Universidade de Aveiro
Departamento de Eletrónica, Telecomunicações e Informática
2019

Miguel
Lopes Rêgo

POSICIONAMENTO POR LUZ VISÍVEL ATRAVÉS DE UMA CÂMARA

VISIBLE LIGHT POSITIONING USING A CAMERA

Universidade de Aveiro
2019

Departamento de Eletrónica,
Telecomunicações e Informática

Miguel
Lopes Rêgo

POSICIONAMENTO POR LUZ VISÍVEL ATRAVÉS DE UMA CÂMARA

VISIBLE LIGHT POSITIONING USING A CAMERA

Dissertação apresentada à Universidade de Aveiro para cumprimento dos requisitos necessários à obtenção do grau de Mestre em Engenharia Eletrónica e Telecomunicações, realizada sob a orientação científica do Doutor Pedro Nicolau Faria da Fonseca, Professor auxiliar do Departamento de Eletrónica, Telecomunicações e Informática da Universidade de Aveiro, e do Doutor Luís Filipe Mesquita Nero Moreira Alves, Professor auxiliar do Departamento de Eletrónica, Telecomunicações e Informática da Universidade de Aveiro.

o júri / the jury

presidente / president

Professor Doutor António José Ribeiro Neves
Professor Auxiliar, Universidade de Aveiro

vogais / examiners committee

Professor Doutor Luís Manuel de Sousa Pessoa
Professor Auxiliar Convidado, Instituto de Engenharia de Sistemas e Computadores do Porto

Professor Doutor Pedro Nicolau Faria da Fonseca
Professor Auxiliar, Universidade de Aveiro

agradecimentos /
acknowledgements

Agradeço aos meus orientadores, Professor Pedro Fonseca e Professor Luís Nero Alves, por todo o apoio, conselhos e ensinamentos transmitidos ao longo da elaboração desta dissertação. Ao Instituto de Telecomunicações e à Universidade de Aveiro, pelas condições disponibilizadas e por todo o apoio logístico providenciado.

Aos meus pais e à minha irmã por me terem acompanhado nesta viagem, por toda a ajuda que me deram e por sempre serem compreensivos.

Por fim, um agradecimento a todos os colegas e amigos que me apoiaram neste percurso, de uma forma especial ao Luís Rodrigues e ao Pedro Rodrigues pelas produtivas discussões, sugestões e, por vezes, paciência demonstrada.

Palavras Chave

posicionamento por luz visível, sistemas de posicionamento em interiores, comunicação com câmara, câmara, sensor de imagem, comunicação por luz visível.

Resumo

Esta dissertação tem como objetivo o desenvolvimento de um sistema de posicionamento em espaços interiores através de uma câmara. O sistema proposto utiliza a câmara como o único sensor, obtendo a sua posição numa sala com luminárias LED devidamente moduladas. O trabalho começa com uma investigação inicial das tecnologias atualmente existentes em sistemas de posicionamento. Para além disso, sistemas de posicionamento baseados em luz são explorados e a utilização da câmara neste contexto é explicada e discutida. De seguida, o princípio the funcionamento de uma câmara é apresentado e a possibilidade de utilizar um sensor deste tipo para localização em espaços interiores é desenvolvida. Tendo em conta que é necessário nesta situação, um método de comunicação das luminárias para a câmara é ainda discutido. Seguidamente, o modelo geométrico da câmara é explicado, por forma a clarificar os princípios matemáticos necessários para os cálculos de posicionamento. Na sequência, um script MATLAB é desenvolvido, com o objetivo de produzir fotografias simuladas. De seguida, outra simulação foi desenvolvida, com o objetivo de estudar o impacto de erros na estimativa final da posição num sistema deste tipo. A estrutura do programa é apresentada e os resultados são discutidos. Por fim, o sistema completo é implementado. Após uma explicação deste, os erros finais de posicionamento experimentais são apresentados e discutidos.

Keywords

visible light positioning, indoor positioning systems, optical camera communication, camera, image sensor, visible light communication.

Abstract

This dissertation focuses on the development of an indoor positioning system using a camera. The proposed system uses the camera as the only sensor, calculating its position in a room with properly modulated LED light fixtures. An initial investigation on the current technological advancements on positioning systems is performed. Moreover, light-based positioning systems are explored and the usage of a camera in such a context is explained and discussed. After that, the working principles of a camera are presented and the possibility of using this type of sensor to perform indoor location is developed. Since the application requires it, a method of transmitting data from the light fixtures to the camera is also discussed. Going further, the camera geometrical model is explained, in order to clarify the mathematical principles required to perform the positioning calculation. After that, a MATLAB script is developed, in order to produce simulated photographs. After this, another simulation was developed, with the objective of studying the impact of errors in the final position estimation in such a system. The program structure is presented and the results are discussed. Finally, the complete system is implemented. After an explanation, the final experimental positioning errors are presented and discussed.

Contents

List of Figures

List of Tables

Acronyms

IPS	indoor positioning system
CMOS	complementary metal–oxide–semiconductor
VLP	Visible Light Positioning
GPS	Global Positioning System
LOS	line-of-sight
LED	light-emitting diode
RF	radio frequency
VLC	Visible Light Communication
RSS	Received Signal Strength
TDOA	Time Difference of Arrival
AOA	Angle of Arrival
ID	identifier
FOV	field of view
CCD	charge-coupled device
MIMO	multiple-input and multiple-output
OCC	Optical Camera Communication
UPAMSM	under-sampled pulse amplitude modulation with subcarrier modulation
UPSOOK	under-sampled phase shift ON-OFF keying
OOK	ON-OFF keying
PWM	pulse width modulation
2D	two-dimensional
3D	three-dimensional
PnP	Perspective-n-Point
IMU	Inertial Measurement Unit
ROI	region of interest

CHAPTER 1

Introduction

This chapter presents a contextualization of the developed work. It consists of an introduction to the subject, a consideration of the reasons for exploring it, a listing of the work objectives and a description of the dissertation structure.

1.1 Problem statement and context

For a long time and for a variety of different reasons, science has been finding and exploring localization methods, being either for locating a ship on the sea, finding someone's or something's position on the surface of the earth, finding someone's location inside a building or finding an object's location in space. For the scope of this dissertation, we will divide the positioning problem into two main categories — outdoor and indoor.

For the outdoor situation, the problem is extensively explored and, despite still having margin for improvement, technologies such as Global Positioning System (GPS) reasonably solve the problem, specially for consumer-oriented solutions. For indoor scenarios, however, the situation is quite the opposite. Multiple technologies have emerged throughout the years but most systems are based on radio frequency (RF) and require line-of-sight (LOS), which renders them not suitable for indoor positioning, as is the case with GPS [1]. Some technologies can, indeed, be used for this purpose, such as systems that use, for example, the signal from wireless routers to estimate the receiver position, but are often neither precise nor reliable enough. Dedicated RF-based solutions also exist, often using specialized hardware for the positioning beacons, but are usually too expensive for consumer-oriented implementations. The goal of this dissertation is to explore a type of indoor positioning system (IPS) which solves these problems, through the implementation of a positioning system using artificial light sources as the beacons and a camera as the receiver.

This work is closely related to other projects developed and being developed at Instituto de Telecomunicações, in Aveiro. As such, it was developed in the Integrated Circuits group at this institution, with its full support. This group and, by extension, this project, was also

supported by the European project VisIoN, funded by the European Union's Horizon 2020 research and innovation programme under the Marie Skłodowska-Curie grant agreement.

Finally, the final experimental results obtained in this dissertation were gathered in collaboration with Pedro Rodrigues, also a student at Universidade de Aveiro and working in the Integrated Circuits group. His master's dissertation, developed in parallel with this one, focused on the development of an experimental setup for testing Visible Light Positioning (VLP) systems, ideal for the type of system explored in this work.

1.2 Motivation

As stated previously, the general problem of location is not new and, more importantly, is still an open research subject. Despite the problem of outdoor positioning being reasonably solved with systems like GPS, most systems are not suitable for indoor positioning. Some technologies can, indeed, be used for this purpose, but not a single one proves to be both cheap to implement and precise enough.

To solve this problem, in recent years, a new type of IPS is being explored, consisting in using a camera as the main sensor. By definition, a camera is a device that captures, into an image, a two-dimensional (2D) projection of the three-dimensional (3D) world in front of it. Using this concept, if we have a set of known reference beacons visible in a photograph, we can use the position in the image where the beacons are projected with their world position to calculate, under certain circumstances, the camera's position and orientation. Such a design has the advantage of being easily implemented, from the user side, in a commercial smartphone without any additional hardware, since almost every one is equipped with one or more cameras. This translates in no additional costs to the user, which is a major advantage for this type of system and crucial in allowing a future widespread use.

In order to also reduce infrastructure costs, the idea is to use the existing light fixtures on a building to serve both as illumination sources and beacons for the positioning. In the last years, light-emitting diode (LED) lighting systems have been gradually replacing the conventional incandescent or fluorescent indoor lights [1]. Given the fact that the LEDs can be modulated by, for example, being switched on and off at high frequencies — particularly high enough so that the human eye can not perceive it — this opens up a lot of possibilities for using the LED light fixtures for, at the same time, illuminating a room and transmitting information.

One of the sensors that can be used for receiving such information is, again, a camera. Since a camera can be seen as a device that successively captures incident light information, it can be used to receive data modulated through this medium. This can be done in a variety of ways, which will be explored in this dissertation. Considering this possibility, we now have a way of distinguishing the multiple light fixtures. After this, we can successfully calculate the camera's position and orientation, by relating each fixture's projection coordinates in the image with their world coordinates, as described above. Again, this allows the full implementation of an IPS using just a camera as the receiver sensor.

Given all of the stated above, the deployment of a camera-based IPS, based on visible light, does not involve major infrastructure changes or additions. Provided that the target location already uses LED light fixtures for general illumination, the only modification required for implementing the transmitters is the placement of a specific driver on each LED. As for the receiver side, on a typical smartphone, no additional hardware should be necessary, rendering such a system quite appealing and worth exploring.

1.3 Work objectives

As stated previously, the final goal of this work is to implement a complete camera-based IPS. In addition to this, the objective is for this system to be able to work completely in free space, meaning that the camera can simply be hand-held in any position, without the need for it to be placed, for example, horizontally or at a constant height.

Besides the development of the aforementioned system, theoretical work should be developed in order to understand and plan such a system. This way, besides building the system itself, this work also focuses on understanding the working principles of the main sensor used — a camera. Finally, a study of the nature of errors and noise in the system in particular should be conducted, with the objective of, not only better characterizing the system in question, but also providing a reference study for future camera-based IPSs.

1.4 Methodologies

Besides the investigation, most of the practical implementation of the work contained in this dissertation is software-based. As such, for the programming, MATLAB [2] was used as the main tool for the simulations and the rest of the code was written in C++. Regarding the latter, the open-source library OpenCV [3] was used for the image processing and related algorithms. MATLAB was also used to render most of the plots included in this dissertation.

For the hardware side, an actual smartphone camera was used to acquire the photos used for the experimental testing. As such, the Xiaomi Mi A2 frontal camera was used as the main image sensor.

Finally, for the final experimental tests, as stated previously, the VLP setup developed by Pedro Rodrigues was used. This setup is presented and described in further detail later on, but essentially provides an array of light fixtures placed along the ceiling of a room, each with a configurable Visible Light Communication (VLC) driver, which are essential, since we need to individually modulate the multiple light sources.

1.5 Dissertation structure

This dissertation is divided into six chapters. The second chapter, Visible Light Positioning Systems, introduces the required concepts to understand and work on, as the title suggests, a VLP system. An introduction to position estimation is presented, followed by an introduction to the concept of using light as the main medium for this task. Finally, the possibility of using

a camera as the main sensor in this context is presented and an explanation of the complete process of estimating the position with such a system is given.

Chapter three presents the camera model, as well as the mathematical principles behind it and the expressions involved. These principles are used to develop a simulation with the purpose of producing a simulated photograph of the light fixtures, given the camera intrinsic parameters and coordinates and the fixtures location. This simulation, not yielding any relevant or new results on its own, serves as the starting point of the work, both for debugging expressions used later on and serving as the basis for the simulation developed in the fourth chapter.

Chapter four focuses on the system error. It starts by presenting the possible sources of error for the system in question, followed by an explanation of the algorithm used to estimate the final position of the sensor — in this case, a camera. After this introduction, the developed simulation is explained in detail. The goal is to study the effect of having errors in the multiple stages of the system and analyse their impact on the final position estimation. Finally, the results are presented, followed by an analysis and reflection on the impact of the multiple error sources.

Going further, chapter five presents in detail the complete positioning system. After giving an overview of the complete system, the image processing algorithms are detailed, followed by an explanation of the multiple stages involved in the decoding of each fixtures identifier (ID), used to extract the fixtures coordinates in the world. After that, the final process of finding the camera position is further explained. Finally, the experimental setup is described in detail and the experimental results are presented.

Ending the dissertation, chapter six draws conclusions from the previously presented results. Besides an analysis of the final experimental results, the errors predicted in chapter four and the ones obtained in chapter five are compared. The chapter ends with final remarks on the system error sources and possible future work.

CHAPTER 2

Visible Light Positioning Systems

In this chapter, a brief introduction to positioning systems is presented, followed by an explanation of how to use visible light sources to implement a complete positioning system, how a camera can be used as the only receiver sensor and how, from a single photograph, we can determine the sensor's position on a given referential.

2.1 Positioning systems

The purpose of any positioning system is to provide the location coordinates of a given object in a given environment. As such, and for the purpose of this dissertation, we can separate them in two different categories: indoor and outdoor. When considering outdoor positioning systems, there are already reasonably reliable solutions available. The highlight goes to the GPS technology which, through the aid of a network of satellites, can provide fairly reliable positioning, with an error in the order of meters.

On the other hand, when considering the indoor situation, the solution to the problem is not so trivial. The main limiting factor is the fact that most systems rely either on satellite-assisted positioning or, in some way, radio frequency, which has difficulty passing through obstacles such as the walls on a building. Furthermore, RF-based systems often use the Received Signal Strength (RSS) as a measure of the distance between the transmitter and the receiver, which, again, gives erroneous estimates when an obstacle is placed between the two. Other IPSs have been proposed, based, for example, on the Earth's magnetic field or ultrasound signals. However, both these solutions involve additional hardware and considerable costs which is not desirable, particularly for consumer-oriented systems. For comparison and perspective, an overview of IPS technologies is presented in Table 2.1, including their accuracy and main limitations.

Starting with the technologies with a low user and installation cost, we have audible sound, Wi-Fi and Bluetooth. Audible sound makes use of sound waves to estimate the distance between the beacons and the object to be located. Wi-Fi and Bluetooth base the positioning on the estimation of the distance to the access points through the intensity of

Technology	Accuracy	IC	UC	Weaknesses
Infrared	57cm–2.3m	H	L	Sunlight interference
Ultrasonic	1cm–2m	H	H	Cost and interference
Audible sound	Meters	L	L	Low precision
Wi-Fi	1.5m	L	L	Vulnerable to access point changes
Bluetooth	30cm–meters	L	L	Intrusive; needs signal mapping
ZigBee	25cm	L	H	Low precision and high user cost
RFID	1–5m	H	L	Very low precision
UWB	15cm	H	H	High cost

Table 2.1: Comparison of main indoor positioning technologies. UC: end user cost; IC: installation and maintenance cost; H: high; L: low [4].

the received signal. Both these methods are highly susceptible to interference from obstacles placed between the object and the beacons, besides not providing particularly high accuracy for an indoor scenario. Similarly to Wi-Fi, RFID and ZigBee also use the received signal strength to estimate the distance to the object, but employ dedicated beacons, which increases the cost. Infrared technology uses infrared light waves and, besides also requiring dedicated transmitters, needs line-of-sight clearance between the transmitters and the object, which is a major disadvantage.

Finally, the more precise are UWB and Ultrasonic. The first uses electromagnetic wave forms consisting of a sequence of very short pulses using a large bandwidth. The travel time of the waves is used to estimate the position. Ultrasonic uses principles similar to audible sound but with ultrasounds, giving the technology more flexibility, given that sound can be emitted without the inconvenience of being heard. Both are, besides precise, expensive to implement, both in infrastructure and in user cost.

As we can observe, there is not a single design technology that can be seen as the best solution, since each has its own limitations. Usually, we have to choose between accuracy, installation cost and user cost, always sacrificing one of them. This is the main reason for the interest in VLP systems in the last years. Compared with the other solutions, they can provide high accuracy at a low cost and complexity, by using the already existing LED lighting infrastructure as the basis for the positioning.

2.2 Light-based positioning systems

With the current technological advancements, LED technology as become more and more promising and established and, as such, most indoor illumination systems are being replaced by LED-based ones. The use of LEDs brings many advantages over traditional lighting, being the most relevant:

- High energy efficiency: When compared with traditional light sources, LEDs can save up to 85% energy when compared with incandescent lamps [1].
- Long life expectancy: LEDs can last up to 10 years without having to be replaced [1].
- Cost effective: LEDs have become quite cheap, in addition to their high energy efficiency and long life [1].

- Environment friendly: no mercury or other toxic elements are generally present in LEDs [1].

For communication purposes, LED lighting also brings a variety of new possibilities over traditional lighting systems, being the most relevant:

- Data modulation: LEDs can be switched on and off at high frequencies, allowing for data modulation. Additionally, their intensity can also be modulated, allowing for even further communication techniques.
- Harmless: Visible light is harmless to humans and is *visible*, naturally allowing us to deviate our eyes from it, if necessary.
- Unrestricted usage: the visible light part of the spectrum is not regulated and does not cause electromagnetic interference, allowing its usage in sensitive environments.
- Security: since light cannot penetrate thick walls, this type of systems are inherently more secure that RF-based ones. Besides, this allows for completely independent systems across different rooms.

All these reasons render LED-based lighting systems ideal for adaptation into hybrid systems for illumination and communication. Given all these, light-based positioning systems are ideal for application in certain scenarios, such as [5]:

- Industry and warehouses: useful for robot navigation, object and people location, product detection and inventory management.
- Indoor public spaces: can be used to help, for example, in finding available parking spots, navigation systems for finding specific shops or points of interest and in implementing emergency systems to help find and rescue persons. Such systems can also help the staff to identify and prevent overcrowded places and to control the visitors flow.
- Museums: allows the implementation of visitors guiding systems, both by guiding the user through the multiple exhibitions and artefacts and by displaying additional information regarding the objects the user is looking at.
- Shopping centers: can be used for targeted offers and advertising. Furthermore, the user can benefit from guidance in specific shop or product finding or in finding information about his surroundings.
- Office and school spaces: some applications include the tracking of a person inside a building, which can be used to detect unauthorized access to certain areas. Given the precision, these systems can also aid in finding vacant work stations in a library or laboratory.
- Hospitals: Since LED-based systems do not generate electromagnetic interference, they can be freely used in controlled environments and near sensitive equipment. In hospitals, this can be used to implement patient navigation systems or fall and collision avoidance guides for handicapped or elderly people.
- Transport stations: in large public places such as airports, IPSs can be used to precisely guide users to specific locations, such as specific terminals, train platforms, nearest restrooms and dining areas.

Light-based positioning systems can be classified into two large categories, depending on the receiver sensor: photodiode-based and camera-based. Photodiode-based systems

can be classified according to the type of signal received, which includes Received Signal Strength (RSS), Time Difference of Arrival (TDOA) and Angle of Arrival (AOA). Camera-based systems can be classified as using just a camera or as having auxiliary equipment, such as additional sensors to improve the robustness or accuracy of the positioning. These classifications are summarized in Fig. 2.1.

