Pervasive Reflective Sensing with Visible light and Beyond

Wang, Weizheng

DOI
10.4233/uuid:46ee909c-6323-426a-a173-e2940801b771

Publication date
2024

Document Version
Final published version

Citation (APA)
Wang, W. (2024). *Pervasive Reflective Sensing with Visible light and Beyond*. [Dissertation (TU Delft), Delft University of Technology]. https://doi.org/10.4233/uuid:46ee909c-6323-426a-a173-e2940801b771

PERVASIVE REFLECTIVE SENSING WITH VISIBLE LIGHT AND BEYOND

PERVASIVE REFLECTIVE SENSING WITH VISIBLE LIGHT AND BEYOND

Dissertation

for the purpose of obtaining the degree of doctor
at Delft University of Technology,
by the authority of the Rector Magnificus Prof. dr. ir. T.H.J.J. van der Hagen,
chair of the board for Doctorates,
to be defended publicly on
Wednesday 1 May 2024 at 10:00 o'clock

by

Weizheng WANG

Master of Science in Radio Physics, Wuhan University of Technology, China,
born in Ezhou, China.

This dissertation has been approved by promotors.

promotor: Prof. dr. K.G. Langendoen
copromotor: Dr. M.A. Zúñiga

Composition of the doctoral committee:

Rector Magnificus,	Chairman
Prof. dr. K.G. Langendoen,	Delft University of Technology, promotor
Dr. M.A. Zúñiga,	Delft University of Technology
Independent members:	
Prof. dr. A. Bozzon,	Delft University of Technology
Prof. dr. B. C. Renner,	Hamburg University of Technology, Germany
Prof. dr. A. Gomez,	Technical University of Braunschweig, Germany
Prof. dr. N. Meratnia,	Eindhoven University of Technology
Prof. dr. ir. G. J. P. M. Houben,	Delft University of Technology, reserve member
Other member:	
Dr. Q. Wang,	Delft University of Technology

Prof. dr. K. G. Langendoen heeft in belangrijke mate aan de totstandkoming van het proefschrift bijgedragen.

Keywords: pervasive, reflective sensing, repurposing

Printed by: UIN printing

Front & Back: Eric Wang

ISBN 000-00-0000-000-0

An electronic version of this dissertation is available at
`http://repository.tudelft.nl/`.

To my parents for their endless support without any reservation.

CONTENTS

SUMMARY

Modern societies depend on critical infrastructures, like electricity, illumination, and communication networks. During the last decade, the deployment of billions of Internet-of-Thing (IoT) devices has led to the establishment of a new *sensing* infrastructure. This pervasive system presents advantages and challenges. On the one hand, the massive deployment can capture unprecedented levels of monitoring data, from small-scale smart homes to large-scale smart cities. On the other hand, the infrastructure consumes an enormous amount of resources, not only in terms of energy demands and ecological footprint, but also in terms of management and economic costs.

To optimize resource utilization, researchers are exploring a new sensing paradigm that trades off sensor quality for cost. Instead of adding high-end sensors for new applications, new systems are either *(i) using low-cost sensors* with lower signal quality, or *(ii) re-purposing sensors* to cover applications that they were *not* originally designed for. For example, nowadays, WiFi devices not only transmit data but also use their antennas as sensors to recognize gestures performed by nearby users.

A key underlying principle used with low-cost and re-purposed sensors is *reflective sensing*. Within this principle, an emitter –like a WiFi station or a light bulb– radiates waves that are reflected by the users or objects in the surroundings. The reflected waves are measured by sensors co-located with the emitters to obtain information about the environment. The advantage of this approach is that it is not invasive, the object or subject of interest does not need to carry any special electronic device.

The advantages of reflective sensing have made it a popular, but highly heterogeneous, research area in two main aspects. First, applications. Reflective sensing covers topics that range from gesture recognition to health monitoring. Second, spectra. The (reflected) signals used in those applications exploit a wide range of the electromagnetic spectrum, from microwave to visible light. Each application has its own unique set of challenges, but we are interested in identifying general macro-challenges.

Given the extensive variety of sensors, methods, and topics used in reflective sensing, this thesis focuses on the following research question: *How can we improve the performance of low-cost and re-purposed sensors to approach that of high-end sensors?* The key challenge is that using low-end sensors exposes a large 'sensing gap' because the quality of the received signal can decrease dramatically.

To tackle this challenge, we consider three different applications –crowd monitoring, biometric authentication, and indoor positioning– relying on three different types of reflected waves –microwave, infrared, and visible light. Based on these studies, we identify a framework that has two parts. The first part measures the sensing gap between low-end and high-end sensors. The second part bridges the sensing gap with various alternatives aimed at enhancing the signal quality. Considering that framework, the thesis provides four contributions.

The first contribution, in Chapter 2, *measures the sensing gap* between mmWave (radar) sensors and cameras for crowd monitoring. Due to privacy regulations, some areas are restricting the use of cameras for crowd monitoring. mmWave sensors are a promising low-cost alternative based on reflective sensing, but the gathered data is less precise. After performing a thorough outdoor evaluation over three months, we show that mmWave sensors provide an error rate from 15% to 40% higher than camera-based alternatives, indicating that further research is needed to balance cost and accuracy.

The second contribution, in Chapter 3, covers a different application to show that *enhancing the data processing methodology* is a powerful mechanism to bridge the sensing gap. Recent studies demonstrate that cardiac signals can be used to identify people, but its application requires purposed-design sensors. We propose a methodology aimed at achieving cardiac-based authentication with low-cost and re-purposed sensors. These low-end devices emit light that is reflected by the fingertips of users. The light reflections capture unique blood flow patterns that allow identifying the subject.

The third contribution, in Chapter 4, covers the same application as the prior chapter (cardiac-based authentication), but uses a complementary approach to bridge the sensing gap. Instead of enhancing only the methodology, it also *enhances the re-purposed sensor*. The focus of this study is to re-purpose smartphone cameras as cardiac sensors. To achieve that goal, we propose optimizing the camera parameters and the pressure of the user's fingertip to gather signals that are as close as possible to the ones obtained with cardiac sensors.

The fourth contribution, in Chapter 5, proposes a different application to highlight another alternative to reduce the sensing gap: *enhancing the object*. Considering an indoor positioning system with visible light, we modify slightly the surface of a reflective object to obtain simultaneously location and identification. Our evaluation shows, in a scaled-down scenario, that an object can be tracked with cm-level accuracy using only light reflections, offering similar performance to systems that require placing an active receiver (including a photosensor) on the object.

Overall, the diverse range of topics covered – from crowd monitoring to biometric authentication and indoor localization – underscores the breadth of challenges in pervasive reflective sensing. A major takeaway from this thesis is the need for design alternatives that match the performance of high-end sensors with low-end and repurposed ones. This approach is vital to expand IoT capabilities without significantly increasing its economic and management costs.

1

INTRODUCTION

1

People have an inherent desire to deepen their understanding of the world, a pursuit that requires the collection of extensive data. For most of our history, the data gathered by our senses was sufficient to satisfy our needs. However, the continuous expansion of human knowledge requires tackling problems of a scale and complexity that surpasses the capabilities of our senses. For instance, to understand climate change, scientists need information from a large number of weather stations providing heterogeneous weather data, such as temperature, wind speed, and precipitation.

To tackle modern sensing challenges, researchers have proposed the concept of the Internet of Things (IoT), which aims to operate a diverse array of sensors synergistically. The advancements in IoT technologies have interconnected today's world at an unprecedented scale. IoT sensors, serving as the extension of human senses, are deeply embedded in various facets of our daily lives, from temperature measurements in smart homes [1] to traffic sensing in smart cities [2]. More recently, some researchers are even pushing the boundaries of IoT to space [3].

The proliferation of sensors in both scale and variety owes largely to an important paradigm shift in sensing, moving from invasive to non-invasive methods. Invasive methods, as the name suggests, involve processes that could modify or damage the object under examination. Figures 1.1a and 1.1b depict two scenarios showcasing invasive sensing: brix evaluation[1] and hemoglobin concentration tests. In brix evaluation, workers need to *open the fruit* and squeeze its juice on a refractometer to evaluate its sugar level [4]. In hemoglobin concentration tests, doctors need to *pierce the fingertip* and collect blood [5]. On the other hand, non-invasive methods can perform the same tasks without interfering much with the object of interest. Figure 1.1c shows how researchers in the company fresco exploit mmWave radio signals to evaluate the fruit brix level *without opening it* [6]. Figure 1.1d shows that a hemoglobin test can be achieved *by merely placing a fingertip on top of a smartphone camera* [7].

1.1. NON-INVASIVE SENSING

Non-invasive sensing is emerging as a widely adopted method in various industries. While numerous types of non-invasive sensing techniques exist, in this thesis, we categorize them into two groups: radiative sensing and reflective sensing, as depicted in Figure 1.2.

1.1.1. RADIATIVE SENSING

The concept of radiative sensing is shown in Figure 1.2a, where sensors gather information about the object based on the signals radiated by the object itself. Given that many objects around us emit energy, radiative sensing is a popular approach. A commonly used spectrum is infrared (thermal), which cannot only detect people and animals, but can also detect internal voids, delaminations, and cracks in concrete structures [8].

Radiative sensing, however, has inherent limitations because it cannot control the strength of the signal emitted by the object. A low Signal-to-Noise Ratio (SNR) may require sophisticated implementations, leading to high-cost sensors.

[1] Brix evaluation is used to measure the sugar level in fruits.

Figure 1.1: The change of sensing paradigm from invasive to non-invasive. The top row displays invasive solutions. (a) Shows a brix evaluation that requires *opening the fruit* to extract its juice [4]. (b) Shows a test that requires *piercing the fingertip* to extract blood [5]. The bottom row shows alternative non-invasive solutions. (c) Shows a brix evaluation that relies on mmWave signals and *does not require opening the fruit* [6]. (d) Shows a hemoglobin concentration test that simply requires placing a fingertip *over a smartphone camera* [7].

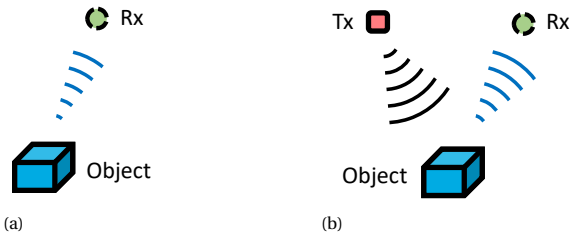


Figure 1.2: Two categories of non-invasive sensing: (a) radiative sensing and (b) reflective sensing.

1.1.2. REFLECTIVE SENSING

To overcome the limitations caused by low SNR signals radiated by the object, researchers exploit external emitters to boost the energy level. Figure 1.2b demonstrates the basic concept. Instead of relying solely on the object's energy, an emitter or transmitter sends a signal that is backscattered by the object. Based on the reflected signal, the sensor can obtain various types of information. Reflective sensing [9, 10] can be divided into two categories.

First category: setups using **passive emitters**. In this type of scenario, the emitter is outside the control of the intended application. The most representative applications are those related to photo-sensing. For example, sunlight reflections can be exploited by

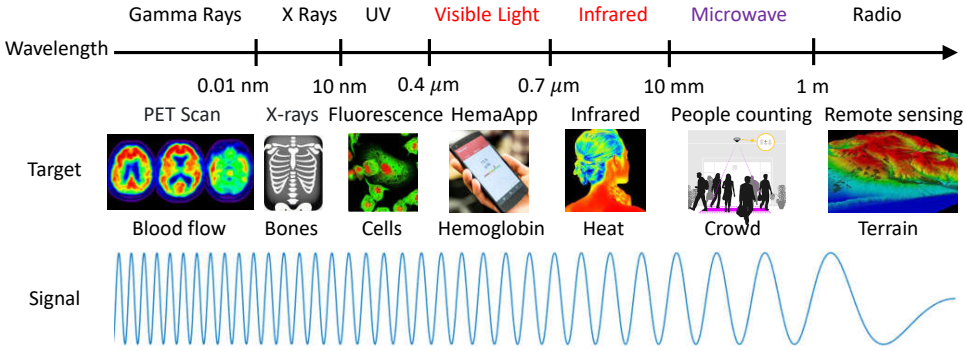


Figure 1.3: Different wavelengths correspond to different types of monitoring targets [23].

cameras to monitor the structural health of a base station. By measuring the changes in light intensity in pixels, researchers can quantify the level of structural vibration [11]. Another study exploits the daylight reflected from vehicles' surfaces to identify the type of vehicle; by placing a simple photodiode in an outdoor parking lot and it shows that car reflections can be unique enough to be recognized [12]. Even though an ambient emitter offers higher power than the object's naturally inherent radiation, this sensing approach is dependent on the presence of the ambient source. When the emitter is absent (dark environments) or weak (low light environments), reflective sensing also becomes vulnerable to low-quality signals.

Second category: setups using **active emitters**. In this type of scenario, the emitter is under the control of the intended application. A popular approach in this domain is related to radar systems, where low-cost modules are being used to identify gestures [13, 14, 15], body structures [16, 17], and cardiovascular parameters [18, 19, 20]. Active emitters add more overhead because they have to be installed, but they allow controlling the signal strength. The latter is a significant advantage because it reduces the complexity of the receiver. No sophisticated boosting modules or signal processing methods are needed at the receiver, resulting in lower-cost sensors while maintaining accuracy. Additionally, unlike systems using *passive* emitters, which rely solely on the default spectrum from the ambient source, *active* emitters can harness spectra from low-frequency audio [21] to high-frequency mmWave [22].

1.2. WAVELENGTH & PERVASIVE CONSIDERATIONS

In reflective sensing, the wavelength is critical since it determines which type of target can be detected. Figure 1.3 shows the mapping from different wavelengths to their corresponding popular targets. To describe the properties of the different signals, we first divide them into two groups. The first group includes Gamma rays, X-rays, and UV; and the second group includes Visible light, Infrared, Microwave, and Radio.

The first group has the advantage of providing fine-granular sensing, but they pose several disadvantages as well. First, they can increase the risk of cancer under long exposure [24, 25, 26]. Second, they have a short sensing range. Third, sensors with short wavelengths are expensive, for example, an entry-level X-ray machine costs more than

40k euros. Due to these disadvantages, this type of signal is not used **pervasively**.

The second group of signals does not have harmful effects on human health. Thus, they have become central to two main pillars of modern societies, illumination and communication. Among these signals, the Radio and Microwave spectra (up to 5 GHz) have been explored widely for reflective sensing. By analyzing the reflections of the surrounding waves, radio tomography can detect the location of people without requiring users to wear any special device [27, 28, 29]. Further research demonstrates that WiFi signals can be used to detect over-the-air commands [30, 31, 32, 33, 34] and monitor the activity of people in a privacy-preserving manner [35, 36, 37] (without cameras).

The above advances in reflective radio sensing are notable, but the Radio spectrum is getting crowded and costly. Due to this reason, researchers are also exploring the Visible Light, Infrared, and Microwave (above 5 GHz) spectra for reflective sensing. The focus of this thesis is on these three spectra since they provide several advantages for ***pervasive sensing***. First, a large part of that spectra is safe, open, free, and has relatively fine granularity. Second, the emitters and receivers are either low-cost, like in the case of mmWave radar; or they are already widely deployed, like in the case of illumination (LEDs, photodiodes, and cameras).

1.3. RESEARCH PROBLEM

Considering that the IoT is giving rise to new sensing infrastructures with stringent requirements, we are posed with new trade-offs. On one hand, there is a need to develop systems that have a low cost, complexity, and ecological footprint, while being privacy-aware. On the other hand, despite these constraints, there is a need to still maintain a high accuracy.

Since deploying a large number of precise but expensive sensors is not an option, the research community is investigating alternatives that require either deploying low-cost sensors or re-purposing existing sensors for other applications. For example, instead of using cameras for indoor monitoring, which infringes privacy, a new generation of low-cost mmWave radars are used for that purpose [38, 39]; and instead of deploying hospitals in remote areas, researchers are re-purposing smartphone cameras to perform health checks such as blood pressure [40]. The main challenge solved by these studies is achieving good accuracy with low-cost or re-purposed sensors. This thesis follows that same line of research: *expanding the pervasiveness of reflective sensing systems in the Visible Light, Infrared, and Microwave spectra.* To contribute to tackling that challenge, we need to answer the following research question:

> ***What design alternatives are available to approach the performance of high-end sensors with either low-cost or re-purposed sensors?***

This thesis argues that the above research question can be explored through different options. In Figure 1.4, we propose a framework to classify those design options. The figure captures the gap in sensing applications between what is desired (application data) and what is available (object). In conventional approaches, the gap is mainly filled by high-end sensors purposely designed for the required task and some methods on top of

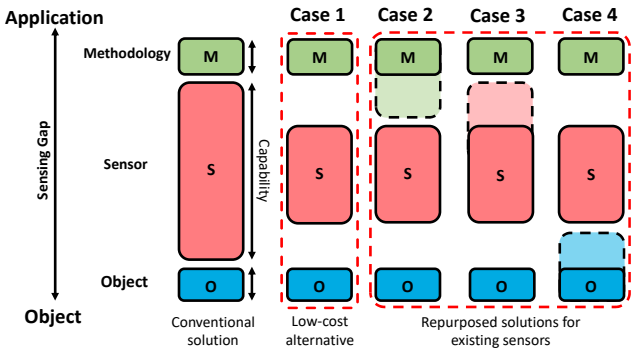


Figure 1.4: Different approaches to fill up the sensing gap. Conventional solutions (leftmost plot), fill up the sensing gap with precise sensors. When low-cost or re-purposed sensors are used, the sensing gap increases (case 1), and it can be filled up by enhancing the methodology (case 2), sensor (case 3), or object (case 4). The vertical length of each block indicates its capability. The shaded color indicates our **main** improvements.

the sensor to perform data processing. If, on the other hand, the system relies on low-cost or re-purposed sensors, the sensing gap is exposed.

The sensing gap can be investigated at two levels. The *first level* is to solely quantify the gap, without trying to bridge it (case 1). From a research perspective, such an approach allows exposing the magnitude of the problem raised by using low-power or re-purposed sensors. The *second level* is to bridge the gap (cases 2, 3, and 4). For this approach, we identify three alternatives.

The first alternative is to *enhance the methodology* (case 2). For scenarios where the sensor is re-purposed, an enhanced methodology could overcome the low quality of the received signals. Advanced signal processing and machine learning techniques are particularly valuable for this alternative.

The second alternative is to *enhance the re-purposed sensor* (case 3). The best example of this alternative is the use of smartphones for health-related applications. Several studies show that optimizing different camera parameters or performing minor physical modifications to the microphone can allow checking auditory or cardiovascular issues [41, 42].

The third alternative is to *enhance the object* (case 4). Given that reflective sensing is fundamentally determined by the properties of the object's external surface, for some applications, it may be possible to perform minor modifications to the object's surface to facilitate sensing.

1.4. THESIS CONTRIBUTIONS & USE CASES

Given the breadth of research on reflective sensing, not only based on the different spectra but also based on the wide range of applications, this thesis considers four case studies. These studies rely on sensors using the Microwave, infrared, and Visible Light spectra, and cover different types of applications, from people and vehicle monitoring to biometric authentication. The details of each case study, their respective sub-research questions, and the contributions of each chapter of this thesis are presented next.

1.4.1. CHAPTER 2 - MEASURING THE SENSING GAP

AN APPLICATION OF PEOPLE MONITORING WITH MICROWAVE SENSORS

This chapter investigates the feasibility of using low-cost microwave sensors instead of precise optical sensors. The application of interest is outdoor people counting, which is a key application for smart cities [43]. Currently, reliable solutions for people counting depend on camera-based techniques [44, 45]. However, cameras violate new GDPR regulations and their field-of-view is easily blocked by obstacles.

New mmWave sensors overcome some of these problems. These sensors, operating in the 60-64 GHz spectrum, use signal reflection for object detection, like any other radar system. Their wavelength is sufficiently broad to prevent facial recognition, aligning with GDPR requirements, while still being effective for crowd monitoring. Moreover, mmWave sensors are less expensive and more energy-efficient than camera systems. For instance, the IWR6843 mmWave sensor, priced at around 200 euros, is cheaper than the 800-euro XOVIS PC2SE Outdoor face-blurring camera.

A shortcoming of mmWave sensors is that they have been mainly evaluated in *indoor* environments [46, 47]. This chapter investigates the performance of mmWave sensors in *realistic outdoor* scenarios. In our setup, we deploy two nodes on our university campus with a weather-proof casing. The setup tracked the flow of pedestrians and cyclists over a period of three months considering different types of weather conditions and environmental dynamics. The study in this chapter allows us to analyze the shortcomings of mmWave radar with the following research question:

> ***Case 1:*** *What is the sensing gap between mmWave sensors (low-end sensors) and cameras (high-end sensors) for people counting in dynamic outdoor environments?*

Contribution: Our contributions towards analyzing the sensing gap of mmWave sensors are threefold. First, we perform a careful set of experiments to design an appropriate radome (radar cover), considering important parameters such as radome material width and sensor placement. Through tests, we show that our radome outperforms a commercial alternative in terms of coverage. Second, we deploy mmWave sensors at a site frequented by cyclists and pedestrians, collecting data for three months in varying weather conditions (sunny, foggy, rainy, and windy days). Third, through thorough evaluations, our analysis indicates that low-cost radar sensors provide low correlation coefficients for people counting, ranging from 60% to 85%, lagging behind the less than 10% error rate of commercial people-counting cameras. Overall, the sensing gap exposes significant room for improvement in outdoor scenarios, be it enhancing the methodology, sensor or object.

1.4.2. CHAPTER 3 - ENHANCING THE METHODOLOGY

AN APPLICATION OF CARDIAC IDENTIFICATION WITH INFRARED SENSORS

In this chapter, we present a general method for authentication with cardiac signals. Current security methods –such as fingerprint, face, and iris recognition– rely on *external* biometric patterns that can be obtained by attackers since they are constantly exposed. To pursue a higher security level, researchers are shifting their focus to unique internal

features like cardiac patterns [48, 49]. These patterns reside beneath the skin, thereby rendering it more challenging for attackers to acquire them. The primary technique for capturing these patterns is photoplethysmography (PPG), which uses reflective sensing principles to measure blood flow changes and capture cardiac patterns [48, 50].

The performance of cardiac authentication depends on the quality of the sensor. Instead of performing authentication with advanced cardiac equipment [51, 52], our framework relies on cardiac signals gathered with *low-cost infrared* sensors or by *re-purposing* smartphone cameras using the *visible light* spectrum. By placing the fingertip on top of a low-cost PPG sensor or a smartphone camera, a unique user pattern can be identified to unlock a system.

The challenge is that, in realistic scenarios, the quality of cardiac signals obtained with low-cost or re-purposed sensors is low, preventing systems from obtaining stable and distinct features. To bridge the performance gap, one alternative is to design a more advanced method, which leads to the following research question.

> ***Case 2:*** *Can a general methodology be designed to exploit low-quality cardiac signals for authentication?*

Contribution: Our research addresses the sensing gap of low-end cardiac identification with three major contributions. First, we introduce an adaptive filtering technique to obtain stable and distinct features between subjects. Second, instead of assuming a single cardiac morphology per user, we show that users have multiple morphologies, enhancing the system's real-time response and user inclusion. Third, we propose a multi-cluster approach with Mahalanobis distance measurements to counter non-linear effects. Our study considers low-cost and re-purposed sensors (pulse oximeter and cameras), including both healthy and unhealthy individuals. The results indicate that the improvements in our methods reduce the sensing gap between 10% and 15% in terms of identification accuracy.

1.4.3. CHAPTER 4 - ENHANCING THE RE-PURPOSED SENSOR

AN APPLICATION OF CARDIAC IDENTIFICATION WITH SMARTPHONE CAMERAS

This chapter builds upon the previous one. In the prior chapter, we bridge the sensing gap by focusing solely on the methodology (case 2 in Figure 1.4). In this chapter, we experiment with enhancing the sensor itself (case 3 in Figure 1.4).

Considering that cameras are not designed to capture cardiac patterns, camera signals are of lower quality than those collected from a purposedly-designed PPG sensor. The design differences reduce the authentication accuracy. An enhanced methodology can overcome some issues related to low-quality signals, as done in the prior chapter, but it is also valuable if a re-purposed sensor can be improved to get signals that are as similar as possible to the original sensor.

The difference in performance between PPG sensors and smartphones is due to the different emitters, receivers, and spectra used by both devices. As an emitter, the PPG sensor uses an LED that radiates infrared light, while the smartphone uses a flashlight that emits (white) visible light. Infrared light is better suited to track changes in blood

flow. As a receiver, the PPG sensor uses a photodiode tuned for the IR spectrum, while the smartphone uses a camera tuned for pictures and videos. Considering these design differences, we pose the following research question.

> ***Case 3:*** *Despite the different hardware and spectra, can the parameters of smartphone cameras be fine-tuned to resemble a PPG sensor?*

Contribution: We propose a framework that makes smartphone cameras behave as close as possible to PPG sensors. Our solution bridges the sensing gap by enhancing the sensor, with three main contributions. First, we analyze the available camera parameters and optimize their configuration to obtain stable cardiac signals with a high SNR. Second, we investigate different fingertip pressure levels, from no pressure to excessive pressure, and identify the best pressure for each subject during the PPG collection process. Third, we evaluate our camera-based PPG authentication with a comprehensive dataset including 58 subjects. The results show that our system can bridge the sensing gap, achieving comparable authentication performance to PPG sensors.

1.4.4. CHAPTER 5 - ENHANCING THE OBJECT

AN APPLICATION OF INDOOR LOCALIZATION WITH VISIBLE LIGHT

This chapter introduces a method to bridge the sensing gap by enhancing the object's surface (case 4 in Figure 1.4). The application used for this case study is indoor localization, which is widely studied. Among the various indoor positioning approaches, systems using the lighting infrastructure are gaining significant attention. In these systems, light bulbs act as anchors, modulating their intensity to send location information to nearby users or objects. The strength of visible light waves is that, contrary to radio waves, they do not suffer from severe multi-path effects, enabling centimeter-level precision with simple light bulbs [53, 54].

Visible light positioning systems provide high accuracy, but they share a common constraint: the object of interest needs to carry an optical receiver, a processor with a camera or a photodiode. This constraint introduces practical challenges in terms of installation, battery maintenance, and practical operation because the sensor must be always placed on the external surface of the object.

To extend the benefits of localization with visible light, it would be important to analyze if accurate positioning can be attained passively. That is, instead of placing the photosensors on the object, simple photodiodes can be placed on the ceiling –together with the light bulb– to measure light reflections. These reflections can then be processed to provide indoor positioning. Targeting this passive localization scenario leads to the following research question.

> ***Case 4:*** *Can the external surface of some objects be enhanced to attain accurate localization with light reflections?*

Contribution: Our research addresses the sensing gap of passive localization using light with three main contributions. First, we eliminate the active sensor on an object by re-

purposing the object's surface with barcode-like ID patterns. Exploiting reflection geometry, our system can achieve positioning and identification simultaneously in a passive manner. Second, we modify (re-purpose) light bulbs to provide multi-beam coverage to create fine-grained reflection signals. Third, we evaluate our system in a scaled-down scenario, and the results show that the ID and location of an object can be obtained with cm-level accuracy (barring blind spots).

Chapters 2, 3, 4 and 5 are presented in the following papers:

- **Weizheng Wang**, Girish Vaidya, Anup Bhattacharjee, Francesco Fioranelli, Marco Zuniga (2023). A Long-Term Study of mmWave Sensing in an Outdoor Urban Scenario. *In Proceedings of the 19th Annual International Conference on Distributed Computing in Smart Systems and the Internet of Things (DCOSS-IoT).*

- **Weizheng Wang**, Qing Wang, Marco Zuniga (2023). Taming Irregular Cardiac Signals for Biometric Identification. *ACM Transactions on Sensor Networks (TOSN).*

- **Weizheng Wang**, Qing Wang, Marco Zuniga (2022). CardioID: Mitigating the Effects of Irregular Cardiac Signals for Biometric Identification (Best Paper Runner-up). *In Proceedings of the International Conference on Embedded Wireless Systems and Networks (EWSN).*

- **Weizheng Wang**, Marek Vette, Qing Wang, Jie Yang, Marco Zuniga (2022). CamPressID: Optimizing Camera Configuration and Finger Pressure for Biometric Authentication. *In Proceedings of the IEEE International Conference on Mobile Ad Hoc and Sensor Systems (MASS).*

- **Weizheng Wang**, Qing Wang, Junwei Zhang, Marco Zuniga (2020). PassiveVLP: Leveraging Smart Lights for Passive Positioning. *ACM Transactions on Internet of Things (TIOT).*

- **Weizheng Wang**, Qing Wang, Junwei Zhang, Marco Zuniga (2018). Leveraging smart lights for passive localization. *In Proceedings of the IEEE International Conference on Mobile Ad Hoc and Sensor Systems (MASS).*

2

A Long-Term Study of mmWave Sensing in an Outdoor Urban Scenario

The first step in bridging the sensing gap between low-cost and high-end sensors is to measure the difference in performance. In applications related to crowd monitoring, cameras can count the number of people precisely, but a clear image violates privacy regulations. Millimeter Wave (mmWave) sensors, on the other hand, only capture a vague representation of people due to their longer wavelength, complying with privacy regulations. In this chapter, we take mmWave sensors out of indoor scenarios, where they are usually evaluated, and test them thoroughly in an outdoor setup. Our work measures the sensing gap between cameras and mmWave sensors for people counting over three months considering different weather conditions. The results showcase a large sensing gap. Low-cost radar sensors provide low correlation coefficients for people counting, between 60% to 85%, lagging behind commercial people-counting cameras with over 90% accuracy.

2.1. INTRODUCTION

The modernization of the urban infrastructure promises efficient and safe delivery of services in public places. Many of these services require sensing people's flow. For instance, pedestrian monitoring can avoid unsafe crowding situations. Similarly, through active crowd management at tourist destinations, it is possible for civic authorities to intervene in a timely manner [55].

There are two popular mechanisms for sensing people-traffic, viz., device-based and vision-based. The device-based mechanisms count the number of devices, such as mobile phones, to estimate the traffic in a particular area [56]. The main limitation of this method is that it assumes that every individual carries one device. On the other hand, the affordability of cameras and the simultaneous advancement in vision processing algorithms have made vision-based systems a popular choice for monitoring people-movement. The city of Amsterdam alone has more than 1000 registered cameras deployed for crowd sensing [57].

While the deployment of sensors is necessary for monitoring the crowd flow, the increasing number of cameras in public spaces can be problematic. Cities such as London, with more than six hundred thousand surveillance cameras, have triggered concerns amongst privacy advocates [58]. To circumvent this concern, a few privacy-friendly cameras are commercially available. Some cameras blur the faces, while others can be configured not to share any images and instead pass only the people count for further analysis. However, such software mechanisms may still be prone to hacking. More importantly, there is the issue of *perceived privacy*, which cause many citizens to have strong reservations even with privacy-aware cameras [59].

2.1.1. CHALLENGES IN MMWAVE SENSING

Millimeter wave (mmWave) radar is emerging as a promising alternative for privacy-preserving crowd monitoring. mmWave radar detects people and objects as point clouds, making it difficult to collect personal information. The mmWave radars are available as compact single-chip solutions, and this improved availability and affordability has encouraged their wider acceptance and deployment.

While recent works have explored the possibility of mmWave radar deployment for sensing people's flow [60, 61, 62], they have several limitations. First, most deployments have been done under indoor conditions. For urban applications, the mmWave sensor must be deployed outdoors. This implies that the radar must be enclosed in a radome[1]. The design of the radome –which comprises its material, thickness, and placement of radar– significantly influences the performance of the system. Secondly, the studies have been conducted for a short duration, thus not capturing the impact of sustained operations under varying weather conditions. Thirdly, most of the current literature captures the performance of the radar under a controlled flow of people. Through this work, we design a suitable radome and investigate the performance of mmWave sensing with a flow of real-world pedestrians and bikers in the license-free band (60 GHz).

[1] A radome is a structure to protect radar equipment, and it has to be made from material that is transparent to radio waves

2.1.2. OUR CONTRIBUTION

To overcome the above limitations, we perform a thorough evaluation of mmWave radar using the IWR6843ISK platform [63]. In particular, our contribution is three-fold.

- *A careful radome design [Section 2.3].* We perform a careful set of experiments to design an appropriate radome. Our design considers important parameters such as the width of the material and the exact location of the radar sensor inside the radome. We compare the performance of our prototype with a commercial alternative and show that we perform better in terms of coverage.
- *A real-world deployment [Section 2.4].* We deploy two mmWave radar sensors to monitor real-world people's movement and compare their outputs with those from commercial camera-based sensors. The location is regularly used by cyclists and pedestrians. We gather per-minute data over a period of three months. The time frame considers sunny, foggy, rainy, and windy days.
- *A detailed evaluation [Sections 2.5 and 2.6].* We utilize precision, recall, and correlation metrics with different levels of granularity to quantify the performance of mmWave sensing. Our results show that, compared to the high accuracy of indoor scenarios, radar systems outdoors still need to improve. Detecting empty spaces has an accuracy of 85% and estimating the flow of people has correlation coefficients between 60% and 85%, depending on the weather condition. We believe that, compared to the error rate provided by people-counting cameras (below 10%), the current performance of radars still needs to improve.

2.2. BACKGROUND

2.2.1. PROPERTIES OF MMWAVE SENSING

Frequency Modulated Continuous Wave (FMCW) radars are a special class of mmWave radars, that are commercially available in a miniaturized form-factor for wide-scale deployment. FMCW radars have a carefully designed array of transmitter and receiver antennas. The radar sends chirp signals, which are reflected by the environment and processed by the system to provide point clouds for all the *moving* objects, as shown in Figure 2.1. The point clouds provide angular information of the objects, as well as the range and velocity [64].

The fact that people, or any dynamic element, are represented as points is the key strength of mmWave sensors to maintain privacy. In the GHz bands, the wavelengths are in the order of a few millimeters, which are too long to capture the fine granularity present in human faces. Thus, even if the radar platform is hacked to obtain the raw signals, the attacker will be fundamentally limited regarding the depth of information that can be attained.

To capture mobile elements as cloud points, the mmWave radar has algorithms that remove *static* clutter [65]. Since indoor scenarios offer a *stable* static background, mobile elements reflect clear signals. Furthermore, indoor evaluations do not require adding a case, and hence, do not need to consider the attenuation and distortion that is inherent to mechanical enclosures. These two properties, static background and no need for casings, allow mmWave studies to deliver indoor applications that not only provide

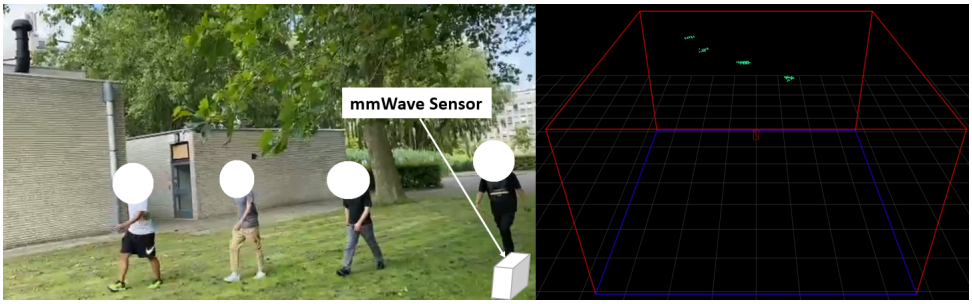


Figure 2.1: Point Cloud Tracking.

tracking [66] but also 3D skeleton reconstruction [67] and vital sign monitoring [68]. Outdoors, the radars need casing and the background is not static. Animals can be under the field of view and weather conditions can affect the quality of the reflected signal (rain or fog). Furthermore, wind also has a pernicious effect because it blows leaves, plants, grass and other objects that appear in the point cloud measurements. It is, hence, more challenging to perform mmWave sensing in outdoor scenarios.

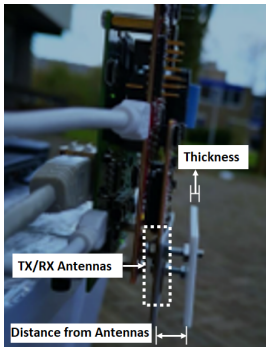
2.2.2. FREQUENCY RANGES FOR MMWAVE SENSING

Another element that is often overlooked in mmWave studies is the operational band. mmWave radars are available in three bands: 24GHz, 60GHz and 77GHz. The 77GHz band is generally reserved for automotive applications. Initially, we wanted to use a sensor on this band (to complement the traffic information gathered by cars), but we consulted a municipality that did not approve of its use. The 24GHz band has two main components, an Ultra-Wide Band (UWB) and a NarrowBand (NB). Spectrum regulations and standards developed by the European Telecommunications Standards Institute (ETSI) and Federal Communications Commission (FCC) prohibit new industrial products from using the 24GHz UWB. The NB offers a limited bandwidth of 250MHz, adversely affecting people-detection in this spectrum. On the contrary, the 60GHz sensor operates in the free ISM band and offers a wide bandwidth of 4GHz, providing richer point clouds and thus better object identification and tracking [69]. In our deployment, we use the IWR6843ISK platform from Texas Instruments (TI) in the 60GHz band.

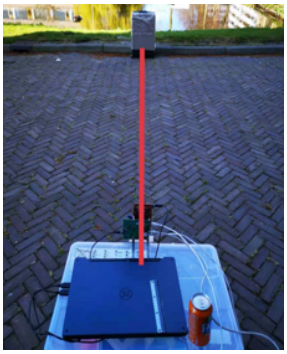
2.3. RADOME DESIGN

An important consideration for any outdoor radar deployment is the design of the casing to protect the hardware from weather-related phenomena. This enclosure is called the radome and it needs to be transparent to radio waves.

