October 2018

MULTI-SENSOR LOCALIZATION AND TRACKING IN DISASTER MANAGEMENT AND INDOOR WAYFINDING FOR VISUALLY IMPAIRED USERS

Zhuorui Yang

MULTI-SENSOR LOCALIZATION AND TRACKING IN DISASTER MANAGEMENT AND INDOOR WAYFINDING FOR VISUALLY IMPAIRED USERS

A Dissertation Presented

by

ZHUORUI YANG

Submitted to the Graduate School of the
University of Massachusetts Amherst in partial fulfillment
of the requirements for the degree of

DOCTOR OF PHILOSOPHY

September 2018

Electrical and Computer Engineering

MULTI-SENSOR LOCALIZATION AND TRACKING IN DISASTER MANAGEMENT AND INDOOR WAYFINDING FOR VISUALLY IMPAIRED USERS

A Dissertation Presented

by

ZHUORUI YANG

Approved as to style and content by:

_____
Aura Ganz, Chair

_____
C. Mani Krishna, Member

_____
Marco F. Duarte, Member

_____
Song Gao, Member

_____
Christopher V. Hollot, Department Head
Electrical & Computer Engineering

ACKNOWLEDGEMENTS

First of All, I would like to express my heartfelt appreciation to my advisor, Professor Aura Ganz, for her wisdom, mentorship, unlimited patience, and encouragement. Without her guidance, I would not have been able to thrive in my doctoral program and succeed in my doctoral research.

To my doctoral dissertation committee: Professor C. Mani Krishna, Professor Marco F. Duarte, and Professor Song Gao, I appreciate the support during my dissertation research, and general advice you provided about academic writing. Special thanks go to Professor Marco F. Duarte for his constructive advice and insightful suggestions during the development of the crowd-resilient vision-based localization algorithm.

I would also like to thank the members of the 5G Mobile Evolution Lab: James Schafer, Hao Dong, Yang Tao, Jun Yi, Jingyan Tang, Dongyi Yang, Sili Wang, Yang Li, Juechen Yi and Quan Shi. Thank you to the fantastic team I have been fortunate to work with over the years.

Finally, I would like to thank my parents, and my loved one, Binru Cao. Your love, understanding, and support motivate me to persist and complete this dissertation.

ABSTRACT

MULTI-SENSOR LOCALIZATION AND TRACKING IN DISASTER MANAGEMENT AND INDOOR WAYFINDING FOR VISUALLY IMPAIRED USERS

SEPTEMBER 2018

ZHUORUI YANG

B.S., KUNMING UNIVERSITY OF SCIENCE AND TECHNOLOGY, P.R. CHINA

M.S., UNIVERSITY OF MASSACHUSETTS AMHERST

Ph.D., UNIVERSITY OF MASSACHUSETTS AMHERST

Directed by: Professor Aura Ganz

This dissertation proposes a series of multi-sensor localization and tracking algorithms particularly developed for two important application domains, which are disaster management and indoor wayfinding for blind and visually impaired (BVI) users. For disaster management, we developed two different localization algorithms, one each for Radio Frequency Identification (RFID) and Bluetooth Low Energy (BLE) technology, which enable the disaster management system to track patients in real-time. Both algorithms work in the absence of any pre-deployed infrastructure along with smartphones and wearable devices. Regarding indoor wayfinding for BVI users, we have explored several types of indoor positioning techniques including BLE-based, inertial, visual and hybrid approaches to offer accurate and reliable location and orientation in complex navigation spaces. In this dissertation, significant contributions have been made in the design and implementation of various localization and tracking algorithms under different requirements of certain applications.

TABLE OF CONTENTS

Page

LIST OF TABLES

LIST OF FIGURES

Figure Page

CHAPTER 1

INTRODUCTION

1.1 Applications

1.1.1 Disaster Management

Catastrophic disasters from natural hazards and terrorist attacks inevitably result in large numbers of causalities. Effectively managing the mass casualty incidents (MCI) is one of the most crucial problems must be addressed. Among five phases in disaster management cycle [1], our focus is on the response phase in which multiple rescue organizations will engage on-site search and rescue missions. During disaster response to MCI, a standard flow including triage, treatment and transport is expected to be accomplished by multiple rescue teams jointly and seamlessly. Therefore, an efficient coordination system, which can streamline information sharing, situation awareness, decision support and collaboration among agencies is desired. Many efforts have been made in attempt to build such disaster management systems, such as DIORAMA [2, 3], WIISARD [4], BMIST [5], TACMED-CS [6], CodeBlue [7], AID-N [8] etc.

In terms of information shared among different agencies, it is widely known that associating geospatial components with entities in real time contributes to clarifying the chaos of the disaster while the disaster scene is in flux; such entities encompass patients, responders and other rescue resources. Moreover, such abilities should considerably alleviate the stress for incident commanders (IC) when the patient identification and tracking arise to be a challenge as search and rescue teams locate increasing numbers of patients [9]. Thereby, assuming that responders' locations are known, the most important issue needs to be solved is how to develop feasible localization techniques which can be

used to determine patients' locations given the requirements of disaster response and the restrictions of disaster scene. Such localization techniques must meet the following requirements: 1) it cannot rely on any pre-deployed infrastructure since it can be disrupted or damaged; 2) it needs to be rapidly deployable because disaster management is a time-critical mission; 3) it must provide adequate accuracy for responders locating patients visually; 4) the devices involved must be reliable and portable; 5) the entire solution has to be scalable since the extent of damage is unpredictable; 6) This solution needs to be cost-efficient due to applicability. Nevertheless, none of the abovementioned systems [4-8] except DIORAMA are capable of tracking patients while meeting all these requirements. As part of the novelties of DIORAMA, we present two localization algorithms that are particularly designed for on-site patient tracking during disaster response with two technologies, RFID and BLE. Due to the difference of their radio propagation characteristics, we design a deterministic approach for the RFID-based algorithm and a probabilistic approach for the BLE-based one.

As we apply RFID-based system to the outdoor disaster scene, we assume that each responder equips an Android smartphone which has access to Global Positioning System (GPS) and a wearable active RFID reader. Unlike RFID-based system, the BLE-based system reduces the requirement of hardware by using smartphones as readers which can measure the signal strength of BLE readings. When the responders start to conduct triage, each patient will be tagged with a D-tag which is composed of a conventional paper triage tag, an embeddable passive RFID tag and an active RFID/BLE tag. Prior to be deployed to the scene, the passive RFID tag and the active RFID/BLE tag are paired to each other and this mapping information is stored in DIORAMA system. Thus, the unique serial number

of passive RFID tag can be leveraged as the identifier of the patient and the readings transmitted from the active RFID/BLE tag carried by this patient. From this moment, the signal readings collected by responders will be bundled with the GPS readings from responders' smartphones. As to the GPS readings, they are leveraged as anchor information for localization algorithms, which denotes the locations where the signal readings are collected. With more and more incoming readings transmitted from responders' smartphones to the server, our localization algorithms will enable to calculate patients' locations in real-time. Once the locations are calculated, DIORAMA will share this information with all rescue members. As shown in Figure 1.1 and Figure 1.2, the up-to-date patients' locations and their corresponding triage priorities can be visualized by responders using smartphones. More importantly, regarding user experience, the only actions for responders to take for using DIORAMA contains only two steps, the first step is to scan the passive tag with smartphones and second step is to select the triage level of the patient on the screen. This entire process is anticipated to be finished in a few seconds.

In regard to urban search and rescue (USAR) mission, it poses another technical challenge to us due to inaccessibility of GPS signal or unavailability of accurate GPS readings. Lacking anchor information from GPS readings will lead to the failure of tracking patients. Thus, we utilize a vision-based localization algorithm as an alternative to the GPS readings so that the anchor information can be recovered. By this means, we extend the applicability of DIORAMA to USAR mission.

With the assistance of tracking patients in disaster response using DIORAMA, not only the time spent on searching the patients during treatment phase is reduced significantly, but also the order of patient being evacuated during treatment phase can follow the severity

of triage levels exactly. As a result, from many experiments and drills conducted by certified responders, DIORAMA demonstrates itself as an effective disaster management system which can be used to save people' lives.

Figure 1.1: Illustration of patient tracking on responder's smartphone application

Figure 1.2: Illustration of patient tracking on IC's tablet application

1.1.2 Indoor Wayfinding for Visually Impaired Users

According to the statistics of visual impairment and blindness from the World Health Organization (WHO), there are 285 million people suffering from visual impairment worldwide [10]. Indoor wayfinding in complex public spaces poses a major challenge to blind and visually impaired (BVI) individuals and negatively affects their

mobility and quality of life. Such challenges for BVI can be 1) Travelling inside an unfamiliar building solely without the assist of signs to indicate how to reach your destination step by step. 2) Having no way to orient themselves in the space and navigate to the destination safely and confidently. 3) When the fire alarm is ringing, how to exit the building in a short time? Given these situations, what sighted people take for granted turns out to be obstacles which BVI individuals have to encounter and try to overcome every day.

By receiving training from Orientation and Mobility (O&M) specialists, the BVI users are taught how to use their senses and cognition to familiarize the spaces with white canes or guide dogs. However, these skills are not sufficient for BVI users and it leaves more room for assistive technology to help them navigation inside the spaces by augmenting their perception to the surroundings using different sensors and algorithms. For example, as illustrated in Figure 1.3 and Figure 1.4, with the assistance of PERCEPT, BVI users enable to navigate themselves independently from one point of interest to another in multiple spaces, such as administrative buildings and subway stations. According to [11], the problem of indoor wayfinding for BVI users can be divided into three sub research areas: 1) indoor positioning 2) path planning 3) communication and interaction. Because indoor positioning is the basis and the key of the entire problem, the primary focus in this dissertation is on developing different indoor positioning algorithms with different sensors. Additionally, substantial contributions have also been made to system integration and testing.

Unlike the mature outdoor navigation using GPS, indoor navigation cannot borrow the solution directly due to its global navigation satellite system (GNSS)-denied attribute

in natural. In the wake of developments of indoor positioning for BVI users, countless attempts have been made, but still, there is no available solution that is universal, affordable, portable and reliable enough for BVI users. In the past decade, a variety of technologies have been applied to solve this problem, such as Radio Frequency (RF)-based, inertial, visual and hybrid approaches. Nevertheless, the solutions mostly are about building some costly and bulky prototype systems that can only work in a lab. Until recently, the smartphone becomes a powerful platform which has enough computing power with useful sensors, it exhibits a promising interface from which the BVI users can use it to sense the world. Towards this direction, we develop several different smartphone-based positioning solutions, which contain BLE-based, visual and hybrid approach, using the knowledge of computer vision, signal processing, statistical inference, and machine learning.

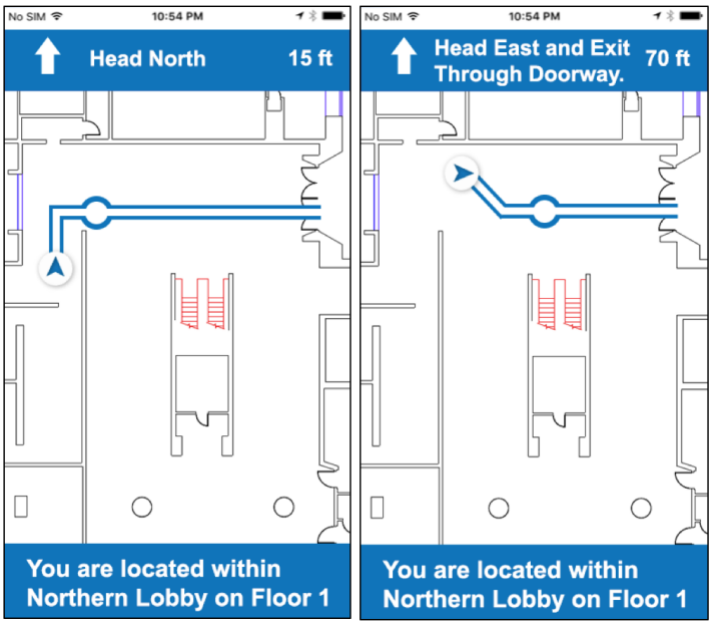


Figure 1.3: Example of indoor wayfinding for BVI users using PERCEPT

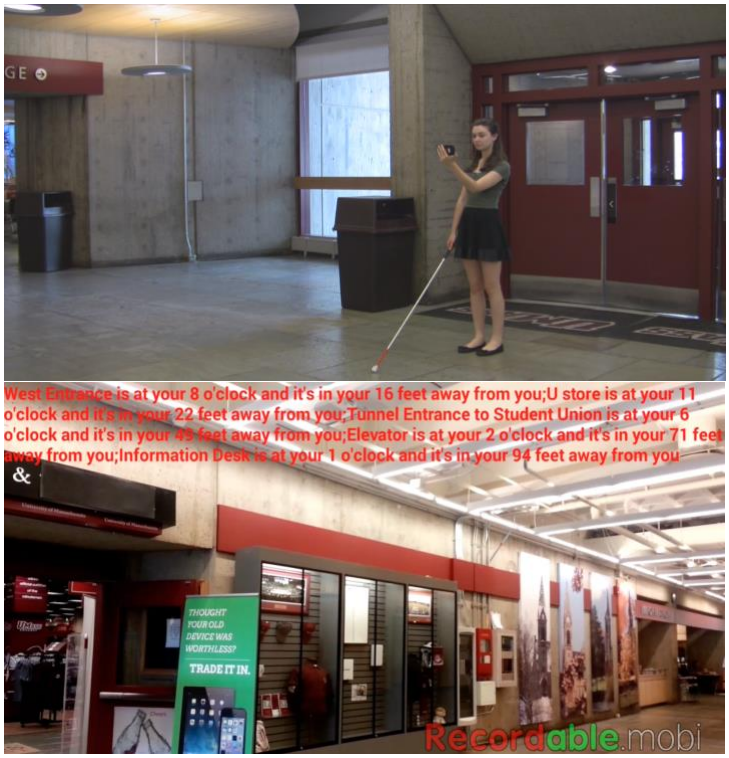


Figure 1.4: Example of indoor wayfinding for BVI users using PERCEPT-V

1.2 Related Work

1.2.1 Patient Tracking in Disaster Management

Real-time patient tracking in disaster scene is critical to disaster management, which contributes to resolving the chaos of flooding information that is overwhelming to rescue personnel. Given the requirements of disaster management and the limitations of the disaster scenes, researchers are facing great challenges to develop suitable solutions for patient tracking. Despite numerous localization and tracking solutions for various purposes having been developed in the past, works relevant to designing an applicable and suitable solution to handle disaster situations are still sparse.

Prior to discussing the related work, we believe that it is necessary to state the reason why the GPS readings from user smartphones does not suffice for the application of interest because the GPS system seems to be a perfect solution already. However, the

solution to patient tracking is not as obvious as it seems to be for two reasons. First, it is almost impossible to hold the assumption that every patient keeps their smartphones with themselves during and after the disaster. It is inevitable that patients' smartphones can be either broken or lost. If rescue teams spend their time on locating the smartphones, their efforts can be in vain. Second, even if patients have their smartphones with them and the smartphones function properly, this idea is still problematic because the existing communication infrastructure is not prepared to handle the large number of requests at the same time. Not to say that currently there is no standard way to share these information with rescue personnel. Thereby, there is no doubt that tracking patients by their GPS readings from their smartphones [14-17] is flawed and it is necessary for the researchers to develop particular solutions of patient tracking.

Among the limited works we found, we categorize them by the technology used to track patients. In [18-20, 2, 21-23], different Radio Frequency (RF) – based approaches are presented. In [18-20], authors introduced an emergency response system based on a location aware wireless sensor network (WSN) to assist responders to track patients. More specifically, this system needs to setup the infrastructure for ZigBee-enabled Disaster Aid Network (DAN) on disaster site, which means the deployment will take quite some time and it makes it unsuitable to the time-critical rescue mission. Apart from that, the cost of deploying such system is high since it requires hundreds of nodes to be distributed across the area. Furthermore, even though the location accuracy shown in the chapter is about 1.64 ft., this result is achieved by simulations with unrealistic parameter settings instead of the more convincible field tests. Hence, the feasibility of such system in disaster management is questionable. In [21], authors developed an RFID-based patient tracking

solution working with a prototype radar. Although the location accuracy is about 7 ft. from field tests, the bulky and expensive radar system doesn't make it suitable to the disaster environment, not to say that the scalability is limited by using the cumbersome devices. Besides, previous patient tracking solutions used by DIORAMA have relied on RFID-based technology. In [22, 23, 2], three different localization algorithms are proposed and showed good result of location accuracy. However, since the purpose of the study is about proof of concept, the scalability, the cost and several other issues of the system are not considered and solved. Thus, these solutions still required significant improvements and tests before they can be ready for the disaster management. We believe that our works, including the new RFID-based localization and BLE-based localization algorithms outperform other approaches in terms of applicability and suitability to patient tracking in disaster management while achieving reasonable accuracy. As shown in Table 1.1, we summarize all RF-based works and use the comparison table to exhibit the advantages of our works [24, 25] compared to others.

Except the RF-based approaches, several cellular-based approaches are discussed in [26-29]. The cellular infrastructure varies from GSM [26, 27] to LTE [28] and 5G [29]. As mentioned, the idea of tracking patients by locating patients' smartphones can be problematic due to various reasons. Other approaches like visual Simultaneous Localization and Mapping (SLAM) [30, 31] are also popular for patient tracking using robotic systems but it is not strongly related to what we have done.

Table 1.1: Comparison of different RF-based solutions of patient tracking

	Hardware	**Accuracy**	**Field Test**	**Scalability**	**Cost**	**Infrastructure required (fixed anchors)**
[18]	WSN, ZigBee	< 1.64 ft.	Simulation	Limited	High	Yes
[21]	Active RFID and radar	7 ft.	Yes	No	High	Yes
[22]	Active RFID	12 ft.	Yes	Limited	Medium	Yes
[2]	Active RFID	11 ft.	Yes	Limited	Medium	Yes
[23]	Active RFID	< 20 ft.	Limited	Limited	Medium	No
[24]	Active RFID	15 ft.	Yes	Yes	Medium	No
[25]	BLE	11 ft.	Yes	Yes	Low	No

1.2.2 Localization and Tracking in Indoor Wayfinding for BVI Users

The crux of the indoor wayfinding for BVI users is to develop an affordable indoor positioning solution which can provide accurate and reliable estimations of location and orientation in different navigation spaces. In the last decade, many attempts have been made using various approaches. Among them, RF-based, inertial, visual and hybrid approaches are the most prevalent ones. And we are going to go over the related work based on the technologies used in each solution. We summarize all the works into Table 1.2.

Regarding RF-based approaches, the methodologies can be categorized into two main types, the fingerprint-based localization and non-fingerprint-based localization. As we go through the literatures published from 2010, there is no fingerprint-based localization reported in developing indoor wayfinding systems for BVI users. The main reason behind this is that the process of building fingerprints takes significant amount of

time and it is hard to maintain the fingerprints. Even though the performance of fingerprint-based localization usually better than that of non-fingerprint-based localization, the environment-dependent property makes it unfavorable to the community. On the other hand, in [32-38], the methodologies of non-fingerprint-based localization contain trilateration, triangulation, proximity and their derivative approaches, which enables to compute a coarse location estimation reliably if the sensors are properly deployed in the space. However, no works covered above except ours have considered the problem of optimal sensor deployment in spite of its strong impact to the performance of localization algorithm. Besides, unlike what we have accomplished, no works propose any means of estimating user's orientation using RF-based techniques. Furthermore, we are one of the few systems that has been tested by BVI users and we are the only one tested in large and open spaces, e.g. multi-floor subway station (North Station at Boston).

In terms of inertial positioning algorithms, the state-of-art methodology, dead reckoning, is known by the community for a long time. Dead reckoning can perform well in a short time, however, the problem for dead-reckoning is that the accumulated drift is growing quickly while time increases. To circumvent this issue, people have proposed to reset the accumulated errors periodically by detecting the strides of the user using a specific IMU mounted on the foot, which makes it uncomfortable to the users. This technique is called Zero Velocity Update (ZUPT). Another way of designing an inertial positioning algorithm is to let dead reckoning work with a highly accurate but expensive IMU, which leads to be an unaffordable solution to the users. In [39-41], the inertial positioning algorithms are used in the development of indoor wayfinding systems for BVI users.

With respect to the visual approaches, the methodologies mostly belong to three categories, Visual Simultaneous Localization and Mapping (V-SLAM), visual odometry and pattern recognition of pre-defined features in the space. SLAM is a technique used to build map for the unknown spaces and localize the agent within the spaces at the same time. In order to obtain reasonable accuracy of map and location, it is necessary to close the loop in SLAM, which means the users have to revisit the same locations on purpose. Unlike robots, the BVI users are unable to work with V-SLAM because of this reason. As what are showed in [42, 43], the devices are usually desired to obtain extra measurements than the visual ones, for instance depth information, using RGB-D cameras which are expensive and unaffordable to the BVI users. As to the pattern recognition approaches in indoor wayfinding for BVI users, there are two types of reference information which can be used. The first one is specifically designed visual markers [44, 69] deployed in the points of interests and the second one is natural appearances already in the spaces selected by the algorithm designers [45-47], such as doors or signage. In order to make it work, both types of patterns will have to be defined deliberately for certain space, which requires significant manual adjustments and limits its applicability to different spaces. In terms of visual odometry, it provides highly accurate location and orientation estimations by comparing images taken by the user at a particular location with those in an image database for the navigation space, covering its entire extent. More specifically, the same keypoint descriptors can be detected in a reference image and the currently observed image so that the camera user can be localized with respect to the reference image by (1) leveraging information about the camera pose of the reference image and (2) applying a suitable spatial transformation to the location and rotation of the features on the reference image. In [48-

50], different algorithms were developed based on the similar principle. However, our works [51, 52] improve upon the current visual odometry approaches by adding crowd resilience ability to the algorithm so that our algorithm is more robust than others when it is running in heavily used indoor public spaces.

Finally, the hybrid approaches, which use different combinations of sensor inputs, are also prevalent to the community. The state-of-art methodologies are the Kalman Filter (KF), Particle Filter (PF) and their derivatives, such as Extended Kalman Filter (EKF), Unscented Kalman Filter (UKF) etc. Many works have been conducted using these methodologies, such as [53-68]. The most challenging problem in these methodologies is to find the proper parameter settings for the entire dynamic system.

Table 1.2: Comparison of different indoor positioning solutions in indoor wayfinding for BVI users

Main Technology	Positioning Methodology	Input	Location	Orientation	Real-time	Environmental Change		Devices	
						Sensors	No sensors	CM	Others
RF-based	Non-fingerprint	BLE	[*]	[*]	[*]	[*]	N/A	[*]	N/A
		others	[32-38]	[35]	[32-35, 37,38]	[32-38]	N/A	[34, 38]	[32, 33, 35-37]
Visual	V-SLAM	Natural visual appearance	[42, 43]	[42, 43]	[43]	N/A	[42, 43]	N/A	[42, 43]
	Odometry	Natural visual appearance	[48-52]	[48-52]	[48-51]	N/A	[48-52]	[49-52]	[48]
	Pattern Recognition	Fiducial markers	[44]	[44]	[44]	[44]	N/A	[69]	[44]
		Natural visual appearance	[45-47]	[46]	[46]	N/A	[45-47]	[46]	[45, 47]
Inertial	Dead reckoning	IMU	[39-41]	[39-41]	[40]	N/A	[39-41]	[39, 41]	[40]
Hybrid	Bayesian filter-based approaches and others	W+I	[53, 54]	[53, 54]	[53]	[53, 54]	N/A	[53, 54]	N/A
		W+V	[55, 56]	[55, 56]	[56]	[55, 56]	N/A	[56]	[55]
		I+V	[57-64]	[57-64]	[57, 59, 62, 64]	N/A	[57-64]	[58, 61]	[57, 59, 60, 62-64]
		W+I+V	[65]	[65]	[65]	[65]	N/A	N/A	[65]
		Others	[66, 68]	[66-68]	[66]	[68]	[66, 67]	N/A	[66-68]

*denotes our BLE works in preparation
CM refers to commercial devices, which includes smartphones and Google Glass only.
W refers to RF-based input, I is short for IMU, V means visual input.

1.3 Our Approaches and Contributions

This dissertation exploits diverse technologies and techniques to provide localization and tracking in disaster management and indoor wayfinding for BVI users. In disaster management, we developed two localization algorithms for patient tracking using RFID and BLE. Both localization algorithms enable responders to locate patients rapidly and conveniently, which can reduce mortality of causalities in the disaster by saving rescue time and keeping treatment order. At the beginning, we built the localization algorithm

with RFID technology while BLE technology is not available on the market at that time. RFID technology, which is a propriety technology working in less noisy band, brings us a decent localization performance due to its reliable propagation characteristics compared to BLE technology. As BLE technology is emerging recently, it provides us the chance to reduce the hardware cost of DIORAMA system even further, so DIORAMA system can be more cost-efficient. However, since BLE technology is working at 2.4 GHz, it is noisier than the band used by RFID technology. To overcome the deficiency of the propagation characteristics of BLE technology, we developed another localization algorithm for BLE technology. As a result, the new localization algorithm achieves comparative performance as that of RFID-based algorithm.

With respect to indoor wayfinding for BVI users, we have explored many possibilities using BLE-based, inertial, visual and hybrid approaches with commercial smartphones. Our aim is to find a suitable, universal, affordable tracking solution for diverse spaces in which BVI users can travel independently with the assistance of PERCEPT system. First, for multi-floor subway station, e.g. North Station in Boston, we developed a BLE-based solution encompassing localization module, orientation module, proximity detection module and optimal sensor deployment module. Second, for large and open space, e.g. UMass campus center main floor, we developed a visual localization algorithm with resilience to crowd, which enables BVI users to use commercial devices, such as smartphones or Google Glass, to navigate themselves in the space. These two spaces are quite distinctive and representative, which make them perfect candidates as testing beds. In the future, we plan to continue test our algorithms using testing datasets collected at these two spaces.

The contributions of this dissertation are the series of localization and tracking algorithms specifically developed for disaster management and indoor wayfinding for BVI users. These localization and tracking solutions are designed appropriately under different requirements of applications.

1.3.1 RFID-based Localization Algorithm for Patient Tracking

Contributions:

Provide real-time patient tracking for responders and Incident Commander (IC) using DIORAMA. The localization solution is infrastructure-less, calibration-free, affordable, portable, reliable and accurate. The accuracy achieved is comparable to GPS accuracy with much lower cost.

1.3.2 BLE-based Localization Algorithm for Patient Tracking

Contributions:

Reduce the cost of hardware significantly, the RFID readers which costs $800 for each are removed from DIORAMA. Moreover, a non-propriety solution is provided. The new design of the algorithm increases the flexibility of the algorithm. The accuracy achieved is as good as what RFID-based solution does.