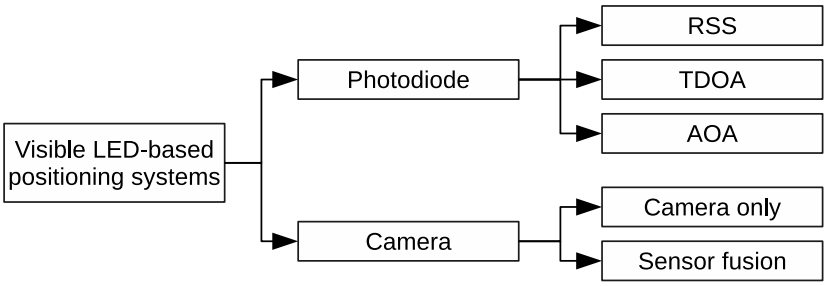


Figure 2.1: Classification of visible light-based IPS designs [1].

For this work, the focus will be on camera-based systems, particularly using just a camera as the sensor. It is relevant, however, to understand the principles behind the photodiode-based ones, particularly because a camera can be seen as closely related to them. First of all, it is important to highlight that there are three main techniques behind position estimation, independently from the technology behind them: trilateration, multilateration and triangulation. A diagram representing the basic principles behind each of them is present in Fig. 2.2.

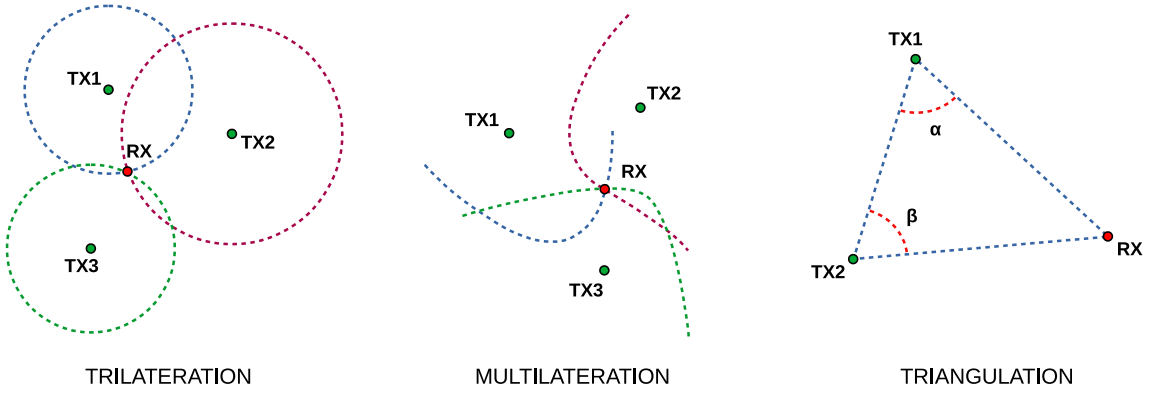


Figure 2.2: Principle of operation of position estimation techniques.

Starting with the RSS, the technique behind it is trilateration, which is the same one used on the GPS system. The signal strength perceived at the receiver is used to estimate its distance to each transmitter. In a two dimensional space, this results in a set of circles which should intersect at a single point — the calculated receiver position. In a three dimensional space, as happens with GPS, it results in a set of spheres which, in a similar fashion, should also intersect at a single point. As can be observed in Fig. 2.2, at least three transmitters are necessary for 2D positioning and four for 3D.

Going over to TDOA, instead of estimating the distances from the receiver to the trans-

mitters directly, these are inferred from the time difference between the reception of a single packet of information from each transmitter, sent at the same time. This technique is called multilateration and is similar to trilateration. The main difference is that, instead of having a circle or sphere of possible locations for each transmitter, we have a hyperbola (2D) or a hyperboloid (3D). Similarly to trilateration, this process also requires at least thee and four transmitters, respectively, for 2D and 3D positioning.

For AOA-based systems, the technique used is triangulation. Instead of estimating, directly or indirectly, the distance between the transmitters and the receiver, the angle formed between the multiple transmitter-receiver pairs is measured. As the name suggests, this technique corresponds to, both in 2D and 3D, having a set of triangles with a known side length and two known angles. From there, the remaining side lengths can be calculated to obtain the receiver position. A simplified version of this process can be observed in Fig. 2.2. It is important to note that, since we are dealing with both position and angles, we have to consider both. This leads to, inherently, having to estimate both the position and orientation of the receiver. In most systems, however, some of these parameters are either fixed (such as a receiver being always in a horizontal position) or can be found with auxiliary methods (by using, for example, a gyroscope to find the receiver orientation). Either way, the number of necessary transmitters depends on the number of unknown parameters for the system in question.

Finally, in camera-based systems, the receiver position is found by relating the location of the transmitters with their projection on the resulting image. This can be seen as close to AOA-based systems. Assuming that we have infinitely small light sources acting as transmitters, they are projected onto a single pixel. The pixel in which they appear directly corresponds to an angle between the light ray coming from that source and the camera normal vector. This process is the basis for the work developed in this dissertation and its particularities will be further explained.

It is important to keep in mind that all these techniques assume that we have a way of knowing the location of each transmitter and of distinguishing them at the receiver. On photodiode-based systems it is quite straightforward to transmit information without interfering with the positioning process given the modulation possibilities of LEDs and the large bandwidth of photodiodes. With a camera, however, the solution is not immediate. In the next section, an explanation of how to accomplish this is presented.

2.3 Optical Camera Communication

As presented before, most light-based positioning systems use photodiodes as the receiver. However, a camera can also be employed in such a task and brings the immediate advantage of being a piece of hardware already existent in almost every smartphone. In order to understand how a camera can be used for this purpose, it is important to review its working principles.

2.3.1 Camera working principles

A camera is a device that captures light wave information into still images, resulting in photographs of a given scene in the world. Since the the focus of this work will be on digital

cameras, only these will be considered. Typically, a digital camera is formed by an image sensor, where the image is produced, and a lens placed in front of it, whose purpose is to focus and direct the light rays into the sensor. A diagram of the components of a basic digital camera is depicted on Fig. 2.3.

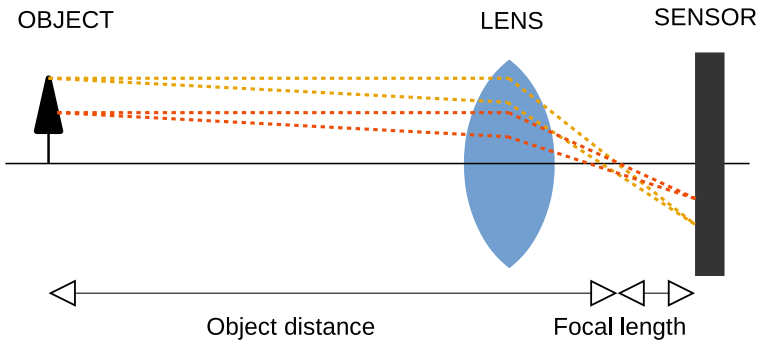


Figure 2.3: Basic diagram of a digital camera structure.

From this structure, it is important to understand the impact of the multiple parameters.

- The distance between the lens and the sensor determines the distance at which the objects will appear completely in focus on the resulting image. Typically, the image plane is fixed on a camera and the lens is slightly moved, as necessary.
- The shape of the lens determines the *focal length*. This translates to how much the image will be affected by perspective distortion. It also directly relates to the field of view (FOV) of the camera. A short focal length gives a wider FOV with more perspective distortion and a long focal length gives a narrower FOV with less perspective distortion (see Fig. 2.4).
- The configuration of the sensor determines the resolution of the resulting image.

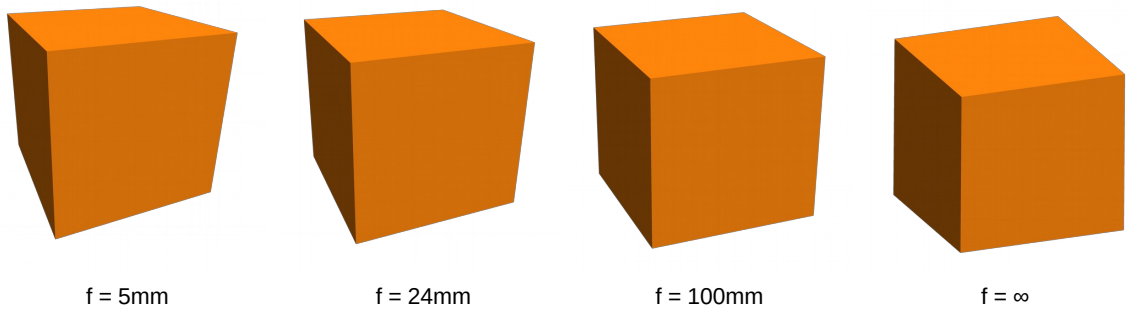


Figure 2.4: Comparison of the impact of the focal length in the resulting photograph. A smaller focal length enhances the perception of depth and increases the perspective distortion, causing far objects to appear smaller on the image. An infinite focal length causes no perspective distortion and no depth.

In a digital camera, the sensor is formed by an array of photodetectors, which translate into pixels on the resulting image. As light reaches each detector, a signal is produced and measured. The two main types of digital image sensors are charge-coupled device (CCD) and complementary metal–oxide–semiconductor (CMOS), being CCD the oldest technology. Both

types consist of a pixel array of photodetectors that deliver an electrical signal proportional to the amount of photons that fall on the pixel surface during a given time (the *exposure time*). Furthermore, both rely on the photoelectric effect in silicon [6]. In a CCD, after the incident photons in a pixel are converted into a charge, that charge is transferred sequentially to a common output structure, were it is converted into a voltage. In a CMOS, however, this same conversion takes place in each pixel.

Without going into further detail about the physics behind each type, from a general point of view, the CCD architecture offers better image quality and has a higher noise immunity whereas CMOS have a lower power consumption, simpler control circuitry and lower size [7]. In order to better understand the reasons for the arising of the CMOS technology over CCD, a comparative view of both is given in Table 2.2.

	CCD	CMOS
Signal out of pixel	Charge	Voltage
Signal out of chip	Voltage	Bits (digital)
Signal out of camera	Bits (digital)	Bits (digital)
Speed	Standard	High
System complexity	High	Low
Sensor complexity	Low	High
System noise	Low	Moderate
Overall image quality	Higher	Lower
Power consumption	Standard	Low

Table 2.2: Comparison between CCD and CMOS characteristics [7].

Given their advantages over CCD, CMOS chips production increased and their price dropped, making CMOS sensors the most commonly used today, mainly in consumer electronics [8]. As such, only this type of sensors will be further explained, in order to understand their characteristics and how they can be used as receivers in communication systems.

2.3.2 The CMOS image sensor

As described before, a CMOS sensor consists of a pixel array of similar structures. A diagram of the sensor is present in Fig. 2.5. In order to capture coloured images, each pixel corresponds, typically, to a set of four aggregated receivers, placed in a 2×2 array. In this array, each receiver has a color filter placed in front of it, separating the incident light ray into color components. Regarding this, there are multiple possible arrangements for the filters in these elements, resulting in different color spaces used for the pixels in the sensor. The most common is the Bayer filter, in which light is separated into its red (R), green (G) and blue (B) components [9]. Since it is a 2×2 pattern, there are two green receivers in each cell, providing additional resolution to this component. This pattern is also illustrated in Fig. 2.5.

On the pixel array, each single structure is responsible for outputting a voltage signal proportional to the number of photons falling onto it during a set period of time. The multiple voltage signals resulting from each pixel of the array are read by additional circuitry embedded on the chip, outside of the pixel array. This architecture is illustrated in Fig. 2.6.

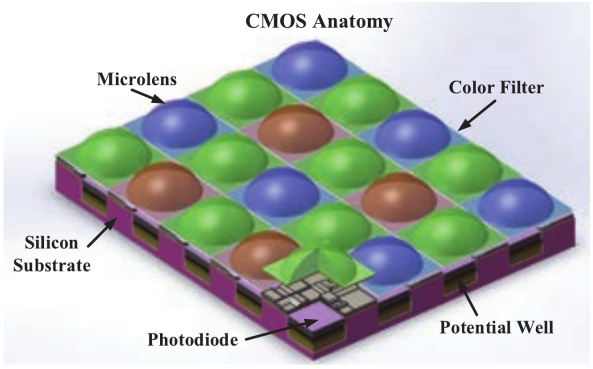


Figure 2.5: CMOS image sensor anatomy [9].

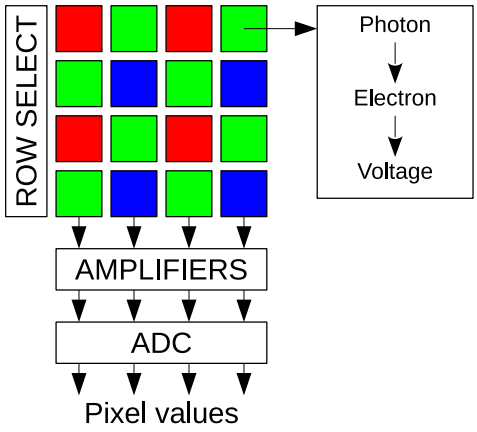


Figure 2.6: CMOS camera architecture.

In order to reduce the complexity and cost of the chip, these additional circuits are typically common for all rows, meaning that the decoding process happens for a single row at a time, sequentially. As such, and as opposed to what the typical case is with CCD cameras, each row is captured at a slightly different time instant [10]. This phenomenon is further developed below, as it constitutes one of the most important features of a CMOS camera from the perspective of VLC. It is also important to note that there are CMOS cameras with both global and rolling shutters. The most common, however, are the later, for the aforementioned reasons [11]. Furthermore, these are the most relevant for VLC and, as such, only the rolling shutter type will be discussed.

2.3.3 The rolling shutter effect

In developing a camera-based IPS, it is fundamental to have the means to distinguish the multiple light sources between them. One possible way of achieving this is, as previously stated, by assigning different colours to the light sources, since cameras are natural multicolor receivers. This, however, is not practical and goes against the principle of using regular LED fixtures to serve both the illumination purpose and the communication necessary for the positioning system. As such, and as emphasized in the previous section, the alternative is to modulate the light sources by turning them on and off at relatively high frequencies, sufficiently high so that the human eye can not perceive the intensity fluctuations.

Regarding the camera sensor, most typical cameras have either an acquisition rate of 30 or 60 frames per second, meaning that relying solely on taking successive photos, receiving a single symbol per photograph, is not a viable option, as we will not be able to encode the data at such speeds without causing light flickering. Despite not existing an established consensus on the minimum frequency required for the flickering effect to be avoided, it is generally accepted that the human eye cut-off frequency is at around 100 Hz. As such, a lower limit of 200 Hz is often considered in VLC systems [12], [13].

However, as evidenced before, most CMOS cameras acquire each row of the image successively, meaning that there is a time difference between each row. This means that a light source being turned on and off at a frequency close to the image sensor's row acquisition rate is captured as consecutive white and black strips [10]. An illustration of this phenomenon is presented in Fig. 2.7.

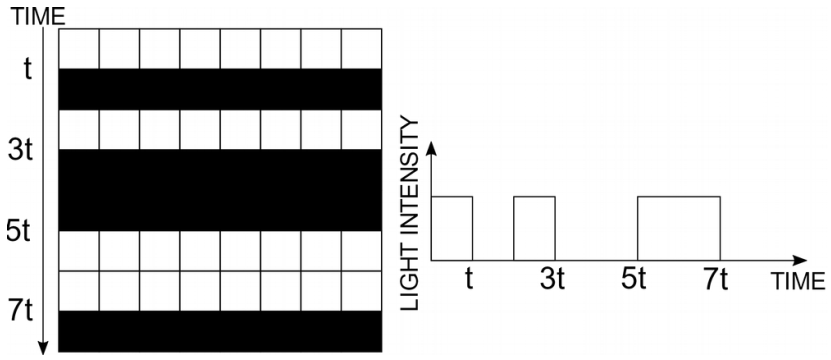


Figure 2.7: Illustration of the rolling shutter effect.

Taking advantage of this, successive bits can be transmitted and received by an image sensor. The only limitation is that, since we can not guarantee the delay between successive pictures taken, we have to capture the whole transmitted frame in a single image, meaning that this process can only be used to transmit a quite limited amount of information [10]. This is not a severe limitation for the work in question, however, since our only intention is to transmit a simple ID in order to distinguish each light source [14]. Each ID requires only a small number of bits, rendering the task of fitting these IDs on a single photograph quite feasible. Modulation schemes for encoding the data will be presented and discussed further in this dissertation.

2.3.4 Why use cameras for communication?

As established above, cameras can, indeed, be used as receivers for VLC systems. Without taking into account their potential for application in positioning systems, there are a variety of reasons for employing a camera for visible-light based communication systems, standing out:

- Widespread technology: Cameras are extremely common nowadays and, given the popularization of smartphones, almost every person already carries one in their pockets.
- multiple-input and multiple-output (MIMO) receiver: Since each pixel in a camera can be seen as capturing a separate angular region inside the camera's FOV, it is a natural

receiver for spatially separated transmitters. In an optimal situation, each pixel can act as a separate receiver.

- Multicolor receiver: Multiple signals can be easily separated by being transmitted in different colors/wavelengths, without the need of any additional filters on the receiver side.
- Low interference: most interference problems found in photodiode-based VLC systems are greatly reduced in camera-based ones. As opposed to a photodiode, a pixel's FOV is greatly narrower, making the background interference light a naturally mitigated problem in an image sensor.

Just for comparison sake, Table 2.3 presents a summary of the differences between a VLC system using a photodiode and a camera as receivers.

Receiver	Photodetector	Image sensor
Interference	High	Low
SNR	Low	High
MIMO Multiplexing	Easy to implement	Difficult to implement
Decoding (complexity)	Signal processing (low)	Image processing (high)
Data rate	High	Low
Range	Near	Far

Table 2.3: Comparison between photodetector-based and camera-based communication systems [9].

As can be observed, the major disadvantages of using cameras are the high decoding complexity and the low data rate. Regarding the decoding complexity, this renders them not suitable for applications such as microcontroller-based systems. However, image processing is still not a prohibitively complex task, meaning that such a system should be easily implemented on, for example, a Raspberry Pi.

Regarding the data rate, it is important to note that there are multiple approaches on how to receive data using a camera, as evidenced before. The techniques for modulating the data into light will be discussed below but, essentially, we can use the camera as a receiver in two modes — by taking successive photographs and receiving a single coded symbol per frame or by taking advantage of the rolling shutter effect, receiving the whole data frame on a single image [15] [16]. Either way, since the camera is a natural MIMO receiver, increasing the number of transmitter LEDs should directly increase the data rate. Similarly, a higher resolution on the camera should also translate in a higher data rate, particularly when using the rolling shutter, since each row or group of rows will translate into a symbol. By having more rows in a single picture, assuming a constant frame rate, we can pack more symbols in the same picture, thus increasing the overall data rate. Finally, a higher frame rate should also increase the data rate, both when using the rolling shutter and when receiving a single symbol per picture.

Considering all this, albeit not perfect, cameras are a remarkable candidate for receivers in VLC systems. Moreover, their natural spatial separation, besides being a useful feature from the Optical Camera Communication (OCC) perspective, can also be taken advantage

of to employ them in VLP systems. This characteristic essentially means that a camera can be seen as an aggregate of light receivers, each with a very narrow FOV. This means that, intrinsically, a camera behaves as an angle detector, by relating each pixel with a different angle between the light rays that fall on it and the camera normal. All these concepts will be further explained in the next section.

