UNIVERSITÀ DEGLI STUDI DI PARMA

Dottorato di Ricerca in Tecnologie dell'Informazione

XXV Ciclo

LASER AND CAMERA INTERCALIBRATION TECHNIQUES FOR MULTI-SENSORIZED VEHICLES

Coordinatore:

Chiar.mo Prof. Marco Locatelli

Tutor:

Chiar.mo Prof. Alberto Broggi

Dottorando: *Matteo Panciroli*

Gennaio 2013

“Everything we hear is an opinion, not a fact.
Everything we see is a perspective, not the truth”

Contents

List of Figures

List of Tables

Chapter 1

Introduction

1.1 Overview

Throughout the years, enormous progress has been made by researchers and automotive suppliers on intelligent vehicle technology, such that vehicles have evolved to become sophisticated technological machines that extend mobility to leisure, comfort, luxury, sports, and, for some, an extension and expression of their image and personality. In one hand, human mobility and the way people live today are completely revolutionized by such vehicles, in the other hand today's vehicles have many sensors and electronic systems to manage that contribute to automatic control of subsystems for a range of functions from controlling vehicle dynamics (i.e.: cruise control) to supporting the driver in trip planning and route selection (e.g., navigation systems)' [1]. Taking into account these two opposing viewpoints, vehicle safety remains a leading topic for intelligent vehicles research, starting from the incorporation of seat belts and airbags from the 1960s up to active safety systems such as brake assistant and traction control, several passive and active systems have been developed contributing to improve safety.

Nowadays, trends of recent technology of active safety systems are moved toward perception-based architectures such as adaptive cruise control, blind spot detection, lane departure detection and pedestrian (obstacles) detection. All these systems are

also known as “advanced driver assistance systems” (ADAS) a set of handling functions that an intelligent vehicle provides to support the driver on the road avoiding human errors. Since ADAS attempt to substitute the driver by sensing the world and taking actions as human, these systems have to be more reliable than a human operator can be. This research is set within the context of modern intelligent vehicles, describing basic methodology to “calibrate” the perception system of such a vehicles needed by every ADAS.

The key world of this PhD thesis will be the generic concept of *calibration*, that will be introduced in chapter 3 than deeply investigated in the next chapters. To introduce the generic therm *calibration*, lets consider a generic definition:

> “operation that, under specified conditions, in a first step, establishes a relation between the quantity values with measurement uncertainties provided by measurement standards and corresponding indications with associated measurement uncertainties and, in a second step, uses this information to establish a relation for obtaining a measurement result from an indication” - *JCM 200:2008 International Vocabulary of metrology*

In other world the calibration task is a comparison between measurements. Where a first measure is made or set with a device that has a known (or assigned) correctness, usually called *standard* device, and a second measure made, in as similar a way as possible, with a second device called *unit under test*.

- NOTE 1 A calibration may be expressed by a statement, calibration function, calibration diagram, calibration curve, or calibration table. In some cases, it may consist of an additive or multiplicative correction of the indication with associated measurement uncertainty.
- NOTE 2 Calibration should not be confused with adjustment of a measuring system, often mistakenly called self-calibration, nor with verification of calibration.
- NOTE 3 Often, the first step alone in the above definition is perceived as being calibration.

1.2 Motivations and Objective

Motivation on this research resides in recent autonomous vehicle project. Studies on the sensors calibration problem are part of the European project *"Open intelligent systems for future autonomous vehicles"* (OFAV).

The research is based on the extended know-how and experience of the University of Parma, which already marked fundamental milestones worldwide in the field of vehicular robotics.

The OFAV objective is the development of an open architecture for future autonomous vehicles to become a standard platform shared by car makers in the design of next generation intelligent vehicles. It is based on 360 degrees sensorial suite which includes perceptual and decision making modules, with the ultimate goal of providing the vehicle with autonomous driving capabilities and/or supervise the driver's behavior. The perception module also includes vehicle-to-vehicle and vehicle-to-infrastructure subsystems, to increment the vehicles sensing capabilities.

Providing clear advantages on safety for road users, the availability of an open architecture will encourage and make possible the sharing of knowledge between public and private research communities (academic and automotive industry) and thus speed up the design of a standard platform for future vehicles. Further research steps will be eased -and therefore made more effective- thanks to the common and open architectural layer proposed by this project.

Preliminary studies on the sensors calibration problem are part of the European project and had been yield during the *VisLab Intercontinental Autonomous Challenge* (VIAC, briefly described in appendix B), where intelligent vehicles have been driven unmanned only for short periods of time. There are two fundamental motivations that lead this research:

- Firstly, only recent intelligent vehicle adopt sensor fusion to improve their ADAS robustness and safety, furthermore with the complementary purpose of coupling camera and laser together, such applications need to achieve 3D environment information from each sensor in fact of that high accuracy between sensor pose must be kept.

- Then, requirement of automatic calibration procedures could let modern vehicles maintain their own perception system.

The aim of this thesis is to describe a set of method and procedure used in the automotive field to calibrate vision and range sensors -common equipment suits of intelligent vehicles- to reach an accurate perception of the surrounding environment and fill the lack of maintenance present in some vehicle prototypes.

1.3 Outlines of the thesis

The thesis have been structured as follows:

- Chapter 2: gives an overview of modern perception system on intelligent vehicles, describing laser range finders and vision systems capabilities.
- Chapter 3: illustrates the available calibration techniques for camera and laser sensors. Further details on maker detection are detailed in this chapter.
- Chapter 4: shows the basic idea of sensor cross-calibration and describes the use of laser and sensor fusion to cross calibrate a vision system. Previous works are described and a new approach is presented.
- Chapter 5: modular calibration procedure proposed for VIAC prototypes involved equipment and maintenance phases.
- Chapter 6: conclusive discussion on the work and analysis on future perspectives.

Chapter 2

Perception Sensors on Intelligent Vehicles

Today, vehicles are equipped with more electronic systems than in the past years, a great quantity of sensing and actuating systems are suitable, such as temperature, tire pressure, accelerometer, and speed sensors, most of them already present on the market. On the other hand, the market penetration rate of some perception sensors, except for ultrasonic sensors, is very low mainly because of their cost. Modern 360° LIDAR or GPS inertial unit could be very expensive in some cases and present only in some prototypes. However, environmental perception systems for almost mobile robot are based on sonar, radar, LIDAR and vision sensors.

In this chapter, principal perception sensors designer for obstacle detection and navigation of autonomous ground vehicles are described. The purpose of this chapter is to give an overview on the sensors used by ADAS and go deeply into detail about the math behind such systems. The selection of these perception sensors is grouped into two different categories: vision (or passive) sensors, and laser scanner (or active) sensors that are the main parts of this research.

2.1 Cameras

Vision perception remain one of the most investigated process in computer science research and it is one of the fastest growing technology being deployed in vehicles for a wide range of function such as general obstacles and lane detection. Commonly, vision systems are based on digital image and there are several sensors used to acquire such information, these sensors can be charge-coupled device (CCD), CMOS, infrared and near infrared (IR and NIR), and stereo vision systems. Among these different sensors two types of image can be acquires:

- *Intensity image* that represent the common photo, it measure the amount of light that impress a planar photosensitive device. All intensity images are treated as 2D arrays of integer value.
- *Range images* these images encoding environment objects in distances, the distance measure can be computed by a stereo rig even with modern Time of Flight (ToF) cameras (that are similar to a laser scanner with global shutter).

From these two different images derives two fundamental consequences [2]:

- The relationship between digital images (range images or intensity images) and the physical world is determined by the acquisition process, which depends on the sensor used.
- Any shapes, distance measure, or objects identity inside images become unavoidably intrinsic information that have to be extract (computed) from 2D numerical arrays in which they are stored.

Cameras are the cheapest and most common sensor used to perceive a 3D world through intensity or range images. However, to recover world information there are several physical parameters that describe the analytical model.

Before introducing the calibration problem in this section the physical model of the camera is described to introduce some notes and notations on the camera model. There are several physical parameters that playing a role during image formation, from world to the 2D numerical array, consider:

- *Photometric parameters* that describes intensity, quantity, direction and reflection of the light that from objects in the scene reach the sensor.
- *Optical parameters of the lens* mounted on cameras to converge ray of light on the photosensitive are described by their focal length, aperture, shutter and field of view.
- *Geometric parameters* to compute the relationship between image point and world point, include type of projections, position and orientation of camera into the world (also called pose), and perspective distortion introduced by image process. Moreover, lens distortion factor have to be considered.
- *Acquisition parameters* that characterize the digital images based on the physical properties of the 2D matrix. They are: image jitter, scale factor (or pixel size), quantization and dynamic of the sensors.

2.1.1 Projective Geometry

Observing a 3D space through a camera, than using a CCD sensor, has a direct consequence of reducing information to a 2D space. This process is defined as *perspective projection*. In this section notation on the perspective (even known as *pinhole* camera model) is given.

The perspective model, shown in figure 2.1.1, is composed of a *image plane* $\mathscr{R}$ and a point O_C, the *center of projection*. The distance f between I and point C is the *focal length* and the line through C and orthogonal to I is the *optical axis*. The intersection of the optical axis and the image plane is defined as *principal point*. To analytical describe the perspective camera model two reference frames have to be defined[1] . First of all a *camera frame* $\{X_C, Y_C, Z_C\}$ is defined to represent 3D points of the world observed by the camera, it is centered in O_C with unit vector Z coincident to the optical axes. The second frame $\{u, v\}$ is placed on the image plane with axes u

[1]The camera and sensor frames adopted in this PhD thesis are one of the several possible solutions that are present in literature. This representation remain one of the simplest representation that is used to describe notations and assumptions

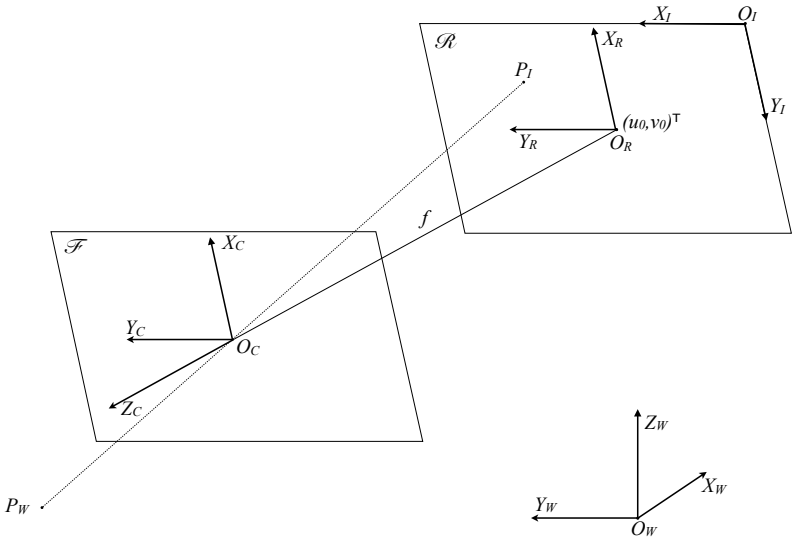


Figure 2.1: Pin-Hole model. $\{O_W\}$ = world reference system, $\{O_C\}$ = camera reference system, $\{O_R\}$ = retina reference system, $\{O_I\}$ = image reference system (axis $(X_I, Y_I) \equiv (u, v)$). $\mathscr{F}$ is the focal plane and $\mathscr{R}$ the image (or retina) plane.

and v oriented like X_I and Y_I respectively. Than, given two point $P = [X, Y, Z]^\top$ in the *camera frame* and $p = [u, v]^\top$ in the image frame its projection on the image plane the perspective projection is derived

$$u = \frac{-f}{Z}X \qquad v = \frac{-f}{Z}Y \tag{2.1}$$

The minus sign represents the coordinates reflection across O_C and Z_C component introduces a non linear component in the model. The non linear system 2.1 can be represented with homogeneous coordinates to become linear as shown in 2.2 . Than,

given $\tilde{P} = [X,Y,Z,1]^\top$ and $\tilde{p} = [u,v,1]^\top$ the perspective projection model become

$$Z\begin{bmatrix} u \\ v \\ 1 \end{bmatrix} = \begin{bmatrix} -fX \\ -fY \\ Z \end{bmatrix} = \begin{bmatrix} -f & 0 & 0 & 0 \\ 0 & -f & 0 & 0 \\ 0 & 0 & 1 & 0 \end{bmatrix} \begin{bmatrix} X \\ X \\ Z \\ 1 \end{bmatrix} \tag{2.2}$$

with matrix notation,

$$\tilde{p} \approx M\tilde{P} \tag{2.3}$$

Equation 2.3 is defined except for a scale factor Z and represent the perspective projection matrix of a camera. Note that a realistic camera model that can precisely map 3D points on the image plane must consider the sensor pose with respect of the world frame (described in 2.1.4) and CCD physics (detailed in 2.1.3).

2.1.2 Lens distortion

Usually, optics employed on intelligent vehicles are wide-angle or have short focal length (e.g., a 6mm focal length for lane detection applications or 3.6mm for lateral views), hence a more accurate camera model have to deal with image distortion introduced by such optics, that become evident at the periphery of the image with a non-linear effect. There are several lens models described in literature to describe analytically the distortion effect. Under normal condition e.g., narrow field-of-view (FOV) and non fish-eye optics the standard model used for such lens is the polynomial model of radial distortion also called *radial distortion model* described in [3] as follows in equation 2.4

$$\begin{aligned} r_d &= r_u + \textstyle\sum_{i=1}^{\infty} k_i r_u^{2i+1} \\ &= r_u + k_1 r_u^3 + \cdots + k_i r_u^{2i+1} + \ldots \end{aligned} \tag{2.4}$$

another representation is the *Division Model* shown in equation 2.5

$$r_u = \frac{r_d}{1 + \sum_{i=1}^{\infty} k_i r_d^{2i}} \tag{2.5}$$

where r_d is the distorted radius with respect a *center of distortion* (u_{d0}, v_{d0}), while r_u is the undistorted radius, and k_i are the model coefficients. Considering the whole

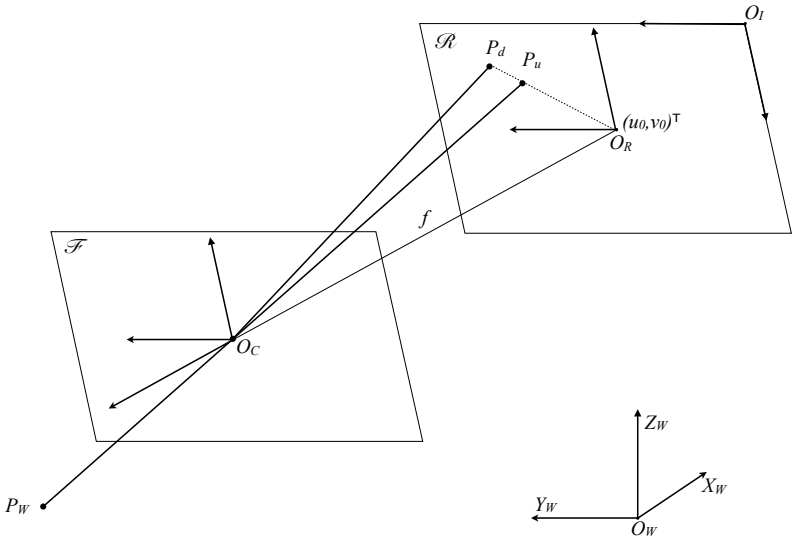


Figure 2.2: Radial distortion component apply to the Pin-Hole model. Distorted radius $r_d = \overline{O_R P_d}$ and undistorted radius $r_u = \overline{O_R P_d}$, O_R represent the distortion center.

distorted image, (u_{d0}, v_{d0}) is the singular point where no distortion effect is present, thus $(u_{d0}, v_{d0}) \equiv (u_u, v_u)$. Usually the distortion center point does not correspond to the principal point (u_0, v_0), even if they can be approximated. To be able to define a simple distortion model a new reference frame have to be considered with respect the center of distortion such that

$$\begin{cases} \overline{u} = u - u_{d0} \\ \overline{v} = v - v_{d0} \end{cases} \tag{2.6}$$

and a distortion effect can be modeled as displacement from center of distortion in u, v components 2.7 as described in Brown-Conrady [4]

$$\begin{cases} u_u = u_d f_d(r_d) \\ v_u = v_d f_d(r_d) \end{cases} \tag{2.7}$$

with

$$r_d^2 = u_d^2 + v_d^2 \tag{2.8}$$

(u_d, v_d) are coordinates of the distorted point. As shown in equation 2.7 the radial displacement is computed by a function f_d which, with the center of distortion, models the lens distortion (e.g., even function expressed by a Taylor series, such that can model only symmetrical distortion with respect the center of distortion). $r_d^2 = u_d^2 + v_d^2$ is the radius of point (u_d, v_d) used to compute the radial distortion, the function f_d depends only by r_d^2 in the form

$$f_d(r^2) = 1 + k_1 r^2 + k_2 r^4 + \dots \tag{2.9}$$

While polynomial models are easy to be implemented and embedded it has -as every polynomial model- the practical problem that there is no analytical method to invert 2.4.

Further representation of optics distortion consider radial, tangential and prismatic distortion with distortion center is detailed in equations 2.10

$$\begin{aligned}
\overline{u} &= u - u_{d0} \\
\overline{v} &= v - v_{d0} \\
r &= \sqrt{\overline{u}^2 + \overline{v}^2} \\
\tilde{u} &= (p_0 + r^2 p_2)(r^2 + 2\overline{u}^2) + 2(p_1 + r^2 p_3)\overline{uv} + (1 + k_1 r^2 + k_2 r^4)\overline{u} + s_0 r^2 + s_2 r^4 \\
\tilde{v} &= (p_1 + r^2 p_3)(r^2 + 2\overline{v}^2) + 2(p_0 + r^2 p_2)\overline{uv} + (1 + k_1 r^2 + k_2 r^4)\overline{v} + s_1 r^2 + s_3 r^4
\end{aligned} \tag{2.10}$$

where

- Radial distortion parameters are: k_1, k_2.
- Tangential distortion are described by p_0, p_1, p_2, p_3.
- Thin prism distortion described by s_0, s_1, s_2, s_3 model the possible displacement between optic lens and sensor.

As mentioned 2.4 and 2.5 are used for automotive application since they can describe distortion in standard non fish-eye optics. However they can not model sever distortion and it needs more degree of freedoms. Therefore, a first polynomial approach to describe fish-eye optics is described in [5, 6] by equation 2.11

$$r_d = \sum_{i=1}^{\infty} k_i r_u^i \tag{2.11}$$

Fish-eye model not based on polynomial models benefit of an analytical inverse representation and can model strong non-linear effect in the peripheral area of the image. Basu and Licardie proposed a *Fish-Eye Transform* in [7] with equation

$$r_d = s\ln(1+\lambda r_u) \tag{2.12}$$

the logarithmic function model the behavior of a fish-eye that has high density of information in the foveal area of the image and it decrease toward the border of the image with an exponential trend. Th inverse of this model is given by

$$r_u = \frac{e^{\frac{r_d}{s}} - 1}{\lambda} \tag{2.13}$$

Devernay and Faugeras in [8] describe the *Field-of-View* model based on the fish-eye model 2.14.

$$r_d = \frac{1}{\omega}\arctan\left(2r_u \tan\frac{\omega}{2}\right) \tag{2.14}$$

with inverse 2.15

$$r_u = \frac{\tan(r_d\omega)}{2\tan(\frac{\omega}{2})} \tag{2.15}$$

where ω represents the angular field of view of the ideal fish-eye optic. In fact of that, in their outcomes had shown that this model may not be always enough accurate to model the complex distortion.

2.1.3 Intrinsic parameters

Intrinsic parameters link world point in the camera reference frame with the corresponding pixel coordinate of an image

The intrinsic parameters characterize the inner components of a camera and can be defined in two main sets:

- Pinhole model: which is composed by the *focal length* to model the perspective projection and the transformation parameters that transform coordinates from camera frame to pixel in image reference frame.
- Geometric distortion model: which is used to evaluate distortion introduced by lens.

With respect figure 2.1.1, conversion from camera point P_W to pixel p in image plane O_I is computed by an affine transformation that consider the image center and a scale factor on axis u and v that depends on the pixel size. The resulting relation follow in 2.16[2]

$$\begin{cases} u = k_u \frac{-f}{z} x + u_0 \\ v = k_v \frac{-f}{z} y + v_0 \end{cases} \tag{2.16}$$

where (u_0, v_0) describe the optical center of the image (also called principal point), and $k_u(k_v)$ represent the effective size of the pixel in $[pixel \cdot m^{-1}]$ in the horizontal (vertical) direction respectively. A more general camera model consider one more parameter θ that model the angle between u and v axis (usually assumed $\theta = \pi/2$). It is also called *skew*. The matrix representation of the intrinsic parameters used in this PhD thesis is described by a matrix $\mathbf{K}$

$$\mathbf{K} = \begin{bmatrix} -fk_u & fk_u \cot\theta & u_0 \\ 0 & -fk_v/\sin\theta & v_0 \\ 0 & 0 & 1 \end{bmatrix} = \begin{bmatrix} \alpha & \gamma & u_0 \\ 0 & \beta & v_0 \\ 0 & 0 & 1 \end{bmatrix} \tag{2.17}$$

[2] X_I and ,Y_I axis represent metric units, while u and v axis represent pixel units.

2.1.4 Extrinsic parameters

The *extrinsic parameters* define the camera reference frame also called *pose* with respect to a well known reference frame (i.e., a generic vehicle reference frame, or more in general a world reference frame).

Usually, the camera pose is used to determine distance information through fundamental equations of the perspective projection described in equation 2.2 and depicted in figure 2.1.1, despite the camera pose is unknown or measured under uncertainty, it gives geometrical information with respect the sensor reference frame. The main problem in the calibration procedure is to compute the geometric transformation that uniquely represent the camera frame with respect to a known reference frame. The most common way to represent such transformation between camera and vehicle reference frames is to use matrix representation

$$\mathbf{T} = [\, \mathbf{R}_\Theta \mid \mathbf{t} \,] \tag{2.18}$$

where a 3×4 transformation matrix $\mathbf{T}$ is composed by:

- a translation vector $\mathbf{t}$ describing the relative position from the vehicle reference frame origin to the camera reference frame origin;
- a 3×3 rotation matrix $\mathbf{R}$ that is an orthogonal matrix which is usually represented in the automotive field by Euler angles in a *roll-pitch-yaw matrix*. Hence, a rotation matrix can be represented by a rotation vector.

$$\Theta = [roll, pitch, yaw]^\top$$

A relation between coordinates of a generic point $\mathbf{P} = [X, Y, Z]^\top$ in vehicle reference frame and a camera reference frame is computed by 2.18 as shown in figure 2.3.

$$\mathbf{T} = [\, \mathbf{R}_\Theta \mid {}^v\mathbf{t}_c \,] \tag{2.19}$$

With respect to figure 2.3 the point $\mathbf{P}$ can be represented by ${}^v\mathbf{p}$

$${}^v\mathbf{p} = {}^v\,\mathbf{t}_c + {}^v\,\mathbf{R}_c\ {}^c\mathbf{p} \tag{2.20}$$

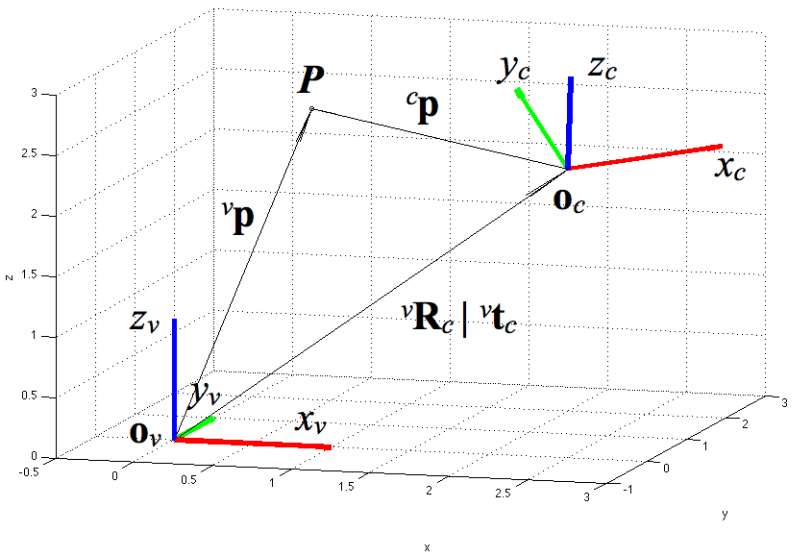


Figure 2.3: Vehicle $\mathbf{O}_v$ and camera $\mathbf{O}_c$ reference frame relations.

either with inverse transformation

$$^{c}\mathbf{p} = -^{c}\mathbf{R}_v \, ^{v}\mathbf{t}_c + ^{c}\mathbf{R}_v \, ^{v}\mathbf{p} \tag{2.21}$$

through a matrix representation and using homogeneous vectors $^{v}\tilde{\mathbf{p}} = [\mathbf{p},\ 1]^{\top}$ the homogeneous transformation matrix becomes

$$^{v}\mathbf{T}_c = \begin{bmatrix} ^{v}\mathbf{R}_c & ^{v}\mathbf{t}_c \\ \mathbf{0}^{\top} & 1 \end{bmatrix} \tag{2.22}$$

such that

$$^{v}\tilde{\mathbf{p}} = ^{v}\mathbf{T}_c \, ^{c}\tilde{\mathbf{p}} \tag{2.23}$$

and it inverse transformation

$$ {}^{c}\tilde{\mathbf{p}} = {}^{c}\mathbf{T}_{v}\ {}^{v}\tilde{\mathbf{p}} = \left({}^{v}\mathbf{T}_{c}\right)^{-1}\ {}^{c}\tilde{\mathbf{p}} \tag{2.24} $$

where

$$ {}^{c}\mathbf{T}_{v} = \begin{bmatrix} {}^{c}\mathbf{R}_{v} & -{}^{c}\mathbf{R}_{v}^{v}\,\mathbf{t}_{c} \\ \mathbf{0}^{\top} & 1 \end{bmatrix} \tag{2.25} $$

2.2 Laser scanners

Laser scanners as active range sensors get distance measures according to the *time of flight* concept, essentially a laser pulse is emitted and it will be reflected by objects, a photodiode coupled with a optoelectronic circuit measure time between the outgoing and the returning signal, the distance is calculated considering the speed of light. Furthermore, the 3D laser scanning is based on a controlled steering of laser beams followed by a range measurement at every pointing direction. Most laser scanners employed in mobile robotics use a rotating mirror to deflect laser beams with several solution (e.g., Light Detection and Ranging LIDAR with rolling or global shutter) it is possible to achieve 3D object information as a cloud points.

Laser scanners have better range and lateral resolution than radar or sonar sensors, hence they can exhibit more reliable measures than wave sensors. However, despite their reliability with respect to sonar sensors, laser scanner performances might de-

Distance measurement range	0.3 . . . 50 m
Scan planes	4
Vertical aperture angle	3.5°
Horizontal aperture angle	85°
Scanning frequency	12.5 Hz (0.125 . . . 0.5°)
	12.5 Hz (0.25°)
	25 Hz (0.25°)
	50 Hz (0.5°)
Measurement resolution	40 mm
Statistical error (σ_1)	100mm
Beam divergence	horizontal 0.08° per layer
	vertical 0.8° per layer
	per laser 1.6°
Light sources	2 laser diodes, infrared light (895 ... 915 nm)

Table 2.1: Laser scanner specifications. Example with SICK LD-MRS

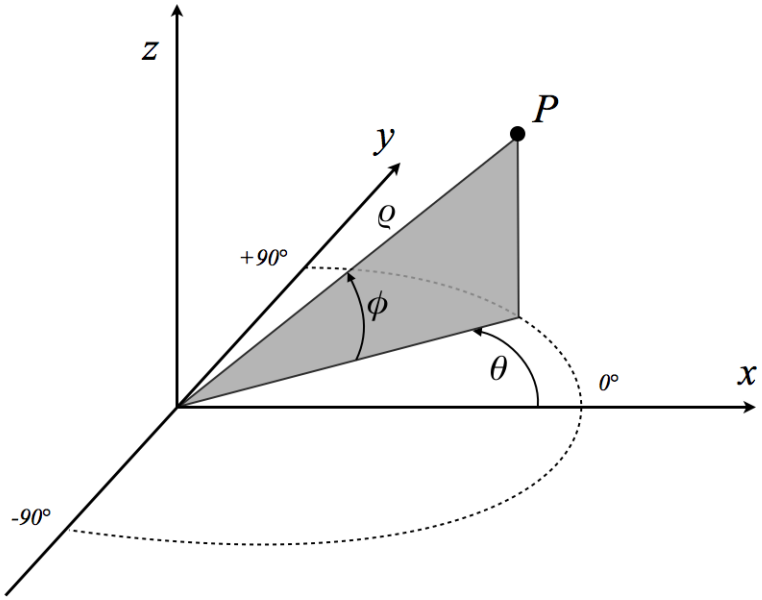


Figure 2.4: LIDAR reference frame.

crease with hard weather condition. For that reason they must be coupled with other kind of sensors to enforce their perception. To give an example of these laserscanners table 2.1 shown main technical specification of a four-layer laser scanner adopted on vehicle prototype involved in this research. Sometimes accuracy specification given by laser scanner dealer differ from real accuracy sensor, especially for instruments built in small series where accuracy depends on individual calibration. For that reason, several institution have published method and results concerning laser scanners accuracy tests [9], here follow a general dissertation on laser scanners parameter accuracy. Generic laser scanner gives information as tuple $P_l = [\rho, \theta, \phi]^\top$, where range information ρ is given for certain azimuth θ and elevation ϕ. Conversion from po-

lar coordinates P_l to Cartesian coordinate $P_c = [x, y, z]^\top$ is given by relation 2.2 and shown in figure 2.4

$$\begin{cases} x = \rho \cos\phi \cos\theta \\ y = \rho \cos\phi \sin\theta \\ z = \rho \sin\phi \end{cases} \tag{2.26}$$

2.2.1 Angular and Range accuracy

Since laser beams are deflected by rotating mirror devices the desired steering angle may be changed by mechanical components. Any unwanted deviation angle will result in errors perpendicular to the propagation path. For the reason that it is hard to determine precise 3D laser point position, errors in this situation can be detected by measuring vertical and horizontal range distances from a spherical object. With similar test LIDARs show for distances up to 100 m constant ranging accuracy and a systematic scale error that can be easily removed when distance differences are determined. Noise of range measurement is achieved by scanning planar surfaces perpendicular to the laser, through 3D points the standard deviation of the range measures is computed. Another important technical specification is the laser resolution used in different context to detail laser scanner performances. Resolution describes the minimum increment of angle(range) between two consecutive points (or the laser spot size on the incident object).