1.3.3 Crowd-resilient Visual Positioning with Background Extraction

Contributions:

Build the prototype of PERCEPT-V system. Provide highly accurate and reliable estimations of location and orientation in the presence of crowd in public spaces by taking a picture/video using smartphones or Google glasses. Develop the path planning strategy for large and open space. Provide real-time audible navigation instructions and vision-free interaction for BVI users.

1.3.4 Intelligent Sensing Framework for Indoor Spatial Awareness for BVI

Contributions:

Provide real-time indoor positioning including location and moving direction estimation for PERCEPT system. Augment users' cognitions with landmark proximity detection algorithm, which enables PERCEPT to detect the nearby landmarks for BVI users. Optimize the deployment strategy of BLE sensors for navigation spaces, which balance the performance of indoor positioning algorithm and the hardware cost.

1.3.5 Sensor Fusion for Pedestrian Indoor Localization

Contributions:

Expand the applicability of indoor positioning algorithm to different spaces. Make localization algorithm more flexible to the navigation spaces by learning the unique features of the spaces. Perform accurately and reliably in different spaces by fusing different sensors input together.

1.4 Dissertation Outline

The remaining of the dissertation is organized as follows. In Chapter 2, we start by introducing the **RFID-based localization algorithm for patient tracking** used in DIORAMA. In Chapter 3, we develop a different **BLE-based localization algorithm for patient tracking** used in DIORAMA to lower the hardware cost of system. Then, the context will switch from DIORAMA to PERCEPT. In Chapter 4, to track the location and orientation of users and increase the robustness of the algorithm in large and open spaces, e.g. shopping malls or public transportation hubs, we develop a **crowd-resilient visual localization algorithm with background extraction**, which increases the resilience to the crowd in the space. In Chapter 5, we present a **BLE-based intelligent framework for**

indoor spatial awareness for BVI, it contains four components, which are the optimal sensor deployment generation, localization, moving direction estimation and landmark proximity detection. In Chapter 6, we develop a **sensor fusion algorithm using vision, BLE and smartphone sensors** to improve the accuracy of BLE-based localization. The multi-modal data and complementary techniques are fused using Unscented Kalman Filter (UKF) [12]. In Chapter 7, we discuss the future work can be done in outdoor patient tracking in disaster management, pedestrian indoor localization and indoor wayfinding assistance for BVI. The dissertation outline is shown in Figure 1.5.

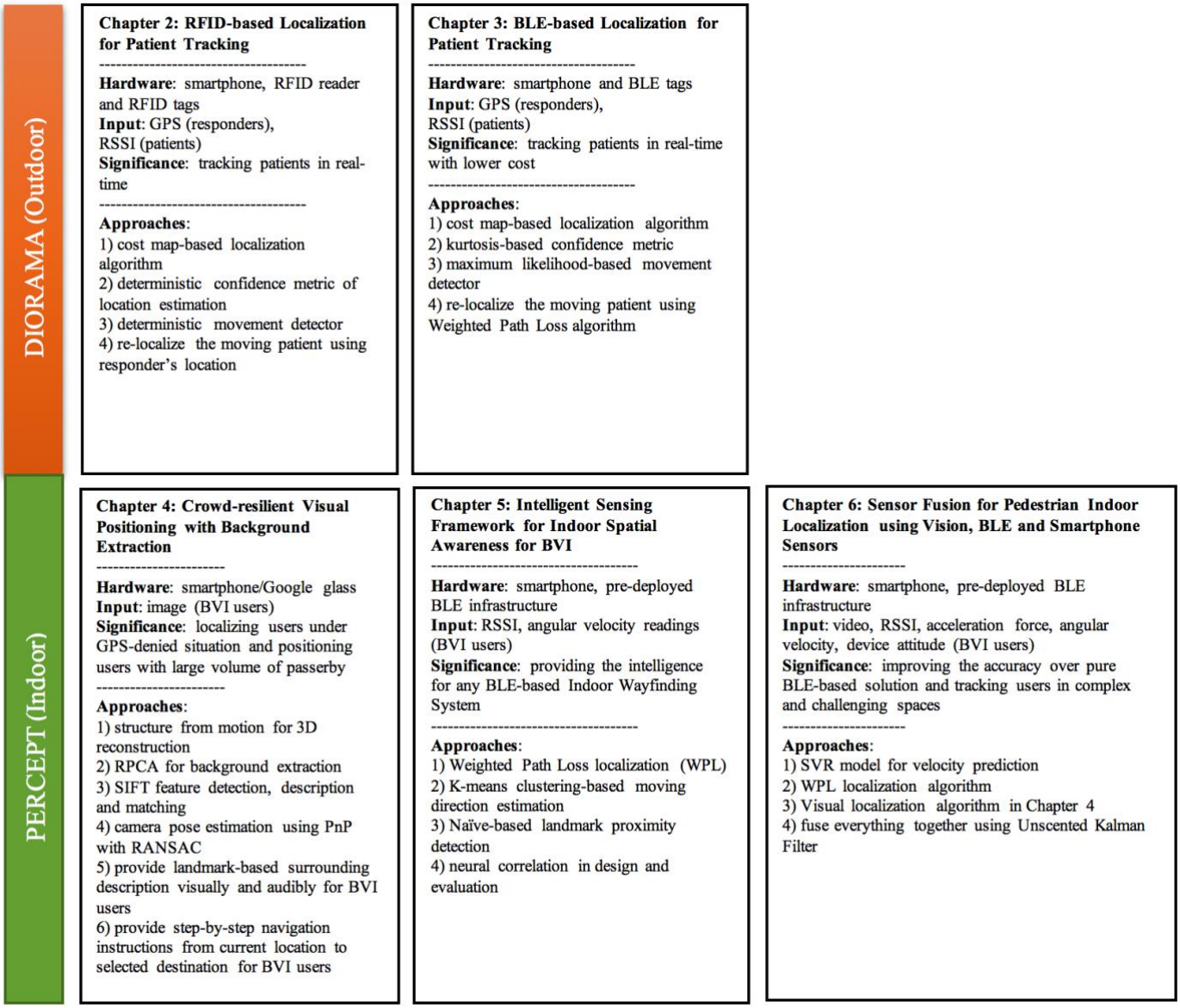


Figure 1.5: Dissertation outline

CHAPTER 2

RFID-BASED LOCALIZATION ALGORITHM FOR PATIENT TRACKING

2.1 Introduction

When a Mass Casualty Incident (MCI) is declared the number of people killed or injured can overwhelm the local EMS resources. EMS responders who join the search and rescue operations will face difficulties to keep track of the victims' location and their corresponding priority levels, the information that can help them prioritize victims evacuation. Moreover, the Incident Commander (IC) who receives information through chaotic radio communication from multiple responders faces challenges to build precise situational awareness of the MCI. Currently, the IC organizes this information and coordinates the responders/teams by either writing down on paper or on a whiteboard. We have developed DIORAMA system to deal with all these challenges and enhance disaster situation awareness.

One of the most useful functionalities provided by disaster response systems is to keep track of responders and victims in the field. While several disaster response systems were developed, only DIORAMA system is able to keep track of the static victims continuously and the moving victims opportunistically without relying on any infrastructure. We assume that each responder will carry a GPS-enabled Smartphone and an active RFID reader, and the victims will be tagged with active RFID tags during the triage process. This active RFID-based algorithm which we will present in detail in this chapter does not require any calibration or pre-deployed infrastructure to reach an accuracy comparable to the GPS system but bringing the cost down significantly. Moreover, it is a

mobile and scalable solution which means that it can be used anywhere anytime as long as the GPS is available.

The responders use the information regarding the location and severity of injury of the victims provided by DIORAMA along with their own geo-location information on their Smartphones to orient and navigate themselves towards the victims. After the victim with the appropriate severity of injury is located, the responder will transport the victim to the treatment area. We have shown that by using DIORAMA system we can reduce the transportation time and decrease the victims mortality significantly.

2.2 System Overview

The on-site rescue operations focus on two phases in disaster response: triage and treatment. Each responder carries a GPS-enabled Android smartphone, an active RFID reader and certain numbers of D-tags. Moreover, each D-tag is composed by an active RFID tag, a passive RFID tag and a traditional paper triage tag. The active RFID devices that are manufactured by RF Code [70] include 433 MHz M160 Wristband tags which are battery-powered RF transmitters and M220 active RFID readers. The detailed hardware specification of RFID tag is listed in Table 2.1.

Table 2.1. Hardware specification of RFID tag

Manufacturer	RF Code
Operation Frequency	433.92 MHz
Transmission Range	up to 300 ft.
Time Interval	2 seconds
Battery Life	> 4 years (normal)
Unit Cost	$ 25

In triage phase, the responders evaluate the patients' severity of injury using Simple Triage and Rapid Treatment (START) system [71], which is a quick triage strategy for

patient assessment exploited in MCI by responders. The responders will tag each patient based on their symptoms using the D-tag indicating the specific severity of injury. From this moment, the localization engine will begin to keep track of the location and movement information of patients using the Radio Signal Strength Indicator (RSSI) measurements collected by the active RFID readers carried by the responders. In addition, the smartphones can provide the anchors' real-time locations with GPS coordinates. The data flow between the patients, responders, incident commander (IC) and localization engine are depicted in Figure 2.1.

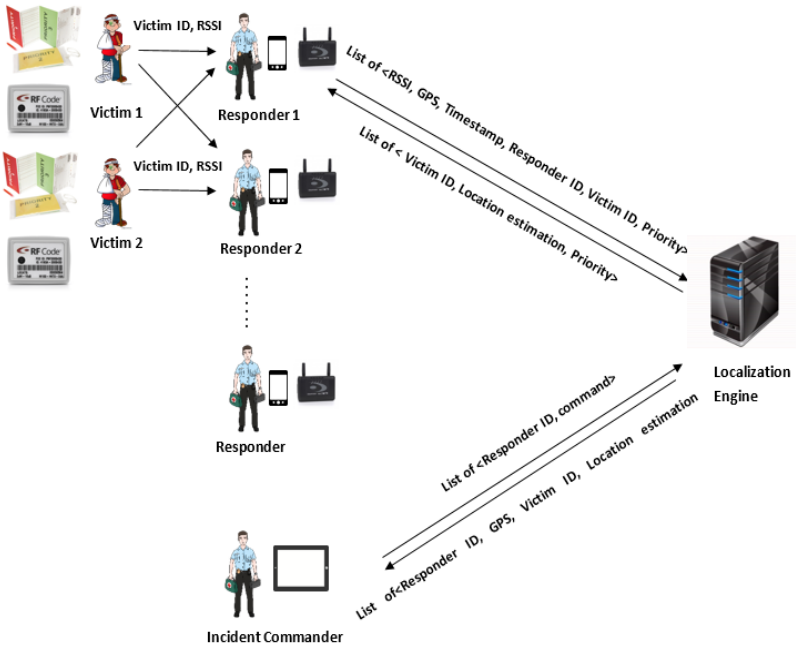


Figure 2.1: Data flow of localization in DIORAMA

During the treatment phase, the responders are searching for the triaged patients and moving them to the designated treatment area in the disaster scene ceaselessly. Meanwhile, the RSSI will be collected while this process. With more and more RSSI measurements, the localization result can be calculated and improved for the stationary patients. Furthermore, regarding the moving patients, which are the minority of patient population, their movements are very likely to be detected by the localization engine and

they can be re-localized if they are staying in the range of the moving responders. Still, there is a chance that the moving patients cannot be tracked since our approach is opportunistic. By applying the confidence metric of the location estimation to evaluate the goodness of the result, we enable to update patients' locations to responders and IC only when there is a potentially better estimation coming up. Therefore, not only the responders and IC are receiving reliable results of patients' locations from the localization engine, but also the workload of communication is reduced in DIORAMA.

2.3 Methodology

The proposed algorithm is comprised of three important components, recursive tracking of patients, confidence metric for evaluating the location estimation and opportunistic re-localization of moving patients.

2.3.1 Recursive Tracking

First, we determine the relationship between RSSI and the distance between the sender and receiver. We collected a training dataset to build a signal strength attenuation model for our localization algorithm by using curve fitting techniques and obtained the following equation:

$$rssi = 0.71 \times log(distance)^2 + 8.49 \times log(distance) + 39.3 \qquad (2.1)$$

where $rssi$ is defined as the expected RSSI in dB measured at $distance$ in feet between the sender and receiver.

Second, we consider the limited movement ability of patients and several other technical issues, such as computation and storage cost. The localization algorithm keeps track of the patients in a 100 ft^2 area centered at the initial triaged location provided by GPS.

Then the engine uses a 100 by 100 grids mapping to this 100 ft^2 geographical area as shown in Figure 2.2.

Based on the attenuation model in (2.1), a cost function defined in (2.2) is used to estimate the location of the patients on the grid-based cost map. Since each grid on the map is mapping to a known GPS coordinate, the distance between the real-time GPS coordinates from responders and the GPS coordinates on each grid is easily computed. So the expected RSSI on each grid can be calculated using (2.1). Thus, the cost value can be computed by (2.2) recursively on the fly. An example of the cost function is depicted in Figure 2.3.

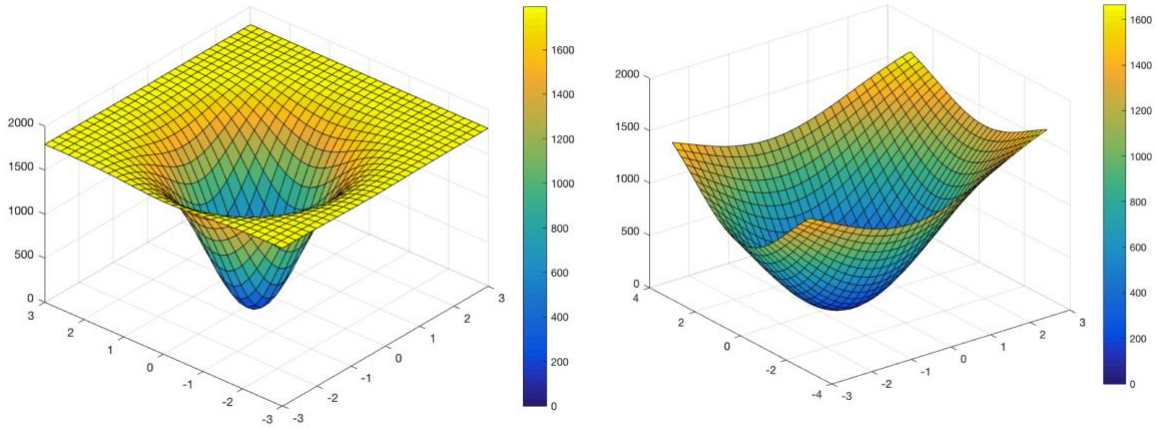
$$Cost_{(x,y,t)} = \sum_{i=1}^{i=t} |S_{expected(x,y,t_i)} - S_{measured(t_i)}|^2 \tag{2.2}$$

where $Cost_{(x,y,t)}$ is defined as the accumulated cost value at *(x, y)* up to time *t*, and $S_{expected(x,y,t_i)}$ is the RSSI expected to be measured at responders' current GPS coordinates from *(x, y)* on the map at time t_i. $S_{measured(t_i)}$ is the measured RSSI at time t_i.

Third, the localization algorithm estimates the location from the cost map by returning the GPS coordinate on the grid with the minimum cost value as shown in (2.3). By combining it with the confidence metric which will be discussed later, the reliable location estimation can be delivered to the responders and IC successfully. $[\hat{x}, \hat{y}]_t$, which is the location estimation at time *t*, is given by:

$$[\hat{x}, \hat{y}]_t = \underset{(x,y,t)}{argmin}\, Cost(x, y, t) \tag{2.3}$$

Figure 2.2: An example of grid-based geographical map



(a) (b)

Figure 2.3: Illustration of different cost maps with associated confidence levels of estimation (a) higher confidence (b) lower confidence

2.3.2 The Confidence Metric of Location Estimation

Since the location estimation is computed from the cost map, the confidence metric is also achieved using this map. The shape of the cost map determines the quality of the location estimation. The sharper the cost map, the higher the estimation confidence. Figure 2.3 illustrates two cases: Figure 2.3b illustrates a cost map with lower confidence level than the cost map depicted in Figure 2.3a.

2.3.3 Track the Moving Patients Opportunistically

In the disaster scenario, it is possible that there are a small number of patients who will move intermittently. The localization engine should be able to deal with such case by detecting the movement of the patients based on the RSSI measured by the responders and re-localize the patients opportunistically if they are in range.

2.3.3.1 Detection of the Movement

To introduce movement detection into the proposed localization algorithm, we take many factors into account, such as GPS accuracy level, estimation accuracy level and RSSI reliability. The engine will detect that patients are moving from the known locations to somewhere new when a certain number of reliable RSSI readings are collected consistently at places where such readings should not be measured. Figure 2.4 illustrates the principle of movement detection. If the area of the current location estimation does not intersect with the area centered at responder's current position with uncertainty from both GPS and signal noise, this RSSI reading will be treated as an indicator that the patient has moved.

2.3.3.2 Re-Localize the Moving Patients

After the movement is detected, the engine starts to predict the new location of the patient. This re-localization functionality is built based on an opportunistic approach because it is necessary that the responders pass by the new location of the patient and collect RSSI around that area.

In a sliding window of T seconds, if there are more than a certain number of RSSI readings that are less than 85 dB (about 40 ft.) the engine will start to re-localize the moving patient. By using this mechanism, it is highly likely that the moving patient is nearby since

the readings in short range (less than 40 ft.) are reliable and noise-resistant. Moreover, we can control the sensitivity and reliability of this functionality by adjusting the threshold of the number of RSSI readings at short range.

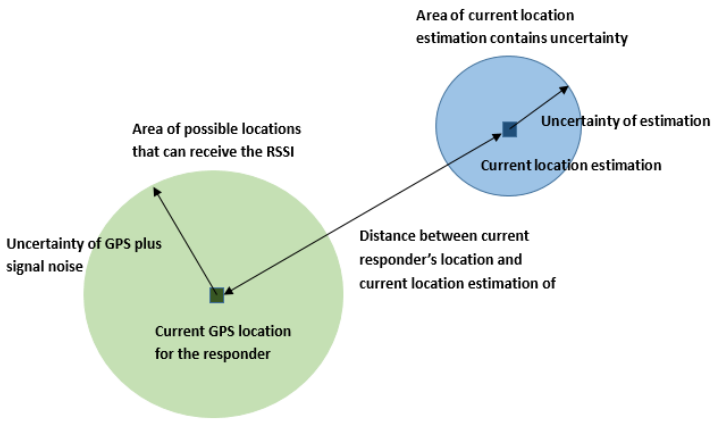


Figure 2.4: Illustration of movement detection

2.4 Performance Evaluation

In the experiments, we investigate the localization engine performance in terms of localization accuracy for the following cases:

1. the relationship between the localization accuracy and the number of tags per patient
2. the relationship between the localization accuracy and the number of responders
3. the localization accuracy for moving patients.

Number of tags per patient: in this experiment, we have 11 static patients represented by cones randomly placed in a 100×100 ft. area. There is only one responder that are constantly moving at random in the area. The responder carries an active RFID reader and a Smartphone. The responder tags each cone/patient with a number of active RFID tags. From Figure 2.5, we observe that increase in number of tags does not improve the localization accuracy. This is because we have enough diversity of readings from one tag since the active

RFID reader is moving. Therefore, we decide to tag each patient with one RFID tag, which is a more affordable solution.

Number of responders: we investigate the relationship between the localization accuracy and the number of responders. From Figure 2.6, as expected, we observe that the localization accuracy improves as the number of responders increases.

Moving patients: we conduct a number of trials as follows.

- Trial 1: in this experiment, we have 25 patients in total, which include 3 moving patients and 22 static patients represented by cones randomly placed in a 260×300 ft. area. There are two responders who are constantly moving by performing triage and treatment.
- Trial 2: in this experiment, we have 3 moving patients in a 260×300 ft. area. There are two responders who are moving constantly at random.
- Trial 3: in this experiment, we have 4 patients in total, 3 moving patients and one 1 static patient in a 260×300 ft. area. There are two responders who are moving constantly at random.
- Trial 4: in this experiment, we have 6 patients in total, 5 moving patients and 1 static patient in a 320×420 ft. area. Two responders are constantly moving at random.
- Trial 5: In this experiment, we have 25 patients across an area as big as 309, 000 ft^2. Two moving patients are involved. And two responders are performing primary triage and evacuation.
- Trial 6: In this experiment, we have 25 patients across an area as big as 465,000 ft^2. Two moving patients are involved. And two responders are performing primary triage and evacuation.

In this section, we would like to examine the localization engine for moving patients in terms of false positives (patients flagged as moving while stationary) and the re-localization accuracy. We notice that in none of the experiments we obtained false positives, i.e. no static patient is detected as moving from the original location.

Figure 2.7 displays the relationship between the re-localization accuracy and the number of RSSI readings below 40 ft. (85dB) radius zone. This is due to the fact that the RSSI readings in this zone are highly reliable. We observe that as the number of readings in this zone increases the re-localization accuracy increases. Of course, this is an opportunistic localization algorithm and the responders may not get in this zone (within 40 feet from the patient) resulting in worse localization accuracy. However, we observe that it is enough to have a few readings (at least 7 readings) in this zone to obtain GPS like localization accuracy.

In summary, from the experiments conducted we conclude that for both static and moving patients we can obtain a GPS like localization accuracy. Note that since this is an opportunistic algorithm using mobile anchors the re-localization accuracy can be worse than GPS like accuracy if the responders are not in the vicinity of the patients' new locations.

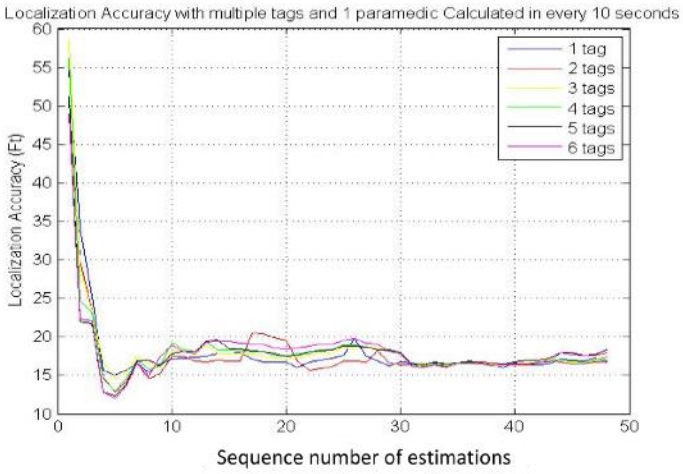


Figure 2.5: Localization accuracy with different number of tags and one responder in the field

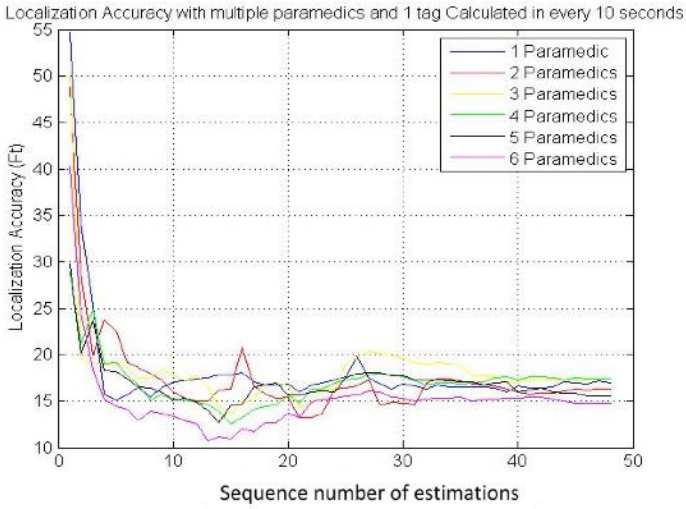


Figure 2.6: Localization accuracy with different number of responders in the field and one tag on each patient

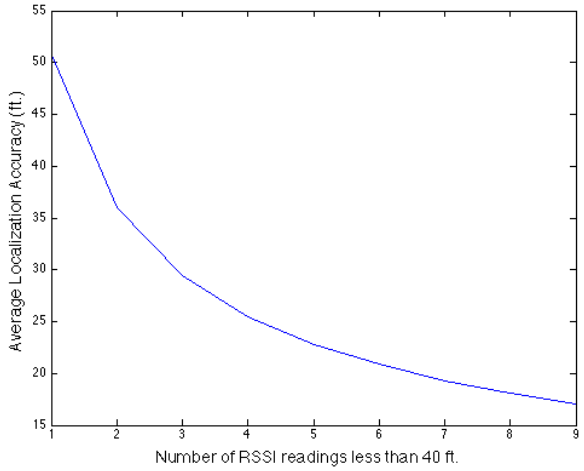


Figure 2.7: Relationship between re-localization accuracy and average number of RSSI measured in 40 feet radius

2.5 Conclusion

We introduced an active RFID based mobile localization algorithm used in DIORAMA system. The localization algorithm obtains reliable localization accuracy commensurate with GPS (under 20 feet localization accuracy). For moving victims, the algorithm is opportunistic since only victims that are within the range of responders will be re-localized.

CHAPTER 3

BLE-BASED LOCALIZATION ALGORITHM FOR PATIENT TRACKING

3.1 Introduction

Natural disasters (e.g. floods, earthquakes) and man-made disasters (e.g. terrorist acts) cause a significant number of victims and overwhelm emergency resources. To maximize the effectiveness of emergency resources, a number of disaster management systems were developed.

DIORAMA system is a real-time, scalable and infrastructure-less disaster management system. DIORAMA provides visualization and collaboration tools for responders and incident commander. As shown in the previous experiments, DIORAMA can reduce the victims' evacuation time by up to 43%.

Using active RFID technology, DIORAMA provides localization of victims. This technology requires proprietary equipment, such as active RFID readers and tags from RFCode. While the localization results obtained are excellent, we seek to develop a localization system with non-proprietary and inexpensive technology. In this chapter we introduce a victims' localization algorithm that uses Bluetooth Low Energy (BLE) technology which enables us to use non-proprietary BLE tags and a Smartphone that supports Bluetooth 4.0 as a reader.

The localization engine should meet the following requirements. First, the engine should provide accurate victims' location after they are triaged. Second, the localization engine needs to detect the movement of victims and localize them. To meet the first requirement, we utilize a cost map-based localization algorithm along with a Kurtosis-

based confidence metric. For the second requirement, a Maximum Likelihood (ML)-based movement detector is adopted to detect the movement of a victim.

3.2 System Overview

Similar to RFID-based system, the data flow of BLE-based system is illustrated in Figure 3.1. Bluetooth Low Energy (BLE) tags, also known as iBeacons, are manufactured by a number of companies. In this study, we chose Tough beacon manufactured by Kontact.io [72]. Tough beacon's specification is given in Table 3.1.