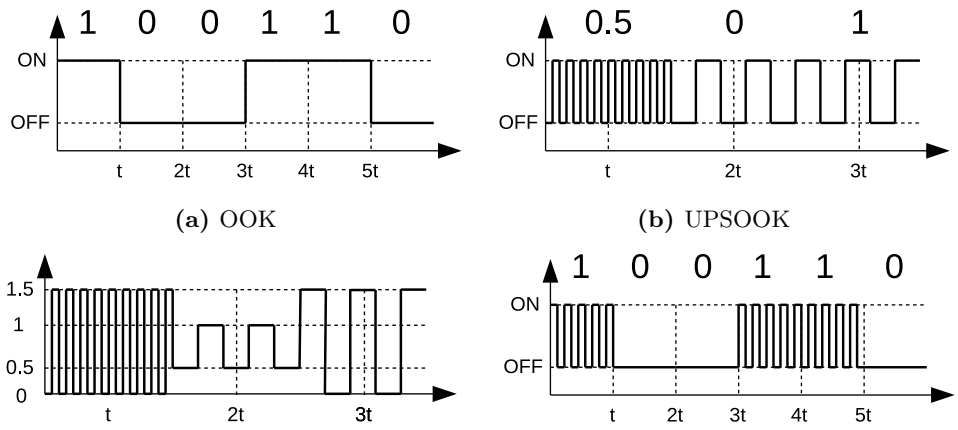
2.3.5 Modulation techniques

After presenting the working principles of the camera and how to make use of the rolling shutter effect, we have to analyse the possible techniques for modulating the light in order to achieve the communication necessary for this work. In order to clarify this process, a summary of the most relevant contributions in OCC is present in Table 2.4. Since the purpose of this work is to implement a room-size VLP system using OCC, only contributions featuring a distance of over 2 meters were considered. Existing works with a distance smaller than this often focus on data rates higher than necessary and are often not suitable for the implementation of this work. Furthermore, only contributions focusing on LED to camera communication were considered, since the implementation of this work requires the multiple light fixtures to be independent, each transmitting a separate message.

Ref.	Data rate (kbps)	Distance (m)	Modulation technique	Rolling shutter
[17]	0.002	25	OOK	No
[18]	0.1	50	UPAMSM	No
[19]	0.15	12	UPSOOK	No
[20]	0.62-1.35	1.5-5.5	Hybrid OOK-PWM	Yes
[21]	84	4	PWM	Yes

Table 2.4: Contributions summary on OCC [15].

First of all, all works are based on ON-OFF keying (OOK), illustrated in Fig. 2.8a, meaning that the light is modulated by being switched on and off, without any intermediate intensity levels. This makes sense, since it renders the LED drivers much simpler to implement. In [17] an LED is simply turned on and off according to the data bits. While reliable, this technique does not allow for light fixtures to serve both the illumination and communication purpose. Furthermore, the rolling shutter effect is not taken advantage of, meaning that a single bit is obtained from each photograph, hence the low data rate. The work developed in [19], on the other side, takes measures to prevent visible light flickering. Instead of pure OOK, under-sampled phase shift ON-OFF keying (UPSOOK) is used, consisting of modulating the light using high frequency square waves. In this contribution, two frequencies are used: a frequency much higher than the camera frame rate, which causes the camera to perceive it as a half-on state and a frequency only slightly higher than the frame rate. The second causes the light to be perceived as ON or OFF, depending on the phase of the transmitter, as illustrated in Fig. 2.8b. This solution is also not suitable for the present work, since a full data packet requires multiple successive photographs, which adds unnecessary delay to the positioning, assuming that a typical smartphone camera is used. The work conducted in [18] uses under-sampled pulse amplitude modulation with subcarrier modulation (UPAMSM),



(a) OOK **(b)** UPSOOK

(c) UPAMSM **(d)** OOK-PWM

Figure 2.8: Modulation techniques for OCC.

which is similar to UPSOOK but, instead of a single LED, two sources are combined to provide more intensity levels, without the need of added driver complexity. The technique is illustrated in Fig. 2.8c. This increases the data transmitted per camera frame, but still requires multiple photographs for a complete transmission. Furthermore, the usage of two independently-modulated light sources to transmit a packet of information is not suitable for the current work.

Finally, the two remaining contributions make use of the rolling shutter effect, allowing for a complete packet of information to be captured in a single photograph. In [20], the data is modulated simply using OOK. An additional pulse width modulation (PWM) modulation is added, however, on top of it, in order to allow the adjustment of light intensity, as illustrated in Fig. 2.8d. From the perspective of the camera, the PWM is high enough so that the camera can't detect the multiple pulses, only seeing it as a single, dimmer, pulse. Contribution [21] focuses on car-to-car communication and, while using similar principles to [20], encodes the multiple zeros and ones as short or long PWM pulses. This improves the noise immunity of the system, which makes sense, since it focuses on an outdoor scenario. For indoor, however, such precautions should not be as important, considering the much lower lever of background illumination noise.

From these contributions, the one more closely related to the objectives of the current work is [20]. As such, the adopted modulation technique will be OOK, considering its simplicity of implementation. The light intensity control, however, will not be considered. In order to maintain the light intensity level independently from the transmitted data, Manchester coding will be used on top of OOK, guaranteeing a constant average signal value. More sophisticated intensity control techniques, however, are considered out of the scope of this work.

2.4.1 Overview

As explained previously, a necessary requirement for implementing an IPS with a camera is to have a method of distinguishing the multiple light fixtures. The essential reason for this requirement is the fact that, in order to calculate the sensor's location, we must know the location (in the world coordinates) of each fixture. Given this, our system consists of a set of light fixtures and a receiver which, for illustration purposes, will be considered as being, for example, the camera included in a smartphone. A diagram illustrating this setup is present in Fig. 2.9. Moreover, the goal is to allow a completely unconstrained receiver, meaning that the camera can be hand-held, without the need to be placed, for example in a certain orientation or at a certain height.

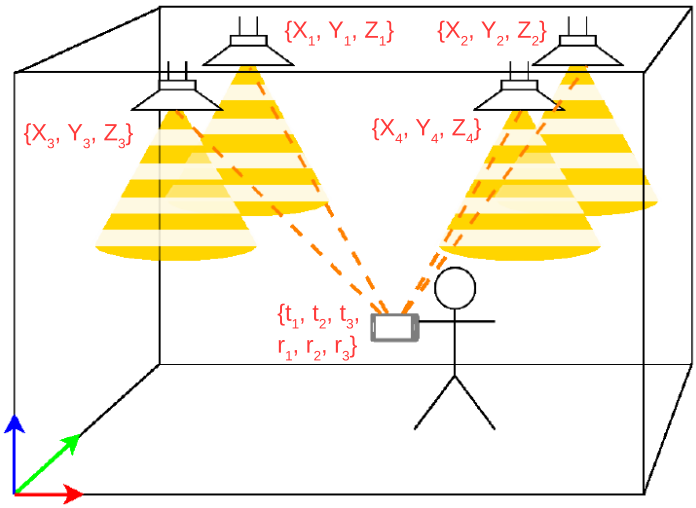


Figure 2.9: Minimal camera-based VLP system setup with an unconstrained receiver.

Regarding the position calculation, the basic principle behind it is the fact that a photograph is a 2D projection of the 3D world. As such, if the image sensor has a known geometry and parameters and the real-world coordinates of the light fixtures are known in advance, we can infer the sensor's coordinates based on the image coordinates at which the light sources appear. The whole purpose of using the rolling shutter effect to receive an unique ID for each fixture is to provide the means to find the world coordinates of each fixture, which can be done with a lookup table with previously measured locations for the light fixtures, after successfully decoding each ID. The mathematical principles behind this whole process will be, again, discussed later on. A diagram of the received image is present in Fig. 2.10.

As can be observed in the figures, our goal is to find the position and orientation of the camera. As explained previously, since this system involves angular information, even though we are only interested in the camera position, we always have to calculate the camera orientation, whether or not this value is a desired output. As such, the camera is represented by a set of six unknown variables:

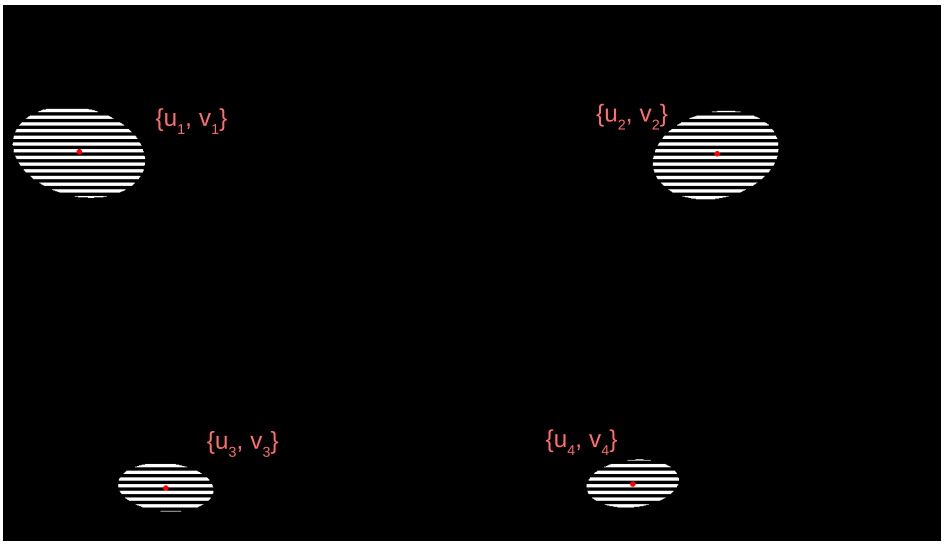


Figure 2.10: Typical image acquired by the camera in a camera-based VLP system.

- $\{t_1, t_2, t_3\}$ represents the camera position, relative to a fixed referential in the considered room;
- $\{r_1, r_2, r_3\}$ represents the camera orientation, relative to the same referential.

From the transmitters side, each light fixture is represented by a set of three coordinates — $\{X, Y, Z\}$ — representing their position, relative to the fixed room referential. These coordinates can be seen as constants, since their value does not change with the camera's movement, and are obtained, again, through the IDs obtained from the usage of the rolling shutter effect. Furthermore, on the image acquired by the camera we have the projections of the light fixtures. Here, each is represented by a pair of coordinates — $\{u, v\}$ — which correspond to the location, in pixel coordinates, of each light fixture's projected center. In order to simplify the image processing algorithm, the light fixtures are seen as punctual light sources, without any shape or volume. This assumption also enables the development of a system without any restrictions regarding the light fixtures' shape which, again, contributes to out objective of using already existing illumination infrastructure with minor adaptations.

Considering that we have six unknown coordinates, following the mathematical principles behind system solving, we need at least six independent known variables to be able to calculate all of the unknowns. As such, considering that each light fixture visible in the photo gives us two known variables, we would need at least three fixtures visible in the photograph in order to calculate the position of the camera. In this situation, however, up to four real solutions can be found for the camera position and orientation [22]. As such, an additional fixture is necessary in order to remove the ambiguity, rendering the minimum number of light fixtures visible in each photograph as four. It is also important to note that these fixtures can not be collinear.

The problem of finding a camera pose given a set of point correspondences between the image and the 3D space is called the Perspective-n-Point (PnP) problem. This mathematical problem is not new and there are multiple algorithms already implemented to solve it.

Ultimately, however, we will use an already ready-to-use algorithm included in the OpenCV library, which will be presented later.

2.4.2 Existing work

Multiple contributions have already been made regarding camera-based VLP systems. Most of these works, however, impose constrains for the camera movement, usually limiting its angular movement. Furthermore, most experimental setups are made on a small scale when compared with typical rooms inside buildings. The most relevant contributions on camera-based IPSs are summarized and compared in Table 2.5.

Ref.	Axis	Testbed size (m)	Accuracy (cm)	LEDs	Additional sensors
[23]	2D	–	Room	–	–
[24]	2D	50 (distance)	<100	–	IMU
[25]	2D	5.4×7.5×2.1	10	6	Fish-eye lens
[26]	2D	0.71×0.74×2.26	7	5	–
[27]	2D	0.75×0.35×1.8	6.6	3	–
[28]	2D	±5×5×2.5	6	4	–
[29]	2D	0.5×0.5×0.6	5	3	–
[30]	2D	0.6×0.52×0.6	2.4	3	–
[31]	3D	2.4×1.8×3	30	2	Polarization filters
[32]	3D	5×3×3	25	3	Accelerometer
[33]	3D	–	10	–	–
[34]	3D	1.4×1.4×1.6	8.5	3	–
[35]	3D	0.6×0.6×2.6	6.1	2	Dual camera
[36]	3D	1.8 (height)	4.3	3	IMU
[37]	3D	±5×5×2.5	3.2	4	–
[38]	3D	0.7×0.7×1.8	2.67	1	Gyroscope

Table 2.5: Comparison of localization and navigation systems based on OCC [5], [15]. Ref.: reference; 2D - two dimensions (no height consideration); 3D - three dimensions (height considered); Testbed size is width×depth×height; LEDs: minimum number of visible LEDs required for the positioning.

As we can see, there are contributions focusing both on 2D and 3D positioning. Starting with the 2D scenarios, works [23] and [24] use the LED fixtures as beacons and estimate the position, when possible, based on the visible beacon on a given instant. For this reason, the first system can only place the user in a given room. The second, however, in addition to the latter, uses an Inertial Measurement Unit (IMU) to provide a rough position estimation when no beacons are within the camera's FOV and LOS. The work developed in [25] makes use of a fish-eye lens, providing a much wider FOV for the camera, reducing the number of situations in which an insufficient number of light fixtures appear on the image. The techniques used in these three works, despite interesting, do not overlap with the objectives for the current work, as the goal is to implement an absolute positioning system without the aid of additional sensors.

The remaining contributions for the 2D scenario require a minimum of 3 LEDs visible in the image, as expected. The most accurate is [30], despite having a considerably small

testbed size. The most generously-sized setup is the one used in [28], providing an average accuracy of 6cm. This, of course, corresponds to the receiver placed at a fixed height.

For the 3D situation we find, again, the use of auxiliary sensors to either improve the accuracy of the system or to allow the system to work with a smaller number of visible fixtures [38]. A particular technique is used in [31], consisting of taking advantage of light polarization for the fixture identification, instead of directly modulating the light intensity and, thus, not requiring the use of the rolling shutter effect. Work [35] uses a dual camera setup which, again, goes against the goals of the current work. Regarding this, it is important to note that, despite dual camera smartphones becoming increasingly more popular, this camera is usually placed on the opposite side of the screen, rendering it non-practical for indoor positioning scenarios.

The remaining contributions for 3D — not using any additional sensors — require a minimum of 3 LEDs, again, as expected. The most accurate is [37], with an average error of 3.2cm. The largest setup is the one used in [37] which, interestingly, is the second most accurate system considering 3D.

From the listed contributions, it is import to note that the only works that, besides considering the 3D scenario, explicitly allow the receiver to have an arbitrary orientation and test such scenario are [32], [36], [37] and [38]. From these, the only one which does not use auxiliary sensors is [37].

On a final remark, the work most closely related to the current one is [37]. Its design allows an unconstrained placement of the receiver and requires at least 4 LEDs present in each picture taken. Its main downside, however, is that the light fixtures are differentiated by color, rendering the system substantially simpler to implement. The objective of this dissertation, however, is to not only implement accurate positioning using just a commercial camera but also to use light fixtures working both as conventional illumination sources and signal transmitters, something that can not be verified by using color as a differentiation means.

CHAPTER 3

Camera imaging

This chapter focuses on the working principles of a camera form a geometrical point of view, including the mathematical expressions behind it. Using these same expressions, a MATLAB script was developed in order to generate a simulated photograph of the multiple light fixtures, as seen by the camera. The results of this are presented at the end.

3.1 The pinhole camera model

As previously explained, a camera is a device used to capture incident light into still images, resulting in a 2D projection of the 3D world in front of it. Typically, it is formed by an image sensor and a lens placed in front of it, whose purpose is to focus and direct the light rays into the sensor. This set can be described, in an ideal situation, by the pinhole camera model. This model is presented in Fig. 3.1.

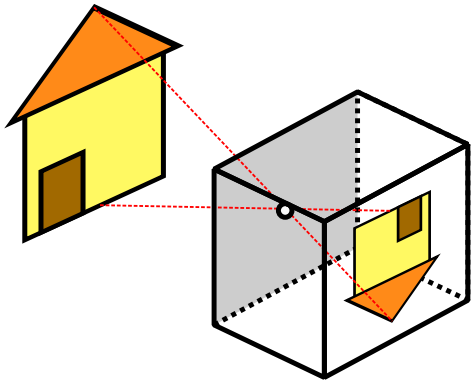


Figure 3.1: Pinhole camera diagram. Light rays pass through an infinitely small hole, causing the captured object to be projected on the image plane, flipped in both axis.

In this model, the lens of the camera are seen as an infinitely small hole, placed at a certain distance from the image plane. The image plane is an infinite plane with an infinite sampling resolution. Given these characteristics, a pinhole camera does not have a *focal distance*, as an

image is always projected in focus, as a result of the infinitely small hole in the place of a lens, as would happen on a real camera. It still has, however, a *focal length* which, similarly to a real camera, corresponds to the distance between the image plane and the pinhole. In a similar fashion, the focal length determines the amount of perspective distortion present in the image and the FOV of the camera. Finally, the scene is projected on the image plane flipped on both axis, as a direct result of the light rays passing through the pinhole.

The geometrical representation of the pinhole camera is present in Fig. 3.2.

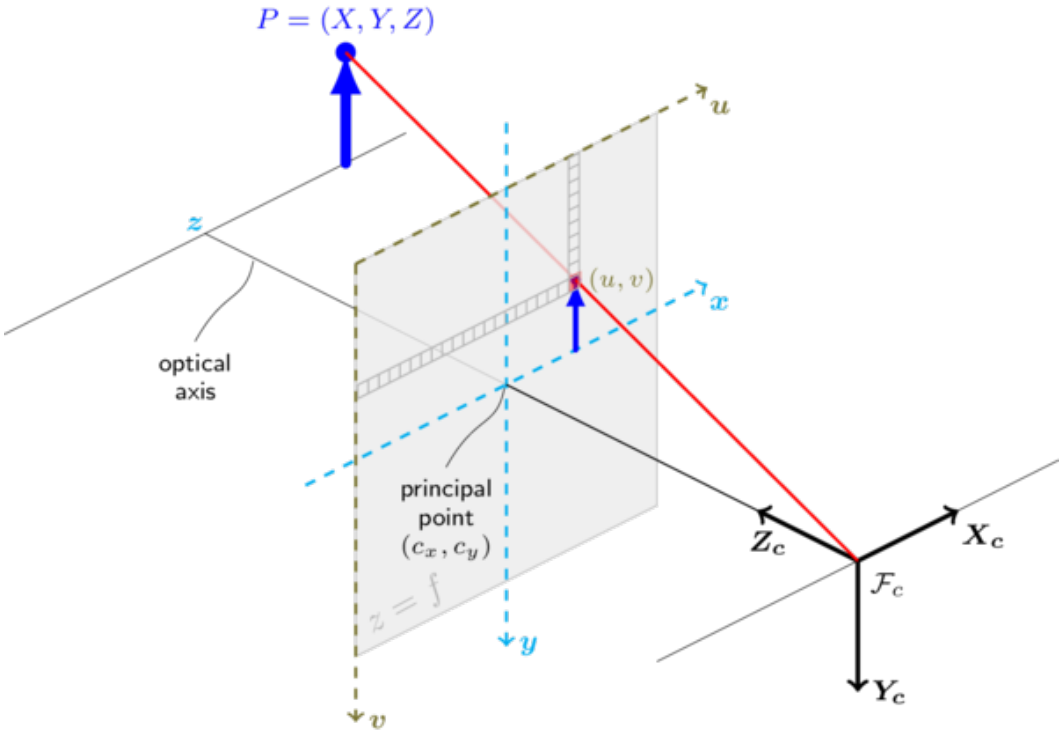


Figure 3.2: Geometrical model for the pinhole camera [39].

