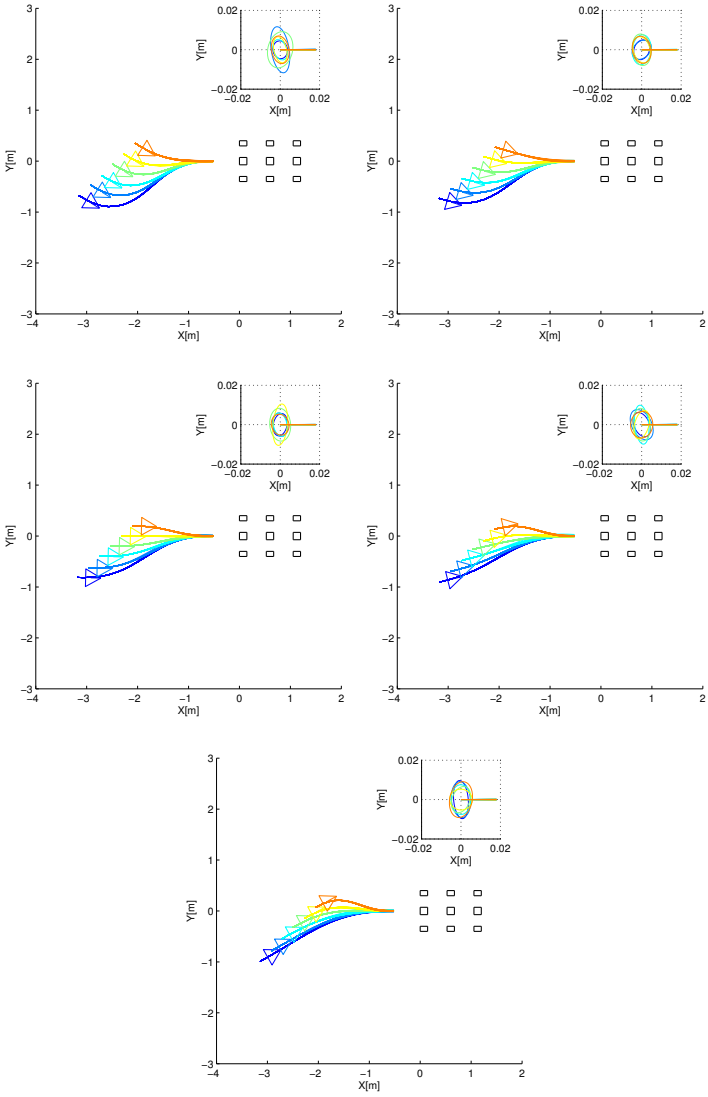
Table A.8: 3 Scan Overview

Theoretical Id. Numb.	**3000**		
Requests	3044		
True Id. Numb.	2997	**Identification Ratio**	**99,90%**
False Id. Numb.	0	**False Detection Ratio**	**0,00%**
No Id.Numb.	44	**Unsuccessful Elaboration**	**1,47%**
Zero Id.Numb.	3		

Table A.9: 5 Scan Overview

Theoretical Id. Numb.	**3000**		
Requests	3034		
True Id. Numb.	2998	**Identification Ratio**	**99,93%**
False Id. Numb.	0	**False Detection Ratio**	**0,00%**
No Id.Numb.	34	**Unsuccessful Elaboration**	**1,13%**
Zero Id.Numb.	2		

Paths



BIBLIOGRAPHY

Amarasinghe, Dilan; Mann, George K. I., and Gosine, Raymond G. Integrated laser-camera sensor for the detection and localization of landmarks for robotic applications, 2008.

Andrews, Stuart; Hofmann, Thomas, and Tsochantaridis, Ioannis. Multiple instance learning with generalized support vector machines, 2002.

Baglivo, L; Biasi, N; Biral, F; Bellomo, N; Bertolazzi, E; Da Lio, M, and De Cecco, M. Autonomous pallet localization and picking for industrial forklifts: a robust range and look method, 2011.

Baglivo, Luca; Bellomo, Nicolas; Miori, Giordano; Marcuzzi, Enrico; Pertile, Marco, and De Cecco, Mariolino. An object localization and reaching method for wheeled mobile robots using laser rangefinder, 2008.

