2022

Detecting and Tracking Vulnerable Road Users' Trajectories Using Different Types of Sensors Fusion

Zhongchuan Wang
University of Central Florida

DETECTING AND TRACKING VULNERABLE ROAD USERS' TRAJECTORIES USING DIFFERENT TYPES OF SENSORS FUSION

by

ZHONGCHUAN WANG
B.S. University of Central Florida, 2021

A thesis submitted in partial fulfillment of the requirements
for the degree of Master of Science
in the Department of Civil, Environmental, and Construction Engineering
in the College of Engineering and Computer Science
at the University of Central Florida
Orlando, Florida

Fall Term
2022

Major Professor: Mohamed Abdel-Aty

ABSTRACT

Vulnerable road user (VRU) detection and tracking has been a key challenge in transportation research. Different types of sensors such as the camera, LiDAR, and inertial measurement units (IMUs) have been used for this purpose. For detection and tracking with the camera, it is necessary to perform calibration to obtain correct GPS trajectories. This method is often tedious and necessitates accurate ground truth data. Moreover, if the camera performs any pan-tilt-zoom function, it is usually necessary to recalibrate the camera. In this thesis, we propose camera calibration using an auxiliary sensor: ultra-wideband (UWB). UWBs are capable of tracking a road user with ten centimeter-level accuracy. Once a VRU with a UWB traverses in the camera view, the UWB GPS data is fused with the camera to perform real-time calibration. As the experimental results in this thesis have shown, the camera is able to output better trajectories after calibration. It is expected that the use of UWB is needed only once to fuse the data and determine the correct trajectories at the same intersection and location of the camera. All other trajectories collected by the camera can be corrected using the same adjustment. In addition, data analysis was conducted to evaluate the performance of the UWB sensors. This study also predicted pedestrian trajectories using data fused by the UWB and smartphone sensors. UWB GPS coordinates are very accurate although it lacks the other sensor parameters such as accelerometer, gyroscope, etc. The smartphone data have been used in this scenario to augment the UWB data. The two datasets were merged on the basis of closest timestamp. The resulting dataset has precise latitude and longitude from UWB as well as the accelerometer, gyroscope and speed data from smartphone making the fused dataset accurate and rich in terms of parameters.

The fused dataset was then used to predict GPS coordinates of pedestrian and scooter using LSTM.

ACKNOWLEDGMENTS

I would like to express my deepest appreciation to professor Dr. Aty, my supervisor, for his constant encouragement and guidance. He has walked me through all the stages of the writing of my study. Without his patient instruction, insightful criticism, and expert guidance, this thesis would not have been possible. Additionally, I would like to appreciate my committee members: Dr. Zubayer and Dr. Guo, for their encouragement and suggestions.

Secondly, I would like to express my heartfelt gratitude to my beloved family for their loving consideration and great confidence in me all through these years.

Finally, I also owe sincere gratitude to my friends Jingwan and Xing Di who gave me their help and time by listening to me and helping me work out my problems during the difficult course of the study. Last but not least, I had the pleasure of working with Ou Zheng.

TABLE OF CONTENTS

LIST OF FIGURES viii

LIST OF TABLES x

CHAPTER ONE: INTRODUCTION 1

CHAPTER TWO: LITERATURE REVIEW 4

2.1 Sensing Technologies 4

2.1.1 Camera 4

2.1.2 Ultra-Wideband (UWB) 9

2.2 Sensor Fusion Algorithms 13

CHAPTER THREE: METHODOLOGY 24

3.1 The UWB Sensor Tests 24

3.1.1 Data Collection 25

3.1.2 Performance of the UWB sensor 29

3.1.3 Trajectory Prediction 30

3.2 UWB-Smartphone Fusion 31

3.2.1 LSTM Model for Trajectory Prediction 33

3.2.2 Sensor priority 35

3.3 UWB-Camera Fusion 37

3.3.1 Objective Detection and Tracking 39

3.3.2 Coordinate Transformation 41

3.3.3 Sensor Adjustment Method 42

CHAPTER FOUR: RESULTS 44

4.1 UWB Sensor Data Analysis and Results 44

4.2 UWB and Smartphone Fusion Results 50

4.3 UWB and Camera Adjustment Results 52

CHAPTER FIVE: CONCLUSIONS 57

LIST OF REFERENCES 60

LIST OF FIGURES

Figure 1 The architecture of autonomous vehicle [30] 14

Figure 2 Classical approaches for sensor fusion algorithms [41] 17

Figure 3 Common deep learning sensor fusion algorithms [41] 22

Figure 4 UWB Sensor 24

Figure 5 Studied Intersection for UWB Sensor 26

Figure 6 GNSS Device 26

Figure 7 Landmarks on the crosswalk 27

Figure 8 Landmark locations 27

Figure 9 Anchor 28

Figure 10 Console 29

Figure 11 Participant with the UWB Sensor 29

Figure 12 Latency, Processing Time, and Response Time Flow Chart for UWB 30

Figure 13 Sensor Priority Flowchart 32

Figure 14 Fused Dataset 33

Figure 15 LSTM Model 35

Figure 16 Sensor Priority Flowchart 36

Figure 17 Proposed scheme for VRU using camera and UWB sensor adjustment 39

Figure 18 YOLOv5 network structure [55] 40

Figure 19 DeepSORT algorithm overall process 41

Figure 20 The flowchart of Coordinate Transformation 42

Figure 21 Ground Truth and UWB Reading Comparison for One Trial 45

Figure 22 Video Image from Roadside 45

Figure 23 Test Results Between Latency, Processing Time, and Response Time 48

Figure 24 Pedestrian Trajectory Prediction Error 49

Figure 25 Scooter Trajectory Prediction Error 49

Figure 26 The results of the experiments 53

Figure 27 Sidewalk by area 55

LIST OF TABLES

Table 1 Summary of sensor data fusion methods [51] 19

Table 2 Validation Results for UWB (Pedestrian) 46

Table 3 Validation Results for UWB (E-scooter) 47

Table 4 Evaluation Results Summary 50

Table 5 Results for Trajectory Prediction 51

Table 6 UWB and Smartphone Transferability Latency 52

Table 7 Comparison of the pedestrian experiments 54

Table 8 Adjustment results for trajectory by zone 55

CHAPTER ONE: INTRODUCTION

Vulnerable road user (VRU) detection plays a crucial role in traffic safety. VRU crashes have been increasing at an alarming pace in recent years. Between 2009 and 2016, annual U.S. pedestrian fatalities increased by 46%, and bicyclist fatalities increased by 34% [1]. Previous studies show that vehicles are obvious and easily detected by the sensors and have various avenues to share their intent, such as braking lights and turning indicators [2]. However, the VRUs may change their direction of movement quickly, not follow the specific sidewalks and be too small to be captured by the sensors, especially cameras. In addition, it is still absent a commonly understood mode to show or observe the intent of pedestrians and cyclists[3]. Therefore, in order to ensure the safety of the VRU, it is necessary to more accurately discover and track the dynamic trajectory of the VRU and provide corresponding warnings to drivers.

To achieve accurate and stable VRU detection, a large variety of sensors have been used, including radar, lidar, ultrasonic detectors, and cameras. Nevertheless, these sensors are not always reliable because of darkness, fog, rain, etc [4]. Versatility is the most distinctive merit of radar sensors because they can operate during the day and nighttime. Most studies, however, reported that radar is not effective in recognizing object details or distinguishing pedestrians from other objects such as rain, snow, and trash bins, leading to many inevitable false detections [5]. Lidar systems are sensitive to precipitation and fog, as well as blinded by direct sunlight [6]. Ultrasonic detectors do not have the potential to be used for VRU detection. It was found that ultrasonic detectors have an effective detection range of six meters or less and are typically

utilized in parking assistance systems [4]. Cameras can also be used to detect and track VRU based on various computer vision algorithms.

There are still several challenges to using cameras to detect VRU. First, in previous studies, it was found that cameras are susceptible to environmental conditions, and sometimes lenses may be distorted and blurred [7]. Second, many studies generate new descriptive features or use complicated data preprocessing techniques to calibrate cameras. However, these processes increase the computational complexity and make them hard to apply in real-time [8]. Third, sensor fusion may improve the camera's performance, but some sensors have high battery consumption and computation costs [9].