As can be seen, the focal length f corresponds to the distance between the focal center F_c and the image plane. A given point P in the 3D space is projected onto the image plane at (u, v). It is also important to note that the camera axis follows the standard coordinate system for pictures, with the x axis pointing to the right, the y axis downwards and the z axis pointing to the front (from the perspective of someone looking through the camera). The resulting image has a similar coordinates system, but without the z component. Finally, the camera center is considered to be the focal center, F_c.

Since we will be working with geometric transformations, all the operations considered in this section will be represented by matrices. Moreover, we will use the homogeneous coordinates system to represent the multiple points, both in the 2D and 3D space. In the next section, the equations that describe the camera's behaviour will be presented, always associated with the model presented in Fig. 3.2.

3.2 Camera transformation matrix

In the pinhole camera model, the picture is formed by projecting the 3D points onto the image plane using a perspective transformation. The matrix that represents this transformation is:

$$M = \begin{bmatrix} f_x & 0 & 0 & c_x \\ 0 & f_y & 0 & c_y \\ 0 & 0 & 1 & 0 \end{bmatrix} \tag{3.1}$$

where f_x and f_y are the focal lengths expressed in pixel units and (c_x, c_y) is the principal point, representing the central pixel in the image plane. The focal length itself, as explained before, corresponds to the distance, in metres, between the pinhole and the image plane. The values in pixel units can be obtained with:

$$\begin{aligned} f_x &= F\frac{N_x}{W} \\ f_y &= F\frac{N_y}{H} \end{aligned} \tag{3.2}$$

where F is the focal length, in metres, N_x and N_y are the number of pixels in both axis and W and H are the sensor width and height, respectively, in metres. Considering this, given an arbitrary point $P(X, Y, Z)$, we can calculate the coordinates of its projection on the image plane $P'(u, v)$ using:

$$\begin{aligned} P' &= M \times P \\ \begin{bmatrix} u \\ v \\ 1 \end{bmatrix} &= \begin{bmatrix} f_x & 0 & 0 & c_x \\ 0 & f_y & 0 & c_y \\ 0 & 0 & 1 & 0 \end{bmatrix} \times \begin{bmatrix} X \\ Y \\ Z \\ 1 \end{bmatrix} \end{aligned} \tag{3.3}$$

Since we are using projective geometry, the perspective transformation is valid assuming that the camera is placed at the origin point of the world coordinate system, with all the orientation angles set to zero. As such, in order to obtain the image coordinates of the projection of a set of points, the camera's position is represented by applying a rotation and translation operation to the world itself. A visual representation of these operations is present in Fig. 3.3. It is important to note that a rotation, as an elementary operation, is always applied around a given axis, with the center of rotation at the origin of the referential.

Regarding the translation, this operation is represented by the matrix:

$$T(t) = \begin{bmatrix} 1 & 0 & 0 & t_x \\ 0 & 1 & 0 & t_y \\ 0 & 0 & 1 & t_z \\ 0 & 0 & 0 & 1 \end{bmatrix} \tag{3.4}$$

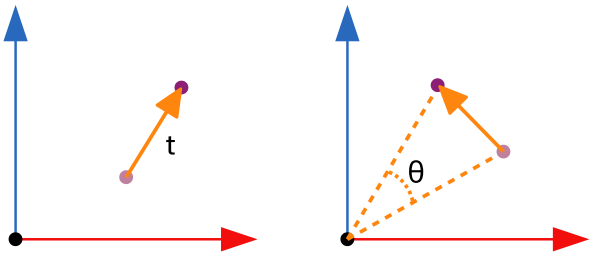


Figure 3.3: Visual representation of the translation and rotation geometric transformations. On the left: translation, following a vector t. On the right: rotation of an angle θ, with the center at the origin.

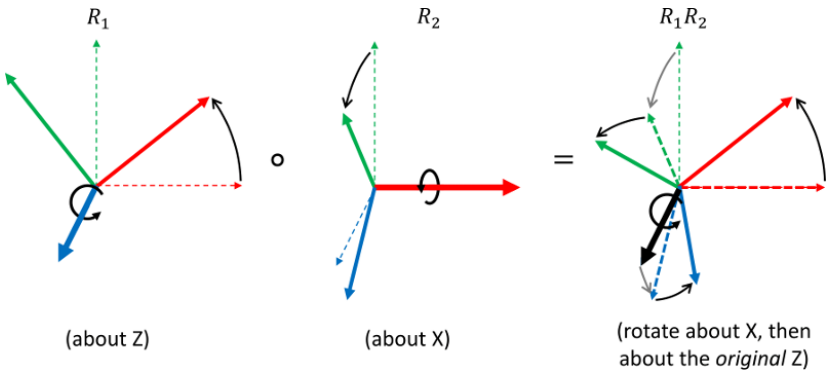


Figure 3.4: The Euler Angles rotation system. The rotations are decomposed into three successive rotations around single axis. Each rotation is applied relative to the original referential, meaning that changing the order by which they are applied yields different results [40].

where (t_x, t_y, t_z) is the desired translation vector. In this scenario, the translation vector directly correspond to the position of the camera in the world.

For the rotation, the situation is slightly more complex. A rotation in the 3D space, contrarily to a translation, can be represented in multiple ways. Without explaining all the multiple conventions, the Euler Angles system will be used in this dissertation. This consists in representing a 3D rotation as the result of three successive rotations, each around one world axis. A representation of this system is present in Fig. 3.4.

Since the rotation referential is always the *world axis*, the order of the Euler Angles change the resulting transformation. Moreover, the same axis can be used twice, adding even more possibilities to this order. Euler Angles can be divided into two groups: proper Euler angles and Tait-Bryan angles. The first use the same axis in the first and third components of the rotation ($z-x-z$, $x-y-x$, ...) and the last uses a different axis for each component ($x-y-z$, $y-z-x$, ...). In this dissertation, a specific combination of the Tait-Bryan angles will be considered — $z-y-x$ — since this is the order typically used when describing the orientation of a ship or aircraft and is the one used in most IMUs. The three matrices that

represent the rotations around each angle are:

$$
\begin{aligned}
R_x(\theta_x) &= \begin{bmatrix} 1 & 0 & 0 & 0 \\ 0 & \cos(\theta_x) & -\sin(\theta_x) & 0 \\ 0 & \sin(\theta_x) & \cos(\theta_x) & 0 \\ 0 & 0 & 0 & 1 \end{bmatrix} \\
R_y(\theta_y) &= \begin{bmatrix} \cos(\theta_y) & 0 & \sin(\theta_y) & 0 \\ 0 & 1 & 0 & 0 \\ -\sin(\theta_y) & 0 & \cos(\theta_y) & 0 \\ 0 & 0 & 0 & 1 \end{bmatrix} \\
R_z(\theta_z) &= \begin{bmatrix} \cos(\theta_z) & -\sin(\theta_z) & 0 & 0 \\ \sin(\theta_z) & \cos(\theta_z) & 0 & 0 \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 \end{bmatrix}
\end{aligned}
\tag{3.5}
$$

where the angles θ directly represent the camera's orientation, relative to the world referential, in radians. The rotation direction follows the right hand rule. Given all this, the camera's location on the world, P_C, can be described with:

$$
P_C = T \times R_x \times R_y \times R_z \times \begin{bmatrix} 0 \\ 0 \\ 0 \\ 1 \end{bmatrix} \tag{3.6}
$$

As a side note, it is important to remember that, when representing geometrical operations with matrices, the transformations are applied in the reverse order, meaning that the first matrix to take effect will be R_x, followed by R_y and so on. This follows the chosen Tait-Bryan sequence — $z - y - x$.

Since the projection equation considers the camera to be at a fixed position and orientation at the origin of the coordinate system, the angles θ and the coordinates t have to be multiplied by -1. This corresponds to the change of referential from the camera coordinates to the world coordinates, given that these vectors represent the camera's coordinates — assuming that the camera moves — and we need the reverse scenario — assuming that the world moves, instead of the camera. As such, the final world translation and rotation matrices T_W, R_{Wx}, R_{Wy} and R_{Wz} will be given by:

$$
\begin{aligned}
T_W &= T(-t) \\
R_{Wx} &= R_x(-\theta_x) \\
R_{Wy} &= R_y(-\theta_y) \\
R_{Wz} &= R_z(-\theta_z)
\end{aligned}
\tag{3.7}
$$

In a similar manner, the operation order also has to be reversed to represent the referential change. Given all this, the complete equation for obtaining the image coordinates of the projection of an arbitrary point P is given by:

$$\begin{bmatrix} u \\ v \\ 1 \end{bmatrix} = P \times R_{Wz} \times R_{Wy} \times R_{Wx} \times T_W \times \begin{bmatrix} X \\ Y \\ Z \\ 1 \end{bmatrix} \tag{3.8}$$

This equation will be the basis for the calculations involved in the developed simulation, as explained in the next section.

3.3 Camera acquisition simulation

As a preliminary step for the work contained in the next chapters, a MATLAB simulation was developed, intended to produce simulated photographs of a set of light fixtures, as seen by a parametrizable camera. The rendering of the simulated photograph consists in taking a series of points and, using the expressions presented above, finding their projection coordinates on the image plane. As such, the inputs for the simulation are the camera parameters (which directly correspond to the values present in the matrix M, described in the previous section), the size and position of the light fixtures and the camera position and orientation. Additionally, the possibility of assigning a different color for each fixture was included, as a tool for quickly distinguishing them in the picture. Finally, the rolling shutter effect is also illustrated, in a rather crude way, by applying a mask over the multiple fixtures, representing their successive ON-OFF states. As such, a parameter for controlling the frequency of this mask's transitions was also included.

A block diagram illustrating the simulation structure is present in Fig. 3.5. The following subsections contain a detailed description of each block. It is important to emphasize that, for the purpose of this simulation, the light fixtures will always be considered to be circular and placed horizontally, meaning that their normal vectors align with the y axis. This assumption greatly simplifies the simulation and is made considering that this will be the case with the experimental setup, presented later in this dissertation. Finally, the fixtures are also assumed to be identical, all having the same radius.

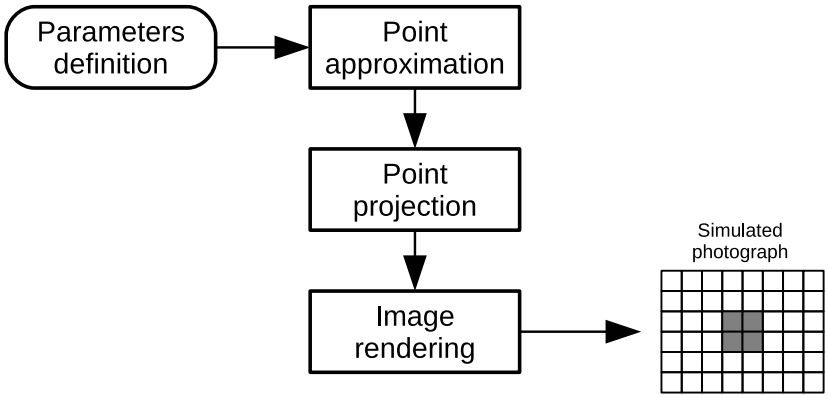


Figure 3.5: Block diagram of the developed simulation.

3.3.1 Parameters definition

As explained before, the inputs of the simulation are:

- The camera parameters matrix;
- The diameter of the light fixtures;
- The center position of the light fixtures;
- The render color for each fixture;
- The camera position;
- The camera orientation.

The simulation 3D setup is illustrated in Fig. 3.6. The number and placement of the light fixtures were chosen to match the experimental setup which, again, will be detailed later. The diameter of each fixture is 15cm, which also corresponds to their actual size in the experimental setup.

The camera parameters matrix will not be calculated using the variables described above. Instead, the matrix for the *actual* smartphone camera that will be used to acquire the experimental results was considered. This was obtained directly by taking a series of photographs to calibrate the camera, with the aid of image processing. This whole process is going to be described later but, for the purposes of this section, the matrix is:

$$M = \begin{bmatrix} 3912 & 0 & 0 & 2592 \\ 0 & 3912 & 0 & 1940 \\ 0 & 0 & 1 & 0 \end{bmatrix} \tag{3.9}$$

Regarding the camera position, as stated previously, it is completely unconstrained, meaning that the position can be anywhere in the 3D space. The case is similar with the orientation, meaning that the camera can face any direction desired.

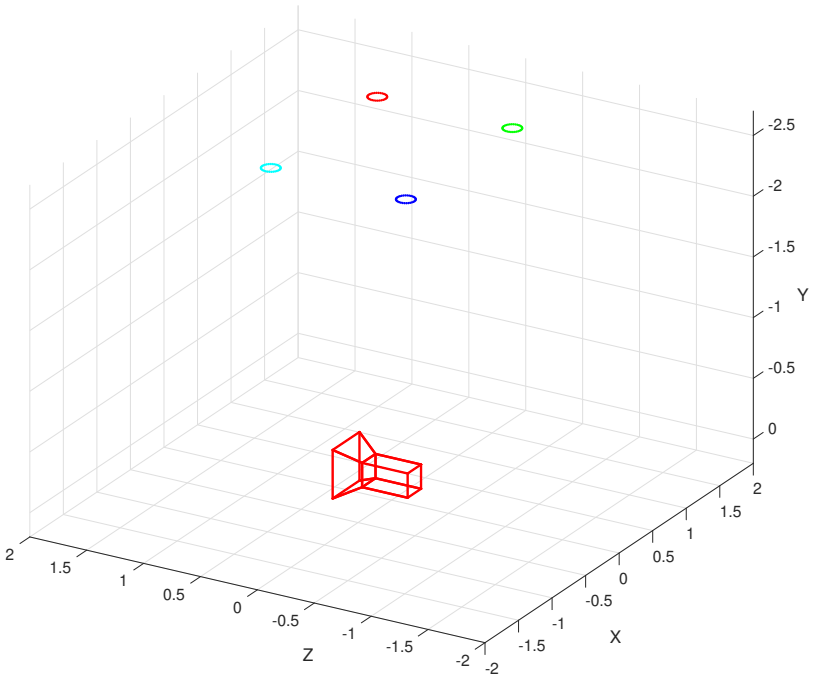


Figure 3.6: 3D representation of the simulation parameters. The camera is placed at the origin, with all the angles set to zero. The light fixtures are placed at a constant height, with their centers at $(0.7935, -2.7, 0.5935)$, $(0.7935, -2.7, -0.5935)$, $(-0.7935, -2.7, -0.5935)$ and $(-0.7935, -2.7, 0.5935)$. All units are in metres.

3.3.2 Point approximation

Without any additional mechanisms, projective geometry is used to calculate the projection coordinates of single points. Since the light fixtures are considered to be circles (as opposed to single points in space), in order to render them, we have to decompose them into a series of points, describing a circular approximation. When the fixtures are assumed to be placed horizontally, this process is greatly simplified. The process of decomposing each fixture into a group of points is illustrated in Fig. 3.7.

First, a point with coordinates $(0, 0, r)$ is considered, where r is the radius of the fixture. Then, a series of new points are calculated, which correspond to the first point with an applied rotation operation, around the y axis. The expression for this operation is:

$$P'_{circle} = R_y(n\theta) \times P_{circle}, 0 < n < N \tag{3.10}$$

where α is the rotation angle, in radians. Assuming an even spacing between the points, this angle is constant and multiplied by n, which corresponds to the instance number of each new point (starting at 1, for the first new point). The total number of steps to achieve a full circle is given by $N = 2\pi/\alpha$. After all the instances are calculated, the group is shifted with a

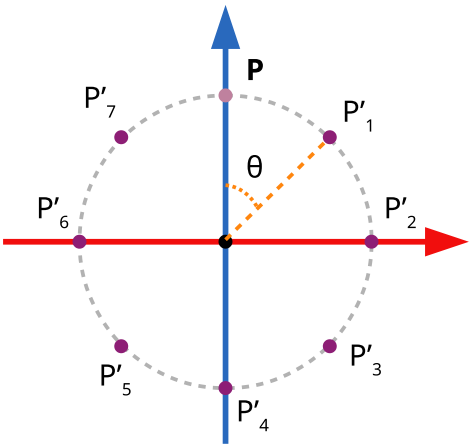


Figure 3.7: Illustration of the decomposition of each fixture into a group of points. In this scenario, the circle is decomposed into eight equally spaced points, each calculated trough successive rotations of θ.

translation operation corresponding to the fixture's center:

$$\begin{aligned} x' &= x + A_x \\ y' &= y + A_y \end{aligned} \tag{3.11}$$

where (A_x, A_y) is the fixture's center position.

3.3.3 Point projection

After performing the operations described above for each light fixture, we are left with a series of points in the 3D space. The next step is to calculate the pixel coordinates of the projection of each of these points, using the matrix operations described in the previous section, present in Equation 3.8.

It is worth mentioning that, since we are dealing with homogeneous coordinates, in order to obtain the final coordinates in the image (u, v), we have to divide the both the first and second elements of the result vector by the homogeneous term, the third element. As such, the resulting vector from Equation 3.8 will have the format $\begin{bmatrix} uw \\ vw \\ w \end{bmatrix}$, where w is the homogeneous component. As previously evidenced in the equation, our goal is to have the third component equal to 1, thus requiring the mentioned division.

3.3.4 Image rendering

After all the point projections calculated, the next step is to *actually* paint them in the corresponding pixels of an image. The first step is to crate a blank image, with the specified resolution. Again, the experimental results will be gathered with a camera with a resolution of $5184px \times 3880px$ and, as such, this value was used. This image is painted with a single color, which corresponds to the background. In the results above, the chosen color was black.

After the blank image prepared, the projection coordinates of each of the previous points are rounded to the nearest integer. This process corresponds to the pixel rounding that happens as a result of the finite resolution of the image sensor. The rounded coordinates are used as indexes for the corresponding pixel. Since our goal is to represent the fixtures as circles, instead of simply filling the corresponding pixels with the desired color, each group of pixels was approximated by a polygon, in order to represent the area delimited by them. This polygon was found using the MATLAB function `poly2mask`.

After calculating the aforementioned polygon for each light fixture, the pixels inside each of them are painted with the color chosen for the fixture in question. To complete the process, a mask with successive white and black rows is generated, according to the defined frequency ratio between the fixtures and the camera, to simulate the rolling shutter effect. Finally, the image is saved to a file, in the `PNG` format.