This specification is important for calibration purposes, since it determine object resolution and the ability to describe small features.

2.2.2 Edge effect and Influence of surface reflectivity

The laser beam generated by laser scanners has approximately a cone shape rather than an ideal straight line, consequently -even if well focused- the laser spot incident on any target object will have a certain size. When laser spot hit an object edge part of it will be reflected, while rest of it will reach other targets (even nothing whether out of detection range). In this case LIDARs produce wrong points in edge

neighborhoods. Boehler et al. in [9] detail such behavior that can vary the range measure from millimeters up to several decimeters.

Another important aspect rely on reflective abilities[3] of target surface, since signal strength perceived by the laser scanner is influenced by object material.

It is known that different albedo result in a systematic error in range, it is also observed that bright surfaces reflect the laser beam better than dark surfaces and can be detected over larger distances. From SICK laser scanners specifications brilliant white plaster reflects approximately 100% of the incident light, black foam rubber approximately 2.4%. On very rough surfaces, part of the energy is lost due to shading. This reduces the scanning range of the laser scanner. Experiments on reflectivity have been shown that objects composed by different materials or painted rather than coated (i.e., vehicle body) are always affected by serous errors.

[3] Albedo or reflection coefficient, derived from Latin albedo "whiteness" (or reflected sunlight), in turn from albus "white", is the diffuse reflectivity or reflecting power of a surface. It is defined as the ratio of reflected radiation from the surface to incident radiation upon it

Chapter 3

Camera calibration for automotive systems

Camera calibration is the first step toward 3D perception of the scene. In other words, calibrate a camera means to measure in some way its parameter described by the pinhole model detailed in 2.1. Although some information from 3D object can be computed with uncalibrated camera [10], metric measurements need a calibrated camera. There is a main categorization of camera calibration that distinguish between *photometric calibration*: in which a target object with known geometry is used to achieve camera parameters, and *self calibration* that assume a static scene (e.g., structure from motion). The first category is preferred fro the automotive field.

3.1 Methods overview

Camera calibration include internal and external parameter measurements and there are several classification in literature to distinguish all methods. Here follow a first classification based on involved lens distortion model then a second kind of classification that split intrinsic parameters and extrinsic parameters. These methods are all used for intelligent vehicle equipment and each one has specific used based on particular conditions.

3.1.1 Linear regression

This method is based on a linear representation of camera perspective distortion. The transformation matrix relates 3D world points $\mathbf{X}_i$ to their 2D projection on the image $\mathbf{x}_i$. The calibration procedure exploits least squares method (A.1) to obtain matrix $\mathbf{M}$ that is composed by perspective projection $\mathbf{K}$ mentioned in 2.1 and pose $\mathbf{T}$(2.1.4)

$$\mathbf{x}_i \approx \mathbf{K}[\mathbf{R}|t]\mathbf{X}_i \tag{3.1}$$

substituting $\mathbf{K}$ from equation 2.17

$$\begin{bmatrix} u \\ v \\ 1 \end{bmatrix} \approx \begin{bmatrix} \alpha & 0 & u_0 \\ 0 & \beta & v_0 \\ 0 & 0 & 1 \end{bmatrix} \begin{bmatrix} r_{11} & r_{12} & r_{13} & t_1 \\ r_{21} & r_{22} & r_{23} & t_2 \\ r_{31} & r_{32} & r_{33} & t_3 \end{bmatrix} \begin{bmatrix} X_w \\ Y_w \\ Z_w \\ 1 \end{bmatrix} \tag{3.2}$$

than obtaining relation

$$\begin{bmatrix} u \\ v \\ 1 \end{bmatrix} = \begin{bmatrix} m_1 & m_2 & m_3 & m_4 \\ m_5 & m_6 & m_7 & m_8 \\ m_9 & m_{10} & m_{11} & 1 \end{bmatrix} \begin{bmatrix} X_w \\ Y_w \\ Z_w \\ 1 \end{bmatrix} \tag{3.3}$$

Linear regression can directly estimate the 11 unknowns in the $\mathbf{M}$ matrix by using known 3D points $[X_i, Y_i, Z_i]^\top$ and measured features into the image space $[u_i, v_i]^\top$. By matrix multiplication the expression 3.3 can be represented in u_i, v_i

$$u_i = \frac{m_1 X_{wi} + m_2 Y_{wi} + m_3 Z_{wi} + m_4}{m_9 X_{wi} + m_{10} Y_{wi} + m_{11} Z_{wi} + 1} \quad , \quad v_i = \frac{m_5 X_{wi} + m_6 Y_{wi} + m_7 Z_{wi} + m_8}{m_9 X_{wi} + m_{10} Y_{wi} + m_{11} Z_{wi} + 1} \tag{3.4}$$

The 11 unknowns m_i $i = 1 \dots 11$ can be arranged in a parameter vector
$\mathbf{A} = (m_1 m_2 m_3 m_4 m_5 m_6 m_7 m_8 m_9 m_{10} m_{11})^\top$ and a coefficient matrix $\mathbf{Q} \in \mathbb{R}^{2n \times 11}$ for n

pair of points in a relation $\mathbf{QA} = \mathbf{B}$. Where,

$$\mathbf{Q}_{2i-1} = \begin{vmatrix} X_w i \\ Y_w i \\ Z_w i \\ 1 \\ 0 \\ 0 \\ 0 \\ 0 \\ -u_i X_w i \\ -u_i Y_w i \\ -u_i Z_w i \end{vmatrix}^\top \quad , \quad \mathbf{Q}_{2i} = \begin{vmatrix} 0 \\ 0 \\ 0 \\ 0 \\ X_w i \\ Y_w i \\ Z_w i \\ 1 \\ -v_i X_w i \\ -v_i Y_w i \\ -v_i Z_w i \end{vmatrix}^\top \quad , \quad \mathbf{B}_{2i-1} = u_i \quad , \quad \mathbf{B}_{2i} = v_i \tag{3.5}$$

solution vector $\mathbf{A}$ is obtained by pseudo-inverse

$$\mathbf{A} = ((\mathbf{Q}^\top \mathbf{Q})^{-1})\mathbf{Q}^\top \mathbf{B} \tag{3.6}$$

Note that approximation $m_{12} = 1$ can be assumed since the transformation matrix $\mathbf{M}$ is defined up to a scale factor. The first limitation of this method is that it can not model lens distortion and it is useful only if such distortion is previously removed. Whether lens distortion is present the whole calibration parameters will be compromise and accuracy will be related to the optic distortion. Furthermore, matrix $\mathbf{M}$ is mixes internal and external parameters than it is pose specific and camera can not be moved. The main advantages is in the simplicity of the model and the calibration procedure. Related works on this method can be found in Hall et al. [11] and Toscani-Faugeras [12]

3.1.2 Non-Linear regression

The non-linear equations arise when lens distortion are taken into account in the camera model. These models are estimated through an iterative optimization technique that minimize certain loss function. Usually, the minimization function measure the distance between the image points and their back projection obtained by the searched

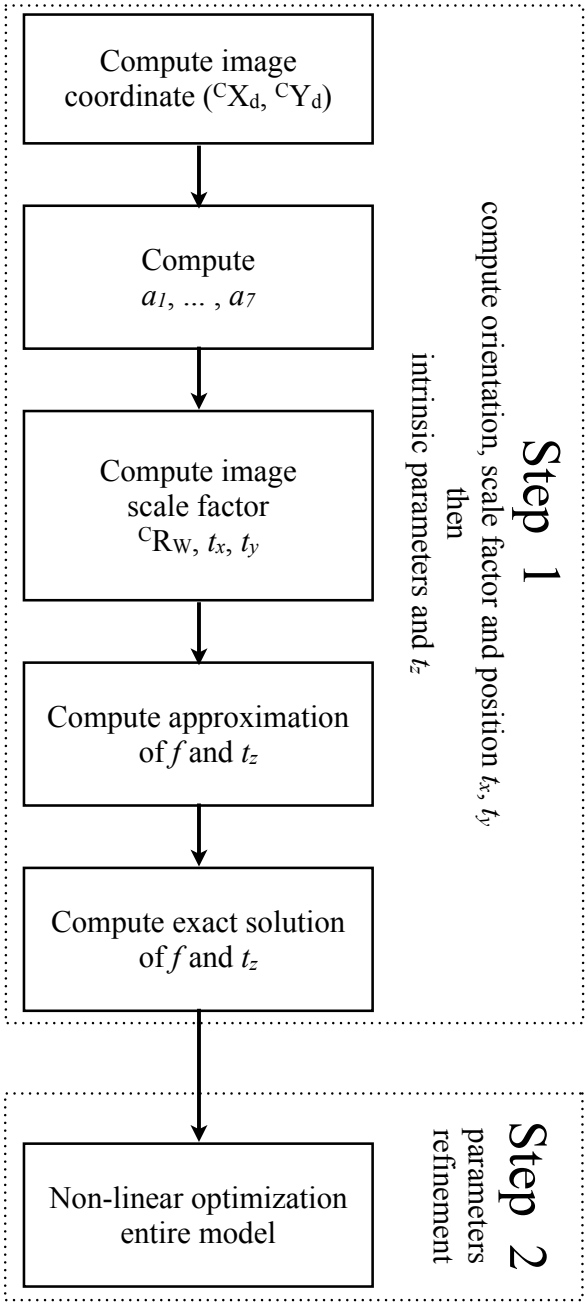


Figure 3.1: Scheme of two-steps Tsai method.

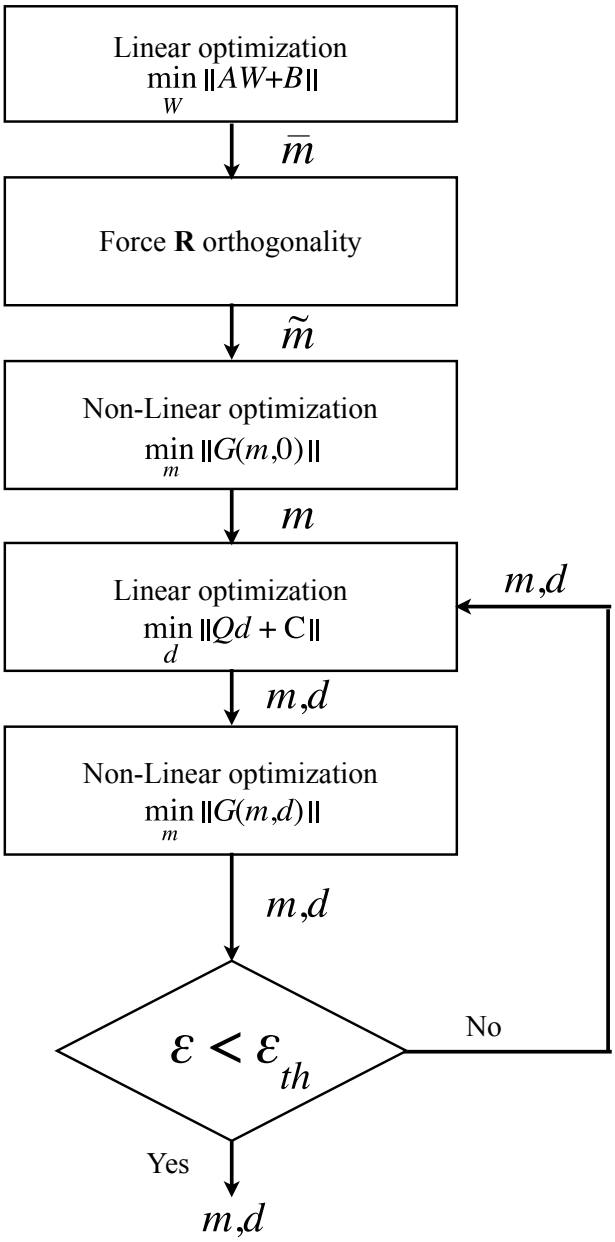


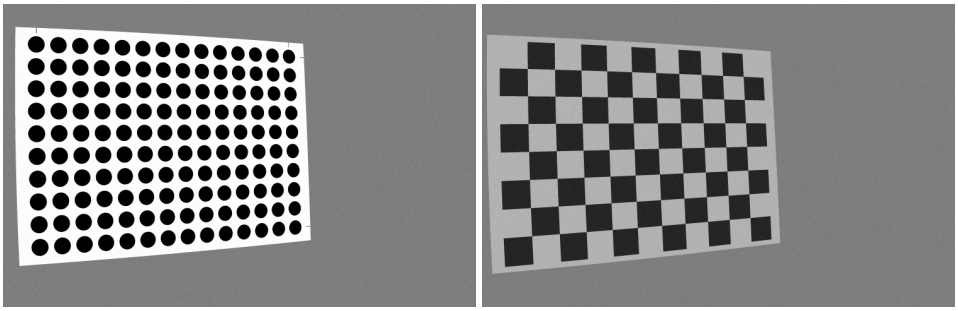
Figure 3.2: Scheme of Weng method.

model. As non-linear optimization problem described in A.2 these techniques need a good initial guess in order to converge to correct solution, for that reason it is preferred a *two-steps* technique where a linear guess is obtained in the first step then a non-linear optimization is achieved in the second step. The most representative works on these procedures can be found in Faugeras-Toscani with distortion model [13], Tsai [14], Weng [15], and Sturm-Maybank-Zhang [16, 17]. The most used method in computer vision is the *Tsai two-steps method* shown in figure 3.1 , although this procedure is very flexible and accurate it can model only radial distortion. In the automotive field it is observable that such distortion model can not be enough to correctly calibrate a camera mounted on a vehicle due to asymmetrical geometry of some windshield, hence different lens distortion model have to be considered. The model of *Weng*, proposed in 1992, include three distortion components to improve camera calibration and it is shown in figure 3.2.

3.2 Determine Lens Distortion

Almost 3D computer vision applications for automotive rely on the pinhole camera model to extract geometrical information from a scene, usually such application are coupled either with wide-angle or short focal length optics as application constrain. Due to severe non-linear lens distortion these applications must compensate such distortion to be able to apply correctly the perspective projection of the pinhole model. Thus, image undistortion (or image correction) is an important preprocessing. For automotive application the best way to manage lens distortion is to generate a look up table (LUT) which removes distortion effect on the image plane, then it can be apply to incoming frames quickly with low computational power as preprocessing. Consequently, the main problem to sole is to measure the distortion model parameters to be able to build a proper LUT.

Therefore, the correction of lens distortion involves a calibration procedure used to determine the parameters of the model selected from one described in 2.1.2. Then the distortion can be corrected by inverting the model and recover the undistorted position of each pixel.



(a) (b)

Figure 3.3: Checkerboards involved in the de-distortion process. Based on blob (*a*) and corner (*b*) detection.

First of all, it is necessary to choose a model that best fit the distorted image. The best practice to do this is to test different of them and select the one that returns the lowest residual (i.e. back projecting marker from undistorted image to the distorted image). Devernay and Faugeras describe this process where, for a given model (e.g., polynomial) this method will almost always select the model with the highest number of parameters. However, it is still possible to select different model with the same number of parameters.

At this point it is possible to select a specific calibration procedure, a largest part of calibration procedure on this topic use specific calibration grids with known 3D geometry, (e.g. shown in figure 3.3(a), 3.3(b)) for this research method such as [18, 14] and most used [8] had been implemented to perform lens distortion recovery. Those methods measure image distortion from features like corners, dots, and lines that can be easily detected in the image.

Basically these method try to optimize the distortion parameters (such as distortion center and polynomial coefficients) by minimizing the total distortion error, that is computed as the sum of the distortion errors on the selected features and their undistorted position (e.g., Devernay and Faugeras compute distortion error as distance between detected segments an straight lines).

3.3 Determine Intrinsic

Pixel focal length, optical center and skew parameters (described in 2.1.3) represent physical and optical camera parameters that are involved during image formation (as detailed in 2.1.3). They are defined intrinsic since they are an internal characteristic that is not directly measurable.

Measurement of these parameters has been achieved in several calibration procedure of this thesis by using Zhang procedure [17] that is an improvement of Tsai calibration technique (3.1). The Zhang method is based on multiple views of a plane (at least three), rather than using a single view with more then one plane. This is automatic calibration with the main difference that it is supposed to known the geometry of the observed calibration object. Basically, it assumes that is possible to compute the homography transformation that map a plane present in the scene to the image plane. To be able to compute such transformation a planar calibration target with a grid superimpose on it must be used (e.g., in figure 3.3(a) and figure 3.3(b)). With a camera modeled by pinhole, the relation between world point $\mathbf{P} = [X,Y,Z]^\top$ and its image projection $\mathbf{p} = [u,v]^\top$ is given by relation

$$s\tilde{\mathbf{p}} = \mathbf{K}[\mathbf{R} \quad \mathbf{t}]\tilde{\mathbf{P}} \tag{3.7}$$

where generic vector $\tilde{\mathbf{x}}$ represent the homogeneous coordinates of vector $\mathbf{x}$, thus $\tilde{\mathbf{P}} = [X,Y,Z,1]^\top$ and $\tilde{\mathbf{p}} = [u,v,1]^\top$, s is an arbitrary scale factor, $\mathbf{K}$ is the camera intrinsic matrix (2.1.3), and $[\mathbf{R} \quad \mathbf{t}]$ are the extrinsic parameters.

To simplify equations -and without loss of generality- it is possible to assume the model plan on $Z = 0$ of the world reference system $\{O_W\}$ (shown in figure 2.1.1). Obtaining

$$s\begin{bmatrix} u \\ v \\ 1 \end{bmatrix} = \mathbf{K}[\mathbf{r}_1 \quad \mathbf{r}_2 \quad \mathbf{r}_3 \quad \mathbf{t}]\begin{bmatrix} X \\ Y \\ 0 \\ 1 \end{bmatrix} = \mathbf{K}[\mathbf{r}_1 \quad \mathbf{r}_2 \quad \mathbf{t}]\begin{bmatrix} X \\ Y \\ 1 \end{bmatrix} \tag{3.8}$$

therefore, each model point P of the plane and its image position p is related by homography $\mathbf{H}$ composed with the first, second and fourth column of matrix $[\mathbf{R} \quad \mathbf{t}]$.

Since Z coordinate is always equal to zero $\mathbf{P} = [X,Y]^\top$, and $\tilde{\mathbf{P}} = [X,Y,1]^\top$ are coupled in a relation such that

$$s\tilde{\mathbf{p}} = \mathbf{H}\tilde{\mathbf{P}} \tag{3.9}$$

where $\mathbf{H} \in \mathbb{R}^{3\times 3}$ defined up to a scale factor s, thus with

$$\mathbf{H} = \lambda \mathbf{K}[\mathbf{r}_1 \quad \mathbf{r}_2 \quad \mathbf{t}] \tag{3.10}$$

matrix $\mathbf{H}$ can be estimated by taking advantage of the orthogonality between $\mathbf{r}_1$ and $\mathbf{r}_2$ obtaining constraints on the intrinsic parameters using relations 3.11.

$$\begin{aligned} \mathbf{r}_1 &= \lambda \mathbf{K}^{-1}\mathbf{h}_1 \\ \mathbf{r}_2 &= \lambda \mathbf{K}^{-1}\mathbf{h}_2 \end{aligned} \tag{3.11}$$

Furthermore, considering orthogonality $\mathbf{r}_1^\top \mathbf{r}_1 = 0$ and using equations 3.11 follow that $\mathbf{h}_1^\top \left(\mathbf{K}\mathbf{K}^\top\right)^{-1} \mathbf{h}_2 = 0$ or equivalently

$$\mathbf{h}_1^\top \mathbf{B}\mathbf{h}_2 = 0 \tag{3.12}$$

In the same way, considering relation $\mathbf{r}_1^\top \mathbf{r}_1 = \mathbf{r}_2^\top \mathbf{r}_2$ a second constraint can be represented as follows in equation 3.13.

$$\mathbf{h}_1^\top \mathbf{B}\mathbf{h}_1 = \mathbf{h}_2^\top \mathbf{B}\mathbf{h}_2 \tag{3.13}$$

Finally,3.12 and 3.13 are the two basic constraints on the intrinsic parameters, given one homography. Because a homography has 8 degrees of freedom (element $\mathbf{h}_3 = 1$) and there are 6 extrinsic parameters (3 for rotation and 3 for translation), we can only obtain 2 constraints on the intrinsic parameters. Note that $\mathbf{B} = \mathbf{K}^\top \mathbf{K}^{-1}$ actually describes the image of the absolute conic [19].

Matrix $\mathbf{B}$ is a 3×3 symmetrical matrix, hence with 6 unknown. Each pose gives two equation in six unknown then with n observations (with different orientation and position) it is possible to derive $2n$ equations for a linear system $\mathbf{A}\mathbf{b} = \mathbf{0}$. Where $\mathbf{A}$ is a $2n \times 6$ matrix. If $n \geq 3$, there is in general a unique solution of unknown vector b defined up to a scale factor. If $n = 2$, solution can be achieved by imposing zero skew $(\gamma = 0)$ constraint (i.e., $[0,1,0,0,0,0]\,b = 0$), which is added as an additional

equation to the linear system. (If $n = 1$, only two camera intrinsic parameters can be solved , e.g., K_u and K_v, assuming u_0 and v_0 are known (e.g., at the image center) and $(\gamma = 0)$). The solution to the linear system is well known as the eigenvector of $\mathbf{A}^\top\mathbf{A}$ associated with the smallest eigenvalue (equivalently, the right singular vector of $\mathbf{A}$ associated with the smallest singular value). Once b is estimated, all camera intrinsic matrix $\mathbf{K}$ can be easily computed. Let $\mathbf{b} = [B_{11}, B_{12}, B_{22}, B_{13}, B_{23}, B_{33}]^\top$

$$\begin{aligned} v_0 &= (B_{12}B_{13} - B_{11}B_{23}) / \left(B_{11}B_{22} - B_{12}^2\right) \\ \lambda &= B_{33} - \left[B_{13}^2 + v_0 (B_{12}B_{13} - B_{11}B_{23})\right] / B_{11} \\ \alpha &= \sqrt{\lambda / B_{11}} \\ \beta &= \sqrt{\lambda B_{11} / \left(B_{11}B_{22} - B_{12}^2\right)} \\ \gamma &= -B_{12}\alpha^2\beta / \lambda \\ u_0 &= \gamma v_0 / \beta - B_{13}\alpha^2 / \lambda \end{aligned} \tag{3.14}$$

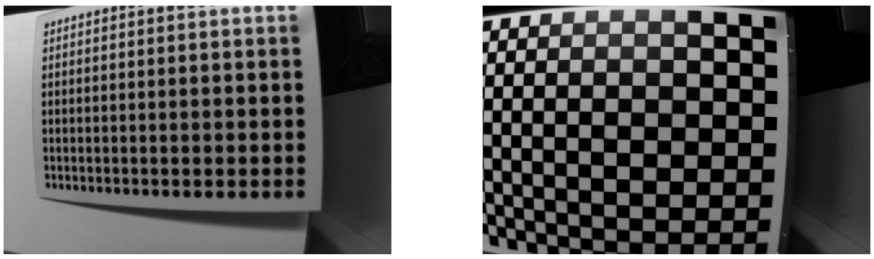
3.4 Results and contributions

There are several process during camera calibration that can have bad effect on parameter estimation two main sources of error that affect 3D measurement results depends on:

- The hypotheses of planar target with ideal patterns of conventional camera calibration techniques (e.g., checkerboard). The defects of the calibration target lead to inaccurate results.

- The uncertainty in locating the control points directly using the geometries of the calibration patterns suffers from lens distortion as well as perspective distortion from raw images.

Results on marker detection accuracy are presented in the following sections. Going into detail, this topic had been correlated with an algorithm implementation for a calibration grid detection, developing also two marker detectors, both based on analytical approach to refine a set of rough detected features.



(a) (b)

Figure 3.4: Calibration target. Based on blob (*a*) and corner (*b*) grids.

3.4.1 Marker detection

In this research a plane grid panel is used as calibration target to be able to determine optical sensor parameters as mentioned in previous sections 3.2 and 3.3, i.e., Zhang procedure requires that a camera observes a planar pattern in several poses to obtain such parameters. Usually the specific pattern employed in calibration procedures is printed or even superimposed on a rigid plane surface (e.g., figure shown makers used in VisLab laboratories). Different poses are obtained by moving either the camera or the calibration target manually. Therefore this calibration technique become more flexible and robust than a classic *photogrammetric calibration* which requires specific tools and a very accurate setup. Then Zhang procedure is used to perform camera calibration with different pattern in the following process:

1. Print specific pattern as shown in figures 3.4(a), 3.4(b) on a plane foam board of size A4 (210×297mm) or even A3 (297×420mm), depending on the camera FOV and the feature size on the image plane.

2. Acquire several photos of the scene (at least two to compute pinhole parameters) with different pose of the planar pattern

3. Compute the marker detection algorithm on each acquired image on the specific feature.

4. Estimate the intrinsic and extrinsic camera parameters using Zhang calibration method, then estimate the distortion model.

5. Finally an optimization phase based on nonlinear optimization is computed to refine model parameters.

The contribution on this research is focused on the marker detection (point 3 of previous calibration procedure). Marker detectors are software capable to recognize known features into the image and compute their position with respect to the image reference frame. Basically, the marker detector developed in this thesis are analyzed and compared with different maker detectors. To be able to measure detection accuracy an compare it with different detectors, a specific synthetic image generate had been developed.

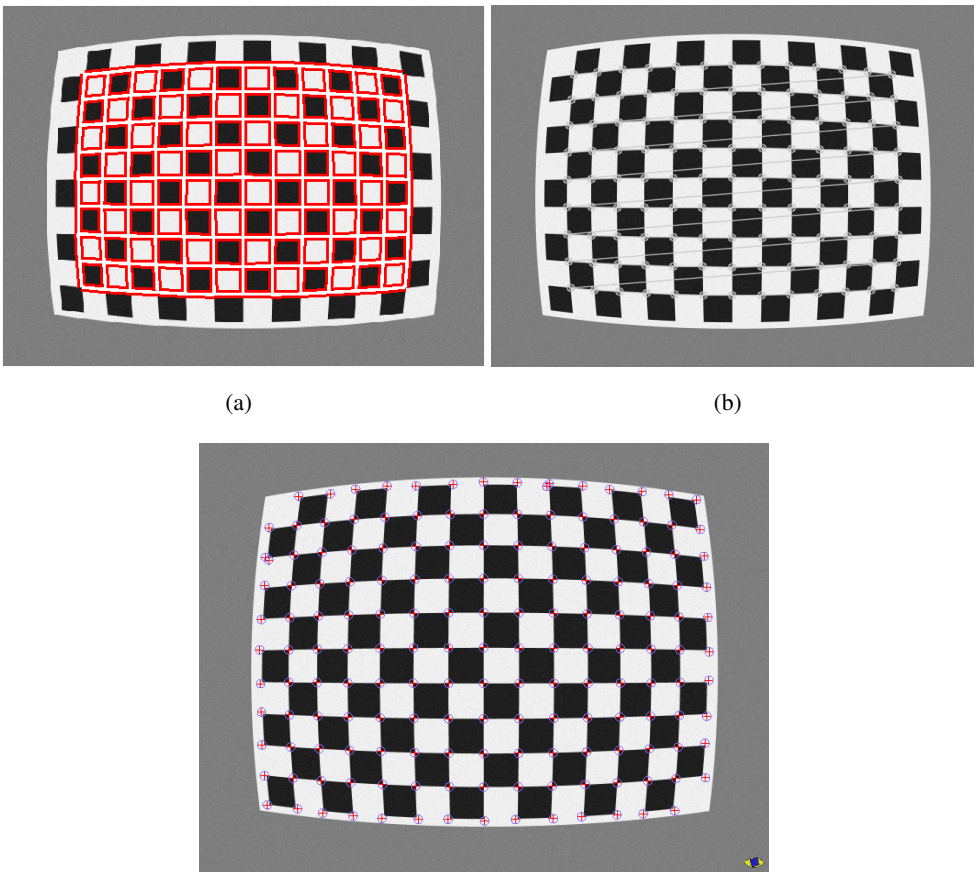
The two investigated pattern investigate in this thesis are:

- grids of circular blobs. A flood fill algorithm is used to recognize small circular pattern;

- checkerboards. Particular *X-junction* detection algorithm in employed using saddle points associated to corner features [20].