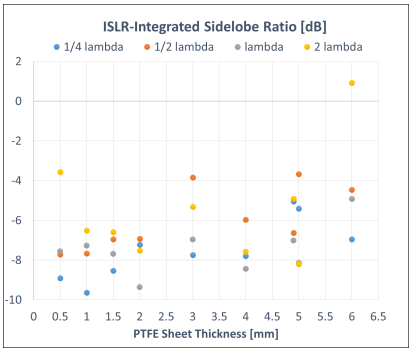
We considered using a commercial product that already has a radome but each node costs 600 $ and the node is designed for indoor scenarios [70]. Given that the price of the mmWave chip is 120 $, significant savings can be achieved if we design our own casing. Later in this section, we compare our design with the commercial product and show that our prototype not only saves costs but also attains a better performance.



(a) Key parameters for a radome's design



(b) Experimental setup for radome design, the radar has a front aluminum plate



(c) Results of the tests with PTFE (Teflon) sheets of different thicknesses and distances from the radar antennas

Figure 2.2: The experiments for radome design.

2.3.1. PARAMETER ANALYSIS

To design a radome, one needs to take into account two main aspects. The first one is simple: To consider the mechanical and electronic requirements for placing the radar chip, the microcontroller, and the power circuitry. The second aspect is more complex: To analyze electromagnetic effects in order to minimize distortions to the emitted and received signals.

A thorough design of a radome requires full electromagnetic simulations [71, 72]. Our objective is not to make an optimal design from an electromagnetic perspective, but to build an inexpensive casing that does not hinder performance. With that goal in mind, we focus on three key parameters: the type of material, its thickness, and the distance of the radome to the radar's antennas. These parameters are captured in Figure 2.2a.

Material. Following the suggestion from the manufacturer (Texas Instruments), we use PTFE (Teflon) as the material for our casing. Teflon has a relatively low cost and it is easy to cut (manipulate) into the desired shape. Furthermore, there are off-the-shelf sheets available with different *thicknesses*, which is important because the thickness is another critical parameter.

Thickness. In mmWave systems, the signal's bandwidth usually spans several GHz. This means that the wavelength value will vary across such bandwidth and thus the optimal radome thickness will vary as well. Thus, in our evaluation we test sheets with nine thickness values, equal to 0.5, 1.0, 1.5, 2.0, 2.9(~ 3.0), 4.0, 4.9(~ 5.0), 5.1, and 6.0 mm.

Radome-antenna distance. The *radome-antenna distance* is determined by the signal's wavelength. Theoretically, it can be demonstrated that the optimal distance to minimize reflections caused by the radome is an integer multiple of half the wavelength in air [71, 72]. Using the central frequency as the reference point, our evaluation considers four distance values: 1.25 mm, 2.5 mm, 5 mm, and 10 mm, corresponding in terms of wavelength to $\lambda/4$, $\lambda/2$, λ and 2λ.

Shape. For imaging applications, the best shape is a spherical casing to ensure that the waves' propagation distance traveled within the dielectric radome is constant with respect to the angle of the radar's field of view [71, 72]. Similar to the commercial product we use as a benchmark (explained next), we use a flat surface due to its simplicity and

the fact that we are not focused on exact imaging applications, but on simple tracking.

2.3.2. EVALUATION.

To obtain the final radome's design, we test Teflon sheets considering all 36 combinations of thicknesses (9 values), and antenna-teflon distances (4 values).

To test the different configurations we use a flat aluminum plate as a target (due to its strong reflection). This plate is placed in front of the radar at the boresight, at a distance of approximately 6m. Figure 2.2b captures the setup used for the radome design.

As a quantitative metric to compare the different configurations, we use the Integrated Side Lobe Ratio (ISLR). This is the ratio between the energy of the sidelobes in the radar range profile and the energy of the main lobe associated with the target response. Lower values indicate better performance in the sense that the target lobe will be more easily detectable.

Figure 2.2c shows the ISLR results. It can be seen that the best values are obtained for PTFE sheets of thickness equal to 1 mm and a distance from the antennas of $\lambda/4$, i.e. 1.25 mm. These are the values used for the final fabrication of the radome, shown in Figure 2.4. It is important to highlight the importance of the empirical evaluation because the theoretical values for thickness and distance are not optimal.

2.3.3. COMPARISON WITH COMMERCIAL PRODUCT

We bought one unit of a commercial product, viz., WAYV Air from Ainstein, to benchmark the performance of our system. It uses the same mmWave chipset we use in the 60 GHz band. The brochure states that the sensor is designed for indoor scenarios with a maximum range of 6 m. Considering that the sensor is not designed for outdoor use (the casing does not appear to be weatherproof), our evaluation is done on a clear day in the same setup as our platform, c.f. Figure 2.3. In this outdoor evaluation, the range was around 2.5 m, and the system had trouble detecting more than three people. The cost of the Ainstein sensor is 600 $, while the total cost of our system is 310 $. The itemized costs are presented in Table 2.1. Besides reducing our costs by half, we also triple the range. In the next section, we describe the configuration used to achieve a range around 8 m. We cannot make conclusive remarks about the reasons for the differences in the design of our systems since all the design details of the Ainstein sensor are not available.

Table 2.1: Costs of Materials.

Component	Price($)/piece(excl. Taxes)
mmWave sensor: IWR6843ISK	120.87
Processor: RaspberryPi 4B	49.77
Casing: Front Radome (Teflon)	101.12
Casing: Back (Plastic)	12.74
Other (adapters, cables, etc.)	≈26.50
Total Price($) (excl. taxes)	**311**

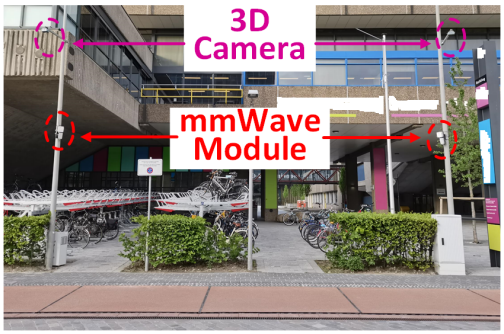


Figure 2.3: Scenario. The gray path is for pedestrians and the orange path is for bikes. The scenario has two cameras (outside our control) and two of our mmWave sensors.

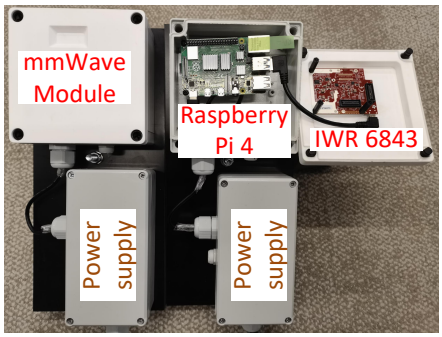


Figure 2.4: The mmWave modules installed on poles.

2.4. EXPERIMENT SETUP

Two mmWave sensors are deployed on our university campus. The data analyzed for this submission corresponds to three months: from January 2022 until April 2022. During this time, the sensors experienced diverse weather conditions. We collected the temperature, humidity, wind, precipitation and fog parameters through a weather station located in proximity to our deployment.

2.4.1. INSTALLATION SETUP

Figure 2.3 shows our deployment. The radome-enclosed sensors are placed on lamp posts adjacent to a road frequently used by pedestrians and cyclists. The nodes are installed close to two people-counting cameras that were present in the scenario to benchmark the performance of the mmWave sensors. The nodes are placed at a height of 3.54 m to avoid any pilferage. The orientations have an elevation angle of 25°, and azimuth angles of 30° and 150°. The casings have rubber gaskets, cable glands and pressure valves to ensure the radomes are resilient to water leakage.

2.4.2. SYSTEM CONFIGURATION

Sensor node: Figure 2.4 shows two sensor nodes. The left sensor is closed and ready to be deployed, and the right sensor is open. Each node consists of a TI IWR6843ISK as the mmWave chipset and a Raspberry Pi (RPi) 4B as the computational unit. The mmWave sensor computes the point clouds and counts the number of passing objects. This data is shared with the RPi through a UART link, and the RPi uploads the data to the cloud via WiFi. The power supply unit is connected to the AC mains and generates a regulated DC voltage suitable for the mmWave module and RPi.

The computation of point clouds, static-noise removal, and overall tracking is done by the firmware provided by the manufacturer (Texas Instruments). The firmware allows selecting different configurations, through empirical evaluation, we choose the configuration for 'Sense and Direct HVAC Control'. The configuration generates a 2-dimensional

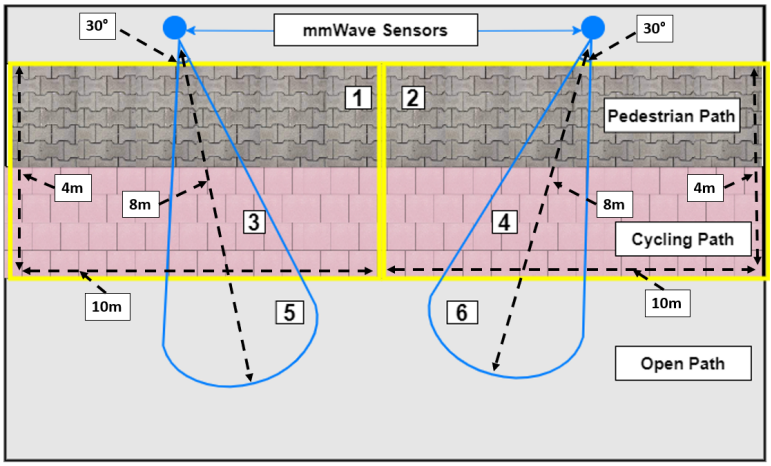


Figure 2.5: Eagle-eye view of deployment.

point cloud and covers the range and velocity parameters for our use-case. The range and velocity are 10.35m and 6.065 m/s respectively, while their resolutions are 0.089 m and 0.047 m/s.

2.4.3. AN INDIRECT MEASUREMENT OF ACCURACY

As stated before, we co-locate the mmWave sensors with people counting cameras. The cameras, however, are not under our control. These cameras only report the number of people in the Field-of-View (FoV), but their FoV differs in shape and size from the mmWave's FoV. Figure 2.5 shows that while the FoV of the cameras is rectangular (two yellow boxes), the mmWave sensor has a conical FoV (blue cones). Thus, the objects seen by the cameras and mmWave sensors differ. For instance, while objects 3 and 4 are detected by mmWave sensors as well as the corresponding cameras; only the cameras will observe objects 1 and 2; and only the mmWave sensors will observe objects 5 and 6. Overall, due to their bigger coverage, the cameras will see more people on average.

One approach to overcome this disagreement would be to modify the orientation of the cameras and utilize basic signal processing to bound the cameras' coverage to match the radars' coverage. However, due to stringent privacy regulations, we were not allowed to access the camera's images or modify their FoV in any way[2].

The inability to change the parameters or configuration of infrastructure that is already deployed is an important and frequent challenge in real urban setups. These constraints are in place to avoid jeopardizing in any way the safety, security or privacy of citizens.

To alleviate the disagreement between the different coverages, in the next section we propose a framework with different metrics to compare the performance of these two systems.

[2]The Ethics Review Board gave us permission to deploy the mmWave nodes.

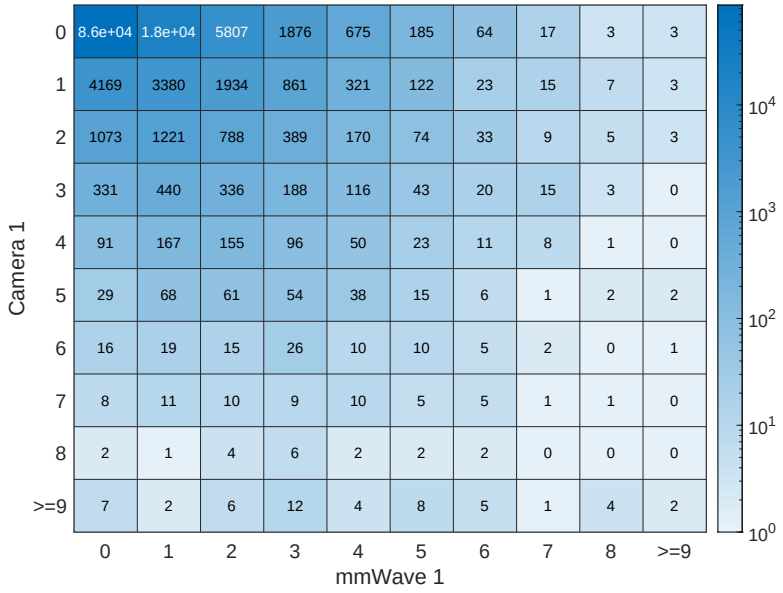


Figure 2.6: Confusion matrix. The heatmap follows a log scale.

2.5. EVALUATION FRAMEWORK

Considering that the cameras and mmWave sensors have different coverages, we first need to identify the appropriate metrics, periods and quantization levels to make sure that we do not distort the comparisons further. We start by formally defining the metrics, and then, we use those metrics to analyze the effect of different monitoring periods and quantization levels. Our analysis first focuses on empty areas and then on scenarios with one or more people.

2.5.1. ANALYSIS OF EMPTY AREAS

Our sensors have a sampling rate of one image per minute. Thus, during the three-month period, we gather around 130K samples for each sensor. Figure 2.6 shows the confusion matrix for one of the cameras and its corresponding mmWave sensor. Given that the coverage of the sensors is different, this matrix has to be considered carefully because the camera does *not* provide the ground truth, only an approximation of it. In spite of these differences, we are able to obtain one important insight.

Insight 1: Outdoors, mmWave sensors have an error of 15% detecting empty scenarios. Despite the differences in coverage, the fact that the camera's view is greater than the radar's, allows us to state –with high probability– that if the camera does not detect any person, the radar should not detect any either [3]. Thus, using the confusion matrix in Figure 2.6, we could use precision and recall to quantify the performance of mmWave sensors in detecting if an area is indeed empty. The precision of the system is 94%, that

[3] mmWave sensors can sense people in areas 5 and 6 outside of the cameras' FOV (yellow boxes) in Figure 2.5. However, almost all people commute on the paved road (pedestrian and cycling paths). Hence, when cameras cannot sense people, radar sensors probably cannot sense people either.

is, if the mmWave sensor indicates that the area is empty, very likely it is. The recall, on the other hand, is 76%, which means that almost a quarter of the empty scenarios are deemed as busy, i.e. the mmWave sensor indicates that there are one or more people when there are none in fact. The low recall is an artifact of the high background noise present in outdoor scenarios, any mobile element –an animal or a moving object – may be confused as a person. Overall, combining the precision and recall of the mmWave sensor into the F1 metric gives an accuracy of 84%, which is lower than the high accuracy reported for indoor scenarios (error ≤ 1 person for 97.8% [22]).

2.5.2. ANALYSIS OF OCCUPIED AREAS (ONE OR MORE PEOPLE)

The analysis done for empty areas cannot be applied to busy areas (one or more people). Given that the camera has a bigger coverage, it can detect the presence of persons that are not under the radar's coverage. Partly due to this smaller coverage, the radar's recall for presence detection drops to 67%, that is, around 30% of the time when the camera reports a person (or persons), the radar reports an empty scenario. The confusion matrix in Figure 2.6 also captures this underestimating behavior due to the larger coverage of the camera: the lower triangle has more points than the upper triangle. Next, we propose using correlation metrics to overcome the different coverages for scenarios with one or more people.

To analyze occupied periods, we first filter out all the samples where the camera indicates zero presence, and then, use correlation coefficients to analyze the count-similarity between the sensors. The use of a correlation metric builds upon the assumption that, over time, the flow of people observed by the mmWave sensor is proportional to the flow observed by the camera.

CORRELATION COEFFICIENT

The Pearson correlation coefficient is a widely accepted measure to evaluate the correlation between two variables, X and Y. The coefficient has a range between [-1,+1]. Values around ±0.8 are considered to have a high correlation, values around ±0.4 are deemed as a medium correlation and 0 implies that there is no dependency between the variables. Equation 2.1 gives the mathematical definition.

$$\rho_{X,Y} = \frac{\sum_{i=1}^{N}(x_i - \bar{x})(y_i - \bar{y})}{\sqrt{\sum_{i=1}^{N}(x_i - \bar{x})^2}\sqrt{\sum_{i=1}^{N}(y_i - \bar{y})^2}} \tag{2.1}$$

where, N is the sample size in X and Y. x_i, y_i are the i^{th} individual sample points of X and Y. $\bar{x}$, $\bar{y}$ are sample means for X and Y, i.e., $\bar{x} = \frac{1}{N}\sum_{i=1}^{N} x_i$.

PEOPLE-FLOW

Most crowd-sensing applications are not interested in an instantaneous view, but the flow over a period of time. Thus, instead of correlating the individual samples, we correlate the flow of people. The people's flow is the total number of people detected over a certain period of time. Formally, if we denote a_j as the number of people sampled at time j, the flow (x_i) for a period τ is given by: $x_{i=t} = \sum_{j=t}^{t+\tau} a_j$.

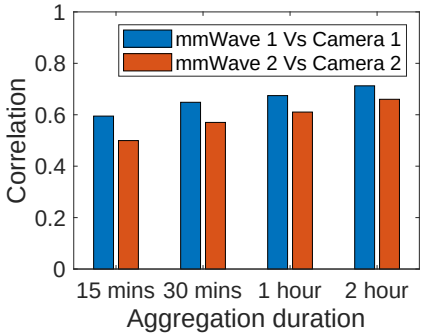


Figure 2.7: Correlation under different aggregation durations.

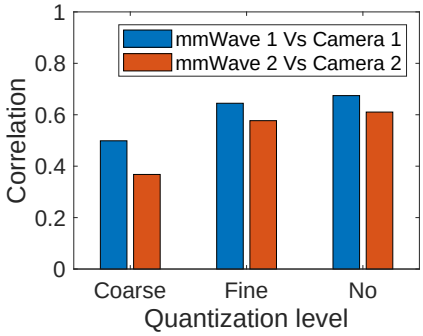


Figure 2.8: Correlation under different quantization levels.

In our analysis, we denote x_i^τ and y_i^τ as the flows measured at the mmWave sensor and camera, respectively, for a period τ. These flows are inserted in Equation 2.1 to obtain the correlation between the two types of sensors.

Figure 2.7 shows the comparison of correlation coefficients for different durations, viz., 15 minutes, 30 minutes, 1 hour and 2 hours. We see that the correlation coefficients increase as the period increases, but the difference is not significant. Thus, for simplicity, we use only the hourly flow for the remainder of our evaluation.

QUANTIZATION LEVELS

Often, crowd-monitoring applications do not require the exact number of people but labels such as 'Sparse' or 'Busy'. Thus, we compare the correlation of three granularity levels, viz., (i) No quantization, where the original x_i^τ and y_i^τ values are used in the correlation; (ii) Fine-quantization, where the crowd levels are divided into five categories and only the crowd levels are correlated; and (iii) Coarse-quantization, with three crowd levels. Tables 2.2 and 2.3 define the thresholds for the fine and coarse quantization, respectively.

Figure 2.8 shows the correlation coefficients for the two mmWave radars and their corresponding cameras at different quantization levels. We observe that the correlation increases for more fine-granular levels. The coarse quantization shows a correlation of around 0.4 while the fine- and no-quantization are around 0.6. This difference is not due to limitations of the hardware or the algorithms to measure low granularity levels, but it is rather an artifact of Pearson's equation, which penalizes the inaccuracy introduced by reducing the maximum range of the coarse quantization, capped at 31 people in Table 2.3. To avoid this artificially generated error, we consider the *no-quantization* in our next results.

Table 2.2: Threshold levels for fine-quantization.

Flow Levels	Empty	Sparse	Normal	Busy	Crowded
Hourly people flow	0	1-12	13-30	31-60	$\geqslant 61$

2.6. WEATHER OBSERVATIONS

Until now we have analyzed the aggregated data. In this section, we evaluate the impact of individual weather conditions. This evaluation is imperative for outdoor deployment. Using the information from a weather station close to our university, we divide the samples based on different weather parameters: temperature, humidity, precipitation, solar radiation, wind and fog. There were also a few days with snow but the samples were too few to analyze that condition. It is important to highlight that the weather station only provides data every six hours. Hence for true representation, we label only the samples close to the reporting time (± 1 hour). In effect, this means that over a period of six hours, we only have two samples for the hourly flow, instead of six.

To measure the correlation between both sensors, we first filter out the samples where the camera indicates zero presence. This is done for two reasons. First, 85% of the samples show no people, and hence, we do not want to skew the analysis. Second, the assessment of basic presence has been already done with precision and recall. After filtering the data, we use the hourly flow of people considering no quantization. The unquantized data for the radar and camera are then evaluated with the correlation coefficient.

To assess the impact of each weather condition, we segment them into bins. While the temperature, humidity, solar radiation and wind speed are divided based on different intervals, precipitation and fog are binned depending upon their presence. For example, Figure 2.10b shows four bins for temperature, but Figure 2.10d only shows two bins for precipitation: rain and no rain.

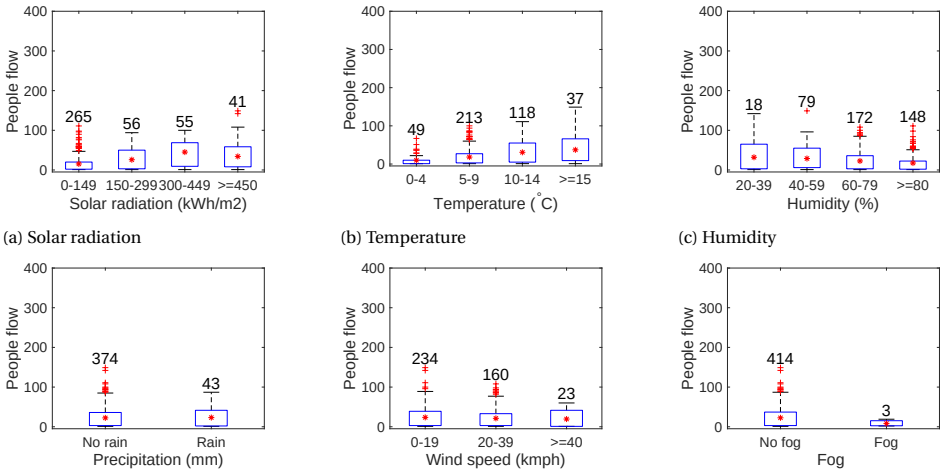
There are two important aspects that need to be considered for correlation analysis. First, we need a sufficient number of samples because, statistically, having few samples may incorrectly represent the impact. Second, the range of the values needs to be similar, otherwise, the Pearson equation adversely skews the correlation for lower ranges (as we saw during the analysis of different quantization levels, c.f. Figure 2.8).

To obtain the number of samples and ranges for all weather conditions (and their bins), we use the data from camera-1. Figure 2.9 show the box-plots for temperature, humidity, solar radiation, precipitation, wind speed and fog; and Figure 2.10 plot their respective correlation coefficients. The boxplots capture the quantiles and mean (red asterisk), and at the top of the plot, we state the number of samples. The correlations for the two sensors are represented in different colors: blue for radar/camera 1 and red for radar/camera 2 in Figure 2.10.

We divide the results into two main groups: benign and harsh conditions. Benign conditions are those where we do not expect the sensor to be impacted much, such as changes in temperature, humidity or solar radiation. Harsh conditions capture more challenging parameters, such as precipitation wind and fog.

Table 2.3: Threshold levels for coarse-quantization.

Flow Levels	Empty	Normal	Crowded
Hourly people flow	0	1-30	$\geqslant 31$



(a) Solar radiation (b) Temperature (c) Humidity

(d) Precipitation (e) Wind speed (f) Fog

Figure 2.9: People flow under different levels of weather conditions. In each level of a weather condition (each box-plot in a figure), we show the mean, quantiles, extremes and total sample number (on top of the box-plot).

2.6.1. BENIGN CONDITIONS

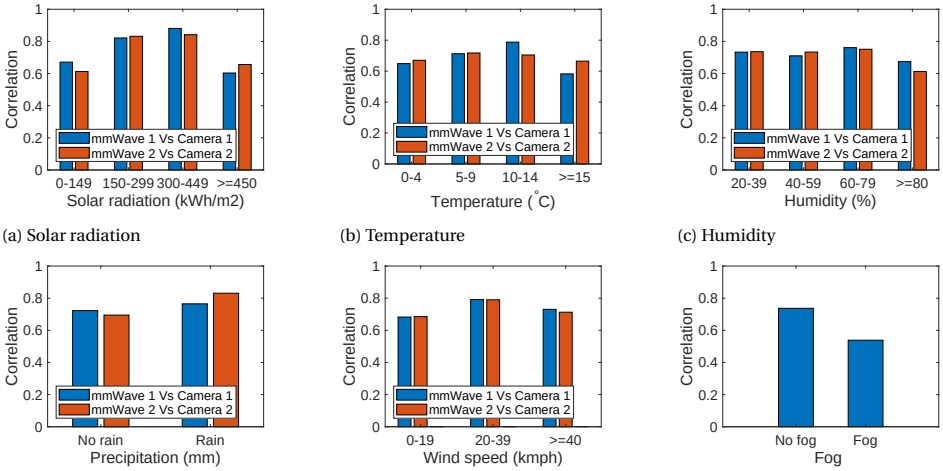
For temperature, humidity and solar radiation, the correlation values in Figure 2.10 follow the mean (red asterisks) in Figure 2.9. For example, for solar radiation the correlation increases for the first three bins and then drops (Figure 2.10a), following the same pattern of the mean in Figure 2.9a. There are some cases where the correlation coefficient does not follow the mean, but this effect can be explained by the lower number of samples. For example, for temperature, the correlation coefficients of the first three bins (Figure 2.10b) follow the mean of Figure 2.9b, but the last point drops because the number of samples decreases significantly from 118 to 37. The trends for humidity are more stable, the correlation coefficients (Figure 2.10c) capture the slow decreasing trend of the mean (Figure 2.9c) but without major differences. Overall, the above observations provide a second important insight.

Insight 2: Under benign weather conditions, radar sensors provide correlation coefficients that lay mainly between 60% and 85%. Correlation coefficients cannot be used strictly as metrics for accuracy, i.e. we cannot state that radars are 60%-85% accurate, but they indicate that even though radar sensors cannot track the people-flow in a fine-grained manner (as they do indoors), they can provide coarse-grained flow information.

2.6.2. HARSH CONDITIONS

We now focus our attention on more challenging weather conditions: precipitation, wind and fog. A positive result is that precipitation does not seem to have a negative impact on the radar sensor. Even though we have more sample points for the no-rain case, there is still a sufficient number of samples for the rain case (Figure 2.9d), and the correlation coefficients remain stable across the board, between 70% and 85% (Figure 2.10d).

The wind speed has a slightly higher impact than precipitation but the effect is not so detrimental. All wind categories have a similar mean (Figure 2.9e), and in spite of the



(a) Solar radiation (b) Temperature (c) Humidity

(d) Precipitation (e) Wind speed (f) Fog

Figure 2.10: Correlation coefficients.

differences in the number of samples, the correlation coefficients remain between 70% - 80% across the range (Figure 2.10e).

For the case of fog, the situation is more complex, when there is no fog, the correlation reaches a value higher than the aggregated data, around 70% (Figure 2.10f) compared to the 60% with no quantization in Figure 2.8. This better correlation makes sense because without fog all measurements are clearer. The problem we faced is that with fog we have limited data (only seven samples) and the pair of sensors-2 (red) were malfunctioning during some of this time. For sensors-1 (blue) the correlation is around 50%, but no conclusions can be made because of the limited amount of data and because cameras are less accurate on foggy days as well. Overall, the evaluation under harsher conditions provides a third important insight.

Insight 3: Precipitation and wind speed do not seem to affect the performance further, compared to the benign scenarios. Rain seems to have little to no effect, compared to clear days. Wind can cause false positives with radars, and thus, an important research problem for the community is the removal of background clutter with mmWave sensors, but the correlation coefficient during busy periods remains stable. Fog, on the other hand, is the situation where cameras are more vulnerable and a longer study with accurate ground truth data is required to assess the performance of radar sensors.

All in all, our outdoor evaluation provides a fourth overarching insight.

Insight 4: mmWave radars in the ISM band (60 GHz) are not yet ready to provide accurate information in outdoor setups. The research community could balance the effort put into analyzing radar sensors between indoor and outdoor setups. The tracking accuracy of radars indoors is high (above 95%) and hence there is no major gap compared to using cameras. Outdoors, however, the gap between cameras and radars increases to levels that prevent mmWave being used for fine-granular analysis.

2.7. RELATED WORK

To the best of our knowledge, there is no study evaluating the performance of mmWave radar for crowd tracking in the ISM band (60 GHz). Thus, we position our work within the following related areas.

Outdoor people counting: The majority of outdoor people counting solutions are implemented with video-based methods, which can detect a large number of people with good accuracy. Sacchi et al. exploit novel image processing methods for real-time estimation of tourist-flows [73]. Under different weather conditions, they report a counting error of about 10% for up to 300 tourists. Hou et al. estimate the number of people in complicated outdoor scenarios (up to 100 people) [74], with a best average error of around 10%. Overall, camera-based solutions are a reliable alternative for people counting but they raise up serious privacy concerns.

Another alternative for people counting is based on characteristics of Wifi signals. Depatla et al. develop an analytic model mapping the probability distribution of the received signal amplitude to people count [75]. Their accuracy with count error less than 1 is 92% for up to nine people. This WiFi solution preserves privacy but requires dedicated antennas and extensive training data.

mmWave for indoor applications: mmWave radars are finding a wide range of applications, from vital sign monitoring [68] and activity recognition [76] to people identification [60] and 3D pose reconstruction [67]. All these applications however are performed in *controlled* indoor setups with minimal background clutter and specific instructions to the users. A more related area to our work is precise indoor people counting. Through digital beamforming, multi-target detection and a robust clustering technique, Wu et al. can count people (up to 5) with an error ≤ 1 person for 97.8% of the time [22]. Weiss et al. use an adaptive OS-CFAR peak detection algorithm and a vital sign verification algorithm to detect and check targets. They can achieve 85.4% accuracy for indoor counting for 0-4 people [46]. Similarly, the study by Gross et al. demonstrates an accuracy of nearly 97% for up to 5 people in indoor scenarios [47]. All these indoor counting studies provide high accuracy but cannot be directly mapped to outdoor environment.

Automotive radar for outdoor people tracking: The 77-81 GHz frequency band is licensed for automotive applications. Thus, there is not as much work as in the 60 GHz band, but some studies report the use of this type of radar for people-detection. For example, Scheiner et al. propose a method to detect and track an object/person around corners with an automotive radar [77]. Through an image formation model for Doppler radar non-line-of-sight (NLOS) measurements, they can derive the position and velocity of people via reflections. Our work complements the studies done in the automotive bands, as we target a long-term urban scenario providing relevant information for traffic management.

2.8. CONCLUSION

We analyze the performance of mmWave sensing by deploying them outdoors for a period of three months and under different weather conditions. We provide a step-by-step design for a radome to enable outdoor deployment. Our study shows that the performance of mmWave sensors drops in the outdoor deployment as compared to the indoor

case. In general, our study demonstrates a wide sensing gap of an averaged 20% accuracy loss between low-cost mmWave sensors and precise cameras, indicating that radar sensors need to be further studied outside controlled indoor setups.

In this thesis, we do not look at further studies to bridge the sensing gap of mmWave sensors. Motivated by the work in this chapter, other members of our research team are investigating methods to improve the accuracy of crowd monitoring with radar sensors. In the following chapters, we introduce other applications and present alternatives to bridge their respective sensing gaps.

3

CardioID: Taming Irregular Cardiac Signals for Biometric Identification

In this chapter, we focus on cardiac identification and show that the sensing gap of this application can be reduced by enhancing the methodology. Cardiac signals are unique and can be used to identify people, but they are subtle to detect. Precise cardiac sensors can identify people with high accuracy, they can even distinguish twins [78]. We investigate cardiac identification with re-purposed PPG sensors and smartphone cameras. These devices can emit light that is reflected by the fingertip of people, and the reflected light carries cardiac information. However, under *realistic* scenarios, the reflected signals suffer high irregularity. In this chapter, we propose a general framework (methodology) to obtain stable and distinct features on cardiac signals. Our approach reduces the sensing gap of both re-purposed sensors, allowing them to achieve cardiac identification with balanced accuracy (BAC) of over 90% for a PPG sensor and over 80% for a smartphone camera.

3.1. INTRODUCTION

Biometrics play a fundamental role in human identification, and the most popular systems rely on external features, such as fingerprints, iris patterns, and face contours. These systems have excellent precision but they are vulnerable to attacks: fingerprints can be recreated in latex from touched objects [79]; iris patterns can be scanned and emulated [80]; and pictures from the Internet can be used to obtain renditions that can fool face recognition systems [81].

To overcome the fundamental weakness of *external* features, i.e., the fact that they can be easily captured because they are constantly exposed, researchers are investigating *internal* biometric signals, which are hidden under our skin, and hence, they are hard to obtain and forge.[1] An approach that is gaining interest is the use of cardiac patterns since they are uniquely defined by the heart, lung and vein structures of an individual [78]. These cardiac patterns can be obtained with a photoplethysmogram (PPG), which measures changes in blood volume via light absorption. PPG signals can be acquired with simple inexpensive sensors that are widely available on wearable devices. For example, one option is to use a pulse oximeter on a finger, which consists of a small LED and a simple photosensor [48]; another option is to place a finger on top of the flashlight and camera in a smartphone [49]. With both types of sensors, researchers have shown that PPG signals can provide between 85% and 95% identification accuracy for groups consisting of tens of people [50, 79, 84, 49].

Challenge. The results obtained so far for PPG identification are promising, but they have been obtained mainly under ideal situations: accurate sensors used in controlled environments. These two factors (sensors and environment) determine how similar cardiac cycles are for the same individual. The higher the similarity of the cardiac cycles, the higher the identification accuracy. We show that when PPG signals are gathered in a more natural (uncontrolled) manner, the cardiac cycles can be highly irregular, significantly decreasing the accuracy of state-of-the-art (SoA) approaches.

Our contributions. Considering the above challenge, we analyze the pernicious effects of irregular cardiac cycles on biometric identification and propose a novel framework to overcome those effects. In particular, our work provides four main contributions:

Contribution 1: Morphology Stabilization [section 3.3]. The biometric information present in cardiac cycles is restricted to a narrow spectrum of the signal. A key limitation of the SoA approaches is that their filters target the *same* spectrum for *all* individuals. This one-size-fits-all approach leads to either information loss (if the default spectrum is too narrow for a particular individual) or insufficient noise filtering (if the spectrum is too broad). We propose an adaptive technique that fine-tunes the filtering parameters based on the individual cardiac properties. This approach allows us to obtain more stable and distinctive features per user.

Contribution 2: Morphology Classification [section 3.4]. SoA studies assume that the cardiac pulses of individuals have a *single* dominant morphology (shape). Assuming a single morphology means that several "non-conforming" cardiac periods can be unnec-

[1] A few studies show that it is possible to spoof electrocardiography (ECG) authentication systems [82, 83], but this assumes access to a compromised medical database. That process is harder than obtaining a victim's fingerprint through the objects he/she has touched or obtaining a victim's face image by searching online.

essarily discarded, affecting the responsiveness of the system. More importantly, we find out that in some cases, the strict SoA assumption of considering a single dominant morphology, leaves out users that rarely have such cardiac morphology, rendering the SoA methods futile for those users. We show that a *single user* can have *multiple* valid morphologies. Our ability to consider a wider range of morphologies reduces the system's response time, increases user inclusion (to serve more people), and facilitates identifying the rightful individual even when his/her cardiac periods are different from each other.

Contribution 3: *Analysis of non-linear effects [section 3.5].* The SoA utilizes PPG signals to perform two types of biometric applications: identification and authentication. For identification, the SoA uses linear (PCA [49, 85], LDA [86, 84]) and non-linear approaches (NN-based [48, 87]), but there is no analysis determining what approach is better and why. We show that if we tackle the non-linear effects of cardiac cycles at an early stage, both approaches, linear and non-linear, render similar results. For authentication, we identify two main shortcomings in SoA methods: the use of Euclidean distances and the assumption that the features of a subject form a *single* cluster. To ameliorate these non-linear effects, we propose a *multi*-cluster approach, together with the use of the Mahalanobis distance.

Contribution 4: *Thorough multi-sensor and multi-application evaluations [Section 3.6].* The evaluation of cardiac signals for biometric applications can be divided into four quadrants: based on the type of sensor (pulse oximeter or camera) and application (identification or authentication). Most studies evaluate a single quadrant with healthy subjects (usually, identification with pulse oximeters), *no study has evaluated all four quadrants or considered unhealthy individuals*. Our evaluation assesses both applications relying on three datasets. The datasets consider both types of sensors (pulse oximeter and camera) and two types of individuals (healthy and unhealthy). Overall, the results show that in uncontrolled scenarios the average balanced accuracy (BAC) of the SoA drops beyond 15%, depending on the complexity of the uncontrolled scenario. This is a significant drop for identification and authentication applications. For the less dynamic dataset (using pulse oximeter sensors with healthy individuals), our method improves the accuracy by 15%. For the most challenging datasets (using smartphone cameras or monitoring patients in ICU), our methods improve the BAC with values above 10%.

3.2. PRELIMINARIES

3.2.1. PPG BASICS

The cardiac cycle represents the change in blood pressure determined by our hearts and blood circulation systems. Given that people have different heart structures in terms of volume, surface shape and motion dynamics [88, 89, 18], and different tissue thickness and blood vessel distribution [89], the cardiac signal has been used to obtain unique biometric signatures [52, 85]. A cardiac cycle can be measured in various manners, the simplest option is to obtain a PPG signal by measuring the amount of light absorbed by our body as blood flows through. A PPG signal can be measured with sensors containing inexpensive LEDs and photodiodes, or with the flashlight and camera in smartphones. The geometric relations among the various peaks and valleys present in a PPG signal

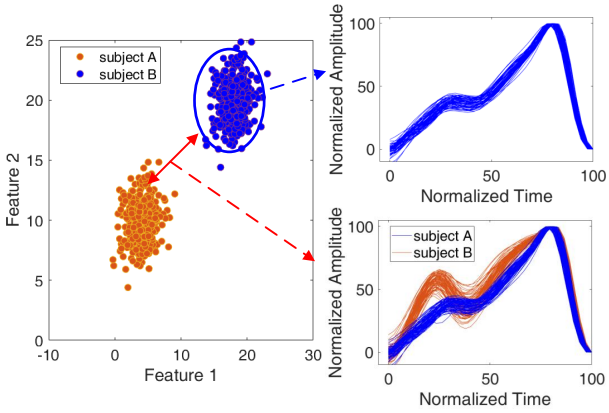


Figure 3.1: Sample application with *controlled* PPG signals.