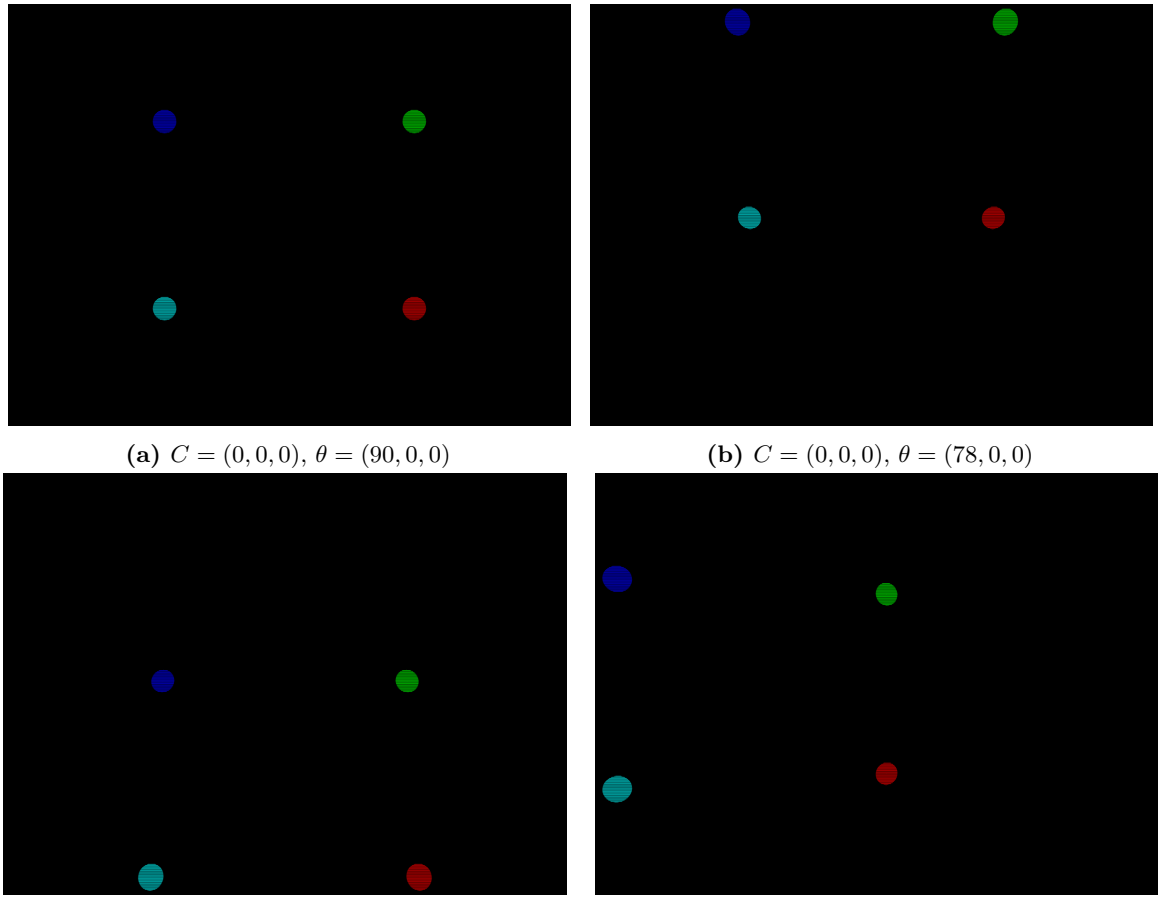
3.4 Results

As mentioned before, the results for this chapter are more illustrative that value-oriented, specially considering that they correspond to a series of images. As such, a series of poses were chosen to simulate the multiple photographs, with the intent of illustrating key situations that will arise when acquiring the experimental results.

The results are divided in three parts: impact of angle variation, perspective distortion and raising the camera height. The first one aims at illustrating the angular boundaries of the camera, while maintaining the four light fixtures visible in the resulting image. The second illustrates the impact of perspective distortion, showing the impact of higher tilting angles and a higher distance between the camera and the fixtures. The third illustrates the impact of raising the camera height, as opposed to keeping it on the floor. All units are in metres and degrees, respectively for the positions and the orientations. Furthermore, C corresponds to the position of the camera in the world coordinates and θ to the orientation around the x, y and z axis, respectively.

The set of images illustrating the impact of the angle variation are present in Fig. 3.8. From these, we can observe that, when placed at the origin, the camera can tilt around 12° around the x axis (Fig. 3.8b and 3.8c) and 15° around the y axis (Fig. 3.8d), in both directions. This angle is quite small and will greatly limit the possibilities for the experimental results. This problem, however, would be minimized in scenarios with a higher ceiling level, which is the case in most public spaces. Furthermore, the difference between both axis is expected, since neither the camera sensor nor the fixtures placement correspond to a perfect square.

The impact of perspective distortion is illustrated in Fig. 3.9. As our intuition led us to believe, with an increase in the tilting angle of the camera, the fixtures become less circular and more oval-shaped. Furthermore, with an increase in the distance between the camera and the fixtures, the fixture size decreases. It is important to highlight, however, that the strips width does not depend on this distance. Instead, it is only affected by the ratio between the modulated frequency and the camera's shutter speed.



(a) $C = (0, 0, 0)$, $\theta = (90, 0, 0)$

(b) $C = (0, 0, 0)$, $\theta = (78, 0, 0)$

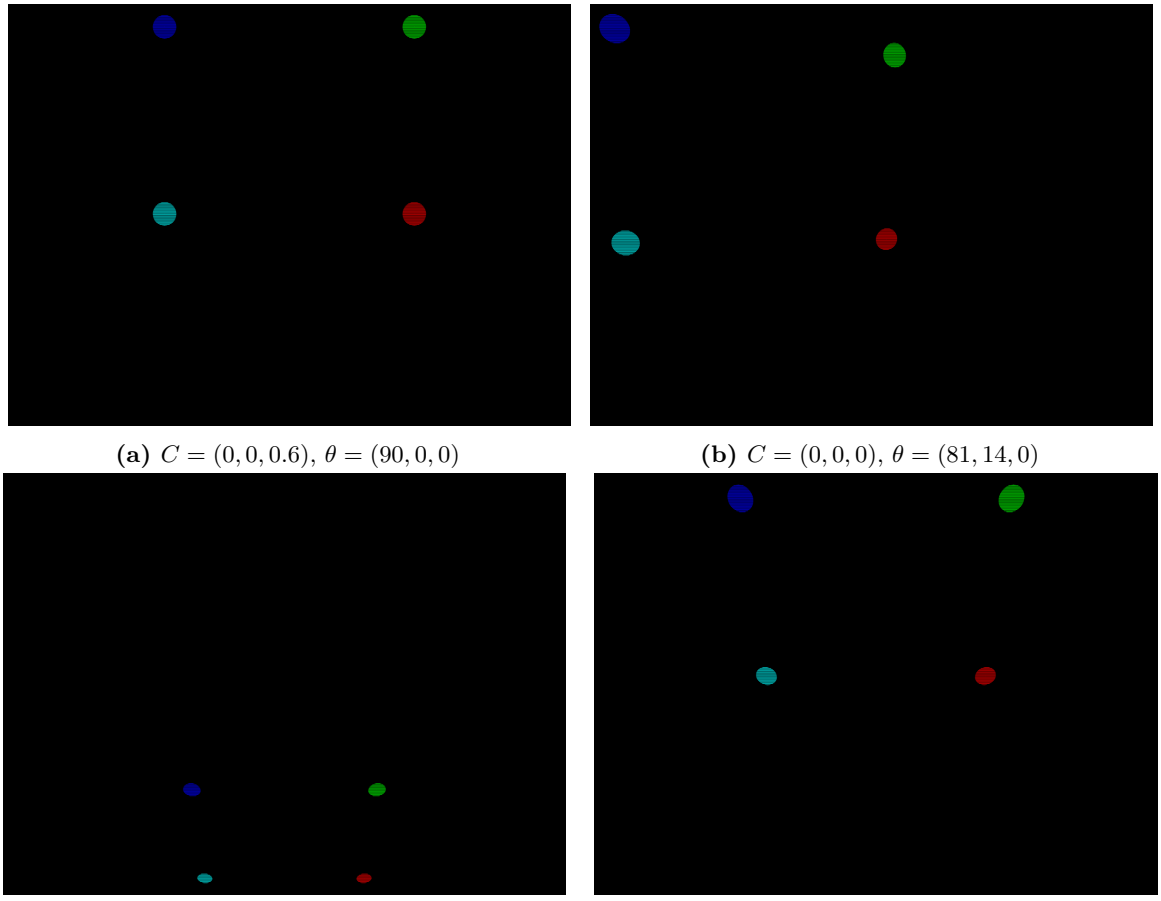
(c) $C = (0, 0, 0)$, $\theta = (102, 0, 0)$

(d) $C = (0, 0, 0)$, $\theta = (90, 15, 0)$

Figure 3.8: Impact of the angle variation on the captured image.

Regarding the impact of the tilting, we can also conclude that, as the angle increases, so does the floor area in which we can place the camera, while maintaining all the fixtures visible in the picture. This can be observed when comparing Fig. 3.8a with 3.9d and 3.9c with 3.9d.

Finally, the impact of raising the level of the camera to 1 m above the floor is present in Fig. 3.10. This value was chosen as the typical height at which the smartphone is usually held when being looked at by its user, hand-held. As we can observe, the four light fixtures are still visible, but both the angular and linear margins of movement are considerably smaller. The maximum tilting angles become about 4° around the x axis and 6° around the y axis.



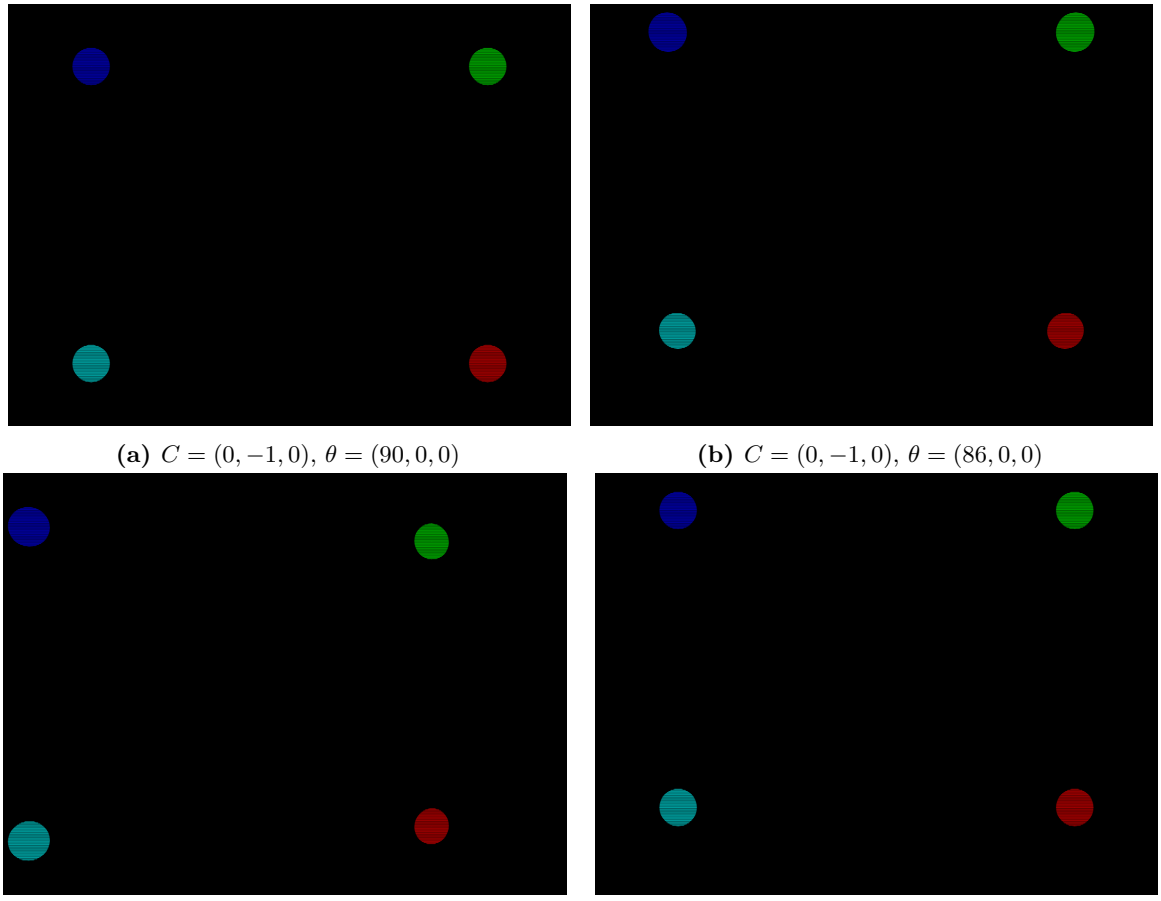
(a) $C = (0, 0, 0.6)$, $\theta = (90, 0, 0)$

(b) $C = (0, 0, 0)$, $\theta = (81, 14, 0)$

(c) $C = (0, 0, -3.2)$, $\theta = (60, 0, 0)$

(d) $C = (0, 0, -0.9)$, $\theta = (60, 0, 0)$

Figure 3.9: Impact of perspective distortion on the captured image.



(a) $C = (0, -1, 0)$, $\theta = (90, 0, 0)$

(b) $C = (0, -1, 0)$, $\theta = (86, 0, 0)$

(c) $C = (0, -1, 0)$, $\theta = (90, 6, 0)$

(d) $C = (0, -1, 0.1)$, $\theta = (90, 0, 0)$

Figure 3.10: Impact of raising the camera level on the captured image.

CHAPTER 4

Positioning error simulation

This chapter presents a simulation developed to study the impact of the error on the system. An introductory reflection on the sources of error is given, followed by an explanation of the multiple simulation components. Finally, the results are presented and discussed.

4.1 Possible sources of error

As is the case with every system, there are a number of factors that contribute to errors in the final position estimation. For the system in question, the sources identified are:

- The camera resolution: since the image is formed by an array of pixels, the number of pixels present directly relates to the final error. A larger number corresponds to a smaller error;
- The camera matrix: as explained in the previous chapter, the camera has to be calibrated in order to obtain its intrinsic parameters. Errors in this calibration directly translate to errors in the position estimation;
- The detection algorithm: in order to estimate the camera position from an image, the multiple light fixtures have to be correctly identified, with an image processing algorithm. If the results of such algorithm are not accurate, the position estimation will also have a larger error;
- The measurement of the fixtures position: as explained previously, we need to relate the location of each fixture in the resulting image with their position in the world. If their position in the world is measured incorrectly, it will also cause deviations in the final position estimation.

Regarding the errors introduced by an inaccurate calibration of the camera, these will not be considered in this work. Since this calibration only needs to happen once for a specific camera, it was considered that it can be done with sufficient accuracy in order to not introduce relevant errors in the final estimation. The remaining three sources — the *camera resolution*, the *detection algorithm* and the measurement of the *fixtures position* — are going to be studied with the developed simulation, described below.

Starting with the camera resolution, this parameter was considered to be the most relevant source of error from the sensor itself. Despite the fact that an image produced by a camera has noise and distortion, the positioning should be affected mainly by the rounding that happens as a result of the pixel resolution. Furthermore, considering the type of pictures that we want for this scenario, the noise in the image should have minimal impact. As described before, we will photograph the light fixtures with the camera directly pointing at them, while they are switched on and off. As a result, there should be no problem in identifying the ON and OFF states of each fixture, considering that the camera is pointing directly at them. Moreover, the spatial separation that occurs naturally in an image should also minimize the impact of these parameters. It is important to note, however, that there is visible distortion when considering the rolling shutter effect. This will be explained later and will not be considered in this chapter, since our goal here is to study the impact on the position estimation and such distortion mainly affects the communication part, which will be considered later in this dissertation.

For the detection algorithm, its impact on the position estimation is going to be studied by forcing an error in the fixture detection on the image. Similarly, the impact of the measurement of the fixtures' position is going to be simulated by forcing a deviation on this value and finding the error in the estimation that it causes.

4.2 The solvePnP function

As explained in Chapter 2, the problem of finding the position of a camera given a set of correspondences between points in the image and their position in the world in not new and is called the PnP problem. Considering that it is out of the scope of this dissertation to study the algorithms used to solve this problem in depth, an already developed solution was used — the `solvePnP` function. This method is included in the OpenCV library and outputs a set of six coordinates, which correspond to the camera's position and orientation in the world[41]. As for the inputs, the most important are:

- Object points;
- Image points;
- Camera matrix;
- Solving method.

The object points property corresponds to an array with the world coordinates for each point. In our situation, this corresponds to the coordinates of the center of each fixture, in the world referential. The image points are the coordinates of the center of each fixture in the image, in the same order as the object points array. The camera matrix corresponds to the intrinsic parameters of the camera, as explained in the previous chapter.

For the solving method, there are three main possibilities: `ITERATIVE`, `P3P` and `EPNP`. Starting with the P3P method, it requires exactly four points, which is not ideal for this work, since we want to use as many fixtures as possible for the position estimation, when possible,

and not be limited by four. For the Iterative and EPnP methods, both require at least four points and allow for a larger number.

Regarding the Iterative method, according to the official OpenCV documentation, it is based on the Levenberg-Marquardt optimization. The basic principle of this algorithm is that it tries to find a pose that minimizes the reprojection error, given by the sum of the squared distances between the initial coordinates of each image point and the reprojected coordinates obtained with each guessed pose. As a side note, the points are reprojected using the same expressions described in the previous chapter, similarly to our picture simulation. Finally, the EPnP method has a working principle similar to the Iterative method, but focuses on efficiency. The final chosen method was the Iterative, for two main reasons. The first is that efficiency is not the main concert for this work, since we are not working with a sensitive real-time system, as would be the case with augmented reality, for example. With the Iterative method, while slower that EPnP, we can get a result within tenths of seconds, which is enough for this application. The second is that, through testing, the Iterative method yielded more accurate results, which is our main goal for this work.

4.3 Simulation structure

A block diagram presenting the structure of the developed simulation is visible in Fig. 4.1.

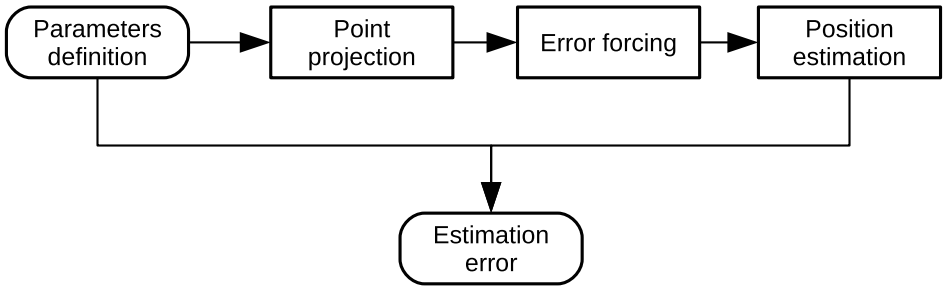


Figure 4.1: Block diagram of the developed simulation.

Starting with the parameters definition, similarly to the previous chapter's simulation, these correspond to the camera's position and orientation, the camera matrix and the fixtures coordinates. Additionally, the error intervals to be forced in each simulated parameter should also be provided. As stated previously, the parameters that will have a forced error are the image and world coordinates of each fixture. It is important to note that, for this simulation, the fixtures are considered punctual sources. Since no image generation *actually* occurs, only the fixture centers are considered.

The point projection block, similarly to the image simulation, calculates the image coordinates of each fixture's projection. Again, since we are considering the fixtures as points, only a single point per fixture is calculated, as opposed to the circular approximation previously described. It is important to note that the calculated coordinates are still rounded to the nearest integer, in order to simulate the pixelization effect. Finally, in these calculations, the camera is considered to have an infinite image plane. This does not affect the errors produced

by the pixel rounding, as the intrinsic parameters remain the same. Instead, it removes the FOV limitation, allowing us to have results in situations where, in a real scenario, the fixtures would not appear within the camera's resulting image.

The error forcing block corresponds to adding a series of values to the intended parameters, in order to force errors in them. For the detector error simulation, the operation in this block corresponds to adding a series of integers to the resulting coordinates of the previous block, which simulate a deviation, in pixels, of the image coordinates at which the fixture centers are detected. For the fixture position error, the operation in this block corresponds to adding a series of values representing a deviation of the measured world location of each fixture, in metres.

The position estimation block takes the values previously calculated (and eventually with a forced error) and passes them to the `solvePnP` function. From this, the camera position and orientation is estimated. Finally, the estimated position and the real position provided in the initial parameters are compared, resulting in the estimation error.