Baglivo, Luca; Bellomo, Nicolas; Marcuzzi, Enrico; Pertile, Marco; Bertolazzi, Enrico, and De Cecco, Mariolino. Pallet pose estimation with lidar and vision for autonomous forklifts, 2009.

BALLARD, D.H. Generalizing the hough transform to detect arbitrary shapes, 1981.

Barrow, Harry G; Tenenbaum, Jay M; Bolles, Robert C, and Wolf, Helen C. Parametric correspondence and chamfer matching: Two new techniques for image matching, 1977.

Biasi, N. Riconscimento di un pallet attraverso uns sistema laser-camera, 2010.

Biegelbauer, Georg; Vincze, Markus, and Wohlkinger, Walter. Model-based 3d object detection, 2010.

Biswas, Joydeep and Veloso, Manuela. Depth camera based localization and navigation for indoor mobile robots, 2011.

Blanc, Christophe; Trassoudaine, Laurent, and Gallice, Jean. Ekf and particle filter track-to-track fusion: a quantitative comparison from radar/lidar obstacle tracks, 2005.

Bok, Yunsu; Choi, Dong-Geol; Jeong, Yekeun, and Kweon, In So. Capturing city-level scenes with a synchronized camera-laser fusion sensor, 2011.

Bostelman, Roger; Hong, Tsai, and Chang, Tommy. Visualization of pallets, 2006.

Bouguerra, A.; Andreasson, H.; Lilienthal, A.; Astrand, B., and Rognvaldsson, T. Malta: A system of multiple autonomous trucks for load transportation, 2009.

Bouguet, J. Y. Camera calibration toolbox for matlab, 2008.

Censi, Andrea. An accurate closed-form estimate of icp's covariance, 2007.

Chiabrando, Filiberto; Chiabrando, Roberto; Piatti, Dario, and Rinaudo, Fulvio. Sensors for 3d imaging: metric evaluation and calibration of a ccd/cmos time-of-flight camera, 2009.

Cucchiara, Rita; Piccardi, Massimo, and Prati, Andrea. Focus based feature extraction for pallet recognition, 2000.

Cui, Guang-zgao; Lu, Lin-sha; Yao, Li-na; Yanf, Cun-xiang; Huang, Bu-yi, and Hu, Zgi-hong. A robust autonomous mobile forklift pallet recongition, 2012.

Dalal, Navneet and Triggs, Bill. Histograms of oriented gradients for human detection, 2005.

De Cecco, Mariolino; Baglivo, Luca, and Angrilli, Francesco. Real-time uncertainty estimation of autonomous guided vehicle trajectory taking into account correlated and uncorrelated effects, 2007a.

De Cecco, Mariolino; Bertolazzi, Enrico; Miori, Giordano; Oboe, Roberto, and Baglivo, Luca. Pc-sliding for vehicles path planning and control-design and evaluation of robustness to parameters change and measurement uncertainty., 2007b.

Douillard, B.; Fox, D., and Ramos, F. Laser and vision based outdoor object mapping, 2009.

Felzenszwalb, Pedro F.; Girshick, Ross B.; McAllester, David, and Ramanan, Deva. Object detection with discriminatively trained part-based models, 2010.

Fischler, Martin A and Bolles, Robert C. Random sample consensus: a paradigm for model fitting with applications to image analysis and automated cartography, 1981.

Fornaser, Alberto. Taratura di sistemi robotici equipaggiati con laser range finder e telecamere, 2009.

Früh, Christian and Zakhor, Avideh. An automated method for large-scale, ground-based city model acquisition, 2004.

Fuchs, Stefan. Multipath interference compensation in time-of-flight camera images, 2010.

Fuchs, Stefan and Hirzinger, Gerd. Extrinsic and depth calibration of tof-cameras, 2008.

Garibotto, Giovanni; Masciangelo, Stefano; Ilic, Marco, and Bassino, Paolo. Robolift a vision guided autonomous fork-lift for pallet handling, 1996.

Garibotto, Giovanni; Masciangelo, Stefano; Ilic, Marco, and Bassino, Paolo. Service robotics in logistic automation: Robolift: Vision based autonomous navigation of a conventional fork-lift for pallet handling, 1997.

Gidel, S.; Blanc, C.; Chateau, T.; Checchin, P., and Trassoudaine, L. Non-parametric laser and video data fusion: Application to pedestrian detection in urban environment, 2009.