(heights, widths, etc.) [79], or the spectral information in the frequency domain [48], are optional features used to perform identification.

3.2.2. APPLICATIONS, MORPHOLOGIES AND METRICS

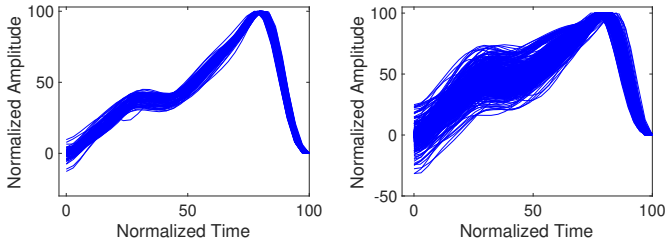
We analyze the performance of two applications: identification and authentication. In ***identification***, the population size is *known* and the training phase requires gathering data from *all* individuals. The goal is to determine classification boundaries among the *various* subjects. In ***authentication***, the population size is *unknown* and the training phase *only* gathers data from the user of interest. The goal is to determine the best authentication boundary for a *single* subject.

No study in the SoA has tested its methods with both applications: they only focus on one, usually on identification, which is simpler than authentication. Our study analyzes both. Independently of the target application, achieving high biometric accuracy with PPG signals requires attaining a delicate balance between two competing goals:

- ***Challenge 1***: *reduce intra-cluster variance.* We need cardiac cycles that are as homogeneous as possible for the **same individual**, in order to obtain stable features.
- ***Challenge 2***: *increase inter-cluster distance.* We need cardiac cycles that are as different as possible **among individuals**, to define clear identification boundaries.

Several SoA studies focus on controlled PPG signals [84, 49], which means no finger movements and fingertip pressure change during PPG collection, such as the situation in Figure 3.1 (we reconstruct these similarly clear PPG signals with 2 users). Under these favorable circumstances, it is simpler to tackle the above challenges and to differentiate the individuals.

Morphologies. We use the term *morphology* to refer to the shape of a cardiac cycle, and *stable morphology* to refer to cardiac cycles that have *(i)* the same numbers of peaks and valleys, and *(ii)* a small signal variance. For example, subjects A and B in Figure 3.1 have stable morphologies with two peaks and three valleys. In uncontrolled environments, gathering distinct and stable morphologies for each user becomes significantly more complicated.



(a) Controlled environment (b) Uncontrolled environment

Figure 3.2: Cardiac periods collected for the same subject.

Metrics. There is no common metric in the SoA to measure accuracy. Some studies use the equal error rate (EER) [48], others use F1-score [79] or BAC [49]. All these metrics are derived from true/false positive/negative results. Our evaluation can be presented with any of these metrics. We decide to use BAC because our datasets are unbalanced. BAC is the average of the true positive rate (TPR, sensitivity) and the true negative rate (TNR, specificity):

$$\text{BAC} = \frac{\text{TPR} + \text{TNR}}{2}, \tag{3.1}$$

where

$$\text{TPR} = \frac{\text{True Positive}}{\text{True Positive} + \text{False Positive}}, \tag{3.2}$$

and

$$\text{TNR} = \frac{\text{True Negative}}{\text{True Negative} + \text{False Positive}}. \tag{3.3}$$

3.2.3. THE DETRIMENTAL EFFECT OF IRREGULAR CYCLES

Multiple PPG studies report a high identification accuracy, ranging from 85% to almost 100%, depending on various evaluation parameters and scenarios [79, 49, 84, 50, 87, 48, 90]. Most of those studies, however, follow a *well-controlled* data gathering process, which results in limited distortions across cardiac periods, and thus, a good performance. The *controlled* process is reflected in two factors: 1) the conditions under which the dataset is gathered, and 2) the diversity of individuals in the dataset. Controlled datasets typically focus on healthy individuals with a narrow age range, between 20 and 40, as discussed in subsection 3.6.1. Furthermore, for each individual in the dataset, the measurements seem to be taken without considering the small but normal finger movements that affect the pressure between the fingertip and sensor. This type of control groups, consisting of young and healthy people without considering finger movements, leads to more stable signals.

In contrast to the *controlled* process, an uncontrolled process covers a more realistic scenario: a wider age range, including different health conditions, and considering minor (unconscious) finger and hand movements. Figure 3.2 depicts PPG signals for a single individual collected in controlled and uncontrolled environments. The small variance observed in Figure 3.2a is similar to the ones observed in Figure 1 in [84] and

Table 3.1: Performance of SoA.

	Controlled	Uncontrolled
Signal Variance	1.83	2.96
BAC for identification [79]	91%	72%
BAC for authentication [49]	93%	69%

Figure 3 in [49].[2] While it is not unreasonable to assume that PPG signals are collected in controlled environments, such assumptions constrain the ubiquitous applicability of PPG-based biometrics.

Differences in signal regularity can have a major impact on the performance of SoA methods. Table 3.1 shows a preliminary evaluation with four subjects, for whom we collected PPG signals in controlled and uncontrolled environments. The exact description of the SoA methods used as baselines for identification [79] and authentication [49], and the means used to calculate the signal variance, are explained in section 3.6. For now, *the important takeaway is that when the SoA is tested with controlled data, the performance is high (above 90%), as reported in the original studies; but when tested with highly variable signals, the accuracy drops significantly (around 70%).*

3.3. MORPHOLOGY STABILIZATION

A major shortcoming of the SoA is to use the *same spectrum* to filter the PPG signals of *all subjects*. In this section, we propose a novel adaptive filtering method. Figure 3.3 depicts a macro view of our approach and its relation with the SoA. First, we describe the methods we borrow from the SoA (subsection 3.3.1), and then, we describe their limitation and present our contributions (subsection 3.3.2).

3.3.1. SOA METHODS: BASIC FILTERING AND DERIVATIVES

Figure 3.3a depicts an ideal PPG signal. The biometric signature of an individual is captured by four fiducial points: diastolic (highest valley), systolic (lowest peak), dicrotic notch (which form a small peak in the middle of the period) and second wave. Figure 3.3b shows a raw PPG signal $s(t)$, which has two undesirable properties. First, a significant amount of noise distorts the location and intensity of the fiducial points, and in some cases, the noise level is high enough to erase the second wave and dicrotic notch completely, affecting the system's accuracy severely. Second, even in the ideal case, when all fiducial points are present, the signal's morphology is too simple and generic. Given that features are obtained based on the relative duration, heights, and slopes between fiducial points, the limited number of fiducial points limits the number of features. To overcome these effects, the SoA proposes a basic filtering step and the use of the second derivative of the PPG signal.

Filtering. To mitigate the noise in PPG signals, the SoA has identified the spectrum over which cardiac information is contained. For biometric purposes, the lowest meaningful frequency of a PPG signal is the heart rate. Considering that athletes can have heart rates as low as 0.5 Hz [91], the lower cut-off frequency f_l is usually set to that value. Regarding the upper cut-off frequency f_h, according to [92], sampling frequencies

[2] These studies do not post their PPG data. To infer the variance of their signals, we have to rely on their figures.

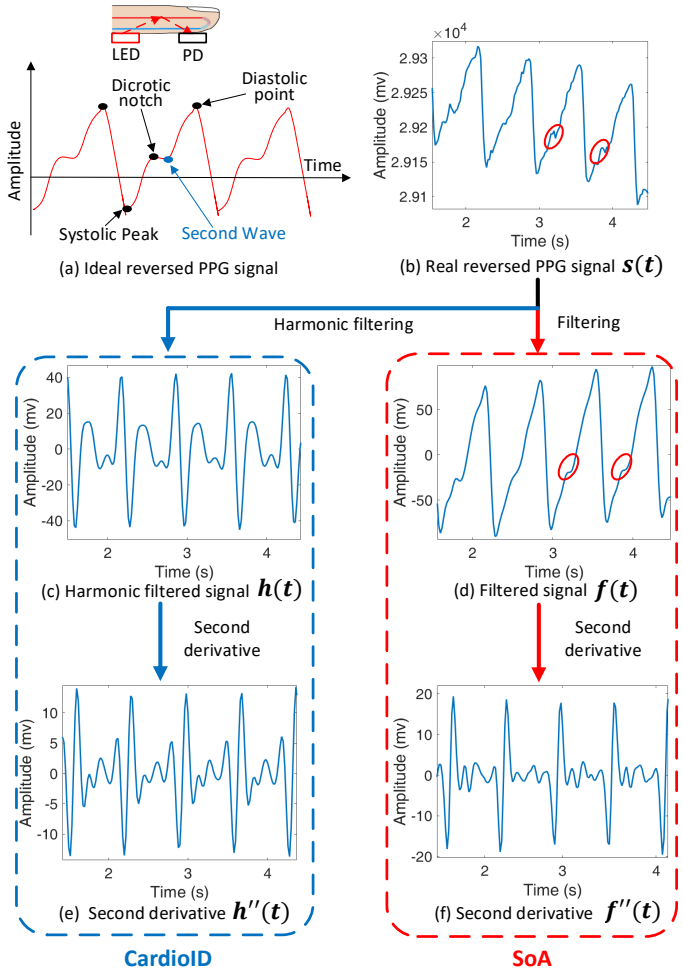
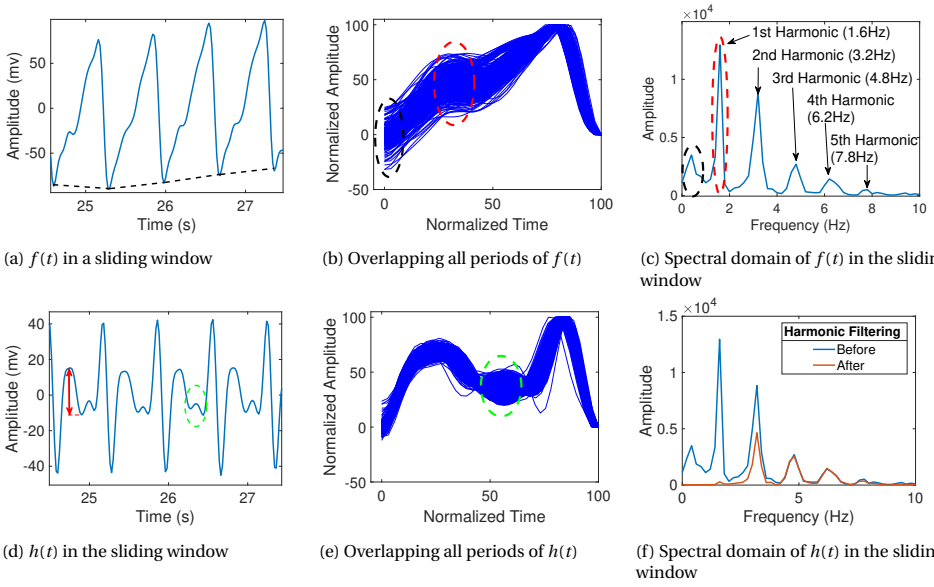


Figure 3.3: Morphology stabilization.

above 25 Hz do not provide any extra information, hence, f_h can be set to 12.5 Hz (due to the Nyquist-Shannon sampling theorem). Some studies use other filtering bands [79, 49], but the overall filtering process is similar. Figure 3.3d shows a PPG signal $f(t)$ after being filtered with a second-order Butterworth bandpass filter with bandwidth 0.5-12.5 Hz [93].

Derivatives. Filtering alleviates noise, but it also eliminates valuable information. For instance, the raw PPG signal $s(t)$ in Figure 3.3b contains faint but detectable second waves (red circles). After filtering, however, those fiducial points no longer exist (corresponding red circles in Figure 3.3d). To overcome this issue, researchers obtain features not only from $f(t)$ but also from its second derivative $f''(t)$ [79]. Figure 3.3f depicts the second derivative of the filtered cardiac signal, which exhibits more fiducial points than $f(t)$.



(a) $f(t)$ in a sliding window

(b) Overlapping all periods of $f(t)$

(c) Spectral domain of $f(t)$ in the sliding window

(d) $h(t)$ in the sliding window

(e) Overlapping all periods of $h(t)$

(f) Spectral domain of $h(t)$ in the sliding window

Figure 3.4: Frequency analysis to determine the lower cut-off frequency f_l.

3.3.2. Contribution: Harmonic Filtering

We also use the filtering and derivative stages, but we do not utilize the same parameters for all users. We propose a harmonic filtering phase that adapts its parameters to every individual. This process allows us to obtain more stable morphologies for every user (Challenge 1) and distinct fiducial points among users (Challenge 2). Our harmonic filtering depends on the subject's heart rate, which can change over time, thus, we track the heart rate using a 5-second sliding window with 1-second steps.

Determining the lower cut-off frequency f_l

The SoA usually uses a lower cut-off frequency that is too low, which increases signal variance and makes it hard to identify the most vulnerable fiducial points (second wave and dicrotic notch). Figure 3.4a shows the filtered signal $f(t)$ using SoA methods and Figure 3.4b shows the overlapping cardiac cycles using the endpoint of periods as an alignment anchor. We can observe a large variance in the starting points (black ellipsoid in Figure 3.4b) and significant instability in the dicrotic notch (red ellipsoid in Figure 3.4b).

Thus, *the fundamental question is how high should f_l be?* To obtain this optimal value, we analyze $f(t)$ in the spectral domain in Figure 3.4c. Our analysis leads to two important insights. First, the wide variance occurs because an $f_l = 0.5$ Hz does not filter important dynamics such as heart rate variability, the effect of respiration (slow changing frequency component) and subtle unconscious pressure changes on the fingertip, which are common phenomena in uncontrolled scenarios. Those dynamics generate a fluctuating envelope in the time domain (black dashed line in Figure 3.4a), which causes the height differences between the starting and end points in periods. Considering that the endpoints are the alignment anchors, those height differences among periods will

lead to a significant variance in the starting points. Second, an $f_l = 0.5$ Hz obscures the dicrotic notch. The energy of PPG signals is concentrated around the harmonics of the heartbeat, in particular the first harmonic (red ellipsoid in Figure 3.4c). The SoA does not filter the first harmonic because it uses the heart rate period as a feature, which is good, but the spectral energy of the heart rate overwhelms the second wave and dicrotic notch, which are the most vulnerable fiducial points.

Our analysis indicates that to lessen the dampening effects of the heart rate period, we need to filter out the first harmonic. We noticed, however, that for some subjects the second harmonic is as high (and as dampening) as the first harmonic and should be attenuated too. Therefore, denoting the frequency of the first harmonic as f_{1h}, we set $f_l = 2f_{1h}$. Note that with our approach we do *not* lose the heart rate data because it is contained in the other harmonics.

DETERMINING THE UPPER CUT-OFF FREQUENCY f_h

High-frequency noise modifies the location of fiducial points, which in turn, affects the stability of features and the overall performance of the system. Depending on the individual, a $f_h = 12.5$ Hz may be too high. For example, in Figure 3.4c the spectral energy is almost negligible beyond 10 Hz. Considering this situation, how low should f_h be?

As stated earlier, it is central to preserve the most vulnerable fiducial points on PPG signals (second wave and dicrotic notch). We use Figure 3.5, which zooms into those two vulnerable points, to illustrate the derivation of f_h. Denoting t_1 as the duration between the second wave and the dicrotic notch, the sine wave in the FFT containing the spectral energy of these points has a period of $2t_1$, which means that f_h must be higher than $1/2t_1$, else those two fiducial points would be filtered out. Now, denoting t_p as the period of a cardiac cycle, we observed empirically that $2t_1 > t_p/5$, and consequently, in the frequency domain $1/2t_1 < 5/t_p$. Finally, considering that $5/t_p$ represents the fifth harmonic of the heart rate, we set $f_h = 5.5f_{1h}$ to preserve all fiducial points while removing high-frequency noise. The negligible frequency components beyond the fifth harmonic in Figure 3.4c prove the correctness of our analysis.

ADAPTIVE FILTERING

The frequency response of our filter is solely based on the value of the first harmonic, $2f_{1h}$ to $5.5f_{1h}$, which is simple to obtain from the signals. More importantly, our approach is based on the subject's heartbeat instead of fixed parameters, allowing us to perform accurate adaptive filtering per subject. Figure 3.4d, Figure 3.4e and Figure 3.4f show the signals filtered with our method, their overlapping cycles and spectral domains. We can observe that, compared to the filtered signal $f(t)$ in the SoA, $h(t)$ has three advantages: *(i)* the signal variance is much lower throughout the entire cycle, Figure 3.4e; *(ii)* the difference between the second wave and dicrotic notch is accentuated significantly, red arrow in Figure 3.4d; and *(iii)* our method exposes another fiducial point, green ellipsoid in Figure 3.4d, which we can exploit to obtain more features as explained next.

DERIVATIVES

As described earlier, the SoA uses derivatives to accentuate the presence of fiducial points. We borrow that idea to obtain the second derivative of our harmonic signal $h(t)$. Fig-

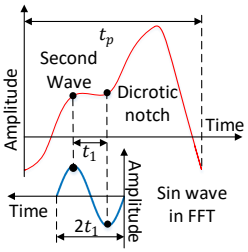
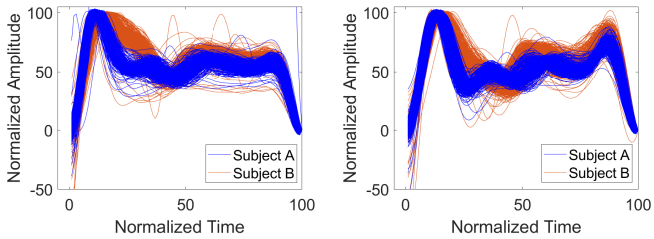


Figure 3.5: Frequency analysis to determine the upper cut-off frequency f_h.



(a) SoA: $f''(t)$ (b) CardioID: $h''(t)$

Figure 3.6: Overlapping periods with uncontrolled data.

ure 3.6 plots overlapping cycles for $f''(t)$ and $h''(t)$ for two sample subjects with uncontrolled data. Our second derivative $h''(t)$ has two important advantages compared to the SoA's $f''(t)$. First, even though $f''(t)$ is more stable than $f(t)$ because the derivative removes offsets, $h''(t)$ is still less variable because it inherits the stability of $h(t)$. The variance of $f''(t)$ for subjects A and B are 2.8 and 3.0, respectively, while for $h''(t)$ are 2.2 and 2.8. This lower variability helps to tackle *Challenge 1*. Second, thanks to the tailored cut-off frequencies of our adaptive filter, $h''(t)$ can exploit the specificity of $h(t)$ to obtain more distinctive morphologies for different users, tackling *Challenge 2*. Compared to $f''(t)$, the fiducial points of $h''(t)$ are more distinctive and conspicuous across the *entire* time domain. Furthermore, subject A (blue) in Figure 3.5b shows that the second derivative disentangles the 'knot' caused by the new fiducial point captured by the green ellipsoid in Figure 3.4e.

Summary. Overall, our approach also follows the two basic steps of the SoA, filtering and second derivatives, but using a novel filtering method leads to a more stable morphology for each user (Challenge 1) and more distinctive morphologies for different users (Challenge 2). The only input parameter required by our filter is the first harmonic (heart rate period), which can be easily obtained from any PPG signal. The SoA obtains its features from $f(t)$ and $f''(t)$, and we obtain them from $h(t)$ and $h''(t)$. An exact description of the selected features is presented next.

3.4. MORPHOLOGY CLASSIFICATION

Existing studies share a common underlying assumption: all cardiac signals have a single dominant morphology. That, however, is not necessarily the case. We show that a single user can have multiple *valid* morphologies. Without this insight, a system would need to either discard periods that do not conform to a pre-defined morphology (introducing latency), or consider all periods with different morphologies, but at the risk of obtaining widely different features for the same user (reducing accuracy).

In this section, we first describe the segmentation method to obtain cardiac periods, then we show that cardiac periods can have multiple morphologies, and finally, we describe the features used in those various morphologies.

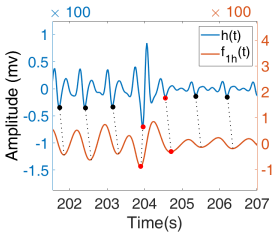
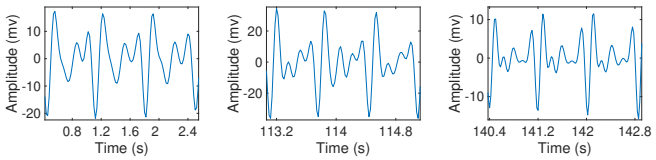
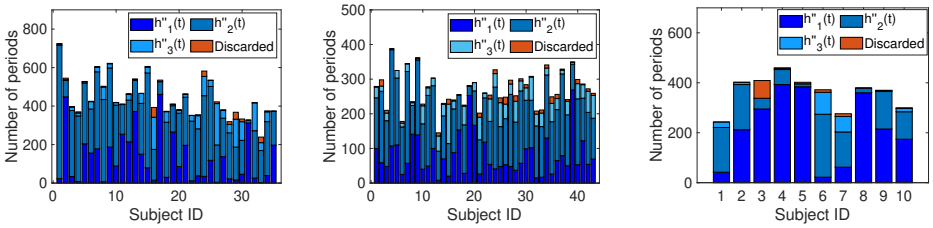


Figure 3.7: Segmentation method.



(a) $h''_1(t)$: 3 peaks, 4 valleys (b) $h''_2(t)$: 4 peaks, 5 valleys (c) $h''_3(t)$: 5 peaks, 6 valleys

Figure 3.8: Dominant morphologies for $h''(t)$.



(a) Public (pulse oximeter) (b) Private (camera) (c) Public (pulse oximeter, MIMIC-III)

Figure 3.9: Frequency of occurrence of morphologies in three datasets. The datasets use pulse oximeters and cameras.

3.4.1. SIGNAL SEGMENTATION

Several methods can be used to segment periodic signals. Many approaches use amplitude-based thresholds to detect periodic peaks or valleys [94], however, given the strong distortions present in our signals, we decided to use a spectral approach [95, 96]. Considering that our harmonic filter $h(t)$ relies on the first harmonic f_{1h}, we design a segmentation method that also relies on f_{1h}. We use a filter with bandwidth $[0.5, 1.5] * f_{1h}$ to isolate the heartbeat period. A sample harmonic signal $h(t)$ and the corresponding signal used for segmentation $f_{1h}(t)$ are shown in Figure 3.7. In spite of the distortions, our approach can accurately map the valleys from $f_{1h}(t)$ to $h(t)$ and perform segmentation.

Our harmonic signal $h(t)$ can cope with movements and changes in finger pressure, but sometimes the movement of the finger is so strong that a period becomes invalid. Our segmentation method has the ability to discard those events. For example, for the signal in Figure 3.7, the first three and the last three periods (black dots) are valid, even if the finger pressure is different, but the middle period (red dots) captures a drastic finger movement that should be invalid. We introduce two criteria to verify a period. First, on the x-axis, the interval between two adjacent valleys on $h(t)$ must be similar to the period corresponding to the heartbeat. Second, on the y-axis, the values of the valleys at the start and end of a period should be similar. For the signal in Figure 3.7, the period with the red dots is discarded because it violates the second criteria. The segments for $h''(t)$ are obtained following a similar mapping approach.

3.4.2. MULTIPLE MORPHOLOGIES

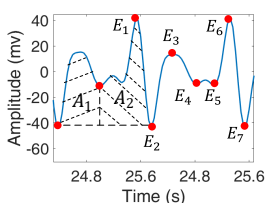
Currently, all studies using fiducial points assume a single macro morphology for all subjects. That is a valid approach in controlled scenarios, but in uncontrolled scenarios various factors can cause the appearance of multiple morphologies: unintended fingertip pressure [97], significant differences in the cardiac profiles of subjects, etc. When we perform the second derivative of our harmonic signal $h(t)$, we observe multiple morphologies. Figure 3.8 depicts the three most dominant macro morphologies observed in $h''(t)$: $h''_1(t)$, $h''_2(t)$, $h''_3(t)$.

Our evaluation considers three datasets, and those dominant morphologies account for (i) 98.4% of the periods measured in a *public dataset* with 35 subjects [98], 15301 out of 15557 periods; (ii) 97.5% of the periods measured in a *private dataset* with 43 subjects that we collected from volunteers, 11328 out of 11617 periods; and (iii) 97.2% of the periods measured from ten subjects in ICU (intensive care unit) from the public *MIMIC-III dataset [99]*, 3509 out of 3612 periods. Figure 3.9 shows the presence of the three macro morphologies in those datasets. There are other macro morphologies in $h''(t)$, but we do not consider them because they are rarely present ("Discarded" label in the figures).

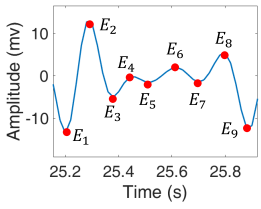
If we would choose only one morphology as the template for all subjects, as conventional methods do, the system could face two major problems. *First*, it may take a long time to identify a subject because the system will need to wait for the right morphology to arrive. For example, for the public dataset, our system can obtain 1.229 (15301/12444) periods/s, compared to much lower speeds if would only use $h''_1(t)$ (0.3584 periods/s), $h''_2(t)$ (0.7427 periods/s), or $h''_3(t)$ (0.1283 periods/s). *Second*, and perhaps more critical, the right morphology may never arrive for some of the subjects or it may be so rare that there would be insufficient samples to train the system properly. In practice, such a limitation would render an identification system futile because the basic premise is that it should be able to identify *all* members in a target group. In uncontrolled scenarios, no morphology is dominant. Even $h''_2(t)$, which is the most common, may be rarely active in some subjects, such as user 31 in the public dataset, and subjects 5 and 8 in the MIMIC-III dataset. Hence, the key advantages of considering multiple morphologies are decreasing latency and eliminating the risk of excluding some types of subjects.

3.4.3. FEATURE EXTRACTION

We extract features from $h(t)$ and $h''(t)$. Like several other studies in the area [79, 18, 49], our features are largely based on the geometric relations amongst fiducial points. Figure 3.10 and Table 3.2 provide a pictorial representation and the notation for all the features. In our notation, $E_i(t)$ and $E_i(a)$ denote the time and amplitude of fiducial point i. For $h(t)$, shown in Figure 3.10a, we collect three types of features: *1)* the duration of a period, *2)* the ratio of the areas inside a period, and *3)* the differences in duration, height and slope between consecutive fiducial points in one period. The total number of features for $h(t)$ is 14. For $h''(t)$, we only consider the third type of feature (differences between contiguous fiducial points). Figure 3.10b shows the features for $h''_2(t)$, and the same principle is applied to extract the features from $h''_1(t)$ and $h''_3(t)$. In the end, the number of features for $h''_1(t)$, $h''_2(t)$ and $h''_3(t)$, are 18, 24 and 30, since they have 3, 4 and 5 peaks, respectively (Figure 3.8). Features based on duration and height are susceptible



(a) Fiducial points on $h(t)$



(b) Fiducial points on $h_2''(t)$

Figure 3.10: Fiducial points used to extract features.

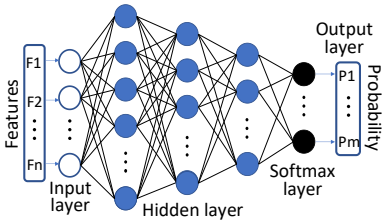


Figure 3.11: The structure of our neural network.

Table 3.2: Features extracted from $h(t)$ and $h''(t)$.

	Feature	Description
$h(t)$	Period duration	$E_6(t) - E_1(t)$
	Area ratio	A_1 / A_2
	Duration (D_i)	$E_{i+1}(t) - E_i(t), i = 2,3,5,6$
	Height (H_i)	$\lvert E_{i+1}(a) - E_i(a)\rvert, i = 2,3,5,6$
	Slope	$(-1)^{i+1} * H_i / D_i$
$h''(t)$	Duration (D_i'')	$E_{i+1}''(t) - E_i''(t), 1 \le i \le 8$
	Height (H_i'')	$\lvert E_{i+1}''(a) - E_i''(a)\rvert, 1 \le i \le 8$
	Slope	$(-1)^{i+1} * H_i'' / D_i''$

to heartbeat variance. In [52, 49], the authors state that normalizing the features makes them immune to heart rate changes. Therefore, we also normalized the duration and height of $h(t)$ and $h''(t)$. [3]

3.5. IDENTIFICATION AND AUTHENTICATION

As stated earlier, SoA studies only evaluate one type of application: identification or authentication (mainly identification). We consider both. Our system relies on the same set of features for both cases. Upon receiving a raw PPG period, we first obtain $h(t)$ and $h_i''(t)$, and derive their features. The combined features, $h(t)$+$h_i''(t)$, are given as inputs to two different processing branches depending on the type of application. Considering that performing identification is simpler, we first present that system, and later we focus on authentication.

3.5.1. IDENTIFICATION

Identification requires gathering training data from *all subjects*, and during the testing phase the aim is to match an incoming cardiac sample to the right subject. As with many other classification problems, PPG identification requires two main components:

[3] For $h(t)$ we also observed two types of morphologies: one where E_4 and E_5 are present and the other where those two points merge into a single valley. However, contrary to the multi-morphology approach used for $h''(t)$ in subsection 3.4.2, we decide to use a single morphology for $h(t)$ because the features obtained from the fiducial points E_4 and E_5 did not have any impact on the system's accuracy. The information from those two points gets disentangled and captured in one of the three morphologies present in $h''(t)$. Due to this reason, we do not evaluate $i = 4$ for $h(t)$ in Table 3.2.

dimensionality reduction, to identify the most informative features; and *decision boundaries*, to perform accurate classification.

The SoA utilizes two supervised learning methods, linear and non-linear, but does not provide insights about which one is better and why. In our evaluation, we consider both methods. The most representative linear method is linear discriminant analysis (LDA) [86], which simultaneously reduces dimensionality and draws decision boundaries. The most well-known non-linear methods are based on neural networks (NN) [48]. As it is customary with NN [100], we first use an autoencoder for dimensionality reduction, blue layers in Figure 3.11, and then, we add a softmax layer to perform classification (decision boundaries), black layer.

Considering that we use three morphologies, we need three LDA and NN pipelines running in parallel for each morphology (each pipeline receives the corresponding set of features presented in Table 3.2). Since LDA is an analytical solution, the LDA module is the same for all three pipelines (but with different training data). In contrast to LDA, due to the influence of the network structure and parameter values, we tailor three different NN modules for each morphology. The hidden (blue) layers for morphologies one, two and three are 128-64-32, 170-85-42 and 128-64-32, respectively. The activation functions are sigmoids to guarantee the non-linearity of the system, and parameters such as L2 and sparsity regularization are tuned for each morphology.

3.5.2. AUTHENTICATION

Contrary to identification, in authentication systems, the training set only consists of samples from the *legitimate subject*, while its testing set can include samples from *any subject*. Authentication also requires dimensionality reduction and boundaries, but given that we lack information about other users, drawing an optimal boundary for that single legitimate user becomes more complex. Next, we first explain the methods used in the SoA for dimensionality reduction, and then, some techniques to improve the definition of boundaries.

Dimensionality reduction can be performed with linear and non-linear methods. There are two mainstream linear techniques: principal component analysis (PCA) [101] and non-negative matrix factorization (NMF) [102]. NMF requires non-negative features, but the slopes in our feature set can be negative. Hence, similar to prior studies [49], we adopt PCA. Even though there are several non-linear dimensionality reduction techniques –such as Isomap [103], local linear embedding (LLE) [104], t-distributed stochastic neighbor embedding (t-SNE) [105], and autoencoder [100]– we did not find SoA studies using them for PPG authentication. Isomap, LLE and t-SNE share a common disadvantage for PPG-authentication: they must perform an entire recalculation every time a new test point is added. Autoencoders, on the other hand, do not have that shortcoming. We performed a preliminary evaluation of authentication with autoencoders but the performance was not good. We hypothesize that it is due to the limited data, autoencoders are usually trained with at least thousands of training points[4]. PPG-based systems are trained with a few minutes of cardiac data in one subject, which maps to a few hundred cardiac periods. In identification, NN methods can be trained with sev-

[4] At a popular Quora forum discussing "How large should be the data set for training a Deep auto encoder?" Yoshua Bengio states the need for having large amounts of training data [106].

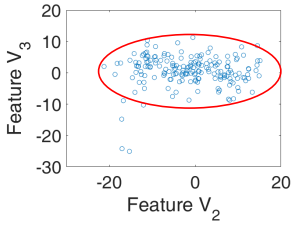


Figure 3.12: Mahalanobis.

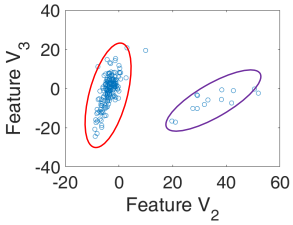


Figure 3.13: Multi-cluster.

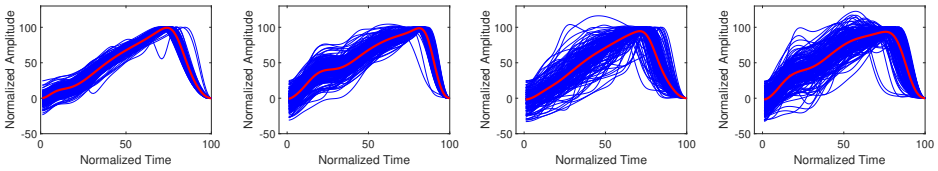
eral thousand samples coming from *all users*, but in authentication, we only have a few hundred samples coming from *the legitimate user*. Due to this finding, in our evaluation section, we only consider PCA for authentication.

Mahalanobis distance. After dimensionality reduction, the most significant features of a subject usually form a cluster similar to the one shown in Figure 3.12. When a new test sample arrives, the system calculates the average distance of this new point to the cluster. If the distance is below a threshold, the user is deemed legitimate. Many studies use Euclidean distances to measure proximity [49]. But Euclidean distances are fundamentally ill-equipped to deal with feature spaces that have widely different variances. For example, in Figure 3.12, using Euclidean distances, with any threshold, leads to a boundary that has the shape of a circle. The circle will be either too long for feature v_3, causing numerous false positives; or too short for v_2, causing significant false negatives. Therefore we adopt the Mahalanobis distance [107], which considers the standard deviations in each dimension and can be used to define tight boundaries such as the red ellipsoid shown in Figure 3.12.

Multi-cluster approach. Current PPG authentication systems assume that the features of a user converge to a *single* cluster [86, 49]. However, with uncontrolled data, we observed that a single subject can form two or more clusters for a single morphology, as depicted in Figure 3.13. We need an authentication system that can identify multiple clusters and then use the Mahalanobis distance to set an appropriate threshold for each cluster.[5]

For our purposes, the clustering method should meet three requirements: *(i)* be resilient to the presence of outliers, *(ii)* able to detect clusters with arbitrary shape, and *(iii)* fast. Hierarchical clustering methods, such as BIRCH [108], and centroid-based methods like K-means [109] are vulnerable to outliers and cannot detect arbitrary shapes. Most grid-based clustering methods, like CLIQUE [110], and density-based methods, like OPTICS [111] and DBSCAN [112] do not have shortcomings (i) and (ii), but they need a relatively long computation time. Due to the above reasons, we decide to use WaveCluster [113], which exploits the multi-resolution property of wavelet transforms. WaveCluster can identify arbitrary shape clusters at different degrees of accuracy, it is insensitive to outliers and has a low time complexity $O(n)$.

[5] To calculate the Mahalanobis distance, the number of samples must be greater than the number of features (dimensions). If a cluster has few samples, like the purple one in Figure 3.13, we use spline interpolation to add the extra necessary points.



(a) Variance 1.8 (b) Variance 3.9 (c) Variance 5.8 (d) Variance 6.6

Figure 3.14: Different signal variances.

Table 3.3: Details of the three datasets used in our evaluation.

	# subjects	# female	Age (Mean, Variance, Range)	RD (mins)	CTE
Public (pulse oximeter)	35	12	28.4, 14.04, 10-75 Y/O	5-6	4.29
Private (camera)	43	16	36.7, 14.93, 12-79 Y/O	4	5.97
MIMIC-III	10	8	Encrypted	5	4.72

3.6. Performance Evaluation

In this section, we describe the datasets we use, the studies taken from the SoA as baselines for comparison, and the results for the evaluation of identification and authentication.

3.6.1. Datasets

We use three datasets to evaluate the performance of CardioID. The first dataset uses a *pulse oximeter* and is public [98]. All subjects are sitting during the signal collection. The second dataset is collected by us from volunteers that are also sitting.[6] We use the *camera & flashlight* of an iPhone-7 to record 60-FPS videos of the volunteers' fingertips. In each frame, we focus only on the red channel of the pixels covered by the fingertip. The method to select the covered pixels is the same as in [49]: $I_{\text{red}} > 80\% \times (I_{\text{red}} + I_{\text{blue}} + I_{\text{green}})$. We average the red-channel intensities among the selected pixels to represent one data point of a PPG signal. To maximize the peak-to-peak amplitude of cardiac periods, we carefully set the three parameters affecting the camera's exposure: the aperture and ISO are set to the lowest values, -2 and 20, respectively, and the shutter speed to 200. Other parameters like white balance, focus and zoom are set to auto. The third dataset is obtained from patients in the MIMIC-III waveform database [99]. All subjects from this dataset were ICU patients (intense care unit) at a Medical Center in Boston, USA. Their PPG signals are gathered with a pulse oximeter in multiple sessions. We select 10 patients from that dataset and their information is shown in Table 3.4.

Significance of datasets. The parameters of the three datasets –the number of subjects, gender, age distribution (average, variance and range), the recording duration (RD) and variability (cross-track error, CTE)– are given in Table 3.3. There are three important points to highlight about the selection of our datasets.