4.4 Simulated parameters

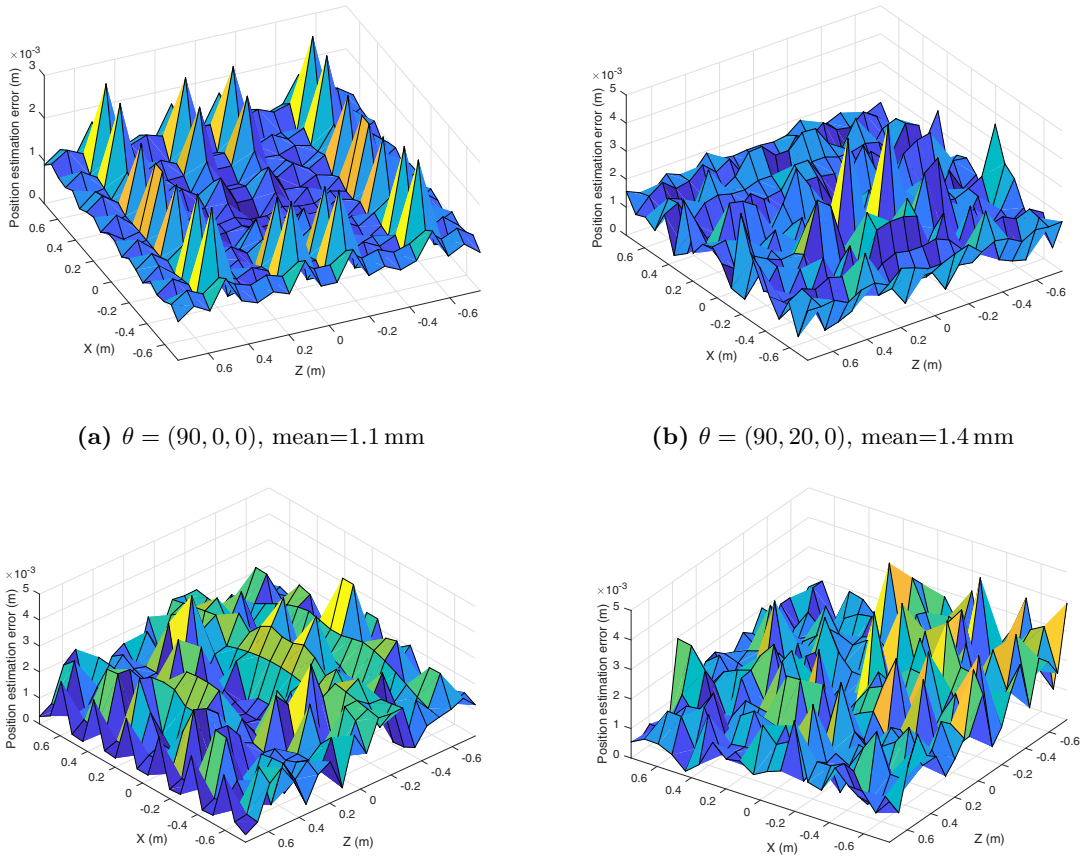
In order to evaluate the impact of the pixel rounding in the position estimation error, the first set of results will be acquired without any forced error in the two previously mentioned parameters. For the next results, as stated previously, we will simulate the position estimation impact of forcing errors in two parameters: the fixtures location in the image and the world coordinates of the fixtures. The first gives us an estimation of the impact of the accuracy of the fixture detection algorithm and the second the impact of imprecisions when measuring their location in the world.

Regarding the fixture location in the image, we will simulate an error in both the x and y axis, varying from $-20\,\text{px}$ up to $20\,\text{px}$. For the location in the world, we will simulate an error in the x and z axis, varying from $-1\,\text{m}$ up to $1\,\text{m}$.

Throughout the results, as was the case in the previous chapter, the camera parameters will remain constant and correspond to the calibrated values with the camera that is going to be used to obtain the experimental results. Furthermore, the setup and fixture placement will, again, be the same as in the previous chapter.

4.5 Results

Before presenting the results, it is important to note that all units, similarly to the previous chapter, are in meters (for distances), degrees (for angles) or pixels (for image coordinates), unless specifically mentioned otherwise. Moreover, captions mentioning *Image x coordinate* refer to *The x image coordinate of the projection of a single fixture on the image*. Similarly, *World x coordinate* refer to *The measured x world coordinate of a single fixture on the world*. Finally, the error is always forced on the fixture located at $(0.7935, -2.7, 0.5935)$, both when applied to the projection coordinates and the world coordinates.



(a) $\theta = (90, 0, 0)$, mean=1.1 mm

(b) $\theta = (90, 20, 0)$, mean=1.4 mm

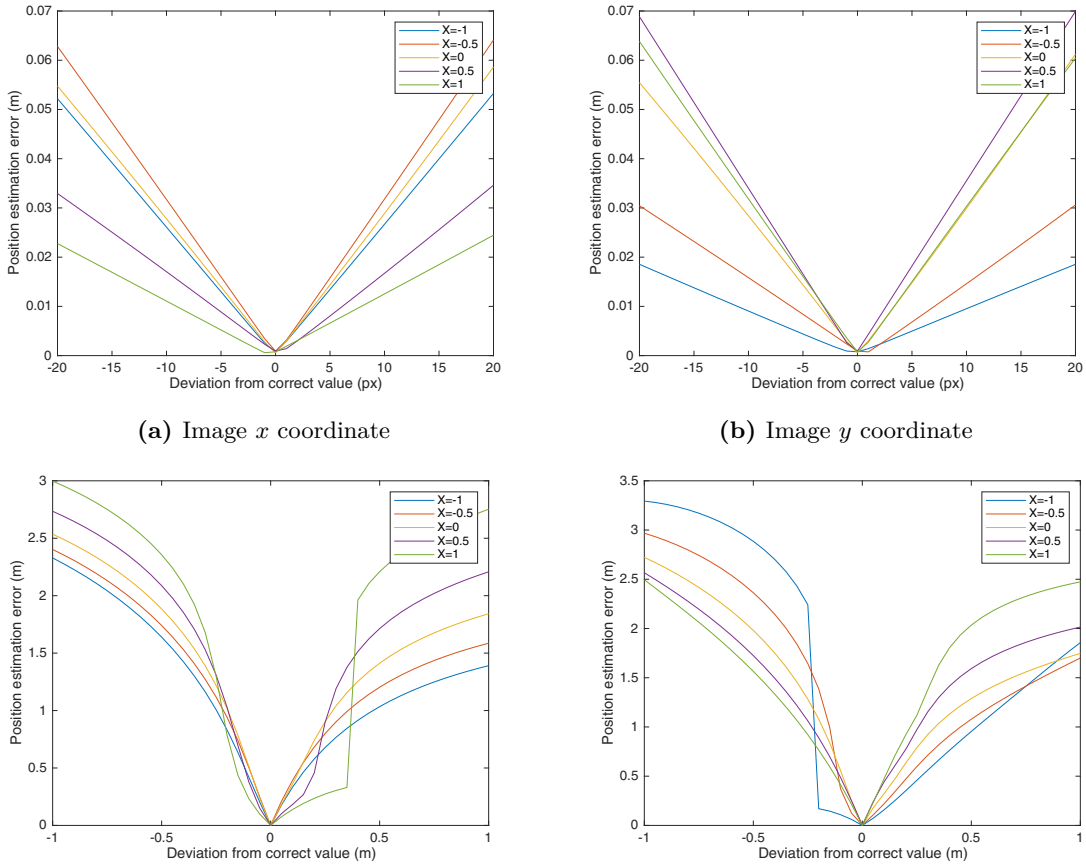
(c) $\theta = (70, 0, 0)$, mean=1.50 mm

(d) $\theta = (70, 20, 0)$, mean=1.6 mm

Figure 4.2: Position estimation error along the ground surface. The camera orientation is constant in each situation, as indicated in the corresponding caption. The obtained mean value for the error is also given.

In Fig. 4.2 the impact of the pixel rounding in the position estimation error is presented. In all these situations, the camera orientation was kept constant (facing upwards) and its position was changed along a grid on the ground surface. the only error present, besides the position estimation algorithm itself, is the pixel rounding. As such, the positioning error is quite small, being always in the order of millimetres. Considering that the real camera has a resolution of 5184 px × 3880 px, the pixel rounding error was expected to be quite low, as can be verified in the figure. Moreover, we can see that increasing the angle between the fixtures plane and the image plane increases the average error (Fig. 4.2b and 4.2c). This effect is as expected, since the increase in the angle difference also leads to perspective distortion. This effect also translates into each pixel corresponding to a larger area on the fixtures plane, effectively decreasing the spatial resolution for the same plane when captured by the camera. Finally, increasing the angle difference in both axis has an even more profound impact on the error, since the perspective distortion increases even further (Fig. 4.2d).

In the next series of results, the impact of forcing an error on the aforementioned parameters on the position estimation error is presented. In Fig. 4.3, the camera angle was kept constant,



(a) Image x coordinate

(b) Image y coordinate

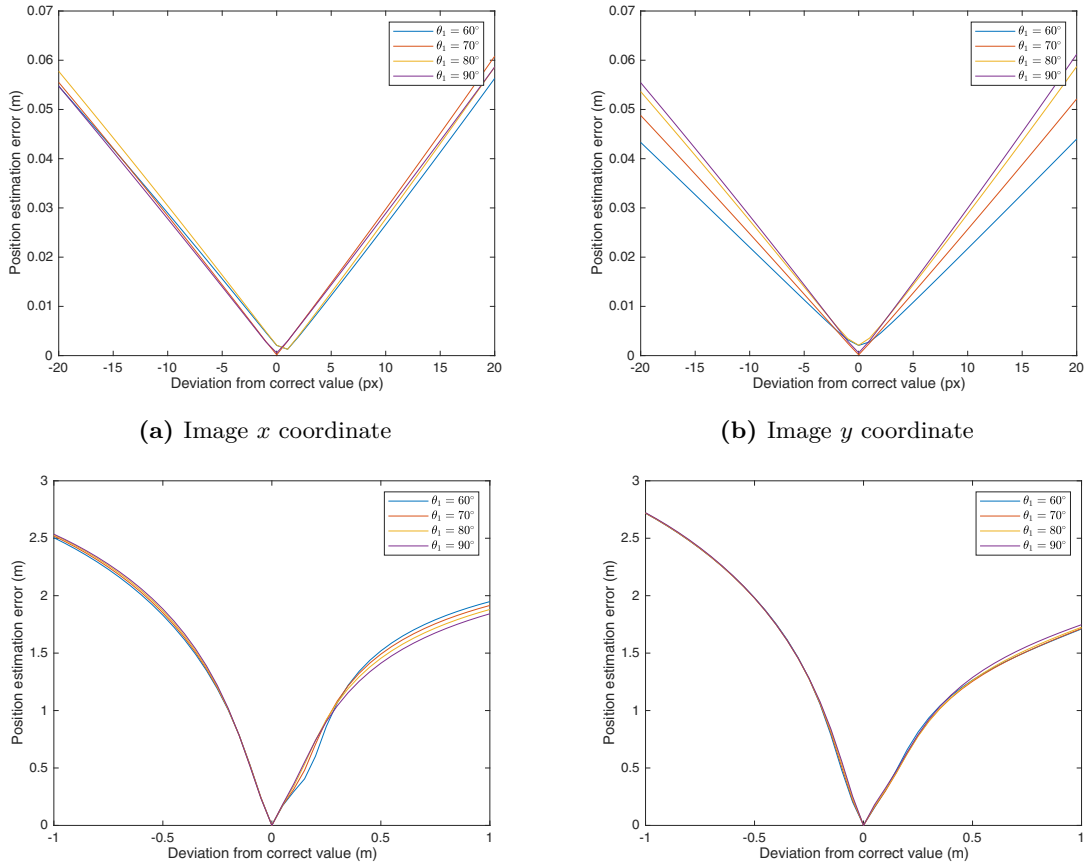
(c) World x coordinate

(d) World z coordinate

Figure 4.3: Impact of the error sources on the final position estimation, keeping the camera angle constant, with $\theta = (90, 0, 0)$. In each plot, five different camera positions along the x axis are tested, as indicated in each legend. The parameter where the error is forced is indicated in each caption.

facing upwards. In every situation, as expected, increasing the forced error on each parameter increases the final position estimation error. However, this increase is linear when forcing an error on the image coordinates and non-linear on the world coordinates. Moreover, we can observe that the camera location has no direct correlation with the final error. This makes sense when comparing with Fig. 4.2a: the error variation with the camera position follows the almost-random pattern found in the first images. Regarding the impact between both axis, the small difference is expected, considering that the fixtures are not arranged in a square, but rather a rectangle. Finally, we can observe that a small error in the fixtures world location measurement yields a large error in the final value, with a deviation of 50 cm causing a final error of more than 2 m.

In Fig. 4.4, similarly to Fig. 4.3, the impact of forcing an error on both parameters is presented. In this situation, however, the camera position is kept constant at the center of the referential and the angle is varied. Generally, the results are similar to the previous figure. Fig. 4.4a and 4.4b display a behaviour similar to the previous scenario: the orientation of the camera has no direct correlation with estimation error that arises from the error forcing



(a) Image x coordinate

(b) Image y coordinate

(c) World x coordinate

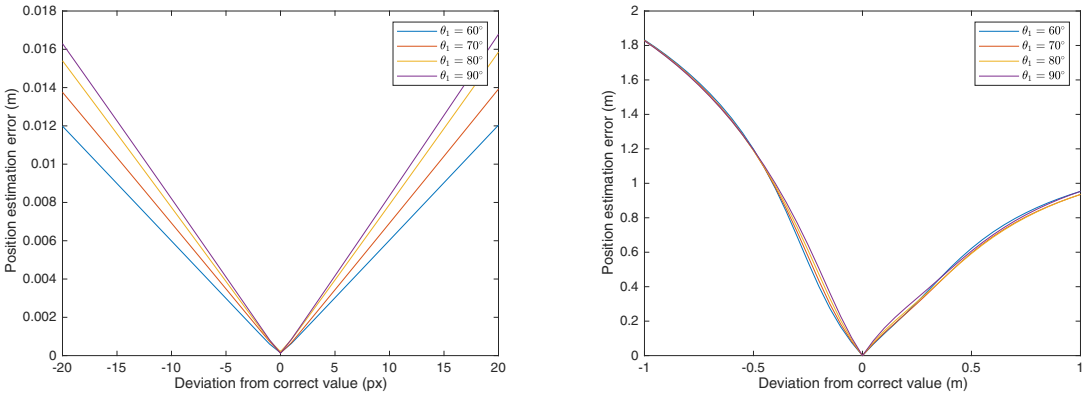
(d) World z coordinate

Figure 4.4: Impact of the error sources on the final position estimation, keeping the camera position constant, at $C = (0, 0, 0)$. In each plot, four different camera orientations around the x axis are tested, as indicated in each legend. The parameter where the error is forced is indicated in each caption.

in the image coordinate. For the forced error in the world coordinates (Fig. 4.4c and 4.4d), this phenomenon is visible even more clearly, as the lines between the multiple angles almost completely overlap. The same non-linear pattern observed in Fig. 4.3 is present and the magnitude of the final error is similar. We can also observe more clearly that the world coordinates error impact is larger when along the z axis (Fig. 4.4d) than along the x axis (Fig. 4.4d). This, again, might me due to the non-square disposition of the fixtures.

In Fig. 4.5, the same parameters are, again, tested, but with the camera placed 1m above the ground. As would be expected from intuition, the magnitude of the errors decreases in both situations. The remaining effects are extremely similar to the previous observations, with the angle variation still having no evident correlation with the error. Interestingly, the impact of the error forcing seems to vary linearly with the distance between the camera and the fixtures. For a deviation of 1 m in the measurement of the world coordinates of the fixtures, the final error is around 1 m, as opposed to around 2 m in the previous scenario.

As a final remark, we can conclude that errors in the measurement of the fixtures position in the world are the critical factor for the error in the positioning. Even though we can



(a) Image y coordinate

(b) World z coordinate

Figure 4.5: Impact of the error sources on the final position estimation, with the camera placed 1 m above the ground, at $C = (0, -1, 0)$. In each plot, four different camera orientations around the x axis are tested, as indicated in each legend. The parameter where the error is forced is indicated in each caption.

not directly compare the error in pixels with the error in meters, it can be concluded that, for commercial spaces, we can expect several centimetres of error in the measurement of the fixtures world location, whereas the detection in the image can be expected to be quite accurate. In the same manner, even with the tested deviation of 20 px in the detected image coordinate the final error still remains in the order of centimetres, which further proves the lower impact of eventual errors in the image processing.

CHAPTER 5

Complete positioning system

This chapter focuses on the complete positioning system. The developed software is presented and an explanation for each step is given. Finally, the experimental results are discussed.

5.1 System overview

The complete developed positioning system is summarised with a block diagram in Fig. 5.1. As explained throughout this dissertation, the system receives a single photograph as the input and, from there, calculates the camera position. As mentioned in Chapter 2, the camera orientation is also calculated but, since we are building a positioning system, this component is discarded.

Despite not evidenced in the diagram, similarly to the previous simulations, it is assumed to be previously known to the system:

- The camera matrix of intrinsic parameters;
- The fixtures coordinates in the world;
- The association between the ID transmitted by each fixture and the corresponding coordinates.

Each component of the system is explained in the following sections. Despite the results being presented at the end, intermediate experimental values are given to illustrate the developed algorithm.

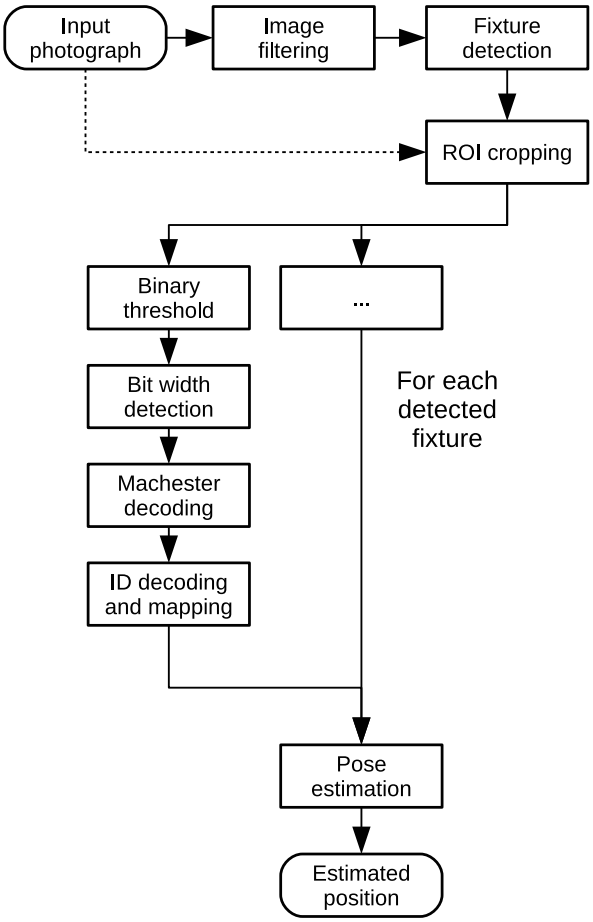


Figure 5.1: Block diagram of the complete positioning system.

5.2 Detecting the fixtures

Since we have no interest in color information, before any operation is done, the original image is converted to grayscale. As such, from now on, when referring to the *original image*, the grayscale version is considered.

The first step in the algorithm is the detection of the fixtures in the image. In order to do this, an initial filtering is applied to the image. This filtering consists in applying a binary threshold and two consecutive morphological operations: dilation and erosion. Since each fixture is captured as a series of black and white strips, the goal of this process is to eliminate those strips and transform each fixture into a filled circle.