GIDEL, Samuel; CHECCHIN, Paul; BLANC, Christophe; CHATEAU, Thierry, and TRASSOUDAINE, Laurent. Parzen method for fusion of laserscanner data: Application to pedestrian detection, 2008.

GIDEL, Samuel; CHECCHIN, Paul; BLANC, Christophe; CHATEAU, Thierry, and TRASSOUDAINE, Laurent. A method based on multilayer laserscanner to detect and track pedestrians in urban environment, 2009.

Huttenlocher, Daniel P.; Klanderman, Gregory A., and Rucklidge, William J. Comparing images using the hausdorff distance, 1993.

Jinshi, Cui; Zha, Hongbin; Huijing, Zhao, and Ryosuke, Shibasak. Multi-modal tracking of people using laser scanners and video camera, 2008.

Kahlmann, Timo; Remondino, Fabio, and Ingensand, H. Calibration for increased accuracy of the range imaging camera swissrangertm, 2006.

Kanazawa, Yasushi and Kanatani, Ken-ichi. Do we really have to consider covariance matrices for image features?, 2001.

Karaman, Sertac; Walter, Matthew R.; Frazzoli, Emilio, and Teller, Seth. Closed-loop pallet engagement in an unstructured environment, 2010.

Kassir, Abdallah and Peynot, Thierry. Reliable automatic camera-laser calibration, 2010.

Kelly, Alonzo and Nagy, Bryan. Reactive nonholonomic trajectory generation via parametric optimal control, 2003.

Kleinert, Steffen and Overmeyer, L. Using 3d camera technology on forklift trucks for detecting pallets, 2012.

Koren, Yoram and Borenstein, Johann. Potential field methods and their inherent limitations for mobile robot navigation, 1991.

Lecking, Daniel; Wulf, Oliver; Viereck, Volker; T"odter, Joachim, and Wagner, Bernardo. The rts-still robotic fork-lift, 2005.

Lecking, Daniel; Wulf, Oliver, and Wagner, Bernardo. Variable pallet pick-up for automatic guided vehicles in industrial environments, 2006.

Li, Hao and Nashashibi, Fawzi. Comprehensive extrinsic calibration of a camera and a 2d laser scanner for a ground vehicle, 2013.

Lindner, Marvin; Schiller, Ingo; Kolb, Andreas, and Koch, Reinhard. Time-of-flight sensor calibration for accurate range sensing, 2010.

Linzmeier, D.T.; Skutek, M.; Mekhaiel, M., and Dietmayer, K.C.J. A pedestrian detection system based on thermophile and radar sensor data fusion, 2005.

Lo, Tsz-Wai Rachel and Siebert, J Paul. Local feature extraction and matching on range images: 2.5 d sift, 2009.

Mahalanobis, P.C. On the generalized distance in statistics, 1936.

Mian, Ajmal S; Bennamoun, Mohammed, and Owens, Robyn. Three-dimensional model-based object recognition and segmentation in cluttered scenes, 2006.

Monteiro, Goncalo; Premebida, Cristiano; Peixoto, Paulo, and Nunes, Urbano. Tracking and classification of dynamic obstacles using laser range finder and vision, 2006.

Nygards, J.; H"ogstrom.T., , and Wernersson, A. Docking to pallets with feedback from a sheet-of-light range camera, 2000.

Nygards, Jonas and Wernersson, Ake. On covariances for fusing laser rangers and vision with sensors onboard a moving robot, 1998.

Okada, R. Discriminative generalized hough transform for object dectection, 2009.

Oliveira, Luciano; Nunes, Urbano; Peixoto, Paulo; Silva, Marco, and Moita, Fernando. Semantic fusion of laser and vision in pedestrian detection, 2010.

Pagés, J.; Armangue, X.; Salvi, J.; Freixenet, J., and Marti, J. A computer vision system for autonomous forklift vehicles in industrial environments, 2001.

Papazov, Chavdar and Burschka, Darius. An efficient ransac for 3d object recognition in noisy and occluded scenes, 2011.

Papazov, Chavdar; Haddadin, Sami; Parusel, Sven; Krieger, Kai, and Burschka, Darius. Rigid 3d geometry matching for grasping of known objects in cluttered scenes, 2012.