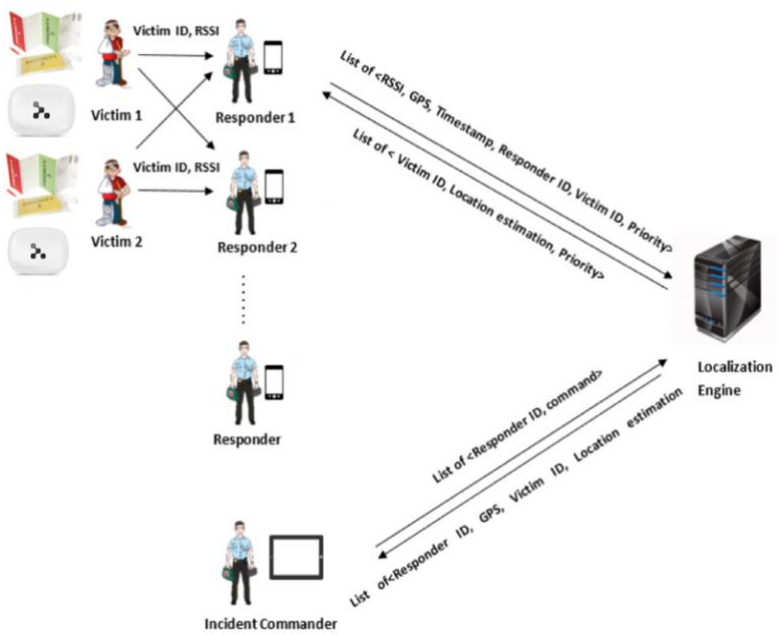


Figure 3.1: Data flow of localization in DIORAMA

Table 3.1. Hardware specifications of BLE tag

Manufacturer	Kontact.io
Operating Frequency	2.4 GHz
Transmission Range	Up to 100 ft.
Time interval	350ms
Battery Life	Up to 2 years with default setting
Unit cost	$27

To determine the tags propagation characteristics, we conducted a number of experiments in an open area using Samsung Galaxy S6 as the receiver, which collects the RSSI from BLE tags. 200 RSSIs are collected at every 3 ft. from 3 ft. to 66 ft. Note that all measurements are collected under line-of-sight (LOS) situations. The transmission power of each BLE tag is set as the default value, which is -12 dB.

We employ the log-distance path loss model:

$$PL(d) = PL_0 + 10 * \gamma * log_{10} \frac{d}{d_0} + X_g \tag{3.1}$$

where PL is the total path loss measured in Decibel (dB), PL_0 is the path loss at the reference distance d_0, γ is the path loss exponent and X_g is a normal random variable with zero mean.

The propagation model of the Tough beacon is given in equation (3.2) and is plotted in Figure 3.2.

$$PL(d) = -76.278 + 10 * -1.511 * log_{10} \frac{d}{3} \tag{3.2}$$

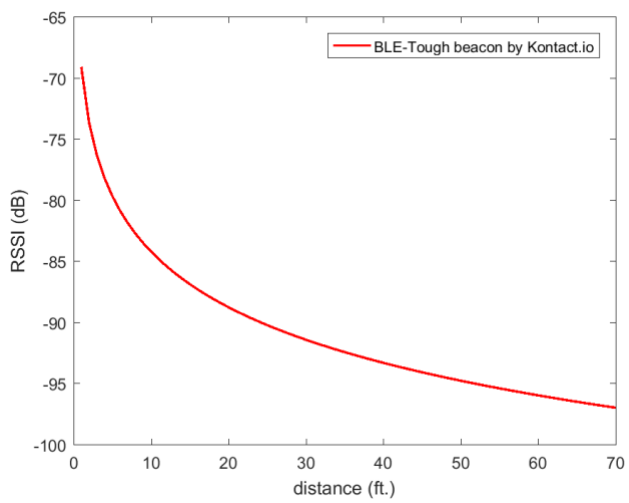


Figure 3.2: BLE tag propagation model

3.3 Methodology

The localization algorithm includes the following four parts:

- localize stationary patients using a cost map-based localization algorithm
- evaluate location estimation using a kurtosis-based confidence metric
- detect movement of patients using a maximum likelihood-based movement detector
- re-localizing moving patients using weighted path loss algorithm

Details of each part are provided below.

3.3.1 Cost Map-based Localization Algorithm

The purpose of employing cost map-based localization algorithm is to compensate for the inaccuracy of GPS as well as the noise in RSSI readings. The cost map-based localization algorithm is built based on a quadratic cost function defined in (3.3).

$$Cost(x,y)_{patient_j,T} = \sum_{i=1}^{i=T} |ExpRSSI(x,y,responder_k)_{t_i} - RSSI_{responder_k,t_i}|^2 \quad (3.3)$$

where $Cost(x,y)_{patient_j,T}$ is defined as the accumulated cost value of vertex (x,y) on sample grid of $patient_j$ up to time T, and $ExpRSSI(x,y,responder_k)_{t_i}$ is the expected RSSI computed from $responder_k$'s current GPS coordinates and the geolocation of vertex *(x, y)* at time t_i. $RSSI_{responder_k,t_i}$ is the measured RSSI from $responder_k$ at time t_i.

As shown in Figure 2.2, for each patient, we have a 100 x100 sample grid (each grid represents a 1 ft. x 1 ft. area) centered at the initial triaged geolocation. Since each vertex maps to a certain geolocation, the location estimation of patient_j at time T, $(\hat{x},\hat{y})_{patient_j,T}$, is given by equation (3.4).

$$(\hat{x},\hat{y})_{patient_j,T} = \underset{(x,y)}{argmin}\, Cost(x,y)_{patient_j,T} \quad (3.4)$$

3.3.2 Kurtosis-based Confidence Metric

It is necessary to find a proper metric to evaluate the goodness of the cost map-based localization estimation. It is known that the sharper the cost map, the higher the confidence of the localization estimation. Therefore, we will use kurtosis to design the metric, which is a measure of sharpness or peakedness of the probability distribution. Similar to [73], we compute the confidence value of the current location estimation of $patient_j$ at time T, $Conf(\hat{x},\hat{y})_{patient_j,T}$ as follows:

$$Conf(\hat{x},\hat{y})_{patient_j,T} = -3 + \frac{1}{\sigma^4}\sum_x\sum_y[(x-\hat{x})^2 * (y-\hat{y})^2] * Cost(x,y)_{patient_j,T} \quad (3.5)$$

where σ is the standard deviation of distances from the geolocation of each vertex on the sample grid to the location estimation.

3.3.3 Maximum likelihood-based Movement Detector

We introduce a maximum likelihood (ML) based movement detector. Using the propagation model, we generate the probability density function of the RSSIs measured at each sample distance. We compute the likelihood of receiving RSSI at a specific responder's location. By comparing the likelihood of receiving RSSI at a current location estimation with the likelihood of receiving RSSI at another unknown location, the ML-based movement detector will determine whether the patient moves from the previous location or not.

3.3.4 Weighted Path Loss Algorithm

After the moving patient is detected, the localization framework starts to track it by applying weighted path loss algorithm. Basically, the weighted path loss localization algorithm [74] computes the location estimation in (3.6) by using a weighted average of the responders' locations and the corresponding RSSI values collected at each location.

Given N responders' information, the location estimation of the moving patient, $(\hat{x}, \hat{y})$ is given by:

$$(\hat{x}, \hat{y}) = \sum_{i=1}^{n} w_i * (x_i, y_i) \tag{3.6}$$

where (x_i, y_i) denotes the location of $responder_i$ and w_i is the weighting factor given by:

$$w_i = \frac{1/d_i}{\sum_{i=1}^{n} 1/d_i} \tag{3.7}$$

where d_i denotes the distance between $responder_i$ and the moving patient, which is computed using equation (3.2) with the measured $RSSI_i$.

3.4 Performance Evaluation

We conducted the following experiments with three responders carrying Samsung Galaxy S6 Smartphones which run the localization application:

- Three experiments with 17 stationary victims and 4 moving victims generating datasets 1, 2 and 3.
- Two experiments with 15 stationary victims and 3 moving victims generating datasets 4 and 5.

In all experiments, stationary patients were represented by cones at the locations shown in Figure 3.3. Each experiment lasted 20 minutes.

In order to evaluate the performance of the proposed localization algorithm, we use the mean value, the rooted mean square error (RMSE) in (3.8) and the cumulative

distribution function (CDF) to measure the localization accuracy of stationary patients. Since the re-localization algorithm is opportunistic (patients can be re-localized only if there are responders nearby), for moving patients we evaluate the movement detector performance rather than the re-localization accuracy. To evaluate the movement detector performance, we use metrics used to measure the performance of binary classification tests, such as Recall in (3.9), False Positive Rate (FPR) in (3.10), and movement detector accuracy (MACC) in (3.11). P is the number of positives (i.e., the patients move), N is the number of negatives (i.e., the patients did not move), TP is the total number of estimated true positives, TN is the total number of estimated true negatives, FN is the total number of estimated false negatives, FP the total number of estimated false positives.

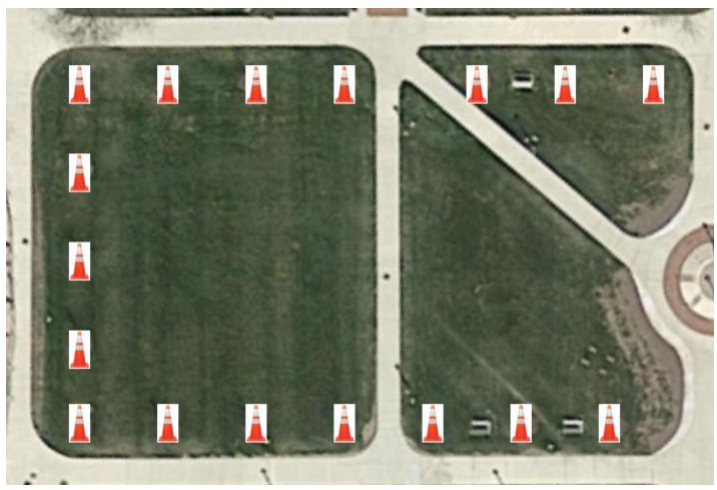


Figure 3.3: Experiment setup

$$RMSE = \left(\frac{1}{n}\sum_{i=1}^{n}(x - \hat{x})^2\right)^{1/2} \tag{3.8}$$

$$Recall = \frac{TP}{TP+FN} \tag{3.9}$$

$$FPR = \frac{FP}{FP+TN} \tag{3.10}$$

$$MACC = \frac{TP+TN}{P+N} \tag{3.11}$$

Figure 3.4 depicts a box plot of the localization accuracy across all experiments. We observe that the mean localization accuracy of 81 stationary patients is about 11.08 ft. and the RMSE is 13.88 ft. As shown in Figure 3.5, over 97% of the stationary patients are localized within 25 ft., which is adequate for our system. The recall (Recall) of the detector is 33%, the false positive rate (FPR) is as low as 1.2% and overall movement accuracy of the detector (MACC) is 87%.

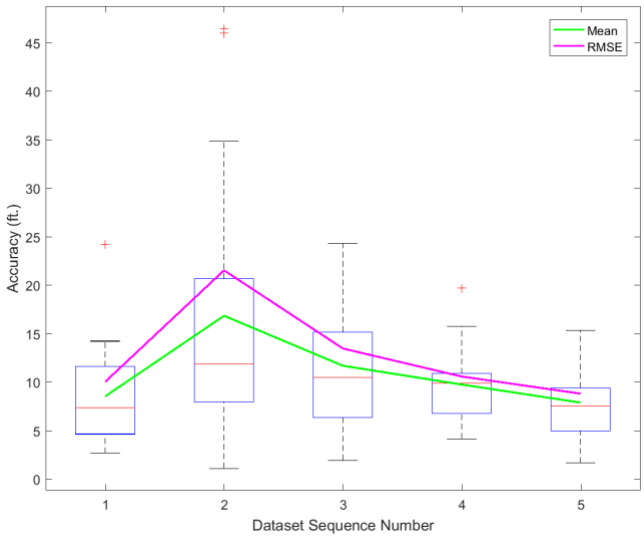


Figure 3.4: Localization accuracy of stationary patients

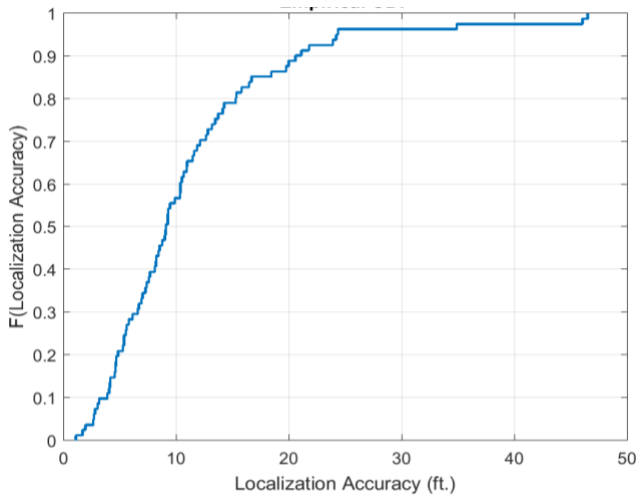


Figure 3.5: CDF of localization accuracy for stationary patients

3.5 Conclusion

We introduced the first outdoor localization algorithm for a disaster management system that uses BLE technology. The use of BLE technology allows us to use commercial off-the-shelf BLE tags as well as use a Smartphone as a receiver, significantly reducing the system cost and maintenance. The localization algorithm achieves a mean localization accuracy of 11 ft. For moving victims, the proposed maximum likelihood-based movement detector keeps a low false positive rate of 1.2%.

CHAPTER 4

CROWD-RESILIENT VISUAL POSITIONING WITH BACKGROUND EXTRACTION

4.1 Introduction

According to visual impairment and blindness statistics from the World Health Organization (WHO), there are 285 million people suffering from visual impairment worldwide [93]. Indoor wayfinding in complex public spaces poses a major challenge to blind and visually impaired (BVI) individuals and negatively affects their mobility and the quality of life. To increase the BVI individuals' ability of travelling independently, we developed the PERCEPT indoor navigation system [37,38] using Near-Field Communication (NFC) tags; it was proved to be beneficial to the BVI users. From the experiments conducted with BVI subjects, PERCEPT has shown significant effectiveness on indoor wayfinding by delivering step-by-step audible navigation instructions to users. Although the PERCEPT system provides reliable localization and orientation to users by scanning the tags, the deployment of NFC tags requires changes in the environment, which can be costly.

To make PERCEPT system scalable and cost-effective, we propose to develop an organic computer vision-driven smartphone-based indoor navigation system, which we name PERCEPT-V. For this system, we show that the visual localization algorithm [51] can determine the BVI user's location and orientation in real-time using image or video captured by commercial devices, such as smartphones or wearable cameras. Moreover, the accuracy of the location and orientation estimates is sufficient for BVI users to navigate themselves safely in the space. While existing visual localization algorithms provide sufficiently accurate estimation of location and orientation, we find that there is a new

technical challenge to PERCEPT-V: we must increase the reliability of the visual localization algorithm when crowds are present in the observed environment, and our algorithm must be resilient to instability in the framing and view of the images acquired by the BVI users.

In order to address the two aforementioned issues, we propose to integrate a background extraction algorithm into the image processing pipeline to improve the resilience of visual localization to the presence of crowds in the observed scene. In contrast to most existing background subtraction algorithms, which simply identify a mask that identifies and extracts the foreground in the image, we focus on the use of background subtraction algorithms that create a model for all pixels of the background. We refer to such algorithms as background extraction algorithms. Removing the foreground helps prevent spurious matches between features corresponding to crowds (foreground) and the reference navigation space (background).

4.2 System Overview

We implemented the vision-based indoor wayfinding system for BVI users, which we call PERCEPT-V, using a client-server architecture. The client can be either Google Glass or smartphone used by BVI users. For the Google Glass application, we implemented user interaction with vocal commands, such as “localize myself”. For the Smartphone application, we implemented a vision free interface, which allows the user to choose the building they enter and deliver the guidance instructions via an audible interface. At the server side, we implemented the positioning algorithm, path planning algorithm along with the guidance instructions generation algorithm.

The system includes an offline phase and an online phase as shown in Figure 4.1. The offline phase is responsible for preparing four kinds of necessary information, which contains: the SIFT descriptors for the point cloud, the locations of the point cloud on the image plane, 3D locations of the point cloud used to represent the space and the manually labeled locations and descriptions of landmarks which will be used to generate the guidance instructions.

In the online phase, we first compute the user location and orientation using the image captured by Google Glass or smartphone and other spatial information prepared in the offline phase. The user location and orientation along with the locations of the landmarks defined in the offline phase will serve as input to the guidance instruction generation module. For each landmark in the surrounding of the user, the guidance instruction module will generate information that includes the landmark description along with the relative distance from the user's location to the specific landmark and the relative orientation from user's direction to the specific landmark using clock phase.

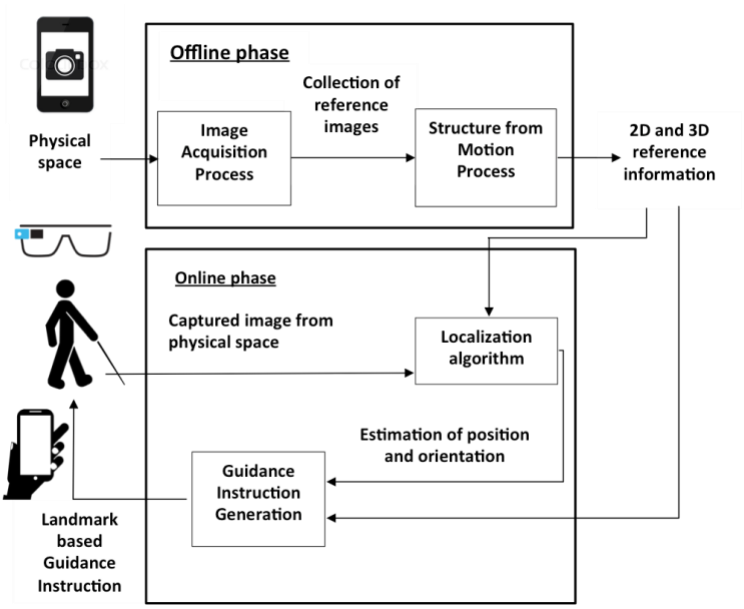


Figure 4.1: Offline phase and online phase of PERCEPT-V

4.3 Methodology

4.3.1 Offline Phase

In this phase, we collect unordered images captured in the physical space using a smartphone camera. Using this image collection, we adopt Structure from Motion (SFM) technique by using BundlerSFM [78] or VisualSFM [79] to produce the 3D reconstruction of sparse scene geometry as output. The output is parsed by our Matlab script to extract reference information including 2D locations of the point cloud on the image plane, SIFT descriptors of the point cloud and 3D locations of the point cloud. We use OpenCV coordinate system instead of OpenGL coordinate system since OpenCV is a powerful library for computer vision algorithms.

In addition, since our positioning algorithm in the online phase uses a calibrated camera model, we calculate the intrinsic parameters of the user's device using Zhang's camera calibration algorithm [80] and Matlab Toolbox [81]. The intrinsic parameters matrix *KK* is defined in equation (4.1).

$$KK = \begin{bmatrix} f_c(1) & \alpha_c * f_c(1) & cc(1) \\ 0 & f_c(2) & cc(2) \\ 0 & 0 & 1 \end{bmatrix} \tag{4.1}$$

where f_c is a 2×1 vector for the focal length in pixels, *cc* is a 2×1 vector for principle point, α_c is the skew coefficient defining the angle between x and y pixel axes.

4.3.2 Online Phase

As shown in Figure 4.2, in the standard approach (without the shaded block), after keypoint descriptors are extracted from the acquired and reference images, a search finds the best match between the descriptors among the reference images to those from the acquired image. Consequently, a registration module calculates the most likely geometric

transformation between the images, providing an estimate of the location of the camera given the location of the registered image. The workflow is illustrated in Figure 4.3. And we describe each step of the algorithm in following sections.

Our proposed architecture adds the one shaded block: a background extraction scheme to remove activity from passerby before keypoint descriptors are obtained. Through extracting the background, the proposed localization algorithm reduces the likelihood of mismatches from features for the foreground to the reference images and increases the likelihood of recovering more useful features about background that can match with the reference images and benefit the localization performance simultaneously.

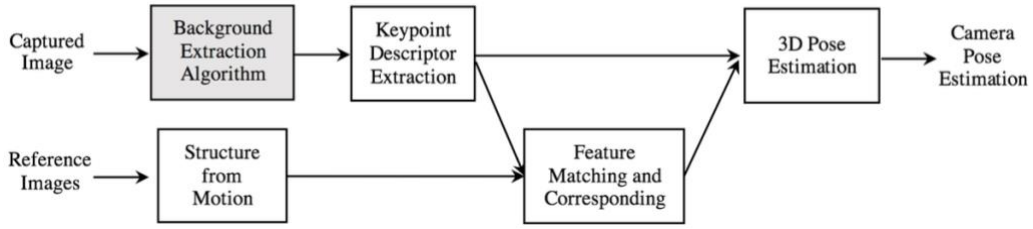


Figure 4.2: Flow chart of data processing pipeline for PERCEPT-V

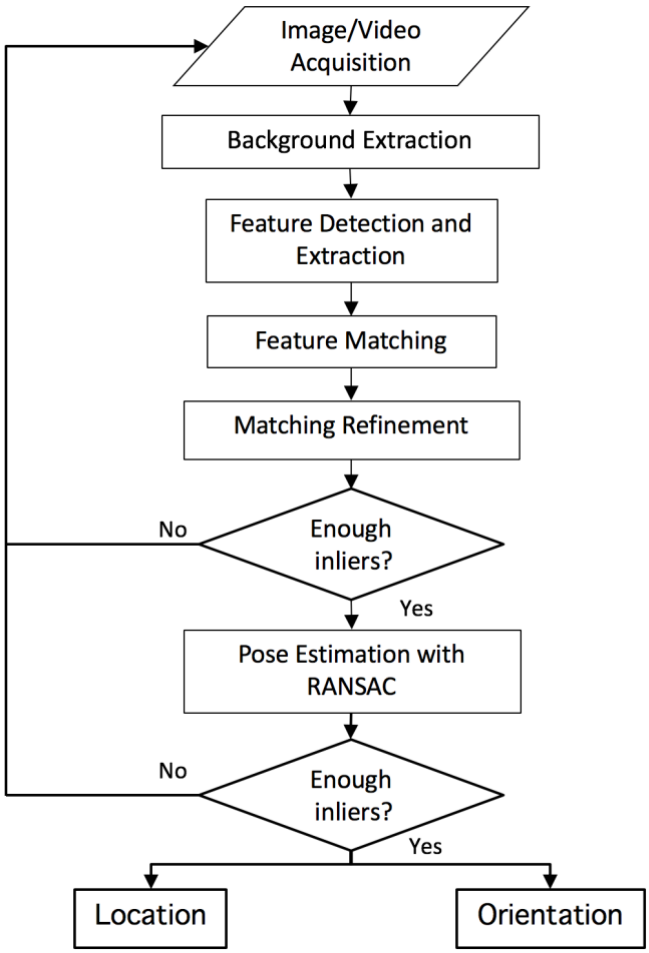


Figure 4.3: Algorithm workflow in online phase

4.3.2.1 Feature Detection and Extraction

To account for different indoor environments, the feature extraction algorithm needs to be invariant to scale, rotation, illumination, blur and motion blur. Following the comparison studies published in [82, 83], we decided to use SIFT algorithm [84] since it is the best candidate for our environmental conditions described above.

4.3.2.2 Feature Matching

In computer vision algorithms, usually the most computationally expensive component is to find the nearest neighbor matches in high dimensional data that represents the training data set. In our case, SIFT algorithm will generate a 128-dimensional vector

for each keypoint. To solve this problem, Muja and Lowe [85] proposed an approximate nearest neighbor library which enables us to find the optimal nearest neighbor algorithm and its parameters. We use the fast library for approximate nearest neighbors (FLANN) and randomized K-D tree. In our system, we need to find the nearest two neighbors.

4.2.2.3 Voting Tally

In order to remove the outliers from the matching pairs, the locations of the matched pairs of SIFT descriptors on the image plane will be used to calculate the fundamental matrix. To find these locations we apply a voting process to obtain the reference image which contains the highest number of matched reference SIFT features.

4.3.2.4 Outliers Removal

4.3.2.4.1 Distance Ratio

Before the voting process, Lowe [84] suggests finding reliable matches by setting a maximum threshold of 0.8 on the ratio between the distances of the closest neighbor to that of the second-closest neighbor depicted in equation (4.2).

$$\begin{cases} inlier & \frac{\|descriptor_i - descriptor_{the\ closest\ neighbour}\|}{\|descriptor_i - descriptor_{the\ second\ closest\ neighbour}\|} < 0.8 \\ outlier & otherwise \end{cases} \tag{4.2}$$

4.3.2.4.2 Fundamental Matrix

After the voting process, we use Random Sample Consensus (RANSAC) [86] to estimate the fundamental matrix from the matched pairs of descriptors which pass the distance ratio test but still contains outliers. Fundamental matrix F is defined in epipolar geometry as a unique 3×3 rank 2 homogeneous matrix. All pairs of corresponding points hold $x'^T F x = 0$, where x and x' are the corresponding points in a stereo image pair. Only the inliers used to estimate the fundamental matrix proceed to pose estimation.

4.3.2.5 Pose Estimation

Given the 3D-to-2D point correspondences, the estimation of the pose of a calibrated camera used by the users is known as Projective-n-point (PnP) problem. PnP obtains the extrinsic parameters of the camera which describe the user's location in the world coordinates and the user orientation. The 3x3 rotation matrix, R and the 3x1 translation vector t are obtained from equation (4.3).

$$\begin{bmatrix} u \\ v \\ 1 \end{bmatrix} = KK \cdot [R \quad t] \cdot \begin{bmatrix} X \\ Y \\ Z \\ 1 \end{bmatrix} \tag{4.3}$$

where *(u, v)* are the locations of the projection point in pixels on image plane, *(X, Y, Z)* are the locations of 3D point cloud in world coordinate space, *KK* is the matrix that represents the intrinsic parameters of the camera defined in (4.1).

Given R and t, the location of the camera P is calculated by equation (4.4) and the orientation of the camera O which represents the direction that the camera is facing is calculated by equation (4.5).

$$P = -R^T \cdot t \tag{4.4}$$

$$O = R^T \cdot [0 \quad 0 \quad 1]^T \tag{4.5}$$

where P and O are 3×1 column vectors.

To make the pose estimation resistant to outliers, we adopt RANSAC. We compare the performance between the two algorithms in this area, one of them is an iterative method based on Levenberg-Marquardt optimization to find the pose that minimizes the re-projection error, another is a non-iterative solution called EPnP developed by Lepetit [87] that can reduce the computation time from $O(n^5)$ to $O(n)$.

4.3.2.6 Background Extraction via RPCA

Let the acquired video be comprised of n frames of size m = width × height denoting the total amount of pixels on each frame. We store this video in a matrix $D \in R^{m \times n}$; each column of the matrix D corresponds to a video frame and each row represents the evolution of a specific pixel over the acquisition time. In RPCA, we consider the following decomposition for the video matrix [88]:

$$D = A + E \tag{4.6}$$

where A is a low-rank matrix corresponding to the background and E is a sparse matrix corresponding to the activity of passerby. This decomposition is motivated by the small number of degrees of freedom for the background and the localized and highly concentrated passerby activity.