The binary threshold is necessary because, by definition, morphological operations can only be applied to binary images. Considering that the images captured in this scenario have an almost-black background and the fixtures, as a result of the direct LOS, appear as almost pure white, a middle value of 100 was chosen for the threshold (considering 0 to be black and 255 to be white). Regarding the morphological transformations, an elliptical kernel with $16\,\text{px} \times 16\,\text{px}$ was chosen. This value was calibrated experimentally but it is important to note that, since the fixtures are typically quite spaced in the image, a larger kernel should not be a problem and, as such, the best option is to err to the large side, rather that the small.

An example result of all these operations, including the two morphological transformations, is visible in Fig. 5.2.

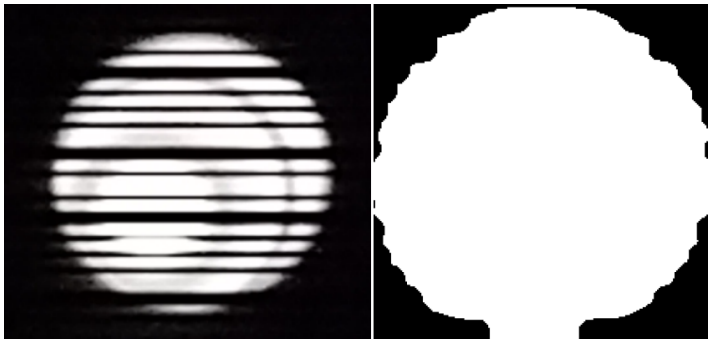


Figure 5.2: Fixture before (left) and after (right) the initial filtering.

After the filtering, a circular blob detector was employed to detect the resulting circular shapes in the image. This was done with OpenCV's `SimpleBlobDetector`. The parameters for the detector were, again, calibrated by experimentation. After applying the detector, a series of resulting region of interests (ROIs) are obtained. The original image is then cropped according to these regions, hence isolating each fixture. An example result of the fixture detection is shown in Fig. 5.3. In it we can observe the detected ROI highlighted on top of the original image, as a red circle.

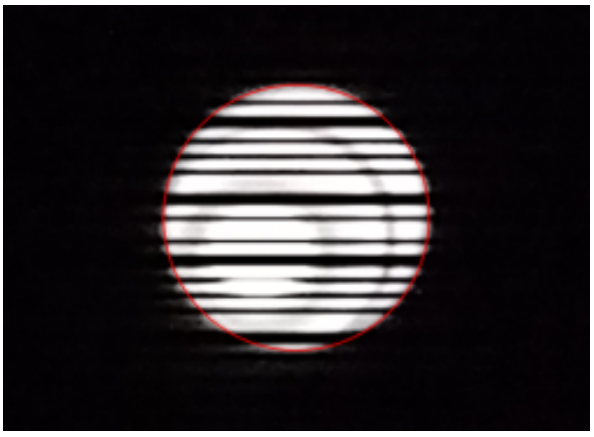


Figure 5.3: Fixture detection example.

5.3 Finding the binary threshold and bit width

After the multiple fixtures are cropped into several ROIs, each one is decoded individually. The first step in this process is, again, the binarization of the region. The binarization process described above, applied to the complete image, may not have the ideal threshold for each fixture. As such, an individual value is calculated for each region, in order to reduce, as much as possible, the number of wrongly decoded situations.

The first step in this process is to reduce each row in the region to a single value. This is done with an averaging of the grayscale values in each row. It is important to note that this

operation takes into consideration the binary masks obtained previously to distinguish, inside each row, the values belonging to the detected fixture from those belonging to the background. With these averages, an histogram of the resulting row average intensity values is calculated, as displayed in Fig. 5.4.

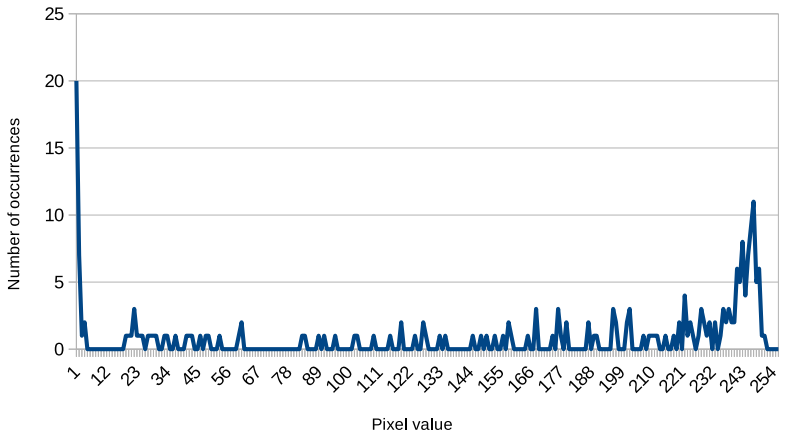


Figure 5.4: Histogram used to calculate the optimal threshold for the fixture binarization.

From this histogram, the two largest peaks are extracted, which should correspond to the average intensity value for the black and white lines. The considered threshold for the binarization is the mean value between these two peaks.

After the binarization of each region, we have to find the width of each received symbol. Before explaining this process, it is important to remember that the fixtures transmit the ID with Manchester coding, meaning that, in the resulting bitstream, we never find more that two consecutive ones or zeros. This means that we can only have two line widths: 1 bit or 2 bits. Taking this particularity into consideration, the width of each black and white line is extracted. With these values, two separate histograms are calculated, one for the black lines and another for the white ones. An example of these histograms is visible in Fig. 5.5.

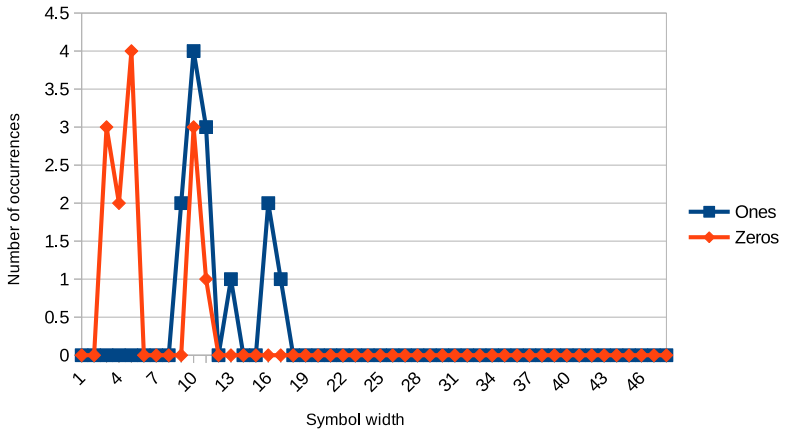


Figure 5.5: Histogram used to calculate the bit width for each fixture.

The reason for having two separate histograms for the black and white strips is because

the while lines have a bleeding effect, causing them to be larger than the black ones for a single bit. This effect, as introduced in Chapter 2, arises from the fact that, on a typical camera, the exposure time for each row in the picture slightly overlaps the adjacent ones, thus rendering the white regions larger than the black ones. In each of these histograms, on a typical scenario, we find two peaks: one from strips representing a single bit and another for strips representing two consecutive bits received. From these peaks, the mean value between them is considered as the optimal threshold for deciding if each of the lines extracted earlier represents one or two bits. After this, since we already know if the line in question corresponds to zeros or ones (given that black corresponds to zero and white corresponds to one), we can extract a bitstream from each fixture. A visual representation of an extracted bitstream is presented in Fig. 5.6.

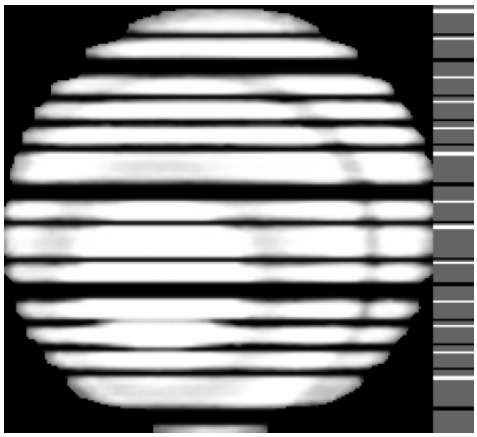


Figure 5.6: Bitstream extracted from a single fixture. The narrow lines on the right represent a single bit detected and the large lines represent two consecutive bits.

5.4 Manchester and ID decoding

After successfully decoding the bitstream in each fixture, the next step is to decode the Manchester code to obtain the original frame. As a reminder, when a message is coded using Manchester, a 0 is converted into 1-0 and a 1 into 0-1 (the reverse situation is also possible, but this was the polarity chosen in this work). As such, in the decoding process, a 0-1 transition corresponds to a 1 and a 1-0 transition to a 0. Since we have no guarantee about the phase at which we receive the message and given that each pair of received bits correspond to a single decoded bit, we have to consider the possibility of receiving the message at the middle of a transmitted symbol. In order to overcome this situation, we simply test the two situations: we start by assuming that the reception started at the beginning of a symbol; if, while decoding, we find a sequence of 1-1 or 0-0, it means that our original assumption was wrong and the process is started again, this time discarding the first received bit. If the message was received without errors, one of these situations yields a successfully decoded message. If both of them result in an error, the fixture is classified as invalid.

Given that the fixture is constantly transmitting the same ID, we must have a mechanism to identify the boundaries of the bits representing that same value. In order to do this, the transmitted ID value is packed into a frame, formed by:

- A preamble: 3 consecutive bits with values 1-1-1
- The ID: 4 bits
- A postamble: 2 consecutive bits with values 0-0

This configuration allows us to, in most situations, unequivocally identify the received ID by finding a matching frame inside the received bitstream. If the frame can not be found, the fixture is classified as invalid. Finally, the decoded ID can be associated with the corresponding world coordinates for the fixture, previously stored in a database.

5.5 Position calculation

Assuming that the previously described process resulted in a successfully decoded ID for at least four fixtures, the only remaining step it to pass the correspondences between the image and world coordinates for each fixture to the `solvePnP` function, as described in the previous chapter. Again, the camera intrinsic parameters is assumed to be previously known and already stored, since it only has to be calibrated once.

5.6 Experimental setup

A photograph and description of the experimental setup is visible in Fig. 5.7. It consists of four modulated LED fixtures, placed at a distance of 2.7 m from the floor, with a nominal value of 18 W each. Their centers are located at (0.7935, -2.7, 0.5935), (0.7935, -2.7, -0.5935), (-0.7935, -2.7, -0.5935) and (-0.7935, -2.7, 0.5935), all in metres. The origin of the world referential was placed at the geometrical center of the rectangle formed by the four fixtures. For the acquired results, the fixtures were modulated using OOK with a duty-cycle of 50%. The implementation details for the fixture driver circuits, as highlighted in the introduction, are out of the scope of this work, since this setup is the result of another dissertation, conducted in parallel with this one.

The camera used to acquire the images was the frontal camera included in the smartphone Xiaomi Mi A2. A diagram evidencing the coordinate system for the smartphone camera is present in Fig. 5.8. The maximum resolution of the camera was used, with an image size of 5184 px × 3880 px. The camera ISO setting was kept at 1600 and the exposure was reduced, in each situation, to the lowest possible value. The camera intrinsic parameters were obtained using the calibration program included in the OpenCV library [42]. The resulting camera matrix of intrinsic parameters was:

$$M = \begin{bmatrix} 3912 & 0 & 0 & 2592 \\ 0 & 3912 & 0 & 1940 \\ 0 & 0 & 1 & 0 \end{bmatrix} \tag{5.1}$$

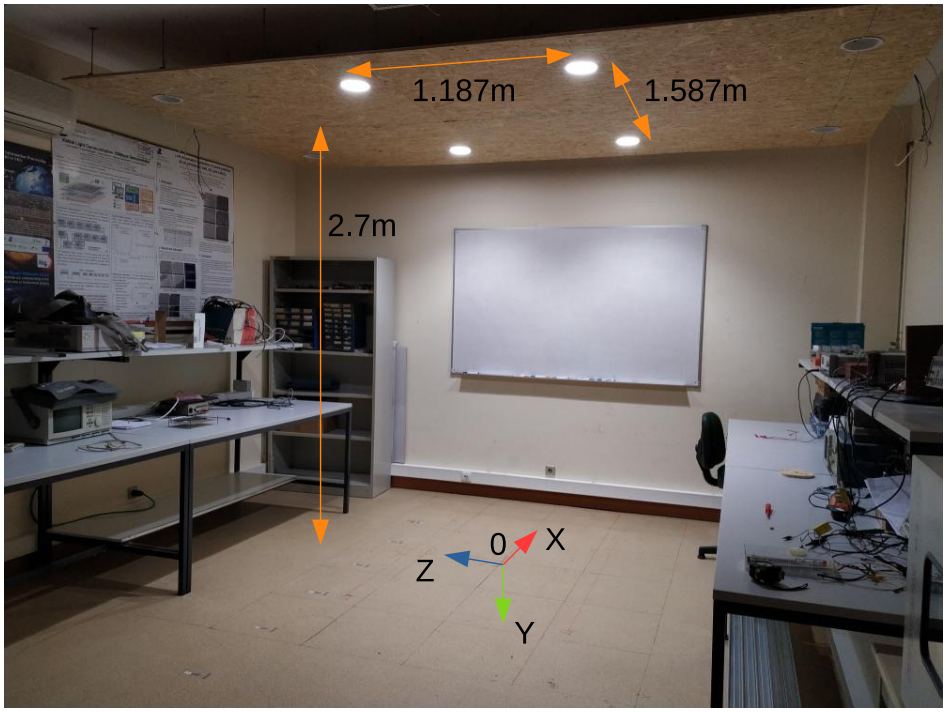


Figure 5.7: Experimental setup used to test the system.

Figure 5.8: Coordinate system for the smartphone camera.

Finally, the fixtures were modulated at a frequency of 10 kHz. This value was obtained through experimentation and was the highest frequency at which the messages were still received correctly in most pictures taken.

5.7 Positioning error results

The complete positioning system was tested using the described setup and smartphone camera, by testing it with photographs taken along a grid placed on the floor surface. The camera orientation was kept at $\theta = (90, 0, 0)°$. The obtained positioning error is presented in Fig. 5.9. One of the photographs taken for a single position is visible in Fig. 5.10.

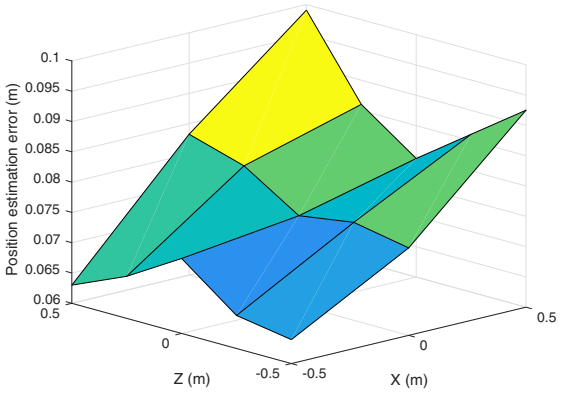


Figure 5.9: Experimental results for the position error. The camera was kept at a constant angle, with $\theta = (90, 0, 0)°$.

From these results, we can observe that the error remains under 10 cm, which is considered a good accuracy for the system in question. Moreover, the average error obtained was 7.8 cm which, again, corresponds to a relatively high positioning accuracy.

In Fig. 5.11 we can observe the difference between the camera location points and the estimated positions. As can be observed, the estimated positions were always placed at around 6 cm from the floor level. Furthermore, the configuration of the estimated positions seems considerably regular, as opposed to random, indicating a systematic error source. This might suggest errors in the measurements of the setup, particularly in the placement of the grid on the floor. The difference in the y axis, however, might suggest an error in the measurement of the fixtures position in the world.

As a final remark, if we consider that the positioning error originates only as a result of errors in the measurements of the fixtures world coordinates, the 8 cm average error would correspond to roughly a 2 cm error in those measurements, according to the plots obtained in the previous chapter. This error magnitude in the fixture world coordinates would be perfectly acceptable for a system with the described dimensions. If such assumption is correct, it further highlights the fact that the accuracy of the system is mainly limited by the same measurement, as previously speculated in the last chapter.

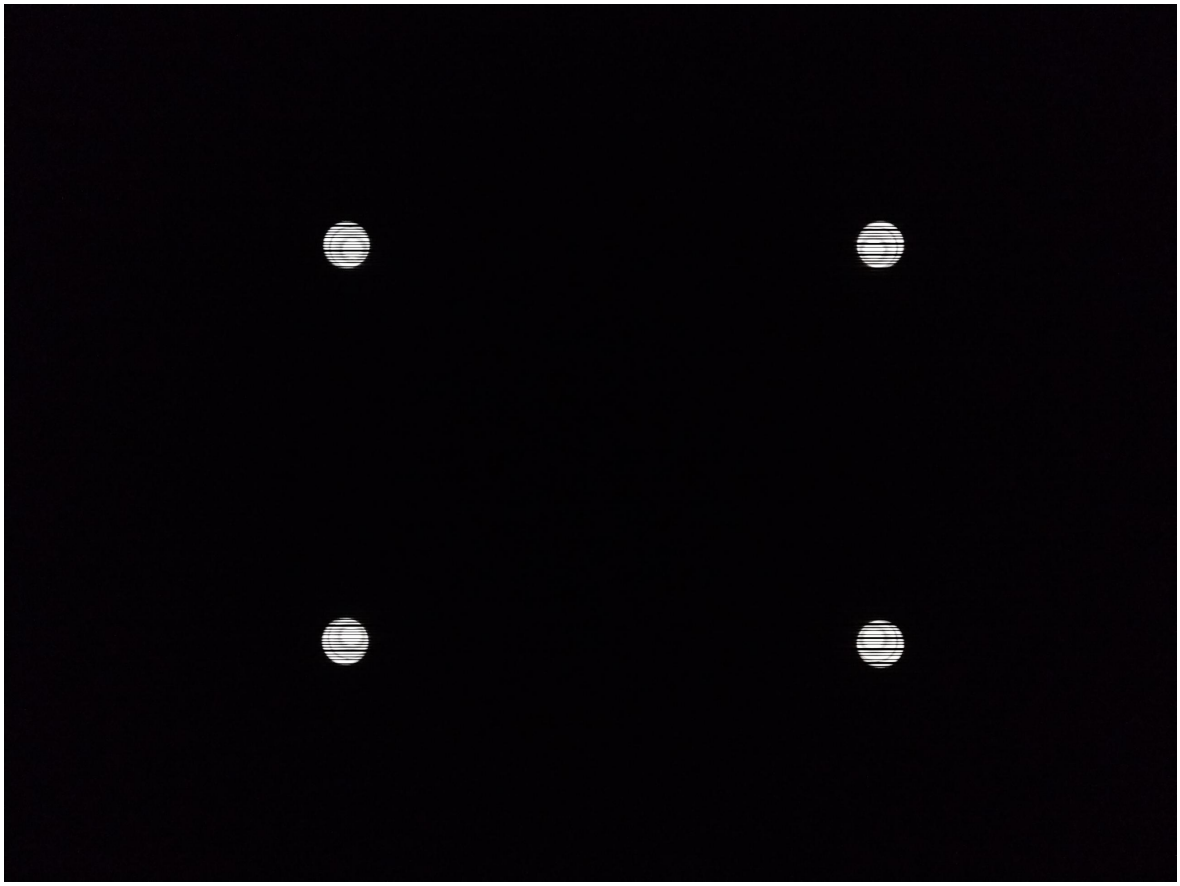
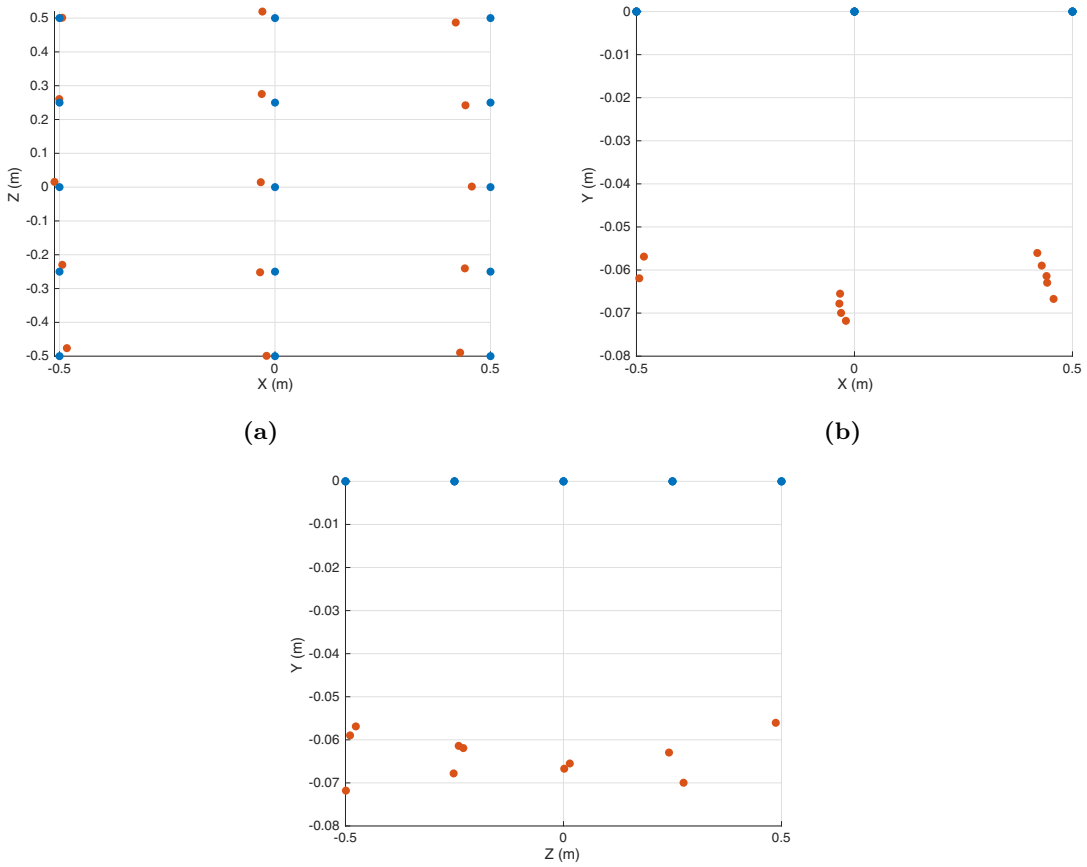