For detection and tracking with the camera, it is necessary to perform calibration to obtain correct GPS trajectories. This method is often time-consuming and necessitates accurate ground truth data. In addition, if the camera performs any pan-tilt-zoom function, it is usually necessary to recalibrate the camera. In this thesis, we propose camera calibration using an auxiliary sensor: ultrawideband (UWB). Previous studies also showed that UWB allocates a unique ID tag and the detection range is up to 200 meters [10]. In addition, this sensor has been built into both the Android and iOS systems, which has led to a high market penetration rate [11]. However, UWB is an active detection sensor, which means that each user needs a sensor, and object classification cannot be achieved [10]. Therefore, cameras are able to be calibrated by the UWB sensor due to the UWB's cm-level accuracy and real-time data acquisition.

Furthermore, two types of sensors were evaluated in this study, including Ultra-Wideband Band (UWB) and camera. A corridor within the University of Central Florida (UCF) campus was selected for field tests and demo purposes. Extensive field tests were conducted at

the selected location in order to evaluate the sensor performance. The various outputs were provided by the different sensors: UWB and camera sensors were used to provide localization information of road users, such as vehicles, pedestrians, E-scooters. Besides, pedestrian trajectories were predicted by fusing UWB and smartphone data. The author adopted LSTM algorithm to predict pedestrian trajectories.

The rest of the thesis is organized as follows: The existing studies are reviewed in Chapter two. In Chapter three, the related methodologies are presented. Chapter four provides findings and the experimental results. The conclusions and future research are introduced in Chapter five.

CHAPTER TWO: LITERATURE REVIEW

Current studies about VRU detection used in this thesis can be divided into two groups based on the devices, which are cameras and UWB sensors. Besides, the following section also discusses the details regarding the notable work of sensor fusion methods.

2.1 Sensing Technologies

2.1.1 Camera

Cameras are widely adopted sensors for object detection and tracking. They are promising sensors, which can provide rich information such as position and orientation. It was found that cameras have several features, including low cost, low weight, limited power consumption, and reasonable size [12]. In most cases, cameras at a signalized intersection are used to detect and locate pedestrians. Their superiority is to identify road users (e.g., vehicles, motorbikes, and pedestrians) by computer vision and pattern recognition techniques [13]. Camera-based methods could achieve high accuracy. Nevertheless, several defects in cameras are inevitable. For instance, cameras are typically assumed to be stationary in their original position unless operators control the camera. Information is extracted by this assumption using computer vision algorithms, otherwise, information may be lost [14]. Furthermore, when a vehicle makes a turn, cameras lose the observation position as well as the detection region [15]. Then, the performance of visual tracking can be easily influenced by lighting, weather, and pavement conditions, and it may suffer from distortion in fast motion and odometry drifts [16]. The amount or visual detail of road users is limited due to the low resolution of cameras,

depending on environmental conditions. In particular, compared to daytime, special camera systems are needed at night and the detection rate is inevitably reduced [17]. All these limitations make the camera-based methods less flexible and convenient.

Traffic cameras are used to monitor target objects such as vehicles, drivers, pedestrians, and landmarks on roads. It could clearly capture the feature information (e.g., license plate, color, shape, contour, etc.) of target objects in real time. The cameras could be installed outside and adapt to the brightness, weather and monitoring range changes, high or low temperature and dust, as well as continuous work for 24 hours. The main applications are summarized as follows:

2.1.1.1 Road Monitoring

The camera sends real-time images back to the center for storage via the internet for playback or real-time display. Managers can grasp the traffic conditions of intersections in real time, such as traffic flow, signal light status, vehicle violations, and crashes. Edge computing is an emerging technology that enables users to exploit mobile devices or Internet-of-Things (IoT) in close proximity. Recently, video analytics edge computing exploiting IoT cameras has gained significant attention. However, since video and image transmissions are both time consuming and bandwidth-hungry, running such tasks on remote servers is still very challenging. The video data generated by cameras could be processed through edge computing in order to reduce the transmission delay between vehicles and infrastructures, and several additional services can be achieved. For example, surveillance cameras using edge computing technologies are available for intrusion detection and crowd monitoring, as well as wearable cameras for lifelogging and healthcare monitoring [18].

2.1.1.2 Information Collection

Traffic violations such as speeding and red-light running could be captured, identified, and stored to provide violation evidence and road traffic analysis. Furthermore, cameras could be used to monitor road traffic conditions and estimate traffic parameters such as volume, speed, etc.

2.1.1.3 Image Processing

The camera could analyze the collected images in real time and recognize information such as license plate information. It could analyze the image stream to judge the entry and exit of the vehicle according to the change of the background gray value, calculate the trace of the vehicle and track license plates and measure accurate speed.

Different types of cameras could be utilized for different situations and objectives. In recent years, many companies provide video analysis solutions by integrating cameras with AI chips. The following types of cameras are widely utilized for various traffic applications:

1. Fisheye Camera

A fisheye camera refers to a camera with a fisheye lens, with an angle of view close to or equal to 180-degree and a focal length of 16 mm or less. To maximize the photographic angle of view, the diameter of the front lens of this photographic lens is parabolic and protrudes toward the front of the lens, which is similar to the eyes of a fish called "fisheye lens." The fisheye lens is a special kind of ultra-wide-angle lens, and its view could reach or exceed the range that the human eye can see. Therefore, fisheye lenses are very different from the real scene in human eyes. The scenery we see in real life is a regular and fixed scenery, but the picture effect produced by fisheye lens is not this form.

In general, the shorter the focal length, the larger the viewing angle and the stronger the deformation due to optical principles. In order to achieve a 180-degree ultra-large viewing angle, the designer of the fisheye lens has to allow the deformation (barrel distortion). Thus, the further the object is from the center of the camera view, the stronger the barrel distortion.

For autonomous vehicles, implementing robust visual localization using only cameras and vehicle sensors is a challenge. Muehlfellner et al. integrated four monocular, wide-angle, fisheye cameras on a vehicle and implemented a mapping and localization pipeline [19]. Visual features and odometry are combined to build and localize against a keyframe-based 3D map.

Traditionally, the fisheye image is transformed to a flat image before subsequent processing, which is time-consuming. The high-resolution flat image with low quality at the edge will also slow down the post-processing. Zhu et al. described a method of fisheye object detection and localization on the unrestored fisheye image to shorten the processing time [20]. A single-stage neural network is built for object detection. To improve the detector performance, its submodules are designed specifically by combining the central rotational property and severe distortion of the fisheye image. After that, the detected objects are localized with the assistance of data fusion on the fisheye model and micro aerial vehicle (MAV) sensory data.

Another drawback of fisheye cameras is that the pixels in the central region of a fisheye image have lower angular resolution compared to those of a pinhole image with the same image resolution. To maintain both accuracy and efficiency, Cui et al. presented a real-time dense mapping system for self-driving vehicles and a new strategy for fisheye depth map recovery using images with different resolutions [21]. Several filtering methods are adopted to filter noisy and unreliable depth estimates in texture-poor and low-resolution areas. Subsequently, they fused

the fisheye depth maps directly into a 3D map model. To increase the system's scalability, they reduced both memory usage and run-time with local map pruning and only store map data in the vehicle's vicinity. To fulfill the requirement of real-time mapping for self-driving vehicles, a fast object detection pipeline to handle potentially moving objects were adopted.

2. Closed-Circuit Television (CCTV)

Closed-circuit television (CCTV) cameras are widely deployed and are mainly used for surveillance. CCTVs provide long-distance monitoring and control. They can also provide a reliable way to collect road users' positions in time and space. Zhang et al. attempted to extract the road user locations using CCTV, identify the pedestrian-involved potential collisions, and analyze the interaction behavior between pedestrians and vehicles [22]. The results show that the pedestrian–vehicle interaction behavior in many types of intersections can be analyzed via CCTV. Hashmi and Keskar suggested a statistical-based approach to determine the traffic parameters at heavily crowded urban intersections. In this study, CCTVs are used to estimate vehicle count at a high traffic density T-intersection [23]. The proposed approach could guarantee a high success rate for vehicle detection at heavily crowded urban Intersections. CCTV in traffic flow could monitor at a T intersection using a detection zone-based approach. Thus, CCTV can be implemented for traffic analysis at multi-directional, multi-lane traffic flow.