The exact recovery of the low-rank matrix A of interest from the sum D can be solved by the following convex optimization problem:

$$\arg\min_{A,E} \| A \|_* + \lambda \|E\|_1, \quad subject\ to\ D = A + E, \tag{4.7}$$

where $\|\cdot\|_*$ denotes the nuclear norm of a matrix, $\|\cdot\|_1$ represents the L_1 norm of a matrix, and λ is a positive weighting parameter.

We apply the accelerated proximal gradient method to solve the optimization problem in (4.7), as described in [89].

4.4 Performance Evaluation

4.4.1 Algorithm Performance without Background Extraction

We evaluate the performance of the proposed algorithm using the following metrics: accuracy of location, accuracy of orientation, and computation time.

As mentioned, the location and orientation results will serve as input to the guidance instructions generation module. This guidance module will provide the distance and orientation to specific landmarks surrounding the user. Using this information, the user will discover the landmarks using his white cane.

Figure 4.4 depicts a blind user of height 5.4 feet [90] carrying a white cane. Using the cane, the user can detect objects in front of him in a radius of 5.65 feet. Thus, a location accuracy of less than 5 feet will suffice. The user orientation is conveyed in clock face directions, i.e. the coffee shop is at your 5 o'clock, in increments of hours, i.e., 30 degrees. Therefore, the orientation accuracy obtained by our algorithm can deviate by plus or minus 15 degrees from the exact orientation without impacting the user orientation.

The algorithm will generate two possible outputs:

- provide the location and orientation of surrounding landmarks
- unable to estimate (will prompt the user to retake the picture)

We define successful positioning estimation when the algorithm estimates the user location within 5 feet and the orientation angle is less than 15 degrees.

The reliability of positioning estimation is the likelihood of providing correct positioning estimation which is defined as the summation of the likelihood of successful positioning estimation and the likelihood of not being able to position the user from the query image.

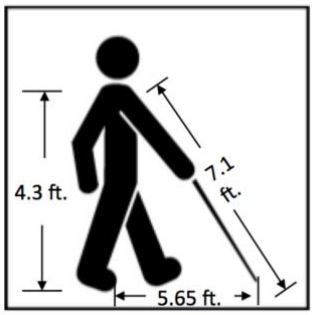


Figure 4.4: Illustration of BVI users with white cane

We collect the following datasets that correspond to a large open indoor environment, UMass Campus Center (over 12,000 ft^2):

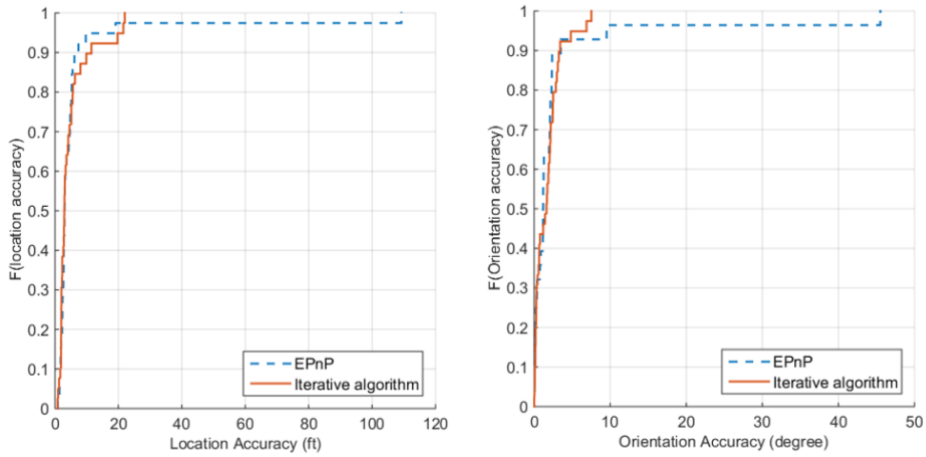
- Dataset that includes 67 test images.
- Dataset that includes 393 images taken in the off-line phase and were used to build the 3D space using SFM software. A sparse point cloud that contains over 20,000 data points is used to represent the reconstructed space as reference information.

In subsection 4.4.1.1, we study the performance of the proposed positioning algorithm using two pose estimation algorithms. In addition to the original datasets we will also test our algorithm performance for distorted images.

- different levels of blurriness (subsection 4.4.1.2)
- different levels of motion blurriness (subsection 4.4.1.3)
- different intensity of illumination (subsection 4.4.1.4)
- different levels of the crowd obstruction. (subsection 4.4.1.5)

4.4.1.1 Baseline Datasets – Comparison between Two Pose Estimation Algorithms

In this subsection, we use the baseline dataset to compare the performance of two positioning algorithms: 1) iterative pose estimation algorithm and 2) EPnP. The mean



(a) (b)

Figure 4.5: Empirical CDF of positioning accuracy (a) location accuracy (b) orientation accuracy

values are provided in Table 4.1 and the empirical CDF of location accuracy and orientation accuracy are shown in Figure 4.5.

Table 4.1: Comparison of performance

Pose estimation algorithm	**Mean location accuracy (ft.)**	**Mean orientation accuracy (degree)**	**Mean computation time (ms)**
Iterative	4.88	1.73	4146
EPnP	7.76	3.09	3821

We note that for both approaches over 90% of the position estimations are located within 10 ft. and less than 5 degrees.

As evidenced by these results, we conclude that the iterative approach shows better performance over EPnP without significant time penalty. Therefore, we decided to apply the iterative pose estimation algorithm in the following performance tests with distorted datasets.

4.4.1.2 Performance Test for Blurriness

In this subsection, we explore the influence of image blurriness on the algorithm performance. To simulate the level of blurriness, we use Gaussian blur with increasing blur radius. The radius of the Gaussian kernel ranges between 0.5 and 9 [82].

As shown in Figure 4.6a and 4.6b, we observe that the likelihood of the positioning success decreases as the level of blurriness increases. This is due to the fact that our algorithm rejects images that do not contain sufficient information required for accurate positioning. We note that the algorithm reliability is higher than 75%.

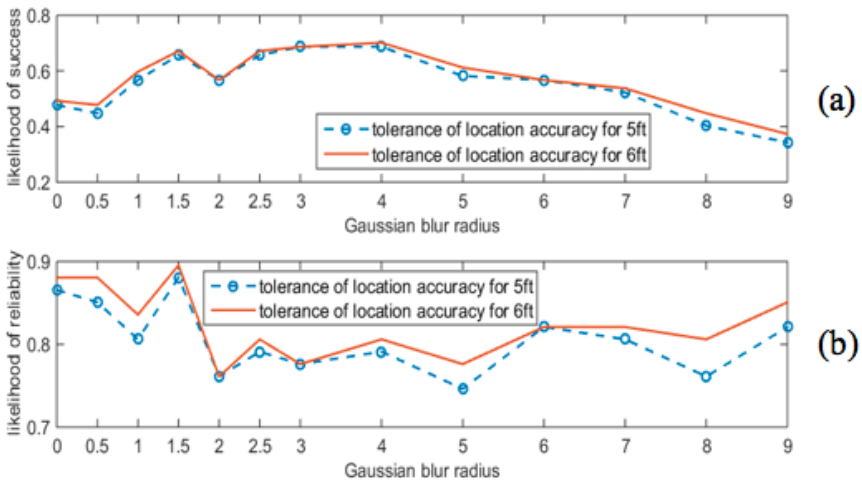


Figure 4.6: Influence of blurriness on positioning accuracy for different success criteria (a) likelihood of success (b) likelihood of reliability

4.4.1.3 Performance Test for Motion Blurriness

We explore the influence of motion blurriness on the algorithm performance. Motion blur is emulated as the number of pixels shifted in a random direction and ranges between 1 and 30 pixels.

As shown in Figure 4.7a and 4.7b, we observe that the likelihood of the positioning success decreases as the level of motion blurriness increases. We note that the algorithm reliability is higher than 75%.

4.4.1.4 Performance Test for Illumination Changes

We explore the relationship between illumination changes on the algorithm performance. The intensity of illumination ranges between 0.1 and 0.9 [82].

As shown in Figure 4.8a and 4.8b, we observe that the likelihood of the positioning success decreases as the intensity of illumination decreases. We note that the algorithm reliability is higher than 75%.

4.4.1.5 Performance Test for Crowd Obstruction

We explore the relationship between the levels of crowd obstruction and the algorithm performance. Level of crowd obstruction is defined by the percentage of the area covered by crowds in a random location on the image plane. The level of crowd obstruction ranges between 0.1 and 0.95.

As shown in Figure 4.9a and 4.9b, we observe that the likelihood of successful positioning decreases as the level of crowd obstruction increases. We note that as the level of crowd obstruction increases above 50% the algorithm reliability deteriorates fast, making it unusable for this range.

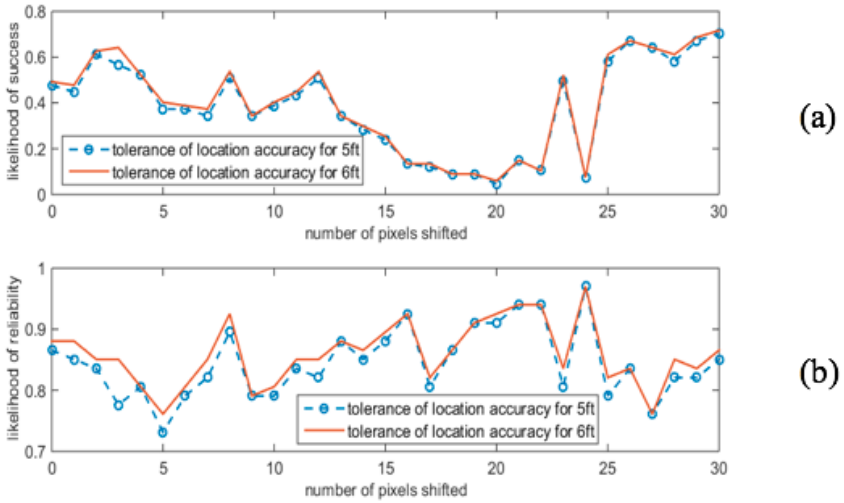


Figure 4.7: Influence of motion blurriness on positioning accuracy for different success criteria (a) likelihood of success (b) likelihood of reliability

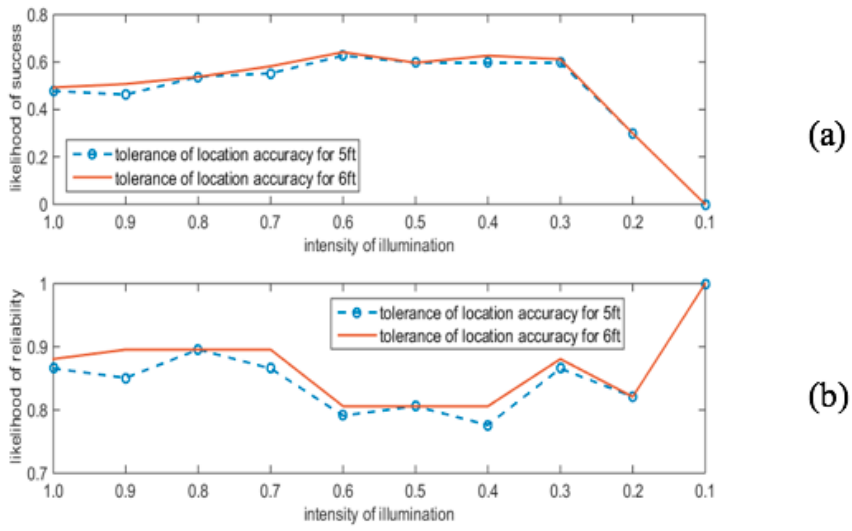


Figure 4.8: Influence of illumination on positioning accuracy for different success criteria (a) likelihood of success (b) likelihood of reliability

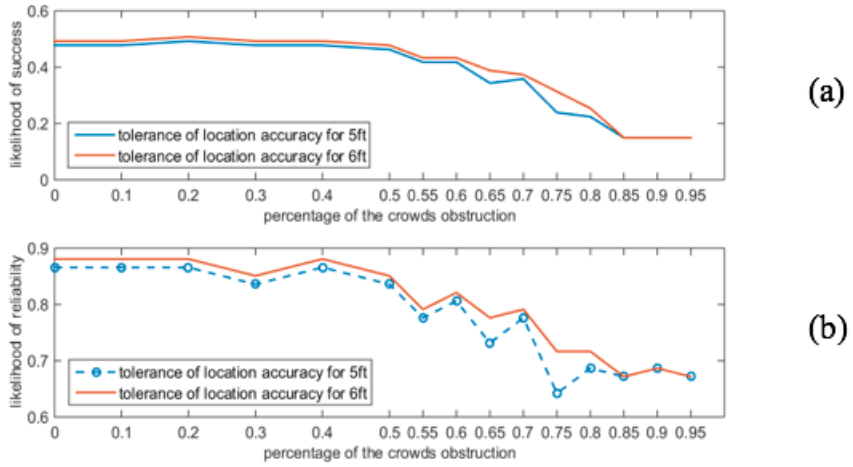


Figure 4.9. Influence of crowds on positioning accuracy for different success criteria (a) likelihood of success (b) likelihood of reliability

4.4.2 Algorithm Performance with Background Extraction

In this section, we will show the performance improvement of new visual localization algorithm including the RPCA background extraction using two experiments. In our first experiments, we will test the advantages of using RPCA background extraction over FuzzySOBS background subtraction [91], one of the best-performing background subtraction algorithms in the literature [92], as the new block in PERCEPT-V. In our

second experiment, we will test the performance improvement (in terms of localization reliability) achieved by using RPCA background extraction to recover higher-quality keypoint descriptors that can be useful for localization purposes.

4.4.2.1 Background Subtraction vs. Background Extraction

For our first experiment, we use data from video sequences captured at the UMass Amherst Campus Center. The reference dataset consists of images of the center's first floor space that were taken while the center was closed. The test dataset consists of 63 video sequences taken in two groups: 26 sequences were taken during low levels of activity (winter break) and 37 sequences were taken during high levels of activity. The number of frames in each video sequence is 200 and the resolution of each frame is 192 by 108. We performed SIFT feature extraction [84] and subsequent feature matching for three different versions of each video sequence: (i) the original video sequence frames, (ii) the background frames extracted from the FuzzySOBS algorithm, and (iii) the background frames extracted with the RPCA model.

We consider not only the number of keypoint descriptors obtained from each frame of the video sequence, but also the percentage of keypoint descriptors from the tested image that are successfully matched to descriptors in the registered image.

Figure 4.10 shows the average number of keypoint descriptors extracted from each of the video frames as a function of the frame index, as well as the average percentage of those descriptors that are successfully matched during the feature matching process. These quantities are averaged over the 26 video sequences containing low activity. The results show that the quality of FuzzySOBS is poor, resulting in a very low percentage of features being matched. Moreover, the quality of the background degrades as further frames are

processed. Furthermore, the percentage of keypoint descriptors matched in the RPCA background image is higher than that obtained from the original images, which is indicative of the higher quality background keypoint descriptors obtained from RPCA.

We also processed video sequences with high levels of activity, which poses a more challenging setting for background subtraction algorithms. The percentage of features that are matched between the captured images and the reference images is very small, as shown in Figure 4.10, due to the large number of features obtained from the foreground activity. Since the background extraction algorithms are not completely successful, the percentage of matched SIFT features stays low. Nonetheless, it is still the case that the RPCA background image provides higher-quality keypoint descriptors (in aggregate) than the two alternatives.

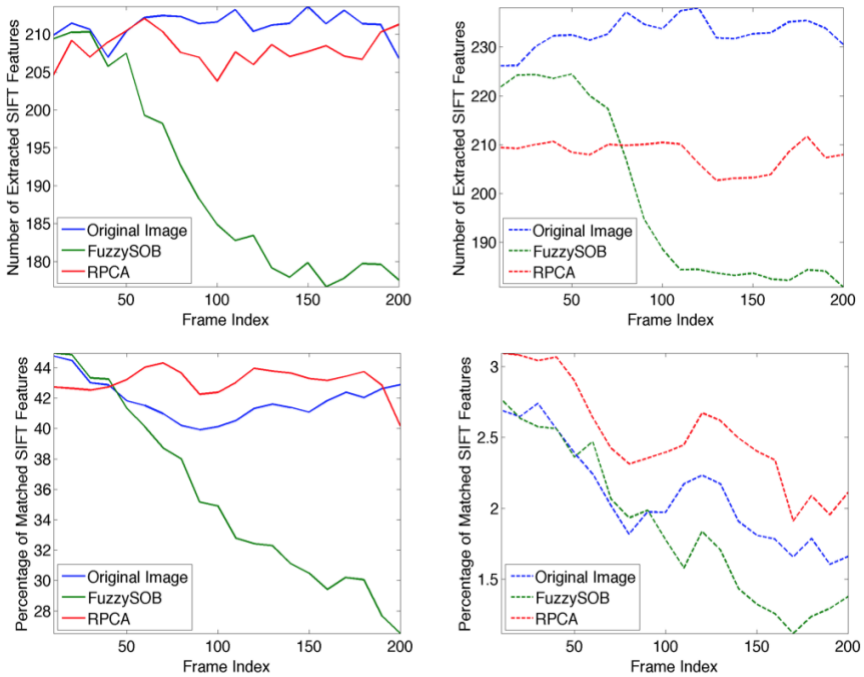


Figure 4.10: Average performance of SIFT descriptor-based image matching for (i) original video sequences, (ii) backgrounds obtained from the FuzzySOBS algorithm, (iii) backgrounds obtained from the RPCA algorithm.
Top row: Average number of SIFT features extracted for each frame of the video sequence. Bottom row: Average percentage of SIFT features matched during camera localization. Left Column: Sequences with low and medium-level activity. Right Column: Sequences with high-level activity

4.4.2.2 Improvement of Localization Reliability

For our second experiment, we collected an additional set of test data at the UMass Amherst Campus Center. The new test dataset contains 77 video sequences and there are 15,400 frames in total. The number of frames in each video sequence is 200 and the resolution of each frame is 480 x 270. The goal for this experiment is to determine the impact of RPCA background extraction on the performance of visual localization algorithm as measured by the number of 2D-to-3D correspondences in the inlier set returned by the pose estimation (last block in Fig. 4.2) using RANSAC. An increase in the number of inlier correspondences is indicative of improved localization performance, given that there are more higher-quality keypoint descriptors that are recovered by RPCA and represent a single perspective hypothesis. We checked this impact by plotting the number of correspondences from the background image obtained by RPCA to the number of correspondences from the original image without processing by RPCA.

As shown in Figure 4.11, we observe significant increments on the number of correspondences for many images after applying RPCA background extraction, as evident by the large number of points on the upper triangle of the figure. For the rest of the frames, almost all of the marks in the figure are close to the diagonal, implying that any negative effects on a frame from applying RPCA are minor. Based on our observation, the slight reduction in number is due to the presence of the blurriness caused by either the capturing or the background extraction. Among 15,400 frames, there are 9,339 frames (60%) whose number of inlier correspondences after RPCA is larger or equal than that before RPCA.

Furthermore, since our interest focuses on recovering higher-quality keypoint descriptors as many as possible for the localization algorithm when the background is

almost covered by passerby (i.e., frames containing less than 10 correspondences, reflecting such scenarios), we analyzed the result particularly for these worst cases to check if RPCA background extraction can benefit the visual localization algorithm for this type of circumstances. Figure 4.12a shows a histogram for the increment of the number of correspondences due to the use of RPCA. The figure shows that the benefits of applying RPCA are much larger than the losses, since the range of increments is from -20 to 140. Among the 737 worst-case frames, 716 frames (97%) have equal or higher number of correspondences after applying RPCA background extraction.

Figure 4.12b shows the cumulative error distribution function for localization with and without RPCA background subtraction in the processing pipeline. The figure presents the substantial improvement of localization performance after RPCA is applied due to the increment of number of 2D-to-3D correspondences, which occurs thanks to the higher number of higher-quality keypoint descriptors recovered by the RPCA background extraction. Besides, the average localization error for these worst cases reduces from 12.50m to 8.14m after RPCA is applied, reflecting that 35% of the localization error is reduced in average. These results imply that the increase in the number of correspondences after applying RPCA results in an improvement of localization performance.

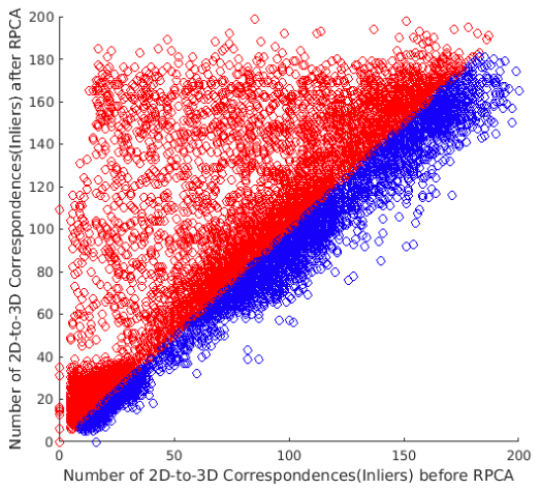


Figure 4.11: Comparison of number of 2D-to-3D correspondences (inlier set) before and after RPCA-based background extraction is applied to the frames of the video sequences

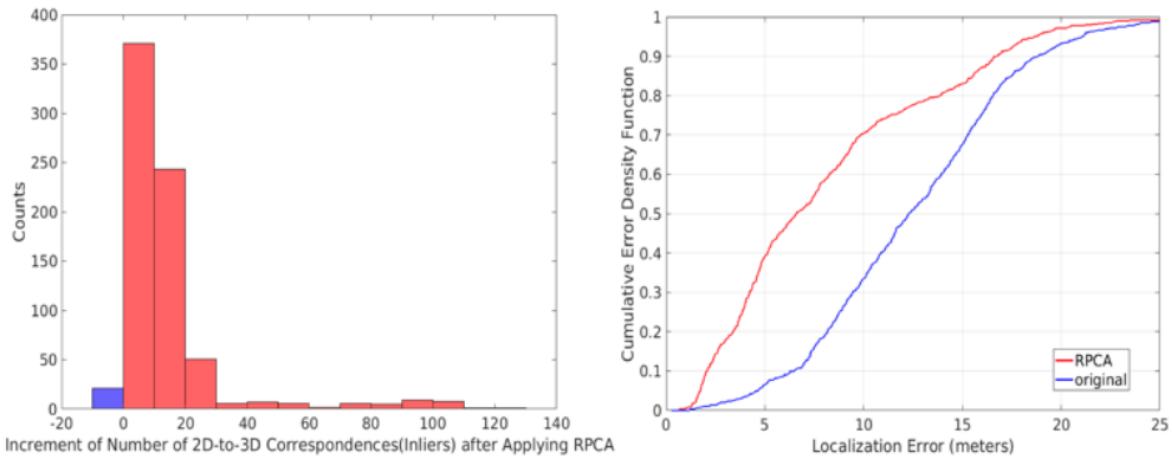


Figure 4.12: Left: Histograms for increment in number of 2D-to-3D correspondences (inlier set) after RPCA. Right: CDF of localization error in meters.

4.5 Conclusion

We introduced the first egocentric vision-based positioning technique developed on Google Glass that processes reference information from point cloud instead of images and uses a noniterative pose estimation algorithm. Moreover, the algorithm is the first to successfully perform with realistic image datasets captured by visually impaired users that include deformations such as severe blurriness, motion blurriness, low illumination

intensity and crowd obstruction. When we have mint query images our algorithm obtains mean location accuracy within 5 ft., mean orientation accuracy less than 2 degrees and reliability above 88%. After applying deformation effects to the query images such blurriness, motion blurriness and illumination changes, we observe that the reliability is above 75%. In case of crowds' obstruction, we notice that when the level of crowds' obstruction increases above 50% the algorithm reliability deteriorates fast, making it unusable for this range.

In addition, we proposed PERCEPT-V, an indoor navigation system for the BVI users based on a novel visual localization algorithm resilient to crowds in the observed scene. We addressed the new challenges in localization faced by the system using RPCA background extraction to increase the localization reliability. Unlike popular background subtraction algorithms, RPCA background extraction enables us to model the background more accurately by leveraging a low-rank-plus-sparse decomposition of a matrix representation of the tested video sequence. With more useful information available in the extracted background, the proposed visual localization algorithm can increase its reliability in the presence of crowds.

Our experimental results indicate two positive findings. First, we show that RPCA background extraction outperforms FuzzySOBS in obtaining an accurate background model for PERCEPT-V. Second, we demonstrate an improvement of localization reliability when using RPCA background extraction that is due to the recovery of more background-generated keypoint descriptors that can be matched to those from the reference images. We anticipate future work in the direction of finding the best-performed RPCA algorithm implementation by comparing among all candidates.

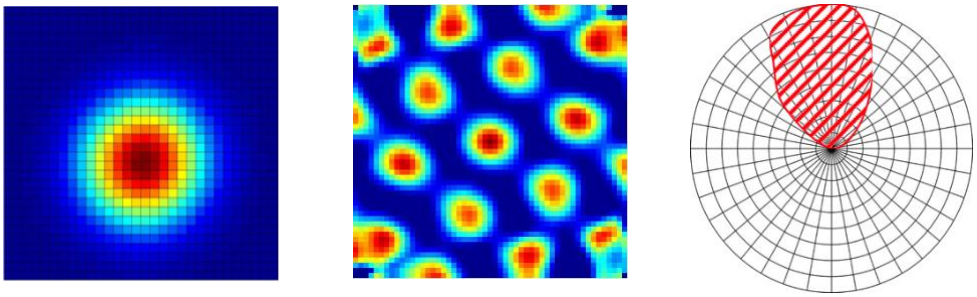
CHAPTER 5

INTELLIGENT SENSING FRAMEWORK FOR INDOOR SPATIAL AWARENESS FOR BVI

5.1 Introduction

Traveling independently in unfamiliar large public venues is a very challenging task for blind and visually impaired (BVI) people [93]. In this chapter we introduce a framework that provides spatial cognition to BVI users enabling them to navigate independently in large public venues.

In cognitive neuroscience, there is a consensus [94] that one's ability of navigation depends on one's capability to build the cognitive map of the space. With the mental representation of space, humans can position and navigate themselves inside the space. To identify the needs of humans in spatial positioning and navigation and develop corresponding navigation aids, it is necessary to know how the brain encodes the space from a neuroscientific viewpoint [95]. There are four types of cells found in the spatial representation in the brain (see Figure 5.1).



(a) Place cells (b) Grid cells (c) Head-direction cells

Figure 5.1: Illustration of firing patterns. In (a) and (b) the color code demonstrates the rate of firing: red shows high activity and blue shows low activity.