Figure 5.10: Sample photograph acquired in the practical tests. The camera was placed at the origin of the world referential, with $\theta = (90, 0, 0)^\circ$.



(a) **(b)** **(c)**

Figure 5.11: Difference between the real (blue) and estimated (orange) camera positions, as seen from top and both sides.

CHAPTER 6

Conclusions

This chapter presents final remarks on the work developed. The results obtained are reflected upon and possible future work is discussed.

6.1 Results discussion

This dissertation started with an introduction to the general problem of indoor positioning. After that, an introductory view of camera-based systems for this purpose was given, in Chapter 2. In it, the concept of using a camera for such a task was explained, with general remarks about its particularities.

Going further, Chapter 3 presented the mathematical tools involved in the process, detailing the camera geometrical model and the expressions associated with it. As a result, a simulation was developed with the objective of testing and employing those same expressions in order to generate simulated photographs, given the scene parameters — the camera position and orientation and the multiple fixtures position in the world. The photos presented there were used to anticipate the camera's field of view and, therefore, infer the possible test regions for the experimental results. Furthermore, the impact of changing the camera's position and orientation on the final acquired picture was analysed, anticipating, again, the experimental scenarios.

In Chapter 4, the impact of introducing error in the system was studied. As such, another simulation was developed, joining the expressions presented in Chapter 3 with the chosen position estimation algorithm. Through the simulation, the final position estimation accuracy was studied, by considering the pixel rounding effect and forcing error intervals on two parameters — the known world location of the multiple fixtures and the image coordinates at which those same fixtures are detected. The results produced by this simulation allowed us to analyse the impact of error in these same parameters, providing a general idea of the main causes for deviations in the final position estimation, on a real system. Finally, the magnitude of the error introduced by the multiple parameters was also inferred, providing further insights regarding the expected final accuracy observed in the experimental results.

In Chapter 5, the complete system for performing indoor positioning with a camera was presented. Its purpose is to calculate the camera position from a single input photograph, provided that a minimum of four fixtures are visible in the image. Furthermore, the fixture identification was done by modulating an unique ID through OOK and Manchester for each one. Similarly, the developed system also included steps for performing the ID extraction. The multiple system blocks were explained, with experimental results illustrating each step. Finally, experimental results were gathered and analysed.

The implemented positioning system yielded valid and reasonably accurate results, with an average positioning error of 7.8 cm. This value was considered quite satisfactory, particularly given the dimensions of the practical setup and the typical required accuracy for an IPS. Moreover, through the comparison of these results with the ones from Chapter 4, the sources of error in the experimental results were reflected upon. Through this process, the obtained accuracy was considered reasonable and justified for the scenario in question, specially when compared with the predicted positioning errors in the aforementioned results.

6.2 Future work

The work contained in this dissertation, by itself, proved the feasibility of a complete camera-based indoor positioning system, including the usage of existing LEDs for serving both the illumination and ID transmission purposes, the unconstrained camera coordinates and the usage of a typical smartphone camera as the receiver. Regardless of this, a number of possible future improvements and extensions for this work were identified.

The first one is performing tests in larger spaces. Despite this dissertation having results for a room-sized system, the acquisition area was still quite small, mainly given the fact that only four light fixtures were available, severely limiting the testing surface. Furthermore, testing different fixture placement configurations would also be interesting, particularly given the variety of light fixture placement possibilities found in most commercial spaces.

Regarding the algorithm, despite being functional, not all acquired photographs were successfully interpreted, due to errors in the decoding of the ID for each fixture. Despite limitations imposed by the used camera, it was still visible in the images that the implemented algorithm has margin for improvement, which would provide a more robust decoding process.

Another interesting future development would be to implement this system directly on a smartphone, through a custom-made application. Besides rendering the results acquisition a more straightforward process, variables such as processing time could be measured. The possibility of performing real-time tests is also appealing.

Finally, the usage of a custom-made application would also provide the possibility of incorporating the smartphone's IMU into the algorithm, eventually increasing its accuracy through sensor fusion techniques. Moreover, with such an improvement, the system could also be improved to be able to estimate a position even when less than four light fixtures are visible in the photograph. Such addition would increase the versatility and the number of possible test scenarios for this type of system.

References

[1] N. UL Hassan, A. Naeem, M. Pasha, T. Jadoon, and C. Yuen, "Indoor Positioning Using Visible LED Lights: A Survey", *ACM Computing Surveys*, vol. 48, Nov. 2015. DOI: 10.1145/0000000.000000.

[2] *MATLAB - MathWorks*, en. [Online]. Available: https://www.mathworks.com/products/matlab.html (visited on 12/13/2019).

[3] *OpenCV*. [Online]. Available: https://opencv.org/ (visited on 12/13/2019).

[4] R. F. Brena, J. P. García-Vázquez, C. E. Galván-Tejada, D. Muñoz-Rodriguez, C. Vargas-Rosales, and J. Fangmeyer, *Evolution of Indoor Positioning Technologies: A Survey*, en, Research article, 2017. DOI: 10.1155/2017/2630413.

[5] Y. Zhuang, L. Hua, L. Qi, J. Yang, P. Cao, Y. Cao, Y. Wu, J. Thompson, and H. Haas, "A Survey of Positioning Systems Using Visible LED Lights", *IEEE Communications Surveys Tutorials*, vol. 20, no. 3, pp. 1963–1988, 2018, ISSN: 1553-877X, 2373-745X. DOI: 10.1109/COMST.2018.2806558.

[6] P. Magnan, "Detection of visible photons in CCD and CMOS: A comparative view", en, *Nuclear Instruments and Methods in Physics Research Section A: Accelerators, Spectrometers, Detectors and Associated Equipment*, Proceedings of the 3rd International Conference on New Developments in Photodetection, vol. 504, no. 1, pp. 199–212, May 2003, ISSN: 0168-9002. DOI: 10.1016/S0168-9002(03)00792-7.

[7] A. Luštica, "CCD and CMOS image sensors in new HD cameras", in *Proceedings ELMAR-2011*, ISSN: 1334-2630, Sep. 2011, pp. 133–136.

[8] T. Hirayama, "The evolution of CMOS image sensors", in *2013 IEEE Asian Solid-State Circuits Conference (A-SSCC)*, Nov. 2013, pp. 5–8. DOI: 10.1109/ASSCC.2013.6690968.

[9] Z. Wang, Q. Wang, W. Huang, and Z. Xu, "Optical Camera Communication: Fundamentals", in *Visible Light Communications: Modulation and Signal Processing*, IEEE, 2018, pp. 239–290. DOI: 10.1002/9781119331865.ch8.

[10] C. Danakis, M. Afgani, G. Povey, I. Underwood, and H. Haas, "Using a CMOS camera sensor for visible light communication", in *2012 IEEE Globecom Workshops*, ISSN: 2166-0077, Dec. 2012, pp. 1244–1248. DOI: 10.1109/GLOCOMW.2012.6477759.

[11] D. Litwiller, "CCD vs. CMOS: Facts and Fiction", 2001.

[12] C. S. Herrmann, "Human EEG responses to 1–100 Hz flicker: Resonance phenomena in visual cortex and their potential correlation to cognitive phenomena", en, *Experimental Brain Research*, vol. 137, no. 3, pp. 346–353, Apr. 2001, ISSN: 1432-1106. DOI: 10.1007/s002210100682.

[13] Z. Ghassemlooy, L. N. Alves, S. Zvanovec, and M.-A. Khalighi, *Visible light communications: theory and applications*, English. 2017, OCLC: 990156401, ISBN: 978-1-4987-6754-5 978-1-4987-6753-8 978-1-315-36733-0. [Online]. Available: https://doi.org/10.1201/9781315367330 (visited on 11/25/2019).

[14] N. T. Le, M. Hossain, and Y. M. Jang, "A survey of design and implementation for optical camera communication", en, *Signal Processing: Image Communication*, vol. 53, pp. 95–109, Apr. 2017, ISSN: 0923-5965. DOI: 10.1016/j.image.2017.02.001.

[15] N. Saeed, S. Guo, K.-H. Park, T. Y. Al-Naffouri, and M.-S. Alouini, "Optical camera communications: Survey, use cases, challenges, and future trends", en, *Physical Communication*, vol. 37, p. 100 900, Dec. 2019, ISSN: 1874-4907. DOI: 10.1016/j.phycom.2019.100900.

[16] Z. Wang, Q. Wang, W. Huang, and Z. Xu, "Optical Camera Communication: Modulation and System Design", in *Visible Light Communications: Modulation and Signal Processing*, IEEE, 2018. DOI: 10.1002/9781119331865.ch9.

[17] Z. Ong and W.-Y. Chung, "Long Range VLC Temperature Monitoring System Using CMOS of Mobile Device Camera", *IEEE Sensors Journal*, vol. 16, no. 6, pp. 1508–1509, Mar. 2016, ISSN: 1530-437X, 1558-1748, 2379-9153. DOI: 10.1109/JSEN.2015.2506907.

[18] L. Song, P. Luo, M. Zhang, Z. Ghassemlooy, D. Han, and H. L. Minh, "Undersampled Digital PAM Subcarrier Modulation for Optical Camera Communications", EN, in *Asia Communications and Photonics Conference 2015 (2015), paper AM1E.2*, Optical Society of America, Nov. 2015, AM1E.2. DOI: 10.1364/ACPC.2015.AM1E.2.

[19] P. Luo, Z. Ghassemlooy, H. Le Minh, X. Tang, and H.-M. Tsai, "Undersampled phase shift ON-OFF keying for camera communication", in *2014 Sixth International Conference on Wireless Communications and Signal Processing (WCSP)*, Oct. 2014, pp. 1–6. DOI: 10.1109/WCSP.2014.6992043.

[20] J. Hao, Y. Yang, and J. Luo, "CeilingCast: Energy efficient and location-bound broadcast through LED-camera communication", in *IEEE INFOCOM 2016 - The 35th Annual IEEE International Conference on Computer Communications*, ISSN: null, Apr. 2016, pp. 1–9. DOI: 10.1109/INFOCOM.2016.7524511.

[21] Y. Imai, T. Ebihara, K. Mizutani, and N. Wakatsuki, "Performance evaluation of high-speed visible light communication combining low-speed image sensor and polygon mirror in an outdoor environment", in *2016 Eighth International Conference on Ubiquitous and Future Networks (ICUFN)*, ISSN: 2165-8536, Jul. 2016, pp. 51–55. DOI: 10.1109/ICUFN.2016.7536978.

[22] M. A. Fischler and R. C. Bolles, "Random Sample Consensus: A Paradigm for Model Fitting with Applications to Image Analysis and Automated Cartography", *Commun. ACM*, vol. 24, no. 6, pp. 381–395, Jun. 1981, ISSN: 0001-0782. DOI: 10.1145/358669.358692.

[23] N. Rajagopal, P. Lazik, and A. Rowe, "Visual light landmarks for mobile devices", in *IPSN-14 Proceedings of the 13th International Symposium on Information Processing in Sensor Networks*, Apr. 2014, pp. 249–260. DOI: 10.1109/IPSN.2014.6846757.

[24] S. Hyun, Y. Lee, J. Lee, M. Ju, and Y. Park, "Indoor positioning using optical camera communication pedestrian dead reckoning", in *2015 Seventh International Conference on Ubiquitous and Future Networks*, ISSN: 2165-8528, 2165-8536, Jul. 2015, pp. 64–65. DOI: 10.1109/ICUFN.2015.7182499.

[25] Y. Nakazawa, H. Makino, K. Nishimori, D. Wakatsuki, and H. Komagata, "Indoor positioning using a high-speed, fish-eye lens-equipped camera in Visible Light Communication", in *International Conference on Indoor Positioning and Indoor Navigation*, Oct. 2013, pp. 1–8. DOI: 10.1109/IPIN.2013.6817855.

[26] Y.-S. Kuo, P. Pannuto, K.-J. Hsiao, and P. Dutta, "Luxapose: Indoor positioning with mobile phones and visible light", *Proceedings of the Annual International Conference on Mobile Computing and Networking, MOBICOM*, Sep. 2014, ISSN: 978-1-4503-2783-1. DOI: 10.1145/2639108.2639109.

[27] B. Lin, Z. Ghassemlooy, C. Lin, X. Tang, Y. Li, and S. Zhang, "An Indoor Visible Light Positioning System Based on Optical Camera Communications", *IEEE Photonics Technology Letters*, vol. 29, no. 7, pp. 579–582, Apr. 2017, ISSN: 1041-1135, 1941-0174. DOI: 10.1109/LPT.2017.2669079.

[28] B. Zhu, J. Cheng, J. Yan, J. Wang, and Y. Wang, "VLC Positioning Using Cameras with Unknown Tilting Angles", in *GLOBECOM 2017 - 2017 IEEE Global Communications Conference*, Dec. 2017, pp. 1–6. DOI: 10.1109/GLOCOM.2017.8254786.

[29] Y. Han, Q. Cheng, and P. Liu, "Indoor positioning based on LED-camera communication", in *2016 IEEE International Conference on Consumer Electronics-China (ICCE-China)*, Dec. 2016, pp. 1–4. DOI: 10.1109/ICCE-China.2016.7849732.

[30] H.-S. Kim, D.-R. Kim, S.-H. Yang, Y.-H. Son, and S.-K. Han, "An Indoor Visible Light Communication Positioning System Using a RF Carrier Allocation Technique", *Journal of Lightwave Technology*, vol. 31, no. 1, pp. 134–144, Jan. 2013, ISSN: 0733-8724, 1558-2213. DOI: 10.1109/JLT.2012.2225826.

[31] Z. Yang, Z. WANG, J. Zhang, C. Huang, and Q. Zhang, "Polarization-Based Visible Light Positioning", *IEEE Transactions on Mobile Computing*, vol. 18, no. 3, pp. 715–727, Mar. 2019, ISSN: 1536-1233, 1558-0660, 2161-9875. DOI: 10.1109/TMC.2018.2838150.

[32] M. Yasir, S.-W. Ho, and B. N. Vellambi, "Indoor Positioning System Using Visible Light and Accelerometer", *Journal of Lightwave Technology*, vol. 32, no. 19, pp. 3306–3316, Oct. 2014, ISSN: 0733-8724, 1558-2213. DOI: 10.1109/JLT.2014.2344772.

[33] A. Jovicic, "Qualcomm® Lumicast™: A high accuracy indoor positioning system based on visible light communication", en, p. 12,

[34] R. Zhang, W.-D. Zhong, K. Qian, and D. Wu, "Image Sensor Based Visible Light Positioning System With Improved Positioning Algorithm", *IEEE Access*, vol. 5, pp. 6087–6094, 2017, ISSN: 2169-3536. DOI: 10.1109/ACCESS.2017.2693299.

[35] M.-g. Moon, S.-i. Choi, J. Park, and J. Kim, "Indoor Positioning System using LED Lights and a Dual Image Sensor", *Journal of the Optical Society of Korea*, vol. 19, pp. 586–591, Dec. 2015. DOI: 10.3807/JOSK.2015.19.6.586.

[36] R. Zhang, W.-D. Zhong, D. Wu, and K. Qian, "A Novel Sensor Fusion Based Indoor Visible Light Positioning System", in *2016 IEEE Globecom Workshops (GC Wkshps)*, Dec. 2016, pp. 1–6. DOI: 10.1109/GLOCOMW.2016.7848823.

[37] B. Zhu, J. Cheng, Y. Wang, J. Yan, and J. Wang, "Three-Dimensional VLC Positioning Based on Angle Difference of Arrival With Arbitrary Tilting Angle of Receiver", *IEEE Journal on Selected Areas in Communications*, vol. 36, no. 1, pp. 8–22, Jan. 2018, ISSN: 0733-8716, 1558-0008. DOI: 10.1109/JSAC.2017.2774435.

[38] H. Huang, L. Feng, G. Ni, and A. Yang, "Indoor imaging visible light positioning with sampled sparse light source and mobile device", *Chinese Optics Letters*, vol. 14, Sep. 2016. DOI: 10.3788/COL201614.090602.

[39] *OpenCV: Camera calibration With OpenCV*. [Online]. Available: https://docs.opencv.org/master/d4/d94/tutorial_camera_calibration.html (visited on 12/05/2019).

[40] *3drotations*. [Online]. Available: http://motion.pratt.duke.edu/RoboticSystems/3DRotations.html (visited on 12/06/2019).

[41] *Camera Calibration and 3d Reconstruction — OpenCV 2.4.13.7 documentation*. [Online]. Available: https://docs.opencv.org/2.4/modules/calib3d/doc/camera_calibration_and_3d_reconstruction.html (visited on 12/13/2019).

[42] *OpenCV: Camera calibration With OpenCV*. [Online]. Available: https://docs.opencv.org/3.4/d4/d94/tutorial_camera_calibration.html (visited on 12/13/2019).