Piatti, D. Time of flight cameras: tests, calibration and multi-frame registration for automatic 3d object reconstruction, 2011.

Pradalier, Cédric; Tews, Ashley, and Roberts, Jonathan M. Vision-based operations of a large industrial vehicle: Autonomous hot metal carrie, 2008.

Pratikakis, I; Spagnuolo, M; Theoharis, T, and Veltkamp, R. A robust 3d interest points detector based on harris operator, 2010.

Premebida, C.; Ludwig, O., and Nunes, U. Lidar and vision-based pedestrian detection system, 2009.

Ramos, F.T.; Nieto, J., and Durrant-Whyte, H.F. Recognising and modelling landmarks to close loops in outdoor slam, 2007.

Rousseeuw, Peter J. Least median of squares regression, 1984.

Rusu, Radu Bogdan and Cousins, Steve. 3d is here: Point cloud library (pcl), 2011.

Scaramuzza, Davide; Harati, Ahad, and Siegwart, Roland. Extrinsic self calibration of a camera and a 3d laser range finder from natural scenes, 2007.

Schnabel, Ruwen; Wahl, Roland, and Klein, Reinhard. Efficient ransac for point-cloud shape detection, 2007.

Seelinger, Michael and Yoder, John-David. Automatic pallet engagment by a vision guided forklift, 2005.

Seelinger, Michael and Yoder, John-David. Automatic visual guidance of a forklift engaging a pallet, 2006.

Smith, Randall C and Cheeseman, Peter. On the representation and estimation of spatial uncertainty, 1986.

Sungmin, Byun and Minhwan, Kim. Real-time positioning and orienting of pallets based on monocular vision, 2008.

Susnea, Ioan; Minzu, Viorel, and Vasiliu, Grigore. Simple, real-time obstacle avoidance algorithm for mobile robots, 2009.

Szarvas, M.; Sakai, U., and Ogata, J. Real-time pedestrian detection using lidar and convolutional neural networks, 2006.

Tavernini, Mattia. Study and application of motion measurement methods by means of opto-electronics systems, 2013.

Teller, Seth; Walter, Matthew; Antone, Matthew; Correa, Andrew; Davis, Randy; Fletcher, Luke; Frazzoli, Emilio; Glass, Jim; How, Jonathan; Huang, Albert S.; Jeon, Jeong Hwan; Karaman, Sertac; Luders, Brandon; Roy, Nicholas, and Sainath, Tania. A voice-commandable robotic forklift working alongside humans in minimally-prepared outdoor environments, 2010.

Vasconcelos, Francisco; Barreto, Joao P., and Nunes, Ubano. A minimal solution for the extrinsic calibration of a camera and a laser-rangefinder, 2012.

Von Wahlde, Raymond; Wiedenman, Nathan; Brown, Wesley A, and Viqueira, Cezarina. An open-path obstacle avoidance algorithm using scanning laser range data, 2009.

Wang, Jianguo Jack; Hu, Gibson; Huang, Shoudong, and Dissanayake, Gamini. 3d landmarks extraction from a range imager data for slam, 2009.

Weichert, Frank; Skibinski, Sebastian; Stenzel, Jonas; Prasse, Christian; Kamagaew, Andreas; Rudak, Bartholomäus, and ten Hompel, Michael. Automated detection of euro pallet loads by interpreting pmd camera depth images, 2013.

Weyer, Christoph A; Bae, Kwang-Ho; Lim, Kwanthar, and Lichti, D. Extensive metric performance evaluation of a 3d range camera, 2008.

Winkelbach, Simon; Molkenstruck, Sven, and Wahl, Friedrich M. Low-cost laser range scanner and fast surface registration approach, 2006.

Zhang, Q. and Pless, R. Extrinsic calibration of a camera and laser range finder (improves camera calibration), 2004.

Zhen, Chen; Liang, Zhuo; Kaiqiong, Sun; , and Congxuan, Zhang. Extrinsic calibration of a camera and a laser range finder using point to line constraint, 2012.

Zhendong, He; Xinjin, Wan; Jie, Liu; Junman, Sun, and Guangzhao, Cui. Feature-to-feature based laser scan matching for pallet recognition, 2010.

Zisserman, A. and Hartley, R. Multiple view geometry in computer vision, 2000.