1. **Place cells** (Figure 5.1a) are pyramidal neurons inside the hippocampus which fire when an individual (animal or human) visits a particular place (small region) in the environment, thereby exhibiting a representation of the place with respect to the environment [96]. This

area of high firing rate is known as the cell's 'place field'. Such place fields are considered to be allocentric, implying that they are defined by the external recognizable cues in the environment, for instance, landmarks. While visual input comprises a key element in the formation of place fields, it was shown in [97] that in the absence of visual input, both humans and other vertebrates studied in this context, are capable of generating very effective spatial representations using other sensory input.

2. **Grid cells** (Figure 5.1b) are neurons within the medial entorhinal cortex (MEC) which exhibit firing at multiple locations in the environment. The spatial firing fields are positioned regularly in a grid across the environment comprised by equilateral triangles. Unlike place cells, grid cells seem to be the internal cognitive representation of the external Euclidean space. Moreover, it is found that grid cells play a critical role in path integration (i.e. navigation or wayfinding) since their firing depends on the ego-motion of the individuals, such as moving direction.

3. **Border/boundary vector cells (BVC)** are neurons found in the hippocampal formation which fire when the individual is at a specific distance and direction relative to the environment boundaries such as walls, low ridges or vertical drops

4. **Head-direction cells** (Figure 5.1c) are neurons which fire in a range around the head direction. Head direction cells fire maximally when an animal's head faces a particular direction in the azimuthal (horizontal) plane. The firing relies on the angular position of environmental cues [98-101]. Like place cells, the firing of head direction cells has been shown to rely on the angular position of environmental cues and generate a lobe of a certain width.

These extraordinary findings that confirm the spatial nature of the entorhinal-hippocampal system led to the award of the 2014 Nobel Prize in Physiology or Medicine to O'Keefe and the Mosers for their discovery of "a positioning system in the brain." [102].

It is important to mention that researchers [103-106] have shown that the hippocampus can use non-visuospatial input such as spatial olfactory and tactile information, to generate spatial representations. In spite of the fact that olfactory input is less precise than visual input, it can substitute for visual inputs to enable the acquisition of metric information about space. However, for BVI users traveling through large public venues it is difficult or sometime impossible to use only olfactory and/or tactile information to form the cognitive map of the space. Therefore, in [37, 38] we introduced PERCEPT system which complements these senses and helps the BVI user to build the cognitive map which helps them to independently navigate through large public venues. PERCEPT provides users with audio/text information about their location in space relative to landmarks, proximity to landmarks as well as detailed instructions to their chosen destination. In order to provide such information, PERCEPT system incorporates the sensing framework which we introduce in this chapter. As reported later in this chapter, the framework was tested within the entire PERCEPT system and shown that it provides excellent information that helps the BVI users to build a cognitive map of the space and reach the chosen destination independently [107]. Figure 5.2 illustrates the architecture of PERCEPT system which includes the sensing framework presented in this chapter, the navigation and instruction generation module and the user interface.

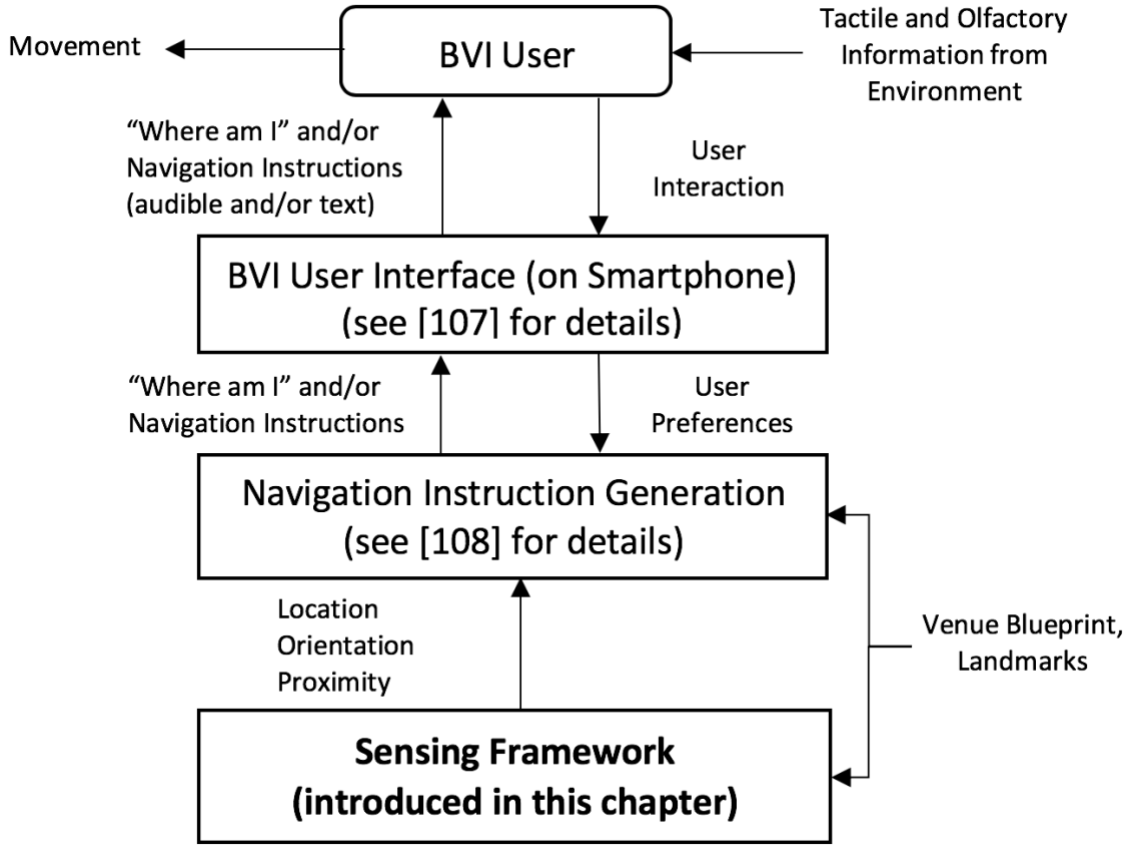


Figure 5.2: PERCEPT system architecture

5.2 System Overview

The sensing framework introduced in this chapter includes the following modules that correspond to the abovementioned place, border and head-direction cells: a localization module, a moving direction estimation module and a landmark proximity detection module. Grid cells correspond to the graph representation of space included in the Navigation Instruction Generation Module (see Figure 5.2) [108]. As shown in Figure 5.3, the proposed framework includes two phases, the offline phase and the online phase. In the offline phase, we generate the optimal sensor deployment strategy for the indoor space, minimizing the cost of the deployment while considering the requirements of the localization algorithm. During the online phase, we develop intelligent sensing algorithms

that provide location estimates, moving direction estimates and detect landmarks next to the user.

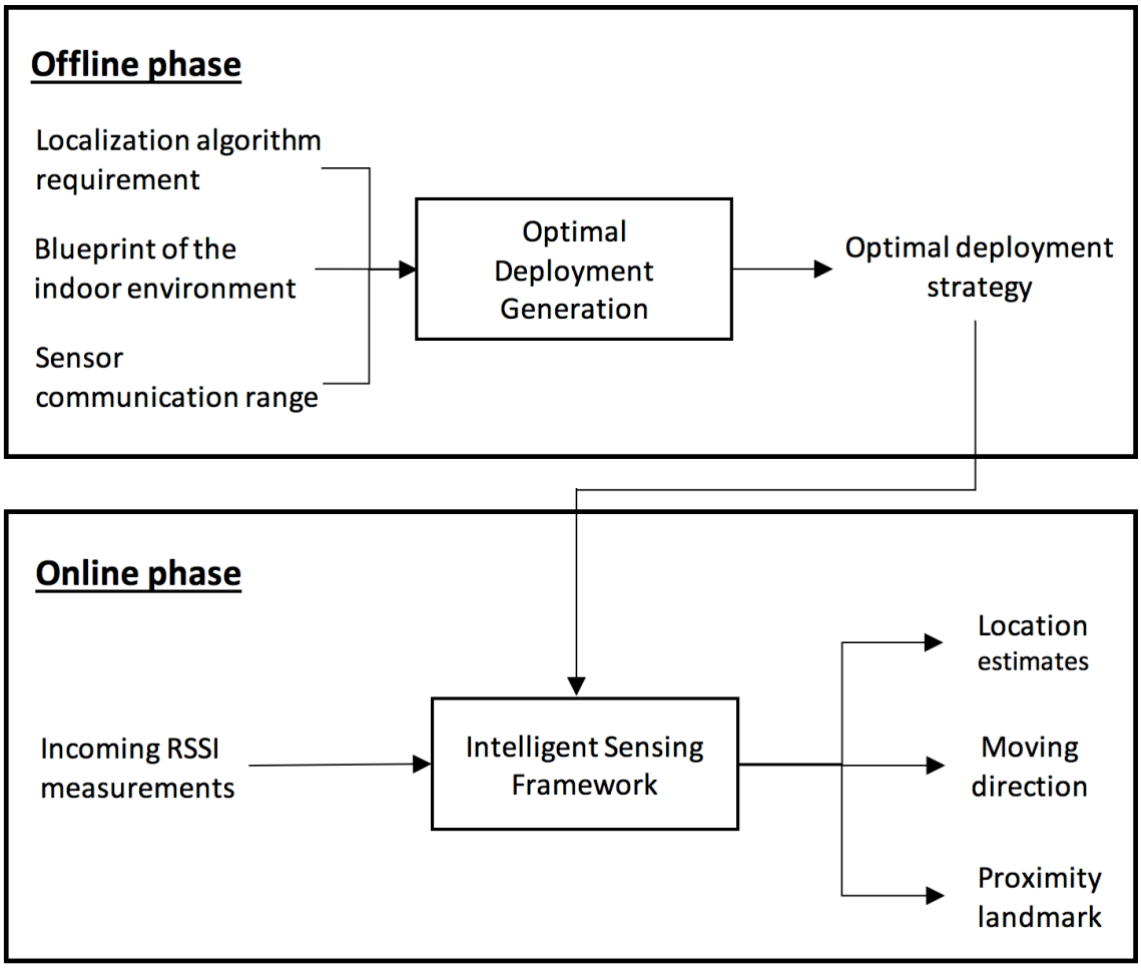


Figure 5.3: Sensing framework overview

In contrast to the established literature, our framework will not seek to achieve the exact coordinates (e.g. singular point) of the user location or exact value of user orientation in degrees. Neuroscience reveals that the human cognitive system for positioning and navigation uses a region to understand the location instead of a singular point. Therefore, we propose to evaluate the performance of the localization algorithm using the success rate of region detection for different region sizes. In addition, as shown in Figure 5.1c and reported in [109], head-direction cells fire in a range around the preferred firing direction

(e.g. the direction at which neurons fire maximally). Thus, instead of evaluating the moving direction algorithm using deviation of the estimated moving direction from the ground truth, we evaluate its performance using success rate of estimating a 4-way or 8-way cardinal direction.

Table 5.1 summarizes the spatial information represented in each type of neuron and the corresponding modules provided in our framework.

Table 5.1: Analogy between neural representation of space and proposed framework

Neuron Type	**Neuron Firing Condition Given environment input**	**Technical Equivalence (modules) In Sensing Framework (synthesizing input)**
Place Cells	Fired in small specific regions (landmarks) of the environment	Localization module
Border Cells	Fired in small regions around the environment boundary	Landmark proximity detection module (provides border detection as well as proximity to other landmarks)
Head-direction Cells	Fired in a range around the head direction	Moving direction module

5.3 Methodology

5.3.1 Offline Phase

In this phase we generate an optimal sensor deployment strategy for an indoor environment. The input includes: the blueprint and its associated scale, the sensor communication range and the number of sensors, k, that should cover each point in the blueprint (dictated by the localization algorithm). To ensure the coverage, we use the optimal deployment pattern that guarantees k-covering [110]. If k equals to 3, the optimal deployed pattern is shown in Figure 5.4.

The sensor deployment algorithm is implemented in Matlab and the graphical user interface (GUI) is shown in Figure 5.5, which includes the following parts:

- *Blueprint (Top):* displays the blueprint as background and the superimposed locations of the sensors obtained from the deployment algorithm using red dots.
- *Parameter settings (Input-bottom left):* sets the value of k, BLE communication range and the scale of the blueprint.
- *Interaction (center):* marks the deployment area with a blue box, triggers the action of computing the optimal deployment or resets the current parameters.
- *Results (Output-bottom right):* the size of the area covered by the BLE infrastructure, the total number of beacons required and the density of the beacons in the deployment region (i.e. the size of area covered by one beacon).

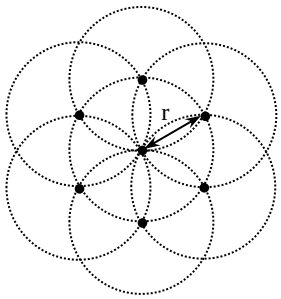


Figure 5.4: Optimal pattern when k=3 (r is BLE communication range in ft.)

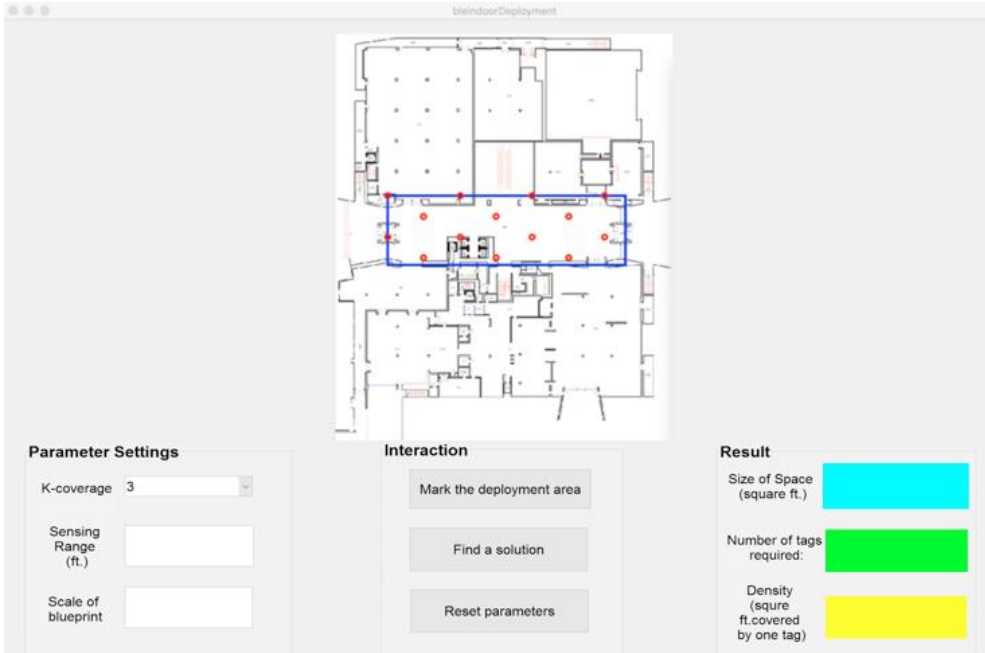


Figure 5.5: GUI of the optimal deployment strategy generation

Note that the generated optimal sensor deployment may not be followed exactly while deploying the sensors in the environment due to the physical limitations of each suggested location. Nevertheless, the optimal deployment strategy still provides a valuable insight for guaranteeing the performance of the localization algorithm across the entire environment.

5.3.2 Online Phase

The online phase includes the following modules: localization (Section 5.3.2.1), moving direction estimation (Section 5.3.2.2), and landmark proximity detection (Section 5.3.2.3).

5.3.2.1 Localization

We assume that the user's smartphone can collect RSSI measurements from k nearby BLE sensors simultaneously. Distance d_i from BLE sensor i to user's smartphone is given by:

$$d_i = 10^{\frac{s_i - PL_0}{10*\gamma}} \tag{5.1}$$

where s_i refers to the RSSI collected from BLE sensor i, γ is path loss component and PL_0 is the path loss at the reference distance d_0.

Eq. (5.2) describes the Log-distance path loss propagation model. Given a certain *distance*, $RSSI$ can be computed by:

$$s_i = PL_0 + 10 * \gamma * log_{10}(d_i/d_0) \tag{5.2}$$

The weighting factor w_i is given by:

$$w_i = \frac{1/d_i}{\sum_{i=1}^{k} 1/d_i} \tag{5.3}$$

The location estimates (u, v) of the user can be obtained by:

$$u = \sum_{i=1}^{k} w_i * x_i \tag{5.4}$$

$$v = \sum_{i=1}^{k} w_i * y_i \tag{5.5}$$

where (x_i, y_i) is the 2D location of BLE sensor i.

5.3.2.2 Moving Direction Estimation

To determine user's moving direction from the sequence of location estimations, we use the K-means clustering algorithm in conjunction with a sliding time window.

Given a set of location estimates computed in the past *N* seconds, $e^{(1)} \ldots, e^{(N)}$, we group the data into two cohesive clusters, extracting the moving trajectory from two centroids. The pseudocode implementation of the algorithm is provided in Table 5.2.

Table 5.2: Pseudocode of moving direction estimation

```
Input: N = 10 seconds (size of the sliding window)
       E = {} (set of entities to be clustered)
       k = 2(number of clusters)
       MaxIters = 300

Output: C = {c_past, c_present} (set of start and the end of the moving trajectory)

Do the following forever:
if a new location estimate e_i is received, then
    add e_i into E
end
if E contains more than N elements, then
    foreach c_i ∈ C do
        c_i ← e_j ∈ E (e.g. random selection)
    end
    foreach e_i ∈ E do
        l(e_i) ← argminDistance( e_i, c_j ) j ∈ {1...k} (e.g. Euclidean distance)
    end

    changed ← false;
    iter ← 0;
    repeat
        foreach c_i ∈ C do
            updateCluster(c_i);
        end
        foreach e_i ∈ E do
            minDist ← argminDistance(e_i, c_j) j ∈ {1...k};
            if minDist ≠ l(e_i) then
                l(e_i) ← minDist;
                changed ← true;
            end
        end
        iter++;
    until changed = true and iter ≤ MaxIters;

    c_previous ← findMostCommonLabel({l(e)|e = 1, 2, ..., 5});
    c_present ← findMostCommonLabel({l(e)|e = 6, 7, ..., 10});
    C = {c_previous, c_present};
end
```

As shown in Figure 5.6, we cluster the location estimates in a sliding time window into two groups using K-means algorithm. Using the centroids of the two clusters, we determine the moving direction by finding the trajectory from the past centroid to the present one.

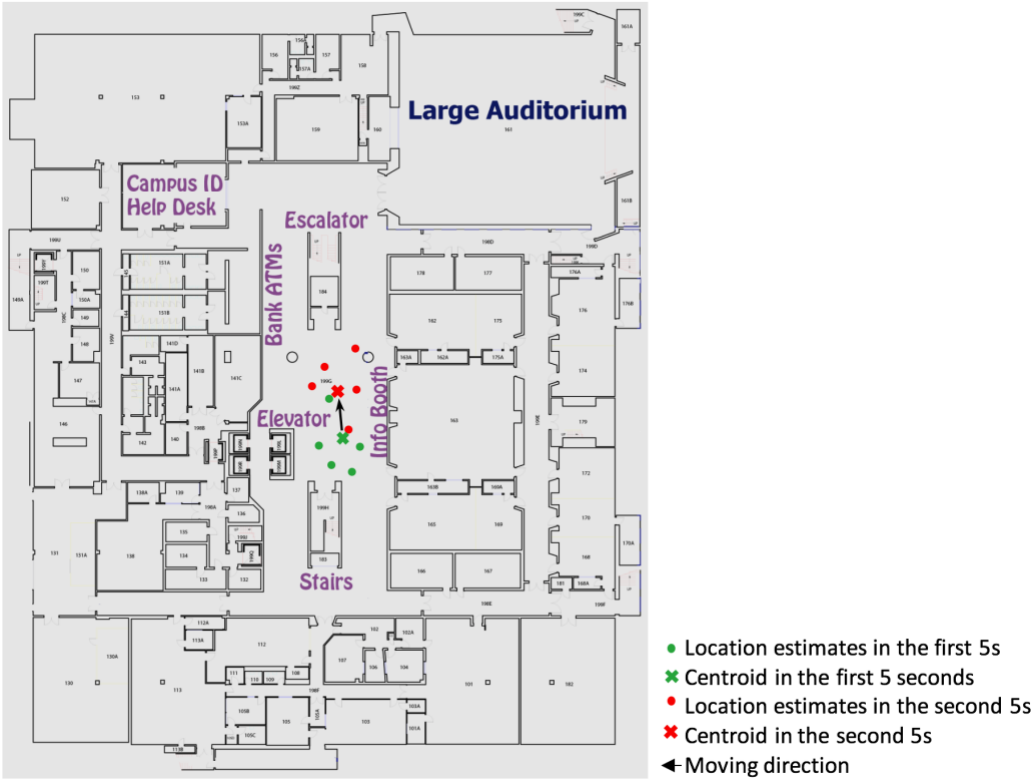


Figure 5.6: Illustration of moving direction estimation

5.3.2.3 Landmark Proximity Detection

In addition to location and moving direction estimation, the proposed framework also offers the functionality of landmark proximity detection. In conjunction with this contextual information related to the navigational space, this function can enhance users' perception by detecting the landmark next to the user, which can be either a landmark along the boundary of the space or a landmark in the middle of the space.

Since our detection problem can be treated as a binary classification problem, we leverage the naïve Bayes classifier to detect the landmark next to the user. For each landmark, the two possible outcomes are either the user stays in close proximity to the landmark or the user stays out of a certain radius of the landmark. To be more specific, we first define the proximity radius for each landmark. Second, we train the probabilistic model that will be used in naïve Bayes classifier using the RSSI measurements collected

at different distances. For the labelling process, all the RSSI measurements collected within the radius are labelled with 1, and other measurements are labelled with 0.

Mathematically, the conditional probability model for each landmark can be calculated using (5.6).

$$p\left(C_i^k \middle| x_i^1, \dots x_i^n\right) \propto p(C_i^k, x_i^1, \dots x_i^n) \tag{5.6}$$

where x_i^j denotes the j th RSSI measurement from sensor i in the sliding window, and C_i^k denotes the outcome for landmark i. Each landmark is paired with a certain sensor.

The joint model can be expressed as follows:

$$p\left(C_i^k, x_i^1, \dots x_i^n\right) = p\left(C_i^k\right) p\left(x_i^1 \middle| C_i^k\right) p\left(x_i^2 \middle| C_i^k\right) \dots$$

$$= p(C_i^k) \prod_{j=1}^{n} p(x_i^j | C_i^k) \tag{5.7}$$

Finally, the decision can be made using:

$$\widehat{y_i} = arg \max_{k \in \{0,1\}} p(C_i^k) \prod_{j=1}^{n} p(x_i^j | C_i^k) \tag{5.8}$$

where n refers to the number of RSSI measurements collected from the sensor that is paired with landmark i, $\widehat{y_i}$ denotes the estimated class for landmark i.

5.4 Performance Evaluation

To evaluate the performance of the algorithms included in the proposed framework, we deployed our testbed at the basement level of UMass Campus Center, which is a 9,000 ft^2 area (see Figure 5.7). We use Bluetooth Low Energy sensors manufactured by Kontact.io and different models of iPhones (iPhone 6, iPhone 6 plus and iPhone 6s plus) to collect our datasets. The hardware specifications [111] of the BLE sensor are provided in Table III. The sensors' transmission power is set to medium level, -12 dBm. Due to the default Bluetooth communication setting in iOS, the frequency of collecting the RSSI

signal is 1 Hz. The BLE infrastructure is deployed following the guidance of the optimal deployment strategy with minor adjustment to the environment.

Table 5.3: Hardware specifications of Tough beacon

Model	Tough Beacon
Operating Frequency	2.4 GHz
Transmission Power	-30 dBm to 4 dBm
Transmission Range	Up to 100 ft.
Time Interval	350 ms
Battery Life	Up to 2 years with default setting

Our dataset contains 35 and 34 groups of RSSI measurements along Route 1 and Route 2, respectively. As shown in Figure 5.7, we collected RSSI measurements following the testing routes so that the ground truth walking trajectory is known to us. Along each route, we pressed a button on our data collection application when we passed by the marked checkpoints (6 for Route 1 and 4 for Route 2). The recorded information is used as the ground truth for evaluating the proximity landmark detection module. Since we also known that the moving direction for each walking trajectory, the ground truth moving direction can be collected as well.

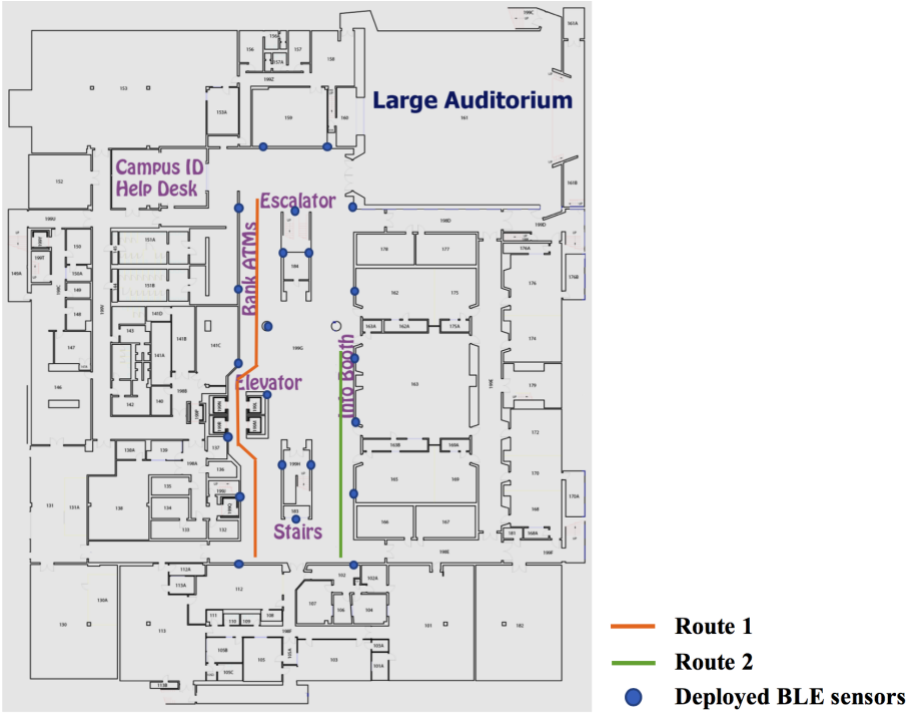
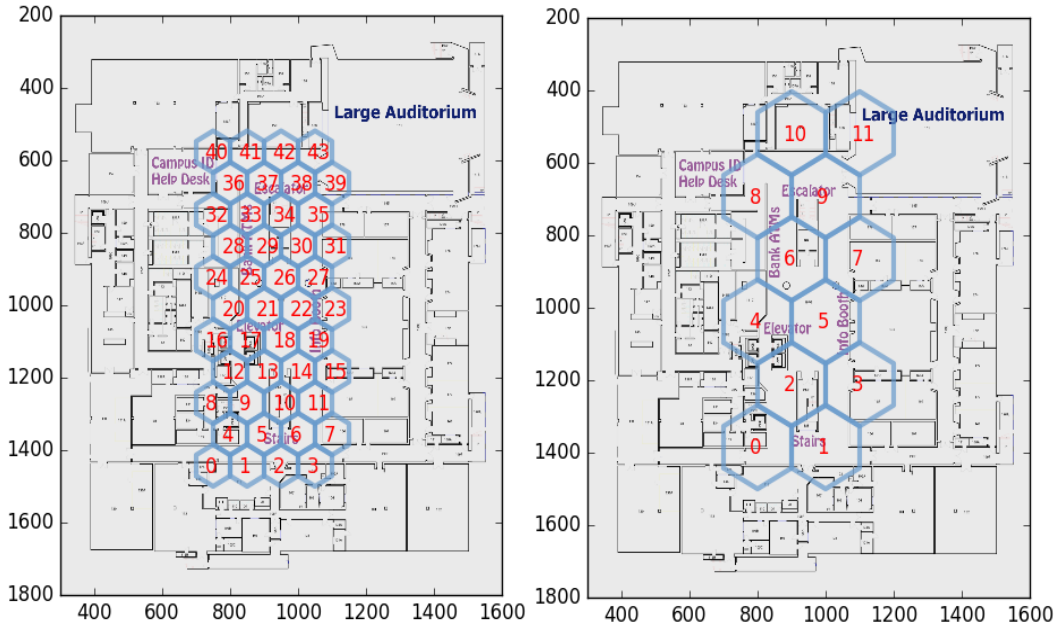


Figure 5.7: Testbed at UMass campus center (180 ft x 50 ft)

5.4.1 Localization

In equations introduced in Section 5.3.2.1 we use PL_0 = -63.5379 dB, γ = -2.086 and d_0 = 3ft.