First, *no SoA study has analyzed the performance of their methods using both types of sensors, pulse oximeter and camera.* In general, a pulse oximeter is more precise than a camera because its infrared spectrum can enhance the signal quality, and its finger clip can reduce the noisy motion artifacts [94]. This is one reason why the CTE (variability)

[6] This dataset was gathered under the approval of the Ethics Committee (HERC) of our university.

Table 3.4: Information of the subjects in the MIMIC-III dataset.

Subject ID	Gender	Ethnicity	Training data collection time	Testing data collection time	Interval (hours)	Disease
10045	F	White	23:32	9:31	10	Fever
10083	F	White	12:28	23:28	11	Hypetension
10124	F	Jewish	9:34	18:05	9	Congestive heart failure
41976	M	Hispan	23:18	20:50	21	Pneumonia
42033	F	White	17:29	13:38	20	Shortness of breath
42199	F	White	12:41	20:09	8	Chest pain
42302	F	White	12:42	19:32	7	Asthma COPD Flare
42367	F	White	11:59	19:59	8	Seizure; status epilepticus
43798	M	White	18:27	5:34	11	Esophageal CA/SDA
43827	F	Unknown	13:29	20:47	7	MI congestive heart failure

in Table 3.3 is higher for the camera dataset.

Second, *our datasets consider a more diverse group of people.* Even for *healthy* people, which is the main focus of the SoA, the morphology of cardiac signals can vary significantly based on the age group and skin tone. This is another reason why the CTE in Table 3.3 is higher for the camera dataset. The studies we use as baselines, [49] and [79], focus on a narrow age segment of the adult population, 22-33 and 18-46, respectively. The age ranges of the first two datasets we use are 10-75 (public) and 12-79 (private), including children, teenagers, adults and elders. Moreover, we also consider a bigger population: 70% more for the camera sensor (43 people vs. 25, [49]) and 17% more for the pulse oximeter (35 vs. 30, [79]). In terms of skin tones, the public dataset includes only medium-toned skin, and our private dataset includes 51% medium, 33% light and 16% dark skin.

Third, *no study considers unhealthy individuals or long periods of time between the training and testing periods.* To identify the limits of PPG-based identification and authentication, we consider a dataset (MIMIC-III) with people having various health related-problems from asthma and seizures to heart failures and hypertension (Table 3.4). For these subjects, we select signals for the training and testing phases that are **at least 7 hours apart**. Since these individuals are ICU patients, they are subject to the effects of medicine, which can change their physiological and psychological conditions even within a session.

Overall, to the best of our knowledge, these datasets consider –by a wide margin– the most demanding conditions in the SoA.

Signal variance analysis. Gathering data from different sensors, while considering motion artifacts and a broader range of people, provides us with more realistic (less controlled) PPG signals. To quantify the variance of these signals, we first obtain the average signal for a user, red signals in Figure 3.14; and then, we calculate the cross-track error (CTE)[7] from every (blue) PPG signal to its average. Denoting e_i as the CTE for signal i, the signal variance for a subject is the mean absolute error for all e_i's. The average signal variances for our datasets are 4.29, 5.97 and 4.72, shown in Table 3.3. *In order to put these values in context, it is important to note that the variance found in SoA plots is a*

[7] The CTE is used in GPS systems to measure the difference between the given and followed paths. We tried different similarity metrics, including dynamic time warping, and we found that the CTE captures the similarity of PPG signals in a more precise manner.

bit lower than what is shown in Figure 3.14a, i.e. less than 2. The majority of users in the public dataset have a variance in the range [2,6]; in the private dataset in [4,6] and in the MIMIC-III dataset in [2,6]. Hence, our evaluation handles situations with a wide spectrum of signal variability.

3.6.2. BASELINES USED FOR COMPARISON

We utilize two SoA studies as baselines for comparison, one for identification [79] and the other for authentication [49]. The reasons for selecting those baselines are presented in section 3.7. In this subsection, we quantity the improvement in the acquisition rate for CardioID and the difference in acuraccy between our work and the SoA baselines.

QUANTIFYING ACQUISITION RATES

Every study in this research area, including ours, removes periods that do not conform to the required morphologies. The goal is to discard as few periods as possible, while maintaining high accuracy. Denoting S as the cardiac periods from all users and S' as the periods used by a system (after discarding non-conforming morphologies), the acquisition rate is given by S'/S. With controlled data, the acquisition rate is high, $S' \approx S$; but with uncontrolled data, the rate can be very low, $S' \ll S$. As stated in section 3.4, a low rate can increase the system's delay and in some cases exclude the participation of some users. Hence, before even assessing the accuracy of the system, we need to make sure that a method has the capability to recognize 100% of the users.

Considering that S is equal to 14347, 10728 and 3612 periods for the public, private, and MIMIC-III datasets, respectively, we first need to find S' for the SoA baselines. The morphology and features used by Kavsaouglu *et al.* are presented in Figures 7 and 8 in [79], and the corresponding information for CardioCam is provided in Figure 8 in [49].[8] We use that information to discard the morphologies that do not conform to their requirements because the right morphology is necessary to obtain their features. After discarding the non-conforming morphologies in the SoA, we obtain the following acquisition rates: 74.6% (public dataset), 64.5% (private dataset) and 66.3% (MIMIC-III dataset) for [79]; and 59.2% (public), 32.8% (private) and 37.4% (MIMIC-III) for [49]; significantly lower than the 98.4% (private), 97.5% (public) and 97.1% (MIMIC-III) obtained for CardioID. *More importantly, in the camera dataset there were three users that* ***did not have a single cardiac period*** *resembling the morphology required by [49], and thus, there is no possibility to authenticate them with that method.*

Moreover, the SoA studies [79, 49] show a long acquisition delay. There are 12444, 10320, and 3000 seconds of data for the public, private and MIMIC-III, correspondingly. The acquisition speeds are 0.932 (public), 0.726 (private) and 0.798 (MIMIC-III) period/second for [79]; and 0.740 (public), 0.369 (private) and 0.450 (MIMIC-III) period/second for [49]; significantly lower than the 1.229 (private), 1.097 (public) and 1.169 (MIMIC-III) periods/second obtained for CardioID. The SoA acquisition speeds are below 1 period/second. Some of the acquisition speeds are even lower than 0.5 period/second, which means that users would need to wait a few seconds to be identified.

[8] Upon close inspection, we notice that, in both SoA studies, the second derivative of $f(t)$ is a signal similar, but not exactly the same, as morphology-2 in our case, $h_2''(t)$.

CARDIOID VARIANTS

To assess the impact of our contributions –*morphology stabilization*, *morphology classification* and *the reduction of non-linear effects for authentication*– we perform an ablation study creating different variants for CardioID. For the SoA approaches, we implement them based on the morphologies and features provided in their respective studies.

For identification, we consider four variants.

- Variant I.1 (MS): we use **morphology stabilization** to obtain $h(t)$ and $h''(t)$ with their respective features, cf. Table 3.2. This variant only considers morphology-2 periods, $h''_2(t)$. The classification method is K-NN, the same as in [79].
- Variant I.2 (MC): we add **morphology classification** to the MS variant. Here, we include periods with morphologies $h''_1(t)$ and $h''_3(t)$ but we still use K-NN for the classification.
- Variants I.3 and I.4 (CardioID.LDA and CardioID.NN): we replace K-NN in the MC variant with **LDA** and **NN**, respectively. We consider these variants the final implementations of CardioID for identification.

For authentication, we also consider four variants.

- Variant A.1 (MS): similar to the MS variant used for identification, but instead of K-NN, it uses PCA and Euclidean distance to achieve authentication, as in [49].
- Variant A.2 (MC): it adds **morphology classification** to the MS variant.
- Variant A.3 (Mahal): it replaces the Euclidean distance with **Mahalanobis distance** in the MC variant.
- Variant A.4 (CardioID): it adds the **multi-cluster** approach to the Mahal variant.

EMULATING A WIDE RANGE OF SIGNAL VARIANCES

Our aim is to evaluate the SoA baselines and CardioID variants under a wide range of signal variances. Collecting that type of data would require asking users to steadily increase the level of finger movement and pressure from low to high. That would be a complex process, instead we use our datasets to create (emulate) subsets with increasing levels of variance. To generate the emulated subsets, we perform the following process. For every user, we only include signals that lead to a variance less than t, where $t = 2,4,6$. If after this filtering process, a variant cannot collect 20 periods from a user, we leave the user out of the emulated set because we would not have sufficient training data for that user. A macro view of the emulated subsets is presented in Table 3.5, Table 3.6, and Table 3.7. For example, if we look at Table 3.5 for reference [79], we can see that if we set $t = 2$, *(i)* only 21.6% of the periods would have a variance less than 2 **and** satisfy the morphology requirement of [79]; and *(ii)* only 29 subjects, out of 35 (82.9%), have more than 20 periods satisfying the conditions in (i).

Table 3.5 and Table 3.6 provide two important insights. First, for all variance levels, the MC variant has the best performance in terms of including more users and having the highest acquisition rate (because it accepts three different morphologies). The SoA

Table 3.5: Percentage of detectable subjects and periods in the public (pulse-oximeter) dataset.

Public	Signal variance	2	4	6	All
[79]	Subject %	82.9	100	100	100
	Period %	21.6	54.2	67.3	74.6
[49]	Subject %	68.6	80	82.9	82.9
	Period %	19.5	45.9	55.2	59.2
MS	Subject %	57.1	97.1	97.1	97.1
	Period %	8.1	29.8	44.9	59.4
MC	Subject %	74.3	100	100	100
	Period %	14.8	50.6	75.1	98.39

Table 3.6: Percentage of detectable subjects and periods in the private (camera) dataset.

Private	Signal variance	2	4	6	All
[79]	Subject %	37.2	90.7	100	100
	Period %	5.5	33.1	51.2	64.5
[49]	Subject %	25.6	41.9	58.1	69.8
	Period %	4.7	18	25.9	32.8
MS	Subject %	18.6	81.4	95.3	100
	Period %	3.4	22.3	37.5	55.5
MC	Subject %	30.2	93	100	100
	Period %	5.5	37.6	64.3	97.5

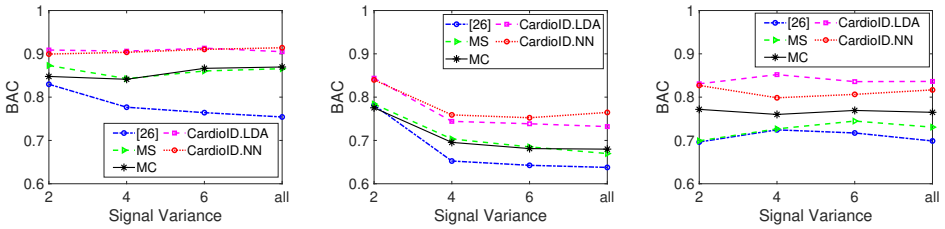
baselines and the MS variant have lower performance because they consider only one morphology. Second, when we consider all the data (last column in the tables), one of the baselines [49] cannot include all users in either dataset. This is an important point because it means that the (only) morphology allowed by [49] has requirements that are so stringent that some users may rarely (or never) show the required morphology. *In fact, with the camera dataset, three users did not have a single cardiac period satisfying the morphology requirements, and two of those users were above 50 years old (an age bracket that was not considered in the SoA).* The broad type of users in our dataset enables us to expose this type of limitation. The variants LDA, NN, Mahal and Cluster, have the same % of subjects and periods as MC because they are derived from that variant.

PROVIDING THE RIGHT CONTEXT FOR PERFORMANCE EVALUATION

Until now, we have only evaluated the impact of CardioID on the acquisition rate. In the remainder of this section we evaluate its accuracy, and it is important to consider the following points: (i) for each emulated subset, we use the first 80% of periods for training and the rest for testing; (ii) the variance observed in the SoA signals is around 1.5, hence, the variance considered in our emulated subsets ($t = 2, 4, 6$) poses a greater challenge; (iii) for CardioID and the SoA baselines, we use a *single period* to perform identification or authentication, using more periods would increase the performance but also latency; (iv) using a single cardiac period with *controlled signals*, the SoA reports a BAC of 0.95 [49], which can be translated, in expectation, to the rightful user having a probability around 95% to get access to the system (sensitivity), and an attacker a probability of around 5% of being successful (specificity). Thus, our goal is to try to get as close as possible to a BAC of 0.95 with *uncontrolled signals*. In authentication and identification, *improvements in the order of 5%, or above, are considered significant.*

Table 3.7: Percentage of detectable subjects and periods in the MIMIC-III dataset.

MIMIC-III	Signal variance	2	4	6	All
[79]	Subject %	70	90	90	90
	Period %	28.4	53.1	60.6	66.3
[49]	Subject %	40	80	80	80
	Period %	11	29.3	34.2	37.4
MS	Subject %	70	80	90	90
	Period %	8.6	20.7	26.3	31.8
MC	Subject %	80	100	100	100
	Period %	25.2	59.8	77.9	97.1



(a) Public (pulse oximeter) (b) Private (Camera) (c) MIMIC-III (pulse oximeter)

Figure 3.15: Identification performance.

3.6.3. IDENTIFICATION

Public dataset. Figure 3.15a shows the results for the pulse oximeter dataset. The MS variant provides the most significant improvement, above 10% for most approaches. This occurs because our harmonic filter adapts to every user, enabling more distinct and stable features. The MC variant does not really improve the BAC, but –as described before– the fact that it accepts multiple morphologies improves inclusion (more subjects are accepted) and the acquisition rate (40% higher than the MS variant and 20% higher than the SoA baseline [79] for the entire dataset, cf. last column in Table 3.5). As for the *LDA* and *NN* variants, the performances are similar. The improvement of both solutions over the MC variant is around 5%. The final BAC of CardioID reaches 91.6%, close to the 95% target, while the SoA drops to 76.1% [79], leading to a total improvement of exceeding 15%.

Private dataset. Figure 3.15b shows the results for our camera dataset. We can see that, compared to the prior dataset, the overall performance is lower, but the same general trends appear, showcasing the ability of CardioID to improve the identification performance across different sensors and users. Considering the subsets with variances four and above, we obtain the same benefits as for the pulse oximeter: the MS variant provides about a 4% improvement, the MC variant offers a marginal improvement in terms of accuracy but significant improvements on subject inclusion and period acquisition rate (cf. Table 3.6), and the classification methods (LDA and NN) provide an improvement of around 7%.

It is important to note that even though CardioID performs significantly better than the SoA with the camera dataset (12% better), it is still far from the 95% BAC target, which leaves some chances for attackers. At the end of this section, we discuss some ways to overcome this problem.

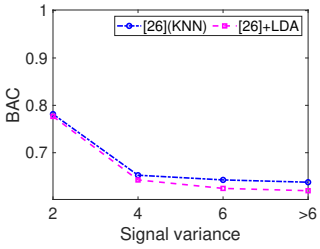


Figure 3.16: LDA contribution.

MIMIC-III dataset. Figure 3.15c shows the result for the patients in ICU. The cardiac signal is obtained with a pulse oximeter, but *the time difference between the training and testing for all subjects is at least 7 hours. In the prior two datasets, as well as for all the other SoA studies, the training and testing sets are contiguous.* The overall performance of the MIMIC-III dataset is better than with the camera dataset. The difference between the SoA baseline [79] and the MS variant is about 3%. The MC variant, whose multiple morphologies allow to nearly triple the number of cardiac periods and consider 10% more subjects (cf. Table 3.7), improves the performance by 3% with respect to the MS variant. Although the NN classification method is inferior to LDA, it still provides a 5% improvement. The LDA method, which is the one that performs best, can obtain a BAC of 83%, which is a 15% improvement compared with the SoA [79].

Linear vs. non-linear methods. SoA studies have been using linear [86, 90, 84] and non-linear [48, 87] methods to perform identification, but no study has compared both approaches or stated why one is preferable over the other. Our evaluation shows that both methods have a similar performance (Figure 3.15). We hypothesize that this occurs because our morphology stabilization and classification provide cardiac periods with stable and distinct features, and hence, the role of the classification method is less prominent. To highlight the importance of our morphology variants (MS and MC), we replace the K-NN classifier used in [79] with LDA in the camera dataset (while maintaining everything else the same). The results in Figure 3.16 show that LDA even degrades the performance a bit. Without the stable and distinct PPG signals provided by our morphology stabilization and classification, a machine learning method cannot do much on its own to overcome the high variance present in uncontrolled scenarios.

3.6.4. AUTHENTICATION

Public dataset. Before discussing the authentication results, we need to highlight a critical difference compared to identification: for authentication, the SoA [49] **fails** to include all types of users. In Table 3.5, we can observe that both, the SoA baseline used for identification [79] and the MC variant, consider all 35 subjects for $t = 4$ and above, and thus, the comparison is unbiased because the population size is the same. However, *the maximum number of subjects considered by [49] (baseline for authentication), is only 29 (82.9%).* This occurs because our evaluation requires at least 20 periods per subject, but the morphological requirements of [49] are too stringent, which does not allow getting enough periods for 6 subjects. Hence, for the results we present with authentication,

CardioID faces a more challenging scenario than the SoA because a bigger population increases the likelihood of errors.

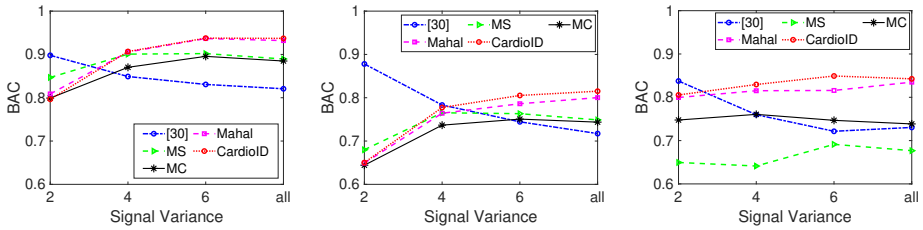
With that clarification, we can now discuss the main insights for the private dataset (pulse oximeter) in Figure 3.17a. First, at $t = 2$, all the approaches have a similar performance. This occurs due to the limited data. For $t = 2$, the emulated subset filters out most samples. Contrary to identification, where the system can exploit the samples from *all the other users*, in authentication, the system can only use the limited samples belonging to a *single* user. Thus, for $t = 2$, the performance of the system is largely determined by the small number of relatively well controlled signals, leaving little room for the methods to showcase their respective strengths. For $t = 4$ and above, the MS and MC variants play the same role as in identification: MS increases the performance, while MC increases the participation (number of subjects) and the acquisition rate (reduces delay). Overall, CardioID achieves a 93.7% BAC with 35 subjects, while [49] achieves 11% less BAC with only 29 subjects.

Private dataset. Figure 3.17b shows the results with the camera dataset. Due to the higher signal variability, the SoA baseline filters even more periods than with the prior dataset: the percentage of subjects for the SoA is less in Table 3.6 than in Table 3.5 (69.8% vs. 82.9%), and as stated before, three subjects did not have a single period satisfying the morphological requirements in [49], which would make the SoA system invalid for this target group. Even if we leave that critical point aside, CardioID still outperforms the SoA by 10%, but it is still not able to reach the desired 95% BAC target. A counter-intuitive trend with the camera dataset is that CardioID's performance increases significantly with the signal variance. But this is not due to the increase in variance per se (which adds noise), but due to the increase in data (when more variability is allowed, the subset contains more cardiac periods and subjects).

There is one result, however, that we did not expect and it highlights a particular strength of the SoA with the camera. For $t = 2$, the baseline [49] has a strong performance compared to all of our variants, and it is in accordance with what the authors report in the original paper.[9] Initially, we thought that it was because, at $t = 2$, the SoA considers only 11 out of 43 subjects (25.6% subject participation in Table 3.6), but our MS variant, which relies only on morphology-2, also considers the same amount of subjects and performs worse. The SoA has strict morphological requirements that filter out many cardiac periods, but this conservative standard allows them to have more similarity among their periods when the data is controlled, which is particularly useful in authentication because the training phase utilizes a single user. The stability of our features relies solely on the harmonic filter. After that, the conditions to consider a morphology valid are rather permissive: simply counting the number of peaks and valleys present.

MIMIC-III dataset. Figure 3.17c shows the results for the ICU patients. Similar to the other datasets, the benchmark [49] cannot include all subjects (80%). The performance of the MS variant is noticeably below other methods, due to the limited amount of data (cf. Table 3.7). Among the other variants, the Mahalanobis distance contributes to the largest improvement, around 10%. Overall, CardioID improves the BAC performance by 12% with respect to the benchmark [49].

[9] In [49], the authors report a 95% BAC using a single sample. Their signals seem to have a variance of 1.5. If we extrapolate the performance of [49] in Figure 3.17b, we will obtain the reported result.



(a) Public (pulse oximeter) (b) Private (camera) (c) MIMIC-III (pulse oximeter)

Figure 3.17: Authentication performance.

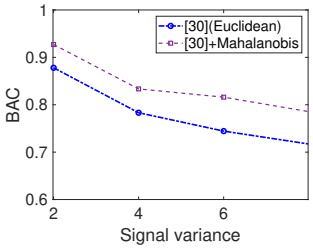


Figure 3.18: Mahalanobis contribution.

Mahalanobis contribution. A final point to discuss is the role of the Mahalanobis distance. Among all our variants, the Mahalanobis variant provides the biggest improvement,[10] which leads to the following question: If we simply replace the Euclidean distance with the Mahalanobis distance in [49] with the camera dataset, would the performance get comparable to CardioID? The result is shown in Figure 3.18, where we can see that there is an improvement across all signal variances but the gain is not as significant as when Mahalanobis runs on top of MS and MC. When all the signal variance is considered, the BAC of the modified SoA ([49]+Mahalanobis) is below 80%, while that of CardioID is around 85%. Similar to what happened with identification, where we simply modify the SoA to bypass MS and MC (cf. Figure 3.16), this result proves that all the foundational blocks of CardioID are important to obtain a maximum performance gain.

Summary. Based on the results with all the datasets, we can summarize the following takeaway lessons. Overall, morphology stabilization (section 3.3) and the methods to overcome the non-linear effects of authentication (section 3.5) improve performance, while morphology classification (section 3.4) improves the acquisition rate (more users and less latency). For the public dataset with uncontrolled data, our methods can bring back the accuracy of identification and authentication close to the desired target of 95% BAC using a single cardiac period for testing. With the private and MIMIC-III dataset, however, even though our methods still have a better performance than the SoA, they do not reach the desired 95% target. To ameliorate this problem, one could place stricter constraints on the types of morphologies that can be accepted and use more periods for

[10] We also tried the variant MC+MultiCluster, without including the Mahalanobis distance, but the result of that variant was lower than the MC+Mahalanobis variant.

Table 3.8: The most relevant studies in the SoA. Some studies evaluate multiple datasets, if the dataset is public, a reference is given in the '# subjects' column. The studies in bold are used as comparison baselines in our work.

	Application	Sensor Type	# Subjects	# Features	Decision Method	# PT	Accuracy
[50], 2003	Identification	pulse oximeter	17 (own)	4 (fiducial)	discriminant function	–	94,0%
[90], 2013	Ident.	pulse oximeter	44 14	1 (entire period)	correlation functions	16	5.3% (EER) 14.5% (EER)
[79], 2014	Ident.	pulse oximeter	30	40 (fiducial)	KNN	1	87.0% (F1)
[84], 2016	Ident.	pulse oximeter	32 , [114]	15 (fiducial)	LDA & QDA	5	92.5% (Rank-1)
[48], 2017	Identification	pulse oximeter	42 , [115]	DWT (non-fiducial)	SVM & NN	–	100%
[86], 2018	Auth.	pulse oximeter	42 , [115] 32 , [114] 41 , [116]	CWT (non-fiducial)	Direct-LDA	8-40	1.6% (EER) 3.8% (EER) 8.8% (EER)
[49], 2019	Auth.	camera	25	30 (fiducial) & 36 (non-fiducial)	PCA	1	95.8% (BAC)
CardioID	Both	Both	35 , [98] 43 10	32,38,44 (fiducial)	LDA & NN (Ident.) PCA (Auth.)	1	93.2% , 95.5% 78.2% , 85.3% 83.6% , 84.3% (BAC: Iden. , Auth.)

testing, but that would increase the system training overhead and delay while testing.

3.7. RELATED WORK

We divide the related work into three main phases, highlighting the elements we build upon from the SoA and the novelty of our work. A summary of the most relevant studies is presented in Table 3.8.

Phase 1: Basic identification. Gu *et al.* report the first results for PPG identification using just four features. They achieve an accuracy of 94% using a discriminant function [50]. Later, researchers found that the derivatives of a PPG signal can provide more stable and unique features [117]. Motivated by those initial results, researchers performed further experiments and found that the reported high accuracy is strongly dependent on the data-gathering process. Spachos *et al.* [87] considers fiducial features with two data sets. They report widely different performances for each set, EERs (Equal Error Rate) of 0.5% vs. 25%, leading them to state that PPG signals can be used for identification ***"given that [they] are collected under controlled environments and with accurate sensors"***. Bonissi *et al.* [90] also find significant differences in EER depending on the databases they use, 5.3% vs. 14.5%.

One of the most comprehensive evaluations is performed by Kavsaoğlu *et al.* [79]. They use 40 features from PPG signals and their derivatives (first and second), and utilize a *single period* for testing to obtain an F1 accuracy of 87.2% (F1 is a stricter metric compared to BAC). From that work, we borrow the idea of using the second derivative and multiple features. We implemented this method and used it as a comparison baseline for identification.

Phase 2: non-fiducial features and more challenging PPG signals. An important motivation for our study comes from [84] and [86]. Sarkar *et al.* [84] analyze PPG signals with subjects that undergo various emotions. To enforce the same morphology, they normalize signals in time and amplitude so fiducial features can maintain a common pattern. This approach, however, requires 20 cardiac periods per testing sample (~15 sec,

too long of a delay). Yadav *et al.* [86] also look into PPG signals that consider different levels of emotions and physical exercise, but they propose to use *non*-fiducial features based on continuous wavelet transform (CWT). CWT considers the spectral response of a signal, which is more resilient to noise than geometric (fiducial) features, but they still require long testing sequences, between 8 and 40 cardiac periods (~6 to 60 sec). Karimian *et al.* [48] also utilize non-fiducial features (discrete wavelet transformation, DWT), and report an accuracy of 100%, compared to a 95-98% accuracy obtained with fiducial features, but does not report the number of periods used for testing.

From the above studies, we take two insights. First, the normalization of PPG signals in time and amplitude to overcome distortions caused by emotions. Second, we do not use spectral features due to the many cardiac periods needed for testing, but we do perform a thorough spectral analysis (harmonic filtering) to obtain more stable and distinct fiducial features.

Phase 3: cameras and authentication. Most studies focus on performing *identification* with *pulse oximeters*, but a recent work uses smartphone *cameras* to attain *authentication* [49] (CardioCam). Authentication is more challenging than identification because it trains the system with a single user. CardioCam achieves a high BAC (95.8%) using a single cardiac period for testing. Motivated by CardioCam, we also collect data with a smartphone camera and implement the *signal processing chain* proposed in that work (filters, features and PCA method). Our results show that its performance decreases with irregular PPG signals. Furthermore, for some users, the requirements of CardioCam's morphology are so strict that they would not be able to use the system.

There are also *some other studies related to our work.*

Cardiac health applications. Several cardiac health applications use a smartphone camera. Seismo [42] develops and evaluates a smartphone-based blood pressure monitoring application. It uses the smartphone's accelerometer to measure the vibration caused by the heart valve movements to measure the pulse at the fingertip. The Pearson correlation coefficient of the blood pressure estimation is 0.2–0.77. Chandrasekaran *et al.* [118] combines sound information from the chest and video from a fingertip to measure people's blood pressure. Their estimation accuracy is above 95%. HemaApp [7] infers hemoglobin levels based on the light absorption detected by a smartphone camera, and achieves sensitivity and precision of 85.7% and 76.5%. Zhang et al. [119] adopt a smartphone camera to monitor blood glucose. They use Gaussian functions to fit the camera PPG signal and GSVM-based (Gaussian support vector machine) classification to identify the blood glucose level. The overall accuracy is 81.49%. Siddiqui et al. [120] capture the PPG signal on the red channel to estimate the pulse rate with an accuracy of 98.02%. All these systems aim at vital signs, which only require information on few fiducial points on cardiac signals. while identification needs to consider the whole morphology of cardiac signals. Therefore, we cannot adopt their methods.

Identification with ECG sensors. The most common cardiac sensor is ECG. These sensors are widely available in hospitals and have also been used for identification. Safie *et al.* uses ECG features to obtain an AUR and EER of 0.9101 and 0.1813, respectively [121]. Silva *et al.* explore a less invasive form of ECG sensors for authentication, finger-ECG [122]. They utilize pre-processed templates as inputs for K-NN and SVM, and obtain an EER below 9.1%. Similarly, Singh *et al.* [123] use just one electrode to extract 19 fiducial fea-

tures, and apply an adaptive threshold to perform authentication with a 99% accuracy. ECG sensors are more accurate than PPG sensors and smartphone cameras, but they are less pervasive and their filtering and identification methods are similar to the SoA studies using pulse oximeters.

3.8. CONCLUSIONS

Motivated by the fact that PPG biometrics have been largely limited to controlled setups, we analyze the impact of more realistic *uncontrolled* signals. We identify three main limitations in the SoA: the same filtering parameters are used to obtain the features for all users, a single dominant morphology is assumed, and there are important non-linear effects that have not been considered. These limitations create a wide sensing gap, preventing low-end sensors from performing authentication. Our novel methodology, named CardioID, overcomes those limitations with a new morphology stabilization and classification mechanism, and by using the Mahalanobis distances with a multi-cluster approach. Overall, considering a wide variety of cardiac cycles from three datasets, our novel *methodology* (first alternative) increases the *average* BAC from around 70% to 85%, significantly reducing the sensing gap.

4

CamPressID: Optimizing Camera Configuration and Finger Pressure for Biometric Authentication

This chapter builds upon the prior one, targeting cardiac identification. To reduce the sensing gap even further, we focus on *enhancing the re-purposed sensor*. In the previous chapter, the smartphone camera yields a lower performance compared to the PPG sensor. This situation stems from two facts. First, a difference in the optical components. PPG sensors use emitters and receivers in the *infrared* spectrum because it is better suited to capture blood flow. Smartphones, on the other hand, use white flashlights and cameras tuned for picture-taking rather than PPG collection. Second, a lack of add-ons for pressure control. PPG sensors have a finger-clip to control the user's pressure, but smartphones do not have such mechanical design. This chapter introduces methods to overcome these two limitations, re-purposing smartphones to capture PPG signals that are comparable to PPG sensors.

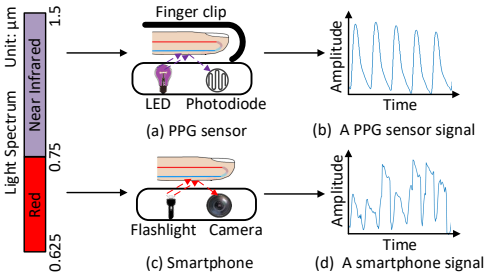


Figure 4.1: Cardiac signal obtained with a PPG sensor and a smartphone.

4.1. INTRODUCTION

As described in the prior chapter. Cardiac signals –as hard-to-forge biometrics– are being studied as potential alternatives for authentication systems [79, 49, 18]. Cardiac signals can be captured with a photoplethysmogram (PPG) sensor, which measures the changes in blood volume via the absorption of light. PPG sensors rely on two key factors: 1) the use of LEDs and photodiodes operating in *the near infrared spectrum*, and 2) *a steady pressure* guaranteed by a finger clip, as shown in Figure 4.1(a). These two properties allow for capturing stable cardiac signals, as illustrated in Figure 4.1(b). The flashlight and camera on smartphones can also be used to capture PPG signals, as shown in Figure 4.1(c). However, the signal quality degrades significantly, as depicted in Figure 4.1(d), because the required light spectrum and finger clip are no longer present. Some SoA studies using PPG signals *obtained from a camera* report unsatisfying performances, with an Equal Error Rate (EER) around 20% [124]. Although some studies have optimized part of the camera parameters to gather more stable cardiac patterns, such as in [49], they still need to collect PPG signals in a *controlled manner* to achieve a good performance. None of these studies have conducted a comprehensive analysis of the *camera parameter configuration* or the amount of *fingertip pressure* over the camera.

In this work, we improve the quality of PPG signals captured with smartphone cameras for authentication. Our study focuses on analyzing the **sensing process** (camera parameters and finger pressure) and on utilizing a bigger and more realistic **dataset**, as presented in Table 4.1. Overall, we provide the following contributions.

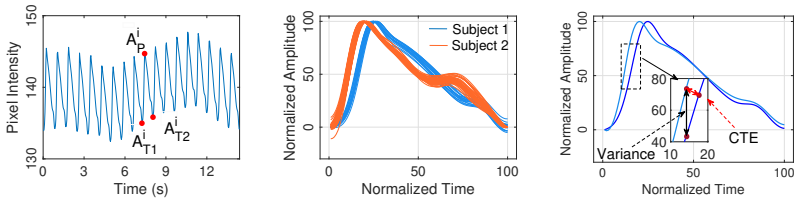
Contribution 1: Camera Configuration [Section 4.3]. Cameras are designed for taking pictures and recording videos. Using cameras to capture PPG information leads to poor signal quality as shown in Figure 4.1(d). We investigate and configure all camera parameters to render a PPG signal that is as similar as possible to the one captured with a dedicated PPG sensor.

Contribution 2: Pressure Control [Section 4.4]. Pressure plays an important role in signal distortion. Without a finger clip, SoA systems using cameras have no control over pressure. We investigate the impact of pressure on the signal's quality and provide feedback to the user to maintain an optimal pressure level. By doing so, we achieve stable PPG signals for every subject *without any hardware add-ons.*

Contribution 3: Thorough Evaluation [Sections 4.6 & 4.7]. We collect a dataset with 58 subjects. This dataset is bigger and more balanced than other datasets in the SoA [49, 125]. Moreover, we collect *single-session* and *double-session* datasets. The single-session

Table 4.1: Our contributions compared to the SoA: CardioCam [49] and CardioID [125]. The fields in orange depict the novelty of our study.

	Category	CardioCam [49]	CardioID [125]	CamPressID
Sensing	**Camera Configuration**	×	×	✓
	Finger Pressure	×	×	✓
Dataset	# Users	25	43	58
	# Female	6	16	22
	Age range (μ/σ)	25-33 (-/-)	12-79 (36.7/14.9)	15-80 (40.2/14.6)
	Interval (h)	0	0	≥ 7 hours



(a) PPG amplitude

(b) Subjects with distinct morphologies

(c) Similarity metrics comparison

Figure 4.2: The metrics and morphology description for PPG authentication.

dataset does not leave any interval of time between the testing and training phases, which is what the SoA does [125, 49, 124]); while the double-session dataset leaves at least seven hours between the testing and training phases, providing a more realistic and challenging scenario to test authentication. To showcase the benefit of improving the sensing process, we test our framework using the same authentication method as the SoA [125]. The results show that our more principled way of gathering PPG signals increases the average balanced accuracy (BAC) of CamPressID by more than 10% compared to the SoA (10% higher for the double-session dataset, and about 20% higher for the single-session dataset). Furthermore, in the single-session dataset, the BAC gets above 90%, which approaches the performance of a system using a dedicated PPG sensor.

4.2. SYSTEM OVERVIEW

To improve the quality of the signals captured from cameras, we propose two methods that can be added to existing authentication solutions: *camera configuration* and *pressure control*. First, we analyze the effect of all camera parameters in order to optimize the smartphone for PPG signal collection (as opposed to optimizing it for pictures and videos). Second, the system asks the subject to use different pressure levels. Based on these levels, the system identifies the optimal pressure and provides feedback to the subject to maintain that level. After performing these two steps (camera and pressure optimization), CamPressID processes the PPG signals based on authentication methods already reported in the SoA [125].

A high authentication accuracy of PPG-based systems requires two conditions: i) the shape of PPG periods *from the same subject* should be similar (short intra-class distance), and ii) the shapes of PPG periods *from different subjects* should be distinct (long inter-

Table 4.2: Camera configurations in the SoA [124, 49, 125]. The table includes the default values obtained for light and dark skin. The parameters in orange are optimized in CamPressID. (DC: default configuration; γ: gamma correction; DPS: dynamic pixel selection; –: hardware-limited; ×: available but not controlled; ✓: controlled but the value is not reported).

Configuration	Smartphone	Frame rate	Frame resolution	Flashlight	Aperture	ISO	Shutter speed	White balance	γ	ROI
[124]	Iphone X	240 fps	1280*720	×	–	✓	✓	✓	×	×
[125]	Iphone 7	60 fps	×	×	–	20	200	×	×	Selected pixels
[49]	Iphone 7	60 fps	1280*720	✓	–	✓	×	×	×	DPS
DC for light skin	Moto G7 plus	30 fps	320*240	–	–	100	1/729	4474 K	×	All pixels
DC for dark skin	Moto G7 plus	30 fps	320*240	–	–	100	1/611	4858 K	×	All pixels
CamPressID	Moto G7 plus	30 fps	320*240	–	–	100	1/400	6600 K	1	Central 1/4

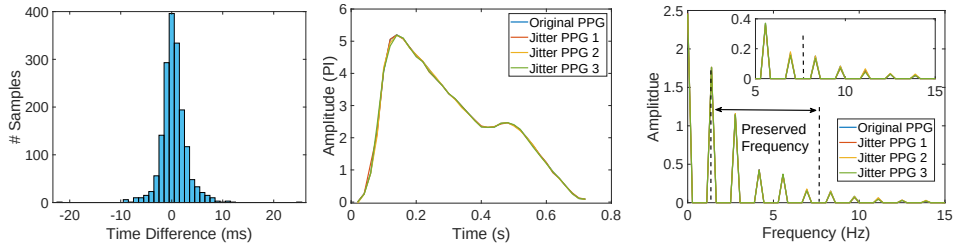
class distance). Considering the cardiac structure of subjects stays almost the same, the intrinsic inter-class distance is fixed. Thus we only deal with the intra-class distance. In this work, *we refer to the shape of PPG periods as* ***morphology***. If the morphologies gathered for the subjects are noisy and/or variable, it is challenging to differentiate them. Therefore, in our system, we utilize two quality indicators for PPG signals: 1) a high signal-to-noise ratio (SNR) and 2) a stable morphology for every individual subject.