One of the challenges of implementing applications using CCTV cameras is the ability of the end user to manually pan, tilt, and zoom into the camera view. The constant movement of the camera's field of vision impedes algorithms that depend on placing virtual zones, lines, or loops for computation.

To summarize, fisheye cameras could cover a larger area that could be suitable for large intersections; CCTVs cameras are widely deployed and can be used for many traffic applications.

2.1.2 Ultra-Wideband (UWB)

With advances in communication technologies, UWB sensors have very promising technology for accurate positioning with pre-installed receivers. Previous studies described that UWB sensors could be used as a replacement for Global Navigation Satellite Systems (GNSS) [24]. UWB sensors possess the following technical characteristics: First, theoretically, UWB data transmission could reach a rate of 1 Gbit/s, thus achieving a rate of more than 100 Mbit/s in practice [25]. Meanwhile, UWB can provide the relative position and precise trajectory of VRUs, and its positioning accuracy can attain the centimeter level, which is a cheap alternative to a lidar-based [26]. This technology has been recognized as a feasible technology for wireless sensor networks (WSNs) applications due to its very good time-domain resolution, allowing for precise location, tracking, low complexity, and coexistence with low-power and low-cost on-chip implementation facilities [27]. Unfortunately, UWB sensors have some drawbacks, which are mainly applied to indoor terrestrial applications and need a sensor tag for each user [28].

UWB positioning technology is used to transmit data with low power spectral density and narrow pulse width. It has the characteristics of high temporal resolution and strong spatial penetration. In line of sight (LOS) environment, the ranging and positioning accuracy is better than the centimeter level. UWB was initially used for military-industrial purposes, and commercial specifications were released in 2002. Compared with the traditional communication system, the UWB positioning technology has the following technical characteristics [11]:

(1) High-speed data transmission: theoretically, UWB can reach the rate of 1 Gbit/s, so it is possible to achieve the rate of more than 100 Mbit/s in practice.

(2) Simple implementation of the system structure: UWB positioning systems can be roughly divided into location-aware layers, network transmission layers, service layers, and positioning application layers.

(3) Low power consumption: UWB transmits data signals by sending nanosecond pulses, which only consume a small amount of energy when transmitting narrow pulses. In short-range applications, the transmission power of UWB is usually less than 1MW (which is also the technical requirement of the Federal Communications Commission (FCC) for UWB in order to avoid interference to other devices).

(4) Low interception rate and detection rate: the transmission power spectral density of the UWB system is very low, the useful information is completely submerged in the noise, the probability of being intercepted is very small, and the probability of being detected is also very low.

(5) Strong anti-jamming ability: UWB communication adopts time modulation sequences, which can resist multipath fading. Multipath fading refers to the random change of signal amplitude at the receiving point caused by the superposition of reflected wave and direct wave, while each pulse transmission time of UWB systems is very short, and the transmission and reception of direct wave have been completed before the arrival of the reflected wave.

(6) High precision positioning: UWB has very high positioning accuracy using UWB radio communication, it is easy to integrate positioning and communication.

Moreover, UWB radio has strong penetration ability and can carry out high-precision positioning indoor and underground. UWB technology can give the relative position, and its positioning accuracy can reach the centimeter level.

UWB location technology can be based on or not based on range. In contrast, the location technology without distance measurement has poor accuracy and needs many anchor points (nodes with known location). Indoor localization requires high accuracy and few anchor points, so location algorithms based on distance are generally used, including time of arrival (TOA), time difference of arrival (TDOA), received signal strength (RSS) and angle of arrival (AOA). AOA locates by acquiring the signal arrival angles from the measured point to two receivers, which needs to configure a complex antenna system, and the influence of angle error on the positioning accuracy is far greater than the ranging error. According to the signal propagation model, RSS uses the relationship between the received signal strength and the signal propagation distance to locate the target. The location coverage distance of this method is relatively short, and it is highly dependent on the channel transmission model. The change of multipath and environmental conditions will seriously deteriorate its accuracy. Meanwhile, the accuracy of range estimation has no relationship with the signal bandwidth, which is the advantage of UWB. Therefore, RSS and AOA methods are generally not used for UWB positioning alone but can only be used as auxiliary means for primary coarse positioning. The accurate positioning of UWB positioning technology mainly depends on precision ranging.

In the process of TOA implementation, the distance information between the UWB location tag and each base station needs to be measured, so the location tag needs to communicate with each base station back and forth, so the power consumption of the location tag

is high. The advantage of this location method is that it can maintain high location accuracy both inside and outside the location area (inside and outside the area surrounded by the base station).

For industry, the UWB positioning technology started from the pulse communication technology that emerged in the 1960s. In 2002, the United States first approved the technology used for civilian communications. In 2015, this technology began to show up in the Microsoft Indoor Positioning Competition (Microsoft, 2015). In 2016, the UWB positioning technology occupied half of the 3D group of the Microsoft Indoor Positioning Competition (Microsoft, 2016). Commercial applications. Compared with consumer-level (To C) applications, enterprise-level (To B) pays more attention to technical performance and is less price-sensitive. Compared with other positioning technologies, UWB positioning technology has high precision and security. It also has indicators such as low latency, high transmission rate, and high stability, which lead UWB to be applied in the enterprise market, mainly including judicial prisons, hospitals, mines, chemical plants, factories and warehouses, and other application scenarios that require material or personnel management.

In order to quickly promote the development of this industry, NXP, Samsung, Sony, Bosch, ASSA ABLOY, HID Global formed the FiRa Alliance in 2019, aiming to use UWB technology promotes a seamless user experience (Wikipedia, 2019). Sony, LitePoint, and the Telecommunications Technology Association (TTA) have also joined the FiRa organization. The alliance is also actively promoting the development and definition of technology in cooperation with other industry participants such as automobiles.

1. Apple U1 chip: The Apple U1 chip uses UWB technology (2019) to make the phone 11 series more spatially aware and can accurately locate other Apple devices equipped with

U1. Unlike traditional narrow-wave signals, UWB uses a wide spectrum range and is not easily interfered with, so it can provide more precise positioning capabilities than Wi-Fi and Bluetooth. The U1 chip uses UWB technology to achieve spatial perception, allowing the iPhone 11pro to accurately locate Apple devices that are also equipped with the U1 chip.

2. Samsung: As early as 2003, Samsung teamed up with chip manufacturer Staccato to deploy UWB technology. In 2006, Samsung exhibited at CeBIT. On display, a prototype mobile phone SGH-i750 based on UWB technology and d-hoc transmission mode was exhibited for data transmission. According to complete statistics, since 2003, Samsung has applied for more than a dozen patents in the United States. Especially since 2013, the number of UWB patents is more intensive than earlier in Samsung.

2.2 Sensor Fusion Algorithms

As more and more sensors are being used for urban streets and traffic safety, it is becoming increasingly important to understand multiple sensors collectively. Sensor fusion is, therefore, an interesting area of research and has been studied extensively for different applications such as intelligent intersections, and the navigation system of autonomous vehicles. Smart corridors can make the intersection more stable and efficient by adding sensor-based technologies and by broadcasting secure communications in near-real-time. Some recent studies rectify three key components for an intelligent corridor, such as the complete sensor set for an intersection, the powerful sensor fusion algorithms that generate the environment model, and the dedicated short-range communication units, both at the intersection and in the vehicle [29]. The

sensors such as camera, LiDAR, SONAR, etc. collect data from the surrounding environment. The main goal of such systems is safety and reliability, but both elements are controlled by the information received from various sensors. It is very difficult to maintain safety and reliability based on information from a single sensor because it might report faulty or incomplete sensor data. Therefore, it is important to extract information from different sources to make a more accurate and complete estimation. Sensor fusion is a technique that integrates the information collected from multiple sensors. It is an integral part of the design of a connected system, which is composed of numerous components according to Fei Liu et al. [30]: algorithms, including sensing, vision, and decision-making; customer systems, including the operating system and the hardware infrastructure; and the cloud platform, including high-definition maps, training of deep learning models, emulation, and data storage as shown in Figure 1.