Both calibration targets had been developed and reviewed during this research and compared with two modern computer vision libraries:

- OpenCV library [21]: The Open Source Computer Vision Library opts for using multiple views of a planar object (a chessboard) rather than one view of a specially constructed 3D object. It uses a pattern of alternating black and white squares (as shown in figure 3.5(b)), which ensures that there is no bias toward one side or the other in measurement. Given an image of a chessboard (or a person holding a chessboard, or any other scene with a chessboard and a reasonably uncluttered background), OpenCV uses function *cvFindChessboardCorners()* to locate the corners of the chessboard. The corners returned by *cvFindChessboardCorners()* are only approximate. Therefore, locations are accurate only to within the limits of imaging device, which means accurate to



(a) (b)

(c)

Figure 3.5: Marker Detectors. LibCBDetect from KIT and based on Matlab library in (*a*), OpenCV in (*b*), and Marker detector based on GOLD framework in (*c*).

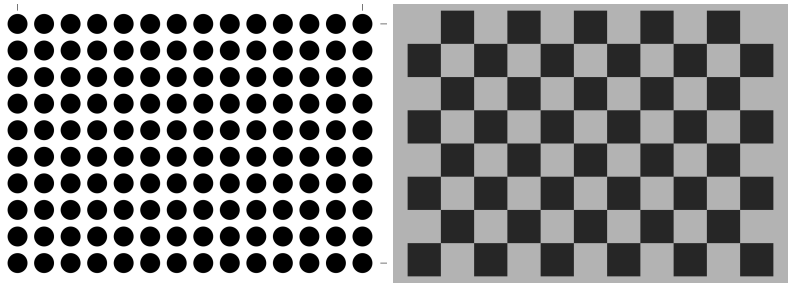
within one pixel. A separate function *cvFindCornerSubPix()* is used to compute the exact locations of the corners to subpixel accuracy, since the chessboard interior corners are simply a special case of the more general Harris corners; the chessboard corners just happen to be particularly easy to find and track. Neglecting to call subpixel refinement after you first locate the corners can cause substantial errors in calibration.

- LibCBDetect [22]: this is a toolbox for Matlab developed by the Karlsruhe Institute of Technology. The algorithm automatically extracts corners to subpixel accuracy and combines them to rectangular checkerboards or chessboard-like patterns. However this library is able to detect corner only and does not provide any marker detector for circular features. It works on a variety of imagery such as pinhole cameras, fisheye cameras, and omnidirectional cameras.

- GOLD: The General Obstacle and Lane Detection [23, 24] is a C++ framework designed by the VisLab to develop applications related to modern intelligent vehicles. GOLD offer a set of API to develop computer vision application. Furthermore, two marker detectors had been reviewed: *BlobMarkerDetector* and *CeckBoardMarkerDetector*. Notice that these two marker detectors use only local information and they do not use any "bundle-adjustment" to refine via nonlinear optimization the final result.

3.4.2 Detection Error analysis

Tests had been conduced using different sets of images. Such sequences had been generated with simulator designed and developed ad hoc specific analysis. The software generates a set of views from an input image (i.e., shown in figure 3.6(a), and 3.6(b)) using the pinhole camera model with radial distortion. Thus, a list of random poses (orientation and position) are generated to compute the relative pose between simulated camera and calibration target. Outcome of this procedure will be a set of distorted images where the input pattern is shown from different point of view in a grey scale background with Gaussian pixel noise added and for each image a list of feature position used as a ground truth.



(a) (b)

Figure 3.6: Synthetic image generator inputs. Circular blob pattern in (*a*), and Checker or corner features in (*b*) are considered as A4 paper format (210×297mm).

As the sequence of images can be "asymmetric" with respect to the camera optical center, since random motions of the camera may led in a set of images where calibration target appear only in part of the image, each generated image is rotated with respect to the optical center axis.

The pinhole parameters adopted in these tests are reported in table 3.1. A set of different test are reported:

1. Subpixel accuracy evaluation with comparison between different detectors with respect measurement accuracy. Test had been yield using both corner and circular features in 400 different poses. X-junction detection accuracy is reported in table 3.2, with graphical detection error reported in figure 3.7 and error distribution in figure 3.8. Circular marker detection and accuracy are summarized in table 3.3, detection error is reported in figures 3.9 and 3.11, with distribution error shown in figure 3.10 and 3.12 for OpenCV and GOLD environments respectively.

2. Detection Error over different distance. Camera is closer and facing the calibration target, it is shifted backward step by step from 20cm up to 2.15m. Results are shown in figure 3.13.

3. Detection Error over different orientation. With fixed distance between camera

pinhole model			
u_0	330.81043	v_0	265.294135
k_u	408.40393	k_v	408.40393
width	656	height	492
radial distortion model			
d_x	331.138	d_y	262.473
k_0	1.58485×10^{-6}	k_1	2.2072×10^{-12}
k_2	8.25425×10^{-18}	k_3	3.09292×10^{-23}

Table 3.1: Intrinsic parameters for virtual camera. Width, height, u_0, v_0 are $[pixel]$, pixel focal length k_u and k_v are $\left[pixel \cdot m^{-1}\right]$.

and calibration target (1m), the virtual camera moves around the target following a bow trajectory. Elevation angle of the camera with respect the calibration target range from 30° to 150° (notice that 90° represents a camera in front of the calibration grid). Results are shown in figure 3.14

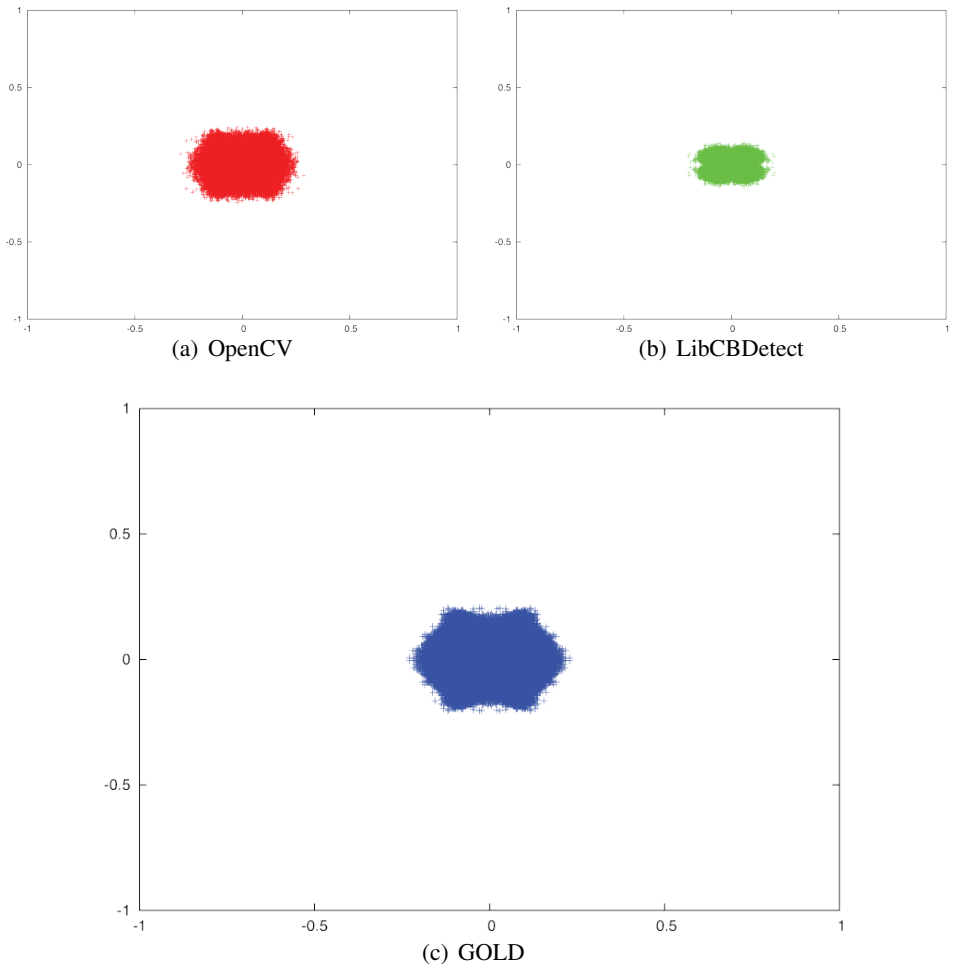
Considering all tests, it can be said that all the marker detectors evaluated (OpenCV, LibCBDetect, and GOLD) conduced to an absolute error smaller than one pixel -as it might be-. In both cases, corner features and circular blob, all three detectors get similar accuracy in marker detection. On the one hand the X-junction is more robust to high relative orientation due to its persistent information, and on the other hand circular spot are badly affected by perspective, thus considering high tilt of calibration target. Although blob detection is badly affected by perspective, whether a calibration marker is facing a camera it can achieve better results with respect to corner detection. This behavior is related to the specific algorithm that involves different information. Blob detectors developed on GOLD library presents different performance with respect to OpenCV algorithm basically for two reason: firstly, it use only local information to detect such marker without knowledge of the grid size (while

	OpenCV	LibCBDetect	GOLD
CD %	100.00	100.00	99.99
μ_x	-10.7641×10^{-05}	-4.6298×10^{-08}	-8.67552×10^{-06}
μ_y	6.31393×10^{-05}	5.31449×10^{-08}	-9.14797×10^{-06}
σ_x	0.0942444	0.0645443	0.083767
σ_y	0.0876403	0.0522401	0.0802072

Table 3.2: X-junction detection accuracy on 46800 features.

	OpenCV	GOLD
CD %	100.00	99.51
μ_x	-3.8147×10^{-09}	-0.00198672
μ_y	-4.76837×10^{-08}	0.00270155
σ_x	0.0506908	0.0881842
σ_y	0.0368875	0.0881886

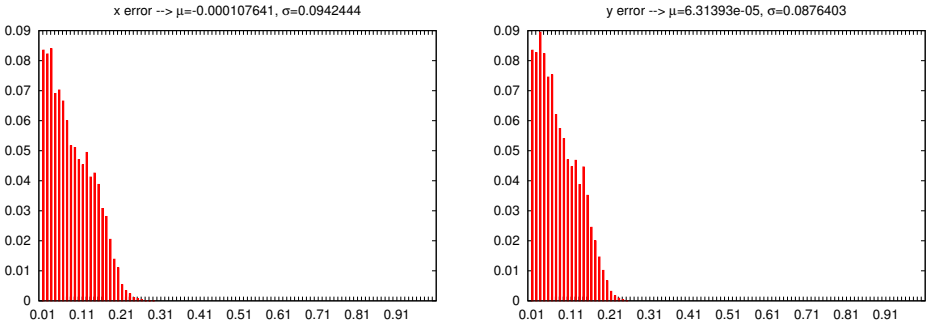
Table 3.3: Circular marker detection accuracy on 56000 features.



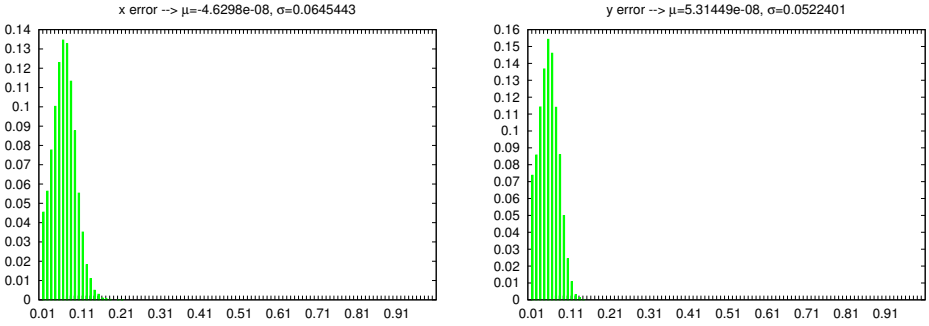
(a) OpenCV (b) LibCBDetect

(c) GOLD

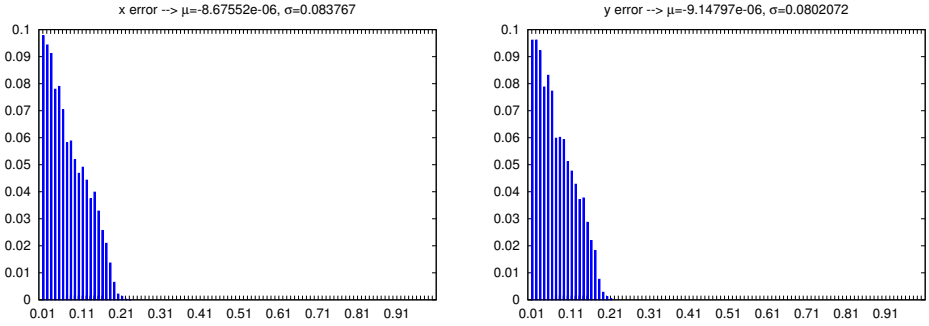
Figure 3.7: X-junction detection accuracy.



(a) x axis (b) y axis



(c) x axis (d) y axis



(e) x axis (f) y axis

Figure 3.8: OpenCV (*a*, *b*), LibCBDetect (*c*, *d*), and GOLD (*e*, *f*) detection error distribution for corner features.

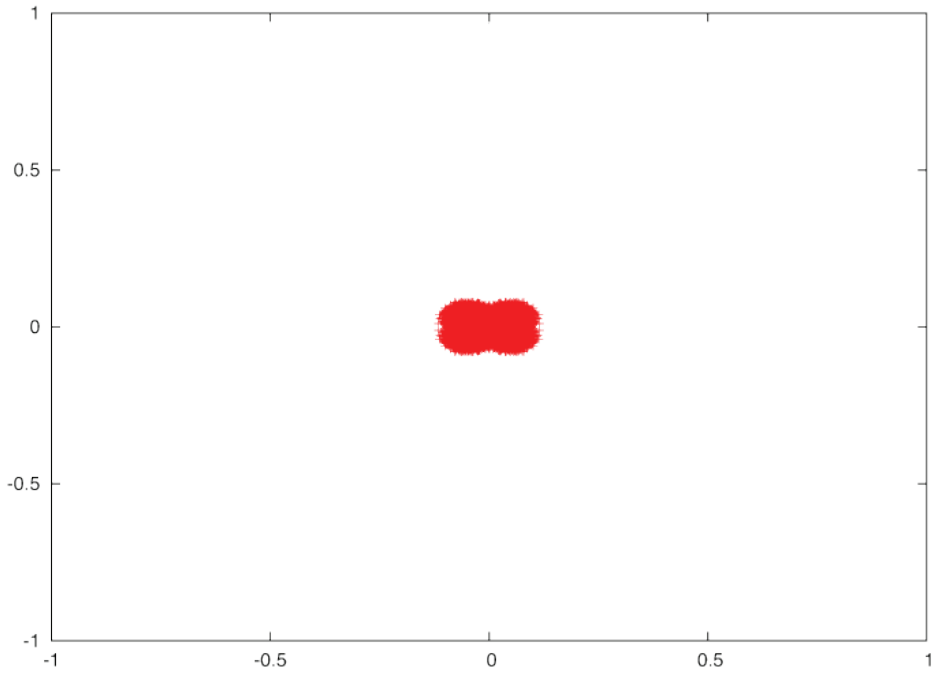
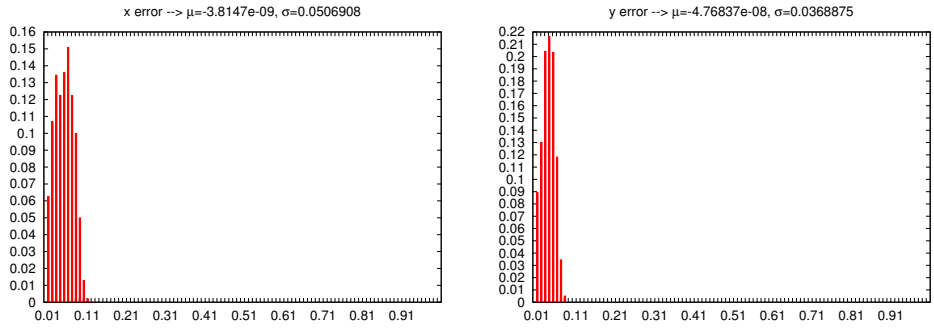


Figure 3.9: Circular marker detection accuracy with OpenCV.



(a) x axis (b) y axis

Figure 3.10: OpenCV detection error distribution for circular markers along x-axis and y axis (*a*, *b*) .

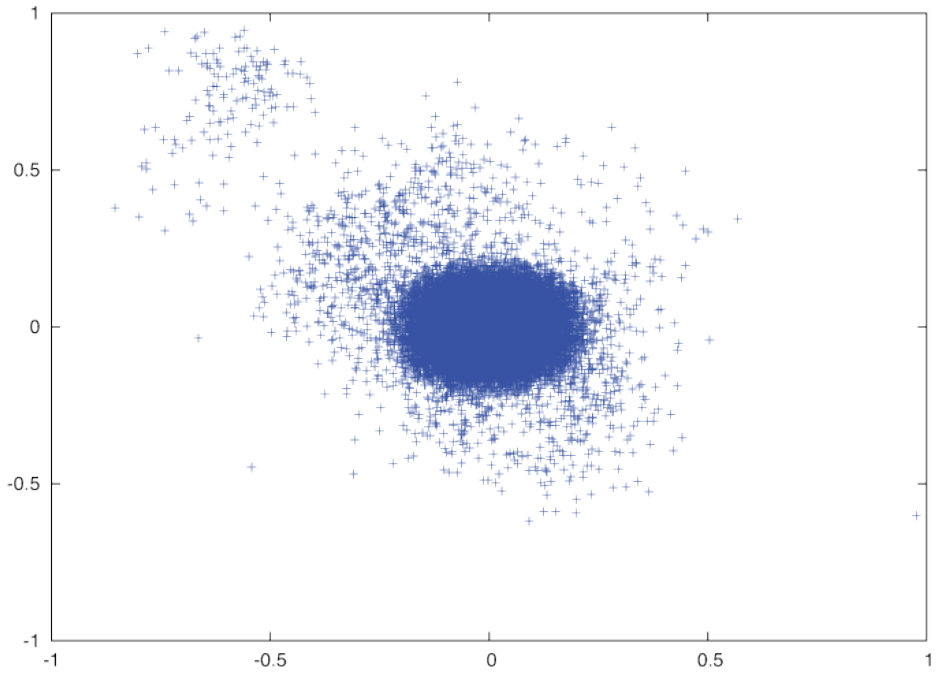
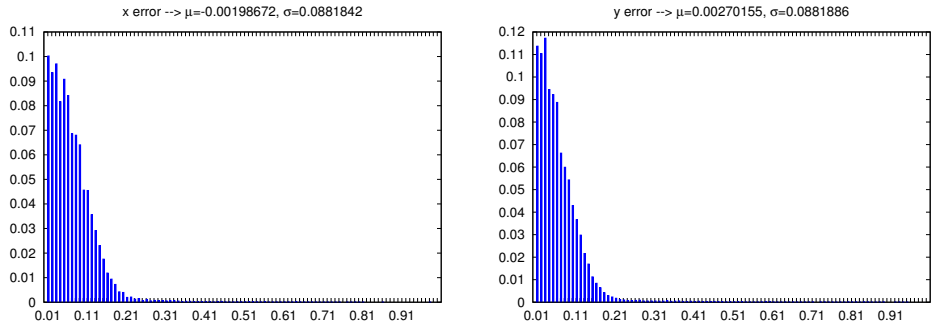
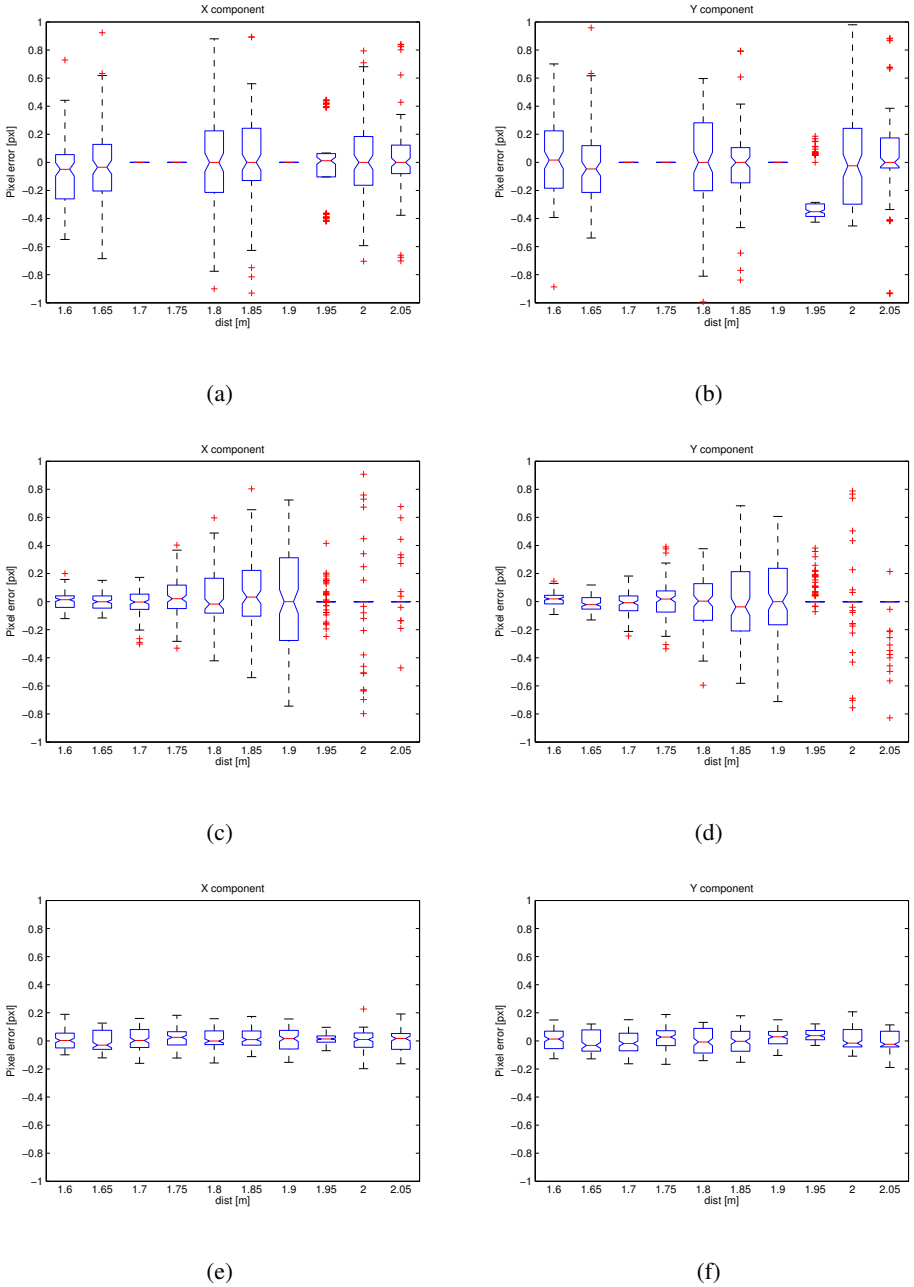


Figure 3.11: Circular marker detection accuracy with GOLD.



(a) x axis (b) y axis

Figure 3.12: GOLD detection error distribution for circular markers along x-axis and y axis (*a*, *b*) .



(a) (b)

(c) (d)

(e) (f)

Figure 3.13: OpenCV (*a*, *b*), LibCBDetect (*c*, *d*), and GOLD (*e*, *f*) detection error over distance using corner features.

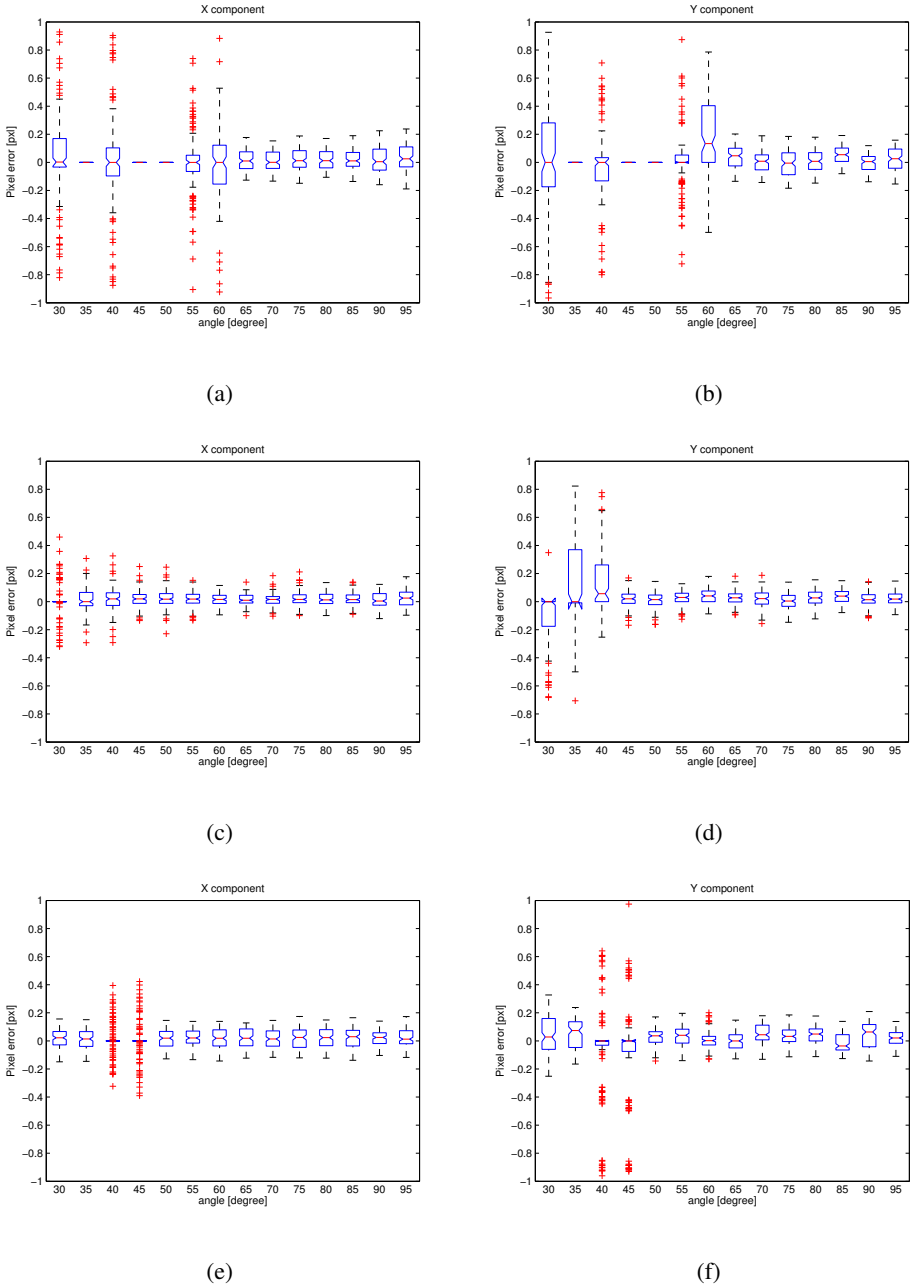


Figure 3.14: OpenCV (*a*, *b*), LibCBDetect (*c*, *d*), and GOLD (*e*, *f*) detection error over orientation using corner features.

OopenCV uses such information), secondly given a definition of correct detection as any detection that is far from the ground truth less than 1pixel it is possible to observe that GOLD detectors do not filter out all wrong detections (i.e., wrong aspect ration or color features).

3.5 Pose Estimation Problem

One of the most important problem to solve during calibration tasks is the pose estimation of our sensors. In computer vision the pose estimation problem has the goal of estimate the rigid transformation of calibrated camera by using set of correspondences. This thesis exploits different solutions of the pose estimation problem, which can be classified in three different kinds related to data types perceived by the sensors:

- 2D: when only projection on two different image planes of 3D points is considered. The rigid transformation from one image to the other have to be computed. Related works are the iterative method of Horn in [25, 26].
- 3D: three-dimensional information of a specific object are given with respect to two different reference frames and the rigid transformation from one sensor to the other have to be computed. The analog rigid body movement problem is treated with a closed-form solution in A.3.3 using SVD decomposition related to Arun et al. [27], other works can be found in Horn [28] a closed-form solution based on unit quaternion, in [29] other closed-form solution based on orthonormal matrices, and Walker et al. [30] with solution based on dual number quaternions.
- P*n*P: 3D points of a specific object and their 2D projection on the image plane are both known and the relative position of the camera with respect to a global reference frame have to be computed.