For the first indicator, the SNR can be mapped to PPG amplitude. Within the duration of a PPG signal, the amplitude captures the average peak-to-peak values of all periods. For example, in Figure 4.2a, the PPG amplitude can be calculated as $\frac{1}{n} \times \sum_i^n (A_P^i - (A_{T1}^i + A_{T2}^i)/2)$, where n indicates the number of periods in the signal. *For the second indicator*, a stable morphology is reflected by the similarity among periods. To visualize the similarity of signals, we overlap all periods in Figure 4.2a to generate the blue normalized curves in Figure 4.2b and we add the normalized curves of a second subject (orange) to show that the cardiac signals can be unique to each subject[1]. To measure the similarity between periods, we use the *cross track error (CTE)*, which calculates the shortest Euclidean distance from points on one signal to a reference signal. For the remainder of the paper, our CTE calculations (for a given subject) use the average morphology of all periods as the *reference signal*. Note that CTE is a better similarity metric than variance, because the variance can penalize heavily small misalignments between periods, as depicted in Figure 4.2c.

4.3. Optimization of Camera Configuration

Cameras are designed to take pictures and record videos. When they are exploited to collect PPG data, it is inevitable to compromise the signal quality due to inappropriate camera configurations. This will lead to a degradation in authentication performance. In Table 4.2, we summarize the camera configurations used in the SoA. We can see that a comprehensive study of the key parameters for biometric applications is still missing. In this section, we provide a thorough analysis of such parameters to enhance the quality of PPG signals.

[1] This overlapping process consists of two steps: First, we normalize the duration and amplitude of each period to 100. Second, we set the starting point of all periods to (0,0) to align them. We can also use other reference points, but given that the alignment does not change the morphology of periods, the performance is not affected. Thus, we just pick the starting point.



(a) Histogram of the frame jitter (b) PPG signal with jitter in time domain (c) PPG signal with jitter in frequency domain

Figure 4.3: The analysis of frame rate jitter on the PPG signal.

4.3.1. CAMERA PARAMETERS OPTIMIZATION

There are mainly three camera configurations that affect getting high-quality cardiac signals: video recording, light exposure, and image processing. Next, we analyze the impact of each part.

VIDEO RECORDING

Video recording consists of two essential parameters: frame rate and frame resolution. With PPG sensors, each sample captures the reflected *near-infrared spectrum* of the finger's blood. With a camera, an entire frame is needed to provide one sample. The color of the multiple pixels in one frame are combined to capture the reflected *red spectrum* from the finger.

The **frame rate** is equivalent to the sampling frequency of the PPG sensor. And there are two parameters we need to consider in the frame rate. First, we need to guarantee that the sampling frequency (frame rate) is sufficient to obtain the necessary spectral information from a cardiac signal. Second, contrary to a PPG sensor, the camera frame rate suffers from a slight jitter. The interval between two frames is not constant. This jitter may distort the shape of the PPG signal, which could lead to unstable morphologies.

Regarding the necessary sampling frequency, according to [92], sampling frequencies above 25 Hz do not provide any extra information. Hence, the customary rate of 30 fps in most commercial smartphones is sufficient. Regarding the jitter, to understand its impact, we collect 1731 periods using a smartphone at 30 fps. The results are shown in Figure 4.3a. The jitter's standard deviation is 2.83 ms and 95% of the samples fall within 5.67 ms. To quantify how these statistical variations distort a cardiac signal, we obtain a *'clean'* period from the PPG sensor and add the *random jitter* shown in Figure 4.3a. This process allows us to emulate PPG signals with jitters, as shown in Figure 4.3b. We can see that jittered signals almost fully overlap with the original one. If we check the frequency domain in Figure 4.3c, the responses are also nearly the same for the range of interest $[2f, 5.5f]$ (described in Section 3.4 of Chapter 2), where f is the heartbeat rate. Thus, in both the time and the frequency domains, jitter causes a negligible distortion, and consequently, we choose 30 fps as the frame rate.

The **frame resolution** determines the number of pixels on each frame. Since we have to combine information from multiple pixels to obtain a single PPG sample, we pick the lowest frame resolution (320*240) in our camera to reduce the computational complex-

Table 4.3: PPG amplitude (in pixel intensity) under different camera settings.

	Shutter Speed			
ISO	**1/200**	**1/400**	**1/800**	**1/2000**
100	10.03	**11.13**	5.59	2.31
200	4.74	6.03	10.67	3.27
400	2.04	4.28	9.52	8.15
800	0.01	2.74	7.40	7.42

ity. In Section. 4.3.2, we describe how we determine a region of interest within a frame to reduce the number of pixels involved in the computation of the PPG sample.

LIGHT EXPOSURE

Four key components affect light exposure: flashlight, aperture, ISO, and shutter speed.

The **flashlight** determines the intensity and spectrum of the light impinging on the fingertip, and the **aperture** controls the opening size of a camera, determining the amount of incoming light. Most smartphones, however, do not allow modifying these components.

The **ISO** and **shutter speed** are the only parameters that can be configured in light exposure. Choosing the right combination can achieve a high PPG amplitude. The ISO controls the pixel's sensitivity to light, and the shutter speed controls the shutter open interval. During the open interval, cameras integrate the light energy received on each pixel. We should assess different ISO and shutter speed values and select the best combination for the system. Taking the camera of the Moto G7 Plus, for example, both ISO and shutter speed contain four values: ISO: 100, 200, 400, 800; shutter speed: $\frac{1}{200}, \frac{1}{400}, \frac{1}{800}, \frac{1}{2000} s$. To measure the effect of these parameters, we use ten subjects to collect 120 seconds of PPG signals for each one of the sixteen ISP/shutter speed combinations. To determine the best combination, we adopt PPG amplitude (SNR) as the metric. The experiment results are shown in Table 4.3. We can see that some light exposures are either too high (ISO: 800 and shutter speed: $\frac{1}{200}$) or too low (ISO: 100 and shutter speed: $\frac{1}{2000}$) resulting in a low PPG amplitude (0.0102 and 2.3122, respectively). In high light exposure, PPG signals saturate most of the time, rendering almost zero PPG amplitude. In low light exposure, pixels are insensitive to light change, which renders a low PPG amplitude. Among all the combinations, ISO 100 and shutter speed $\frac{1}{400}$ provide the highest amplitude. Therefore, for the smartphone Moto G7 Plus, these values are optimal. On other smartphones, the proper ISO and shutter speed configurations can be identified similarly.

IMAGE PROCESSING

Due to the differences between cameras and human eyes, a camera-captured *raw* picture is distinct from the perception of human eyes. To mimic human eyes' perception, engineers introduce the *white balance* and *gamma correction* parameters to modify the *raw* RGB values in a picture. These modifications will inevitably distort camera PPG signals, which results in a high variances (high CTE).

White balance adjusts the color intensity to render "correct" colors to human eyes. It depends on the color temperature. Figure 4.4 shows actual RGB values for the white color in different color temperatures [126]. We can see that except for temperatures in

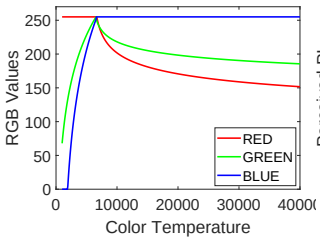


Figure 4.4: The relationship between color temperature and RGB values [126]

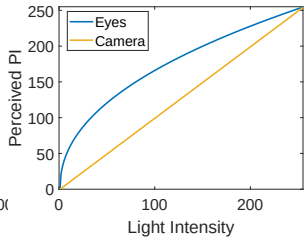


Figure 4.5: Gamma correction delivers the perceived human-eye intensity

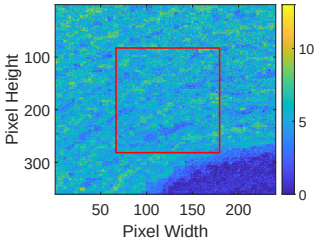


Figure 4.6: Selecting the region of interest (red rectangle)

[6500K, 6600K], this parameter attenuates the intensity of two primary colors. Hence, we set the white balance to 6600K, allowing RGB values to reach 255 under white light.

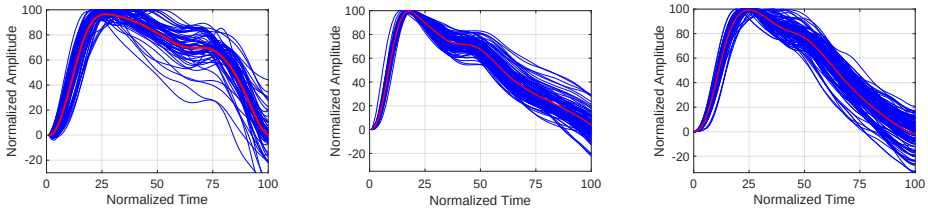
Gamma correction is a non-linear mapping that transfers the light intensities in cameras to those in human eyes. A camera's response to light changes is linear, while human eyes are non-linear, being more sensitive to light changes in a dark environment as shown in Figure 4.5. To approximate the perception of human eyes, smartphones apply gamma encoding to map the camera values (yellow line) to the eyes perception (blue curve) using the equation $Y = 255\,(X/255)^{\gamma}$, where X is the original pixel intensity, γ is a constant and Y is the pixel intensity captured by human eyes. The gamma correction in most cameras use a value of $\gamma = 0.45$ [127]. From the equation, we can observe that the non-linear mapping for human eyes amplifies the light change by *different factors* at *different intensity levels*. To restore the linear mapping, we set γ to 1.

4.3.2. REGION OF INTEREST SELECTION

Although a camera has a large matrix of pixels, only a fraction captures the changes caused by cardiac signals in a clear manner. Some studies select the best areas by analyzing each pixel [49], but such a process is too demanding to attain a real-time response on the smartphone. In this work, we use *only* the central region of each frame. The area of this region of interest (RoI) is 1/4 of the whole frame. To identify this ROI, we measure the difference between the frames at the peak and start (valley) of one period. The difference between these two frames is shown in Figure 4.6. We can see that the bottom right corner has zero difference. This is because that corner is close to the flashlight and thus strong light saturates those pixels at all times. A similar phenomenon is observed with other smartphones. Besides the flashlight effect, the outer regions in general are susceptible to ambient light changes caused by finger movement or pressure change. Due to these reasons, we decided to select the central part (1/4) of the frame as the RoI. Furthermore, to exclude noisy pixels inside the RoI, we use α-trimmed mean filtering [128], with $\alpha = 0.1$ to consider only pixel values between the 10th and 90th percentile.

4.3.3. PRELIMINARY EVALUATION

Now we present a preliminary evaluation to highlight the benefits of optimizing the camera parameters. We use a Motorola G7 Plus with an optimal configuration according to the last row of Table 4.2. Our evaluation considers three setups: (i) the default camera



(a) Overlapping periods under default configuration (CTE=4.6735)

(b) Overlapping periods under our optimized configuration (CTE=3.4207)

(c) Overlapping periods under PPG sensor (CTE=3.4426)

Figure 4.7: Comparison of signal stability. The red lines represent the average of all the overlapping periods.

configuration (optimized for photography like [129]), (ii) our optimized camera configuration, and (iii) a dedicated PPG sensor (SDPPG from APMKorea).

To measure the improvement, we use PPG amplitude and CTE as our metrics. For PPG amplitude, we compare the differences between the two camera configurations[2]. The default configuration (ISO 100 and shutter speed $\frac{1}{721}$ for light skin; and shutter speed $\frac{1}{611}$ for dark skin) has a pixel intensity of around 5.6. Our camera configuration, instead, reaches a pixel intensity of about 11. Thus, our optimized configuration doubles the PPG amplitude.

The CTE measures the similarity between periods. As an example, we demonstrate the CTE comparison for one subject, as shown in Figure 4.7. We can see that the CTE under the default configuration is higher and the periods diverge, Figure 4.7a, making it hard for a system to learn the morphologies of the signals. The CTE under our camera configuration is reduced by 27% and the periods converge, leading to more stable morphologies, c.f. Figure 4.7b. Furthermore, compared to the signals obtained with the PPG sensor, Figure 4.7c, the CTE of our camera configuration is only slightly higher. This demonstrates that our configuration can compensate for the camera disadvantages to obtain a close resemblance to the PPG sensor. More evaluations of our camera configuration will be presented in Section 4.7.1.

4.4. OPTIMIZING THE FINGER PRESSURE ON THE CAMERA

PPG signals are measured by monitoring the changes in blood flow. The pressure of fingertips on the sensor's surface can change the blood flow because it alters the cross-section area of the fingertip's blood vessels. In this section, we study the impact of pressure on *PPG morphologies*.

4.4.1. PRESSURE INFLUENCE

We study the influence of pressure through experiments. We put the smartphone on a scale with the rear camera facing up and ask a subject to put his fingertip on the camera. Initially, no active pressure is applied to the camera. After that, every 45 seconds we add a 3 Newton (i.e., 3 $kg \cdot m/s^2$) pressure to the subject's fingertip until the pressure reaches

[2] We cannot compare the signal amplitude between the sensor and camera because they exploit different spectrums (infrared/visible), process the cardiac signal in different manners and use different units. Thus, we only compare the amplitude between different camera configurations.

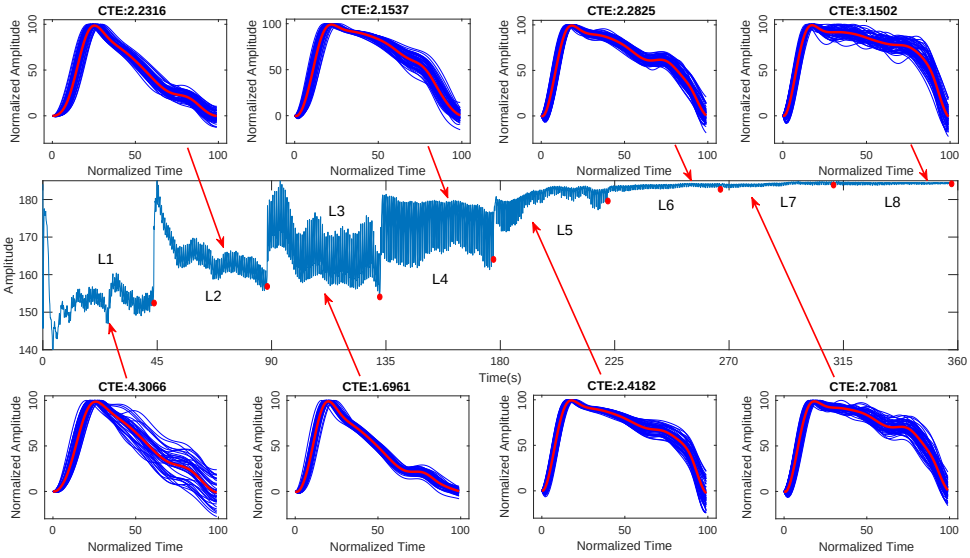


Figure 4.8: Pressure analysis. We apply eight pressure levels on the fingertip of a subject, from 0 N (no pressure) to 21 N (2.1 kg), in steps of 3 N (300 g). Each pressure level lasts for 45 seconds. Red dots on the PPG signal indicate the end of a pressure level.

Table 4.4: Notations used for the pressure levels.

Notations	L1	L2	L3	L4	L5	L6	L7	L8
Pressure (newton)	0	3	6	9	12	15	18	21

21 N. The notation of each pressure level is shown in Table 4.4. This range can simulate the situation from no pressure to overpressure. Figure 4.8 illustrates the pressure's impact on the amplitude and morphology of PPG signals. The corresponding CTE and amplitude values are shown in Figure 4.9.

Under pressure level L1 (unstable surface contact before the first red dot in Figure 4.8), parts of the fingertip's surface detach from the camera when a subtle movement happens. These dynamics affect the light intensity on the camera. Thus, we can see that, within level L1, the PPG amplitude is low and CTE is high (non-ideal periods).

Under pressure levels L2 and L3 (from the first red dot to the third red dot), the added pressure improves the PPG amplitude and CTE. Especially under pressure L3, the resulting amplitude is close to the highest value and the CTE is the lowest. These convergent periods with high SNR will help the system extract the required features.

Under pressure level L4, the amplitude reaches the maximum value, but the CTE increases. Furtermorre, contrary to the triangle-shaped morphology under pressure L2 and L3, the morphology under pressure L4 has a trapezoid shape, which may change the features derived from a cardiac signal.

After pressure level L4, stronger pressure suppresses the vessels, leading to low amplitudes. The low amplitude makes periods vulnerable to noise, amplifying the discrepancy (CTE) between periods. Besides having a low amplitude and high variance (CTE), the morphologies' shape become more "square". These distorted morphologies are far

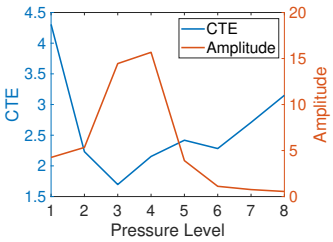
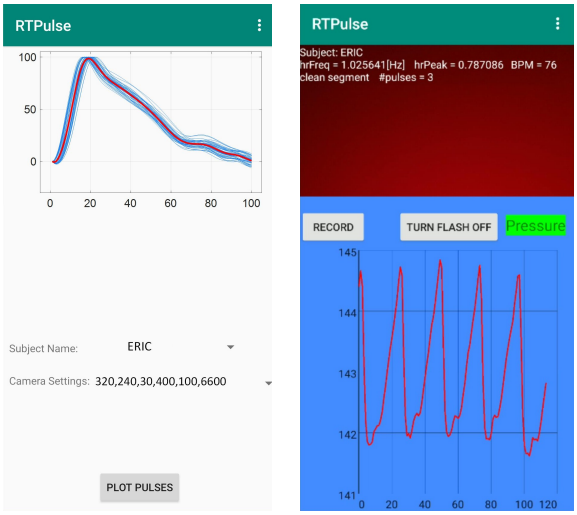


Figure 4.9: The changes in CTE and amplitude at different pressure levels.

Table 4.5: Performance comparison between the triangle and trapezoid morphologies.

	Triangle	Trapezoid
BAC	0.9347	0.8589



(a) Interface with overlapping periods after a data gathering session (b) Online feedback for optimal pressure during signal collection

Figure 4.10: Our smartphone APP for PPG signal collection.

away from reflecting the actual features of the cardiac system.

4.4.2. PRESSURE CONTROL

From the pressure analysis, we see that a subject generates multiple morphologies under various pressures. Thus, we should first determine the best pressure level for authentication, and then provide real-time feedback to the user to maintain that pressure level during the training and testing phases.

From Figure 4.9, we observe that there are two pressure levels with a high amplitude and a low variance (CTE): L3 and L4. The key difference between these levels is that one morphology is triangular (L3) and the other trapezoidal (L4). Hence, a central question is: *what type of morphology is better suited for authentication with smartphone cameras?* To determine the optimal pressure level (and morphology), we perform a preliminary experiment with ten subjects. Each one of these subjects goes through the process shown in Figure 4.8 in order to gather their respective pressure levels L3 and L4. After that, for each pressure level, we divide the data into two parts: the first 80% is for training and 20% for testing. Then, to compare the performance of both pressure considering the final application (authentication), we adopt the same feature extraction and classification methods in the SoA [125], which are described in more detail in Section 4.5). The result in Table 4.5 demonstrates that the triangle shape morphology performs better, with a BAC of 0.93 compared to a BAC of 0.86 for the trapezoidal shape. Due to this reason, we use L3 (6 N) as the optimal pressure level.

To assist users in stabilizing their fingertips at their best pressure *during the data collection* sessions, we design the Android APP shown in Figure 4.10. The App has two main steps. First, to obtain the L3 level for each user in our dataset (53 users in total), we ask them to increase the pressure level until we obtain the amplitude and CTE for level 3,

Table 4.6: Our Methods compared to the SoA: CardioCam [49] and CardioID [125].

	Category	CardioCam [49]	CardioID [125]	CamPressID
Method	Signal processing	SoA_cam.proc	SoA_dual.proc	SoA_dual.proc
	Authentication	SoA_cam.auth	SoA_dual.auth	SoA_dual.auth

Table 4.7: Subject details for different PPG authentication systems.

Studies	**Sensor**	**# Users**	**# Females**	**Age range (mean/STD)**
[129]	PPG sensor	12	4	22–51 (-/-)
[86]	PPG sensor	42/32	-	- (-/-)
[130]	PPG sensor	17	-	23–30 (-/-)
[125]	PPG sensor	35	12	10-75(28.4/14.0)
	Camera	43	16	12-79(36.7/14.9)
[124]	camera	15	2	- (-/-)
[49]	camera	25	6	25–33 (-/-)
CamPressID	camera	**58**	**22**	15–80 (40.2/14.6)

c.f. Figure 4.9a. Then, to gather the data for the training and testing phases, we provide real-time feedback to help the user maintain that optimal pressure, c. f. Figure 4.9b. Second, during the data gathering process of the testing and training phases, the APP gathers information over a window of five seconds and checks two criteria for the periods in that window: (i) their amplitude should reach 80% of the *average amplitude obtained for L3 during the pressure evaluation*; and (ii) the CTE with respect to the *average L3 signal* (red line in Figure 4.9a) should be less than the *average CTE obtained during the pressure evaluation*. If 75% or more of the periods in the five-second window satisfy these two criteria, the App considers the data to be valid and stores it for training or uses it for testing. Else, the App will provide feedback to the user to increase or decrease the pressure until the *sign-pressure* turns green on the interface, middle right sign in Figure 4.9b.

4.5. AUTHENTICATION METHOD FROM SOA

Thus far, camera-based studies used for authentication have focused on proposing novel feature extraction and classification methods [49, 125]. Our work in this chapter does not target those areas, our aim is to improve the camera configuration and pressure levels to showcase that enhancing the data gathering process results in direct gains in overall performance. Due to that reason, we use the features and classification framework presented in Chapter 2, Table 4.6. The details of the signal processing method and authentication method are shown in Sections 3.3 and 3.4 of Chapter 2 and Section 3.5 of Chapter 2, respectively.

4.6. DATASET CONSTRUCTION

To evaluate the performance of our data gathering framework, we need suitable datasets. Existing datasets such as the ones in [49, 125] have a small population and a narrow age range, which are not enough for a thorough evaluation. In this work, we collect PPG signals from 58 subjects (22 are women). The age range is 15–80; and the mean and standard deviation are 40.2 and 14.6 years, respectively. The details are listed in Table 4.7.

Compared to other datasets [49, 125], our subject population (58) is the largest and our age range is wide enough to cover most types of smartphone users (from teenagers to old people).

For all subjects, we use a Motorola G7 plus to extract their PPG signals. The data collection process considers two setups: *single-session* and *double-session*. The *single-session* setup emulates the setups used in the SoA [49, 125]: there is *no interval* between the collection of data used for training and testing. This setup allows us to perform a direct comparison with the SoA. Under this setup, we extract *4-minutes of PPG signal* from each subject. The first 80% cycles of this signal are used for training and the remaining 20% for testing.

The *single-session* setup, however, differs from realistic login systems, which generally have an interval between the data registration (training) and access requests (testing). Hence, we introduce the *double-session* setup to emulate more realistic situations. For the double-session setup, we set the interval between training and testing to *at least 7 hours*. The 4-minute signal gathered for the single-session setup is used for training, and after at least 7 hours, we collect an *additional* 1-minute PPG signal for each subject for testing. Note that for the double-session setup, the ratio between training and testing is still the same as for the single-session setup (80/20).

It is important to highlight that no prior study has considered these more demanding conditions. Due to the interval introduced in the double-session setup, physiological and psychological changes between the training and testing are considered. These changes affect the stability of each subject's morphology. Hence, the data in the double-session setup is more challenging for authentication. It is also important to note that, since the training data come from the single-session recordings, we can quantify the influence of physiological and psychological by comparing the result between these two setups.

For each setup, single and double, we construct three datasets to assess the individual improvements of our camera configuration and pressure control modules [3]:

- *Default* dataset: The data collection is done with *the default camera parameters* and *without pressure control*. This dataset represents the data collection of most SoA authentication systems, like [129].
- *Config* dataset: The data collection is done with our *optimized camera configuration* (Section 4.3).
- *ConfigPress* dataset: The data collection is done under our *optimized camera configuration* and *with pressure control* (Section 4.4).

4.7. EVALUATION

In this section, we evaluate the performance of our approach, CamPressID, and compare it against two baselines: CardioID (SoA_dual) [125] and CardioCam (SoA_cam) [49] selected from SoA camera-based authentication systems in Table 4.7. These two baselines can allow us to assess the improvement of our camera configuration and pressure

[3]The data collection activities related to these datasets have been approved by the Human Research Ethnic Community (HREC) at our University.

Table 4.8: The performance comparison of multi- and single-morphology methods in **single-session** case.

Method/System	Dataset	Acquisition Rate	N_{DS}[3]	PPG Amplitude	CTE
Multi-morphology[1]	Default	0.73 (10167/13920)	55	4.03	4.82
	Config	1.06 (14774/13920)	57	9.41	3.74
	ConfigPress	1.13 (15755/13920)	58	9.41	2.32
Single-morphology[2]	Default	0.25 (3468/13920)	40	4.98	4.73
	Config	0.62 (8695/13920)	48	9.44	3.34
	ConfigPress	0.79 (10977/13920)	48	9.42	2.32

[1] Studies exploit multiple morphologies in PPG authentication, such as CardioID [125] and CamPressID in this chapter.
[2] Studies exploit only one morphology in PPG authentication, such as Seeing Red[124] and CardioCam [49].
[3] The number of detectable subjects.

Table 4.9: The performance comparison of multi- and single-morphology methods in **double-session** case.

Method/System	Dataset	Acquisition Rate	N_{DS}	PPG Amplitude	CTE
Multi-morphology	Default	0.73 (12616/17400)	54	4.3	5.49
	Config	1.07 (18586/17400)	57	10	4.22
	ConfigPress	1.12 (19548/17400)	58	9.52	2.75
Single-morphology	Default	0.25 (4284/17400)	39	5.07	5.18
	Config	0.62 (10825/17400)	47	9.94	3.78
	ConfigPress	0.78 (13600/17400)	50	9.54	2.71

control module across different systems. In addition, SoA_dual also allows us to compare the performance between our CamPressID and the PPG sensor, which can directly demonstrate the effectiveness of our CamPressID.

4.7.1. DATASET EVALUATION

In data collection, there are two branches: *single-morphology* and *multi-morphology*. The *single-morphology merely* considers one morphology of PPG signals. This branch is the mainstream in PPG authentication systems in SoA [124, 49]. In this work, we re-implement the data collection in CardioCam (SoA_cam)[49] as the representative of *single-morphology* systems. Our CamPressID is representative of *multi-morphology* systems. After applying these two kinds of systems to our camera datasets, we can generate Tables 4.8 and 4.9.

There are four metrics for us to evaluate our camera datasets: *acquisition rate, number of detectable subjects, PPG amplitude* and *CTE*.

- *Acquisition Rate*: This metric measures how many valid PPG periods the authentication system can collect per second. The valid periods should contain the dominant morphology h_1'', h_2'', h_3''. This metric indicates the real-time performance of a system.
- *Number of Detectable Subjects (N_{DS})*: This metric measures how many subjects can be detected for more than 20 training periods and 5 testing periods under an authentication system. This metric indicates the pervasiveness of the system (including more users).
- *PPG Amplitude*: This metric indicates the overall SNR of the collected PPG signal. The detail of this metric has been introduced in Section 4.2.

- *CTE*: This metric indicates the overall divergence of the collected PPG signals for each subject. The detail of this metric has been introduced in Section 4.2.

For a good real-time authentication system, the first three metrics should be as high as possible and the CTE should be as low as possible. According to all metric values in these two tables, we have three main insights.

Insight 1: The high difficulty in the double-session datasets. The 7-hour interval introduced in the double-session datasets increases the authentication difficulty substantially. The inconsistency of physiological and psychological status between the training (before the interval) and testing set (after the interval) induces the divergence of PPG signals on each subject. This reflects in the CTE metric, in which the higher value shows subjects have more divergence among their own periods. Compared with single-session setup, the CTE values in double-session are about 10%~20% higher, while other metrics are almost the same. For example, the Default dataset under multi-morphology in the single-session setup has 4.82 CTE, which is lower than that in the double-session setup (5.49). Consequently, the datasets in the double-session setup are more difficult.

Insight 2: The improvement of real-time performance and user inclusion on the multi-morphology system. Contrary to the single-morphology system that has stringent requirements for data collection to filter out periods with non-conforming morphology [4], the multi-morphology system collects periods with various morphologies. Thus, it gains more but a little more diverging periods (higher CTE). For example, in Table. 4.9, comparing the Config datasets between both systems, the acquisition rate and N_{DS} are 44%, 16% lower respectively in the single-morphology system. The low acquisition rate will constrain the real-time performance of the system. Furthermore, the single-morphology system cannot detect all 58 subjects in N_{DS}. This shows that the single-morphology system cannot serve all kinds of users. Whereas, the multi-morphology system has good real-time performance (1.12 periods/second) and perfect user inclusion (58/58=100%).

Insight 3: the improvement of signal quality in terms of the camera configuration and pressure control modules. Regardless of the setups and systems, the camera configuration and pressure control modules improve the signal quality comprehensively, which reflects in our metrics. The camera configuration module contributes to all metrics, while the pressure control module focuses on the CTE. For instance, in Table 4.8, for the multi-morphology system, the improvement from theDefault to Config dataset are remarkable: Acquisition speed increases 45%; N_{DS} increases 4%, PPG amplitude increases 133%; and CTE reduces 32%. While the pressure control module aims at stabilizing the morphology of periods. Hence its CTE declines 38% further from the Config dataset.

4.7.2. AUTHENTICATION EVALUATION

In this section, we aim to address two primary questions. Firstly, can our camera configuration and pressure control module be beneficial for *various camera systems* across *different setups*? Secondly, how close can our *camera system* CamPressID approach the performance of *a system with a PPG sensor*?

[4] The morphology and feature information is provided in Figure 8 in [49].

To analyze these two questions, we generate Figure 4.11 to display the authentication performance across different systems, namely, our CamPressID, CardioID [125] and CardioCam [49]) and different setups, namely, the single-session and double-session setup.

Before interpreting the results, we wish to clarify three points. Firstly, for the first question, we generate results from all systems, *except CardioID (SoA_dual)*, on our Default, Config and ConfigPress datasets under both setups. This is because our CamPressID and SoA_dual share the same feature collection and authentication methods, as outlined in Table 4.6, thereby producing identical results with the same dataset. In this case, we can just display the result of CamPressID to represent both results.

Secondly, for the second question, we directly employ reported results from SoA_dual as the references for a camera system (SoA_dual.CAM) and *a system with a PPG sensor (SoA_dual.PPG)*. Based on the subjects' details presented in Table 4.7, both the camera and PPG sensor datasets in SoA_dual [125] contain a sufficient number of subjects and wide age distributions, ensuring a robust generalization of their results in our single-session setup.

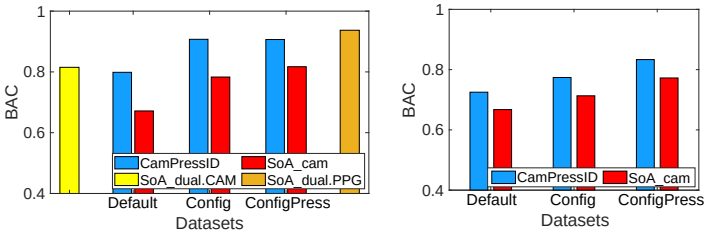
Thirdly, for the results of SoA_cam, we generate them by re-implementing the entire processing pipeline of SoA_cam in both setups (Figure 4.11). This is because the small number of subjects and narrow age distribution in SoA_cam [49] cannot guarantee a good generalization of its result.

THE IMPROVEMENT OF OUR CAMERA CONFIGURATION MODULE

Thanks to the signal quality enhancement by our camera configuration module, we can observe the performance improvement from the Default to Config datasets across different systems under both setups in Figure 4.11.

Under the single-session setup, our camera configuration module can provide about 10% BAC boost for both systems. On the Config dataset, SoA_cam approaches 80% BAC, while CamPressID exceeds 90% BAC. In addition, among SoA_dual.CAM, CamPressID on the Default and Config dataset, we can observe the positive correlations between the level of camera configuration and the performance. Considering that CamPressID and SoA_dual share the same methods of processing and authentication in Table 4.6, the result difference among those three datasets derives from their diverse signal quality, which is caused by various levels of camera configuration. CamPressID on the Default dataset has no camera parameters tuning, thus its result is the lowest. SoA_dual.CAM has some camera configuration, hence its result ranks in the middle. With the comprehensive camera configuration, CamPressID on the Config dataset performs the best. This performance comparison demonstrates the significance of camera configuration.

Under the double-session setup, our camera configuration module merely contributes about 5% BAC boost for both systems. This setup introduces physiological and psychological variations between the training and testing sessions. In each subject, periods from the training session are more likely to have a different morphology from periods from the testing session. Therefore, datasets under the double-session setup are more challenging. Therefore, the overall performances for both systems under this setup are about 7% inferior to those under the prior setup. This harsh environment also attenuates the capability of our camera configuration module. In this case, our camera configuration module cannot provide a comparable improvement as the prior setup.



(a) Performance comparison among CamPressID, CardioCam [49] and CardioID [125] under the single-session setup

(b) Performance comparison between our method and CardioCam [49] under the double-session setup

Figure 4.11: The performance comparison across datasets and systems. SoA_dual.CAM and SoA_dual.PPG stand for the camera and PPG sensor results in [125], respectively.

4

THE IMPROVEMENT OF OUR PRESSURE CONTROL MODULE.

Apart from the signal quality, the period homogeneity within each subject is another key point for authentication systems. In our system, we introduce our pressure control module to handle the period homogeneity. In general, the performance improvements are visible from the Config to the ConfigPress dataset under different systems across various setups. Furthermore, under the single-session setup, our CamPressID on the ConfigPress dataset can approach the performance as the SoA_dual.PPG, which is the authentication system with a dedicated PPG sensor.

Under the single-session setup, compared with about 10% improvement from our camera configuration module, our pressure control module contributes less: SoA_cam gains around 3% BAC increase, while our CamPressID's performances stay nearly the same, from the Config to ConfigPress dataset. Actually, CamPressID still gains a slight BAC boost on the ConfigPress dataset, since the N_{DS} is higher on this dataset in Table 4.8. Moreover, the result of CamPressID on the ConfigPress dataset approximates that of SoA_dual.PPG, which employs a dedicated PPG sensor to collect PPG signals. The approximation cannot be closer, due to the high difficulty of the ConfigPress dataset. The dataset difficulty reflects in two factors: the period homogeneity (CTE) and the number of subjects. Compared with SoA_dual.PPG, although the ConfigPress dataset is 46% lower in CTE (4.29 vs. 2.32), its subject number is 66% more (35 vs. 58). In this case, our ConfigPress dataset is more difficult.

For the negligible improvement on CamPressID from the Config to ConfigPress dataset, we have three reasons: First, the performance limit of our system on single-session datasets should be the SoA_dual.PPG, as we explained in last paragraph. Thus the pressure control module cannot improve the performance further, while SoA_cam can gains some increase under the same setup. Second, the majority pressure on the Config dataset is the best pressure that we use in the ConfigPress dataset. To verify this, we check the pressures that subjects apply on camera on the Config dataset and find that 62.1% (36/58) of subjects apply the best pressure in the majority of the data collection time.

Under the double-session setup, there is more space for our pressure control module to showcase. Our module can provide around a 6% BAC increase for both systems, which is much higher than increases under the single-session. On the ConfigPress dataset, SoA_cam only approaches 80% BAC and Our CamPressID reaches 83% BAC. These lower

authentication performance reflects the difficulty of the double-session setup.

SUMMARY.
Based on all the above results, we could make the following summaries. Our camera configuration and pressure control module can enhance the performance of *different authentication systems under various setups* comprehensively, including system real-time response (acquisition rate), user inclusion (N_{DS}) and authentication accuracy. For our camera configuration module, it provides the significant accuracy gain under the single-session setup. For our pressure control module, it provides the more contribution under the double-session setup. Moreover, under the single-session datasets, the performance of our camera system, CamPressID, can approximate that of the dedicated PPG sensor system. Under double-session datasets, our CamPressID could not achieve 90% BAC because of the impact of physiological and psychological change. To ameliorate this problem, one could combine more periods for testing or use the spectral features, at the cost of additional overhead on training and processing delay.

4.8. RELATED WORK

In Section 3.7 of Chapter 3, we have already discussed studies related to camera-based PPG systems for healthcare and authentication. Next, we describe in more detail the limitations of camera-based authentication systems regarding parameter optimization and pressure control.

Camera-based PPG authentication systems. CardioID [125], as our other baseline, considers a PPG sensor and a smartphone camera for PPG authentication. In the camera case, they did not optimize the camera configuration. CardioID achieves a BAC of 93% with the PPG sensor and 82% using the camera. In CamPressID, we follow the signal processing and feature collection of CardioID, but we optimize the camera configuration and find the optimal pressure for improving the system performance. Besides, Seeing Red [124] studies the influence of cross-session authentication accuracy on camera PPG authentication. The median interval between training and testing sessions is 32.1 hours. They adopt both fiducial and spectral features and achieve 20% EER in authentication with 15 subjects. PPG^{2}Live [129] includes the PPG as an auxiliary authentication factor for face authentication. It combines the remote face and the fingertip PPG signal on smartphone cameras to perform authentication. It can achieve 98.7% accuracy with 12 subjects. Due to their low subject numbers [124, 129], we do not use them as baselines in this work.

Pressure control on cameras for better PPG signals. There is no paper studying the pressure influence on camera-based PPG authentication. We only find two papers for healthcare taking the pressure into consideration during the PPG measurement. Chandrasekhar et al. [97] study the influence of contact pressure on a PPG sensor for blood pressure measurement. They use the pulse arrival time to estimate the blood pressure. They found that the contact pressure leads to at least around 11 mmHg of blood pressure error. PhO$_2$ [131] uses a smartphone mounted by an add-on to measure blood oxygen level. The estimation error is 3.5% compared to FDA-approved gold standard pulse oximetry. To mitigate the negative impacts of user behavior on measurement, they design a light-based pressure detection algorithm to monitor the PPG signal amplitude. In

these papers, instead of morphology, they only focus on a part of the PPG signal. Moreover, their devices are tailored for the infrared spectrum. Therefore, we can not use their studies for authentication on normal smartphone cameras.