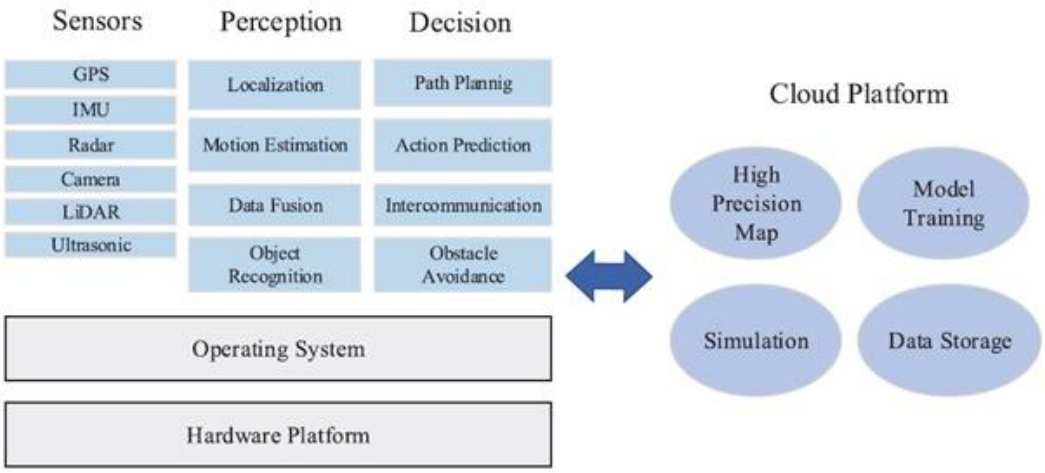

Figure 1 The architecture of autonomous vehicle [30]

Perception and trajectory extraction of individual traffic participants is complex, especially for challenging scenarios with multiple road users or under bad weather conditions. In recent years, sensor fusion methods have been published with the aim of better object detection.

A lidar-radar sensor fusion method for object detection mechanism in transportation during daytime and night is described [31]. Especially, sensor fusion methods could calibrate cameras by fusing radar or LiDAR. For instance, a scheme that applies 3D point clouds returning from the lidar depth sensor to further examination of the object's shape, which the camera detects, is proposed to reduce the false-positive rate and the object occlusion problem [32]. Ishikawa et al. explored the possibility of an automatic targetless calibration method between a fixed camera and LiDAR, based on hand-eye calibration and using sensor-fusion odometry [33]. In another study, Han, X., et al. developed a real-time pedestrian detection system using LiDAR and vision in-vehicle. The regions of interest are gained by clustering lidar point clouds and projecting them onto the images. After that, these regions replace image areas that have no lidar points projected onto them [34]. Similarly, pedestrians are extracted from high-definition LIDAR and a single camera in parallel by SVM-based classifiers for the purpose of realizing efficient and accurate pedestrian recognition [35]. Besides, a fusion of stereo camera and radar targets to significantly improve the tracking of pedestrians indoors is presented, especially suited for surveillance and security applications [36]. Whereas the previous studies did not fuse the UWB sensor and the camera to improve the accuracy of real-time VRU detection and tracking.

Sensor technology can be divided into two types, namely proprioceptive sensors and exteroceptive sensors. The proprioceptive sensors measure values that are intrinsic to the device. Examples in this group include accelerometers, global positioning systems (GPS), inertial measurement units (IMUs), optical flow sensors, gyroscopes, encoders, and gyro-meters. Exteroceptive sensors receive data from the vehicle's surrounding environment. The common

types of exteroceptive sensors are cameras, sonars, Infrared, Radar, LiDAR, Radiofrequency (RF) systems.

Sensors can also be categorized as passive sensors and active sensors [37]. RGB cameras, thermal cameras, and GPS are the passive sensor which generates the output from sensing the surrounding energy. On the other hand, the active sensors, such as Radar and LiDAR produce output from the reflection of the light energy. In order to track or control unique groups of identified traffic sources, Gentili and Mirchandani have introduced a new series of network location concerns that decide where to find active sensors[38]. The authors investigated the performance of multi-sensor fusion, including LiDAR, camera, ultrasonic, radar, global positioning system (GPS), inertial measurement unit [39], and V2X communication [40].

Fayyad et al. classified the sensors fusion algorithm into classical algorithms and deep-learning-based algorithms[41]. The traditional sensor fusion approaches include probabilistic methods, statistical methods, evidential reasoning methods, interval analysis methods, and knowledge-based methods such as fuzzy logic and possibility. Figure 2-2 shows the classical approaches of sensor fusion. Sasiadek classified the sensor fusion algorithm into three different categories such as probabilistic techniques, least-squares methods, and intelligent fusion[42]. The probabilistic algorithms are recursive operators, robust statistics, evidence theory, and Bayesian reasoning. Kalman filtering, regularization, uncertainty ellipsoids and optimal theory are the least-squares approaches. Fuzzy logic, neural networks and genetic algorithms are the intelligent fusion techniques. The combination of the probabilistic and holistic prediction algorithm is used to accurately predict the states of the ego-vehicle and to improve the perception and risk evaluation module performance in an integrated vehicle safety system [43].

Simultaneous Localization and Mapping (SLAM) is a method to combine a set of sensors to make a map. The tracking system of SLAM depends on vision-based sensors, other sensors such as GPS, LiDAR, and sonar [44]. Eckelmann et al. present the possibilities of achieving correct localization and object identification, based on the LiDAR method and the Differential Global Positioning System (DGPS) methodology[45]. The authors proposed a navigation system by combining four sensors (i.e., GNSS, IMU, ODO, LiDAR-SLAM). It can effectively fuse information from various sources to maintain the SLAM process and reduce overall navigation error, particularly in harsh areas where the GNSS signal is severely degraded and LiDAR-SLAM has inadequate environmental characteristics [46].

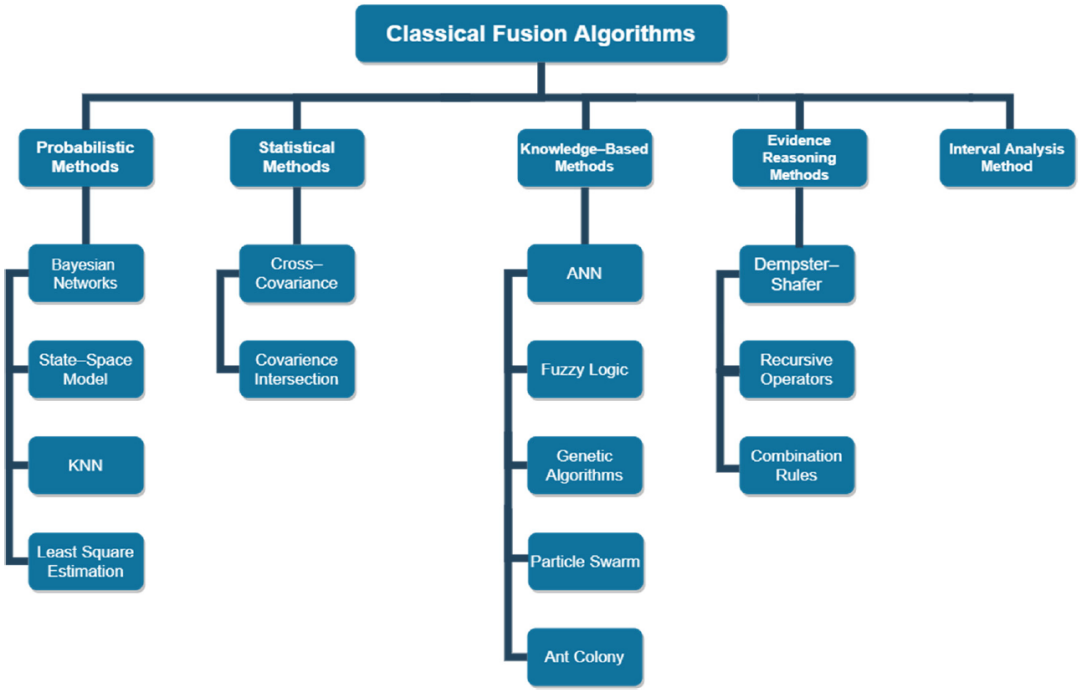

Figure 2 Classical approaches for sensor fusion algorithms [41]

Another strategy is to reduce unintended noise or information by averaging multiple sensor data. The weighted least squares approach has introduced two distributed measurements fusion Kalman filtering algorithms in terms of the weighted average measurements and the average inverse-covariance matrices in sensor networks [47]. In omnipresent localization, an algorithm is proposed based on robust adaptive Kalman filtering to approximate precise headings in order to improve the location tracking performance of the pedestrian dead reckoning system in complex environments [48]. Olfati-Saber and Shamma presented a consensus filter, which plays an important role in solving a data fusion problem [49]. To find an exact location for an autonomous robot, Das et al. developed an intelligence localization model for robots by combining various types of location sensors using a Kalman Filter and an Extended Kalman Filter[50]. In order to predict the reliability factor for each sensor, the authors have followed a fuzzy model. Pires, Ivan Miguel, et al. provided a survey report on different algorithms of sensor fusion [51]. Table 1 shows the summary of sensor fusion techniques.