We evaluate the success rate of region detection which is defined as the percentage of correct region estimations. As shown in Figure 5.8, we generate the hexagon tessellation of the space following the format of spatial representation used in neural cells [95]. As shown in Figure 5.9, while the radius of the hexagon-shaped region increases from 10 ft. to 20 ft., the success rate of region detection increases from 62.5% to 83%.



(a) radius is 10 ft. (b) radius is 20 ft.

Figure 5.8: Hexagon tessellation of the space

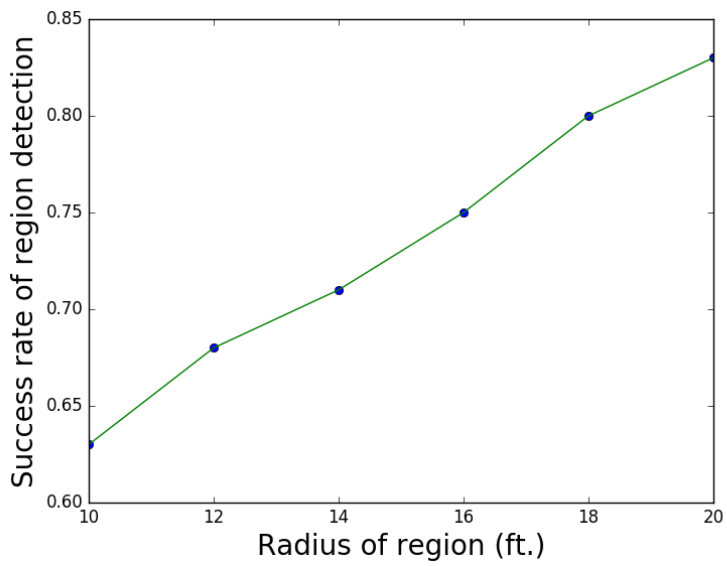
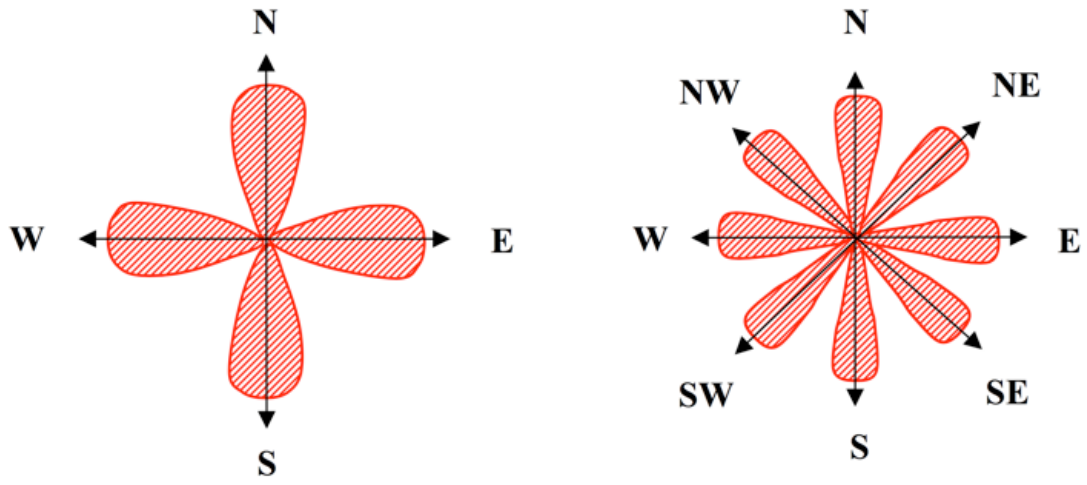


Figure 5.9: Experimental results of region detection

5.4.2 Moving direction

According to a neuroscientific study reported in [109], the head-direction cells used in human cognitive system for orientation will fire in a *range* around the preferred firing direction. The preferred firing direction is defined as the direction at which the neuron fires maximally. Thus, to evaluate the moving direction module, we propose to evaluate

the success rate of estimating a specific cardinal direction (see Figure 5.10) determined by the orientation vector calculated in Section 5.3.2.2. Using 10-seconds sliding window, 1395 estimates are generated from our dataset. Among these estimates, the orientation estimation success rate for 4-way and 8-way cardinal directions is about 94% and 60 %, respectively.



(a) 4-way cardinal directions (b) 8-way cardinal directions

Figure 5.10: Illustration of Head-direction cells firing pattern around cardinal directions

5.4.3 Landmark Proximity Detection

The success rate of landmark proximity detection is defined as the percentage of correct proximity landmarks detected. Given a pre-defined 6-ft. radius around each landmark, if the user is detected in one of these landmark-centered regions, this landmark is detected as the proximity landmark. From 1153 estimates, the success rate of the landmark proximity detection is about 81%.

5.4.4 Integration with PERCEPT Indoor Navigation System for BVI users

The proposed sensing framework was integrated in PERCEPT system (see Figure 5.2) and tested in a large public transportation hub in Boston [107]. Using PERCEPT application installed on users' iPhones, BVI users can: 1) understand where they are in the

space audibly using regions and moving direction calculated by our sensing framework, 2) receive audible navigation instructions from one landmark to another using surrounding landmarks, and 3) receive alerts if they approach some landmarks via proximity landmark detection.

The system has been tested by 6 BVI subjects in subway station [107]. The experiments show that the users can successfully and independently reach their chosen destinations. All participants were very satisfied with the navigational aids provided by PERCEPT. Details of this deployment as well as testing results can be found in [107].

It is important to note that we expect for that the user will make mistakes (i.e., reach wrong landmarks or just get disoriented in the environment) and therefore require rerouting as provided by the application. Rerouting assistance in the application includes the ability to press “Where am I” as well as getting instructions from any landmark to the chosen destination. It was interesting to observe that the participants reported that they have built a cognitive map of the environment using the application routing and rerouting feature as well as the “Where am I” feature.

For completeness of the chapter we include some of the feedback provided by the BVI users which experienced PERCEPT and reported in [107]. After the trials we collected the participants’ feedback and experience using a qualitative questionnaire. Each participant was asked to score their agreement with specific statements related to their experience during the trial. The score followed Likert scale from 1 strongly disagree to 7 strongly agree with, with 4 being neither agreeing or disagreeing with the statement. The averaged scores are provided in Table 5.4. The nine additional trials we performed in a large indoor transportation venue showed similar trends.

Table 5.4: Post-trial questionnaire using Likert scale score

Statements	Average Score	P1	P2	P3	P4	P5	P6
Easy to learn how to use system	6.7	7	7	7	7	6	6
Easy to use the system	6.7	7	7	7	7	6	6
Trial design was easy to complete	6.3	6	6	7	7	7	5
Easy to use User Interface	6.5	7	6	7	7	5	7
System provided sufficient re-orientation information when lost	5.5	5	7	7	6	3	5
I am confident I will reach the destination using this system	6.7	6	7	7	7	7	6

As shown above, we conclude that the information we compute in the proposed sensing framework, i.e., zone localization, region orientation and proximity, can be successfully used by the navigation instructions module to convey the necessary information to allow the BVI user to independently navigate through large indoor venues.

5.5 Conclusion

We introduced a sensing framework that includes optimal deployment strategy, location estimation, moving direction estimation and landmark proximity detection. We note that in spite of the fact that the location, orientation and proximity results computed by our sensing framework are not accurate, they provide the necessary information to the rest of PERCEPT modules and ultimately enable the BVI user to successfully navigate independently in large indoor venues. It is interesting to note that unlike a sensing framework for robots that needs to provide very accurate location and orientation, in the proposed framework used by BVI users high accuracy is not necessary. The reason is that

the movement/navigation decisions made by the user include diverse aspects such as how the navigation instructions are composed, how the user interface is designed and how accurate the user can interpret the instructions and/or manipulate the user interface. Our observation is also aligned with the space encoding methods presented in this chapter which show that the brain encodes zones (place fields) using place cells (i.e., zones, not singular points) as well as orientation regions using head cells with wide lobes (see Figure 5.1c) (i.e., do not record highly accurate directions using very narrow width lobe).

CHAPTER 6

SENSOR FUSION FOR PEDESTRIAN INDOOR LOCALIZATION

6.1 Introduction

Accurate and reliable pedestrian indoor localization can benefit a variety of systems, including personalized advertisement, assistive living, emergency evacuation and etc. Although the GPS can provide sufficient localization accuracy in outdoor environments, it does not work indoors. Therefore, pedestrian localization techniques using wireless signals and/or other sensors are being developed [112-125].

In this chapter we introduce a multi-modal pedestrian localization approach that uses input video as well as wireless and inertial data collected from BLE and IMU sensors, respectively. We provide a brief survey of published multi-modal techniques that include WiFi, computer vision, BLE and PDR.

In this chapter we integrate three technologies with the corresponding algorithms:

1. BLE technology: wireless signals transmitted by the BLEs and received by the Smartphone are noisy, fact that may lead to localization inaccuracies. We will use the WPL localization algorithm [74]. In addition, we also introduce a sensor deployment plan.
2. Pedestrian Dead Reckoning (PDR): it is well known that PDR techniques drift over time and require frequent recalibration [126]. We will use a user-specific velocity regression model using the Supported Vector Regression (SVR) algorithm to estimate the user moving velocity [127].
3. Computer vision: localization using computer vision can be very accurate in determining the user location and orientation if the environment has sufficient

features in the observed frame [52] which is not always the case for a number of reasons. First, the environment may not have sufficient features or the camera view may be obstructed by a passerby. Nevertheless, the location and orientation estimates computed from our visual localization algorithm when successful, can still help the proposed algorithm maintain a long-term stability with some opportunistic measurements.

As discussed above, different types of techniques have their own advantages and disadvantages, and none of them can work perfectly without extra assistance. As a result, to increase the accuracy and robustness of the proposed localization algorithm we fuse together this multi-modal data along with the associated data processing techniques, using the Unscented Kalman Filter (UKF) [12].

6.2 System Overview

The architecture of the proposed the algorithm is depicted in Figure 6.1. As illustrated, the entire procedure can be divided into two phases, the offline preparation phase, and the online localization phase. In the offline preparation phase, two types of information will be acquired beforehand, which are the environmental information and user-specific information. First, we will acquire an unordered image collection of the indoor environment so that the Structure from Motion (SfM) technique [128] can use these images to reconstruct the space in 3D. This step is remarkably time-efficient compared to the preparation efforts introduced in [113, 114] because the SfM works without relying on the labeled images. Second, the BLE will be deployed in the space effectively following the optimal deployment plan in order to guarantee the performance of WPL algorithm while minimizing the hardware cost. As for the user-specific information, the proposed

algorithm will train a velocity regression model using the accelerometer and gyroscope readings collected by the pedestrian. Given the user-specific velocity regression model and smartphone attitude, the PDR can estimate the pedestrian's movement directly instead of using step detection and step length estimation. In the online localization phase, the Unscented Kalman Filter (UKF) is exploited to fuse the outputs of PDR, Weighted Path Loss (WPL) and visual localization such that the proposed algorithm can adapt to the variance and dynamics in different regions of the indoor environment.

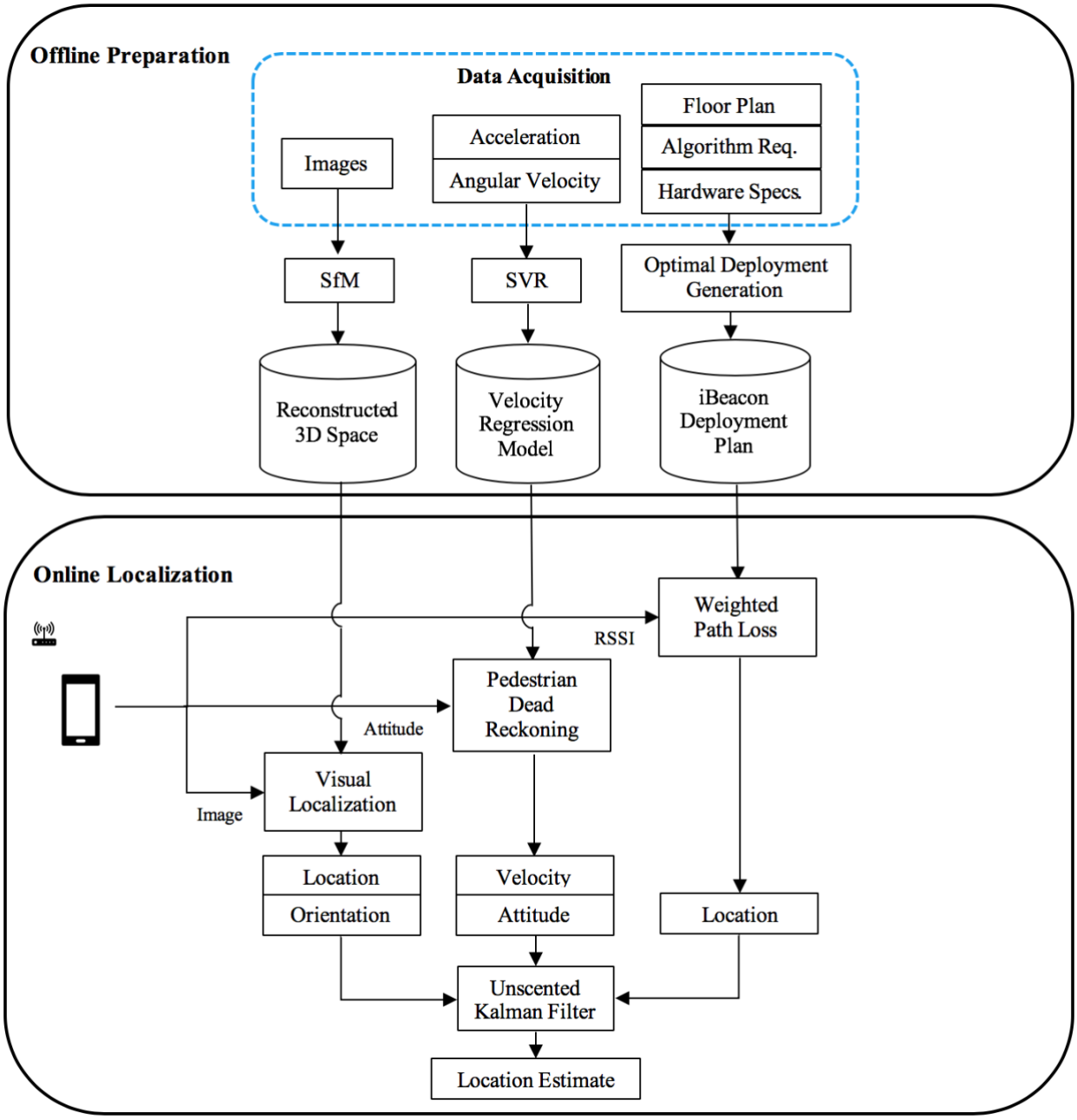


Figure 6.1: Architecture of the proposed algorithm

To make the evaluation process reflect realistic scenarios, we test the proposed algorithm in a very challenging scenario compared to other multi-modal approaches [112-115]. The testing data is collected in a large and open public space, the concourse level of the Campus Center (55 meter x 15 meter) at the University of Massachusetts Amherst, during busy time of the day. To the best of our knowledge, this is the first study of fusing

all three types of multi-modal data and complementary techniques using UKF and evaluating the algorithm performance in a realistic environment.

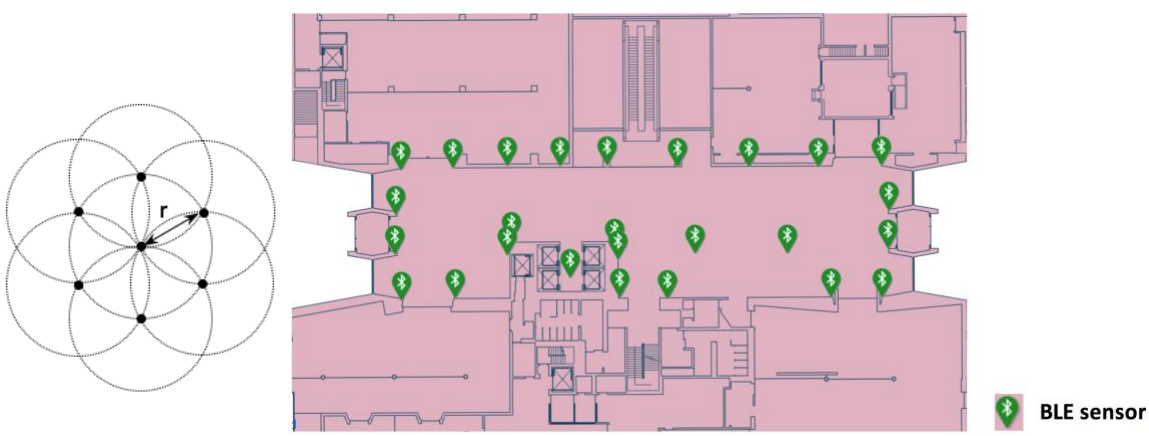
6.3 Methodology

6.3.1 Wireless Localization

6.3.1.1 Sensor Deployment

To balance between the overhead of hardware deployment and localization performance, we adopt the optimal deployment algorithm introduced in [75] which minimizes the number of sensors while considering the sensor communication range, the floor plan and the coverage requirement.

Using the optimal deployment pattern illustrated in Figure 6.2, we guarantee k-coverage for the entire space such that each point will be covered by k=3 BLE sensors in our case. The optimal deployment patterns for $1 \leq k \leq 9$ can be found in [75].



(a) (b)

Figure 6.2: Optimal sensor deployment
(a) Optimal deployment pattern when *k* is 3, *r* refers to the communication range of sensor
(b) Actual deployment plan at UMass Campus Center, the locations for each BLE sensors are adjusted to the physical conditions, accordingly

6.3.1.2 Weighted Path Loss

For each BLE RSSI reading on the Smartphone, the distance between the pedestrian and the BLE sensor can be calculated by (6.1) and it is derived from the Log-distance path loss propagation model as described by (6.2).

$$d_i = 10^{\frac{s_i - PL_0}{10*\gamma}} \tag{6.1}$$

$$s_i = PL_0 + 10 * \gamma * log_{10}(d_i/d_0) \tag{6.2}$$

where s_i refers to the RSSI collected from BLE sensor i, γ is path loss component and PL_0 is the path loss at the reference distance d_0. In our case, PL_0 is -63.5379 dB, γ is -2.086 and d_0 is 1 meter.

Given the distance calculated between the pedestrian and BLE sensor *i*, the weight factor can be computed by (6.3).

$$w_i = \frac{1/d_i}{\sum_{i=1}^{k} 1/d_i} \tag{6.3}$$

The location estimate in world coordinate system, (x_w, z_w) of the pedestrian can be obtained by:

$$\mathrm{x}_w = \sum_{i=1}^{k} w_i * x_w^i \tag{6.4}$$

$$z_w = \sum_{i=1}^{k} w_i * z_w^i \tag{6.5}$$

where (x_w^i, z_w^i) is the 2D location of BLE sensor i with respect to the world coordinate system. (see Figure 6.3) We use x and z axis to express the 2D plane because of the setting of the camera coordinate system (see Figure 6.5).

6.3.2 Visual Localization

The visual localization includes two parts. At first, Structure from Motion technique will reconstruct the space in 3D. With the prepared visual information, the online visual

localization will process the query frame taken by pedestrian and estimate pedestrian's location and orientation using algorithm illustrated in Figure 6.4.

6.3.2.1 Structure from Motion

Structure from Motion (SfM) [128] is the process of estimating the 3D structure of a scene and the camera poses of the sequence from a set of unordered 2D images. The main steps in SfM are feature extraction and matching, image registration, triangulation and bundle adjustment. As a result, either the sparse or the dense point cloud will be used to represent the space in 3D. In our case, the sparse point cloud is incorporated into the visual localization since the included information is sufficient. To reconstruct the space, we use 2315 frames from several videos captured by a regular smartphone during the time when the space is nearly empty and process them using the Visual SFM [129], a SfM software that can compute and visualize the result. The entire collection process takes about 10 minutes. Given the raw output from the SfM software, we further parse it into multiple types of information that will be ready to use in the online localization phase. The detailed information is described in Table 6.1 and the reconstructed 3D space of UMass campus center is exhibited in Figure 6.3.

Table 6.1: Multiple types of information

Data Point (Feature)	**Image ID**	**2D Location**	**3D Location**	**SIFT Feature Descriptor**
F_1	I_1	$(x_p, y_p)_1$	$(x_w, y_w, z_w)_1$	d_1
F_2	I_2	$(x_p, y_p)_2$	$(x_w, y_w, z_w)_2$	d_2
…	…	…	…	…
F_n	I_n	$(x_p, y_p)_n$	$(x_w, y_w, z_w)_n$	d_n

Note that SIFT refers to the Scale-invariant Feature Transform

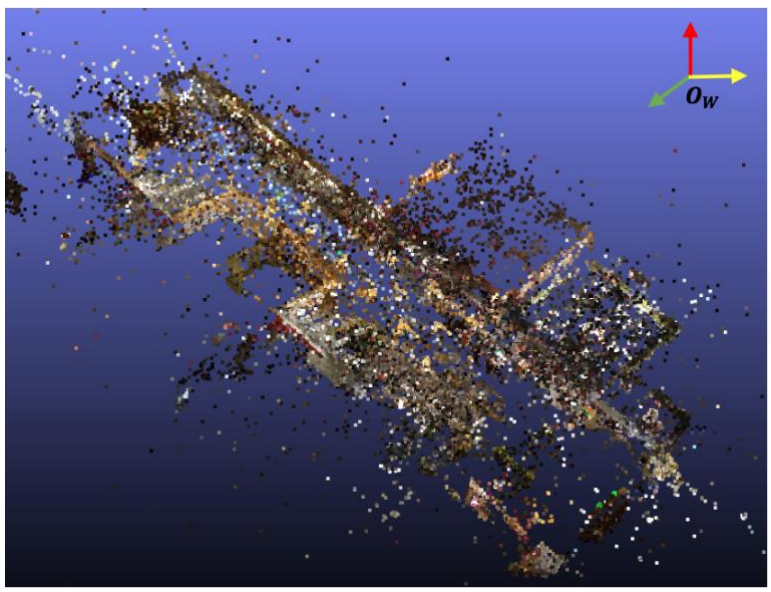


Figure 6.3: 3D reconstruction of UMass campus center

6.3.2.2 Visual Localization Algorithm

The pipeline of visual localization is shown in Figure 6.4. Starting from the captured first-person videos with inevitable motion blur, the feature selection module will select the sharpest frame in every 30 frames when the video is shooting at 30fps. The sharpness is measured using Variance of Laplacian (VoL) [130]. The higher the value of VoL of the frame is, the sharper the image is supposed to be. After the ideal query frame is found, the features on the frame will be extracted by the algorithm called scale-invariant

feature transform (SIFT). Compared to those obtained from reference images and registered in the 3D world coordinate system *W*, the best matches for the extracted features can be found by measuring the Euclidean distance between the two SIFT descriptors. After checking the epipolar constraints using the fundamental matrix, the accuracy of the 2D-to-2D matches can be improved by removing the false putative matches. Afterward, the 2D-to-3D correspondences can be found easily from the visual database. To estimate the pose of the camera when the 2D-to-3D correspondences are available (see Figure 6.5), most methods use random sampling and consensus (RANSAC) [86] to solve the Prospective-n-Point (PnP) problem, which randomly selects the smallest necessary subset of the putative 2D-to-3D correspondences and finds the best geometrical transformation to match the correspondences; the transformation found is then evaluated on all remaining data, selecting the best overall transformation over a fixed number of random draws. The process is repeated until sufficient agreement is observed between different trials or, alternatively, until the number of 2D-to-3D correspondences that agree with the transformation (known as the inlier set) is sufficiently large. The entire computation involves two coordinate systems (see Figure 6.5), one is the camera coordinate system, *V*, and another one is the world coordinate system, *W*. And within these two coordinate systems, camera pose is described using R_w^v and t_w^v, where R_w^v and t_w^v refer to the rotation matrix and translation vector from the world coordinate system *W* to camera coordinate system *V*, respectively. To compute the location and orientation estimates for the fusion framework, we first compute the orientation vector *D* of the camera (i.e. smartphone) with respect to the world coordinate system *W* using (6.6).

$$D = (R_w^v)^T \cdot [0 \;\; 0 \;\; 1]^T \tag{6.6}$$

where $[0\ \ 0\ \ 1]^T$ refers to the unit vector along z axis in camera coordinate system. Then the heading direction θ_w can be derived from 3x1 orientation vector $[D_x, D_y, D_z]$ using (6.7).

$$\theta_w = \tan^{-1}({}^{D_z}/_{D_x}) \tag{6.7}$$

And the location estimate in world coordinate system W, $[x_w, z_w]$ can be computed as follows.

$$[x_w, z_w]^T = -(R_w^v)^T \cdot t_w^v \tag{6.8}$$

More details about the visual localization can be found in [51, 52].