The first two cases can be represented as an intelligent vehicle equipped with cameras and lasers that perceive a static object (or a moving object perceived by a static vehicles). In the first case an optical sensor is adopted and only features on the

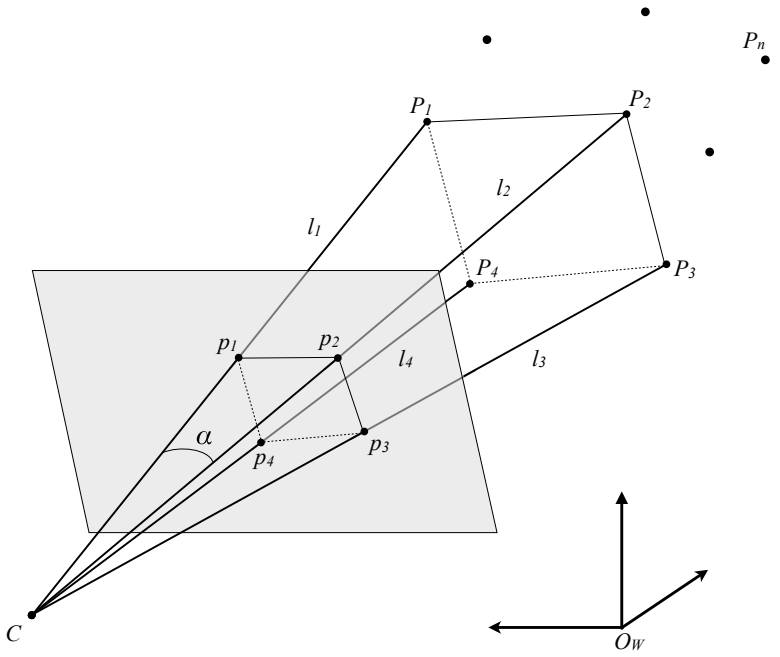


Figure 3.15: Pose estimation problem. Set of n 3D points P_i in the world reference frame O_W are projected on the image plane obtaining corresponding points p_i. The pose of camera C have to be computed with respect to O_W.

image are known, in the second case only range sensors are considered (i.e., stereo rigs and laser scanners).

3.5.1 Perspective *n*-points problem

This can be related to an usual calibration problem where extrinsic parameters of a camera are unknown. A lot of works are described in literature on this subject with different approach that differ in their structure and requirements. Pose estimation solutions can be classified according to the algorithm approach, from the iterative

solutions to the non-iterative solutions.

In the first case main references are: Lowe [31] of 1991 or Dementhon and Davis POSIT algorithm [32] developed in 1995, more recent works are related to 2006 Schweighofer and Paniz [33] and 2000 Lu et al. [34]. All these methods have the benefit of getting accurate solutions, however they can take a long time to converge and a risk of get into a local minima.

In the other case a wide set of non iterative solutions are investigated to avoid risk of local minima with the only side effect of high complexity and consequently a slow processing for large point sets . Mentionable works start from 1981 the Fishler and Bolles method [35], 1999 Quan and Lan linear algorithm [36], 2001 Fiore linear algorithm [37], and simplest DLT algorithm [38] (see also 3.3). In 2009 Lepetit et al. [39] propose an efficient PnP solution that is as accurate as an iterative solution an robust to noise.

Usually in automotive application are available only small set of control points to solve orientation problems, to be able to obtain a valid solution of such problem the smallest value of n is investigate to solve the PnP problem.

Clearly, P1P problem does not provide enough constrained information, thus an infinity of solutions is avoidable. Also the P2P problem admit infinitely many solutions, although solutions are bounded anywhere on a circle that pass through the camera C (in the center of projection) and the two control points (e.g., P_1, P_2) with a diameter of $\overline{P_1P_2}/\sin\alpha_{12}$ where α_{ij} represent the angle between legs $\overline{CP_i}$ and $\overline{CP_i}$ (see figure 3.15), intuitively a camera can see the same segment form different points with different orientation.

The P3P problem with three non collinear points gives up to four real solutions, this is a more interesting case used in this thesis to compute camera extrinsic calibration. As shown in figure 3.15 the P3P requires to determine the length of the three legs $l_i = \overline{CP_i}$ of a tetrahedron, given the triangular base edges and angles α_{ij} between legs on the opposite vertex. This problem can be solved with three polynomial equa-

tions 3.15

$$\begin{cases} d_{12}^2 = l_1^2 + l_2^2 - 2l_1 l_2 \cos\alpha_{12} \\ d_{13}^2 = l_1^2 + l_3^2 - 2l_1 l_3 \cos\alpha_{13} \\ d_{23}^2 = l_2^2 + l_3^2 - 2l_2 l_3 \cos\alpha_{23} \end{cases} \tag{3.15}$$

The system 3.15 can have a maximum of eight solutions (as three independent polynomial equations of second degree have no more solution than $2 \times 2 \times 2$ solutions). However, since every terms in such system is either a constant or of second degree, for every real positive solution there is a geometrical isomorphic negative solution. Therefore, there are at most four positive solution to 3.15. In [35] an explicit algebraic solution for system 3.15 is given. This is accomplished by reducing 3.15 to a biquadratic (quadratic) polynomial in one unknown representing the ratio of two legs of the tetrahedron, in [35] an iterative solution is also given to solve the same problem.

In P4P when all control points lie on the same plane (and there are not three collinear points) a unique solution exists and detailed overview of such solution is given following the Quan and Lan linear algorithm. This procedure involves two separated problems: the perspective-n-points problem itself and the absolute orientation problem. The first problem can be stated as follow:

> With respect to figure 3.15 given the set of correspondences between the 3D points P_i and their 2D projections p_i on camera image plane. Find lengths l_i of line segments between the camera center C and 3D points P_i, $i = 1,2,3,4$ here called legs.

With such l_i it is possible to compute ${}^C P_i$, that is the 3D point in camera reference frame which is P_i represented in the world reference frame O_W. Let define 2D point $p_i = (u,v)^\top$ and its homogeneous coordinates with $\tilde{p} = (u,v,1)^\top$. This image point is related to P_i with intrinsic matrix $\mathbf{K}$ (2.17) with following equation:

$${}^C\hat{P}_i = \mathbf{K}^{-1}\tilde{p} \tag{3.16}$$

where each ${}^{C}\hat{P}_i$ lies on the line that pass through C and the world point ${}^{C}P_i$. Finally each 3D point ${}^{C}P_i$ is obtained by

$$ {}^{C}P_i = \frac{d_i^{C}\hat{P}_i}{\|{}^{C}\hat{P}_i\|_2} \tag{3.17} $$

Therefore, with two sets of 3D points, each of which in different coordinate system (C and O_W), it is needed to find the $[\mathbf{R}|\mathbf{t}]$ matrix which transforms points in O_W to the points in C reference frame, that is goal of the absolute orientation problem:

> Given the set of correspondences between 3D points in different coordinate systems, find the camera transformation matrix $[\mathbf{R}|\mathbf{t}]$ which transforms points in first coordinate system to the second coordinate system.

Solution to such problem is treated in [28]. To be able to fine the legs l_i between camera C and world points P_i let is considered cosines law to define a system of six equations

$$ \begin{cases} d_{12}^2 = l_1^2 + l_2^2 - 2l_1 l_2 \cos\alpha_{12} \\ d_{13}^2 = l_1^2 + l_3^2 - 2l_1 l_3 \cos\alpha_{13} \\ d_{14}^2 = l_1^2 + l_4^2 - 2l_1 l_4 \cos\alpha_{14} \\ d_{23}^2 = l_2^2 + l_3^2 - 2l_2 l_3 \cos\alpha_{23} \\ d_{24}^2 = l_2^2 + l_4^2 - 2l_2 l_4 \cos\alpha_{24} \\ d_{34}^2 = l_3^2 + l_4^2 - 2l_3 l_4 \cos\alpha_{34} \end{cases} \tag{3.18} $$

where $d_{ij} = \|P_i - P_j\|$ that is the euclidean distance between i^{th} and j^{th} 3D point and α_{ij} represent angle between these two legs (i.e., the same angle between $\overline{C^{C}\hat{P}_i}$ and $\overline{C^{C}\hat{P}_j}$).

From definition of dot product it is known that $\langle p,q\rangle = p^{\top}q = \|p\|\|q\|\cos\alpha$ where α is the angle between p and q vectors, thus $\cos\alpha_{ij}$ can be easily computed considering $p = {}^{C}\hat{P}_i$ and $q = {}^{C}\hat{P}_j$, such that

$$ \cos\alpha_{ij} = \frac{\langle {}^{C}\hat{P}_i, {}^{C}\hat{P}_j\rangle}{\|{}^{C}\hat{P}_i\|\|{}^{C}\hat{P}_j\|} \tag{3.19} $$

replacing ${}^{C}\hat{P}$ using relation 3.5.1 to calculate cosines

$$\cos\alpha_{ij} = \frac{\left\langle {}^{C}\hat{P}_i, {}^{C}\hat{P}_j \right\rangle}{\|{}^{C}\hat{P}_i\| \|{}^{C}\hat{P}_j\|} = \frac{\left\langle \mathbf{K}^{-1}\tilde{p}_i, \mathbf{K}^{-1}\tilde{p}_j \right\rangle}{\|\mathbf{K}^{-1}\tilde{p}_i\| \|\mathbf{K}^{-1}\tilde{p}_i\|} \tag{3.20}$$

With knowledge of cosines $\cos\alpha_{ij}$ and segments d_{ij} it is possible to solve the nonlinear system 3.18 with respect to unknowns l_i with $i = 1,2,3,4$. Solution of such system is discussed in [36]. System 3.18 is converted in another non-linear system using Sylvester resultant, obtaining a system of four degree polynomial equations

$$\begin{cases} g_1 = a_4 l^4 + a_3 l^3 + a_2 l^2 + a_1 l + a_0 = 0 \\ g_2 = b_4 l^4 + b_3 l^3 + b_2 l^2 + b_1 l + b_0 = 0 \\ g_3 = c_4 l^4 + c_3 l^3 + c_2 l^2 + c_1 l + c_0 = 0 \end{cases} \tag{3.21}$$

where $l = l_1^2$ from 3.18 and coefficients $\{a_i, b_i, c_i\}$ can be computed from equations of system 3.18. Based on [36] work solution is achieved through system 3.21 represented in a matrix fashion

$$\begin{bmatrix} a_0 & a_1 & a_2 & a_3 & a_4 \\ b_0 & b_1 & b_2 & b_3 & b_4 \\ c_0 & c_1 & c_2 & c_3 & c_4 \end{bmatrix} \begin{bmatrix} x_0 \\ x_1 \\ x_2 \\ x_3 \\ x_4 \end{bmatrix} = \begin{bmatrix} 0 \\ 0 \\ 0 \\ 0 \\ 0 \end{bmatrix} \tag{3.22}$$

with unknown vector $\mathbf{x} = [x_0, x_1, x_2, x_3, x_4]^\top = [1, l, l^2, l^3, l^4]^\top$, and coefficient matrix $\mathbf{A}$. Homogeneous system 3.22 can be solved using SVD decomposition, assuming that $rank(\mathbf{A} = \min(3,5) = 3)$ decomposition will be

$$\mathbf{U}_{3\times5}\, diag(\sigma_1, \sigma_2, \sigma_3, 0, 0)\, \mathbf{V}_{5\times5} \tag{3.23}$$

Solution of system is a Linear combination of the two last column of matrix $\mathbf{V}$ ($\mathbf{v}^4$ and $\mathbf{v}^5$) as follows

$$\mathbf{x} = \alpha\mathbf{v}^4 + \beta\mathbf{v}^5 \tag{3.24}$$

To find α and β coefficients a knowledge on x_i is incorporated such that

$$\begin{cases} x_i x_j = x_k x_l \\ i + j = k + l \\ 0 \leq i, j, k, l \leq 4 \end{cases} \tag{3.25}$$

by substituting x_i from 3.24 in 3.25 a new equation is obtained

$$b_1^2 \alpha^2 + b_2 \alpha\beta + b_3 \beta^2 = 0 \tag{3.26}$$

where

$$\begin{array}{rcl} b_1 & = & v_i^4 v_j^4 - v_k^4 v_l^4 \\ b_2 & = & v_i^4 v_j^5 + v_i^5 v_i^4 - (v_k^4 v_l^5 + v_k^5 v_l^4) \\ b_3 & = & v_i^5 v_j^5 - v_k^5 v_l^5 \end{array} \tag{3.27}$$

There exists up to seven equation of type 3.26 whether all possible combination of $i + j = k + l$ are considered, with i, j and k, l different for each combination, such as:

$$\begin{array}{rcl} 0+2 & = & 1+1 \\ 3+0 & = & 2+1 \\ 4+0 & = & 2+2 \\ 4+0 & = & 3+1 \\ 4+1 & = & 3+2 \\ 4+2 & = & 3+3 \end{array} \tag{3.28}$$

resulting in seven equation for each i, j, k, l solving it in a matrix fashion

$$\begin{bmatrix} b_1^0 & b_2^0 & b_3^0 \\ b_1^1 & b_2^1 & b_3^1 \\ b_1^2 & b_2^2 & b_3^2 \\ b_1^3 & b_2^3 & b_3^3 \\ b_1^4 & b_2^4 & b_3^4 \\ b_1^5 & b_2^5 & b_3^5 \\ b_1^6 & b_2^6 & b_3^6 \end{bmatrix} \begin{bmatrix} \alpha^2 \\ \alpha\beta \\ \beta^2 \end{bmatrix} = \begin{bmatrix} 0 \\ 0 \\ 0 \end{bmatrix} \tag{3.29}$$

System 3.29 can be solved using SVD decomposition on coefficient matrix, solution of such system is the last column of matrix **V** -which correspond to the vector associated to the smallest singular value- and let is denote it as vector $\mathbf{y} = (y_0, y_1, y_2)^\top$

$$\begin{bmatrix} \alpha^2 \\ \alpha\beta \\ \beta^2 \end{bmatrix} = \begin{bmatrix} y_0 \\ y_1 \\ y_2 \end{bmatrix} \tag{3.30}$$

from equation 3.30 it can be computed a ration α/β as it corresponds either to y_0/y_1 or y_1/y_2. Then, considering constraint $x_0 = 1$ and using equation 3.24 it is obtained

$$1 = \alpha v_0^4 + \beta_0^5 \tag{3.31}$$

once obtained α and β a solution **x** of 3.24 is available, hence a solution l from 3.21 as x_1/x_0 or x_2/x_1 or x_3/x_2 or x_4/x_3. Finally, with solution of 3.21 it is given a solution for 3.18. With $l_1 = \sqrt{l}$ substituted in 3.18 it is possible to solve first three quadratic equations by finding l_2, l_3, l_4. Each equation gives more hen one solution (e.g., complex solution are not considered).

Chapter 4

Laser and Camera cross-calibration

4.1 Method Overview

The extrinsic camera calibration problem can be solved using laser range finders by making correspondences between features seen by laser and the same features seen by the camera. However, to produce correspondences between these heterogeneous sensors is complex and arise several problems.

First of all, the collisions between laser beams and objects are not visible, since standard automotive camera systems do not operate in the same frequency spectrum of the Laser Range Finder LRF. Moreover, camera and laser have different error models, since cameras are affected by perspective projection distortion and laser range measurements provide constant precision over distance.

To deal with these known problems there are different methods present in literature that estimate 3D features such as points that lying on depth edges than the camera pose is measured by minimizing a geometric, rather then an algebraic distance either projecting LRF features in the image plane or projecting camera features in the 3D space.

The procedure in [40] describe a extrinsic calibration algorithm by placing a pla-

nar chessboard at different positions and orientations in front of the sensors, the proposed method solves the problem based on 3D reconstruction of the chessboard and geometric constraints between views from the stereo vision system and the LIDAR. The three principle steps of the approach are: 3D corner points triangulation, 3D plane least-squares estimation, solving extrinsic parameters by applying a non-linear optimization algorithm based on the geometric constraints. To evaluate the performance of the algorithm, experiments based on computer simulation and real data are performed. The proposed approach is also compared with a popular calibration method to show its advantages.

In 2007 Li [41] propose an extrinsic calibration method based on line features detection and a Gauss Newton optimization with a geometrical distance function. The extrinsic parameters are obtained by minimizing the distance form the calculated projection of the intersection point to the projected edges of the checkerboard in the image. This procedure takes advantages by getting easily laser point using well known target object, moreover it performs better results than previous procedure (i.e.:Zhang-Pless [42], and Wasielewski-Strauss [43]). Despite this contribution fits for indoor mobile robot and calibration procedure is achieved by short range distances, it employs a camera with 1024×768 pixel resolution and a laser scanner SICK LMS221-30206 that provides 100 FOV and measurement up to 80 m with accuracy $\pm$50mm, which are usable also for outdoor vehicles.

Another important contribution is provided by [44] in 2011 based on the same ideas of Wasielewski [43] and Li [41], in which extrinsic calibration parameters are estimated by minimizing the distance between corresponding features projected on the 2D image plane. The features are lines as mentioned in [41] although, calibration target is a particular two plane panel arranged in a v-shape. Then weighted measure of distances coupled with a penalizing function are used in order to exclude outliers and estimates extrinsic camera parameters. This procedure extends previous works and improves performances by remove bias at depth edges and removing outliers.

A similar platform to VisLab prototypes is considered in [45] 2008. Using multilayer LIDAR and mono camera system, the proposed method employ a circular ring as calibration target and solve extrinsic and intrinsic camera parameters. Further re-

sults on error propagation and confidence intervals are useful to define the scope of this procedure in the automotive field.

To calibrate the extrinsic parameters, a set of features have to be selected on a known object and a set of constraints equations are derived by associating the measurement data of such features seen by the camera and those perceived by laser range finder. Calibration key points then include features selection, derivation and solution of the constraint equations. Three possible features can be selected for the calibration task:

- Points: using a point feature, the camera pose will be constrained on the surface of a sphere while the coordinates of the point with respect the laser range finder is measured, whether the measurement error is not zero, the camera will be bounded to a spherical ring region. Given the 3D coordinates of the point measured by the laser range finder, the center of the sphere is at the feature point and the radius is the range measured by the LRF. For multiple feature points, the position of the camera is constrained to the intersection of the spherical surfaces.

- Lines:the camera position is constrained in the inner space of a cylinder whose axis is the line feature and radius is the measurement distance of the LRF. For multiple line features, the position of the camera must be constrained in the intersection of the inner space of the cylinders.

- Faces: the position of the camera is in the space between the measured plane and the plane that is parallel to the face measured by the LRF and is apart from the face by the measured distance. For multiple face features, the position of the camera is in the intersection of the spaces.

Since light emitted by lasers is not easily detectable by common cameras it is impossible to exploit point features to establish any correspondence. Thus, three main procedures based on line and plane features are detailed in next sections and then the procedure proposed for this PhD thesis is given.

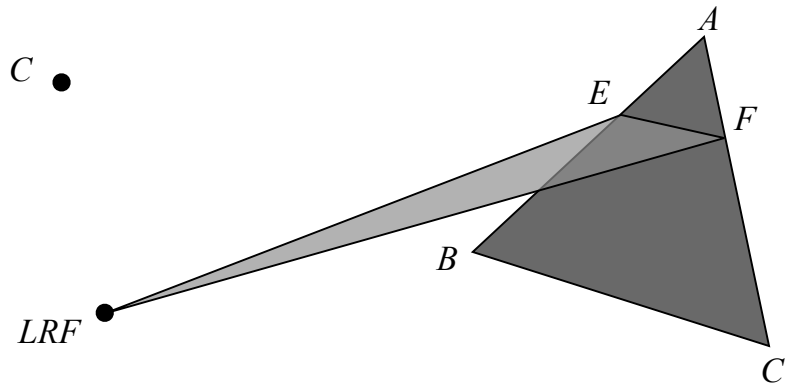


Figure 4.1: The calibration target adopted for line features extraction to cross-calibrate camera C with laser range finder LRF.

4.1.1 Pose from line features

Figure 4.1 shown the classic work condition of such procedure, where LRF slice plane intersects the two edges $\overline{AB}$ and $\overline{AC}$ in two points E and F respectively. The procedure mentioned in [41] uses two kinds of measurement data obtained at different poses of calibration target. One kind of data is the projections of the two perpendicular lines ab and ac on the image plane of the camera, and the other is the measurements of the intersection points E_L and F_L in the LRF reference frame. Assumed a pinhole camera model with intrinsic matrix $\mathbf{K}$ and radial distortion coefficients k (see chapter 2.1 for detail) a point $E_C = (X_E, Y_E, Z_E)^\top$ in camera reference frame can be projected in $mathbfe = (x_e, y_e)^\top$ on the image plane using 2.1.1 and radial distortion

model 2.7. Intersection point E_L is directly measured by the LRF, while its projection point $\mathbf{e}$ must be converted with equation 2.20 to the camera reference frame using unknown rotation matrix $\mathbf{R}$ and translation vector $\mathbf{t}$. Then a distance function $d(\mathbf{e}, \mathbf{ab})$ is defined as a cross product

$$d(\mathbf{e}, \mathbf{ab}) = \frac{\bar{\mathbf{e}} \times \overline{\mathbf{ab}}}{\|\mathbf{ab}\|} \tag{4.1}$$

Same procedure with distance equation 4.1 is applied to the segment AC and intersection F. Then, different distances are obtained for different orientations and positions of the calibration target. The extrinsic parameters calibration problem can be formalized as a optimization problem of finding the optimal solution of the translation vector $\mathbf{t}$ and orientation angles of $\mathbf{R}$ that minimize the sum of the distances, such that

$$\min_{\mathbf{R},\mathbf{t}} \left\{ \sum_i [\, d(\mathbf{e}_i, \mathbf{a}_i\mathbf{b}_i) + d(\mathbf{f}_i, \mathbf{a}_i\mathbf{c}_i) \,]^2 \right\} \tag{4.2}$$

Solution to the nonlinear problem can be achieved by using the Gauss Newton algorithm (A.2.2). With initial guess given by a rough measurement of sensor parameters, otherwise it can also be obtained by the closed-form solution proposed in [42].

4.1.2 Pose from 2D pattern

A planar calibration target can be used for extrinsic camera parameters calibration, the target plane is shown in figure 4.2 on the Z plane of world coordinates. Let $\mathbf{r}_i$ be the i-th column of rotation matrix $\mathbf{R}$, thus vector $\mathbf{r}_3$ represents the surface normal vector $\mathbf{n}$. Without loos of generality, it is assumed the upper-left corner of the calibration target $\mathbf{t}$ as origin of the world reference system and $\mathbf{O}_C$ the center of projection of the camera. Let consider $\mathbf{t}$ as a translation vector that represents relative position of the calibration target upper-left corner in the camera reference frame, since it is visible from the camera and it may not be visible by the LRF. As shown in figure 4.2 points $\mathbf{P}_C$ and $\mathbf{t}$ lie on the target plane from the camera reference frame a vector

$$\vec{v} = \mathbf{P}_C - \mathbf{t} \tag{4.3}$$

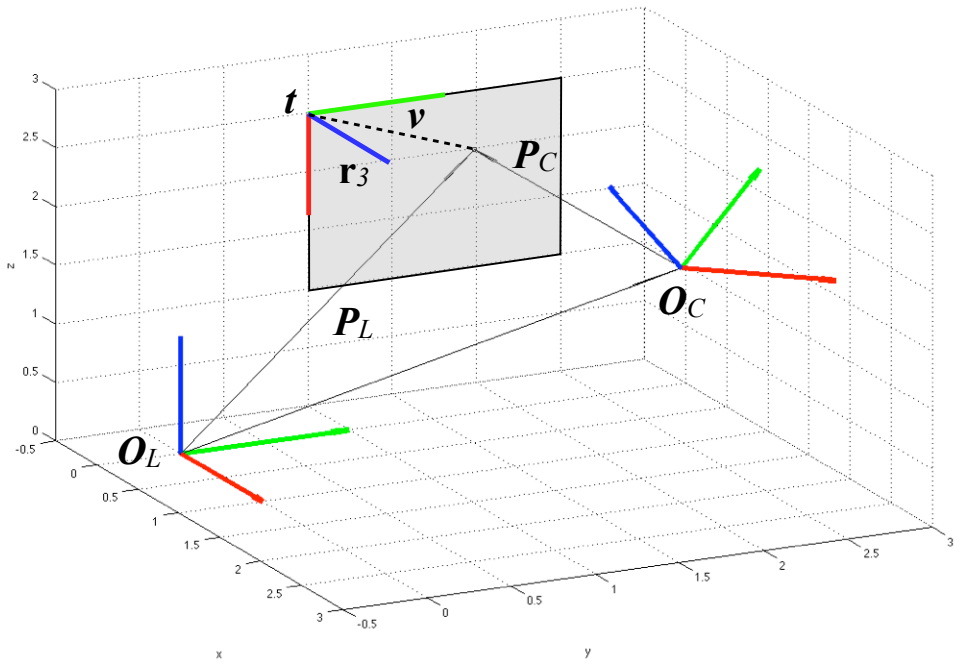


Figure 4.2: Plane calibration target adopted to cross-calibrate camera. Geometrical interpretation is given with a camera reference frame $\mathbf{O}_C$, a laser scanner reference frame $\mathbf{O}_L$ and a plane target defined by a point $\mathbf{t}$ and a normal vector $\mathbf{r}_3$.

since $\mathbf{r}_3$ and v are orthogonal vectors the following equation can be derived

$$\mathbf{r}_3 \cdot \vec{v} = 0 \tag{4.4}$$

Using relation 2.20 of two reference frames and by taking 4.3 in 4.4 results

$$\mathbf{r}_3^\top (\mathbf{R}_l^c \mathbf{P}_l + \mathbf{t}_l^c - \mathbf{t}) = 0 \tag{4.5}$$

by considering laser point $\mathbf{P}_l = [X_l, Y_l, Z_l, 1]^\top$ equation 4.5 can be reviewed in a matrix fashion

$$\mathbf{r}_3^\top [\mathbf{R}_l^c \quad \mathbf{t}_l^c - \mathbf{t}] \begin{bmatrix} X_l \\ Y_l \\ Z_l \\ 1 \end{bmatrix} = 0 \tag{4.6}$$

Here follows a closed-form solution used by Huang in [46] to obtain orientation an translation components $[\mathbf{R}|\mathbf{t}]$ followed by a nonlinear optimization. Supposing that for each pose of the calibration target there are n laser points on the target plane, with $\mathbf{P}_{l,i} = [X_{l,i}, Y_{l,i}, Z_{l,i}]^\top$ with $i = 1, 2, \ldots N$ coordinates respectively.

The closed-form solution becomes a linear system of equations of time $\mathbf{Ax} = \mathbf{0}$ where $\mathbf{A}$ is an $n \times 12$ matrix 4.7 and unknown vector $\mathbf{x}$ defined in 4.8

$$\mathbf{A} = \begin{bmatrix} \mathbf{n}_{31}\mathbf{P}_{l,1}^\top & \mathbf{n}_{32}\mathbf{P}_{l,1}^\top & \mathbf{n}_{33}\mathbf{P}_{l,1}^\top \\ \mathbf{n}_{31}\mathbf{P}_{l,2}^\top & \mathbf{n}_{32}\mathbf{P}_{l,2}^\top & \mathbf{n}_{33}\mathbf{P}_{l,2}^\top \\ & \cdots & \\ \mathbf{n}_{31}\mathbf{P}_{l,N}^\top & \mathbf{n}_{32}\mathbf{P}_{l,N}^\top & \mathbf{n}_{33}\mathbf{P}_{l,N}^\top \end{bmatrix} \tag{4.7}$$

$$\mathbf{x} = \begin{bmatrix} \mathbf{r}_{11} & \mathbf{r}_{12} & \mathbf{r}_{13} & \Delta_x & \mathbf{r}_{21} & \mathbf{r}_{22} & \mathbf{r}_{23} & \Delta_y & \mathbf{r}_{31} & \mathbf{r}_{32} & \mathbf{r}_{33} & \Delta_z \end{bmatrix} \tag{4.8}$$

where $\mathbf{r}_{ij}$ is the element on the i-th row, j-th column in matrix $\mathbf{R}_l^c$, plane normal vector is $\mathbf{n} = [\mathbf{n}_{31}, \mathbf{n}_{32}, \mathbf{n}_{33}]^\top$ and $\Delta = \mathbf{t}_l^c - \mathbf{t} = [\Delta_x, \Delta_y, \Delta_z]^\top$.

Solution can be obtained using the least square method, avoiding null solution $\mathbf{x} = \mathbf{0}$ by the constraint $\mathbf{r}_{31}^2 + \mathbf{r}_{32}^2 + \mathbf{r}_{33}^2 = 1$. This condition is singularity free, since it represents the third row of the rotation matrix $\mathbf{R}_l^c$.

$\|\mathbf{Ax}\|$ can be minimized using Lagrange method. Let split the $\mathbf{x}$ vector in two parts: $\mathbf{r}_3 = [\mathbf{r}_{31}, \mathbf{r}_{32}, \mathbf{r}_{33}]^\top$, and $\mathbf{r}_9$ with remain nine elements. Lagrange equation is given by

$$L = \mathbf{A}_9\mathbf{r}_9 + \mathbf{A}_3\mathbf{r}_3 + \lambda(\mathbf{r}_3^\top\mathbf{r}_3 - 1) \tag{4.9}$$

$$\lambda\mathbf{r}_3 = (\mathbf{A}_3^\top\mathbf{A}_3 - \mathbf{A}_3^\top\mathbf{A}_9(\mathbf{A}_9^\top\mathbf{A}_9)^{-1}\mathbf{A}_9^\top\mathbf{A}_3)\mathbf{r}_3 \tag{4.10}$$

$$\mathbf{r}_9 = -(\mathbf{A}_9^\top \mathbf{A}_9)^{-1} \mathbf{A}_9^\top \mathbf{A}_3 \mathbf{r}_3 \tag{4.11}$$

It is known that $\mathbf{r}_3$ corresponds to the eigenvector of the symmetric positive defined matrix defined in 4.10 associated with the smallest eigenvalue. Then, $\mathbf{r}_9$ is available from 4.11 once $\mathbf{r}_3$ is computed. Finally rotation matrix and translation vector are obtained from $\mathbf{x}$. Due to noise on data the rotation matrix may not be orthogonal and satisfy orthogonality condition $(\mathbf{R})\mathbf{R}^\top = \mathbf{I}$, therefore a new rotation matrix $\hat{\mathbf{R}}_l^c$ have to be computed such that Frobenius norm of the distnace $\|\hat{\mathbf{R}}_l^c - \mathbf{R}_l^c\|_F$ is minimum.