Table 1 Summary of sensor data fusion methods [51]

Sensors	Methods	Achievements
Accelerometer, Gyroscope, Magnetometer, Compass, GPS receiver, Bluetooth, Wi-Fi, digital camera, microphone, RFID readers, IR camera.	Dead-reckoning pedestrian tracking system, Autoregressive-Correlated Gaussian Model, CASanDRA mobile OSGi (Open Services Gateway Initiative) framework, Genetic Algorithms, Fuzzy Logic, Dempster-Shafer, Evidence Theory, Recursive Operators, dynamic time warping framework, CHRONIOUS system, SVM, Random Forests, ANN, Decision Trees, Naive Bayes, Decision Tables, Bayesian analysis.	The use of multiple sensors eliminates noise effects; these approaches also test the precision of the data fusion of the sensor; data fusion can be conducted for mobile devices, analyzing the data and presenting the results in a readable format.
Accelerometer; Gyroscope; Magnetometer; Compass, GPS receiver, Bluetooth, WiFi, digital camera, microphone, low-cost wireless EEG sensors, RFID readers, IR camera	Kalman Filtering, Collaborative-Signal Processing in Node Environment (C-SPINE) framework, Drift and Noise Removal Filter (DNRF) method, sensor weighted network classifier (SWNC) model; GATING technique; Cooperative Framework for Building Sensing Maps in Mobile Opportunistic Networks (COUPON) framework; Complete ambient assisting living experiment (CAALYX) system; high energy-efficient very large-scale integration (VLSI) architecture; sensor-fusion-based wireless walking-in-place (WIP) interaction technique; J3 criterion; DFF method; epidemic routing with fusion (ERF) method; binary spray-and-wait with fusion (BSWF) method; inContexto system; root-mean-square (RMS) frame	These approaches facilitate complicated analysis of the volume of data collected when it is centrally stored through a server-side system; using data from a variety of sources decreases the level of uncertainty of output; conducting a data fusion operation in real time may be difficult due to the vast amount of data that will need to be merged; data merging with mobile applications, accessing, etc.

Sensors	Methods	Achievements
	energy, mel-frequency cepstral coefficients (MFCC), pitch frequency, harmonic-to-noise ratio (HNR) and zero-crossing-rate (ZCR), KNN, Least squares-based estimation methods, Optimal Theory, Regularization, Uncertainty Ellipsoids.	
Gyroscope, Compass, Magnetometer, GPS receiver.	Kalman Filtering, Bayesian analysis.	It is one of the most useful for the context-aware localization systems; multiple recognizer algorithms have been described to perform online temporal fusion on either the raw data or the features.
ECG and others	Kalman Filtering.	The uncertainty level of the performance is decreased by using data from multiple sources; several recognizer algorithms are specified to perform online temporal fusion on either the raw data or the features.

Sensor technology can be divided into two types, namely proprioceptive sensors and exteroceptive sensors. The proprioceptive sensors measure values that are intrinsic to the device. Examples in this group include accelerometers, global positioning systems (GPS), inertial measurement units (IMUs), optical flow sensors, gyroscopes, encoders, and gyro-meters. Exteroceptive sensors receive data from the vehicle's surrounding environment. The common types of exteroceptive sensors are cameras, sonars, Infrared, Radar, LiDAR, Radio Frequency (RF) systems.

The deep learning approaches that have been used for sensor fusion are convolutional neural networks (CNN), recurrent neural networks (RNN), deep belief networks (DBN), and autoencoders (AE). The most popular techniques used in autonomous vehicles are CNN and RNN. Figure 3 Common deep learning sensor fusion algorithms [41]represents the different algorithms of deep learning sensor fusion, for instance, Region-Based CNN (R-CNN), Spatial Pyramid Pooling network (SPP-Net), You only look once (YOLO), Single-Shot Multibox Detector (SSD), Deconvolutional Single-Shot Multibox Detector (DSSD), Long-Short Term Memory (LSTM), and Gated Recurrent Unit [41]. Liu et al. introduced faster R-CNN for multispectral pedestrian detection, where color or thermal images are fused to provide the additional information needed for brightness and darkness detection[30]. It would increase the detection results by providing complementary sensor data; however, selecting the right fusion architecture would produce a better detection outcome. Four Convolutional network fusion models called early fusion, halfway fusion, late fusion, and score fusion were also developed and tested [30].

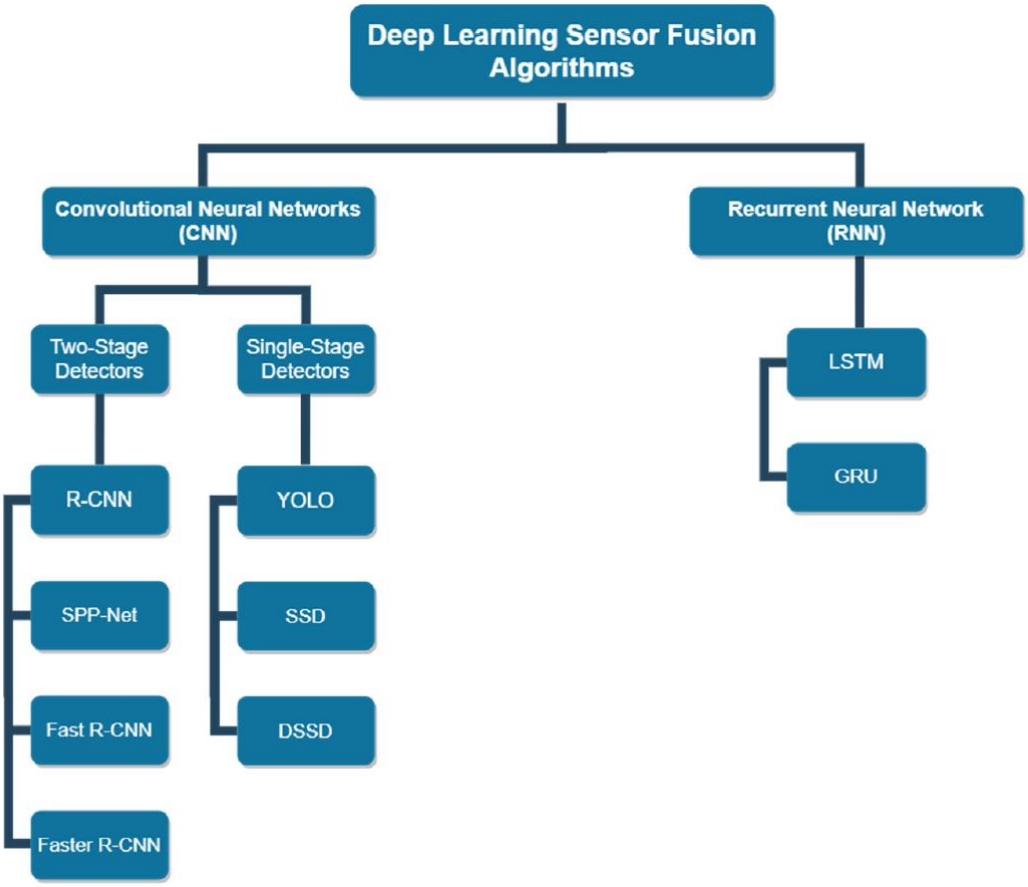

Figure 3 Common deep learning sensor fusion algorithms [41]

Several conclusions can be summarized based on the literature review. First, cameras are the most common sensor to detect and track VRU. Second, UWBs are able to track a road user with centimeter-level accuracy. Third, sensor fusion methods could be used to calibrate the camera. However, there are still some challenges, including camera position, lens distortion and occlusion, time latency, etc. This thesis attempts to address them by proposing a real-time camera calibration method using UWB sensors, which will be elaborated in the following sections.