4.9. CONCLUSIONS

The lack of optimal camera configurations and pressure control prevents current camera-based PPG authentication systems from achieving high accuracy. In CamPressID, we propose solutions to these two limitations , which allow us to re-purpose smartphone cameras for PPG collection. We recruited 58 subjects to build our datasets and evaluated CamPressID's performance. In the end, CamPressID is able to bridge almost completely the sensing gap between a precise PPG sensor and a smartphone camera. CamPressID achieves 90.6% BAC, which approximates the performance of a dedicated PPG sensor with 93.7% BAC.

5

PassiveVLP: Leveraging Smart Lights for Passive Positioning

This chapter explores the last alternative for addressing the sensing gap. Instead of focusing on enhancing the methodology or sensor, we *enhance the object*. To showcase this alternative, we focus on an application related to indoor localization with visible light. Conventional positioning systems exploit light bulbs as transmitters by modulating their intensity to send beacons. This approach provides a high accuracy but requires placing a photosensor on the object, which increases the installation and management overhead. To overcome this limitation, this chapter exploits light reflection. We co-locate a low-cost photodiode with a (light bulb) transmitter to form a transceiver and modify *slightly* the object's surface to provide clear reflection patterns. By leveraging simple reflection geometry, our system achieves accurate object tracking, reducing the sensing gap between active and passive localization methods.

5.1. INTRODUCTION

Thanks to advances in the area of Visible Light Communication (VLC), we now have the ability to piggyback wireless communication on top of LED illumination. This technological breakthrough is creating a new range of exciting applications: localization [132, 133, 134, 135, 136, 137, 138], Internet connectivity via luminaires [139, 140], and a new generation of interactive toys [141], to name a few. Among these applications, positioning is arguably one of the areas that is benefiting the most. This is due to the fact that visible light waves have propagation properties that are well suited for estimating range: they attenuate in a smooth and rather deterministic manner. For example, when simple trilateration methods are used with radio frequency signals such as WiFi –which are notorious for having severe multipath effects–, the positioning accuracy is between 2 m and 6 m [142]. But the same trilateration methods provide sub-meter accuracy when VLC-enabled luminaires are used as anchor emitting beacons [133].

Research problem. Current positioning methods based on VLC [132, 133, 134, 138] provide high accuracy but share an important constraint: they require objects to carry photosensors to decode the beacons sent by luminaires. In PassiveVLP, our target is to unleash the constraint of current visible light positioning methods. We remove this constraint and investigate passive visible light positioning, that is, positioning in scenarios where the object of interest does not carry any photosensor. The case we make to investigate *passive* positioning is simple: when objects do not carry photosensors, the only areas that are *constantly* exposed to light are their external surfaces, thus, it's important to *investigate the interaction between VLC luminaires and the reflections coming from these external surfaces to identify cues for positioning. Passive* positioning with visible light can be tailored for both static objects and moving objects. The positioning for static objects can exploit the idea from [143, 144]. It requires attaching a retro-reflector on the surface of the object. Then the system will pinpoint the object according to the light intensities bounced back from the retro reflector to the receivers. While, the positioning for moving objects is more appealing to us. Considering this type of scenarios, we want to understand under what conditions would such a positioning system work and what would the expected accuracy be.

Fig. 5.1 captures the limitations of a naive implementation of passive positioning with visible light. Consider two luminaires that send periodic beacons $\mathscr{L}_1$ and $\mathscr{L}_2$. Each luminaire has a photodiode (PD) attached to it. In general, the PDs co-located with the luminaires may not be able to hear those beacons back due to the low reflectivity of ground surfaces. But if a mobile object with a highly reflective surface passes by, the PDs will receive their own (or a neighboring) beacon, and thus, be able to detect the presence of the object. For example, if luminaire $\mathscr{L}_1$ hears its own beacon, it will infer that an object is in region A, and if it hears a beacon from luminaire $\mathscr{L}_2$, it will infer that an object is in region B. But this basic configuration has three drawbacks: *poor coverage*, few positions can be detected (regions A, B and C in this case); *coarse-grained accuracy*, the exact location of the object cannot be determined because we do not know the reflecting angles (shape) of the object; and *there is no identification*, this system can only determine the presence of an object but not its ID, if more than one object is present, the objects cannot be distinguished.

Passive visible light positioning could be applied to scenarios where objects move on

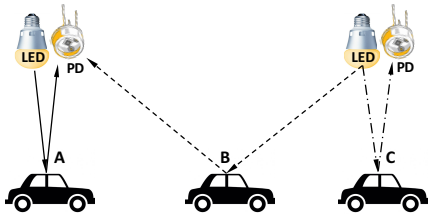


Figure 5.1: The concept of passive positioning with smart lights. Three regions *A, B,* and *C* can be detected under this simple configuration (PD: photodiode).

established paths in illuminated areas, e.g. mining tunnels or underground train and vehicular systems. In those underground scenarios, radio-based solutions face severe multipath effects, rendering most of those approaches inaccurate.

Our contributions. We propose a novel *passive* positioning system to overcome the above described limitations. Our contributions in this new area are listed below:

1. *Geometric model for positioning* (Section 5.2). We develop a geometric model to identify guidelines for the proper design of the transmitters (luminaires), improve the detection coverage, and increase the positioning accuracy. The model gives us insights on how to design the transmitters (luminaires) and on what type of objects can be localized with high accuracy.

2. *Passive identification* (Section 5.3). To identify the objects moving under the luminaires, we embed barcode-like IDs onto the objects' surfaces, and propose a novel framework to decode these IDs via passive reflections.

3. *Implementation* (Section 5.4). We design and implement a testbed to evaluate our system. To test short-range scenarios, we modify an open platform. For medium-range scenarios, we develop our own VLC transmitters using standard off-the-shelf LED bulbs.

4. *Evaluation* (Section 5.5). We evaluate our system under three different scenarios with increasing levels of complexity. Our results show that we can pinpoint with cm accuracy a subset of the object's trajectory and we are able to identify the object's ID in a passive manner.

5.2. MODEL FOR PASSIVE POSITIONING

To gain a better understanding of the properties and limitations of passive positioning with visible light, we use two models: the Lambertian source, which models the radiation pattern of LED lights; and a geometric model based on the laws of reflection. In the rest of this paper, we focus on one-dimensional topologies and assume that (i) *LED luminaires are VLC transceivers* (they can transmit information by modulating their light intensity and they have a photodiode to receive VLC packets), and (ii) the reflective coefficient of objects is higher than that of the surface where the object is moving on.

5.2.1. DETECTING OBJECTS WITHOUT LINE-OF-SIGHT

In VLC, angles play a major role in the received illuminance power. Fig. 5.2 captures this behavior. The wider the irradiation angle ϕ and the wider the incidence angle ψ, the

lower the received illuminance. Also, depending on the LED's optical enclosure, the radiation beam can be long and narrow, or broad and short. Formally, these relationships are captured by the well-known Lambertian model H [145]. Denoting P_t as the illuminance power of the LED, the received illuminance power P_r at the photosensor is given by:

$$P_r = P_t \cdot H \tag{5.1}$$

where

$$H = \begin{cases} \frac{(m+1)A}{2\pi d^2} \cos^m \phi T(\psi) g(\psi) \cos(\psi), & 0 \le \psi \le \Psi_c \\ 0, & \psi > \Psi_c \end{cases} \tag{5.2}$$

where m is the Lambertian order determining the width and length of the beam, a higher m leads to a longer and narrower beam; d is the distance between the LED and the photosensor; A is the detecting area of the photosensor; $T(\psi)$ and $g(\psi)$ are the concentrator and filter gains at the photosensor; and Ψ_c is the field-of-view of the receiver.

The Lambertian model is the path-loss model tailored for line-of-sight communication. It is similar to the Friis pathloss model for radio. Passive positioning relies on reflected beams (no-line-of-sight). To capture the effect of reflections, we modify the above equation based on two properties of the reflective surface; its area A_s and its reflective coefficient ρ.

$$H_{\text{NLOS}} = H f(A_s) \rho \tag{5.3}$$

where $f(A_s)$ is a linear function of A_s, as explained later. To capture the effect of these two new parameters, we perform the following experiment. First, we set a transmitter and a receiver at a two-meter distance with line-of-sight, as shown in Fig. 5.3, points A and B. Then we put a reflective surface at point I, one meter away from the LED, and move the receiver to point C (mirror image of point B). In this experiment we change the area (A_s) and the material (ρ) of the reflective surface. Our results lead to a design guideline for passive positioning.

Guideline 1: *Objects with specular reflection and high reflective coefficients are suitable for accurate positioning.* Passive positioning with VLC relies on receivers being able to detect reflected messages. The stronger the reflection, the higher the likelihood of detecting an object (better coverage). Upon impinging an object, a reflected light beam loses intensity in two ways: by the way the beam is reflected (shininess), and by how much light is absorbed by the object (reflection). We combine these two effects into a single parameter ρ. Table 5.1 shows the coefficients we obtained for three different materials. For the particular setup in Fig. 5.3, only two materials, mirror and aluminum, would be localizable in our system because their reflection coefficients are high enough to allow the decoding of information after reflection.

Guideline 2: *The reflective surface of the moving object A_s should be at least the same size as the receiver's area A.* Fig. 5.4 depicts the effect of changing the area of the reflective surface. For these experiments we use mirrors as the reflecting surface. When the area is small, the received illuminance power is low. Beyond a point however, further increasing the area of the reflective surface does not increase the received illuminance power. This occurs because, with specular reflections, the majority of reflections caused by larger areas do not reach the receiver. Thus, a reflective surface that is smaller than

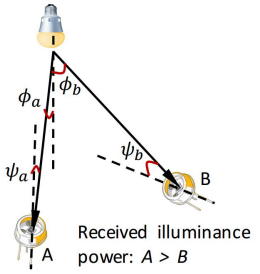


Figure 5.2: The Lambertian model.

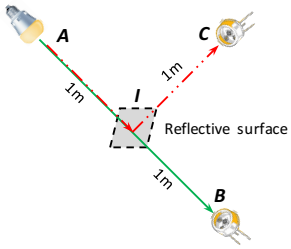


Figure 5.3: Capture of reflection effect.

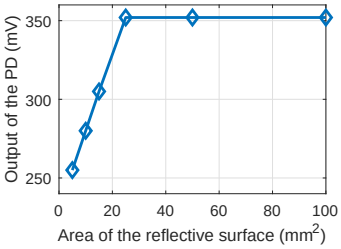


Figure 5.4: Impact of the area of the reflective surface (used material is mirror) under the setup of Figure 5.3 with 0.8-meter height.

Table 5.1: Measured combined coefficient ρ of different materials

Material	Mirror	Aluminum	White cardboard
Coefficient ρ	0.89	0.62	0.30

the photosensor's surface affects coverage, because it reduces the likelihood of detecting the object.

5.2.2. ACHIEVING FULL COVERAGE

The previous subsection indicates that an object only requires a small reflective area to be localized. However, this approach only provides limited coverage. In the simplest case, considering N luminaires we would only be able to identify $2N-1$ points. The N points under the luminaires and the $N-1$ intermediate points. However, since the size of the reflective surface can be small, an object can carry an array consisting of many small reflecting surfaces but tilted at different angles. The different angles will reflect beams toward the receiver at different locations, improving coverage. Based on simple geometry, we can make the following propositions based on Fig. 5.5.

Proposition 1. *Given two neighboring luminaires i and j, with an inter-distance d and height h from the ground; for any given position x_{ij} between the lights, there is a tilted angle that will reflect light towards a neighboring transceiver:*

$$\alpha_{ij} = \frac{\arctan\left(\frac{d-x_{ij}}{h}\right) - \arctan\left(\frac{x_{ij}}{h}\right)}{2}$$

where clockwise turns denote the positive direction of angles.

Proof. We leverage Fig. 5.5 to help visualize the proof. Letting θ be the incidence angle of light on the object, α_{ij} is given by:

$$\alpha_{ij} = \arctan\left(\frac{d-x_{ij}}{h}\right) - \theta$$

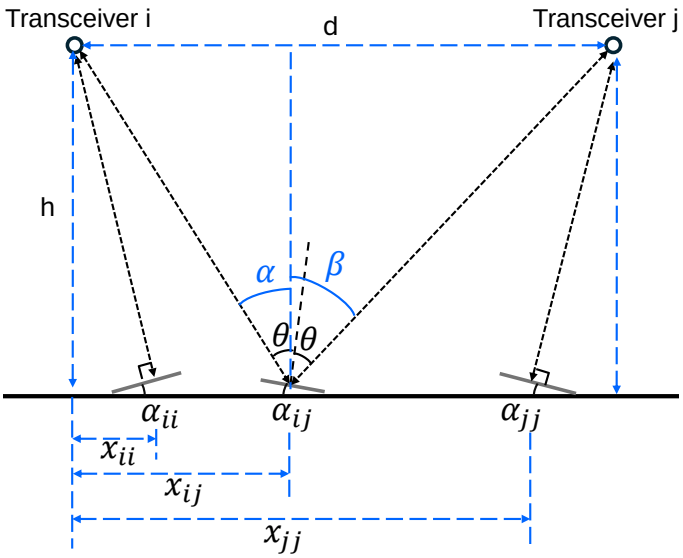


Figure 5.5: The illustration of the existences of α_{ij} (Proposition 2.1) as well as α_{ii} and α_{jj} (Proposition 2.2)

and it is easy to know that

$$\theta = \frac{\alpha+\beta}{2} = \frac{\arctan\left(\frac{x_{ij}}{h}\right) + \arctan\left(\frac{d-x_{ij}}{h}\right)}{2}$$

Hence,

$$\alpha_{ij} = \frac{\arctan\left(\frac{d-x_{ij}}{h}\right) - \arctan\left(\frac{x_{ij}}{h}\right)}{2}$$

□

Based on Proposition 1, we can get:

Proposition 2. *For a given position x, there always exists a tilted angle that can make the object reflect light back towards the same transceiver:*

$$\alpha_{ii} = -\arctan\left(\frac{x_{ii}}{h}\right), \quad \alpha_{jj} = \arctan\left(\frac{d-x_{jj}}{h}\right) \tag{5.4}$$

Proof. We leverage Fig. 5.5 to help visualize the proof. For α_{ii} and α_{jj}, the incidence angle of the light on the object is 0. α_{ii} and α_{jj} are equal to the irradiation angle of the light, which is $\arctan(\frac{x}{h})$ or $\arctan(\frac{d-x}{h})$. Considering the positive direction of angles, we get that

$$\alpha_{ii} = -\arctan(\frac{x}{h})$$

and

$$\alpha_{jj} = \arctan(\frac{d-x}{h})$$

□

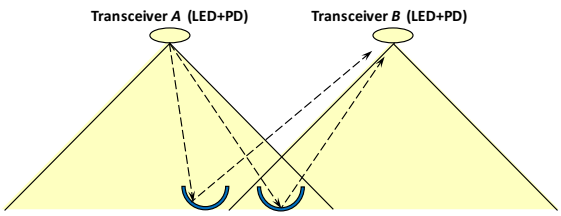


Figure 5.6: The problem of unique signature when each transceiver has a single beam *(the object has many tiny reflective surfaces tilted at different angles (kind of 'retro-reflector') to improve the coverage).*

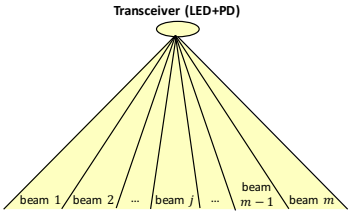


Figure 5.7: A transmitter with m beams *(here we use a triangle instead of the Lambertian shape to represent the coverage of an LED, for simplicity).*

Guideline 3: *A small polyhedral-reflector can be added on top of the object to provide constant coverage.* Consider an object with a reflection coefficient that is too low to reflect a VLC message. Such an object could not be localized in our system. To solve this problem, we can add polyhedral-reflectors to provide better coverage. A polyhedral-reflector can be created to reflect a wide incoming beam (cm^2) into many outgoing narrow beams (mm^2) in different directions. Note that even if an object has a high reflective coefficient, a polyhedral-reflector can be still used to increase coverage, since few objects have surfaces with multiple tilted angles.

5.2.3. OBTAINING UNIQUE SIGNATURES

Until now, our guidelines have focused on the issue of coverage (i.e. increasing the number of detected points). But we have not tackled the issue of obtaining a unique fingerprint for each detected point. To highlight this problem, let us use Fig. 5.6. Independently of the location of the object, a neighboring light will always receive the *same* ID, and thus, it would not be possible to identify the location of an object. For example in Fig. 5.6, Luminaire B receives the *same* beacon for the two positions of the convex surface, and thus, it cannot pinpoint the exact location. One way to discern two locations with the same beacon-ID is to use the RSS. But due to the intricate relation among the irradiation and incident angles, the Lambertian order m and the distance traveled d; we observed three problems. First, except for a few points, most RSS values map to multiple locations. Second, the sensitivity of simple photodiodes is not sufficient to distinguish small changes in RSS. Third, under some lighting conditions, the photodiodes can receive the beacons reflected by the ground, which affects the RSS.

Due to the limitations of RSS, we propose to use luminaires with multiple beams, as shown in Fig. 5.7. Instead of having a single *wide* beam with low Lambertian order, we propose to use multiple *narrow* beams with higher Lambertian order. Notice that the overall power does not need to increase, since each narrow beam requires less energy to attain the same range as a wide beam. In our system, each beam emits a unique ID tuple $< \mathscr{L}, \mathscr{B} >$, where $\mathscr{L}$ denotes the ID of the luminaire and $\mathscr{B}$ the ID of the beam within that luminaire.

Guideline 4: *Luminaires should be designed with multiple beams, the more beams the better.* It is important to highlight that standard off-the-shelf LED lights already consist of multiple internal LED substrates. Many of these LED substrates point in different

directions. Designing VLC luminaires for passive positioning would entail adjusting the angles of some of these beams and providing each beam with a unique ID $\mathscr{B}$.

5.2.4. POSITIONING ALGORITHM

Assuming a set of luminaires and objects following our guidelines, the positioning algorithm works as follows.

Step 1: The algorithm requires as inputs: the inter-node distance d, the height of luminaires h, the FoV of the photosensor ω, the number of beams at each luminaire b, a vector μ with the directional angles of the beams, and a vector containing the k tilted angles of the reflector (surface) $\{\alpha_1, \ldots, \alpha_k\}$.

Step 2: Using Proposition 1, the transceiver calculates the locations $\hat{x_i}$ for all the tilted angles $\alpha_i, i = 1, \ldots, k$. At this point we have k possible locations for the moving object.

Step 3: Upon receiving a beacon from a neighboring light, or from itself, the algorithm computes the region covered by that beam: $[x^1_{\mathscr{L}.\mathscr{B}}, x^2_{\mathscr{L}.\mathscr{B}}] = [h\tan(\mu_{\mathscr{L}.\mathscr{B}} - \omega/2), h\tan(\mu_{\mathscr{L}.\mathscr{B}} + \omega/2)]$. At this point, we know the object is under the coverage of beam $< \mathscr{L}.\mathscr{B} >$, but we do not know exactly where.

Step 4: The only valid locations $\hat{x_i}$ (Step 2) are those that fall in the range $[x^1_{\mathscr{L}.\mathscr{B}}, x^2_{\mathscr{L}.\mathscr{B}}]$ (Step 3). If only one estimation $\hat{x_i}$ falls in this range, $\hat{x_i}$ is given as the object's location. If multiple estimations fall in the range, the algorithm either returns the average as the location, or chooses one of the estimations with a higher probability if the direction and velocity of the target can be estimated with prior points.

5.3. PASSIVE IDENTIFICATION

Positioning an object is not enough. We now introduce how to further *identify* the object *passively* in our system.

Our approach is inspired by a recent work that leverages ambient light for passive communication [146], which adopts the patterns of distinctive reflecting surfaces to modulate ambient light. In this paper, we use a similar method to label objects with unique IDs attached to objects' surfaces, as illustrated in Fig. 5.8. We refer to this 'barcode' as *object-ID*. We use materials with different reflection coefficients to build an object ID, e.g., aluminum (high) and black paper (low). To decode an object ID at the PD according to the light it reflects, we have to tackle two challenges:

- *Overlapping signals containing the beam-ID and object-ID.* As presented in Sec. 5.2, each transmitter sends modulated light containing its *beam-ID*. At certain positions, this modulated light can be *reflected* by the object's surface to the PD and be used to localize the object. But the *object-ID* also *modulates* the impinging light, albeit at a lower frequency. Thus, the PD will receive an overlapping signal containing both the beam-ID and the object-ID, as illustrated in Fig. 5.9. To identify the object, we need to extract the signal that only contains the object ID from the overlapping signal.

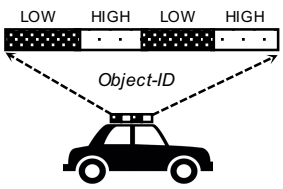


Figure 5.8: Attach a barcode-like ID to the object's surface.

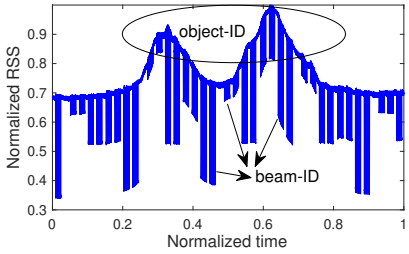


Figure 5.9: Illustration of the overlapped signal.

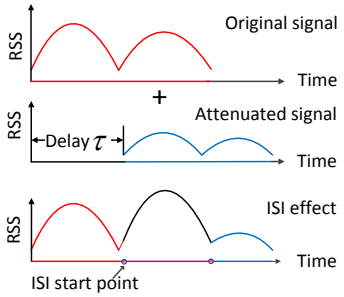


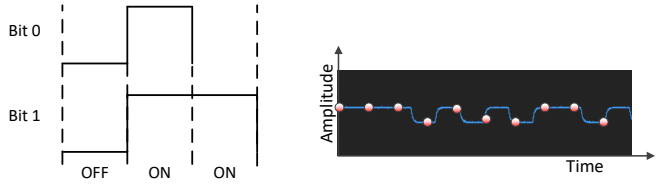
Figure 5.10: Illustration of the inter-symbol interference.

- *Inter-Symbol Interference (ISI)*. At some positions, different parts of the object-ID can reflect concurrently the light coming from two neighboring beams. Therefore, the light modulated by the object ID may contain interference caused by itself, namely, an ISI effect. As illustrated later in Fig. 5.13 of Section 5.3.2, the lights of beam-1 and beam-2 are modulated by the third symbol (at point *A*) and the first symbol (at point *B*) of the object-ID, respectively. This generates ISI at the PD, as shown in Fig. 5.10. Assuming that the object traverses point *A* first, then the signal modulated at point *B* is *actually an attenuated and delayed* version of the original modulated signal at point *A*. In the rest of this paper, we refer to these two types of signals as *attenuated signal* and *original signal*, respectively.

5.3.1. DECOUPLING OF OVERLAPPING SIGNALS

We tackle this problem by proposing a downsampling-based method motivated by the following observations: *(i)* The signal containing the beam-ID is a *high-frequency signal* (created by modulating LEDs at high speed) while the signal containing the object-ID is a *low-frequency signal* (created by the object's movement); *(ii)* When the LED transmits OFF symbols (namely, LED is off), there is a drop in the signal intensity received at the PD, as shown in Fig. 5.9 [1]. However, when transmitting ON symbols (namely, LED is on), no drop exits. *In other words, the overlapping signal only occurs when OFF symbols are*

[1] In Fig. 5.9, the "depth" of one beam ID fluctuates, we will explain this phenomenon using Fig. 5.26.



(a) Tailored modulation (b) Illustration for the downsampling result

Figure 5.11: Proposed downsampling method to decouple overlapping signals.

transmitted. Thus, we just need to remove the OFF symbols from the overlapped signals. The proposed downsampling method works as follows:

1. Guarantee that there are no continuous OFF symbols in the modulated data at the LEDs. This is a necessary requirement to remove the OFF symbols in our method. To achieve this, we use the following modulation at transceivers: use a symbol sequence of OFF-ON to denote a bit 0, and a symbol sequence of OFF-ON-ON to denote a bit 1. This modulation is depicted in Fig. 5.11(a);

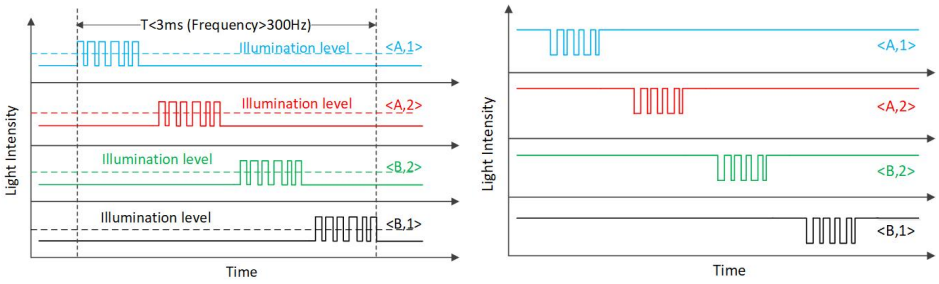
2. Set the downsampling interval slightly wider than the duration of an OFF symbol. By doing this, we guarantee that each OFF symbol is sampled at most once, as depicted in Fig. 5.11(b);

3. Compare each two adjacent samples: if the difference of two adjacent samples is higher than a threshold (i.e., an abrupt drop occurs in the RSS), then we discard the sample that has a lower value (i.e., remove the OFF symbol).

5.3.2. REMOVE THE INTER-SYMBOL INTERFERENCE

As we know, smart lighting has become an effective way to reduce the high energy footprint on artificial lighting. Smart lighting systems adjust the illuminance of artificial lights (normally, LED lights) based on the contribution of ambient natural sunlight in our environment. They are expected to maintain a constant illumination within the area of interest, improving the energy-saving and user-comfort [147]. Taking this practical requirement into consideration, the ISI problem described at the beginning of Section 5.3, can be addressed by different methods according to the requirement on artificial illumination from LEDs. In this paper, the operation modes of the LEDs are divided into two types:

- *Communication mode.* When the ambient light is enough to provide illumination, there will be no demand on artificial LED illumination. Therefore, an LED only needs to be turned on when it is scheduled to send its beacon; otherwise, the LED is turned off. This is depicted in Fig. 5.12(a). When only one LED is turned on at a given time slot, there would not have multiple LED lights adding together to form the ISI effect. Therefore, the ISI problem is avoided.



(a) *Communication mode* of the LEDs

(b) *Communication+illumination* mode of the LEDs

Figure 5.12: Operation modes of the LEDs under different ambient light conditions.

- *Communication+illumination mode.* This mode is enabled when the ambient light is weak. Under this mode, multiple LEDs should be turned on to satisfy the required illumination level. Among these LEDs, one of them is scheduled to transmit its beacon at a given time slot, while the other LEDs are simply turned on to illuminate the surrounding environment, as depicted in Fig. 5.12(b). As we presented at the beginning of Section 5.3, this is the main reason why we have multiple signals overlapped with each other that brings the ISI effect.

Under the *communication mode* of LEDs, the ISI problem can be avoided easily. However, in reality the *communication+illumination mode* of LEDs is more common and also more challenging. Therefore, in the rest of this subsection, we will present our general solution to remove the ISI effect considering the *communication+illumination mode* of LEDs.

To eliminate the ISI effect, we propose a method to remove the attenuated signal from the original one. To do this, we need the following information: (1) *delay* of the attenuated signal with respect to the original signal; and (2) *intensity* of the attenuated signal. Next we present how to obtain these information and how to extract and decode the original signal after that.

- ***Obtaining the delay of the attenuated signal*** Consider the scenario in Fig. 5.13. As presented earlier in this section, the ISI starts when the object reaches point B. At this position, a symbol (e.g., the ith symbol) of the object-ID will generate interference by modulating the light emitted by beam-2. *Note that this modulated light contains the ID of beam-2 and this beam-ID can be decoded by the PD for localization.* Moreover, when reaching point A, this ith symbol has modulated the light emitted by beam-1, and this modulated light carrying the ID of beam-1 has been decoded by the PD as well.

  Let t_0 and $t_0+\tau$ be the time when *for the first time* the PD decodes the IDs of beam-1 and beam-2, respectively. Since the transceiver can modulate light at a very high speed (i.e., it can transmit the frames that contain beam-IDs very frequently), *we can assume that t_0 and $t_0+\tau$ are the time when the object reaches point A and point B, respectively.*

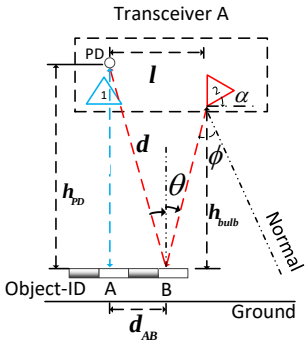


Figure 5.13: Illustration of the geometric model used to remove the ISI.

Therefore, *the difference between t_0 and $t_0+\tau$, namely, τ, is the delay of the attenuated signal.* Note that t_0 and $t_0+\tau$ can be obtained easily in practice. Besides, at some positions (as will be presented in Sec. 5.5.3), τ can be a *negative* value, meaning that the attenuated signal starts *earlier* than the original signal.

- ***Obtaining the intensity of the attenuated signal*** To derive this, we still use Fig. 5.13 as an illustration. Let h_{PD} and h_{bulb} denote the heights of the PD and bulb 2, respectively, and let l be the horizontal distance between them. From the geometric model in Fig. 5.13, we can derive the distance d_{AB}

$$d_{\text{AB}} = l \cdot \frac{h_{\text{PD}}}{h_{\text{PD}} + h_{\text{bulb}}} \tag{5.5}$$

Let θ and ϕ be the reflection angle at point B and irradiance angle of bulb 2, respectively. We have $\theta = \arctan \frac{d_{\text{AB}}}{h_{\text{PD}}}$ and $\phi = \alpha + \theta$. Based on Eq. (5.1), we can obtain the intensity of the attenuated signal modulated by the object-ID at point B:

$$P^B_{\text{att}} = \frac{\rho P_t (m+1) cos^m(\phi)}{2\pi d^4} T(\theta) g(\theta) cos(\theta) \tag{5.6}$$

Similarly, the intensity of the original signal modulated by the object-ID at point A can be expressed as follows:

$$P^A_{\text{orig}} = \frac{\rho P_t (m+1) cos^m(0)}{2\pi h^4_{\text{PD}}} T(0) g(0) cos(0) \tag{5.7}$$

In our model, we can assume that $h_{\text{PD}} \approx h_{\text{bulb}} \gg l$. Then we have the approximation: $h_{\text{PD}} \approx d$. Besides, there is no optical filter or concentrator in our system. Thus, $T(\cdot)$ and $g(\cdot)$ are constant in our model. Let η denote the *attenuation ratio*, defined as the ratio between the intensity of the attenuated signal and that of the original signal, then we have

$$\eta = \frac{P^B_{\text{att}}}{P^A_{\text{orig}}} = cos^m(\phi) cos(\theta) \tag{5.8}$$

- ***Extract and decode the original signal*** To achieve this, from now on we consider the continuous version of the signals. Following convention, we use $f(t)$ to denote a continuous signal. Let $f_{\text{orig}}(t)$ and $f_{\text{att}}(t)$ be the continuous original signal and the continuous attenuated signal, respectively. We have

$$f_{\text{att}}(t) = \eta f_{\text{orig}}(t-\tau) \tag{5.9}$$

Let $f_{\text{sum}}(t)$ be the aggregated signal received at the PD, then

$$f_{\text{orig}}(t) = f_{\text{sum}}(t) - f_{\text{att}}(t) = f_{\text{sum}}(t) - \eta f_{\text{orig}}(t-\tau) \tag{5.10}$$

As presented in Sec. 5.3.2, we can denote t_0 as the time when the object-ID reaches point A and $t_0+\tau$ as the time when the object-ID reaches point B where the ISI appears. Since there is no ISI during the time slot $[t_0, t_0+\tau)$, we have

$$f_{\text{orig}}(t) = f_{\text{sum}}(t), \quad \forall\, t \in [t_0, t_0+\tau) \tag{5.11}$$

Based on Eq. (5.10) and Eq. (5.11), we can now obtain the original signal $f_{\text{orig}}(t)$ (without the ISI effect). To decode it, we use a simple threshold-based decoding method (omitted due to its simplicity, limited novelty, and the page limitation).

5.3.3. OBJECT-ID DESIGN

As mentioned at the beginning of Section 5.3, we use materials with different reflection coefficients to build an object ID, e.g., aluminum (of high reflection) and black paper (of low reflection). An object-ID consists of two parts: *object-ID-preamble* and *object-ID-value*. In these two parts, we use materials with both high reflection coefficients (referred as 'HIGH' in the rest of this paper) and low reflection coefficients (referred as 'LOW' in the rest of this paper) to embed the information. The structure of our designed object-ID is given in Fig. 5.14. The PD's captured signal of the designed object-ID is illustrated in Fig. 5.15. The *object-ID-preamble* plays an important role in decoding the object-ID at the PD. Therefore, we first present our key considerations when designing the preamble.

OBJECT-ID-PREAMBLE DESIGN

The preamble can be used to obtain the threshold which is key to decode the following ID part of the object-ID. Besides, it can help estimate the speed of objects. Remember that we have ISI effect when multiple LEDs operate at the *communication+illumination* mode. Although we can remove the ISI effect with the mechanism introduced in Section 5.3.2, it is better if we can design an *ISI-free object-ID-preamble*. Denoting the length of object-ID-preamble as L_p. An improper L_p will bring ISI effect as well as making it difficult to locate the right ISI starting point in the signal. For instance, as illustrated in Fig. 5.16, if L_p is longer than d_{AB} (cf. Fig. 5.13), then the ISI effect occurs and it is not easy to calculate the ISI starting point, which is crucial for the mechanism we propose in Section 5.3.2.

To avoid the above problem, we design the object-ID-preamble length L_p to be equal to or smaller than d_{AB}. This means that, as illustrated in Fig. 5.16, the attenuated signal starts from the second valley in the original signal (when $L_p = d_{AB}$), or after the second

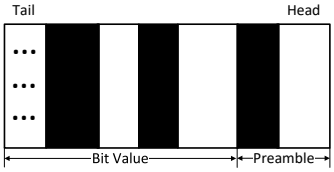


Figure 5.14: The structure of our object-ID.

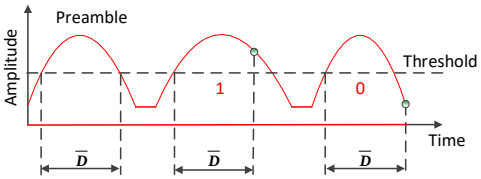


Figure 5.15: Illustration of the PD's captured object-ID signal.

valley (when $L_p < d_{AB}$). *In this work, we set* $L_p = d_{AB}$. By doing this, we not only achieve an ISI-free preamble for the object ID but also know that the ISI starts from the second valley of the original signal. Note that when L_p is too short, it is not easy for the PD to detect the preamble. In passiveVLP, the appearance duration of the Object-ID-preamble at the PD should be much longer than the time needed to transmit a beacon, in order for the PD to decode correctly the beacons. This can be seen clearly from Fig. 5.9, where the duration of an object-ID is much longer than that of a beam-ID.

In the preamble, we use one HIGH and one LOW, starting from HIGH, i.e. the preamble has the pattern of HIGH-LOW. The total length of the HIGH and the LOW in the preamble equals L_p. This design of the preamble is shown in Fig. 5.17. In reality, it is possible that the distance between the object ID and the ceiling PD varies (e.g. the ceiling is not flat in the underground scenario). Therefore, L_p could not always be the same as d_{AB}. In this case, the second valley in the combining signal will no longer be the ISI starting point. Therefore, we need to find the minimum length of LOW to make the second valley as long as possible, to form a flat area. Then, the beginning of the rising trend will become the ISI starting point.

OBJECT-ID-VALUE DESIGN

Following the used pattern in the preamble, in the object-ID-value part, we also use the pattern HIGH-LOW to denote each bit of the object ID. For example, if there are four bits to denote the value of the object ID, then we will have four groups of HIGH-LOW in the object-ID-value part. To distinguish between a bit 1 and a bit 0, we use different durations for the HIGH and LOW in the corresponding group of HIGH-LOW, as illustrated in Fig. 5.17.

5.4. TESTBED

A solid evaluation of passive positioning with VLC requires designing a system where multiple parameters can be adjusted. Below we describe how we use our design guidelines to build a comprehensive evaluation testbed.

5.4.1. LUMINAIRE DESIGN

We design our luminaires based on the Shine platform [148]. Shine was originally designed for omni-directional multi-hop communication with visible light. Each Shine node is equipped with 20 LEDs (HLMP-CM1A-450DD, FoV is 18°); and four PDs (SFH203P, FoV is 90°). The micro-controller is an Atmega 328P, which is low-cost but powerful.

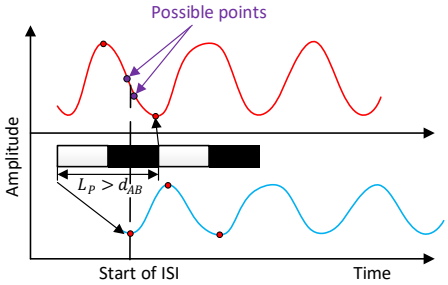


Figure 5.16: Illustration of the starting point of the ISI effect.

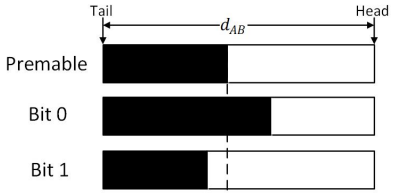


Figure 5.17: Bit design in PassiveVLP.