Closed-form solution is affected by measurement noise, thus result can be refined with a maximum likelihood function using multiple M poses of the calibration target. The euclidean distance 4.5 is minimized such that

$$\min_{\mathbf{R}_l^c, \mathbf{t}_l^c} \sum_i^M \frac{1}{n_i} \left(\sum_j^{n_i} ((\mathbf{r}_3^i)^\top (\mathbf{R}_l^c \mathbf{P}_{l,j}^i + \mathbf{t}_l^c) - (\mathbf{r}_3^i)^\top \mathbf{t}_i)^2 \right) \tag{4.12}$$

where $(\mathbf{R}_l^c \mathbf{P}_{l,j}^i + \mathbf{t}_l^c)$ corresponds to point $\mathbf{P}_{l,j}^i$ in the camera reference system (detail in 2.20).

4.1.3 Pose from circular targets

This method is related to Rodriguez et al. [45] work, it proposes an extrinsic calibration method, vehicle oriented, for sensor suits composed by a multi-layer laser scanner and cameras based on circular calibration target. The particular configuration used for intelligent vehicle is described by two critical conditions.

- The first critical point consists arise considering the necessity to have an important relative distance between the sensors and the calibration target. Therefore, high relative distance between cameras and the calibration target requires a large-size calibration target in order to ensure the accuracy of the estimation pose in the camera frame.
- A second key point is related to the laser scanner detection error (described in refchapter:laser) that measures the surface distance of the calibration target. This increase partially caused by the laser beam impacts on black-white zones.

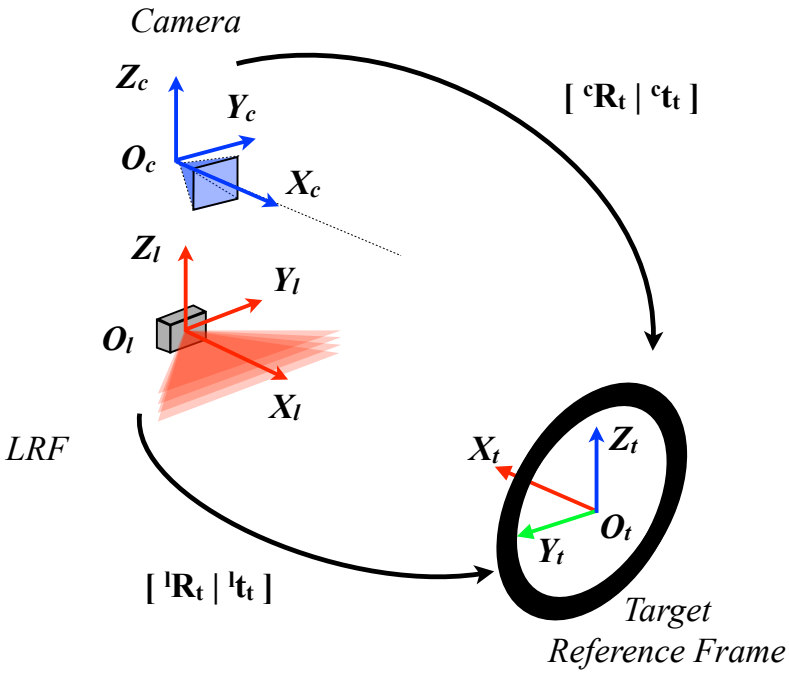


Figure 4.3: Circular calibration target adopted to cross-calibrate camera from a multi-layer laser scanner. Geometrical interpretation is given with a camera reference frame $\mathbf{O}_c$, a laser scanner reference frame $\mathbf{O}_l$ and a target reference frame $\mathbf{O}_t$. Circle target is modeled by its center C and the normal vector of its plane N.

The solution proposed in [45] consists in estimating different poses of the calibration object detected simultaneously by the camera and the multi-layer laser scanner. A minimum of 6 poses have to be estimated in the laser and the camera frame in order to get all degrees of freedom A.3.3. Each pose of the calibration target is parameterized by the 3D coordinates of the circle center and the normal vector of its plane. Then, a first estimation of the rigid transformation is obtained by solving the usual absolute orientation problem 3.5. This solution consists in determining the relationship between the two coordinate frames using sets of corresponded features (i.e.,

circle centers C of each pose). Finally a non-linear 3D minimization is done in order to refine the estimated extrinsic parameters. The whole algorithm is presented in 1.

Algorithm 1 Circle-based Extrinsic camera registration

Require: 4-plane LIDAR.

Require: Camera.

Ensure: Pose of camera frame w.r.t. laser frame

1: **for** $i = 1 \rightarrow \max(6,n)$ **do**
2: Estimate the ith laser calibration pose $[{}^{l}N^{l}C]_i$ (see A.2.4)
3: Estimate the ith camera calibration pose, $[{}^{c}N^{c}C]_i$
4: **end for**
5: Compute a first guess, $[\mathbf{R}_0, \mathbf{t}_0]$, for the lidar-camera rigid transformation using linear solution
6: Match the 3D circle poses estimation
7: **repeat**
8: Non-linear minimization using Levemberge-Marquardt algorithm
9: Robust noise variance estimation σ^2 based in non-linear minimization residuals
10: Weighting matrix W update from M-estimator
11: **until** convergence of $[\mathbf{R}_0, \mathbf{t}_0]$
12: Estimate the confidence intervals

Before estimating the calibration target in the laser frame the authors performs multiple scan acquisitions in order to reduce measure uncertainty and get more robust detection. Then, for each scan only scan points ${}^{l}\mathbf{P}_i$ that lie on the edges of the target are kept for the next fitting phase as stated in appendix A.2.4.

With several references on camera calibration methods based on concentric circles[47, 48], the 3D position of the calibration target is estimated in the camera frame. To compute this point, intrinsic camera parameters have to be known to apply a nonlinear ellipse fitting algorithm stated in [49], thus with two conic matrix that represents the inner and the outer circles shown in figure 4.3 the 3D circle center ${}^{c}C$ and the normal vector ${}^{c}N$ are recovered as stated in [48].

To deal with iterative solutions an initial guess solution is formulated as a classical absolute orientation problem following the closed-form solution developed by Arun et al. [27] that obtain the rigid transformation from the singular value decomposition of the correlation matrix of the cent points

$$\Sigma = \left[{}^{l}C_i - \overline{{}^{l}C} \right] \left[{}^{c}C_i - \overline{{}^{c}C} \right]^{\top} = \mathbf{U}\,\mathbf{S}\,\mathbf{V}^{\top} \tag{4.13}$$

Where ${}^{s}C_i$ are the coordinates of the 3D circle center point set estimated from the ith pose by the sensor measurements, and $\overline{{}^{s}C}$ is the 3D centroid of the target computed as average value. Thus, the rotation matrix is obtained

$$\mathbf{R}_0 = \mathbf{U}^{\top}\,\mathbf{V} \tag{4.14}$$

and translation vector $\mathbf{t}_0$

$$\mathbf{t}_0 = \overline{{}^{c}C} - \mathbf{R}_0 \cdot \overline{{}^{l}C} \tag{4.15}$$

The rigid transformation obtained is a linear minimization of the Euclidean distance error between the 3D circle center point sets that is generally a good starting guess of the extrinsic calibration. To refine this guess n 3D circles are generated with respect the camera frame from n different target poses. It consists in computing m points of each estimated circle pose by using the 3D circle center and an orthonormal base lying in circle's plane. After generating all the initial guess $[\mathbf{R}_0, \mathbf{t}_0]$ on the n poses, a match between the 3D points of the camera and laser transformed estimations for every pose is computed by using the nearest neighbor criterion. Basically, the final results $[\mathbf{R}, \mathbf{t}]$ is obtained by minimizing a nonlinear objective function

$$\varepsilon = \sum_{i=1}^{n} \sum_{j=1}^{m} \mathbf{W} \cdot D_{ij}^{2} \tag{4.16}$$

where $D_{ij} = \| c_{i,j}^{P} - (\mathbf{R}\,{}^{c}P_{i,j} + \mathbf{t}) \|$ is the Euclidean distance residual of the points after applying the rigid transformation (in order to operate in the camera reference frame), and $\mathbf{W}$ is a weighted matrix. The results are obtained by using a robust M-estimator algorithm for calculating the robust weights as stated in [50] and the Levemberge-Marquardt algorithm.

4.2 Proposed Method

This thesis presents a method for solving the extrinsic calibration between camera and multi-layer laser scanner for outdoor multi-sensorized vehicles. The proposed method is designed for intelligent vehicles within the autonomous navigation task where usually distances between sensor and targets become relevant for safety reasons, therefore high accuracy across different measures must be kept.

The calibration procedure takes advantage of triangular shapes still present in scenarios, it recovers three virtual points as target pose in the laser and camera reference frames and then compute extrinsic information of each camera sensor with respect to a laser scanner by minimizing a geometric distance in the image space. To test algorithm correctness, and accuracy a set of simulations are used reporting absolute error results and solution convergence, then tests on robustness and reliability (i.e., outliers management) are based on a wide set of datasets acquired by VIAC prototypes.

In the mobile robotics field and particularly in the intelligent vehicle context, the autonomous navigation task is achieved by heterogeneous suit of perception sensors in order to acquire more rich information content than using only active or passive sensors. For their complementary purposes LIDARs and camera are widely used together on such vehicles, Zehang et al. in [51] reviews the problem of on-road vehicles detection using optical sensors and details pros and cons by introducing active sensor in object perception. In the last year several advanced driver assistance systems coupled camera and laser devices, such as vehicle detection [52] where LIDAR provide good range information to define obstacles while it is hard to recognize vehicles and cameras that allows better recognition but does not provide good range information.

Since laser and camera are complementary sources, in order to fuse their information the extrinsic calibration of these heterogeneous sensors is required. Thus, this approach presents an extrinsic calibration algorithm between slave and master sensors which are a vision system and a multi-layer LIDAR respectively. By detecting a planar triangular board standing either beside the road, or in front of the sensors, the proposed method solves the Pose Estimation Problem 3.5 (or Location Determination Problem) in the laser frame by a 3D reconstruction of the triangular targets,

than minimizing its projection in the image space. Two main steps compose such algorithm: a target detection in each sensor frame and triangle vertex estimation, and compute extrinsic in a optimization phase that minimizing a geometric distance. A preliminary comparison between different extrinsic estimation techniques based on geometric and algebraic minimization functions (e.g., SVD rigid body method or Iterative Closest Point) lead us to chose for a non-linear optimization algorithm based on geometrical constraints in the image space.

4.2.1 Problem statement

Perception system employed on vehicle is composed by a four layer laser scanner and a vision system both placed in front of the vehicle facing the road (i.e. four layer LRF and a stereo system in figure 4.10 detailed in 4.3). The calibration target designed for sensor registration consists of a planar isosceles triangular shape with size $w_t = 0.6$m, $h_t = 1$m (as depicted in figure 4.4) of a specific white bright color to simplify detections of the shape into the image frame. Then, a Target Reference Frame (TRF), Camera Reference Frame (CRF) and Laser Reference Frame (LRF) are defined, notice that a permutation have to be considered to convert from sensor frame to image frame as shown in figure 4.4.

Basic assumption is that triangular calibration target will be fully observable by a laser scanner and a camera, since its geometry is known triangle vertexes can be computed and used as virtual constraints. Partial observed object lead to errors in optimization phase. The transformation of target vertex points from LRF to CRF is represented by a rigid transformation composed by a 3×3 rotation matrix R and a translation vector t

$$P_c = {}^c R_l \cdot P_l + {}^c t_l \tag{4.17}$$

Given two set of N 3D points X_i and Y_i in the LRF frame and camera frame respectively, where correspondences between points are known. The problem statement consist to solve the *rigid pose transformation* $[R_\Theta|t]$ of the camera frame with respect

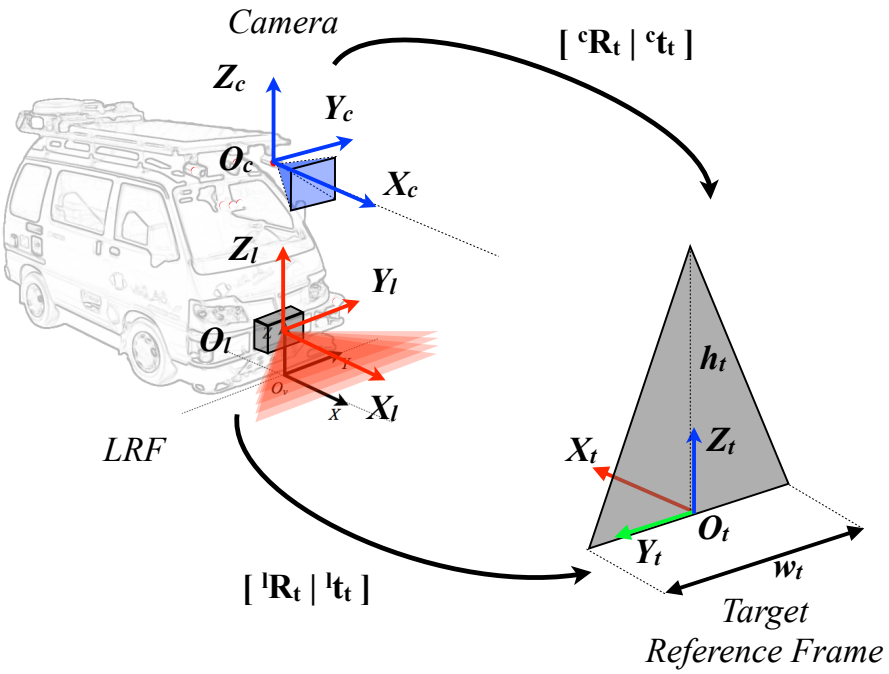


Figure 4.4: Reference frames. CRF and LRF are placed in front of the vehicle facing the same region of interest and TRF is placed in the middle bottom of the calibration target. Target must be seen completely and simultaneously by laser and camera.

to the LRF frame in order to minimize the overlapping error e between X and Y,

$$e = \min_{t,\Theta} \sum_{i}^{N} \|X - s(R_\Theta Y + t)\|^2 \tag{4.18}$$

where $t = (t_x, t_y, t_z)^\top$ is translation vector and $\Theta = (roll, pitch, yaw)^\top$ is a rotation vector.

4.2.2 Extrinsic Calibration

The proposed method is based on three main steps that are explained in the next three sections: acquiring synchronized data from laser and camera (i.e., with the aid of an hardware trigger), then select regions of interest and estimate position of calibration targets with respect to cameras and laser frames, in which targets pose is computed in 2D and 3D respectively. Finally, an optimization procedure based on geometric minimization involves image and laser correspondences simultaneously by fitting virtual points.

Algorithm 2 Extrinsic camera registration

Require: 4-plane LIDAR.

Require: Single or stereo camera.

Ensure: Pose of camera frame w.r.t. laser frame

1: Acquiring simultaneously camera and laser frames
2: Detect targets in laser frame and generate 3D target pose
3: Detect targets in camera frame and generate 2D target pose
4: Estimate rigid transformation $[R|t]$ of cameras frame w.r.t. laser frame

4.2.3 Target Detection in laser frame

In this section detail on the range images acquired by laser scanner is given. In this studies approaches are based on four layer LIDAR SICK or similarly IBEO Lux that have narrow elevation field of views and high scanning resolution.

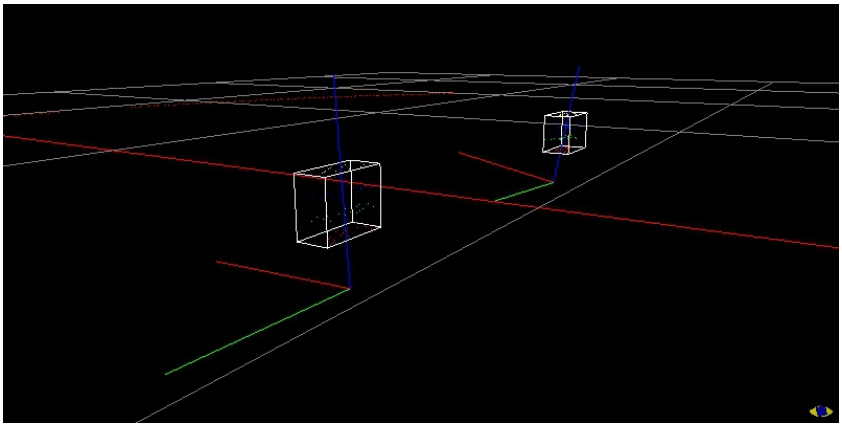


Figure 4.5: Laser point clusterization with Euclidean cluster extraction algorithm.

The first step is to filter out far points (up to 10m), since it is not possible to detect related target objects in image with high accuracy. Then, a clusterization and recognition phase groups raw laser points by an euclidean distance policy and all only opportune planar thin object will be kept (as shown in figure 4.5).

Next, given a set of N potential targets T_j with $j = \dots N$ a procedure to estimate pose of each T_j is computed. RANSAC fitting is used to generate target plane π represent by a centroid C and a normal vector n model.

$$\pi : C + nx = 0 \tag{4.19}$$

All points X_i^j owned by T_j are projected on π, getting a new set of points X_i^π and a plane reference frame $[R^\pi | t^\pi]$ apply to T_j, where $R^\pi = [x_\pi, x_\pi \times n, n]$ and $t^\pi = \overline{X_i^j}$. The virtual point O_t (shown in figure 4.4) in T_j will be computed on π considering only two edge point sets S_1 and S_2 closer to target segments $\overline{AC}$ and $\overline{AB}$ respectively. The procedure to obtain O_t is composed of two steps: first, an intersection point $A = (A_x, A_y)^\top$ is obtained by intersection of two polar lines r_1 and r_2

$$\begin{aligned} r_1 : x\cos(\theta_1) + y\sin(\theta_1) &= \rho_1 \\ r_1 : x\cos(\theta_2) + y\sin(\theta_2) &= \rho_2 \\ \theta_2 = \theta_1 + \alpha & \end{aligned} \tag{4.20}$$

solving a constrained non-linear optimization problem defined as loss function e in equation 4.21 with a constrained angle α,

$$e = LS_{S_1}(r_1) + LS_{S_2}(r_2) \tag{4.21}$$

solution of equation 4.22 is guaranteed by constrain α.

$$\begin{pmatrix} \cos(\theta_1) & \sin(\theta_1) \\ \cos(\theta_2) & \sin(\theta_1) \end{pmatrix} \begin{pmatrix} A_x \\ A_y \end{pmatrix} = \begin{pmatrix} \rho_1 \\ \rho_2 \end{pmatrix} \tag{4.22}$$

Finally point O_t is obtained by shifting of h from A toward $\overline{BC}$ then through the point cloud X_i^j. Resulting a target reference $[R^T|t^T]$ frame as follow in equation 4.23

$$\begin{aligned} R^T &= [n, v, v \times n,] \\ t &= O_t \end{aligned} \tag{4.23}$$

4.2.4 Target Detection in Camera Frames

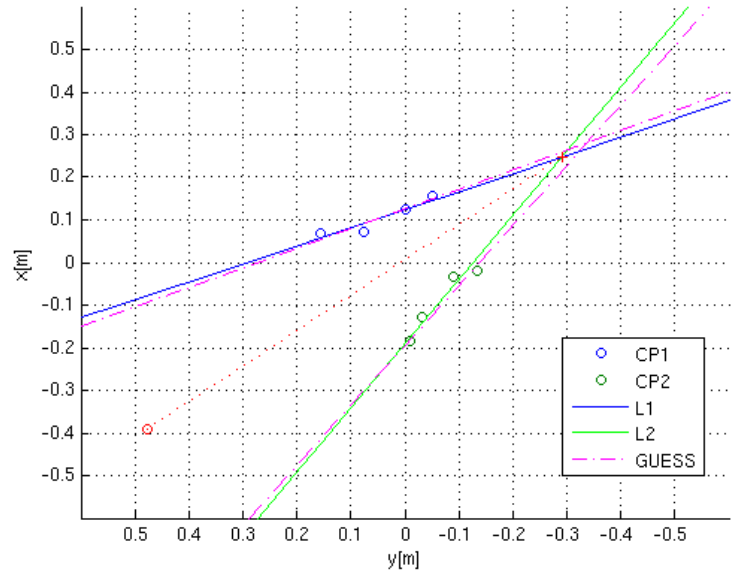
Aim of this procedure is to recover target pose in image frame with less assumptions. Two procedures have been described to recover targets position in camera reference frame based on monocular view and stereo images.

Mono images

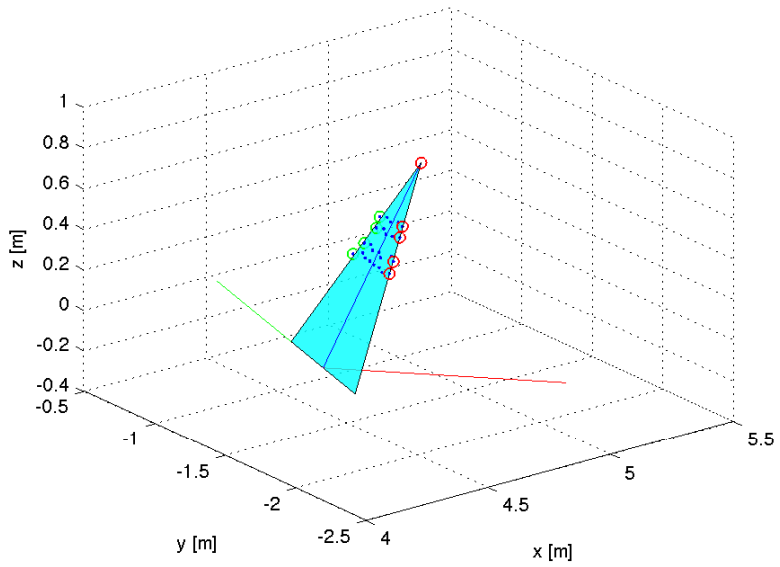
In this section the targets identification procedure based on singular camera is detailed. To find a triangular target in the image is necessary to use a image processing algorithm. There are not assumption on the target orientation and position and moreover it is not possible to compute distances from uncalibrated sensor. Than first step of the automatic calibration procedure is to recognize all triangular shape in the scene with no constraints on the pose of such target.

The overall schema of the algorithm applied to find triangular targets on a 2D image I is showed in algorithm 3.

Firstly a preprocessing stage involves sequentially color filtering, edges extraction, and an Hough transformation that are applied to I generating image E and accumulator space $H(\rho, \theta)$ to be able to detect lines on I (as shown in *Preprocessing*



(a)



(b)

Figure 4.6: Target detection with laser data. Plane projection and temporary reference frame shown in (a). On 2D plane triangle fitting and final result in (b).

function of algorithm 3). At the end of preprocessing stage parametric equations L of linear components are computed from *local maxima* extraction on H and applied on I by relation 4.24.

$$x\cos(\theta)+y\sin(\theta)=\rho \tag{4.24}$$

Second algorithm stage is concerning match between lines L obtained by the previous stage and edges in E. Hence, a set of lines L_i are obtained as intersection between E and L where white pixel amount over each line is higher with respect to a specific threshold set with respect to the target dimensions in pixels. This stage is detailed on the $ExtractLines$ function of the algorithm 3. According to the line extraction algorithm presented in [53], the $SearchTargets$ procedure is able to extract N segments S_i for each line L, and compose it in a large ones with a double thresholded algorithm. Finally, a set of lines S_i, each with a unique segment, is processed by a targets finder step. This step tests each possible triangle composed as line triplet $t=\{S_i,S_j,S_k\}\ \forall i,j,k$ with $i\neq j\neq k$ by matching couple of segment vertex, whether all vertex couples differ each other by less than a Manhattan threshold an intersection point of the lines will be computed. Only when three valid intersection points are obtained, a triangular shape is detected by algorithm.

Algorithm 3 2D triangular shape detection on mono images

Input: 2D image $I=\{x_{i,j},i=1\ldots w,j=1\ldots h\}$.

Output: Set T of n vector T_i with $i=1\ldots n$, where $T_i=[x_1,x_2,x_3]^\top\in I$.

```
1:  procedure SEARCH TARGETS(I,E,H)
2:      [E,H] ← Preprocessing(I)
3:      [L] ← ExtractLinesfrom(H)
4:      Segments S ← L∩E
5:      for all tuple t = {S_i,S_j,S_k}wherei ≠ j do
6:          if Are closer segments ? then
7:              T ← t
8:          end if
9:      end for
10: end procedure
```

Stereo images

In this procedure the relative pose between left and right camera are supposed to be known in order to assume that distances in the sensor frame are correct and only the absolute pose need to be recorded. The target object detection algorithm used for stereo rig systems is equal to the laser procedure detailed in 4.2.3 since laser and camera sensors gather same range information, however a different error distribution is present and must be consider in model extraction as it is displayed in figures 4.7(a) and 4.7(b).

4.2.5 Pose Estimation

Point $\mathbf{x}$ in LRF is represented on image plane with pixel $\mathbf{p}$, as follow

$$\mathbf{p} = F \cdot \mathbf{x}$$

F is composed by a rigid transformation

$$\mathbf{x}' = T \cdot \mathbf{x}$$

with $T = [R_\Theta | t]$ shown in equation 4.17 and a perspective projection K.

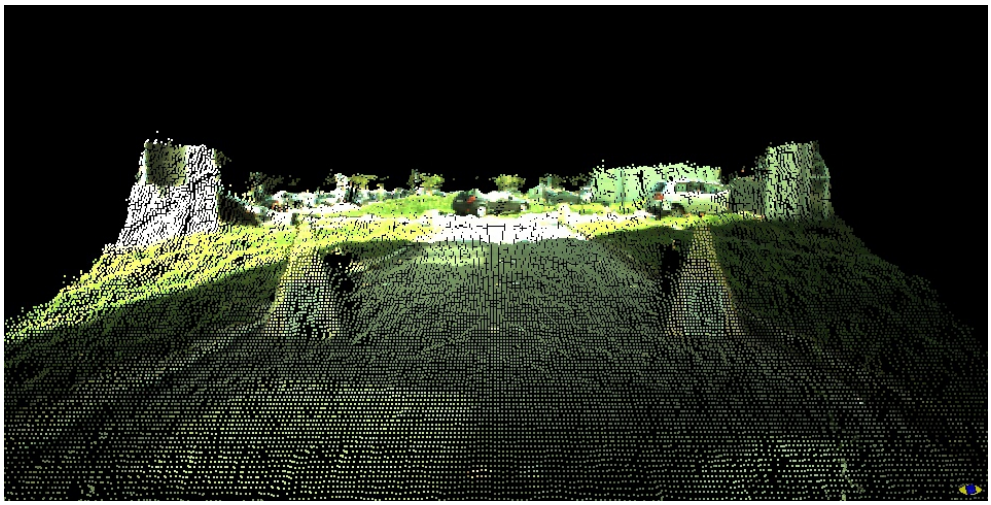
$$\mathbf{p} = K \cdot \mathbf{x}'$$

where K is a pinhole model defined as follow

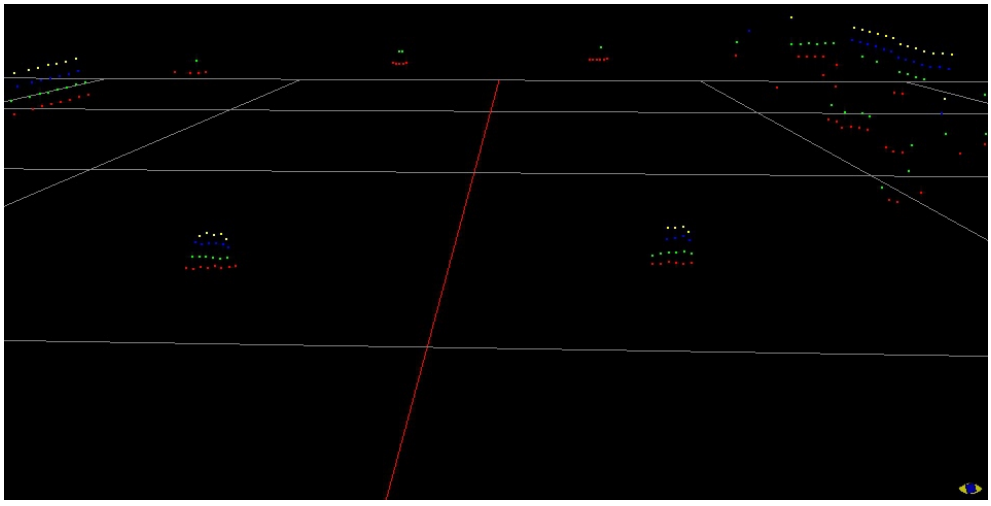
$$K = \begin{pmatrix} k_u & s & u_0 \\ 0 & k_v & v_0 \\ 0 & 0 & 1 \end{pmatrix}$$

To be able to compute the extrinsic parameters in T intrinsic parameters pixel focal length k_u, k_v, principal point (u_0, v_0) and a *skew* s must be known. Extrinsic parameters registration problem can be formulated as a non-linear problem of finding the optimal solution of rotation vector Θ and translation vector t that minimize the euclidean distance between image points.

$$\min_{\Theta,t} \|\mathbf{p}_i - F\mathbf{x}_i\|_2 \tag{4.25}$$

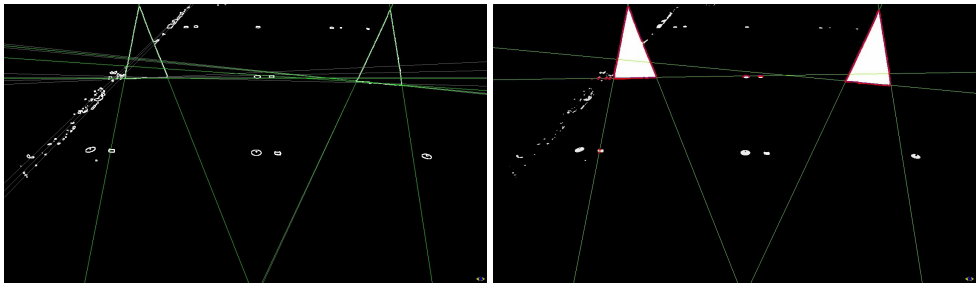


(a) Stereo vision

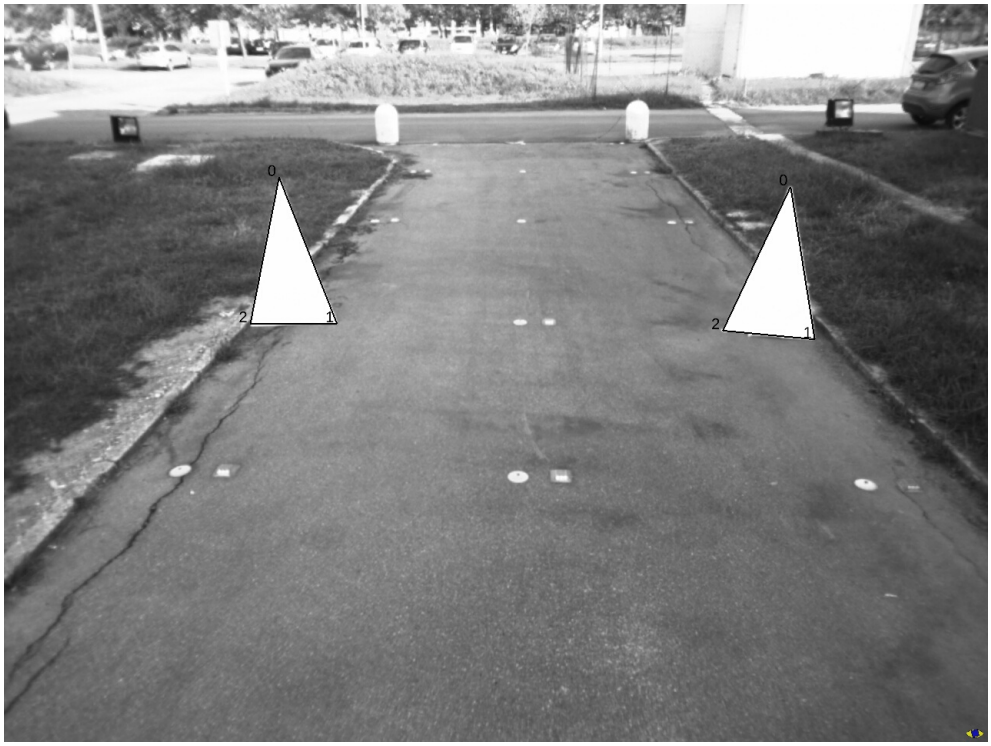


(b) Laser vision

Figure 4.7: 3D point cloud obtained from disparity image in (a) and laser point in (b).