This thesis tries to mitigate the existing research gaps by proposing camera calibration using an auxiliary sensor: UWB. UWBs are capable of tracking a road user with centimeter-level accuracy. Once a VRU with a UWB traverses in the camera view, the UWB GPS data is fused with the camera to perform real-time calibration. In addition, GNSS is used to collect fixed points as the ground truth to calculate the distance between calibrated trajectories and the ground truth.

The main contributions of this research are as follows:

(1) To our knowledge, the UWB sensor in a sensor fusion system has not been previously applied in camera calibration.

(2) It is determined that the UWB sensor optimizes the effectiveness of the camera calibration method.

(3) The trajectories from the camera were calibrated using UWB only in the initial study, and then only the camera could be used to adjust trajectories accordingly, as long as the cameras remained at the same location, height, and angle.

The proposed method requires that the measurement ranges of the camera and the UWB sensor overlap during the initial experiment. To secure overlaps between detections, the experiment was conducted at intersections. The experiment results show that the proposed strategy is better than the previous methods.

CHAPTER THREE: METHODOLOGY

3.1 The UWB Sensor Tests

UWB is a short-range Radio Frequency (RF) technology for wireless communication that can be utilized for localization by transmitting data through radio waves (Figure 4). In this study, the data collection frequency for UWB is 50 Hz.

Figure 4 UWB Sensor

In general, a UWB system includes three parts in order to obtain localization information:

(1) Anchor: four (4) UWB sensors are used as anchors. The anchors are needed as the fixed reference points for the study area.

(2) Console: one UWB sensor is used as the console. The console needs to be connected to the Terminal (e.g., computers, tablets). It is used to monitor the running status of the system and send instructions to other nodes (i.e., anchor, tag).

(3) Tag: each movement object (e.g., pedestrian) should be equipped with one UWB sensor (i.e., tag) for localization purposes.

3.1.1 Data Collection

This experiment aims to evaluate the performance of the UWB sensor for pedestrian and electric scooters (E-scooter) localization. Intersection 1 (i.e., Gemini Blvd East & North Orion Blvd) was selected for data collection for this study. The pedestrian data collection was conducted on November 9th, 2021, and E-scooter data collection was conducted on November 11th, 2021. The E-scooters have become a popular transportation mode, which provides an option for first-/last- mile trips. However, the rapid growth of using E-scoters has also caused safety concerns, and countermeasures are needed to improve E-scooter safety [52]. Meanwhile, UWB sensors could be equipped on the E-scooters to provide real-time localization information, which could be utilized for real-time warnings to improve road safety, such as jaywalking, conflicts.

In this study, the data collection for UWB sensors included the following steps:
Step 1: select anchor locations. Four (4) anchor locations were selected at the four corners of the intersection (Figure 5). Each anchor is installed on the tripod and connected to a battery to provide the necessary power (Figure 7).

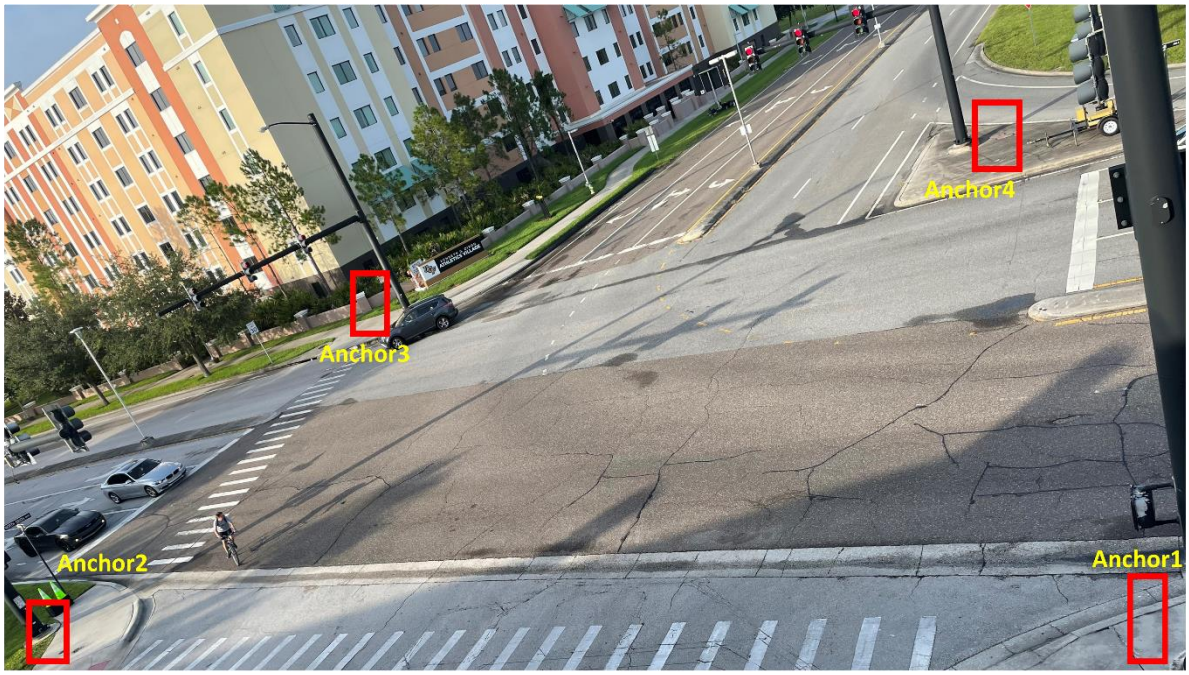

Figure 5 Studied Intersection for UWB Sensor

Step 2: Setup Global Navigation Satellite Systems (GNSS) and Obtain coordinates. The GNSS devices were used to collect the coordinates of the selected locations (Figure 6) Two GNSS devices were used during the experiment. One GNSS device was used as the base station, while the other GNSS device was used as a rover to collect coordinates for anchors.

Figure 6 GNSS Device

Step 3: Selecting landmarks and obtain landmark coordinates using GNSS. Ten landmarks were selected along the crosswalk of the intersection (Figure 8). The coordinates of the landmarks were obtained using the GNSS devices, which were used as the ground truths for UWB sensor evaluation.

Figure 7 Landmarks on the crosswalk

Figure 8 Landmark locations

Step 4: Conduct the experiment. After collecting the coordinates using GNSS, four UWBs were placed at the anchor locations, and one UWB was connected to a laptop as the console (Figure 10). Meanwhile, a pedestrian/E-scooter with a UWB device (i.e., tag) crossed the intersection and passed through the ten landmarks (Figure 11). Meanwhile, a video was recorded from the roadside to obtain the timestamp that the pedestrian/E-scooter arrived at each of the landmarks. Ten trials were conducted for both pedestrian and E-scooter, respectively.

Figure 9 Anchor

Figure 10 Console

Figure 11 Participant with the UWB Sensor

3.1.2 Performance of the UWB sensor

To evaluate the latency, processing time and response time flowchart for the UWB sensor presented in Figure 12 is proposed. The system takes road data from the serial port as input. The output is the position of the tag then transmitted to the cloud server. The processing time is the

duration it takes to process the road data from the serial port, while latency is the round-trip delay to send the output to the cloud.