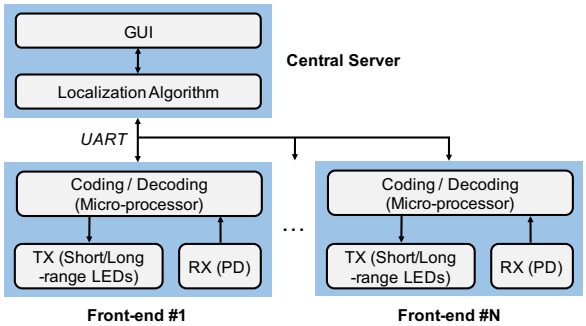


Figure 5.18: The function blocks of our PassiveVLP testbed.

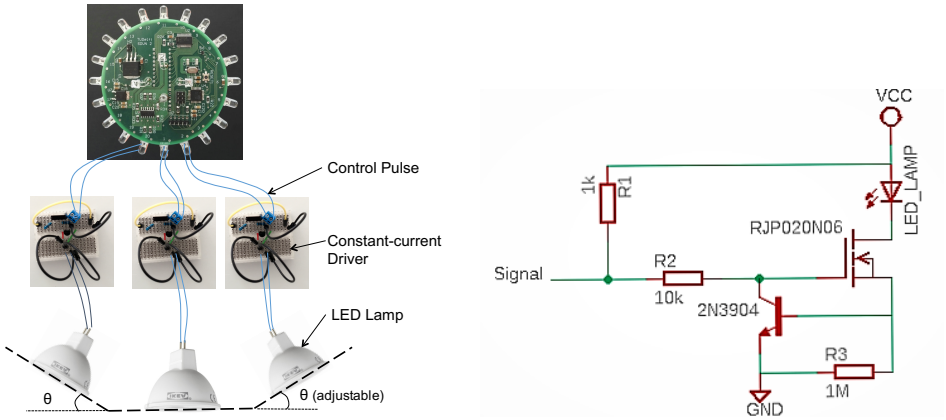
Shine also provides serial communication capability to interface with a PC/laptop. We extend it largely in both the hardware and software to build our testbed in three ways. First, we improve its reception capabilities, a front-end we refer to as *Shine+*. Second, we add more powerful LEDs to build another front-end referred as *Shine++*. Third, we modify its software. The block diagram of our new testbed is given in Fig. 5.18. Next, we present the design details of both front-ends.

A SHORT-RANGE LUMINAIRE WITH MULTIPLE-BEAMS – SHINE+

Shine has multiple narrow beams, which follows *Guideline 3*, but it is designed for line-of-sight communication. Since we rely on *non-line-of-sight reflections*, we must improve Shine's reception capabilities. To achieve this goal, we implement two changes. First, we change the PD from SFH203P to SFH206K, which extend the communication range by 165%. Second, we add a low pass filter to improve the signal-to-noise ratio.

A MEDIUM-RANGE LUMINAIRE WITH LESS-BEAMS – SHINE++

To evaluate our sensing approach with longer ranges in realistic scenarios, we build a new front-end with commercial LED bulbs, dubbed Shine++. We choose the IKEA Ledare LED with a viewing angle of 36 °as the transmitters. We use three LED bulbs, one placed in the middle and the other two on the sides with adjustable inclination angles, as shown



(a) The whole system

(b) The LED driver circuit

Figure 5.19: Our testbed with the front-end Shine++.

in Fig. 5.19(a). Each LED consumes 3.5 W, thus we design a new LED driver circuit to provide higher power, as presented in Fig. 5.19(b).

SOFTWARE

Our software implements the data transmission and reception, the access control for the shared visible light medium, the positioning and object identification algorithms. In our application we need accurate timing to schedule the modulation of light beams. We connect all the nodes to a central server (PC/laptop), where we run a Time Division Multiple Access (TDMA) scheme. This scheme turns *on* all LEDs, but modulates only one *beam* at a time. Adaptive decoding thresholds are implemented in our system to eliminate interference from external light sources, including neighboring LEDs that are *on* but not modulated. The MAC schedules the LEDs 'remotely' through the interface with the micro-controllers. Similarly, the data received from the PDs are decoded at the micro-controllers and sent through the UART to the central server. Upon receiving these frames, the server runs the *positioning algorithm* (cf. Section 5.2.4) and calculates on-the-fly the current position of the mobile object. The outcome of the algorithm is demonstrated in a simple GUI (omitted due to the space limitation).

5.4.2. OBJECT DESIGN

For passive positioning with VLC, the external surface of the objects plays a key role. Three out of the four guidelines in Section 5.2 pertain to the object's surface. In this subsection we describe the surfaces we use for our evaluation and the reasoning behind selecting them.

A PERFECT REFLECTOR

As stated by our guidelines, the ideal surface would use materials with very high reflective coefficients and specular reflection to enable non-line-of-sight communication, and would consist of small reflecting areas (Guideline 1) tilted at different angles (Guideline

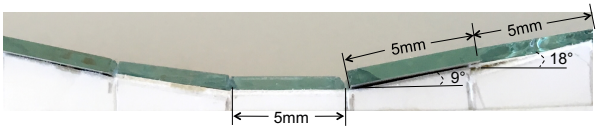


Figure 5.20: Side view of the object with *perfect* reflective coefficient.

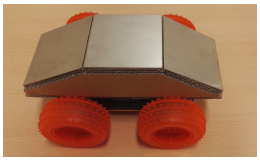


Figure 5.21: Customized toy car *(standard reflector, 28cm× 11cm× 7cm).*

Figure 5.22: The object-ID (preamble + ID value "0") that is attached to all the three parts of the object.

2) to increase coverage. We use thin strips of flat mirrors to satisfy these three guidelines. Fig. 5.20 presents a side view of the reflecting object, which consists of five mirrors with inclined angles of -18°, -9°, 0°, 9°, 18°, respectively. All the mirrors have the same width of 5 mm.

A STANDARD REFLECTOR

To evaluate the performance of passive positioning with more standard objects, we use a customized toy car. Such an object relaxes the requirements of our first three guidelines. The material is metal, whose reflective coefficient is not as good as mirrors. Second, the surfaces are big (relaxes Guideline 1). Third, it has few tilted angles (relaxes Guideline 2). For our object, we consider three parts of a car: the front windshield, the roof, and the back windshield. After a thorough investigation of different cars' shapes, we decide to customize a toy car with inclined angels of -30°, 0°, and 25°. The final customized car is shown in Fig. 5.21. Note that we are aware of the fact that different parts of a real car have different reflection coefficients, but we only use one type of material for a single toy car in this work for simplicity. We build two different toy cars, one with aluminum and the other one with mirrors.

DESIGN OF OBJECT-ID

We label the objects in our system with unique IDs consisting of certain patterns. The ID pattern has been presented in the Fig. 5.17 . An object-ID consisting of the preamble and a bit "0", is shown in Fig. 5.22 where it is attached to all three faces of the customized toy car.

5.5. EVALUATION

In this section, we evaluate our methods under increasingly complex test cases. First we present the evaluation on positioning, followed by the evaluation on identification.

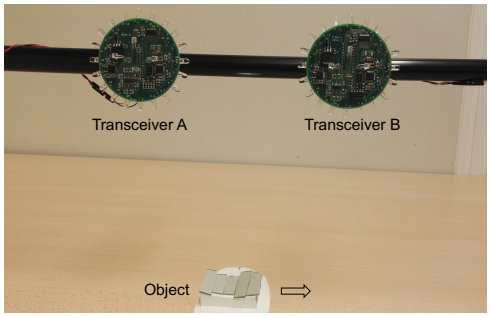


Figure 5.23: Experiment setup in the ideal case (height=15 cm, inter-node distance=20 cm).

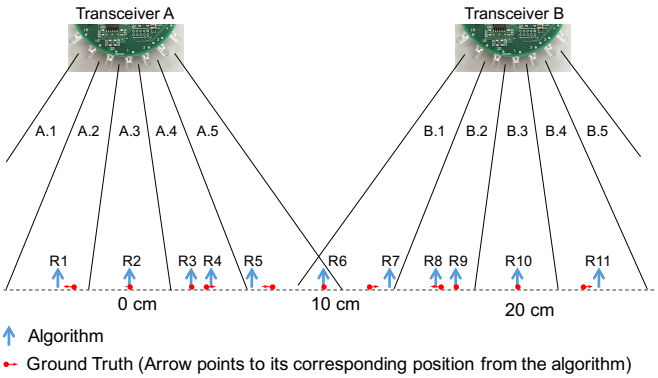


Figure 5.24: Evaluation results in the ideal case.

Note that our passive positioning and identification methods can work at the same time, as described in Sec. 5.5.3.

5.5.1. POSITIONING: IDEAL CASE

For the ideal case we use the best possible setup: the front-end with many beams (Shine+) together with the perfect reflector. The experiment setup is shown in Fig. 5.23. We deploy two nodes (denoted as A and B) at an inter-node distance of 20 cm, and at a height of 15 cm from a desktop. The light intensity measured at the desktop is about 450 lux under the nodes (cf. positions *R2* and *R10* in Fig. 5.24). Shine+ can achieve reliable VLC at a distance of up to 50 cm. By deploying the two nodes as described above, we can assure that the distance travelled by the reflected signals is shorter than the maximal communication distance (i.e., 50 cm). It is also important to highlight that the angles of the perfect reflector are designed based on Propositions 1 and 2 to guarantee that every beam has at least one angle that will reflect the light to a neighboring node or to itself. If these angles are not selected carefully, the system may end up having beams without reflections, and thus, no positioning could be performed inside the area covered by these beams. Different inter-node distances or heights will lead to different tilted angles.

Table 5.2: Details of the localized positions from our algorithm in ideal case *(Number: the detected locations; TX: transmitter; RX: receiver).*

Number	R1	R2	R3	R4	R5	R6	R7	R8	R9	R10	R11
Reflection angle	-9^o	0^o	9^o	-18^o	18^o	0^o	-18^o	18^o	-9^o	0^o	9^o
TX	A.2	A.3	A.4	A.4	A.5	A.5	B.1	B.2	B.2	B.3	B.4
RX	A	A	A	B	B	B	A	A	B	B	B
Calculated position	-4.7	0	4.7	3.3	6.7	10	13.3	16.7	15.3	20	24.7
Measured position	-3.4	0	3.4	4.5	7.9	10	12.1	15.9	16.1	20	23.4

Results. The evaluation results are shown in Fig. 5.24. The red dots represent the ground truth. The experiments were repeated ten times and we did not observe any major variance, which is expected due to the rather deterministic propagation properties of light waves and the nearly specular reflection of mirrors. We can observe that the results from our algorithm match well the ground truth. Nearly half of the locations are localized with almost perfect accuracy, while the other locations are detected with errors up to 1.3 cm.

The detail on how these locations are detected is given in Table. 5.2. Now let us give some more insights on how our algorithm (cf. 5.2.4) leads to these results. The object moves from left to right. The algorithm has all the inputs required in Step 1. The first detected point *R1* is a self reflection, i.e., transceiver A receives the reflection of beam *A.2*. The server calculates all the ground truth locations that the set of self-reflecting angles $\{\alpha_1,\ldots,\alpha_k\}$ can have (Step 2). Then it calculates the valid region for this beam *A.2* (Step 3). The algorithm detects that only one point falls in the valid region, and hence, reports that point as the estimated location. The true location is ≈1 cm to the right. The next two estimated locations, *R2* and *R3*, are also self-reflections. The fourth estimated location *R4* is due to the communication between transceivers A and B (inter-node reflection). Notice that in this case, two locations are within the valid region of beam *A.4*: *R3* and *R4*. But the fact that *R3* is a self reflection while *R4* is not, helps with distinguishing them. Note that *R4* does not need to be under the coverage of transceiver B to achieve inter-node reflection. This is because the photodiodes used in our experiments have a wide field of view (90^o). A more challenging case occurs with the next two locations *(R5, R6)*. These two locations are within the valid region of beam *A.5*, and both are the result of inter-node reflection. As stated in Step 4 of our algorithm, we can either average them out at the cost of increasing the error, or exploit speed and direction information from prior data to select the most likely location. We implement a very simple mechanism to exploit direction information. Every time we detect a new point we set it as the origin, prior data are given negative values based on their distance to this last point, and new data are given positive values. If two or more positive points are estimated as locations, we select the closest one (point *R5* in this case). The remainder half of the path is symmetrical to the first half, and thus, the estimation is similar to what has just been described. We don't have any location estimations for beams A.1 and B.5, because there are no self-reflecting angles for these regions. If transceiver A would have a neighbour to its left, then beam A.1 would have two potential locations (similar to *R6* and *R7*).

Regarding the errors in our estimations, we found that they are due to two main reasons. First, misalignment of the angles in the moving object, e.g. our calculations in

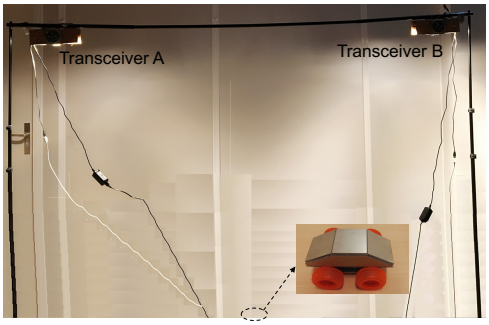


Figure 5.25: Experiment setup in the realistic case.

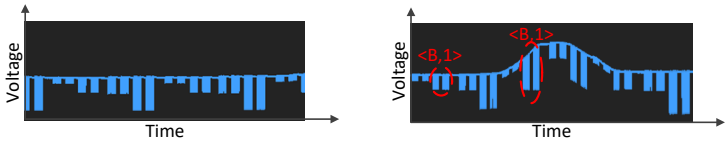
the algorithm are done assuming the tilted angles of the object are 9^o, 18^o, etc., while in practice there are certain errors. Second, we assume that luminaires are single-LED sources, while in practice they have multiple LEDs (5 LED sources in the case of *Shine+*). This difference changes the incidence angles, which in turn affects the estimated location. This latter point is why our errors are more pronounced at the boundaries of two beams and more accurate at the center of beams.

Finally, it is important to note that passive positioning with VLC operates in a fundamentally different manner compared to most positioning methods. Traditionally, after a signal is received, the positioning algorithm provides an estimated location that can be anywhere. In our method, the positioning algorithm only has a limited number of locations to choose from. The number of locations depends on the number of beams and tilted angles. Thus, upon receiving a signal, our method's task is to map the received signal to the most appropriate location, *without requiring a training phase*.

5.5.2. POSITIONING: REALISTIC CASE

We now test our passive positioning in a more realistic case, consisting of less beams and an object with less tilted angels. In this scenario, we use two nodes equipped with the Shine++ front-end. As introduced in Sec. 5.4, Shine++ uses more powerful LEDs, which enable nodes to communicate at a distance of 5 m (the distance can be further by increasing the light intensity of luminairies). The experimental setup in a realistic case is shown in 5.25. In the tests, we set the height of the nodes to various levels: 1 m, 1.5 m, and 2 m. The height (range) can be increased if we add a lens to the photodiode. Meanwhile, the inter-node distance is adjusted between 2 m and 2.5 m. The light intensity measured at the floor is about 600 lux under the nodes (cf. positions *R1* and *R5* in Fig. 5.27) when the height of the nodes is 1.5 m.

Dealing with floor reflection. The floor can reflect the light emitted by LEDs. When the light is modulated to send beacons, the ceiling PD can detect the transmissions of these beacons due to the floor reflection. Fig. 5.26(a) shows this phenomenon. When an object with a reflective surface passes under the LEDs, the PD detects the "depth" change of beacons, as illustrated in Fig. 5.26(b). And these changing beacons show which bulb has its light reflected by the surface. To identify these beacons for object localization, we calculate the depth difference of each beacon between situations in Fig. 5.26(a) and (b).



(a) PD receives beacons reflected by floor (b) PD response when object passes *R*3

Figure 5.26: Dealing with the floor reflection.

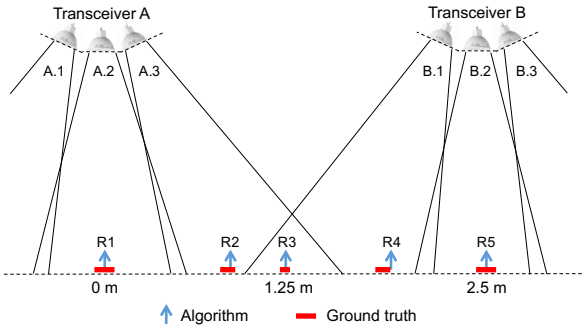


Figure 5.27: Evaluation results in the realistic case (height=1.5 m, distance=2.5 m).

If the depth difference exceeds a threshold, the system can identify the beacon, e.g. the beacon $< B,1 >$ illustrated in Fig. 5.26(b) (the depth of other beacons remains roughly the same), our system will know which light beam has caused the above changes and then localize the object based on this information [2].

Results. The evaluation results are shown in Fig. 5.27, where the height and inter-node distance are set to 1.5 m and 2.5 m, respectively. For this experiment we use the aluminum-car. All the six LEDs work under the *communication+illumination* mode. First, it is important to observe that in this setup we can only detect five locations, which is less than the eleven locations detected under the ideal case. This occurs because we now have less beams and less tilted angles. But all the five estimated locations are still accurate. Note that in Fig. 5.27, red bars are used to represent the ground-truth positions (instead of dots as in Fig. 5.24). This is because the surfaces are wide enough to give a continuous location range, while in Sec. 5.5.1 the mirrors are so narrow that only a 'single' point is detected. Results under different inter-node distances and heights are given in Tables 5.3–5.5. These tables show the estimations of our algorithm and the ranges of the actual locations for the aluminum- and mirror-car. From these results we can observe that the maximum positioning error is around 5.3 cm and the average error is 0.97 cm. The performances with the mirror- and aluminum-car are similar. The only difference is that in Table 5.5, the position *R3* can be detected with the mirror-car but

[2] Except beacons $< B,1 >$, other beacons are constantly reflected by the ground. In the middle of Fig. 5.26(b), beacons $< B,1 >$ (in the second red circle) are reflected by the surface. Hence, when the ON time in $< B,1 >$, the received intensity is higher, while the OFF time in $< B,1 >$, the received intensity returns to the same previous beacons $< B,1 >$ (in the first red circle). Therefore, the received intensity is shown as Fig. 5.26(b)

Table 5.3: Evaluation results in realistic case (height=1 m, inter-node distance=2 m)

No.	R1	R2	R3	R4	R5
Reflection angle	0^o	-30^o	0^o	25^o	0^o
Algorithm	0 m	0.577 m	1 m	1.534 m	2 m
Mirror	-0.02~0.02	0.63~0.66	0.99~1.01	1.48~1.51	1.98~2.02
Aluminum	-0.02~0.02	0.62~0.65	0.99~1.01	1.47~1.5	1.98~2.02

Table 5.4: Evaluation results in realistic case (height=1.5 m, inter-node distance=2.5 m).

No.	R1	R2	R3	R4	R5
Reflection angle	0^o	-30^o	0^o	25^o	0^o
Algorithm	0 m	0.866 m	1.25 m	1.801 m	2.5 m
Mirror	-0.02~0.02	0.85~0.88	1.25~1.26	1.76~1.78	2.48~2.52
Aluminum	-0.02~0.02	0.85~0.88	1.25~1.26	1.77~1.79	2.48~2.52

not with the aluminum-car (due to the lower reflective coefficient of aluminum).

5.5.3. IDENTIFICATION

We now present the evaluation on passive identification. Without loss of generality, we use the object-ID as shown in Fig. 5.22. In the experiment, the height of the transceivers is 1.5 m and the inter-node distance is 2.5 m. The object is moved at a constant speed by a toy train motor from transceiver A to transceiver B. We only use the "standard reflector" in this test because the "perfect reflector" is too small to carry the object-ID.

RESULTS FOR THE SCENARIO WHEN LEDS OPERATE AT COMMUNICATION+ILLUMINATION MODE.

The signal received by the PD of transceiver A is shown in Fig. 5.28(a). *We clearly observe the high-frequency signal that contains the beam-ID, which has been successfully used for positioning in our experiment (similar to Fig. 5.27).* Furthermore, we can observe that the object-ID appears three times: In Fig. 5.28(a), the corresponding signals are marked as *G1, G2* and *G3*. They correspond to the three positions $R1$, $R2$, and $R3$ in Fig. 5.27, where the object is located. Note that the PD on transceiver B captures the signals of the object-ID when the object appears at the location $R3$, $R4$, and $R5$. To decode the object-ID at these positions, we first decouple signals using the downsampling based method presented in Sec. 5.3.1. The resulted signal is shown in Fig. 5.28(b), which is much cleaner. In Fig. 5.28(b), the first two groups of signals (*G1* and *G2*) are affected by ISI. For *G1*, which is mapped to position $R1$, ISI takes effect at the tail. For *G2* (mapped to position $R2$), ISI occurs at the beginning (τ is negative, cf. Sec. 5.3.2). For both *G1* and *G2*, we successfully used the steps presented in Sec. 5.3.2 to remove the ISI. The results are show in Fig. 5.28(c). For *G3* (mapped to position $R3$), the object only modulates the light of one beam. Thus, ISI does not exist, and we can decode the object-ID directly.

RESULTS FOR THE SCENARIO WHEN LEDS OPERATE AT COMMUNICATION MODE.

When LEDs operate at the communication node, there is no ISI effect (cf. Section 5.3.2). However, we cannot directly obtain a clear envelop of the object-ID at the PD that can be

Table 5.5: Evaluation results in realistic case (height=2 m, inter-node distance=2.5 m).

No.	R1	R2	R3	R4	R5
Reflection angle	0^o	-30^o	0^o	25^o	0^o
Algorithm	0 m	1.155 m	1.25 m	1.567 m	2.5 m
Mirror	-0.02~0.02	1.15~1.17	1.25~1.27	1.5~1.52	2.48~2.52
Aluminum	-0.02~0.02	1.13~1.15	none	1.51~1.52	2.48~2.52

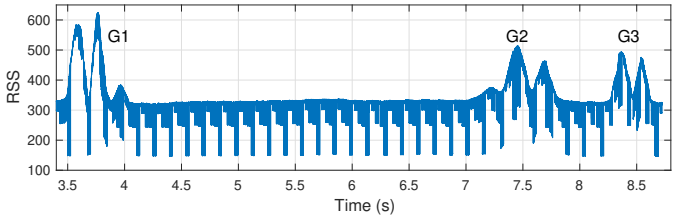
used to decode the object-ID. Fig. 5.29 shows the signal received by the PD of transceiver A when the object passes by the position near R1 (cf. Fig. 5.27). We can observe that there is no ISI effect since no envelop of the object-ID is captured by the PD. To obtain the envelop of the object-ID, we *fit the envelop of the captured signals for the same beam-ID*. For instance, in Fig. 5.29 we highlight the captured signal related to the beam $< A,2 >$. Due to the movement of the object, the captured signal strengths are different over the time. By fitting the maximal amplitudes of these captured beam-ID signals, we can obtain the envelop of the object-ID, as shown by the red-dash line. Similarly, we can obtain the envelop of the object-ID generated due to the object's reflection of the beam $< A,3 >$, as denoted by the yellow-dash line. After obtaining the envelop, we can decode the object-ID, as we present in Section 5.5.3. Note that if the LEDs operate at *communication+illumination* mode, the two envelops mentioned above will overlap with each other in the time domain, starting from the beginning of the yellow-dash line.

5.6. RELATED WORK

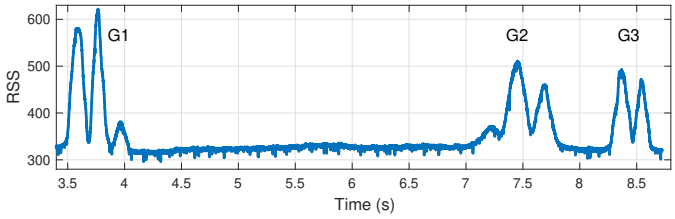
Localization, both active and passive, has been investigated widely. In this section, we summarize the most relevant work.

Passive positioning with radio. M. Youssef *et al.* introduce the concept of Device-free Passive (DfP) positioning using Wi-Fi [149, 150]. They show that changes in radio signals, caused by people, can be harnessed to localize a person. By comparing with the radio map that stores the signal strength in the area of interest [151], the authors show that an average accuracy of 0.3 m can be achieved. Investigations in realistic environments are carried out in a follow-up work [150]. Recently, researchers have also been able to track *multiple* objects passively with existing radio signals [152, 153]. High accuracies are achieved in these studies. We are motivated by these works to analyze the unique properties of visible light waves for *passive* positioning. Compared to radio waves, visible light waves behave in a more deterministic manner (less multipath) but have poorer coverage (because they cannot travel through opaque objects). Our study exposes the opportunities and limitations of exploiting the external surfaces of objects to achieve accurate positioning and identification.

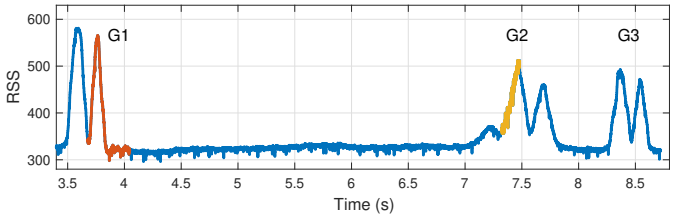
Active positioning with visible light. By leveraging the built-in camera or light sensors in smartphones, researchers have exploited *active* positioning with visible light. Epsilon [154] is one of such pioneer positioning systems. In Epsilon, the smartphone is enhanced with a plugged-in photosensor to detect incident lights emitted by LED lamps. Epsilon requires at least three LED lamps as anchors such that an algorithm based on trilateration can be used to calculate the smartphone's position. It can achieve accuracies of 0.4 m to 0.8 m under different indoor environments. This performance is im-



(a) Raw signal received by the PD at transceiver A



(b) Detected envelope: filtering of high-freq. signal (Sec. 5.3.1)



(c) After eliminating the ISI (the changes are highlighted)

Figure 5.28: Evaluation on passive identification when LEDs operate at the communication+illumination mode.

proved by Luxapose [134], which also adopts smartphone and LED luminaires as receiver and transmitters, respectively. Unlike Epsilon, Luxapose leverages the angle-of-arrival of the signals in its positioning algorithm, and manages to achieve a better accuracy of decimeter-level. A lightweight positioning system with visible light is proposed in [155]. It leverages polarization-based modulation instead of intensity-based to reduce the processing workload in wearable devices. With these resource-constrained devices as receivers, it can achieve an accuracy of 3 m in 90% of the test cases. Complementary to existing work, LiPro [156] provides a solution for scenarios with insufficient reference points (LED bulbs). It tackles the case when there is only one reference point. By rotating the receiver (smartphone) around three orthogonal axes, it continuously records the RSS and magnetic field, and then calculates the receiver's position by exploiting the Lambertian radiation property [157] of LED luminaires. LiPro can achieve a median error of 0.59 m in a corridor. In [158], the authors introduce a VLC method to tackle the underground localization problem. It exploits trilateration to calculate the target position. It can achieve an average error of 3.5 cm. Similar to these studies, we also rely on the Lambertian coverage of LED lights and the ideal attenuation properties of visible light waves to obtain positioning. But contrary to those studies, our method does not require carrying any photodetector with line-of-sight towards the luminaires, we exploit

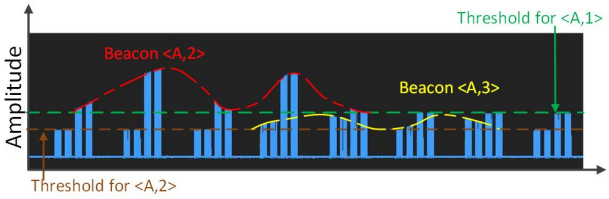


Figure 5.29: Evaluation on passive identification when LEDs operate at the communication mode.

the reflective properties of external surfaces.

Passive sensing with visible light. Okuli [135] was one of the work that inspired our PassiveVLP. Okuli uses an LED and two photodiodes to perform passive *near-field* sensing. It can track a finger's movement with a median error of 0.7 cm within an 8x8 cm pad. Okuli exploits the fact that fingers are round and good diffusers of light to build a model-driven solution, but it requires training data and a lighting system that is *specifically designed* for near-field positioning. Our system harnesses LEDs whose primary function is illumination and does not require any training to obtain only occupancy. Two other related studies are CeilingSee [136, 160] and LocaLight [159]. CeilingSee estimates occupancy by monitoring changes in light reflection caused by people in a room. CeilingSee uses general purpose luminaries, but they require a training phase to obtain only occupancy information. LocaLight deploys photosensors on floors to track people based on the shadows they cast. Compared to our work, the main advantage of these three systems is that they do not need to modify the external surfaces of the elements they track. We, on the other hand, modify the external surfaces, but obtain accurate positioning at longer ranges without requiring extra infrastructure (Okuli, LocaLight) or training phases (Okuli, CeilingSee).

Note that there is also a line of research that leverages light for secure and passive communication [146, 161, 162]. The authors in [161] design a secure system for barcode-based VLC between smartphones. In our previous work [146], we proposed a barcode-like passive VLC system with ambient light such as sunlight. Recently, we further performed a thorough analysis on how to design a barcode to transmit information on a

Table 5.6: Summary of the most relevant work on passive localization/positioning.

State of the art	Purpose	Scenario	Medium	Passive	Accuracy
[150]	Localization	Indoor	Wi-Fi	Yes	0.3 m
[152]	Localization	Indoor	Radio	Yes	0.12 m
Epsilon [154]	Localization	Indoor	Light	No	0.8 m
Luxapose [134]	Localization	Indoor	Light	No	0.4 m
[155]	Localization	Indoor	Light	No	3 m
LiPro [156]	Localization	Indoor	Light	No	0.59 m
[158]	Localization	Underground	Light	No	0.0035 m
Okuli [135]	Localization	Indoor	Light	Yes	0.007 m
CeilingSee [136]	Occupancy	Indoor	Light	Yes	-
Localight [159]	Occupancy	Indoor	Light	Yes	-

moving object [162]. In all those studies, the design of barcodes with visible light is key, but those systems are designed for wireless communication. In this work, PassiveVLP, we leverage visible light barcodes not only for identification but more importantly, for localization in a passive manner.

5.7. LIMITATION & DISCUSSION

Many of the assumptions we make for the potential applications are realistic, such as knowing the height and geometry of luminaries. But our system also has limitations: (*i*) the path should not be bumpy, (*ii*) we can only track a few points in the paths, (*iii*) the system only works for 1-D scenarios, (*iv*) if multiple objects pass the same point simultaneously, they would cause 'reflection collisions', and (*v*) the size of the object's surface determines the maximum number of symbols that can be encoded (maximum number of IDs), (*vi*) the orientation of transceivers should be stable.

Point (*i*) is a strong requirement, the bumpy spots in the detection area will generate outliers in our result. Our system is only resilient to few bumpy spots by eliminating outliers according to the object trace. The other five points can be improved. For point (*ii*), Kalman or Particle filters can be used to provide continuous location information. For point (*iii*), we can create annular FoVs with a single photodiode, cf. Fig. 5.30, to provide 2D positioning, cf. Fig. 5.31. Solving point (*iv*) with a single PD would be challenging, because it is hard to disaggregate colliding signals, a plausible alternative is to add more PDs with a narrower FoV to cover single lanes or tracks. Reducing the receivers' FoV would also help ameliorating point (*v*): a narrow FoV would allow us to use narrower stripes, which would increase the number of IDs that can be encoded on the object's surface. Some results can be found in our latest paper [162]. To alleviate point (*vi*), we can store the light intensity caused by ground reflection after calibration. When there is no object passing by and the ground reflection does not match with the stored value well, we can then recalibrate the orientation of transceivers.

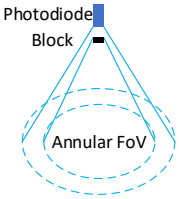


Figure 5.30: Customized annular-FoV.

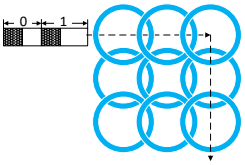


Figure 5.31: 2D passive positioning.

5.8. CONCLUSION

In this work, we took the first step to design a passive localization system based on visible light, where objects are not required to carry photosensors. To achieve our goal, we modify slightly the external surfaces of objects so light reflections can provide information for localization and identification. We define and analyze the elements of our proposed

system, and implement a testbed to benchmark its performance. Our results show that visible light can provide passive identification and localization with cm-level accuracy, bridging the sensing gap between active and passive systems. Passive sensing and localization with visible light is a nascent area, and thus, there is plenty of room for future work. The results from experiments confirmed the feasibility of localizing and identifying objects with visible light. While the current implementation of our system can not localize objects continuously, it succeeded in providing a subset of the objects' trajectory with high accuracy. We hope our work offers new insights in this up-and-coming domain.

6

CONCLUSION

The Internet of Things is enabling the collection of vast amounts of data. This new sensing capability is possible due to multiple factors. In this thesis, we argue that three of those factors are: (i) the ability to use low-cost and re-purposed sensors, (ii) the use of a reflective sensing approach, and (iii) the exploitation of electromagnetic spectra that are safe and pervasive.

6.1. CONTRIBUTIONS

The main challenge tackled in this thesis was measuring and improving the accuracy of a sensing system with low-cost or re-purposed sensors. To analyze this research challenge we investigated four use cases considering different applications.

- **Contribution 1**: *Measuring the sensing gap in crowd monitoring applications with microwave sensors.*

  By performing crowd monitoring in an outdoor public area, we determined the sensing gap between low-cost mmWave sensors and precise cameras. In some public areas, GDPR laws prevent the use of conventional camera-based solutions. One promising alternative is mmWave sensors. Similar to radar, mmWave sensors detect objects via reflected signals. Their relatively long wavelength provides medium ranges (several tens of meters) and vague representations of people, complying with GDPR.

  Chapter 2 measured the sensing gap of mmWave sensors relying on three main contributions. First, we carefully designed a weather-proof radome with high signal quality (-10dB Integrated Sidelobe Ratio) and broad coverage (larger than commercial mmWave sensors). Second, we deployed two mmWave sensors on a real high-traffic area with cyclists and pedestrians. The monitoring lasted three months including various weather conditions (sunny, rainy, windy, foggy and snowy days). Third, via a thorough evaluation, our findings show that low-cost mmWave sensors have an error rate that is between 15-40% higher than that of cameras, indicating that further research is required to enhance the performance.

- **Contribution 2**: *Enhancing the methodology for cardio identification with low-cost and re-purposed sensors.*

  For cardio identification, we bridge the sensing gap between advanced cardiac systems and cheap PPG sensors and smartphone cameras. To achieve that goal we propose a novel methodology. Current authentication systems rely on external features like fingerprint, face, or iris recognition. However, these features are vulnerable because they are constantly exposed. To overcome that limitation, internal biometrics, such as cardiac signals, are being used for authentication. Given that visible light and infrared waves can detect blood flows, low-cost PPG sensors and smartphone cameras could replace high-end cardiac equipment to capture signals for authentication.

  CardioID (Chapter 3) bridged the sensing gap for low-end devices (PPG sensors and cameras) through three contributions. First, to stabilize PPG signals with high variability, we design an adaptive filter that automatically adjusts its parameters according to the subject's heart rate. Second, to accommodate cardiovascular differences, instead of assuming that all subjects have the same cardiac morphology, CardioID considers three dominant morphologies to increase the user coverage and real-time performance of our system. Third, to handle non-linear effects, CardioID adopts Mahalanobis distances with a wavelet clustering approach. Thanks to the novel methodology, CardioID can reduce the sensing (authentication) gap by 10% BAC.

- **Contribution 3**: *Enhancing re-purposed sensors (smartphone cameras) for cardio identification.*

  Enhancing the methodology is central to bridging a sensing gap, but beyond a point, this alternative is fundamentally limited by the quality of the re-purposed sensor. In this case, it is central to enhance the sensor to provide signals that are as close as possible to the original device. For our cardiac application, this implies re-purposing smartphone cameras to operate almost as PPG sensors.

  CamPressID (Chapter 4) extends CardioID to address the lower authentication performance of smartphone cameras. Compared to PPG sensors, cameras use a different spectrum and do not have a finger clip to maintain stable pressure. CamPressID bridges the sensing gap with two novel contributions. First, it considers all camera parameters to generate signals that are similar to what is obtained with an infrared sensor. Second, by analyzing different pressure levels, CamPressID can identify the best pressure for each user and provide feedback to stabilize that pressure. Based on these improvements, CamPressID improves the authentication performance from 80% to above 90% BAC, which is similar to the performance of a PPG sensor.

- **Contribution 4**: *Enhancing the object's surface for indoor positioning with light.*

  Modifying the object's surface can improve the quality of the signals backscattered toward the sensors. This property is particularly useful with light because several materials are good light reflectors. In Chapter 5, we used indoor localization to illustrate how enhancing an object's surface can help with bridging the sensing gap. Unlike radio-based solutions for indoor localization, which suffer from multipath effects, visible light's shorter wavelength mitigate these issues. Additionally, visible light localization can exploit the existing lighting infrastructure, which reduces the system's overhead.

PassiveVLP takes indoor visible light positioning one step further. By placing a small reflective tag on top of the object, PassiveVLP eliminates the common SoA requirement of mounting an active receiver. With our approach, we showed that PassiveVLP can achieve simultaneous identification and localization based on simple reflection laws. Through experiments on a scaled-down testbed, PassiveVLP can track the trajectory of a target with a location error at cm-level without any active device on the object.

6.2. LOOKING BACK

This thesis explored a broad research area, reflective sensing, using various types of sensors and applications. The challenges in this wide landscape are many and diverse. For us, working on topics that range from crowd monitoring to biometric authentication and indoor localization opens the opportunity to identify common macro-challenges. We believe that one of those macro-challenges is captured by the main research question posed in this thesis:

What design alternatives are available to approach the performance of high-end sensors with either low-cost or re-purposed sensors?