(a) (b)



(c)

Figure 4.8: Image processing phases.

between image point $\mathbf{p}_i$ and image projection of laser points $\mathbf{x}_i$, where index i represent the i-th constraint pair $< \mathbf{p}, \mathbf{x} >$ that match a triangle vertex in CRF and LRF respectively. The non-linear optimization is solved by using the Levenberg-Marquardt algorithm.

4.3 Experimental results

Experiments on synthetic and real data have been conducted to evaluate algorithms behaviors.

4.3.1 Experiments on synthetic data

To test algorithm correctness, and accuracy a set of simulations are used reporting absolute error results and solution convergence, then tests on robustness and reliability (i.e., outliers management) are based on datasets acquired by VIAC prototypes. The simulator engine is a separated software provided by TASS, called PreScan® that is used coupled with Matlab Simulink®. Ground truth of sensors relative position $[R_\Theta|t]$ is used to evaluate procedure correctness, with a translation vector $t = (-2.0, 0.1, 0.98)^\top$ and a rotation vector $\Theta = (0.0, 0.0, 0.0)^\top$. Simulated camera had a resolution of 1280×960 pixels, 7.50 mm of focal length, 25 Hz frame rate, with a 1/2" (6.4×4.8 mm) CCD sensor, and without distortion. Simulated laser scanner had 85 degrees of azimuth FOV with 340 beams, and 3.2 degrees of elevation FOV with 4 beams. Laser provided data at frequency of 25 Hz. Concerning system resolution, angular azimuth cell size was 0.125 degree, and angular elevation cell size was 0.8 degrees, with a 0.04 m range cell size.

Relative Pose Error [mm]	80
Relative Orientation Error [degree]	0.3
Number of pose	25

Table 4.1: Results on simulated data.

Table 4.2: Worst case test on simulations.

	Value [m]	$\sigma_1 = 4.2159$	Confidential interval	GT
t_x	-1.819	± 0.096	[-1.916, -1.723]	-2
t_y	-0.022	± 0.224	[-0.246, 0.202]	-0.1
t_z	1.080	± 0.279	[0.801, 1.360]	0.98
	Value [rad]	$\sigma_1 = 4.2159$	Confidential interval	GT
Roll	0.02578	± 0.0241386	[0.00164, 0.0499]	0
Pitch	-0.07278	± 0.0512392	[-0.12402, -0.02154]	0
Yaw	-0.01621	± 0.039511	[-0.05572, 0.02330]	0

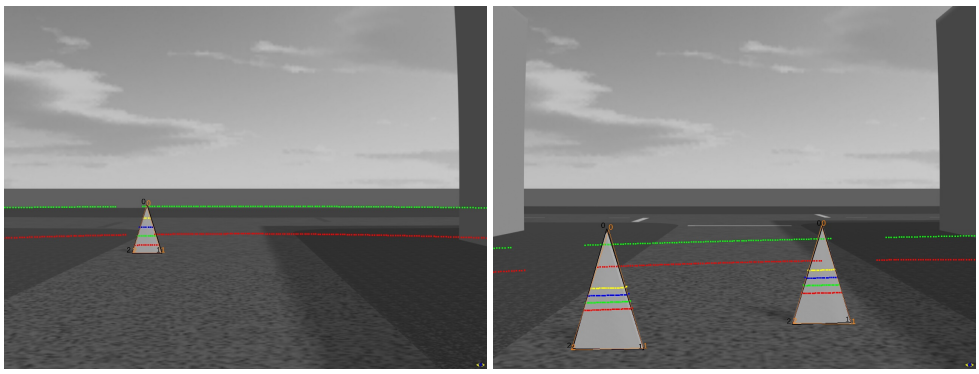
Tables 4.1 and 4.2 refers to an average value of extrinsic parameters that is evaluated with minimum number of poses.

With 2 targets per frame as shown in figure 4.9.(b), position error (figure 4.9.(d)) decreases up to 20 cm when more 5 poses are used. In this case orientation error in figure 4.9.(f) decrease in a range of ± 1 degree for each angles. In the other hand, test in figure 4.9.(a) considers one target per frame and proves that error on yaw angle does not converge, affecting other parameters convergence. This behavior is shown in 4.9.(c)(e).

4.3.2 Experiments on real data

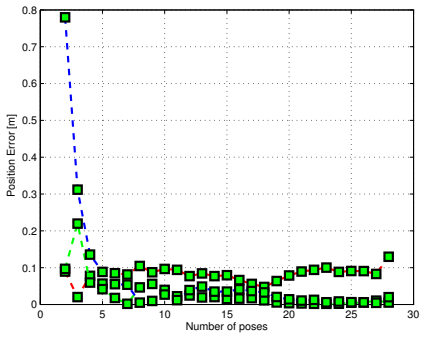
Experiments on real data were made with an intelligent vehicle (figure 4.10.(a)) equipped by VisLab during VIAC expedition. Vehicle was equipped with a autonomous driving state-of-art suite of sensors as detailed in [54]. Concerning computer vision navigation systems, perception tasks were made by a 3 synchronized cameras panoramic vision system, a front and rear stereo cameras and three different mounted laser scanners suites to perform 360 degrees environmental perception.

Two stereo rigs, shown in figure 4.10.(b), was used for experimentation, a large baseline couple of PointGrey Firefly cameras, and a PointGrey Bumblebee X3 for the

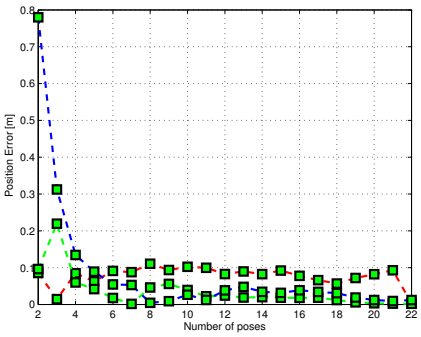


(a) Simulation with a target.

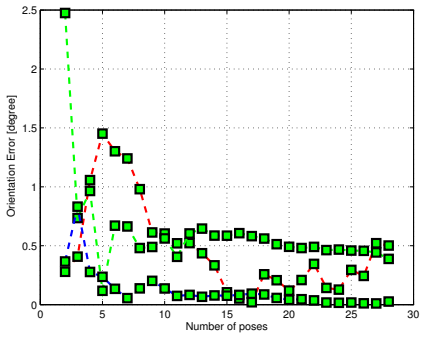
(b) Simulation with 2 targets.



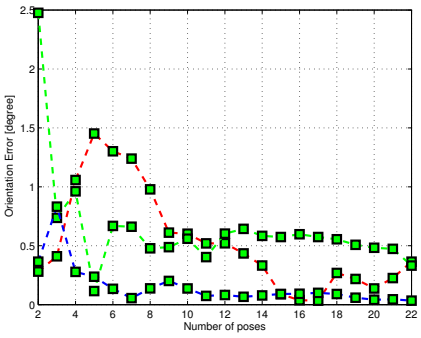
(c) Position errors with 1 target.



(d) Position errors with 2 targets.



(e) Orientation errors with 1 target.



(f) Orientation errors with 2 targets.

Figure 4.9: Errors on position and orientation on simulated data. In the orientation error graphs dash line representing yaw is in blue, roll in red, and pitch in green.

smallest one. The large stereo rig had a 0.80 m baseline, each cameras had 752×480 pixels resolution. Point Grey Bumblebee X3 had a 24 cm baseline, a resolution of 1280×960 pixels, and a 3.8 mm focal length.

Laser scanner used for experiments was a 4-layers Sick LD-MRS-400001. Laser scanner had 3 different angular resolution 0.125°/ 0.25°/ 0.5° with scanning frequency of 12.5 Hz / 25 Hz / 50 Hz respectively. Distance resolution was 40 mm, and the statistical error σ_1 was 100 mm with a range up to 50 m at 10% reflectivity.

Real test was conducted with 2 calibration targets per frame in a fixed position 5 m far from vehicle, detailed in figure 4.10.(c). Results on 20 frames is shown in Table 4.3, that highlights imprecision concerning target pose estimation in LRF due to high level of noise. This problem can be solved acquiring frame with different targets pose.

	Value [m]	$\sigma_1 = 6.45715$	Confidential Range
t_x	-0.162986	±0.0416555	[-0.204642, -0.121331]
t_y	-0.0285926	±0.0884247	[-0.117017, 0.0598321]
t_z	2.47908	±0.0884488	[2.39063, 2.56752]
	Value [rad]	$\sigma_1 = 6.45715$	Confidential Range
Roll	-0.0108902	±0.00814383	[-0.019034, -0.00274638]
Pitch	0.556509	±0.0190649	[0.537444, 0.575574]
Yaw	0.00645968	±0.0184597	[-0.012, 0.0249194]

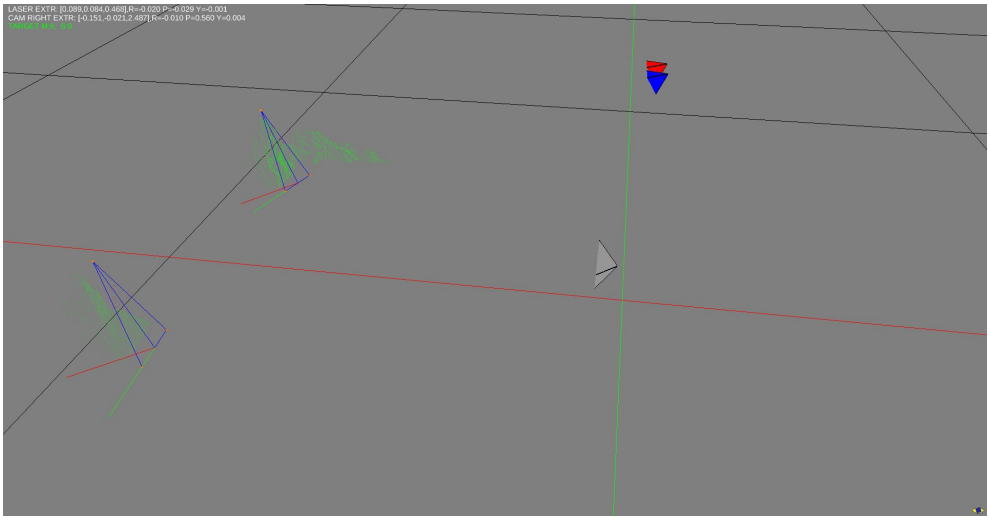
Table 4.3: Test results on real data

(a)

(b)

Figure 4.10: VIAC prototypes used for test on real data and results. (a) Vehicle frontal perception system. (b) Detail on two stereo camera systems with both short and wide baselines.

(a) Camera view



(b) Sensors deployment

Figure 4.11: Data obtained by the perception system after sensor registration. Result observed from camera in (a) and sensors deployment in (b).

Chapter 5

Calibration Procedure for multi-sensorized vehicles

The prototype vehicles developed for the VIAC expedition were designed to drive autonomously along unknown route for an intercontinental journey from Italy to China. The sensor sets chosen for the expedition were based on cameras and lasers, tightly fixed and integrated on the vehicles, as shown in figure 5.1. The calibration procedure has to be applied to seven cameras and four laserscanners with different characteristics. The entire setup task should take several working hours, since it has to deal with six different sensor sets:

- *Panoramic Vision System*: it provides a 180 degrees view of the area surrounding the frontal part of the vehicle by merging 3 images coming from 3 synchronized cameras;
- *Front and Rear Stereo cameras*: the frontal stereo system is used for medium to short range sensing with about 80 cm baseline;
- *Road boundaries laserscanner*: a mono beam laserscanner, pitched down in order to hit ground in front of the vehicle, is used to detect ditches and curbs;
- *Lateral laserscanners*: two single beam laserscanners are mounted right on the

(a)

(b) (c)

Figure 5.1: Front view of the vehicles *(a)* involved in VIAC and Laserscanner *(b)* and vision based *(c)* sensor placements. From left to right: Lateral laserscanners, Off-road laserscanner, Central laserscanner, Front and Rear Stereo Cameras, Panoramic Vision System.

corners of the frontal bumper. Each laserscanner has an aperture of about 270 degrees, while the perception depth is about 30 meters;

- *Central laserscanner*: this 4-planes laserscanner has a perception depth of about 80 meters and aperture of 100 degrees.

The complementary purpose between lasers and cameras in object detection applications leads to a specific extrinsic calibration procedures, as suggested in different

Figure 5.2: Outdoor vehicles measurement site on VisLab facilities.

works [18, 55, 45, 46], that must be taken in order to define the relative poses between the sensors reference systems. Therefore, the calibration procedure for multi-sensorized vehicles becomes a fundamental stage to keep consistence and reliability across raw sensor data. Moreover, many Advanced Driver Assistance Systems were used in VIAC to perceive the environment and all their information were shared between different applications. Depending on the vehicle behavior, such as leader-following rather than way-point following or stop & go, the whole system is able to detect obstacles, lane markings or presence of some preceding vehicles. Some authors [56], for a pedestrian and obstacle detectors, use camera and laserscanner sensors fusion to make hypothesis about the surrounding objects location. Even in leader-follower navigation approach, the data fusion easily aids the vehicle controller system to retrieve the leader vehicle position for the further path planning stage.

Each subsystem requires a specific working range, a specific precision and above a specific maintenance procedure: for those reasons the different calibration techniques impose a modular sensor calibration and discourage the development of a unique calibration procedure.

The work described is motivated by the request of an efficient time management during vehicle maintenance. In the literature [57], an efficient calibration procedure for a removable sensor suite composed by multiple laserscanners and cameras is separated in two main parts: intrinsic calibration and extrinsic calibration. In one hand, the proposed procedure divides intrinsic to extrinsic calibration procedures similarly.

In the other hand, more steps are considered in order to manage unremovable sensors and improve modularity. The first setup stage uses blob markers on plane holders to compute internal camera parameters. At the same time a sensor position coarse measurements is used as initial values that will be refined in the next calibration procedures. This task could be done offline and it might take several hours to be executed. The second stage concerns to the sensor maintenance thus solving, based on previous works [45, 18], the extrinsic parameters estimation problem. In this phase the absolute and relative poses for stereo rigs, laser-to-laser, and laser-to-camera systems are computed.

The outdoor vehicles measurement site on VisLab facilities, shown in figure 5.2, is used during the setup stage and it is composed by markers plugged into the road in a well know position. Moreover, the calibration area permits to perform long range marker detection and correspondences.

In the maintenance stage, the outdoor mobile measurement site requires a plane area in order to lie, on the ground facing the vehicle, a checkerboard sheet of 9×6 meters size shown in figure 5.3. By using homography, situations of non-flat terrain are recognized.

Additionally, to evaluate the marker detector precision a benchmark tool has been developed based on realistic synthetic images: several experiments have been performed to evaluate the calibration procedure.

This chapter is organized as follows. Section 5.1.3 describes the vehicle coordinate systems notation and first procedure steps. In section 5.1 the camera calibration procedure is illustrated indeed with intrinsic and extrinsic parameters evaluation methods, while section 5.2 shows the multi-laser calibration procedure. Section 5.3 faces the camera and laser cross-calibration task. In section 4.3 experimental results are illustrated, the conclusions are drawn and new themes for further works are suggested.

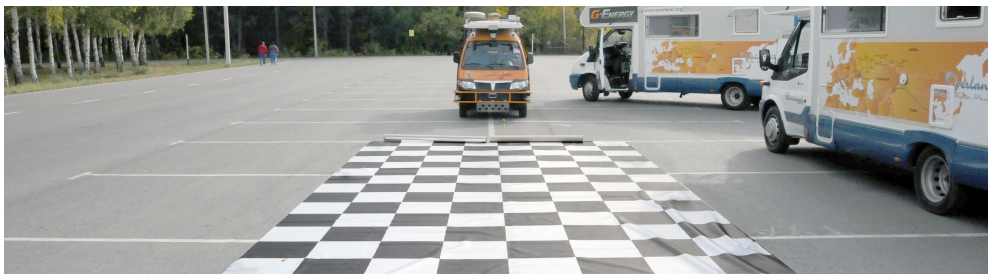


Figure 5.3: Outdoor mobile measurement site used during VIAC.

5.1 Camera Calibration

The pinhole camera model is used to represent the perspective projection that maps 3D world coordinates $\mathbf{x} = (X,Y,Z)^\top$ into 2D image coordinates $\mathbf{p} = (u,v)^\top$. The relationship between $\mathbf{x}$ and $\mathbf{p}$ is described by homogeneous equation

$$\mathbf{p} = \mathbf{K}[\mathbf{R}|\tilde{\mathbf{t}}]\mathbf{x} \tag{5.1}$$

where

$$\mathbf{K} = \begin{bmatrix} k_u & k_\gamma & u_0 \\ 0 & k_v & v_0 \\ 0 & 0 & 1 \end{bmatrix} \tag{5.2}$$

is the 3×3 intrinsic parameters matrix (detailed in 2.1.3), (k_u, k_v) are the focal length in pixel unit, (u_0, v_0) are the coordinates of the principal point, k_γ is the *skew* that in this case is supposed to be zero, $\mathbf{R}$ is the rotation matrix that transforms from world coordinates to camera coordinates and $\tilde{\mathbf{t}}$ is the position of world origin expressed in camera coordinates.

There are several approaches described in the literature facing the camera calibration problem, most of these are based on automated algorithms, applied in structured scenarios, using specific tools to determinate a set of camera parameters that can map 3D world points in 2D image points with a perspective projection.

A modular procedure has been considered, decomposing the camera calibration problem into several steps in order to save time and keep high the measurements

accuracy. The most time-consuming and accuracy-critical tasks are made on vehicles setup that include the lens distortion model estimation and the pinhole model evaluation. Without any previous setup changes that can lead to intrinsic parameter corruption, the extrinsic camera parameters can be computed more efficiently several times during the journey. Furthermore, the extrinsic calibration procedure includes the absolute pose estimation and the relative calibration of the front and rear stereo couples.

Specific calibration target, shown in figure 5.4, made of a rigid foam board with circular or checkerboard markers superimposed on it, has been used. Correspondences between the marker points and the image points has been used to compute the internal parameters.

To minimize the markers detection error, a benchmark tool has been developed to evaluate the accuracy detection of the different types of markers. This tool generates synthetic images representing a virtual calibration target, as shown in figure 5.4, modeling the sensor noise, the relative pose between the board and the observer, the pin-hole camera parameters and the lens distortion. In this way a ground truth has been generated to compare detection errors between different markers.

5.1.1 Intrinsic Parameters calibration

Two sequential steps compose the intrinsic calibration procedure.

The first step is to recover the distortion model coefficients $(k_1, k_2, k_3, \dots)$ and it is done by collecting a small amount of target images with different poses. In images, in fact, only radial distortion has been considered. The undistorted point $\mathbf{p}$, (u_u, v_u) is radially moved to the distorted one (u_d, v_d) through a function $f_d(r)$ of the distorted radius r_d, euclidean distance between the distorted point and the distortion center. Using this function the relationship between the distorted and the undistorted coordinates can be written using the Brown-Conrady model [4]: Since this function can not easily modeled, normally it is replace by a Taylor series expansion like

$$\frac{r_d}{r_u} = f_d(r_d) = 1 + k_1 r_d^2 + k_2 r_d^4 + k_3 r_d^6 + \dots \tag{5.3}$$

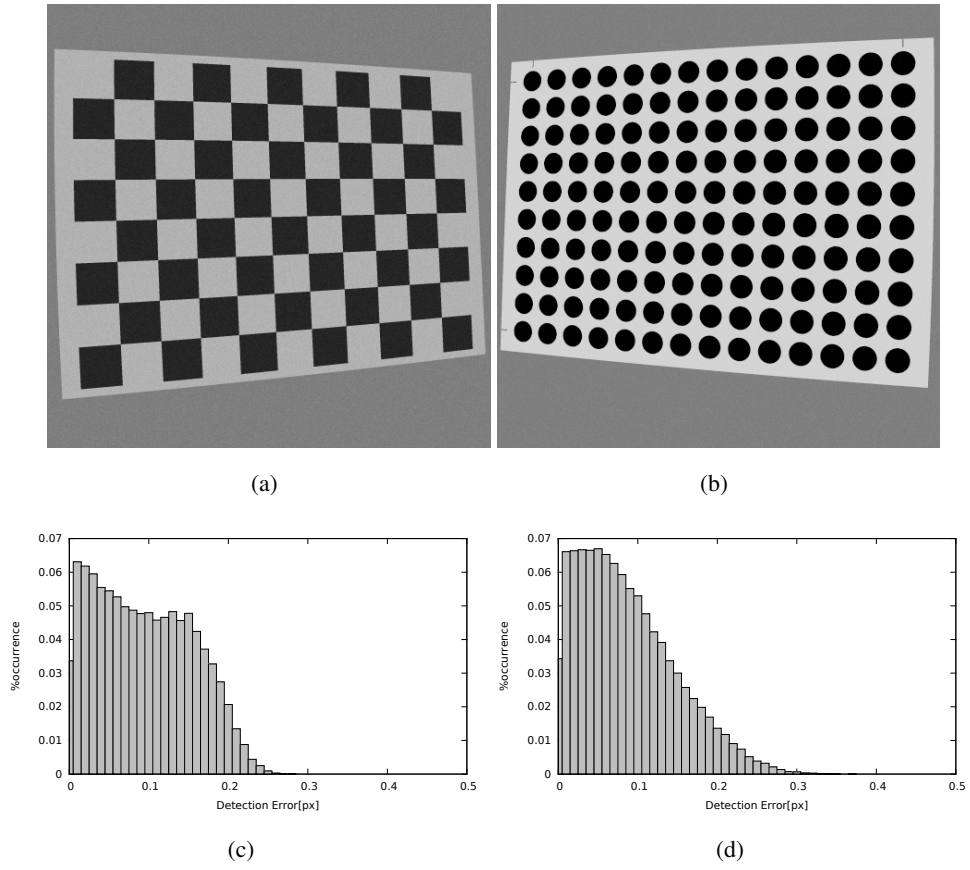


Figure 5.4: Intrinsic calibration foam boards used to estimate internal parameters such as the lens distortion model coefficients, the focal length and the principal point. The checkerboard 5.4(c) and blob 5.4(d) relative detection error, estimated by synthetic images shown more accuracy in the second case. Details are reported in 3.4.2.

where only the even terms of expansion are considered due to symmetry of radial distortion function. k parameters are estimated based on a non-linear regression between the grid points and the image points, minimizing reprojection error

$$s = \sum_{j=1}^{m} \sum_{i=1}^{n_j} \| f(\mathbf{p}_{i,j}) - h_j(\mathbf{x}_{i,j}) \|^2 \tag{5.4}$$

where h_j is the homography projection of j-plane formed by 8 unknowns to be recovered. The Levenberg-Marquardt minimizer is used to find the lens and the homography parameters. This work was developed based on concepts founded in [8]. The developed algorithm generates, for each acquired distorted image, a mesh that connects every marker together and it is used to estimate the distortion model through a non-linear minimization of projected points on the grid. Error information are provided at the end of each processed frame and studied in order to, iteratively, find the models that best fit the acquired data. Furthermore, the *homographic maximum error* evaluated on any grid points is used to index the introduced error and consequently the specific error that could be expected either for the estimated distortion model or for the marker detection error. When an arbitrary small error is reached, a look up table, using the best fit distortion model on date, can be generated and applied to the incoming images to obtain an undistorted scene.

The second step, based on the same images acquired in the previous phase, focus on the $\mathbf{K}$ estimation.

The initial solution for intrinsic parameters calibration is obtained by the Sturm-Maybank-Zhang calibration technique [16, 17]. Since this method does not minimize a geometric distance, but an algebraic one, the initial solution obtained by this method is further refined through the reprojection error:

$$s = \sum_{j=1}^{m} \sum_{i=1}^{n_j} \| \mathbf{p}_{i,j} - \mathbf{K}\mathbf{H}_j \mathbf{x}_{i,j} \|^2 \tag{5.5}$$

where $\mathbf{p}_{i,j}$ are the undistorted image points, $\mathbf{x}_{i,j}$ the points on the grid, $\mathbf{K}$ is the intrinsic parameters matrix and $\mathbf{H}_j = \Pi_{Z=0}[\mathbf{R}_j|\tilde{\mathbf{t}}_j]$ are the matrices for the extrinsic parameters with 6 additional unknown per pose to be recovered.

5.1.2 Relative Pose calibration

The dense stereo detection algorithm has been used for several tasks like lane detection and general obstacle detection; further details on the stereo vision system used on these prototypes can be found in [58]. Indeed, stereo obstacle detection methods need accurate and dense depth maps to locate obstacles. For stereo camera, the relative pose **R** e **t** have to be computed in order to produce good quality disparity and correct distance. In this case, camera A is assumed to be in the origin of the reference frame, aligned along the axis and extrinsic parameters of others cameras (B) are calibrated relative to the master camera (A).

The relationship between points in the reference frames of cameras A and B is

$$\mathbf{R}\mathbf{x}_B + \mathbf{t} = \mathbf{x}_A \tag{5.6}$$

Using again a calibration grid, the above parameters are recovered minimizing the reprojection error

$$\begin{aligned} s = \quad & \textstyle\sum_j (\sum_i \|\mathbf{p}_{A,i} - \mathbf{K}_A \mathbf{R}_{A,j}(\mathbf{x}_i - \mathbf{e}_{A,j})\|^2 + \\ & + \textstyle\sum_i \|\mathbf{p}_{B,i} - \mathbf{K}_B \mathbf{R}_{B,j}(\mathbf{x}_i - \mathbf{e}_{B,j})\|^2) \end{aligned} \tag{5.7}$$

with the constraints

$$\begin{aligned} \mathbf{R}_{A,j}\mathbf{R} &= \mathbf{R}_{B,j} \\ \mathbf{R}_{A,j}\mathbf{t} + \mathbf{e}_{A,j} &= \mathbf{e}_{B,j} \end{aligned} \tag{5.8}$$

where $\mathbf{p}_{A,i}$ are the grid points on image A and $\mathbf{p}_{B,i}$ are the same points on image B relative to marker point $\mathbf{x}_i$, $\mathbf{R}_{A,j}$, $\mathbf{e}_{A,j}$, $\mathbf{R}_{B,j}$ and $\mathbf{e}_{B,j}$ are the poses of each cameras relative to calibration grid.