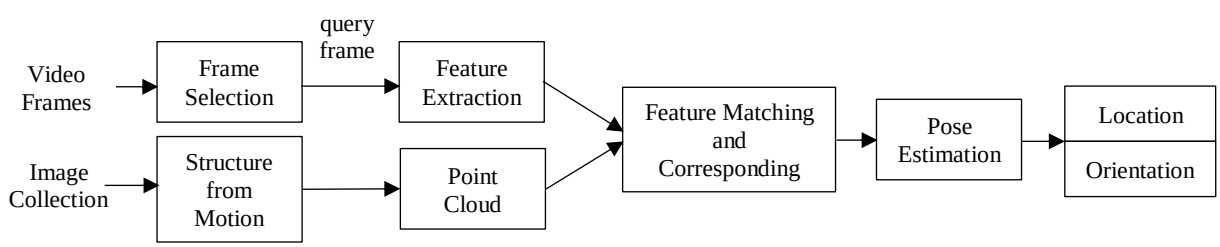


Figure 6.4: Algorithm flow of visual localization

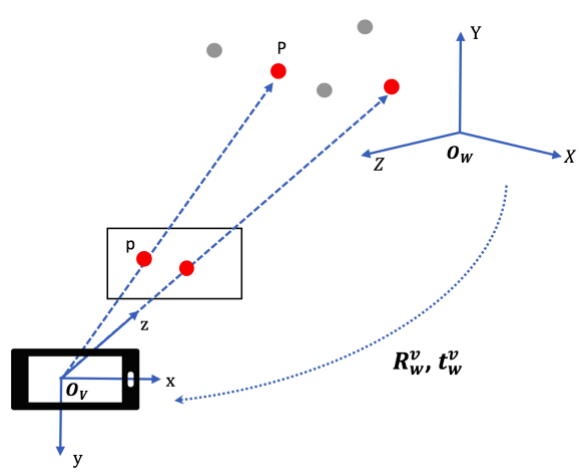


Figure 6.5: Camera pose estimation with 2D-3D correspondences

6.3.3 Pedestrian Dead Reckoning

Traditionally, PDR approaches [131] can be developed based on two principles. On the most straightforward principle, the displacement can be calculated from the integral of

acceleration and angular velocity measured by the IMU. However, the bias and noise from the consumer-grade IMU make it drift drastically in several seconds. To overcome this challenge, researchers try to solve the problem in a different way. Considering the fact that the pedestrian is walking, the steps can be detected by recognizing the oscillatory pattern in the signal of vertical acceleration. With the detected steps, a fixed step length value will be applied to each user for estimating the displacement.

Despite the approach using step information increases the robustness of PDR prominently, it still has several shortcomings. First, the built-in IMU on smartphone is quite noisy, specifically, when pedestrians are walking and holding it freely. As a result, the false detected steps can become a problem for the PDR [126]. Second, such methods are time-consuming on sensor calibration and manual settings of parameters, such as the threshold of the acceleration magnitude. Third, a fixed step length value or a universal model for all pedestrian cannot estimate the step length of a specific pedestrian precisely since the height or the gait varies from one person to another. To solve these issues, we propose to leverage the user-specific regression model that can predict pedestrian moving velocity using inertial data collected on their smartphones.

During data acquisition, we pre-marked several control points on the floor plan and then recovered the timestamps when data collectors passed by them in the captured videos. Assuming that data collectors walk at constant speeds in each segment, the velocities between control points can be calculated given the distances and travelling time. In total, 554 labelled data are collected by iPhone6 and 536 labelled data are collected by iPhone6s. After splitting the dataset into the training and the testing dataset, 70% and 30% of the data are used to train and test the regression model for two testers, respectively.

As shown in Figure 6.6, IMU can provide 3-axes acceleration, a, and 3-axes angular velocity, ω with respect to the inertial coordinate system, I. To reduce the noise, we exploit a sliding time window to smooth the readings along each axis. As an example, the acceleration along x-axis is smoothed using (6.9). And all other readings along different axes are smoothed in the same way. Since the default frequency of reading from IMU is 5Hz, the size of the time window is set to be 1s.

$$\bar{a}_x = \frac{\sum_{i=k}^{k+4} a_x^i}{5} \tag{6.9}$$

where $\bar{a}_x$ is smoothed acceleration along x-axis, a_x^i is the i th acceleration instance along x-axis within the sliding time window.

Given the smoothed readings over last 1 second, $[\bar{a}_x, \bar{a}_y, \bar{a}_z, \bar{\omega}_x, \bar{\omega}_y, \bar{\omega}_z]$, and the labelled velocity, Supported Vector Regression [127] can train the model for each tester and the results are listed in Table 6.2.

Table 6.2: Model evaluation

Training Data	**MSE (meter)**	**MAE (meter)**
iPhone 6	0.10	0.08
iPhone 6s	0.13	0.09

Note that MSE, MAE refers to mean squared error and mean absolute error, respectively.

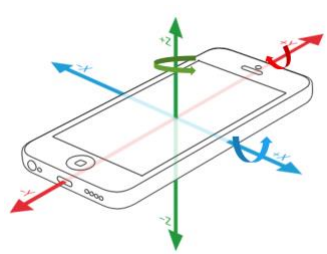


Figure 6.6: Inertial coordinate system used by iPhone

According to the Apple developer manual [132], a software motion sensor, i.e. attitude sensor, can provide real-time attitude data with respect to the reference coordinate

system, *ref*, which is the inertial coordinate system used when the attitude sensor is initiated. The device attitude with respect to this reference coordinate system *ref*, can be represented in many formats, such as rotation matrix, quaternion and Euler Angles. And we use the rotation matrix, R_{ref}^{I}, in our PDR. Given the $R_{v\_k-1}^{w}$, the rotation matrix from camera coordinate system *V* at time *k-1* to world coordinate system *W* and the relative attitude information, we can register the camera coordinate system *V* along with inertial coordinate system *I* at time *k* in the world coordinate system *W* using (6.10) and (6.11).

$$R_{v\_k}^{w} = R_{v\_k-1}^{w} * R_{v\_k-1}^{v\_k} \tag{6.10}$$

$$R_{I\_k-1}^{I\_k} = R_{ref}^{I\_k-1} * (R_{ref}^{I\_k})^{T} \tag{6.11}$$

where $R_{v\_k-1}^{v\_k}$ refers to the rotation matrix of camera coordinate system from time k-1 to k, and it is equivalent to $R_{I\_k-1}^{I\_k}$, the rotation matrix of inertial coordinate system from time k-1 to k because the transformation between camera coordinate system *V* and inertial coordinate system *I* at time *k* is fixed. (See Figure 6.7) Consequently, the heading direction at time *k*, θ_k^w can be extracted from $R_{v\_k}^{w}$ using the same way presented in (6.6) and (6.7).

With the predicted velocity at time k, v_{k-1}^{w}, the location estimates of the pedestrian referring to the world coordination system can be calculated by (6.12).

$$\begin{bmatrix} x_w \\ z_w \end{bmatrix}_k = \begin{bmatrix} x_w \\ z_w \end{bmatrix}_{k-1} + v_{k-1}^{w} \begin{bmatrix} \cos\theta_{k-1}^{w} \\ \sin\theta_{k-1}^{w} \end{bmatrix} \tag{6.12}$$

where $[x_w, z_w]_k^T$ refers to the 2D location estimate in world coordinate system at time *k*, v_{k-1}^{w} denotes the predicted velocity of pedestrian at time *k-1* with respect to the world coordinate system, θ_{k-1}^{w} is the heading direction referring to the world coordinate system, dt is the length of time interval.

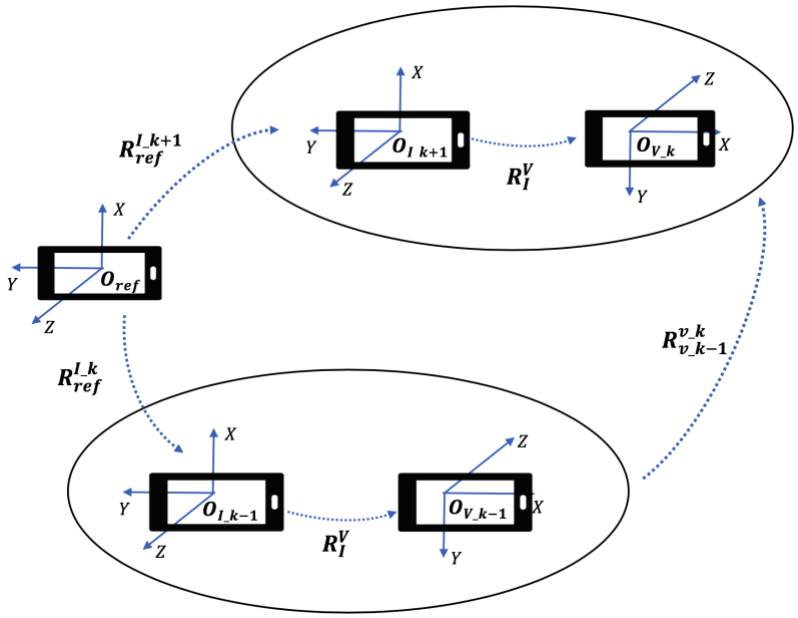


Figure 6.7: Coordinate system alignment

6.3.4 Sensor Fusion

To tackle the nonlinearity of the system, we exploit Unscented Kalman Filter (UKF) [12] to fuse multi-modal data and complimentary techniques discussed in Section 6.3.1, 6.3.2 and 6.3.3.

The state vector X of the dynamic system is defined as

$$X = [x_w, z_w, \theta_w]^T \tag{6.13}$$

Based on the PDR approach, we formulate the system dynamics as

$$X_k = X_{k-1} + u_{k-1} + w_{k-1} \tag{6.14}$$

where

$$u_{k-1} = [v_{k-1}^w \cdot \cos\theta_{k-1}^w, v_{k-1}^w \cdot \sin\theta_{k-1}^w, \alpha_{k-1}]^T$$

and w_k refers to a 3-dimensional system process noise, α_{k-1} refers to the variation of the heading direction at time *k-1* and it can be derived using (6.10) and (6.11).

Since the proposed algorithm will use two types of observations from two modalities, the WPL and the visual localization, it has to deal with the situation in which

the observations are reported in a different format. To address the problem, we can adjust the observation model based on the type of the observation received. When wireless localization algorithm generates a location estimate, the observation model will be set to (6.15) and corresponding measurement noise v_k will be updated as well.

$$Z_k = [x_w, z_w]_k^T + v_k \tag{6.15}$$

where v_k is a 2-dimensional system measurement noise.

When the visual localization yields a location estimate along with an orientation estimate, the observation model will be adjusted to (6.16) and v will become a 3-dimensional observation noise.

$$Z_k = [x_w, z_w, \theta_w]_k^T + v_k \tag{6.16}$$

where v_k is a 3-dimensional system measurement noise.

The UKF will recursively estimate the state of the system in two steps, including prediction step using the process model and update step using the observation model.

6.4 Performance Evaluation

6.4.1 Experiment Setup

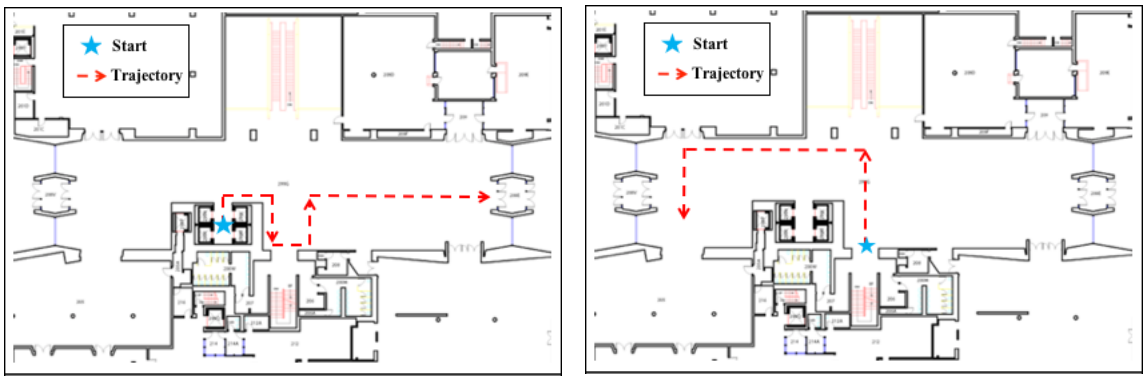
As shown in Figure 6.2(b) and Figure 6.3, we conduct the experiments at the concourse level of the Campus Center, University of Massachusetts, Amherst. This is a large and open public indoor space which works as a hub of student life and full-service convention center. It is a harsh environment for any technology that relies on wireless or electromagnetic signal since it is built using reinforced concrete. The entire testing area used in the experiment is 55 m x 15 m. There are 26 BLE sensors installed in the space. These BLE sensors are manufactured by Kontack.io [111] and the hardware specifications

are described in Table 6.3. To receive the signal from the BLE sensors, an iPhone 6 and an iPhone 6s were carried by two testers during data collection, respectively.

Table 6.3: Hardware specification of BLE sensor

Model	Operating frequency	Transmission Power	Transmission range	Time interval	Battery life
Tough Beacon	2.4 GHz	-30 dBm to 4 dBm	Up to 100 ft.	350 ms	Up to 2 years with default setting

To make the testing scenarios reflect reality, we collected the testing data during the busy time of the day. When the data was collected, many people were walking, standing or doing other things in the space, making the testing scenarios much more challenging compared to an empty laboratory space. With existing crowd, a significant portion of scene-related visual features are covered, and the light-of-sight signal propagation is obstructed. Under such circumstance, testers collected 4 groups of data following 2 trajectories illustrated in Figure 6.8. across different regions of the space. When they were walking along the trajectories, they were holding the smartphones facing front as normal. Given the recorded videos, the ground truth trajectories can be found easily.



(a) (b)

Figure 6.8: Ground truth path of trajectories
(a) Trajectory I (b) Trajectory II

6.4.2 Experimental Results

The sensor fusion algorithm is developed and evaluated on a PC given the multi-modal dataset. Unlike other papers [116-124] assume that an accurate initial state of the location and orientation is given, the proposed algorithm can work properly without relying on it since the initial state can be acquired by the visual localization module included in the fusion framework. This is a significant improvement proposed by our approach compared to others [116-124]. Although the visual localization included in our framework can run independently without any human intervention, the visual localization algorithm is essentially an opportunistic approach due to the practical reasons discussed in Section 6.1. Therefore, we decide to compare the localization performance between two practical algorithms, the proposed algorithm and the WPL algorithm, since both of them can run independently and continuously without given an accurate initial state. Figure 6.9 depicts the trajectories of the ground truth path and the estimated path using the proposed algorithm and the WPL algorithm.

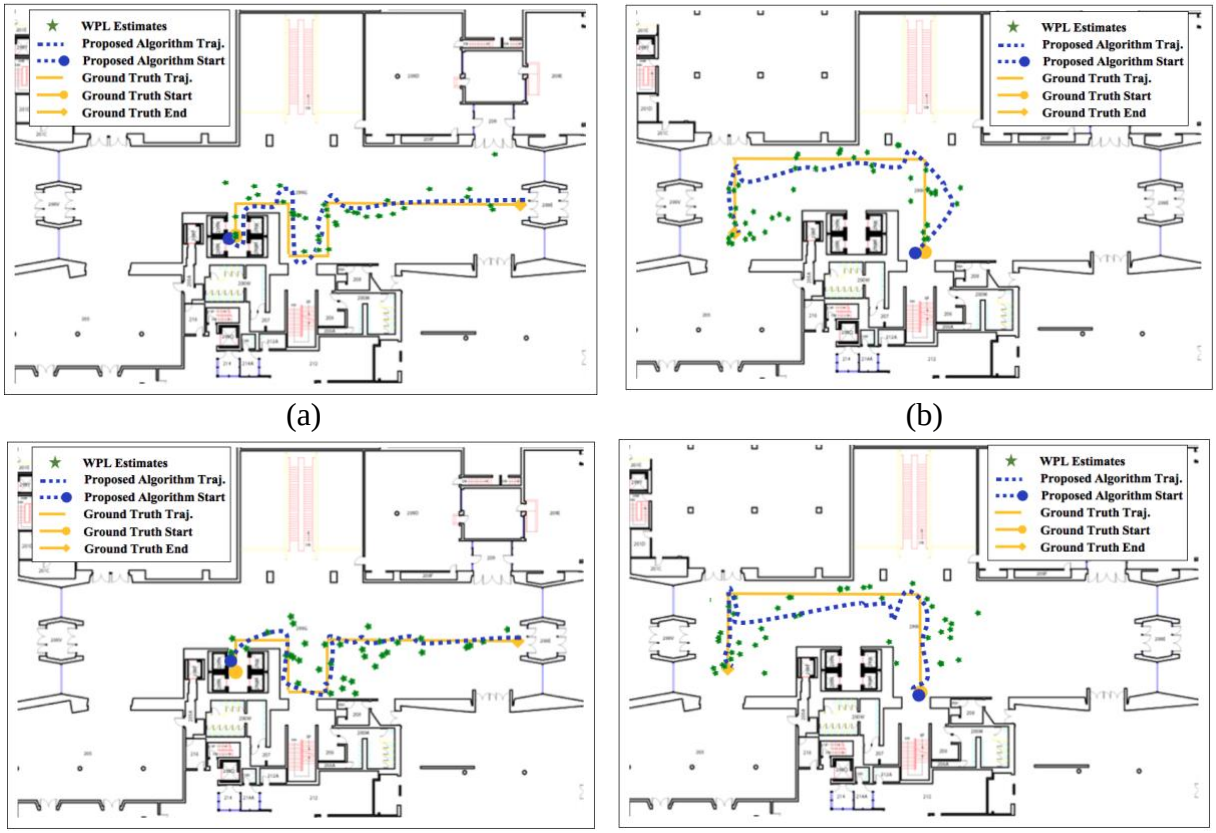


Figure 6.9: Estimated trajectories
(a) Data 1 (b) Data 2
(c) Data 3 (d) Data 4

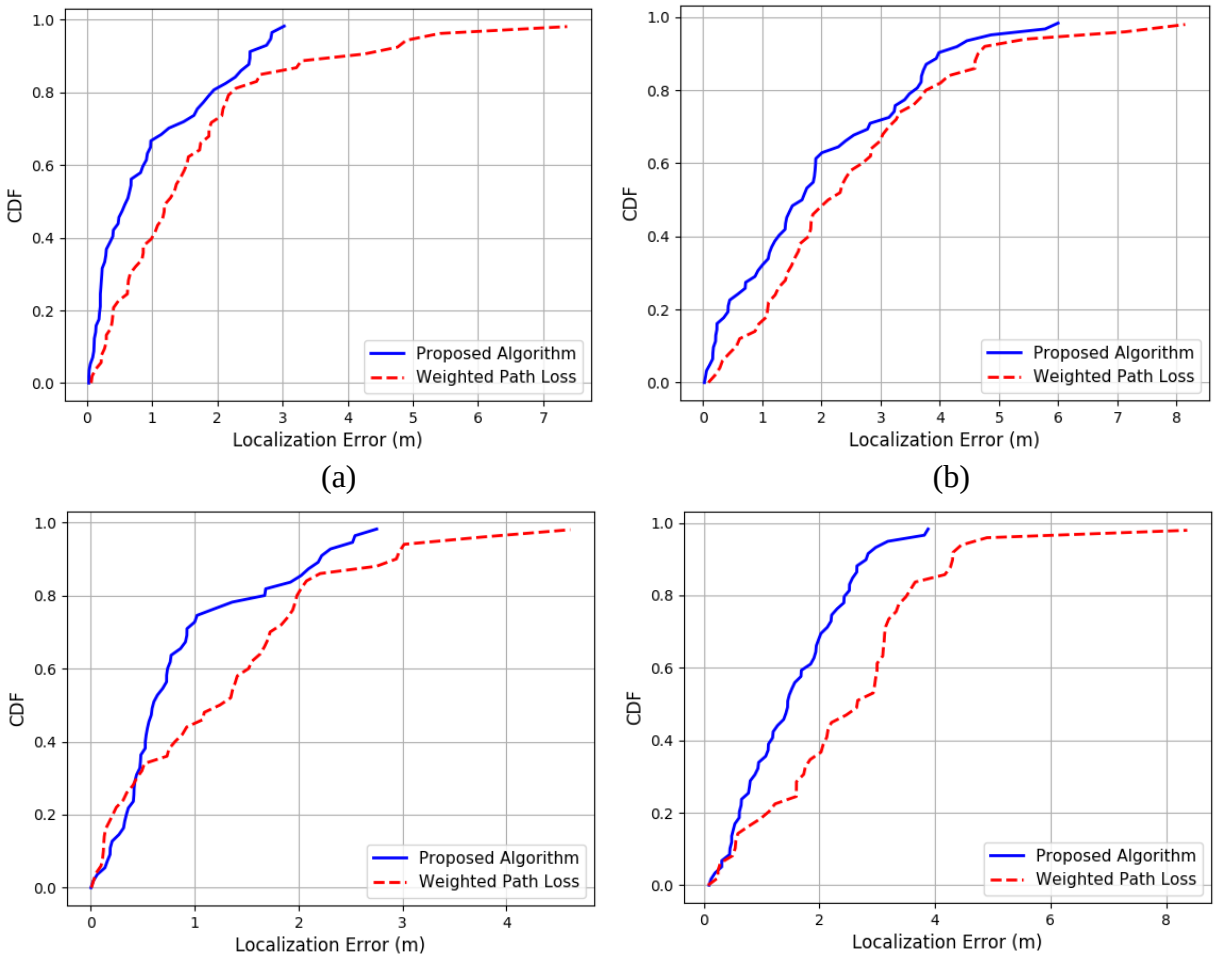


Figure 6.10: CDF of localization errors
(a) Data 1 (b) Data 2
(c) Data 3 (d) Data 4

Table 6.4: Summary of localization errors

Data	**Method**	**50%**	**80%**	**Mean**	**Localization Error Reduction**
1	Proposed Algorithm	0.58m	1.95m	0.94m	41%
	Weighted Path Loss	1.28m	2.17m	1.59m	
2	Proposed Algorithm	1.67m	3.49m	1.92m	22%
	Weighted Path Loss	2.11m	3.76m	2.46m	
3	Proposed Algorithm	0.59m	1.67m	0.87m	31%
	Weighted Path Loss	1.24m	1.98m	1.26m	
4	Proposed Algorithm	1.45m	2.43m	1.52m	39%
	Weighted Path Loss	2.66m	3.50m	2.48m	
Total	Proposed Algorithm	1.07m	2.36m	1.31m	33%
	Weighted Path Loss	1.82m	2.85m	1.95m	

6.5 Conclusion

We propose a novel sensor fusion algorithm for pedestrian indoor localization using multi-modal data and complementary localization techniques. The sensor fusion framework is comprised of a BLE-based localization algorithm, WPL, a PDR with user-specific modelling algorithm (SVR) and a visual localization algorithm that can not only provide the initial state for the system but also determine the location and orientation estimates opportunistically during the walk. We use the Unscented Kalman Filter to integrate all these techniques. From the results achieved in these challenging scenarios, the mean localization errors over all testing data are 1.31m and 1.95m for the proposed algorithm and weighted path loss algorithm, respectively. Using the proposed algorithm, 33% of the localization errors are reduced in contrast to the weighted path loss algorithm.

CHAPTER 7

FUTURE WORK

7.1 Outdoor Patient Tracking in Disaster Management

In this dissertation, we have proposed an RFID-based tracking system and a BLE-based tracking system to help the responders and Incident Commanders (IC) to track the patient locations in real-time without any infrastructure. But our system works relying on the movement of the responders in the disaster scene and it makes it an opportunistic approach.

To improve upon it, we could adopt the drone to strengthen our system. These days, a drone is usually equipped with GPS, IMU, and high-resolution camera, providing a lot of opportunities to disaster management, such as visual imaging, 3D mapping and accessing the damage. Using the motion planning algorithm, drones are expected to cover the entire disaster area with less time and difficulty than the responders. Furthermore, given the real-time visual input from the camera, we could develop and deploy some computer vision algorithms to identify and track the patient continuously.

7.2 Pedestrian Indoor Localization

In the dissertation, we have presented a series of indoor localization techniques varying from BLE-based localization to hybrid tracking. Even though we have reduced the efforts in preparation phase substantially compared to other works, it is still a requisite step to spend some time on acquiring the scene-related knowledge. Other techniques like Simultaneous Localization and Tracking (SLAM) or Deep Learning, will be worth to work on since they have potential to further reduce the efforts involved in the offline preparation

phase and empower indoor wayfinding system with the ability to learn more about the navigation spaces. We think there is a connection between the techniques used in autonomous driving and pedestrian indoor localization. And we believe more improvements can be made if we can transfer the knowledge in autonomous driving.

7.3 Indoor Wayfinding Assistance for BVI

Our indoor localization and tracking algorithms are used to provide location and orientation estimates for the indoor wayfinding assistance for BVI, like what has been done for PERCEPT. However, with computer vision techniques, we think we can even provide more navigational assistance to BVI users, such as describing the objects in front of the user or recognizing the face/expression of whom the user knows.

With the advent of the emerging deep learning techniques, many assistances can be provided audibly to BVI users since the image or the video content can be understood contextually. Such techniques will be able to let them hear the world through the camera. And we believe that more techniques developed using such techniques will benefit the BVI users and facilitate their daily lives in many ways.

BIBLIOGRAPHY

[1] En.wikipedia.org. (2017). *Emergency management*. [online] Available at: https://en.wikipedia.org/wiki/Emergency_management [Accessed 3 Jul. 2017].

[2] A. Ganz *et al*, "Real-time scalable resource tracking framework (DIORAMA): System description and experimentation," in *Technologies for Homeland Security (HST), 2011 IEEE International Conference on*, 2011, pp. 407-412.

[3] A. Ganz *et al*, "Mobile DIORAMA-II: Infrastructure less information collection system for mass casualty incidents," in *Engineering in Medicine and Biology Society (EMBC), 2014 36th Annual International Conference of the IEEE*, 2014, pp. 2682-2685.

[4] L. A. Lenert *et al*, "Design and evaluation of a wireless electronic health records system for field care in mass casualty settings," *Journal of the American Medical Informatics Association*, vol. 18, *(6)*, pp. 842-852, 2011.

[5] T. J. Morris *et al*, "Battlefield medical information system–tactical (BMIST): the application of mobile computing technologies to support health surveillance in the Department of Defense," *Telemedicine Journal & e-Health*, vol. 12, *(4)*, pp. 409-416, 2006.

[6] Anon, (2017). [online] Available at: http://file:///Users/zhuoruiyang/Downloads/ADA477535.pdf [Accessed 3 Jul. 2017].

[7] D. Malan *et al*, "Codeblue: An ad hoc sensor network infrastructure for emergency medical care," in *International Workshop on Wearable and Implantable Body Sensor Networks*, 2004.

[8] Jhuapl.edu. (2017). *Advanced Health and Disaster Aid Network at JHU APL (AID-N)*. [online] Available at: http://www.jhuapl.edu/AID-N/ [Accessed 3 Jul. 2017].