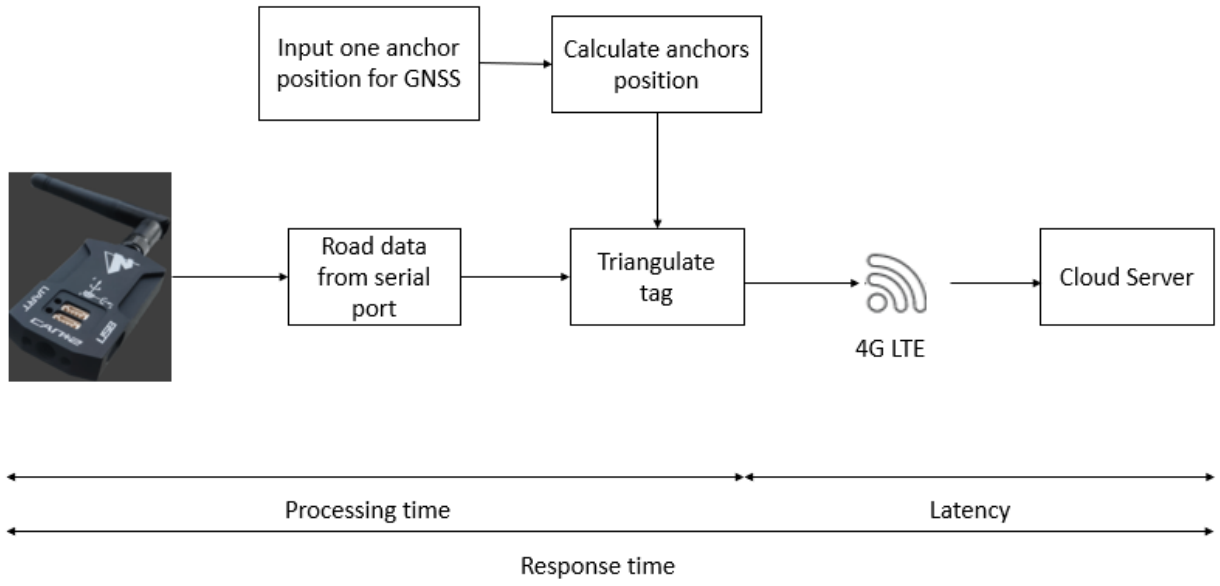

Figure 12 Latency, Processing Time, and Response Time Flow Chart for UWB

3.1.3 Trajectory Prediction

Understanding the probable future coordinates of a road user is pertinent for a smart system making it more robust at detecting conflicts earlier while simultaneously reducing false warnings. For the purpose of predicting trajectories ahead of time, Kalman Filter [53] has been used in this study. It is an algorithm that uses a series of data points that contain noise and other unknown inaccuracies to estimate a few points in the future. It has been widely used in several fields due to its real-time, fast and efficient processing time. Kalman filter estimates the state of a discrete time series governed by the stochastic difference equation given by

$$x_k = Ax_{k-1} + Bu_{k-1} + w_{k-1} \quad (1)$$

Here A is a $n * n$ matrix establishes relation between the two time states of x such as k and $k - 1$ while the matrix B relates an optional control input to the state of x. The Kalman filter can be applied in two steps: a time step and a measurement step [54].

The time update can be given by:

$$\hat{x}_k^- = A\hat{x}_{k-1} + B\, u_{k-1} \tag{2}$$

$$P_k^- = A\, P_{k-1} A^T + Q \tag{3}$$

The measurement update is calculated by:

$$K = P_k^- H^T (H P_k^- H^T + R)^{-1} \tag{4}$$

$$\hat{x}_k = \hat{x}_k^- + K(\hat{z}_k - H\, \hat{x}_k^-) \tag{5}$$

$$P_k = (1 - KH) P_k^- \tag{6}$$

By recursing over these steps, the mean $\hat{x}_k$ and covariance P_k can be estimated.

3.2 UWB-Smartphone Fusion

The UWB sensor data includes latitude and longitude while smartphone data include latitude, longitude, speed, accelerometer and gyroscope data.

The update frequency of UWB and smartphone is different: the frequency of UWB is 50Hz, while the frequency of smartphone is 2Hz. To overcome this issue, the research team

reduced the frequency of UWB for data synchronization. Two data samples were obtained each second from the UWB by averaging every 25 samples.

In the last section, it was shown that UWB GPS coordinates are very accurate although it lacks the other sensor parameters such as accelerometer, gyroscope, etc. The smartphone data have been used in this scenario to augment the UWB data. The two datasets were merged on the basis of closest timestamp. The resulting dataset has precise latitude and longitude from UWB as well as the accelerometer, gyroscope and speed data from smartphone making the fused dataset accurate and rich in terms of parameters. The sensor fusion process is shown in Figure 13. The fused data is shown in Figure 14. The fused dataset was then used to predict GPS coordinates of pedestrian and scooter using LSTM.

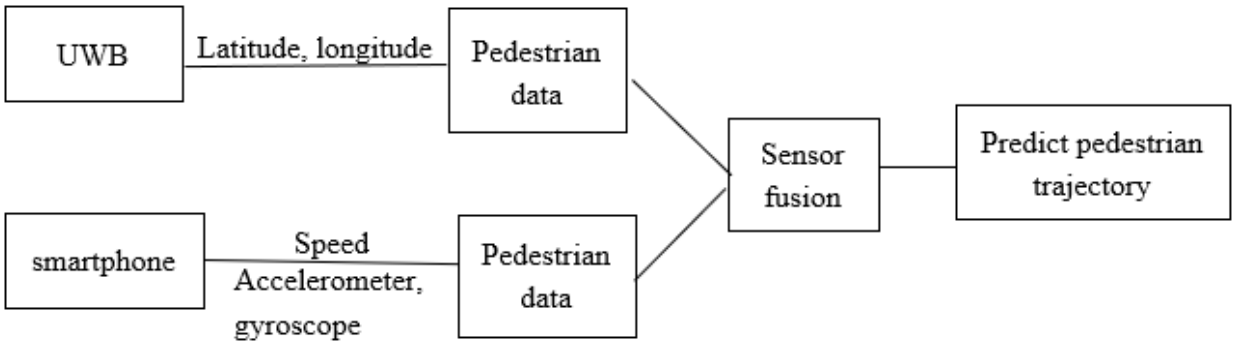

Figure 13 Sensor Priority Flowchart

lat	lon	speed	accx	accy	accz	gyrox	gyroy	gyroz
28.60410372	-81.1960202	0.000805	-1.6495521	9.399306	-0.7947162	-0.0846787	-0.2944284	-0.1063638
28.60410436	-81.19601958	0.0009	-1.6854445	9.497412	-0.9409775	-0.0846787	-0.2944284	-0.1063638
28.60410434	-81.19601897	0.0009	-2.76042	9.528519	1.3956138	-0.30779067	-0.1507266	-0.5719057
28.60410344	-81.19601861	0.0009	-3.4474592	8.477471	0.37357944	-0.30779067	-0.1507266	-0.5719057
28.60410324	-81.19601836	0.0009	-3.708576	9.209974	1.5538392	-0.35452047	-0.06307	0.05657973
28.60410342	-81.19601791	0.190833	-3.4546378	9.368199	1.055833	-0.35452047	-0.06307	0.05657973
28.60410343	-81.19601778	0.190833	-2.382055	8.4248295	3.05713	0.07780667	-0.2719034	0.47394112
28.60410343	-81.19601744	0.190833	-2.1906292	8.896515	2.475375	0.07780667	-0.2719034	0.47394112
28.60410332	-81.19601728	0.005276	0.42622152	8.427222	2.1308086	-0.08651124	0.24563695	0.95246357
28.60410369	-81.19601739	0.005276	0.8428718	8.743374	2.786741	-0.08651124	0.24563695	0.95246357
28.60410346	-81.19601745	0.005276	4.3525443	12.87548	6.485746	-0.46126604	-0.8127323	0.12064233
28.60410334	-81.19601764	0.005276	3.9071803	11.197213	7.493423	-0.46126604	-0.8127323	0.12064233
28.60410315	-81.19601838	0.219317	2.8345976	8.072785	8.724232	-2.5856402	3.002467	-1.5956857
28.60410288	-81.19601843	0.219317	-0.91465646	8.106285	7.1010003	-2.5856402	3.002467	-1.5956857

Figure 14 Fused Dataset

3.2.1 LSTM Model for Trajectory Prediction

In a basic LSTM network architecture, given an input sequence represented by (x_1, $\cdots$,x_T), the output sequence yt can be obtained by iteratively computing Equations 2, 3 for t=1, $\cdots$, T [5]:

$$h_t = \mathrm{LSTM}(h_{t-1}, x_t; W) \tag{7}$$

$$y_t = W_{hy} h_t + b_y \tag{8}$$

where the W terms denote the different weight matrix, b_y denotes the bias vector for the output y_t, and h denotes the hidden state. In the cell of each LSTM, the hidden state is determined by the input gate i, forget gate f, output gate o, and the cell state c via the equations below:

$$i_t = \sigma(W_{xi} x_t + W_{hi} h_{t-1} + W_{ci} c_{t-1} + b_i) \tag{9}$$

$$f_t = \sigma(W_{xf}x_t + W_{hf}h_{t-1} + W_{cf}c_{t-1} + b_f) \quad (10)$$

$$c_t = f_t c_{t-1} + i_t \tanh(W_{xc}x_t + W_{hc}h_{t-1} + b_c) \quad (11)$$

$$o_t = \sigma(W_{xo}x_t + W_{ho}h_{t-1} + W_{co}c_t + b_o) \quad (12)$$

$$h_t = o_t \tanh(c_t) \quad (13)$$

where W_{ab} is the weight matrix from layers a to b;$\sigma(\cdot)$ denotes the sigmoid activation function; each b term with a subscript is the bias vector for the appropriate layer.