5.1.3 Absolute Pose Calibration

The relative pose estimation between cameras guarantees a high degree of precision in 3D reconstruction. However, as each sensor, it is necessary to provide coordinate reference system consistent with the vehicle and in order to permit this, the master camera needs to be properly calibrated.

The absolute pose for a camera is achieved through the use of large, mobile or fixed, infrastructure. With these it is possible to associate, manually in case of non-conventional markers, points in world coordinates with the corresponding points in image coordinates. The cost function minimized in this step is

$$s = \sum_i \|\mathbf{p}_i - \mathbf{KR}(\mathbf{x}_i - \mathbf{t})\|^2 \tag{5.9}$$

where $\mathbf{R}$ and $\mathbf{t}$ are the camera pin-hole position and orientation expressed in vehicle coordinates.

5.2 Laser Calibration

The vehicles are equipped with two kinds of laserscanners as presented in the introduction. The central sensor is a 4-planes laserscanner and the others are single plane laserscanners. The central laser is considered as the master sensor with known pose and all other uncalibrated lasers are considered as slave sensors. The basic idea is to measure slave pose by fusing master and slave data in a common frame. Laserscanner coordinate notation and its reference coordinate system will be described in details. For each sensor the origin of the reference frame is on the laser emitter and the XYZ axes compose a right-hand coordinate system (ISO 8855). A world point $\mathbf{x}_l$ is measured by laser l with a range distance r and an angle α. On the l laserscanner coordinate system, $\mathbf{x}_l$ is formulated as follow:

$$\mathbf{x}_l = (r\sin\alpha, r\cos\alpha, 0)^\top \tag{5.10}$$

considering, for simplicity, a single plane laserscanner.

The laser point $\mathbf{x}_l$ could be finally transformed in world point $\mathbf{x}_w$ as following:

$$\mathbf{x}_w = \mathbf{R}_{lw}\mathbf{x}_l + \mathbf{t}_{lw} \tag{5.11}$$

where $\mathbf{R}_{lw}$ is the sensor to vehicle rotation matrix, and $\mathbf{t}_{lw}$ is the position of the laser sensor in vehicle reference frame.

5.2.1 Assumption

A laserscanner has 6 extrinsic parameters: 3 concerning the relative pose respect to the origin of the coordinate reference system (X, Y, and Z) and 3 concerning the orientation (roll φ, pitch ϑ, and yaw ψ). Laserscanners have less world perception then cameras, hence their physical pose needs more cautions during set up time then camera pose. Even if a misconfiguration of the pose can be software recovered despite the loss of information, due to the few scanning layers, it is important to obtain predetermined plane based on user requirements and avoid sensor signals touching ground in unwanted locations.

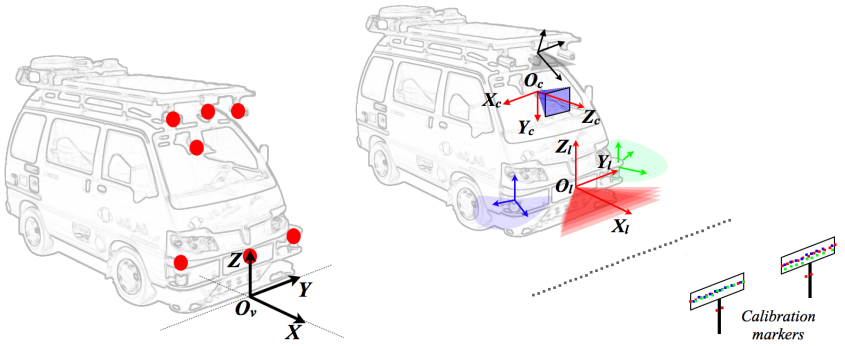
For this reason Z, ϑ, and φ are treated as fixed and well known in order to best fit user requirements. A tripod with a metal bar on the upper part has been used as calibration marker and it is positioned at different location and orientation in front of the vehicle. Figure 5.5(b) gives a conceptual configuration with the perceived area colored with different laserscanner layers. Distance information from the calibration marker and its orientation are used to manually extract laserscanner parameters.

5.2.2 Master Calibration

The frontal laserscanner is defined as master sensor and it is manually oriented toward the horizon plane. A correct orientation should be reached in two phases with at least two tripod markers. In a first step, Z, ϑ, and φ parameters are configured using tripod markers in a well established, manual, procedure. Then, a second step allows automatic measurement of the X, Y, and ψ parameters using a minimization algorithm. Finally, when master pose parameters are measured, the automatic procedure for the slave calibration become available.

5.2.3 Slave Calibration

The automatic slave calibration procedure is applied to every laserscanner, as shown in figure 5.5(b). Without loss of generality, this procedure can be applied to any couple of lasers, formed by a master and a slave. The algorithm requires that all calibration markers must be perceived from both sensors, thus laser FOVs must be



(a) (b)

Figure 5.5: The sensor reference systems and the vehicle global coordinate system (*a*). Camera and laser cross-calibration environment (*b*).

overlapped at least for two calibration markers. Following equation describes the relationship between master and slave points, $\mathbf{x}$ and $\mathbf{x}'$ respectively:

$$\mathbf{x} = \mathbf{R}\mathbf{x}' + \mathbf{t} \tag{5.12}$$

where $\mathbf{R}$ is the rotation matrix and $\mathbf{t}$ is the translation vector. Points acquired by a laserscanner are converted in a 2D coordinate system, projecting their points on the "XY" plane. Firstly, master detects the calibration marker position by clustering points, that compose marker, with a convex hull method. Secondly, a minimization process fits X, Y, and ψ parameters of the slave laser scanner. The non-linear optimization is performed minimizing the cost function

$$s = \sum_i \|\mathbf{x}_i - \mathbf{R}\mathbf{x}'_i - \mathbf{t}\|^2 \tag{5.13}$$

where points $\mathbf{x}_i, \mathbf{x}'_i$ represents the cluster center of mass. The master coordinates $\mathbf{x}_i$ are expressed in vehicle reference frame.

5.3 Camera and Laser Cross-Calibration

The camera and laser data provide complementary information, since laser distance measurements provide constant precision over distance, respect to pixel accuracy on the perspective camera projection. However, it is hard to determine the relative camera and laser pose using only their data without any assumption neither on their settings nor on positions. Related works on this problem are discussed in [45, 18].

With abuse of notation, loss function s could be represented with two components: the euclidean distance between world point $\mathbf{x}$ and its perspective projection f on the current image and, *vice versa*, the euclidean distance between selected marker point $\mathbf{p}$ in image space with its inverse perspective projection in world coordinate system, both normalized by a σ factor that represent the respective sensor error:

$$s = \frac{\sum_i \|\mathbf{p_i} - f(\mathbf{x}_i)\|^2}{\sigma_p^2} + \frac{\sum_i \|\mathbf{x}_i - f^{-1}(\mathbf{p}_i)\|^2}{\sigma_x^2} \tag{5.14}$$

where σ_p is the standard deviation of the detection error and σ_x represents the laser resolution and noise. Since f^{-1} is non-invertible projection function, the Z coordinate, provided by laser, is assigned to $\mathbf{p}_i$ and used to back-project the point.

5.4 Contribution

This approach has been developed to figure out the extrinsic calibration problem for multi-sensorized vehicles through automatic procedures. Cameras and laser scanners extrinsic inter-calibration is performed by the aim of triangular shape targets. Proposed method might be used both for monocular and stereo vision systems, where intrinsic parameters are considered as known. Furthermore, inter-calibration can be coupled with stereo auto-calibration to be able to recover correct distance measure. The proposed method reports good results in simulated environments, however on real scenarios it is strongly affected by laser scanner statistical error. Therefore a multi pose calibration is preferred to come through laser uncertainty limitation as discussed in experimental results section. Further works are focused on using this procedure in urban environment by enforcing the shape detection phase, and using

such method with known objects beside the road, even with road signs whenever a laser scanner can percepts it.

Chapter 6

Contributions and conclusions

This thesis presents the topic of the extrinsic calibration of active and passive sensors which are used on modern intelligent vehicles to get a rich perception of the surrounding environment. An in-depth analysis of the inter-calibration procedure was conduced with respect to the data fusion accuracy.

6.1 Contributions

References on extrinsic intercalibration methods -presented in 4.1- have been studied and tested while a new method, detailed in 4.2, has been developed. Principal calibration approaches have been implemented using the GOLD framework and tested on different prototypes (e.g., BRAiVE and VIAC vehicles) with different sensor suits. Two fundamental contributions compose this work:

- For the VisLab Intercontinental Autonomous Challenge (VIAC briefly described in section B) a calibration procedure has been developed and tested. One of the purposes of this experiment was to optimize and validate the proposed method in order to maintain high precision pose estimation to be able to convert geometrical information between different coordinate systems.
- The calibration procedure defined during VIAC project had been raised two

critical points: the first one concern the support of specific calibration sites with strict constraints on wide flat surfaces as a calibration target, and the second point concern the sensor maintenance ration, i.e., due to the extrinsic parameters variation over time demonstrated by sporadic test on-site (signed in red in figure B.1). Thus a intercalibration procedure based on known triangular targets has been developed and tested after VIAC project to improve these two critical points.

Another important contribution of this thesis is related to the marker detection accuracy (discussed in session 3.4.2). Since camera calibration task is composed by several process that can have bad effect on parameter estimation due to effect such as lens distortion and perspective distortion, the uncertainty in locating the control points on the image frame is investigated. Comparison with different calibration pattern and with different marker detectors have been accomplished in order to derive conclusion on their features and performances.

6.2 Conclusions

Thanks to the VIAC project it has been possible to observe that intercalibration between heterogeneous sensors (i.e., laser and camera) involved during autonomous driving tests for application such as: vision tracking and cruise controls change over time. Furthermore, this discrepancy across sensors degrade reliability and safeness of such applications which, indeed, need to perceive surrounding objects position in the same way from each sensor.

In this thesis, the vision systems have been combined with laser scanners perception to compute the extrinsic camera calibration automatically, thus to be able to update camera parameters whenever a known pattern is perceived from both sensors. As a final results, a procedure based on triangular shapes is presented.

The following conclusion can be deduced relating to the achieved experiences:

- The lasers and cameras procedure for multi-sensorized vehicles defined in chapter 5 can be applied to vehicles equipped with multiple heterogeneous

sensors which are used on board with several ADAS based on sensor fusion (i.e., visual tracking, pedestrian detection, and advanced cruise controls). Its modular architecture has been designed to improve efficiency by saving time and dividing the maintenance process from the whole calibration process in order to apply autonomous intercalibraiton procedures.

- Automated extrinsic laser and camera intercalibration based on triangular targets can take advantage of triangular shapes present in scenarios such as triangular road signs, thus maintenance phase can be repeated more frequently and without human operator interaction whenever a triangular road sign can be perceived from the multi-sensorized vehicle.

The work described in this thesis analyzes the state of the art on intercalibration procedures from indoor mobile robot architectures to recent Intelligent vehicles which adopt laser and camera sensors to perceive the world. Initially, a calibration procedure designed for VIAC prototypes has been described with final goal of improve the common process of sensors calibration as stated in chapter 5. This work gives a fundamental contribute in the intelligent vehicle field since it can provide a specific tool to process sensor calibration autonomously, furthermore this thesis suggests a solution in chapter 4 to maintain 3D perception accuracy across heterogeneous sensors exploiting known environmental object.

Appendix A

Models and Curve Fitting

This appendix face some recurring approximation problems on the computer vision, specifically in sensors calibration tasks. Every perception system that provide information of the world unavoidably introduce some kind of noise in such information, this error must be consider and managed during next signal processing phases. For that reason our applications have to deal with lots of noisily data generating overdetermined systems of equation which have to be solved minimizing the error effects. This notes have the objective to give an overview on the most common fitting methods that can be used in overdetermined systems.

A.1 Least Squares

The linear *Least Squares* (LS) fitting algorithm is a frequently approach to solve overdetermined systems, giving an approximation of the solution by minimizing a sum of squares of an error function. Considering the system of equations $Ax = \bar{b}$. Where, $A \in \mathbb{R}^{m \times n}$ with $(m > n)$, $\bar{b} \in \mathbb{R}^{m \times 1}$ is a perturbed coefficient vector, and x is the solution of the system. The least squares problem in A.1 gives an approximated solution.

$$\min_x \|Ax - b\|_2 \tag{A.1}$$

Given a set of samples that usually in computer vision consists of m points in an euclidean 2D or 3D space, in the form $\bar{x}_i$ for $i = 1 \ldots m$ and a parametric model function of the form $\bar{y} = f(\bar{x}, \bar{p})$ where $\bar{x}$ is the independent component, $\bar{p}$ are he n parameters of the model and $\bar{y}$ is a dependent (measured) component.

The LS target is to find the minimum value of the sum over all m samples of an error function $E(\bar{p})$ that measure the squared residual

$$\min E(\bar{p}) = \min \sum_{i=1}^{m} r^2(\bar{p}) \tag{A.2}$$

where,

$$r_i = \bar{y}_i - f(\bar{x}_i, \bar{p}) \tag{A.3}$$

A.1.1 Linear fitting of 2D points

This is the most simple approach to the least squares fit, commonly used in academic studies to introduce regression techniques. The point sets represents measurement of our model (in this case the model is a straight line) and the y-component is functionally dependent on the x-component in the form $(x, f(x))$. Given a set of samples $\{(x_i, y_i)\}_{i=1}^{m}$, with $(m > 2)$[1], determine parameters A and B for the line $y = Ax + B$ that minimize the sum of the squared errors between the y_i and the estimated value $Ax_i + B$. In this case the residual error is measure in the y-direction.

Define $E(A,B) = \sum_{i=1}^{m} [(Ax_i + B) - y_i]^2$. It is possible observe that this function represent a paraboloid whose absolute minimum correspond to the vertex that occurs when the gradient is set to zero than satisfy $\nabla E = 0$. This condition A.4 leads to a system A.5 of two linear equations in the unknown A and B that can be solved easily.

$$(0,0)^\top = \nabla E(A,B) = \left(\frac{\partial E}{\partial A}, \frac{\partial E}{\partial B} \right)^\top = 2 \sum_{i=1}^{m} [(Ax_i + B) - y_i] (x_i, 1)^\top \tag{A.4}$$

[1] The least squares is an approach that approximate a solution of an overdetermined system, in that sense a set of two points uniquely determine a line without approximation.

with few steps it is possible to obtain the linear system

$$\begin{bmatrix} \sum_{i=1}^{m} x_i^2 & \sum_{i=1}^{m} x_i \\ \sum_{i=1}^{m} x_i & \sum_{i=1}^{m} 1 \end{bmatrix} \begin{bmatrix} A \\ B \end{bmatrix} = \begin{bmatrix} \sum_{i=1}^{m} x_i y_i \\ \sum_{i=1}^{m} y_i \end{bmatrix} \tag{A.5}$$

The solution vector $[A \quad B]^\top$ provides the least squares solution of model function $y = Ax + B$ minimizing the *vertical error* between samples and line. It is also possible to fit a line considering the implicit equation $ax + by + c = 0$ with the condition $a^2 + b^2 = 1$ that can model vertical line that the previous model don't and consider different error function E that minimize the orthogonal distance between the samples and the modeled line. This type of regression is known as *Total Least Squares* (TLS) or Orthogonal least squares line fit. The orthogonal distance d between a sample $(x_i, y_i), i = 1 \ldots m$ and a line can be computed following equation A.6

$$d = \frac{|ax_i + by_i + c|}{\sqrt{a^2 + b^2}} \tag{A.6}$$

The error function to be minimized become

$$E = \frac{1}{2m} \sum_{i=1}^{m} \frac{(ax_i + by_i + c)^2}{a^2 + b^2} \tag{A.7}$$

Note that this function has an absolute minimum and maximum, therefore there are two solution for the line model that are both valid solutions of the system. Considering $\nabla E(a,b,c) = 0$ from the partial derivative $\frac{\partial E}{\partial c}$ it is possible to observe that the line will pass through the centroid $(\bar{x}, \bar{y})$ of our set of samples by the equation $c = -a\bar{x}, -b\bar{y}$ where $\bar{x}$ and $\bar{y}$ represent the x_i and y_i average values respectively. The error function A.7 can be simplified using the previous condition, obtaining the follow

$$E = \frac{a^2 \left(\bar{x^2} - \bar{x}^2\right) + 2ab \left((\bar{xy}) - \bar{x}\bar{y}\right) + b^2 \left(\bar{y^2} - \bar{y}^2\right)}{a^2 + b^2} \tag{A.8}$$

A possible approach can be simplify by subtracting the average component to the data set obtaining $\bar{x} = 0$ and $\bar{y} = 0$ and reintegrate those components only after

the line estimation. The same result can be achieve using the SVD decomposition method that minimize the geometrical error either the algebraic error and it will be discussed in A.3.

A.1.2 Planar fitting of 3D points

The assumptions made in A.1.1 are still valid in this case where the z-component of the data is functionally dependent on the x-component and y-component in the form $(x,y,f(x,y))$. Given a set of samples $\{(x_i,y_i,z_i)\}_{i=1}^{m}$, with $(m>2)$, determine A, B, and C so that the plane $z=Ax+By+C$ best fit all samples, means that the summarized distance between all z_i and $Ax+By+C$ have to be minimized.

Therefore, define an error function $E(A,B,C)=\sum_{i=1}^{m}\left[(Ax_i+By_i+C)-z_i\right]^2$. This function is nonnegative and represent a hyper-paraboloid whose minimum occurs on its vertex where the gradient is equal to zero, or $\nabla E=0$. Such constraint lead to a system in three unknown A, B, and C that can be solved.

$$(0,0,0)^\top=\nabla E(A,B,C)=\left(\frac{\partial E}{\partial A},\frac{\partial E}{\partial B},\frac{\partial E}{\partial C}\right)^\top=2\sum_{i=1}^{m}\left[(Ax_i+By_i+C)-z_i\right](x_i,y_i,1)^\top \tag{A.9}$$

obtaining the followed system

$$\begin{bmatrix}\sum_{i=1}^{m}x_i^2 & \sum_{i=1}^{m}x_iy_i & \sum_{i=1}^{m}x_i\\ \sum_{i=1}^{m}x_iy_i & \sum_{i=1}^{m}y_i^2 & \sum_{i=1}^{m}y_i\\ \sum_{i=1}^{m}x_i & \sum_{i=1}^{m}y_i & \sum_{i=1}^{m}1\end{bmatrix}\begin{bmatrix}A\\B\\C\end{bmatrix}=\begin{bmatrix}\sum_{i=1}^{m}x_iz_i\\ \sum_{i=1}^{m}y_iz_i\\ \sum_{i=1}^{m}z_i\end{bmatrix} \tag{A.10}$$

A.1.3 Fitting a paraboloid to 3D points

The set of points is in the form $\{(x_i,y_i,z_i)\}_{i=1}^{m}$ and it is assumed that these points represent a generic paraboloid in the form

$$z=f(x,y)=p_1x^2+p_2xy+p_3y^2+p_4x+p_5y+p_6=P\cdot Q(x,y) \tag{A.11}$$

where $P = (p_1, p_2, p_3, p_4, p_5, p_6)$ is the parameter vector that need to be computed by minimizing the error distances between samples and model and $Q(x,y) = (x^2, xy, y^2, x, y, 1)$. The error function is defined as shown in A.12

$$E(P) = \sum_{i=1}^{m} (P \cdot Q(x_i, y_i) - z_i)^2 \tag{A.12}$$

The minimum of this function occur when the gradient equal the null vector, hence with condition A.13

$$\nabla E = 2 \sum_{i=1}^{m} (P \cdot Q(x_i, y_i) - z_i)\, Q(x_i, y_i) = 0 \tag{A.13}$$

With some algebra the A.13 is converted in A.14 where $Q_i = Q(x_i, y_i)$

$$\left(\sum_{i=1}^{m} Q_i Q_i^\top \right) P = \sum_{i=1}^{m} z_i Q_i \tag{A.14}$$

The product between vector $Q_i \in \mathbb{R}^{1\times 6}$ and $Q_i^\top$ results a 6×6 symmetric matrix. Define the matrix $A = \sum_{i=1}^{m} Q_i Q_i^\top$ and $b = \sum_{i=1}^{m} z_i Q_i$ as 6 coefficients vector. The least squares solution of the paraboloid is the vector P of equation $AP = b$. A and b components are a sum over samples of specific variable product. Consider a function $s(p) = \sum_{i=1}^{m} p_i$ to simplify the notation A.14 and obtain the final system in six unknown.

A.2 Nonlinear Least Square

As mentioned in previous session least square problem can be stated as finding $\mathbf{x}^\star$, a local minimizer for

$$F(\mathbf{x}) = \frac{1}{2} \sum_{i=1}^{n} (f_i(\mathbf{x}))^2 \tag{A.15}$$

where $f_i : \mathbb{R}^n \to \mathbb{R}$, $i = 1 \ldots m$ are given functions and $m \geq n$. Then, it is necessary to define a model M tuned by a parameter vector $\mathbf{x}$ and residual functions f_i that is computed as difference between measured data y_i and prediction $M(\mathbf{x}, t_i)$ where t_i is

independent variable (or input) and y_i dependent variable (or output). Nonlinear least square sample can be modeled as follow

- model $M(\mathbf{x},t) = x_1 e^{x2t} + x_3 e^{x4t}$
- parameters $\mathbf{x} = [x_1, x_2, x_3, x_4]^\top$
- residual $f_i(\mathbf{x}) = y_i - M(\mathbf{x},t_i) = y_i - x_1 e^{x2t_i} + x_3 e^{x4t_i}$

The global minimizer problem is very hard to sole and a local minimizer is considered to solve a simpler problem.

Local Minimizer
Given a cost function $F : \mathbb{R}^n \to \mathbb{R}$, finding $\mathbf{x}^\star$ so that,
$F(\mathbf{x}^\star) \leq F(\mathbf{x})$ for $\|\mathbf{x} - \mathbf{x}^\star\| \leq \delta$

It is assume that F is differentiable and smooth enough that is representable by its Taylor expansion.

$$F(\mathbf{x}+h) = F(\mathbf{x}) + h^\top F'(\mathbf{x}) + \frac{1}{2} h^\top F''(\mathbf{x}) h + O(\|h\|^3) \tag{A.16}$$

where F' is the gradient in A.17, and F'' is the Hessian matrix

$$F'(\mathbf{x}) = \begin{bmatrix} \frac{\partial F}{\partial x_1}(\mathbf{x}) \\ \frac{\partial F}{\partial x_2}(\mathbf{x}) \\ \vdots \\ \frac{\partial F}{\partial x_n}(\mathbf{x}) \end{bmatrix} \tag{A.17}$$

$$F''(\mathbf{x}) = \left[\frac{\partial^2 F}{\partial x_i \partial y_j}(\mathbf{x}) \right] \tag{A.18}$$

A.2.1 Steepest Descent method

As the name describe this method takes advantage of using gradient information to move toward a local minimum and it will stop when a stationary point is achieved. Steepest descent reach has very good performance in the initial stage of iterations. It is assume that h_d is a descent direction if $F(\mathbf{x} + \alpha\mathbf{h})$ is a decreasing function of α at $\alpha = 0$ and this leads to the following definition.

Algorithm 4 Algorithm Descent method

Input: k max num. iterations; **x** model parameters;

Output: **x**

```
1:  begin
2:  k:=0; x:= x_0; found:=false                          ▷ Starting point
3:  while (not found) and (k < k_max) do
4:      h_d := search_direction(x)                       ▷ From x and downhill
5:      if no such h exists then
6:          found:=true                                  ▷ x stationary
7:      else
8:          α := step_length(x, h_d)                     ▷ From x toward h_d
9:          x := x + α h_d
10:         k := k + 1
11:     end if
12: end while
13: end
```

h is a descent direction for F at $\mathbf{x}$ if $h^\top F'(\mathbf{x}) < 0$

Algorithm 4 shows that when a step is computed with positive α, the relative gain in function value satisfies

$$\lim_{\alpha \to 0} \frac{F(\mathbf{x}) - F(\mathbf{x} + \alpha h)}{\alpha \|h\|} = -\frac{1}{\|h\|} h^\top F'(\mathbf{x}) = -\|F'(\mathbf{x})\| \cos\theta \tag{A.19}$$

where θ is the angle between vector h and $F'(\mathbf{x})$. That shown the maximum gain at $\theta = \pi$ (i.e, in steepest descent method $h_{sd} = -F'(\mathbf{x})$).

A.2.2 Newton's method

This method achieve good performances in the final stage of the iterative process then when solution vector $\mathbf{x}$ is closer to the optimal solution $\mathbf{x}^\star$. It is derived from the condition that $\mathbf{x}^\star$ is a stationary point. According to local minimizer definition A.2 $F'(\mathbf{x}) = 0$. This equation is described by a nonlinear system of equations and it can be approximated by a Taylor expansion in

$$\begin{aligned} F'(\mathbf{x}+h) &= F'(\mathbf{x}) + F''(\mathbf{x})h + O(h^2) \\ &\approx F'(\mathbf{x}) + F''(\mathbf{x})h \qquad \text{for small } h \end{aligned} \tag{A.20}$$

the Newton's method is derived by finding h_n as solution of

$$Hh_n = -F'(\mathbf{x}) \quad \text{with} \quad H = F''(\mathbf{x}) \tag{A.21}$$

Whether H is positive defined, then it is nonsingular (implying that A.21 has unique solution), and $u^\top H u > 0$ for all nonzero u. Thus, multiplying A.21 by $h_n^\top$ results

$$0 < h_n^\top H h_n = -h_n^\top F'(\mathbf{x}) \tag{A.22}$$

A.2.3 Levenberg-Marquardt method

This can be considered as hybrid method between steepest descent method and newton method. When solution $\mathbf{x}$ is far form the correct one it operates as steepest descent method (slow but convergence is guaranteed), then when solution get closer to the $\mathbf{x}^\star$ it becomes a Newton method. Reference in [59]

A.2.4 Fitting a circle to 3D points

Given a set of 3D points $\mathbf{P}_i (i = 1, \ldots, n)$ and assumed that these points represents a plane circle modeled with its center $C = [C_x, C_y, C_z]^\top$, a known radius r and its normal vector $N(\alpha, \beta)$ where α and β are orientation angle of the 3D circle axis vector with respect the **y**-axis and **x**-axis respectively. Solution is given by a nonlinear optimization constrained to the following equations

$$\begin{aligned} {}_i\Pi_1 &= N(\alpha,\beta) \cdot \overline{C\mathbf{P}_i} \\ {}_i\Pi_2 &= N(\alpha,\beta) \times \overline{C\mathbf{P}_i} - r \end{aligned} \tag{A.23}$$

where:

- $\overline{C\mathbf{P}_i} = \mathbf{P}_i - C$
- ${}_i\Pi_1$ corresponds to the Euclidean distance between 3D point $\mathbf{P}_i$ and 3D plane defined by normal vector $N(\alpha, \beta)$ and center C
- ${}_i\Pi_2$ corresponds to the Euclidean distance between 3D point $\mathbf{P}_i$ and the 3D circle axis defined by normal vector $N(\alpha, \beta)$ and center C

Minimizing the following objective function using the Levenberg-Marquardt algorithm A.2.3 the model is obtained

$$e = \sum_{i=1}^{n} \left[{}_i\Pi_1^2 + {}_i\Pi_2^2 \right] \tag{A.24}$$

A.3 Using Singular Value Decomposition

Singular Value Decomposition (SVD) is a useful tool to find patterns in data. On these notes is shown how to use such factorization to solve linear least squares fitting problems.

Let $A \in \mathbb{R}^{m \times n}$ be matrix of rank r. Then there exist orthogonal matrices $U \in \mathbb{R}^{m \times m}$ and $v \in \mathbb{R}^{n \times n}$ such that

$$A = U \Sigma V^\top, \quad \Sigma = \begin{bmatrix} \Sigma_1 & 0 \\ 0 & 0 \end{bmatrix} \tag{A.25}$$

where $\Sigma \in \mathbb{R}^{m,n}$ is a diagonal matrix in the form $\Sigma_1 = (\sigma_1, \sigma_2, \ldots, \sigma_r)$, and $\sigma_1 \geq \sigma_2 \geq \cdots \geq \sigma_r \geq 0$.