[9] G. R. Ciottone *et al*, *Ciottone's Disaster Medicine.* Elsevier Health Sciences, 2015.

[10] Jhuapl.edu. (2017). *Advanced Health and Disaster Aid Network at JHU APL (AID-N)*. [online] Available at: http://www.jhuapl.edu/AID-N/ [Accessed 3 Jul. 2017].

[11] H. A. Karimi, *Indoor Wayfinding and Navigation.* CRC Press, 2015.

[12] E. A. Wan and R. Van Der Merwe, "The unscented kalman filter for nonlinear estimation," in *Adaptive Systems for Signal Processing, Communications, and Control Symposium 2000. AS-SPCC. the IEEE 2000*, 2000, pp. 153-158.

[13] A. Kendall, M. Grimes and R. Cipolla, "Posenet: A convolutional network for real-time 6-dof camera relocalization," in *Proceedings of the IEEE International Conference on Computer Vision*, 2015, pp. 2938-2946.

[14] T. Gao and D. White, "A next generation electronic triage to aid mass casualty emergency medical response," in *Engineering in Medicine and Biology Society, 2006. EMBS'06. 28th Annual International Conference of the IEEE*, 2006, pp. 6501-6504.

[15] T. Gao *et al*, "Vital signs monitoring and patient tracking over a wireless network," in *Engineering in Medicine and Biology Society, 2005. IEEE-EMBS 2005. 27th Annual International Conference of the*, 2006, pp. 102-105.

[16] H. Makino *et al*, "Web-GIS-based emergency rescue to track triage information—System configuration and experimental results," in *Ubiquitous Positioning, Indoor Navigation, and Location Based Service (UPINLBS), 2012*, 2012, pp. 1-4.

[17] M. S. Hossain, "Cloud-supported cyber–physical localization framework for patients monitoring," *IEEE Systems Journal*, 2015.

[18] A. Chandra-Sekaran *et al*, "ZigBee sensor network for patient localization and air temperature monitoring during emergency response to crisis," in *Sensor Technologies and Applications, 2008. SENSORCOMM'08. Second International Conference on*, 2008, pp. 233-238.

[19] A. Chandra-Sekaran *et al*, "Empirical analysis and ranging using environment and mobility adaptive RSSI filter for patient localization during disaster management," in *Networking and Services, 2009. ICNS'09. Fifth International Conference on*, 2009, pp. 276-281.

[20] A. Chandra-Sekaran *et al*, "A comparison of bayesian filter based approaches for patient localization during emergency response to crisis," in *Sensor Technologies and Applications, 2009. SENSORCOMM'09. Third International Conference on*, 2009, pp. 636-642.

[21] L. Vojtech *et al*, "Outdoor localization technique using active RFID technology aimed for security and disaster management applications," in *Carpathian Control Conference (ICCC), 2015 16th International*, 2015, pp. 586-589.

[22] D. T. Vemula, X. Yu and A. Ganz, "Real time localization of victims at an emergency site: Architecture, algorithms and experimentation," in *Engineering in Medicine and Biology Society, 2009. EMBC 2009. Annual International Conference of the IEEE*, 2009, pp. 1703-1706.

[23] X. Yu and A. Ganz, "Scalable patients tracking framework for mass casualty incidents," in *Engineering in Medicine and Biology Society, EMBC, 2011 Annual International Conference of the IEEE*, 2011, pp. 860-863.

[24] Z. Yang and A. Ganz, "Infrastructure-less and calibration-free RFID-based localization algorithm for victim tracking in mass casualty incidents," in *Technologies for Homeland Security (HST), 2016 IEEE Symposium on*, 2016, pp. 1-4.

[25] Z. Yang, J. Schafer and A. Ganz, "Disaster response: Victims' localization using bluetooth low energy sensors," in *Technologies for Homeland Security (HST), 2017 IEEE International Symposium on*, 2017, pp. 1-4.

[26] J. Hunag, R. Harwahyu and R. Cheng, "Study of low cost mobile phone tracking system," in *Next-Generation Electronics (ISNE), 2015 International Symposium on*, 2015, pp. 1-4.

[27] S. Zorn *et al*, "A novel technique for mobile phone localization for search and rescue applications," in *Indoor Positioning and Indoor Navigation (IPIN), 2010 International Conference on*, 2010, pp. 1-4.

[28] A. Hossain, S. K. Ray and R. Sinha, "A smartphone-assisted post-disaster victim localization method," in *High Performance Computing and Communications; IEEE 14th International Conference on Smart City; IEEE 2nd International Conference on Data Science and Systems (HPCC/SmartCity/DSS), 2016 IEEE 18th International Conference on,* 2016, pp. 1173-1179.

[29] L. Wan *et al*, "The critical patients localization algorithm using sparse representation for mixed signals in emergency healthcare system," *IEEE Systems Journal,* 2015.

[30] B. Soni and A. Sowmya, "Victim detection and localisation in an urban disaster site," in *Robotics and Biomimetics (ROBIO), 2013 IEEE International Conference on,* 2013, pp. 2142-2147.

[31] Z. Zhang *et al*, "Finding disaster victims: A sensory system for robot-assisted 3D mapping of urban search and rescue environments," in *Robotics and Automation, 2007 IEEE International Conference on,* 2007, pp. 3889-3894.

[32] E. Di Giampaolo, "A passive-RFID based indoor navigation system for visually impaired people," in *Applied Sciences in Biomedical and Communication Technologies (ISABEL), 2010 3rd International Symposium on,* 2010, pp. 1-5.

[33] B. Amutha and K. Nanmaran, "Development of a ZigBee based virtual eye for visually impaired persons," in *Indoor Positioning and Indoor Navigation (IPIN), 2014 International Conference on,* 2014, pp. 564-574.

[34] K. Liu, X. Liu and X. Li, "Guoguo: Enabling fine-grained smartphone localization via acoustic anchors," *IEEE Transactions on Mobile Computing,* vol. 15, *(5),* pp. 1144-1156, 2016.

[35] A. W. S. Au *et al*, "Indoor tracking and navigation using received signal strength and compressive sensing on a mobile device," *IEEE Transactions on Mobile Computing,* vol. 12, *(10),* pp. 2050-2062, 2013.

[36] A. Ganz *et al*, "INSIGHT: RFID and bluetooth enabled automated space for the blind and visually impaired," in *Engineering in Medicine and Biology Society (EMBC), 2010 Annual International Conference of the IEEE,* 2010, pp. 331-334.

[37] A. Ganz *et al*, "PERCEPT: Indoor navigation for the blind and visually impaired," in *Engineering in Medicine and Biology Society, EMBC, 2011 Annual International Conference of the IEEE,* 2011, pp. 856-859.

[38] A. Ganz *et al*, "PERCEPT-II: Smartphone based indoor navigation system for the blind," in *Engineering in Medicine and Biology Society (EMBC), 2014 36th Annual International Conference of the IEEE,* 2014, pp. 3662-3665.

[39] S. Basso, G. Frigo and G. Giorgi, "A smartphone-based indoor localization system for visually impaired people," in *Medical Measurements and Applications (MeMeA), 2015 IEEE International Symposium on,* 2015, pp. 543-548.

[40] T. H. Riehle *et al*, "Indoor inertial waypoint navigation for the blind," in *Engineering in Medicine and Biology Society (EMBC), 2013 35th Annual International Conference of the IEEE,* 2013, pp. 5187-5190.

[41] V. Prudtipongpun *et al*, "Indoor navigation system for vision-impaired individual: An application on android devices," in *Signal-Image Technology & Internet-Based Systems (SITIS), 2015 11th International Conference on,* 2015, pp. 633-638.

[42] C. Ye, S. Hong and A. Tamjidi, "6-DOF pose estimation of a robotic navigation aid by tracking visual and geometric features," *IEEE Transactions on Automation Science and Engineering,* vol. 12, *(4),* pp. 1169-1180, 2015.

[43] H. Zhang and C. Ye, "An indoor navigation aid for the visually impaired," in *Robotics and Biomimetics (ROBIO), 2016 IEEE International Conference on,* 2016, pp. 467-472.

[44] A. Zeb, S. Ullah and I. Rabbi, "Indoor vision-based auditory assistance for blind people in semi controlled environments," in *Image Processing Theory, Tools and Applications (IPTA), 2014 4th International Conference on,* 2014, pp. 1-6.

[45] C. Manlises *et al*, "Indoor navigation system based on computer vision using CAMShift and D* algorithm for visually impaired," in *Control System, Computing and Engineering (ICCSCE), 2016 6th IEEE International Conference on,* 2016, pp. 481-484.

[46] S. Wang and Y. Tian, "Indoor signage detection based on saliency map and bipartite graph matching," in *Bioinformatics and Biomedicine Workshops (BIBMW), 2011 IEEE International Conference on,* 2011, pp. 518-525.

[47] Y. Bai *et al*, "Landmark-based indoor positioning for visually impaired individuals," in *Signal Processing (ICSP), 2014 12th International Conference on,* 2014, pp. 668-671.

[48] M. Vlaminck *et al*, "Indoor assistance for visually impaired people using a RGB-D camera," in *Image Analysis and Interpretation (SSIAI), 2016 IEEE Southwest Symposium on,* 2016, pp. 161-164.

[49] W. Elloumi *et al*, "Indoor navigation assistance with a smartphone camera based on vanishing points," in *Indoor Positioning and Indoor Navigation (IPIN), 2013 International Conference on,* 2013, pp. 1-9.

[50] B. Taylor *et al*, "Smart phone-based indoor guidance system for the visually impaired," in *Control Automation Robotics & Vision (ICARCV), 2012 12th International Conference on,* 2012, pp. 871-876.

[51] Z. Yang and A. Ganz, "Egocentric Landmark-Based Indoor Guidance System for the Visually Impaired," *International Journal of E-Health and Medical Communications (IJEHMC),* vol. 8, *(3),* pp. 55-69, 2017.

[52] Z. Yang, M. Duarte, A. Ganz, “A Novel Crowd-Resilient Visual Localization Algorithm via Robust PCA Background Extraction,” in *Proceeding of IEEE International Conference on Acoustics, Speech, and Signal Processing (ICASSP),* 2018, pp. 1922-1926.

[53] S. A. Cheraghi, V. Namboodiri and L. Walker, "GuideBeacon: Beacon-based indoor wayfinding for the blind, visually impaired, and disoriented," in *Pervasive Computing and Communications (PerCom), 2017 IEEE International Conference on,* 2017, pp. 121-130.

[54] T. Gallagher *et al*, "Indoor positioning system based on sensor fusion for the blind and visually impaired," in *Indoor Positioning and Indoor Navigation (IPIN), 2012 International Conference on,* 2012, pp. 1-9

[55] A. J. Fukasawa and K. Magatani, "A navigation system for the visually impaired an intelligent white cane," in *Engineering in Medicine and Biology Society (EMBC), 2012 Annual International Conference of the IEEE,* 2012, pp. 4760-4763.

[56] T. Dao *et al*, "Indoor navigation assistance system for visually impaired people using multimodal technologies," in *Control, Automation, Robotics and Vision (ICARCV), 2016 14th International Conference on,* 2016, pp. 1-6.

[57] X. Zhang *et al*, "A slam based semantic indoor navigation system for visually impaired users," in *Systems, Man, and Cybernetics (SMC), 2015 IEEE International Conference on,* 2015, pp. 1458-1463.

[58] D. Croce *et al*, "ARIANNA: A smartphone-based navigation system with human in the loop," in *Control and Automation (MED), 2014 22nd Mediterranean Conference of,* 2014, pp. 8-13.

[59] S. L. Joseph *et al*, "Semantic indoor navigation with a blind-user oriented augmented reality," in *Systems, Man, and Cybernetics (SMC), 2013 IEEE International Conference on,* 2013, pp. 3585-3591.

[60] E. B. Kaiser and M. Lawo, "Wearable navigation system for the visually impaired and blind people," in *Computer and Information Science (ICIS), 2012 IEEE/ACIS 11th International Conference on,* 2012, pp. 230-233.

[61] X. Yu and A. Ganz, "Audible vision for the blind and visually impaired in indoor open spaces," in *Engineering in Medicine and Biology Society (EMBC), 2012 Annual International Conference of the IEEE,* 2012, pp. 5110-5113.

[62] K. Lee, M. Li and C. Lin, "Magnetic Tensor Sensor and Way-Finding Method Based on Geomagnetic Field Effects With Applications for Visually Impaired Users," *IEEE/ASME Transactions on Mechatronics,* vol. 21, *(6),* pp. 2694-2704, 2016.

[63] Y. Zhang *et al*, "Monocular camera and IMU integration for indoor position estimation," in *Engineering in Medicine and Biology Society (EMBC), 2014 36th Annual International Conference of the IEEE,* 2014, pp. 1198-1201.

[64] J. J. Liu, C. Phillips and K. Daniilidis, "Video-based localization without 3D mapping for the visually impaired," in *Computer Vision and Pattern Recognition Workshops (CVPRW), 2010 IEEE Computer Society Conference on,* 2010, pp. 23-30.

[65] W. Simões and V. de Lucena, "Hybrid indoor navigation as sistant for visually impaired people based on fusion of proximity method and pattern recognition algorithm," in *Consumer Electronics-Berlin (ICCE-Berlin), 2016 IEEE 6th International Conference on,* 2016, pp. 108-111.

[66] J. A. Hesch *et al*, "A laser-aided inertial navigation system (L-INS) for human localization in unknown indoor environments," in *Robotics and Automation (ICRA), 2010 IEEE International Conference on,* 2010, pp. 5376-5382.

[67] M. Sekiguchi *et al*, "Development of a wearable system for navigating the visually impaired in the indoor environment-a prototype system for fork detection and navigation," in *Robot and Human Interactive Communication, 2014 RO-MAN: The 23rd IEEE International Symposium on,* 2014, pp. 549-554.

[68] M. Nakajima and S. Haruyama, "Indoor navigation system for visually impaired people using visible light communication and compensated geomagnetic sensing," in *Communications in China (ICCC), 2012 1st IEEE International Conference on,* 2012, pp. 524-529.

[69] C. Chuang, J. Hsieh and K. Fan, "A smart handheld device navigation system based on detecting visual code," in *Machine Learning and Cybernetics (ICMLC), 2013 International Conference on,* 2013, pp. 1407-1412.

[70] Code, R. (2017). *Data Center Management Solutions | RF Code*. [online] Rfcode.com. Available at: http://www.rfcode.com/ [Accessed 3 Jul. 2017].

[71] En.wikipedia.org. (2017). *Simple triage and rapid treatment*. [online] Available at: https://en.wikipedia.org/wiki/Simple_triage_and_rapid_treatment [Accessed 3 Jul. 2017].

[72] Kontakt.io. (2017). *Your solution. Beacon-enabled. - Kontakt.io*. [online] Available at: https://kontakt.io/. [Accessed 3 Jul. 2017].

[73] A. Zamir and M. Shah, "Accurate image localization based on google maps street view," *Computer Vision–ECCV 2010,* pp. 255-268, 2010.

[74] H. Zou *et al*, "An RFID indoor positioning system by using weighted path loss and extreme learning machine," in *Cyber-Physical Systems, Networks, and Applications (CPSNA), 2013 IEEE 1st International Conference on,* 2013, pp. 66-71.

[75] K. Sakai *et al*, "A Framework for the Optimal $ k $-Coverage Deployment Patterns of Wireless Sensors," *IEEE Sensors Journal,* vol. 15, *(12),* pp. 7273-7283, 2015.

[76] F. Pedregosa *et al*, "Scikit-learn: Machine learning in Python," *Journal of Machine Learning Research,* vol. 12, *(Oct),* pp. 2825-2830, 2011.

[77] C. Hsu and C. Lin, "A comparison of methods for multiclass support vector machines," *IEEE Trans. Neural Networks,* vol. 13, *(2),* pp. 415-425, 2002.

[78] Cs.cornell.edu. (2017). *Bundler - Structure from Motion (SfM) for Unordered Image Collections*. [online] Available at: http://www.cs.cornell.edu/~snavely/bundler/ [Accessed 3 Jul. 2017].

[79] Ccwu.me. (2017). *VisualSFM: A Visual Structure from Motion System*. [online] Available at: http://ccwu.me/vsfm/ [Accessed 3 Jul. 2017].

[80] Z. Zhang, "A flexible new technique for camera calibration," *IEEE Trans. Pattern Anal. Mach. Intell.,* vol. 22, *(11),* pp. 1330-1334, 2000.

[81] Vision.caltech.edu. (2017). *Camera Calibration Toolbox for Matlab*. [online] Available at: http://www.vision.caltech.edu/bouguetj/calib_doc/ [Accessed 3 Jul. 2017].

[82] L. Juan and O. Gwun, "A comparison of sift, pca-sift and surf," *International Journal of Image Processing (IJIP),* vol. 3, *(4),* pp. 143-152, 2009.

[83] J. Wu *et al*, "A Comparative Study of SIFT and its Variants," *Measurement Science Review,* vol. 13, *(3),* pp. 122-131, 2013.

[84] D. G. Lowe, "Distinctive image features from scale-invariant keypoints," *International Journal of Computer Vision,* vol. 60, *(2),* pp. 91-110, 2004.

[85] M. Muja and D. G. Lowe, "Scalable nearest neighbor algorithms for high dimensional data," *IEEE Trans. Pattern Anal. Mach. Intell.,* vol. 36, *(11),* pp. 2227-2240, 2014.

[86] K. G. Derpanis, "Overview of the RANSAC Algorithm," *Image Rochester NY,* vol. 4, *(1),* pp. 2-3, 2010.

[87] V. Lepetit, F. Moreno-Noguer and P. Fua, "Epnp: An accurate o (n) solution to the pnp problem," *International Journal of Computer Vision,* vol. 81, *(2),* pp. 155-166, 2009.

[88] E. J. Candès *et al*, "Robust principal component analysis?" *Journal of the ACM (JACM),* vol. 58, *(3),* pp. 11, 2011.

[89] Z. Lin *et al*, "Fast convex optimization algorithms for exact recovery of a corrupted low-rank matrix," *Computational Advances in Multi-Sensor Adaptive Processing (CAMSAP),* vol. 61, *(6),* 2009.

[90] Ergo-Anthropometric Data. (2017) [online] Available at: https://www.bwc.ohio.gov/downloads/blankpdf/ErgoAnthropometricData.pdf [Accessed 3 Jul. 2017].

[91] L. Maddalena and A. Petrosino, "A fuzzy spatial coherence-based approach to background/foreground separation for moving object detection," *Neural Computing & Applications,* vol. 19, *(2),* pp. 179-186, 2010.

[92] A. Sobral and A. Vacavant, "A comprehensive review of background subtraction algorithms evaluated with synthetic and real videos," *Comput. Vision Image Understanding,* vol. 122, pp. 4-21, 2014.

[93] “Blindness Statistics from the American Foundation for the Blind - American Foundation for the Blind.”, Afb.org, [Online] Available at: http://www.afb.org/info/blindness-statistics/2 [Accessed: 28-Jul 2018].

[94] R. A. Epstein et al, "The cognitive map in humans: spatial navigation and beyond," Nat. Neurosci., vol. 20, (11), pp. 1504, 2017.

[95] M. B. Moser, D. C. Rowland and E. I. Moser, "Place cells, grid cells, and memory," Cold Spring Harb Perspect. Biol., vol. 7, (2), pp. a021808, Feb 2, 2015.

[96] J. O'Keefe, "Place units in the hippocampus of the freely moving rat," Exp. Neurol., vol. 51, (1), pp. 78-109, 1976.

[97] V. R. Schinazi, T. Thrash and D. Chebat, "Spatial navigation by congenitally blind individuals," Wiley Interdisciplinary Reviews: Cognitive Science, vol. 7, (1), pp. 37-58, 2016.

[98] J. P. Goodridge and J. S. Taube, "Preferential use of the landmark navigational system by head direction cells in rats." Behav. Neurosci., vol. 109, (1), pp. 49, 1995.

[99] J. S. Taube, "Head direction cells recorded in the anterior thalamic nuclei of freely moving rats," J. Neurosci., vol. 15, (1 Pt 1), pp. 70-86, Jan, 1995.

[100] J. S. Taube, R. U. Muller and J. B. Ranck Jr, "Head-direction cells recorded from the postsubiculum in freely moving rats. I. Description and quantitative analysis," J. Neurosci., vol. 10, (2), pp. 420-435, Feb, 1990.

[101] M. B. Zugaro, E. Tabuchi and S. I. Wiener, "Influence of conflicting visual, inertial and substratal cues on head direction cell activity," Experimental Brain Research, vol. 133, (2), pp. 198-208, 2000.

[102] "The 2014 Nobel Prize in Physiology or Medicine - Press Release," Nobelprize.org. [Online]. Available: https://www.nobelprize.org/nobel_prizes/medicine/laureates/2014/press.html. [Accessed: 10-Aug-2018].

[103] E. Save, L. Nerad and B. Poucet, "Contribution of multiple sensory information to place field stability in hippocampal place cells," Hippocampus, vol. 10, (1), pp. 64-76, 2000.

[104] S. Zhang et al, "Spatial representations of place cells in darkness are supported by path integration and border information," Frontiers in Behavioral Neuroscience, vol. 8, pp. 222, 2014.

[105] E. J. Markus et al, "Spatial information content and reliability of hippocampal CA1 neurons: effects of visual input," Hippocampus, vol. 4, (4), pp. 410-421, 1994.

[106] G. J. Quirk, R. U. Muller and J. L. Kubie, "The firing of hippocampal place cells in the dark depends on the rat's recent experience," J. Neurosci., vol. 10, (6), pp. 2008-2017, Jun, 1990.

[107] A. Ganz, J. Schafer, Y. Tao, Z, Yang, C. Sanderson, L. Haile, "PERCEPT Navigation for Visually Impaired in Large Transportation Hubs", Journal on Technology and Persons with Disabilities, 2018.

[108] Y. Tao, "Scalable and vision free user interface approaches for indoor navigation systems for the visually impaired," MSc thesis, 2015. https://scholarworks.umass.edu/masters_theses_2/172/

[109] J. Taube, "Head direction cells," Scholarpedia. [Online]. Available: http://www.scholarpedia.org/article/Head_direction_cells. [Accessed: 09-Aug-2018].

[110] W. Ku, K. Sakai and M. Sun, "The optimal k-covering tag deployment for RFID-based localization," Journal of Network and Computer Applications, vol. 34, (3), pp. 914-924, 2011.

[111] Beacon, "Tough Beacon", Store.kontakt.io, 2017. [Online]. Available: https://store.kontakt.io/our-products/28-tough-beacon.html. [Accessed: 20- Aug- 2017].

[112] Z. Chen, Q. Zhu and Y. C. Soh, "Smartphone inertial sensor-based indoor localization and tracking with iBeacon corrections," *IEEE Transactions on Industrial Informatics*, vol. 12, *(4)*, pp. 1540-1549, 2016.

[113] H. Xu *et al*, "Indoor localization via multi-modal sensing on smartphones," in *Proceedings of the 2016 ACM International Joint Conference on Pervasive and Ubiquitous Computing*, 2016, pp. 208-219.

[114] M. Liu *et al*, "Scene recognition for indoor localization using a multi-sensor fusion approach," *Sensors*, vol. 17, *(12)*, pp. 2847, 2017.

[115] Y. Zhuang *et al*, "Smartphone-based indoor localization with bluetooth low energy beacons," *Sensors*, vol. 16, *(5)*, pp. 596, 2016.

[116] Z. Chen *et al*, "Fusion of WiFi, smartphone sensors and landmarks using the Kalman filter for indoor localization," *Sensors*, vol. 15, *(1)*, pp. 715-732, 2015.

[117] G. Chen *et al*, "Integrated WiFi/PDR/Smartphone using an unscented kalman filter algorithm for 3D indoor localization," *Sensors*, vol. 15, *(9)*, pp. 24595-24614, 2015.

[118] H. Zou *et al*, "Accurate indoor localization and tracking using mobile phone inertial sensors, WiFi and iBeacon," in *Inertial Sensors and Systems (INERTIAL), 2017 IEEE International Symposium on*, 2017, pp. 1-4.

[119] Y. Zhuang and N. El-Sheimy, "Tightly-coupled integration of WiFi and MEMS sensors on handheld devices for indoor pedestrian navigation," *IEEE Sensors Journal*, vol. 16, *(1)*, pp. 224-234, 2016.

[120] S. He *et al*, "Calibration-free fusion of step counter and wireless fingerprints for indoor localization," in *Proceedings of the 2015 ACM International Joint Conference on Pervasive and Ubiquitous Computing*, 2015, pp. 897-908.

[121] Z. Wu *et al*, "Improved particle filter based on WLAN RSSI fingerprinting and smart sensors for indoor localization," *Comput. Commun.*, vol. 83, pp. 64-71, 2016.

[122] Y. Zhuang *et al*, "PDR/INS/WiFi integration based on handheld devices for indoor pedestrian navigation," *Micromachines*, vol. 6, *(6)*, pp. 793-812, 2015.

[123] X. Li *et al*, "Integrated WiFi/PDR/smartphone using an adaptive system noise extended Kalman filter algorithm for indoor localization," *ISPRS International Journal of Geo-Information*, vol. 5, *(2)*, pp. 8, 2016.

[124] F. Ebner *et al*, "Multi sensor 3D indoor localisation," in *Indoor Positioning and Indoor Navigation (IPIN), 2015 International Conference on*, 2015, pp. 1-11.

[125] M. Luna *et al*, "An indoor pedestrian positioning system based on inertial measurement unit and wireless local area network," in *Control Conference (CCC), 2015 34th Chinese*, 2015, pp. 5419-5424.

[126] R. Harle, "A survey of indoor inertial positioning systems for pedestrians." *IEEE Communications Surveys and Tutorials*, vol. 15, *(3)*, pp. 1281-1293, 2013.

[127] A. J. Smola and B. Schölkopf, "A tutorial on support vector regression," *Statistics and Computing*, vol. 14, *(3)*, pp. 199-222, 2004.

[128] J. L. Schonberger and J. Frahm, "Structure-from-motion revisited," in *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, 2016, pp. 4104-4113.

[129] "VisualSFM : A Visual Structure from Motion System," *VisualSFM : A Visual Structure from Motion System - Documentation*. [Online]. Available: http://ccwu.me/vsfm/. [Accessed: 14-Aug-2018].

[130] J. L. Pech-Pacheco *et al*, "Diatom autofocusing in brightfield microscopy: A comparative study," in *Pattern Recognition, 2000. Proceedings. 15th International Conference on*, 2000, pp. 314-317.

[131] P. D. Groves, "Navigation using inertial sensors [Tutorial]," *IEEE Aerosp. Electron. Syst. Mag.*, vol. 30, *(2)*, pp. 42-69, 2015.

[132] "Understanding Reference Frames and Device Attitude," *Purchase and Activation - Support - Apple Developer*. [Online]. Available: https://developer.apple.com/documentation/coremotion/getting_processed_device_motion_data/understanding_reference_frames_and_device_attitude. [Accessed: 14-Aug-2018].