The motivation for this question comes from the need to expand IoT capabilities with minimal overhead. Given that the IoT is becoming a pervasive sensing infrastructure, it is important to reduce its economic and management costs. The best opportunity to achieve that goal is to re-purpose existing sensors for new applications. Currently, the number of deployed IoT devices exceeds 20 billion [163], leading to abundant chances to identify new ways to use those sensors. The best example of this sensor re-purposing trend is smartphones, where innovators and researchers continue finding novel applications, from health care monitoring to scuba diving communication. For applications that cannot be solved via sensor re-purposing, the research community is aiming towards using low-end alternatives, not only to reduce costs but also to satisfy other requirements like privacy.

Our studies show that the fundamental drawback of adopting a low-cost/re-purposed sensing approach is the low signal quality. That sensing gap needs to be measured and bridged, but the methodology to achieve those goals depends on the type of sensor and application. To measure the sensing gap, the most important consideration is a thorough evaluation to expose all the sensor's shortcomings. A common pitfall of current research efforts is that new applications are developed mainly for controlled scenarios. In our case (Chapter 2), we found that mmWave had not been tested thoroughly outdoors. Our evaluation showed a noticeable sensing gap that opens up research opportunities in that domain.

Once a sensing gap is assessed we need alternatives to bridge it. The most convenient approach is to *enhance the methodology* because it does not require modifying the functionality of the original sensor or object (Chapter 3). Advancements in signal processing and, in particular, machine learning play a key role in the ability of re-purposing sensors for other applications. In cases where enhancing the methodology is not sufficient, the quality of the signal can be improved by *enhancing the re-purposed sensor* (Chapter 4) *or the object* (Chapter 5). A re-purposed sensor can be enhanced in different man-

ners, from the design of an add-on to the (temporal) modification of parameters, which is the approach we follow for our biometric application. The enhancement of the object poses more restrictions. In particular we do not want to add active electronic devices on objects, so we modified the surface instead.

Overall, we believe that the current trend of extending the sensing capabilities of the IoT will continue to open several opportunities for future research.

6.3. FUTURE WORK

This dissertation delved into four case studies to quantify and fill the sensing gap between low-cost and precise sensors with visible light, infrared, and micro waves. The performance of these case-studies can be enhanced, and even further, their underlying repurposing soltions can be applied in widely different application domains. Based on our work, we identify three promising research opportunities.

6.3.1. RE-PURPOSING MMWAVE SENSORS FOR PROTEIN ESTIMATION

Our work assessed mmWave sensors for crowd monitoring out of their comfort zone in indoor environments. However, these sensors are versatile and have been used for various applications such as material and liquid identification. mmWave sensors can detect the material of 21 daily objects made of various components, shapes, and textures at an average 90% accuracy [164] and even distinguish liquors with 1% alcohol concentration difference [165]. Motivated by these studies, we propose another potential application for mmWave sensors: protein content estimation. As an important index for meat quality, protein content determines the price of meat. However, current mainstream approaches like combustion and chemical methods are invasive and their processes are long and complex [166]. With the help of advanced machine learning methods, mmWave systems could be re-purposed to quantify protein content in meat in a non-invasive manner.

6.3.2. RE-PURPOSING NEW SMARTPHONES FOR CARDIAC IDENTIFICATION

A correct camera configuration and pressure control are critical to re-purpose smartphone cameras into PPG sensors. However, current smartphones have two limitations: flashlights with fixed-intensity and no pressure feedback. Fixed-intensity flashlights require a complex compensation of the camera parameters to reach optimal PPG signals for people with different skin thicknesses and tones. And the lack of real-time feedback for the pressure level prevents users from adjusting their fingertip pressure to the optimal point. A new generation of smartphones that place the camera under the screen could solve the above limitations. First, instead of using the constant and white spectrum of the flashlight, the screen could exploit its RGB LEDs to provide a spectrum and intensity that fits better the color and thickness of the user's fingertip, eliminating in that manner the camera parameters compensation. Second, when the fingertip applies pressure on the screen, the camera would be able to detect the pressure based on the number of pixels covered by the fingertip, enabling timely feedback for the users to adjust their pressure level.

6.3.3. ENHANCING OBJECTS WITH DYNAMIC SURFACES

PassiveVLP re-purposes the object's external surface to achieve passive localization and identification under two strong assumptions: the object must move for the ID to be detected and the localization is only done in one dimension and at discrete points. With a stationary object, our system cannot detect any change in light intensity. In addition, the system would not be able to locate an object moving freely in a 2D or 3D space. Overcoming these limitations is central to enable more realistic applications, such as locating a drone flying inside a warehouse using only the lighting infrastructure in the area. To track an object in 3D, we could further modify the surface to modulate the reflected light using a method called retro-VLC [143]. This method proposes tags with two layers: a retro-reflector on the bottom and a liquid crystal (LC) shutter on the top. The LC shutter can control whether the light is reflected or absorbed. These tags are thin, a couple of millimeters, and consume negligible power, less than a microwatt. These properties allow them to be placed on top of multiple surfaces. Overall, these tags can act as a dynamic "surface" that controls the reflection of light over the object's surface, allowing the object to "radiate" communication.

ACKNOWLEDGEMENTS

It has been a long way for me to finish this thesis. During this torturous journey, it is nearly impossible for me to achieve this PhD without support and help from many people around me. Here I want to take this opportunity to show my gratitude to all of you.

First of all, I want to thank my supervising team, even though words cannot fully express my gratitude to them: Koen Langendoen, Marco Zuniga and Qing Wang. As my promoter, Koen showed me how to challenge and justify an idea in a reasonable way. In life, after I made stupid moves every now and then (detailed in Koen's notebook), Koen always had his patience and support for me. I will remember your favorite motto "less is more" (if you give me more time, I will write less words about my gratitude to you). As my supervisor, Marco always has faith in me although I could barely express my ideas at our first meeting. You taught me how to abstract my ideas and think in a structured manner. Qing, as my co-supervisor, is the one who introduced me to our group. You taught me how to do proper research and keep an eye on details.

Along with my supervising team, brilliant minds in our group also offer me countless help. Fernando, Przemysław, VP, Guohao, Qun and Georgios Whose insightful research broadened my horizon and elevated my research tastes. Stef and Belma, thanks for all the food, drink, gathering and meaningful topic discussion together, especially the visit to my hometown and the experience of one typical Chinese day. Jorik and Nikos, thanks for all the late-night talk to get through the frustration time. Vijay and Sujay, thanks for all advice on my problems and cheerful talks. Vito and Belma, thanks for all the joyful moments like my "special" birthday gift and the legal second-hand bike. Miguel, Talia and Koen Godemoundt, thanks for all the meaningful and insightful discussion about politics, news, Latin/Chinese relationships and our directions. Girish, Anup and Keyarash, thanks for the delightful work and useful advice. Gabe, Adrian, Sury, Naram, Mike Fatih and Niels, thanks for all the happy times together. All Chinese fellows, Hanting, Mingkun, Ran, Lingyu, Longyu, Hao, Mengyuan, Yukang, Shenxiu, Gabe, Talia and Miguel (half-Chinese), thanks for all the time playing together and the Chinese festival together making me feel less homesick. Our secretaries Minaksie, Kim and Pam, thanks for your quick response and support for me. Special thanks to my Master Marek, it was a pleasant experience to work with you despite that we had to measure people's heartrate during COVID-19. Finally, I also want to thank our lunch committee members for countless funny moments: "Mega" Koen, Stef, Belma, Vito, Jorik, Nikos, Gabe, Adrian, Fatih and "Micro" Koen.

Besides our group, I feel grateful to my roommates: Chuan, Xiaozhe, Yucheng, Jianping, Peng and JianPeng. You help me a lot with my regular allergy, hay fever and my hopeless cooking skills.

At last, I want to thank my parents. You have been there all the time for me without any reservations, although I showed most of my bad side to you. Without your support, I definitely would not have gone so far to accomplish this thesis.

BIBLIOGRAPHY

[1] Adam Zielonka et al. “Intelligent internet of things system for smart home optimal convection”. In: *IEEE Transactions on Industrial Informatics* 17.6 (2020), pp. 4308–4317.

[2] Youcef Djenouri, Djamel Djenouri, and Jerry Chun-Wei Lin. “Trajectory outlier detection: New problems and solutions for smart cities”. In: *ACM Transactions on Knowledge Discovery from Data (TKDD)* 15.2 (2021), pp. 1–28.

[3] Sujay Narayana et al. “Swans: Sensor wireless actuator network in space”. In: *Proceedings of the 15th ACM Conference on Embedded Network Sensor Systems*. 2017, pp. 1–6.

[4] Fresh Produce Instructment. *Brix Levels Are Crucial to the Fresh Fruit Industry*. URL: https://freshproduceinstruments.com/blog/brix-levels-are-crucial-to-the-fresh-fruit-industry/.

[5] Chris Arnold. *Hemoglobin Testing Devices Market Will Hit USD 5,350 Million By 2026*. URL: https://www.linkedin.com/pulse/hemoglobin-testing-devices-market-hit-usd-5350-million-chris-arnold/.

[6] Fresco microwave sensing. *Brix sensor*. URL: https://fresco.vertigo-tech.com/.

[7] Edward Jay Wang et al. “HemaApp: noninvasive blood screening of hemoglobin using smartphone cameras”. In: *Proceedings of the 2016 ACM International Joint Conference on Pervasive and Ubiquitous Computing*. 2016, pp. 593–604.

[8] V Mohan Malhotra and Nicholas J Carino. *Handbook on nondestructive testing of concrete*. CRC press, 2003.

[9] Yingfeng Shan, John E Speich, and Kam K Leang. “Low-cost IR reflective sensors for submicrolevel position measurement and control”. In: *IEEE/ASME transactions on mechatronics* 13.6 (2008), pp. 700–709.

[10] Katsutoshi Masai et al. “Facial expression recognition in daily life by embedded photo reflective sensors on smart eyewear”. In: *Proceedings of the 21st International Conference on Intelligent User Interfaces*. 2016, pp. 317–326.

[11] Justin G Chen et al. “Video camera–based vibration measurement for civil infrastructure applications”. In: *Journal of Infrastructure Systems* 23.3 (2017), B4016013.

[12] Qing Wang, Marco Zuniga, and Domenico Giustiniano. “Passive communication with ambient light”. In: *Proceedings of the 12th International on Conference on emerging Networking EXperiments and Technologies*. 2016, pp. 97–104.

[13] Haipeng Liu et al. “MTransSee: Enabling environment-independent mmWave sensing based gesture recognition via transfer learning”. In: *Proceedings of the ACM on Interactive, Mobile, Wearable and Ubiquitous Technologies* 6.1 (2022), pp. 1–28.

[14] Haipeng Liu et al. “Real-time arm gesture recognition in smart home scenarios via millimeter wave sensing”. In: *Proceedings of the ACM on interactive, mobile, wearable and ubiquitous technologies* 4.4 (2020), pp. 1–28.

[15] Sameera Palipana et al. “Pantomime: Mid-air gesture recognition with sparse millimeter-wave radar point clouds”. In: *Proceedings of the ACM on interactive, mobile, wearable and ubiquitous technologies* 5.1 (2021), pp. 1–27.

[16] Mingmin Zhao et al. “RF-based 3D skeletons”. In: *Proceedings of the 2018 Conference of the ACM Special Interest Group on Data Communication*. 2018, pp. 267–281.

[17] Xiaotong Zhang, Zhenjiang Li, and Jin Zhang. “Synthesized Millimeter-Waves for Human Motion Sensing”. In: *Proceedings of the 20th ACM Conference on Embedded Networked Sensor Systems*. 2022, pp. 377–390.

[18] Feng Lin et al. “Cardiac scan: A non-contact and continuous heart-based user authentication system”. In: *Proceedings of the 23rd Annual International Conference on Mobile Computing and Networking*. 2017, pp. 315–328.

[19] Zhe Chen et al. “MoVi-Fi: Motion-robust vital signs waveform recovery via deep interpreted RF sensing”. In: *Proceedings of the 27th annual international conference on mobile computing and networking*. 2021, pp. 392–405.

[20] Unsoo Ha, Salah Assana, and Fadel Adib. “Contactless seismocardiography via deep learning radars”. In: *Proceedings of the 26th annual international conference on mobile computing and networking*. 2020, pp. 1–14.

[21] Peter NT Wells. “Ultrasound imaging”. In: *Physics in Medicine & Biology* 51.13 (2006), R83.

[22] Chenshu Wu et al. “mmTrack: Passive multi-person localization using commodity millimeter wave radio”. In: *IEEE INFOCOM 2020-IEEE Conference on Computer Communications*. IEEE. 2020, pp. 2400–2409.

[23] NASA. *The Electromagnetic Spectrum*. 2013. URL: https://imagine.gsfc.nasa.gov/science/toolbox/emspectrum2.html.

[24] Elisabeth Cardis et al. “Risk of cancer after low doses of ionising radiation: retrospective cohort study in 15 countries”. In: *Bmj* 331.7508 (2005), p. 77.

[25] Lane F Donnelly. “Reducing radiation dose associated with pediatric CT by decreasing unnecessary examinations”. In: *American journal of roentgenology (1976)* 184.2 (2005), pp. 655–657.

[26] Alena Rajnochová Svobodová et al. “DNA damage after acute exposure of mice skin to physiological doses of UVB and UVA light”. In: *Archives of dermatological research* 304.5 (2012), pp. 407–412.

[27] Grzegorz Kłosowski et al. “The use of heterogeneous deep neural network system in radio tomography to detect people indoors”. In: *Proceedings of the 28th Annual International Conference on Mobile Computing And Networking*. 2022, pp. 859–861.

[28] Joey Wilson and Neal Patwari. “Radio tomographic imaging with wireless networks”. In: *IEEE Transactions on Mobile Computing* 9.5 (2010), pp. 621–632.

[29] Daniel Romero, Donghoon Lee, and Georgios B Giannakis. “Blind radio tomography”. In: *IEEE Transactions on Signal Processing* 66.8 (2018), pp. 2055–2069.

[30] Raghav H Venkatnarayan, Griffin Page, and Muhammad Shahzad. “Multi-user gesture recognition using WiFi”. In: *Proceedings of the 16th Annual International Conference on Mobile Systems, Applications, and Services*. 2018, pp. 401–413.

[31] Rui Xiao et al. “Onefi: One-shot recognition for unseen gesture via cots wifi”. In: *Proceedings of the 19th ACM Conference on Embedded Networked Sensor Systems*. 2021, pp. 206–219.

[32] Yongsen Ma et al. “SignFi: Sign language recognition using WiFi”. In: *Proceedings of the ACM on Interactive, Mobile, Wearable and Ubiquitous Technologies* 2.1 (2018), pp. 1–21.

[33] Aditya Virmani and Muhammad Shahzad. “Position and orientation agnostic gesture recognition using wifi”. In: *Proceedings of the 15th Annual International Conference on Mobile Systems, Applications, and Services*. 2017, pp. 252–264.

[34] Viet Nguyen et al. “HandSense: capacitive coupling-based dynamic, micro finger gesture recognition”. In: *Proceedings of the 17th Conference on Embedded Networked Sensor Systems*. 2019, pp. 285–297.

[35] Sheng Tan et al. “MultiTrack: Multi-user tracking and activity recognition using commodity WiFi”. In: *Proceedings of the 2019 CHI Conference on Human Factors in Computing Systems*. 2019, pp. 1–12.

[36] Yasha Iravantchi et al. “Privacymic: Utilizing inaudible frequencies for privacy preserving daily activity recognition”. In: *Proceedings of the 2021 CHI Conference on Human Factors in Computing Systems*. 2021, pp. 1–13.

[37] Wei Wang et al. “Understanding and modeling of wifi signal based human activity recognition”. In: *Proceedings of the 21st annual international conference on mobile computing and networking*. 2015, pp. 65–76.

[38] Zhicheng Yang et al. “Monitoring vital signs using millimeter wave”. In: *Proceedings of the 17th ACM international symposium on mobile ad hoc networking and computing*. 2016, pp. 211–220.

[39] Chris Xiaoxuan Lu et al. “See through smoke: robust indoor mapping with low-cost mmwave radar”. In: *Proceedings of the 18th International Conference on Mobile Systems, Applications, and Services*. 2020, pp. 14–27.

[40] Anand Chandrasekhar et al. “Smartphone-based blood pressure monitoring via the oscillometric finger-pressing method”. In: *Science Translational Medicine* 10.431 (Mar. 2018).

[41] Yang Gao et al. “Echowhisper: Exploring an acoustic-based silent speech interface for smartphone users”. In: *Proceedings of the ACM on Interactive, Mobile, Wearable and Ubiquitous Technologies* 4.3 (2020), pp. 1–27.

[42] Edward Jay Wang et al. “Seismo: Blood pressure monitoring using built-in smartphone accelerometer and camera”. In: *Proceedings of the 2018 CHI conference on human factors in computing Systems*. 2018, pp. 1–9.

[43] Xiangjie Kong et al. “Mobile crowdsourcing in smart cities: Technologies, applications, and future challenges”. In: *IEEE Internet of Things Journal* 6.5 (2019), pp. 8095–8113.

[44] Irida Shallari, Silvia Krug, and Mattias O’Nils. “Architectural evaluation of node: server partitioning for people counting”. In: *Proceedings of the 12th International Conference on Distributed Smart Cameras*. 2018, pp. 1–6.

[45] Fabio Dittrich et al. “People counting in crowded and outdoor scenes using a hybrid multi-camera approach”. In: *arXiv preprint arXiv:1704.00326* (2017).

[46] Jonas Weiß et al. “Improved people counting algorithm for indoor environments using 60 GHz FMCW radar”. In: *RadarConf 20*. 2020.

[47] Christian Groß et al. “Towards an Occupancy Count Functionality for Smart Buildings-An Industrial Perspective”. In: *IESES*. 2020.

[48] Nima Karimian et al. “Human recognition from photoplethysmography (PPG) based on non-fiducial features”. In: *ICASSP*. 2017.

[49] Jian Liu et al. “CardioCam: Leveraging Camera on Mobile Devices to Verify Users While Their Heart is Pumping”. In: *MobiSys*. 2019.

[50] Y. Y. Gu, Y. Zhang, and Y. T. Zhang. “A novel biometric approach in human verification by photoplethysmographic signals”. In: *ITAB*. 2003.

[51] Tatiana S Lugovaya. “Biometric human identification based on ECG”. In: *PhysioNet* (2005).

[52] Steven Israel et al. “ECG to identify individuals”. In: *Pattern recognition* (2005).

[53] Hao Ma, Lutz Lampe, and Steve Hranilovic. “Coordinated broadcasting for multiuser indoor visible light communication systems”. In: *IEEE Transactions on Communications* 63.9 (2015), pp. 3313–3324.

[54] Niranjini Rajagopal, Patrick Lazik, and Anthony Rowe. “Visual light landmarks for mobile devices”. In: *IPSN-14 proceedings of the 13th international symposium on information processing in sensor networks*. IEEE. 2014, pp. 249–260.

[55] Keith Still et al. “Place crowd safety, crowd science? Case studies and application”. In: *Journal of Place Management and Development* (2020).

[56] Arvind Thiagarajan et al. “Accurate, low-energy trajectory mapping for mobile devices”. In: (2011).

[57] *Sensorenregister*. `https://sensorenregister.amsterdam.nl/`.

[58] *Big Brother Watch*. `https://bigbrotherwatch.org.uk//`.

[59] Ramnath K Chellappa. “Consumers’ trust in electronic commerce transactions: the role of perceived privacy and perceived security”. In: *Citeseer* (2008).

[60] Peijun Zhao et al. “mid: Tracking and identifying people with millimeter wave radar”. In: *2019 15th International Conference on Distributed Computing in Sensor Systems (DCOSS)*. IEEE. 2019, pp. 33–40.

[61] Tianbo Gu et al. “MmSense: Multi-person detection and identification via mmWave sensing”. In: *mmNets*. 2019.

[62] Shenglei Li et al. “A Field People Counting Test Using Millimeter Wave Radar in the Restaurant”. In: (2021).

[63] Texas Instruments. *60GHz mmWave Sensor EVMs*. May 2022.

[64] Cesar Iovescu et al. “The fundamentals of millimeter wave sensors”. In: *Texas Instruments* (2017).

[65] Keegan Garcia. “Bringing intelligent autonomy to fine motion detection and people counting with TImmWave sensors”. In: *Texas Instruments* (2018).

[66] Xu Huang et al. “Indoor detection and tracking of people using mmwave sensor”. In: *Journal of Sensors* (2021).

[67] Arindam Sengupta et al. “mm-Pose: Real-time human skeletal posture estimation using mmWave radars and CNNs”. In: *IEEE Sensors Journal* (2020).

[68] Yong Wang et al. “Remote monitoring of human vital signs based on 77-GHz mm-wave FMCW radar”. In: *Sensors* (2020).

[69] *Choosing 60-GHz mmWave sensors over 24-GHz to enable smarter industrial applications*. `https://www.ti.com/lit/wp/spry328/spry328.pdf`.

[70] *WAYV Air: Short-Range IoT Radar Sensor*. `https://sensing.ai/products/wayv-air`.

[71] *Mm-Wave Radar Radome Design Guide*. `https://www.ti.com/lit/an/swra705/swra705.pdf?ts=1663590143575`.

[72] *White Paper on Radar wave propagation through materials: Walls and radomes*. `https://www.infineon.com/dgdl/Infineon-Radar_wave_propagation_through_materials-Whitepaper-v01_00-EN.pdf?fileId=5546d462766cbe8601768a120c6da=t`.

[73] Claudio Sacchi et al. “Advanced image-processing tools for counting people in tourist site-monitoring applications”. In: *Signal Processing* (2001).

[74] Ya-li Hou et al. “Automated people counting at a mass site”. In: *IEEE ICAL*. 2008.

[75] Saandeep Depatla et al. “Occupancy estimation using only WiFi power measurements”. In: *IEEE JSAC* (2015).

[76] Yuheng Wang et al. “m-activity: Accurate and real-time human activity recognition via millimeter wave radar”. In: *ICASSP*. 2021.

[77] Nicolas Scheiner et al. “Seeing around street corners: Non-line-of-sight detection and tracking in-the-wild using doppler radar”. In: *CVPR*. 2020.

[78] Emily C Hodkinson et al. “Heritability of ECG biomarkers in the Netherlands twin registry measured from Holter ECGs”. In: *Frontiers in Physiology* (2016), p. 154.

[79] A Reşit Kavsaoğlu, Kemal Polat, and M Recep Bozkurt. “A novel feature ranking algorithm for biometric recognition with PPG signals”. In: *Computers in biology and medicine* 49 (2014), pp. 1–14.

[80] Alex Hern. “Samsung Galaxy S8 iris scanner fooled by German hackers”. In: *The Guardian* 23 (2017).

[81] Debnath Bhattacharyya et al. “Biometric authentication: A review”. In: *IJUNESST* (2009).

[82] Simon Eberz et al. “Broken hearted: How to attack ECG biometrics”. In: (2017).

[83] Nima Karimian, Damon Woodard, and Domenic Forte. “ECG biometric: Spoofing and countermeasures”. In: *IEEE Transactions on Biometrics, Behavior, and Identity Science* 2.3 (2020), pp. 257–270.

[84] Abhijit Sarkar and et al. “Biometric authentication using photoplethysmography signals”. In: *BTAS*. 2016.

[85] Zhidong Zhao et al. “A human ECG identification system based on ensemble empirical mode decomposition”. In: *Sensors* (2013).

[86] Umang Yadav et al. “Evaluation of PPG biometrics for authentication in different states”. In: *ICB*. 2018.

[87] Petros Spachos and et al. “Feasibility study of photoplethysmographic signals for biometric identification”. In: *ICDSP*. 2011.

[88] Marcus Carlsson et al. “Total heart volume variation throughout the cardiac cycle in humans”. In: *AJP Heart and Circulatory Physiology* (2004).

[89] John E Hall. *Guyton and Hall textbook of medical physiology e-Book*. Elsevier Health Sciences, 2015.

[90] Angelo Bonissi et al. “A preliminary study on continuous authentication methods for photoplethysmographic biometrics”. In: *2013 BIOMS*. 2013.

[91] Gerhard Whitworth. *Why Do Athletes Have a Lower Resting Heart Rate?* 2018.

[92] A Choi and H Shin. “Photoplethysmography sampling frequency: pilot assessment of how low can we go to analyze pulse rate variability with reliability?” In: *Physiological measurement* (2017).

[93] Stephen Butterworth et al. “On the theory of filter amplifiers”. In: *Wireless Engineer* (1930).

[94] U. Farooq et al. “PPG delineator for real-time ubiquitous applications”. In: *EMBC*. 2010.

[95] Walter Karlen et al. “Photoplethysmogram processing using an adaptive single frequency phase vocoder algorithm”. In: *BIOSTEC*. 2011.

[96] M. Aboy et al. “An automatic beat detection algorithm for pressure signals”. In: *IEEE Transactions on Biomedical Engineering* (2005).

[97] Anand Chandrasekhar et al. “PPG sensor contact pressure taken into account for blood pressure measurement”. In: *IEEE Trans. Biomed. Eng.* (2020).

[98] Ali Siam and et al. *Real-World PPG dataset*. Mendeley data. http://dx.doi.org/10.17632/yynb8t9x3d. 2019.

[99] Alistair EW Johnson et al. “MIMIC-III, a freely accessible critical care database”. In: *Scientific data* (2016).

[100] Geoffrey E Hinton and Ruslan R Salakhutdinov. “Reducing the dimensionality of data with neural networks”. In: *science* (2006).

[101] Karl Pearson. “LIII. On lines and planes of closest fit to systems of points in space”. In: *The London, Edinburgh, and Dublin Philosophical Magazine and Journal of Science* (1901).

[102] Daniel D Lee and H Sebastian Seung. “Learning the parts of objects by non-negative matrix factorization”. In: *Nature* 401.6755 (1999), pp. 788–791.

[103] Joshua B Tenenbaum et al. “A global geometric framework for nonlinear dimensionality reduction”. In: *science* (2000).

[104] Sam T Roweis et al. “Nonlinear dimensionality reduction by locally linear embedding”. In: *science* (2000).

[105] Laurens van der Maaten et al. “Visualizing data using t-SNE”. In: *Journal of machine learning research* (2008).

[106] Yoshua Bengio. *How large should be the data set for training a Deep auto encoder?* https://www.quora.com/How-large-should-be-the-data-set-for-training-a-Deep-auto-encoder. 2011.

[107] Prasanta Chandra Mahalanobis. “On the generalized distance in statistics”. In: *National Institute of Science of India*. 1936.

[108] Kai Peng and et al. “Balanced iterative reducing and clustering using hierarchies with principal component analysis (PBIRCH) for intrusion detection over big data in mobile cloud environment”. In: *SpaCCS*. 2018.

[109] James MacQueen et al. “Some methods for classification and analysis of multivariate observations”. In: *Proceedings of the fifth Berkeley symposium on mathematical statistics and probability*. Oakland, CA, USA. 1967.

[110] Rakesh Agrawal et al. “Automatic subspace clustering of high dimensional data for data mining applications”. In: *SIGMOD*. 1998.

[111] Mihael Ankerst et al. “OPTICS: ordering points to identify the clustering structure”. In: *ACM Sigmod record* (1999).

[112] Martin Ester et al. “A density-based algorithm for discovering clusters in large spatial databases with noise.” In: *Kdd*. 1996.

[113] Gholamhosein Sheikholeslami et al. “WaveCluster: a wavelet-based clustering approach for spatial data in very large databases”. In: *VLDB Journal* (2000).

[114] Sander Koelstra et al. “Deap: A database for emotion analysis; using physiological signals”. In: *IEEE transactions on affective computing* (2011).

[115] Walter Karlen et al. “Multiparameter respiratory rate estimation from the photoplethysmogram”. In: *IEEE Transactions on Biomedical Engineering* (2013).

[116] *Biosec.lab ppg dataset*. http://www.comm.utoronto.ca/Ëœbiometrics/PPG_Dataset/.

[117] J. Yao and et al. “A Pilot Study on Using Derivatives of Photoplethysmographic Signals as a Biometric Identifier”. In: *EMBC*. 2007.

[118] Vikram Chandrasekaran et al. “Cuffless differential blood pressure estimation using smart phones”. In: *IEEE Transactions on Biomedical Engineering* (2012).

[119] Gaobo Zhang et al. “A noninvasive blood glucose monitoring system based on smartphone PPG signal processing and machine learning”. In: *IEEE Trans. Ind. Inform.* (2020).

[120] Sarah Ali Siddiqui et al. “A pulse rate estimation algorithm using PPG and smartphone camera”. In: *Journal of medical systems* (2016).

[121] Sairul I Safie and et al. “Electrocardiogram (ECG) biometric authentication using pulse active ratio (PAR)”. In: *TIFS* (2011).

[122] Hugo Plácido Da Silva et al. “Finger ECG signal for user authentication: Usability and performance”. In: *BTAS*. 2013.

[123] Yogendra Narain Singh and Phalguni Gupta. “ECG to individual identification”. In: *2008 IEEE Second International Conference on Biometrics: Theory, Applications and Systems*. 2008.

[124] Giulio Lovisotto et al. “Seeing red: PPG biometrics using smartphone cameras”. In: *CVPR Workshops*. 2020.

[125] Weizheng Wang et al. “CardioID: Mitigating the Effects of Irregular Cardiac Signals for Biometric Identification”. In: *EWSN*. 2022.

[126] Tanner Helland. *How to Convert Temperature (K) to RGB: Algorithm and Sample Code*. https://tannerhelland.com/2012/09/18/convert-temperature-rgb-algorithm-code.html. 2012.

[127] *Gamma Correction*. https://en.wikipedia.org/wiki/Gamma_correction.

[128] Sergey Chernenko. *Alpha-trimmed mean filter*. \http://www.librow.com/articles/article-7. Last accessed: Sept 2020.

[129] Jan Spooren et al. “Ppg 2 live: Using dual ppg for active authentication and liveness detection”. In: *IEEE ICB*. 2019.

[130] Y. Y. Gu and Y. T. Zhang. “Photoplethysmographic authentication through fuzzy logic”. In: *IEEE EMBS Asian-Pacific Conference on Biomedical Engineering, 2003.* Oct. 2003. DOI: 10.1109/APBME.2003.1302621.

[131] Nam Bui et al. “Pho2: Smartphone based blood oxygen level measurement systems using near-ir and red wave-guided light”. In: *Proceedings of the 15th ACM conference on embedded network sensor systems*. 2017, pp. 1–14.

[132] Philips. *Philips Lighting deploys LED-based indoor positioning in Carrefour hypermarket*. 2015. URL: https://goo.gl/a0tGIj.

[133] Liqun Li et al. “Epsilon: A visible light based positioning system”. In: *11th USENIX Symposium on Networked Systems Design and Implementation (NSDI 14)*. 2014, pp. 331–343.

[134] Ye-Sheng Kuo, Pat Pannuto, and et. al. “Luxapose: Indoor Positioning with Mobile Phones and Visible Light”. In: *Proceedings of the ACM MobiCom*. 2014.

[135] Chi Zhang et al. “Extending Mobile Interaction Through Near-Field Visible Light Sensing”. In: *Proceedings of the ACM MobiCom*. 2015.

[136] Yanbing Yang et al. “CeilingSee: Device-free occupancy inference through lighting infrastructure based LED sensing”. In: *Proceedings of the IEEE PerCom*. 2017.

[137] Chi Zhang and Xinyu Zhang. “Pulsar: Towards ubiquitous visible light localization”. In: *Proceedings of the ACM MobiCom*. 2017, pp. 208–221.

[138] Tianxing Li and Xia Zhou. “Battery-Free Eye Tracker on Glasses”. In: *Proceedings of the ACM MobiCom*. 2018, pp. 67–82.

[139] PureLiFi. *PureLiFi*. 2012. URL: `http://purelifi.com`.

[140] 802.11.bb. *IEEE 802.11 Light Communications*. 2018. URL: `http://www.ieee802.org/11/Reports/tgbb_update.htm`.

[141] G. Corbellini. “Connecting networks of toys and smartphones with visible light communication”. In: *IEEE Communications Magazine* (2014).

[142] E. Mok and G. Retscher. “Location determination using WiFi fingerprinting versus WiFi trilateration”. In: *Journal of Location Based Services* (2007).

[143] Jiangtao Li et al. “Retro-VLC: Enabling battery-free duplex visible light communication for mobile and IoT applications”. In: *Proceedings of the 16th International Workshop on Mobile Computing Systems and Applications*. 2015, pp. 21–26.

[144] Xieyang Xu et al. “Passivevlc: Enabling practical visible light backscatter communication for battery-free iot applications”. In: *Proceedings of the ACM MobiCom*. 2017, pp. 180–192.

[145] P Fahamuel, J Thompson, and H Haas. “Study, analysis and application of optical OFDM, single carrier (SC) and MIMO in intensity modulation direct detection (IM/DD)”. In: (2013).

[146] Qing Wang, Marco Zuniga, and Domenico Giustiniano. “Passive Communication with Ambient Light”. In: *Proceedings of the ACM CoNEXT*. 2016, pp. 97–104.

[147] Hongjia Wu et al. “SmartVLC: When Smart Lighting Meets VLC”. In: *Proceedings of the ACM CoNEXT*. 2017, pp. 212–223.

[148] L. Klaver and M. Zuniga. “Shine: a Step Towards Distributed Multi-Hop Visible Light Communication”. In: *Proceedings of the IEEE MASS*. 2015.

[149] Moustafa Youssef, Matthew Mah, and Ashok Agrawala. “Challenges: Device-free Passive Localization for Wireless Environments”. In: *Proceedings of the ACM MobiCom*. 2007.

[150] M. Moussa and M. Youssef. “Smart devices for smart environments: Device-free passive detection in real environments”. In: *Proceedings of the IEEE PerCom*. 2009.

[151] M. Youssef, A. Agrawala, and A. Shankar. “WLAN location determination via clustering and probability distributions”. In: *Proceedings of the IEEE PerCom*. 2003.

[152] Fadel Adib, Zachary Kabelac, and Dina Katabi. “Multi-Person Localization via RF Body Reflections”. In: *Proceedings of the USENIX NSDI*. 2015.

[153] I. Sabek and et al. “ACE: An Accurate and Efficient Multi-Entity Device-Free WLAN Localization System”. In: *IEEE Transactions on Mobile Computing* (2015).

[154] Liqun Li et al. “Epsilon: A Visible Light Based Positioning System”. In: *Proceedings of the USENIX NSDI*. 2014.

[155] Zhice Yang and et. al. “Wearables Can Afford: Light-weight Indoor Positioning with Visible Light”. In: *Proceedings of the ACM MobiSys*. 2015.

[156] Shimin Xie Bo and Gong and Guang Tan. “LiPro: light-based indoor positioning with rotating handheld devices”. In: *Wireless Networks* (2016).

[157] J. M. Kahn and J. R. Barry. “Wireless infrared communications”. In: *Proceedings of the IEEE* 85.2 (Feb. 1997), pp. 265–298. ISSN: 0018-9219.

[158] Ali D. Firoozabadi et al. “A Novel Frequency Domain Visible Light Communication (VLC) Three-Dimensional Trilateration System for Localization in Underground Mining”. In: *Applied Sciences* (2019).

[159] Elena Di Lascio et al. “Poster Abstract: LocaLight - A Battery-Free Passive Localization System Using Visible Light”. In: *Proceedings of IPSN*. 2016.

[160] Yanbing Yang et al. “Counting via LED sensing: Inferring occupancy using lighting infrastructure”. In: *Elsevier Pervasive and Mobile Computing* (2018), pp. 35–54.

[161] B. Zhang et al. “SBVLC: Secure Barcode-Based Visible Light Communication for Smartphones”. In: *IEEE Transactions on Mobile Computing* (2016), pp. 432–446.

[162] Rens Bloom et al. “Tweeting with Sunlight: Encoding Data on Mobile Objects”. In: *Proceedings of the IEEE INFOCOM*. 2019, pp. 1324–1332.

[163] Hang Guo and John Heidemann. “Detecting iot devices in the internet”. In: *IEEE/ACM Transactions on Networking* 28.5 (2020), pp. 2323–2336.

[164] Chenshu Wu et al. “msense: Towards mobile material sensing with a single millimeter-wave radio”. In: *Proceedings of the ACM on Interactive, Mobile, Wearable and Ubiquitous Technologies* 4.3 (2020), pp. 1–20.

[165] Yumeng Liang et al. “Fg-liquid: A contact-less fine-grained liquid identifier by pushing the limits of millimeter-wave sensing”. In: *Proceedings of the ACM on Interactive, Mobile, Wearable and Ubiquitous Technologies* 5.3 (2021), pp. 1–27.

[166] Maria Hayes. “Measuring protein content in food: An overview of methods”. In: *Foods* 9.10 (2020), p. 1340.

LIST OF PUBLICATIONS

1. **Weizheng Wang**, Girish Vaidya, Anup Bhattacharjee, Francesco Fioranelli, Marco Zuniga (2023). A Long-Term Study of mmWave Sensing in an Outdoor Urban Scenario. The 19th Annual International Conference on Distributed Computing in Smart Systems and the Internet of Things (DCOSS-IoT).

2. **Weizheng Wang**, Qing Wang, Marco Zuniga (2023). Taming Irregular Cardiac Signals for Biometric Identification. ACM Transactions on Sensor Networks (TOSN).

3. **Weizheng Wang**, Qing Wang, Marco Zuniga (2022). CardioID: Mitigating the Effects of Irregular Cardiac Signals for Biometric Identification (Best Paper Runner-up). International Conference on Embedded Wireless Systems and Networks (EWSN).

4. **Weizheng Wang**, Marek Vette, Qing Wang, Jie Yang, Marco Zuniga (2022). CamPressID: Optimizing Camera Configuration and Finger Pressure for Biometric Authentication. In IEEE 19th International Conference on Mobile Ad Hoc and Sensor Systems (MASS).

5. **Weizheng Wang**, Qing Wang, Junwei Zhang, Marco Zuniga (2020). PassiveVLP: Leveraging Smart Lights for Passive Positioning. ACM Transactions on Internet of Things (TIOT).

6. **Weizheng Wang**, Qing Wang, Junwei Zhang, Marco Zuniga (2018). Leveraging smart lights for passive localization. In IEEE 19th International Conference on Mobile Ad Hoc and Sensor Systems (MASS).