The σ are called singular values of A and if considered

$$U = (u_1, \ldots, u_m), \quad V = (v_{1,\ldots,v_n}), \tag{A.26}$$

the u_i and v_i are respectively, the left and right singular vector associated with $\sigma_i, i = 1, \ldots, r$

A.3.1 Fitting lines to 3D points by orthogonal distance

Assume to fit a line that best fit a point cloud $p_i \in \mathbb{R}^3$ where the orthogonal distance between the modeled line and each point is minimum. First of all, consider the model as a parametric line in the form,

$$l(t) = O + tv, \quad t \in \mathbb{R} \tag{A.27}$$

where O is a point of the line and v is a unit vector that gives a direction to the line. Derived to A.6 the orthogonal distance between point p_i and the line l is obtained by the inner product

$$d_i = (p_i - O)^\top v \tag{A.28}$$

Thus, from A.7 the searching lane can be computed by solving the LS problem

$$\min \sum_{i=1}^{n} d_i^2 \tag{A.29}$$

Considering the partial derivative with respect to line parameter O and equal to zero, it is possible observe that such line pass through the centroid of the data as shown in A.7 with $O = 1/n \sum_{i=1}^{n} p_i$. Lets considering the matrix $A \in \mathbb{R}^{3\times n}$

$$A = \begin{bmatrix} p_1 - O & p_2 - O & \ldots & p_n - O \end{bmatrix} \tag{A.30}$$

the problem can be formulated as follow

$$\min_{\|n\|=1} \|A^\top v\|_2^2 \tag{A.31}$$

Applying the SVD decomposition of $A = USV^\top$ obtain,

$$\|A^\top v\|_2^2 = \|VS^\top U^\top v\|_2^2 = \|S^\top U^\top v\|_2^2 = (\sigma_1 y_1)^2 + (\sigma_2 y_2)^2 + (\sigma_3 y_3)^2 \tag{A.32}$$

where the unit vector $y = U^\top v$. Thus, $\|A^\top v\|_2^2$ is minimized for $y = (0,0,1)^\top$ or equally for $v = U(:,1)$. The minimal value of $\sum_{i=1}^n \left((p_i - O)^\top D\right)^2$ is σ_3^2.

However, the orthogonal least squares line correspond to the largest eigenvalue of the matrix A that correspond to the largest singular value. To summarize the wall procedure, the 3D line is compute by its centroid than by the direction vector that correspond to the largest singular vector.

A.3.2 Fitting planes to 3D points by orthogonal distance

The algorithm to fit a parametric plane model ,

$$n \cdot x + d = 0 \tag{A.33}$$

where n is the normal vector of the plane and d is its distance to the origin is the same of the line fit of A.3.1. Once it is recognize that the centroid of the data set belongs to the plane the orthogonal least squares can be computed on the same set shifted into the origin. Involving use of covariance matrix of the data it is possible to compute largest and smallest eignevectors that match the trend of the data. In practice these vectors are computed using SVD and the best fit plane correspond to the minimum singular value, rather then the maximum singular value that represent the best line.

Such algorithm to fit line or plane using orthogonal distance can be summarized as follow

1. Compute the centroid of the data set $d = 1/n \sum_{i=1}^n p_i$

2. Compose matrix $A = [p_1 - d \; p_2 - d \; \ldots \; p_n - d]$
3. Compute the SVD of A, $USV^\top = A$
4. **plane:** $n = U(:,3)$ representing the smallest singular vector.
   line: $v = U(:,1)$ that is the largest singular vector

A.3.3 Rigid body movement on 3D points

Consider a set of features of a rigid body in a 3D space, with the set $\{x_1, x_2, \ldots, x_n\}$ representing the position of each marker. Than, assume to measure the same body in another position of the space that is described by another set $\{y_1, y_2, \ldots, y_n\}$, that due to some measure uncertainty the roto-translation mapping is not exact so it could be fit by least squares method solving the problem

$$\min_{R,t} \sum_{i=1}^{n} \|[R|t]x_i - y_i\|^2 \tag{A.34}$$

Where solution of the problem is an orthogonal matrix R and a translation vector t that maps the points x_i to y_i. Problem A.34 is linear with respect the vector t but is nonlinear with respect the matrix R (due to the orthogonal constraints). To solve the problem lets consider matrices $A = [x_1 - \bar{x} \; x_2 - \bar{x} \; \ldots \; x_n - \bar{x}]$, and $B = [y_1 - \bar{y} \; y_2 - \bar{y} \; \ldots \; y_n - \bar{y}]$, where $\bar{x}$ and $\bar{y}$ are respectively the average value of x_i and y_i for $i = 1, \ldots, n$. The problem of determine the rotation matrix becomes

$$\min_{R} \|RA - B\|_F \tag{A.35}$$

Where the Frobenius norm of a matrix A is defined as $\|A\|_F^2 = \sum_{i,j} a_{i,j}^2$ and it is shown in literature that this Orthogonal Procrustes problem can be solved by the singular value decomposition of the matrix $USV^\top = BA^\top$. The final solution is given by

$$R = Udiag(1, 1, det(UV^\top))V^\top \tag{A.36}$$

$$t = \bar{y} - R\bar{x} \tag{A.37}$$

Further details are described in [60].

A.4 Estimation of the Homography (DLT)

Given a set of correspondences $\mathbf{M}_i \leftrightarrow \mathbf{m}_i$ where $\mathbf{M}_i$ is the model point in the world reference frame and $\mathbf{m}_i = [u_i, v_i, 1]^\top$ represent its image position with respect to the image reference frame. It is possible to compute matrix $\mathbf{H}$ such that

$$\mathbf{m}_i \simeq \mathbf{H}\mathbf{M}_i \qquad i = 1 \ldots n \tag{A.38}$$

Let considering $\mathbf{H} = [\mathbf{h}_1,\ \mathbf{h}_2,\ \mathbf{h}_3]$ and unknown vector $\mathbf{x} = [\mathbf{h}_1^\top, h\mathbf{h}_2^\top, \mathbf{h}_3^\top]^\top$ equation A.38 can be represented as follow

$$\begin{bmatrix} \mathbf{M}_i^\top & \mathbf{0}^\top & -u\mathbf{M}_i^\top \\ \mathbf{0}^\top & \mathbf{M}_i^\top & -v\mathbf{M}_i^\top \end{bmatrix} \mathbf{x} = \mathbf{0}_{2\times 1} \tag{A.39}$$

From n world points, derive n equations from A.39, which can be written in matrix equation as $\mathbf{A}\mathbf{x} = \mathbf{0}$, where $\mathbf{A}$ is a $2n \times 9$ matrix. As $\mathbf{x}$ is defined up to a scale factor, the solution is well known to be the right singular vector of $\mathbf{A}$ associated with the smallest singular value of SVD decomposition (or equivalently, the eigenvector of $\mathbf{A}^\top\mathbf{A}$ associated with the smallest eigenvalue).

Ideally equation A.38 should be applied to each pair, however this this is not possible due to some uncertainty in extracting image points $\mathbf{m}_i$. It must be assumed that $\mathbf{m}_i$ is corrupted by Gaussian noise with zero mean, in this case the above problem becomes a nonlinear least-squares one such that

$$\min_{\mathbf{H}} \sum_i \| \|\mathbf{m}_i - \hat{\mathbf{m}}_i\| \|^2 \tag{A.40}$$

where

$$\hat{\mathbf{m}}_i = \frac{1}{\mathbf{h}_3^\top \mathbf{M}_i} \begin{bmatrix} \mathbf{h}_1^\top \mathbf{M}_i \\ \mathbf{h}_2^\top \mathbf{M}_i \end{bmatrix} \tag{A.41}$$

The nonlinear optimization can be conduced with the Levemberg-Marquardt method and the initial guess can be commuted with the previous method.

The maximum likelihood estimation of $\mathbf{H}$ can be obtained by considering a covariance matrix $\Lambda_{\mathbf{m}_i} = \sigma^2\mathbf{I}$ for all i with following function

$$\min_{\mathbf{H}} \sum_i \left(\mathbf{m}_i - \hat{\mathbf{m}}_i\right)^\top \Lambda_{\mathbf{m}_i}^{-1} \left(\mathbf{m}_i - \hat{\mathbf{m}}_i\right) \tag{A.42}$$

Otherwise it is possible to consider a geometric residual in the form

$$\min_{\mathbf{H}} \sum_i d\left(\mathbf{H}\mathbf{M}_i, \mathbf{m}_i\right)^2 + d\left(\mathbf{H}^{-1}\mathbf{m}_i, \mathbf{M}_i\right) \tag{A.43}$$

where function $d(.)$ compute the euclidean distance between two points.

Appendix B

VIAC

B.1 From Milan to Shanghai

The VisLab Intercontinental Autonomous Challenge [61, 62] is a test of autonomous driving along an unknown route from Italy to China through, Slovenia, Croatia, Serbia, Hungary, Ukraine, Russia, and Kazakhstan the trip began on July 20, took over three months and the total distance covered was more than 13,000 km. Vehicles travelled from Italy to China through as shown in figure B.1.

The base vehicle is an Electric Porter Piaggio which has been transformed into an Intelligent Vehicle for the challenge, most of the sensing technologies installed on the base vehicle are directly derived from the perception suite of BRAiVE [63]; however, BRAiVE was not designed to drive autonomously in off-road environments, hence it misses all the crosscountry driving skills needed during an intercontinental trip like VIAC. Vehicles were equipped keeping sensors, actuators, and processing units accessible, in order to optimize development time, usability, and ease maintenance in remote locations (detailed in chapter 5).

Throughout the journey the expedition travelled across a plurality of scenarios completely different from each other. Crossing a large part of the Eurasian continent all sorts of situations, environments, roads, and weather conditions were met: mountains, planes, unpaved, and dusty roads. In Europe and in the first part of the Russian

Federation the convoy travelled for many kilometers on highways and drove into the heavy urban traffic of many great eastern Europe cities like Belgrade, Budapest, Kiev, Moscow, then went across Siberia, the flat steppes of Kazakhstan, up the Tien Shan Mountains and finally all the way across China towards its destination, Shanghai. VIAC had his official conclusive event on October the 28th 2010 at the Belgium-EU pavilion inside Shanghai's 2010 World Expo.

The data collected refer to the effective 61 days of autonomous driving on an overall 90 days journey: 191 runs for a total of 214 hours in autonomous mode were completed. Usually the runs ended when no battery power was left, but sometimes logistic needs mandated a stop, such as when crossing a state border. The maximum distance traveled in autonomous mode per run was 96.7 km, with an average of 77.0 km. . No autonomous run were performed in some tracks due to technical or logistic problems. The sum of the tracks gives 8244 km in autonomous mode covered at an average speed of 38.4 km/h and a maximum speed of 70.9 km/h. Maximum distance covered in a single day was of 273 km and the maximum amount of time spent in a day driving in autonomous mode was of 6 h, 26 min.

This project was carried out in the frame of the ERC Advanced Investigator Grant (OFAV) received by Prof. Alberto Broggi. Several technical had been taken part to the VIAC expedition : Piaggio, Topcon, Thales, IBM, Enfinity, and Overland Network, including all other partners that worked for the success of this expedition.

Day	Date	Place	Distance [km]	
1	19/07	Milano		
2	20/07	Milano-Parma	153	
3	21/07	Parma-Roma	470	
4	22/07	Roma-Venezia	530	
5	23/07	Venezia-Trieste	57	
6	24/07	Trieste-Lubiana	98	SLOVENIA
7	25/07	Lubiana-Zagabria	153	CROAZIA
8	26/07	Zagabria-Nova Gradiska	200	SERBIA
9	27/07	Nova Gradiska-Belgrado	186	
10	28/07	Belgrado		
11	29/07	Belgrado-Novi Sad	96	
12	30/07	Novi Sad-Subotica	110	
13	31/07	Subotica-Budapest	192	UNGHERIA
14	01/08	Budapest-Zàhony	313	
15	02/08	Zàhoni-L'viv	287	UCRAINA
16	03/08	L'viv-Novohrad Violyns'kyi	312	
17	04/08	Novohrad Violyns'kyi-Kiev	224	
18	05/08	Kiev		
19	06/08	Kiev-Chorol	237	
20	07/08	Chorol-Kharkov	246	
21	08/08	Kharkov		
22	09/08	Kharkov-Slovjansk	186	
23	10/08	Slovjansk-Border	207	
24	11/08	Border-Rostov	100	RUSSIA
25	12/08	Rostov		
26	13/08	Rostov-Millerovo	214	
27	14/08	Millerovo-Voronezh	360	
28	15/08	Voronezh		
29	16/08	Voronezh-Novomoskovk	327	
30	17/08	Novomoskovk-Mosca	242	
31	18/08	Mosca		
32	19/08	Mosca		
33	20/08	Mosca		
34	21/08	Mosca-Vladimir	185	
35	22/08	Vladimir-Niznij Novgorod	236	
36	23/08	Niznij Novgorod		

37	24/08	Niznij Novgorod-Saransk	289	
38	25/08	Saransk-Saratov	361	
39	26/08	Saratov		
40	27/08	Saratov-Sirzan’	311	
41	28/08	Sirzan’-Samara	166	
42	29/08	Samara		
43	30/08	Samara-Dimitrovgrad	145	
44	31/08	Dimitrovgrad-Kazan	234	
45	01/09	Kazan		
46	02/09	Kazan-Naberesnje Celni	241	
47	03/09	Naberesnje Celni-Ufa	287	
48	04/09	Ufa-Yuryuzan’	182	
49	05/09	YuryuzanâĂŹ-Celiabinzk	250	
50	06/09	Celiabinzk		
51	07/09	Celiabinzk		
52	08/09	Celiabinzk-Snezhinsk	110	
53	09/09	Snezhinsk-Jekaterinburg	92	
54	10/09	Jekaterinburg		
55	11/09	Jekaterinburg-Kamysiov	149	
56	12/09	Kamysiov-Tjumen	189	
57	13/09	Tjumen		
58	14/09	Tjumen-Vagay	146	
59	15/09	Vagay-Ishim	163	
60	16/09	Ishim-Omsk	319	
61	17/09	Omsk		
62	18/09	Omsk-Tatarsk	182	
63	19/09	Tatarsk-Kujbysev	255	
64	20/09	Kujbysev-Novosibirsk	330	
65	21/09	Novosibirsk		
66	22/09	Novosibirsk-Bolotnoe	136	
67	23/09	Bolotnoe-Kemerovo	130	
68	24/09	Kemerovo		
69	25/09	Kemerovo-Novokuznetsk	207	
70	26/09	Novokuznetsk-Barnaul	238	
71	27/09	Barnaul-Rubcovsk	315	
72	28/09	Rubcovsk-Semipalatinsk	150	KAZAKISTAN
73	29/09	Semipalatinsk-Georgiyevka	162	
74	30/09	Georgiyevka-Ayagoz	192	

75	01/10	Ayagoz-Usharal	250	
76	02/10	Usharal-Taldykorgan	270	
77	03/10	Taldykorgan-Almaty	261	
78	04/10	Almaty		
79	05/10	Almaty-Sary Ozek	190	
80	06/10	Sary Ozek-Zharkent	180	
81	07/10	Zharkent-Khorgas	50	
82	08/10	Khorgas: border		CINA
83	09/10	Khorgas-Yining	90	
84	10/10	Yining		
85	11/10	Yining-Qingshuihe-Jinghe	290	
86	12/10	Tuotuoxiang-Hutubi	340	
87	13/10	Hutubi-Shanshan	350	
88	14/10	Shanshan-Hami	340	
89	15/10	Hami-Hongliuyuan	300	
90	16/10	Hongliuyuan-Jiayuguan	320	
91	17/10	Jayuguan-Shandan	300	
92	18/10	Shandan-Yongdeng	335	
93	19/10	Yongdeng-Huining	290	
94	20/10	Huining-Binxian	335	
95	21/10	Binxian-Xian	150	
96	22/10	Xian-Sanmenxia	320	
97	23/10	Sanmenxia-Luohe	320	
98	24/10	Luohe-Xiangcheng	360	
99	25/10	Xiangcheng-Mingguang	300	
100	26/10	Mingguang-Changzhou	300	
101	27/10	Changzhou-Kunshan	260	
102	28/10	Kunshan-Shanghai	90	
103	29/10	Shanghai		
104	30/10	Shanghai		
105	31/10	Shanghai		

Table B.1: Trip schedule involved in VIAC project. Vehicles maintenance is highlighted in gray color.

Bibliography

[1] Azim Eskandarian. *Handbook of Intelligent Vehicles*. Springer London, 2012.

[2] Emanuele Trucco and Alessandro Verri. *Introductory Techniques for 3-D Computer Vision*. Prentice Hall PTR, Upper Saddle River, NJ, USA, 1998.

[3] Chris Mcglone, Edward Mikhail, and Jim Bethel. *Manual of photogrammetry*. American Society for Photogrammetry and Remote Sensing, Bethesda, MD, 5th edition, 1980.

[4] Duane C Brown. Decentering distortion of lenses. *Photogrammetric Engineering*, 32(3), 1966.

[5] S. Shah and J.K. Aggarwal. A simple calibration procedure for fish-eye (high distortion) lens camera. In *Robotics and Automation, 1994. Proceedings., 1994 IEEE International Conference on*, pages 3422 –3427 vol.4, may 1994.

[6] K. Vijayan Asari. Design of an efficient vlsi architecture for non-linear spatial warping of wide-angle camera images. *Journal of Systems Architecture*, 50(12):743 – 755, 2004.

[7] Anup Basu and Sergio Licardie. Alternative models for fish-eye lenses. *Pattern Recognition Letters*, 16(4):433 – 441, 1995.

[8] Frédéric Devernay and Olivier Faugeras. Straight lines have to be straight. *Machine Vision and Applications*, 2001.

[9] *Investigating laser scanner accuracy*, volume XXXIV-5/C15. The International Archives of the Photogrammetry, Remote Sensing and Spatial Information Sciences, 2003.

[10] RichardI. Hartley. Euclidean reconstruction from uncalibrated views. In JosephL. Mundy, Andrew Zisserman, and David Forsyth, editors, *Applications of Invariance in Computer Vision*, volume 825 of *Lecture Notes in Computer Science*, pages 235–256. Springer Berlin Heidelberg, 1994.

[11] E.L. Hall, J.B.K. Tio, C.A. McPherson, and F.A. Sadjadi. Measuring curved surfaces for robot vision. *Computer*, 15(12):42 – 54, dec 1982.

[12] O. Faugeras and G. Toscani. The Calibration Problem for Stereo. In *Computer Vision and Pattern Recognition*, 1986.

[13] Salvi Joaquim. *An approach to coded structured light to obtain three dimensional information*. PhD thesis, Universitat de Girona. Departament d'Electrònica, Informàtica i Automàtica, 1997.

[14] R. Tsai. A versatile camera calibration technique for high-accuracy 3d machine vision metrology using off-the-shelf tv cameras and lenses. *Robotics and Automation, IEEE Journal of*, 3(4):323 –344, august 1987.

[15] J. Weng, P. Cohen, and M. Herniou. Camera calibration with distortion models and accuracy evaluation. *Pattern Analysis and Machine Intelligence, IEEE Transactions on*, 14(10):965 –980, oct 1992.

[16] P.F. Sturm and S.J. Maybank. On plane-based camera calibration: A general algorithm, singularities, applications. In *Computer Vision and Pattern Recognition, 1999. IEEE Computer Society Conference on.*, volume 1, 1999.

[17] Zhengyou Zhang. Flexible camera calibration by viewing a plane from unknown orientations. In *Computer Vision, 1999. The Proceedings of the Seventh IEEE International Conference on*, volume 1, 1999.

[18] Qilong Zhang and R. Pless. Extrinsic calibration of a camera and laser range finder (improves camera calibration). In *Intelligent Robots and Systems, 2004. (IROS 2004). Proceedings. 2004 IEEE/RSJ International Conference on*, volume 3, sept.-2 oct. 2004.

[19] Q.-T. Luong and O. D. Faugeras. Self-calibration of a moving camera from pointcorrespondences and fundamental matrices. *Int. J. Comput. Vision*, 22(3):261–289, March 1997.

[20] L. Lucchese and S.K. Mitra. Using saddle points for subpixel feature detection in camera calibration targets. In *Circuits and Systems, 2002. APCCAS '02. 2002 Asia-Pacific Conference on*, volume 2, pages 191 – 195 vol.2, 2002.

[21] G. Bradski. The OpenCV Library. *Dr. Dobb's Journal of Software Tools*, 2000.

[22] Andreas Geiger, Frank Moosmann, Oemer Car, and Bernhard Schuster. A toolbox for automatic calibration of range and camera sensors using a single shot. In *International Conference on Robotics and Automation (ICRA)*, St. Paul, USA, May 2012.

[23] Massimo Bertozzi, Luca Bombini, Alberto Broggi, Pietro Cerri, Paolo Grisleri, and Paolo Zani. GOLD: A framework for developing intelligent-vehicle vision applications. *IEEE Intelligent Systems*, 23(1):69–71, January–February 2008.

[24] Alberto Broggi, Stefano Debattisti, Matteo Panciroli, Paolo Grisleri, Elena Cardarelli, Michele Buzzoni, and Pietro Versari. High performance multi-track recording system for automotive applications. *Intl. Journal of Automotive Technology*, 13(1), January 2011.

[25] BertholdK.P. Horn. Relative orientation. *International Journal of Computer Vision*, 4:59–78, 1990.

[26] Berthold K. P. Horn. Relative orientation revisited. *J. Opt. Soc. Am. A*, 8(10):1630–1638, Oct 1991.

[27] K. S. Arun, T. S. Huang, and S. D. Blostein. Least-squares fitting of two 3-d point sets. *Pattern Analysis and Machine Intelligence, IEEE Transactions on*, PAMI-9(5):698 –700, sept. 1987.

[28] Berthold K. P. Horn. Closed-form solution of absolute orientation using unit quaternions. *J. Opt. Soc. Am. A*, 4(4):629–642, Apr 1987.

[29] Berthold K. P. Horn, Hugh M. Hilden, and Shahriar Negahdaripour. Closed-form solution of absolute orientation using orthonormal matrices. *J. Opt. Soc. Am. A*, 5(7):1127–1135, Jul 1988.

[30] Michael W. Walker, Lejun Shao, and Richard A. Volz. Estimating 3-d location parameters using dual number quaternions. *CVGIP: Image Understanding*, 54(3):358 – 367, 1991.

[31] D.G. Lowe. Fitting parameterized three-dimensional models to images. *Pattern Analysis and Machine Intelligence, IEEE Transactions on*, 13(5):441 –450, may 1991.

[32] Daniel F. DeMenthon and Larry S. Davis. Model-based object pose in 25 lines of code. *International Journal of Computer Vision*, 15:123–141, 1995.

[33] G. Schweighofer and A. Pinz. Robust pose estimation from a planar target. *Pattern Analysis and Machine Intelligence, IEEE Transactions on*, 28(12):2024 –2030, dec. 2006.

[34] C.-P. Lu, G.D. Hager, and E. Mjolsness. Fast and globally convergent pose estimation from video images. *Pattern Analysis and Machine Intelligence, IEEE Transactions on*, 22(6):610 –622, jun 2000.

[35] Martin A. Fischler and Robert C. Bolles. Random sample consensus: a paradigm for model fitting with applications to image analysis and automated cartography. *Commun. ACM*, 24(6):381–395, June 1981.

[36] Long Quan and Zhongdan Lan. Linear n-point camera pose determination. *Pattern Analysis and Machine Intelligence, IEEE Transactions on*, 21(8):774 –780, aug 1999.

[37] P.D. Fiore. Efficient linear solution of exterior orientation. *Pattern Analysis and Machine Intelligence, IEEE Transactions on*, 23(2):140 –148, feb 2001.

[38] Abdel Y. I. Aziz and H. M. Karara. Direct linear transformation into object space coordinates in close-range photogrammetry. In *Proc. of the Symposium on Close-Range Photogrammetry*, pages 1–18, Urbana, Illinois, 1971.

[39] Vincent Lepetit, Francesc Moreno-Noguer, and Pascal Fua. Epnp: An accurate o(n) solution to the pnp problem. *International Journal of Computer Vision*, 81:155–166, 2009.

[40] You Li, Y. Ruichek, and C. Cappelle. 3d triangulation based extrinsic calibration between a stereo vision system and a lidar. In *Intelligent Transportation Systems (ITSC), 2011 14th International IEEE Conference on*, pages 797 –802, oct 2011.

[41] Ganhua Li, Yunhui Liu, Li Dong, Xuanping Cai, and Dongxiang Zhou. An algorithm for extrinsic parameters calibration of a camera and a laser range finder using line features. In *Intelligent Robots and Systems, 2007. IROS 2007. IEEE/RSJ International Conference on*, pages 3854 –3859, 29 2007-nov. 2 2007.

[42] Qilong Zhang and R. Pless. Extrinsic calibration of a camera and laser range finder (improves camera calibration). In *Intelligent Robots and Systems, 2004. (IROS 2004). Proceedings. 2004 IEEE/RSJ International Conference on*, pages 2301 – 2306 vol.3, sept.-2 oct. 2004.

[43] S. Wasielewski and O. Strauss. Calibration of a multi-sensor system laser rangefinder/camera. In *Intelligent Vehicles '95 Symposium., Proceedings of the*, pages 472 –477, sep 1995.

[44] K. Kwak, D.F. Huber, H. Badino, and T. Kanade. Extrinsic calibration of a single line scanning lidar and a camera. In *Intelligent Robots and Systems (IROS), 2011 IEEE/RSJ International Conference on*, pages 3283 –3289, sept. 2011.

[45] S.A. Rodriguez F, V. Fremont, and P. Bonnifait. Extrinsic calibration between a multi-layer lidar and a camera. In *Multisensor Fusion and Integration for Intelligent Systems, 2008. MFI 2008. IEEE International Conference on*, pages 214 –219, aug 2008.

[46] Lili Huang and M. Barth. A novel multi-planar lidar and computer vision calibration procedure using 2d patterns for automated navigation. In *Intelligent Vehicles Symposium, 2009 IEEE*, june 2009.

[47] Jun-Sik Kim, P. Gurdjos, and In-So Kweon. Geometric and algebraic constraints of projected concentric circles and their applications to camera calibration. *Pattern Analysis and Machine Intelligence, IEEE Transactions on*, 27(4):637 –642, april 2005.

[48] Kenichi Kanatani and Naoya Ohta. Automatic detection of circular objects by ellipse growing. *Int. J. Image Graphics*, 36:20001, 2004.

[49] Walter Gander, GeneH. Golub, and Rolf Strebel. Least-squares fitting of circles and ellipses. *BIT Numerical Mathematics*, 34:558–578, 1994.

[50] Charles V. Stewart. Robust parameter estimation in computer vision. *SIAM Reviews*, 41:513–537, 1999.

[51] Zehang Sun, G. Bebis, and R. Miller. On-road vehicle detection: a review. *Pattern Analysis and Machine Intelligence, IEEE Transactions on*, 28(5):694 –711, may 2006.

[52] Lili Huang and M. Barth. Tightly-coupled lidar and computer vision integration for vehicle detection. In *Intelligent Vehicles Symposium, 2009 IEEE*, pages 604 –609, june 2009.

[53] Rafael Grompone von Gioi, Jeremie Jakubowicz, Jean-Michel Morel, and Gregory Randall. On straight line segment detection. *Journal of Mathematical Imaging and Vision*, 32:313–347, 2008.

[54] L. Mazzei, P. Medici, and M. Panciroli. A lasers and cameras calibration procedure for viac multi-sensorized vehicles. In *Intelligent Vehicles Symposium (IV), 2012 IEEE*, pages 548 –553, june 2012.

[55] Kiho Kwak, Daniel F. Huber, Hernan Badino, and Takeo Kanade. Extrinsic calibration of a single line scanning lidar and a camera. In *Intelligent Robots and Systems (IROS), 2011 IEEE/RSJ International Conference on*, sept. 2011.

[56] P. Cerri, L. Gatti, L. Mazzei, F. Pigoni, and Ho Gi Jung. Day and night pedestrian detection using cascade adaboost system. In *Intelligent Transportation Systems (ITSC), 2010 13th International IEEE Conference on*, sept. 2010.

[57] Huijing Zhao, Yuzhong Chen, and R. Shibasaki. An efficient extrinsic calibration of a multiple laser scanners and cameras' sensor system on a mobile platform. In *Intelligent Vehicles Symposium, 2007 IEEE*, june 2007.

[58] Alberto Broggi, Michele Buzzoni, Mirko Felisa, and Paolo Zani. Stereo obstacle detection in challenging environments: The viac experience. In *Intelligent Robots and Systems (IROS), 2011 IEEE/RSJ International Conference on*, sept. 2011.

[59] D. Marquardt. An algorithm for least-squares estimation of nonlinear parameters. *Journal of the Society for Industrial and Applied Mathematics*, 11(2):431–441, 1963.

[60] A. Lorusso, D. W. Eggert, and R. B. Fisher. A comparison of four algorithms for estimating 3-d rigid transformations. In *Proceedings of the 1995 British conference on Machine vision (Vol. 1)*, BMVC '95, pages 237–246, Surrey, UK, UK, 1995. BMVA Press.

[61] Massimo Bertozzi, Alberto Broggi, Elena Cardarelli, Rean Isabella Fedriga, Luca Mazzei, and Pier Paolo Porta. VIAC Expedition Toward Autonomous Mobility. *Robotics and Automation Magazine*, 18(3):120–124, September 2011. ISSN: 1070-9932.

[62] Massimo Bertozzi, Luca Bombini, Alberto Broggi, Michele Buzzoni, Elena Cardarelli, Stefano Cattani, Pietro Cerri, Alessandro Coati, Stefano Debattisti, Andrea Falzoni, Rean Isabella Fedriga, Mirko Felisa, Luca Gatti, Alessandro Giacomazzo, Paolo Grisleri, Maria Chiara Laghi, Luca Mazzei, Paolo Medici, Matteo Panciroli, Pier Paolo Porta, Paolo Zani, and Pietro Versari. VIAC: an Out of Ordinary Experiment. In *Procs. IEEE Intelligent Vehicles Symposium 2011*, pages 175–180, Baden Baden, Germany, June 2011.

[63] Luca Bombini, Stefano Cattani, Pietro Cerri, Rean Isabella Fedriga, Mirko Felisa, and Pier Paolo Porta. Test bed for Unified Perception & Decision Architecture. In *Procs. 13th Int. Forum on Advanced Microsystems for Automotive Applications*, pages 287–298, Berlin, Germany, May 2009.