As shown in the previous section, latitude, longitude, speed, accelerometer, and gyroscope data were taken as input. A three-layered LSTM network was trained using this data as shown in Figure 15. There are three layers of LSTM model, and the output of the first layer is used as the input of the second layer. The input window is 30 and features number is 5.

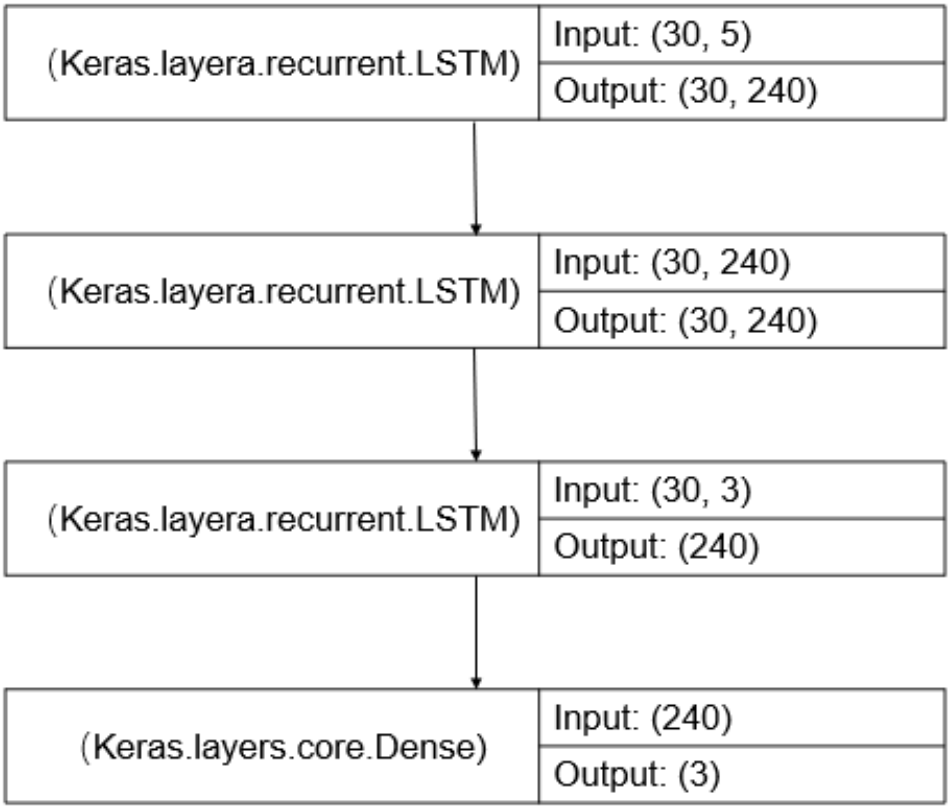

Figure 15 LSTM Model

3.2.2 Sensor priority

The following Figure 16 shows a flowchart explaining sensor priority. Since UWB data is more accurate than smartphone GPS, it is set a higher priority. When both UWB and smartphone is available, the fused data is used. When the UWB data is inaccurate, such as the frequency is too low or most of the values are empty, etc., the smartphone data is directly used. Similarly, when the smartphone data is inaccurate, the UWB data is used.

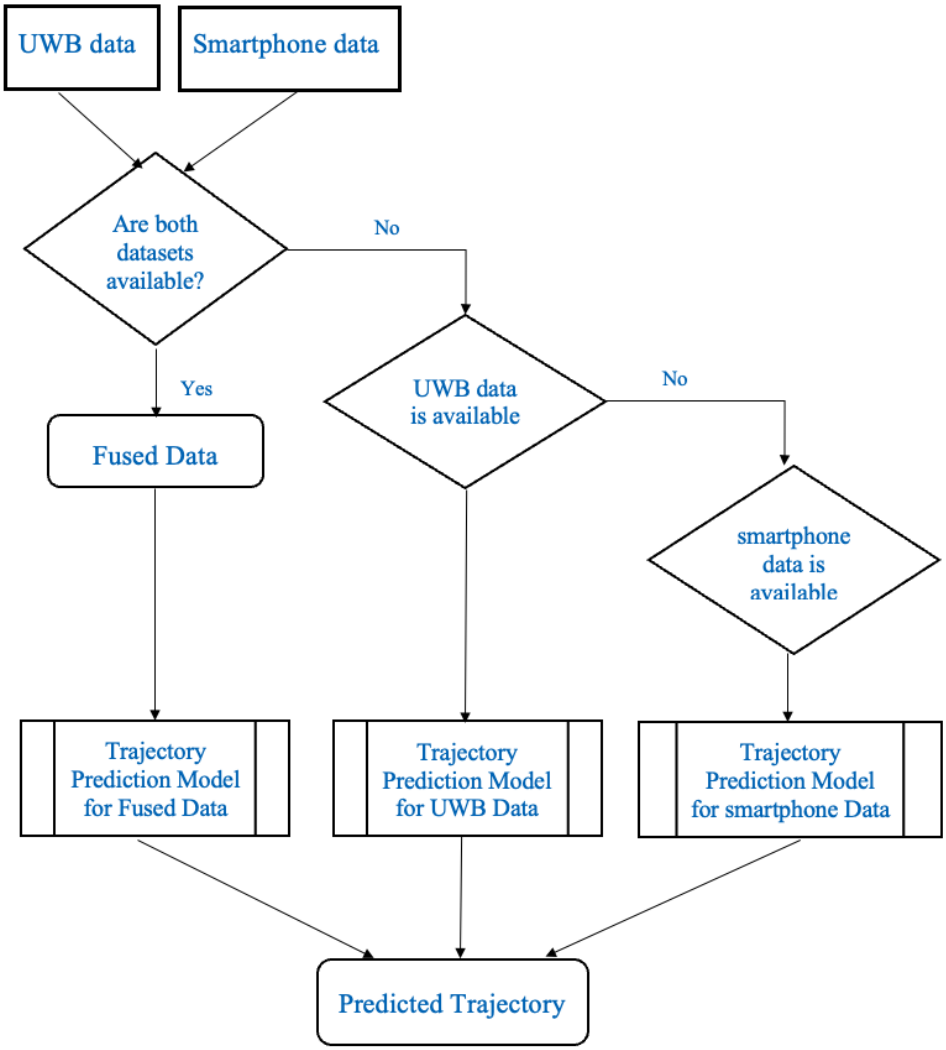

Figure 16 Sensor Priority Flowchart

3.3 UWB-Camera Fusion

To analyze pedestrian safety at intersections, an intersection (i.e., Gemini Blvd East & North Orion Blvd) at the University of Central Florida, Florida, USA was selected as the study site. Video data and UWB data were collected for this study. The spatial map of the site is shown in **Error! Reference source not found.**. The camera was set on the third floor of the garage b uilding.

In this study, the data collection for UWB sensors included the following steps:

Step 1: select anchor locations. Four anchor locations were selected at the four corners of the intersection (Figure 5). Each anchor is installed on the tripod and connected to a battery to provide the necessary power.

Step 2: Setup Global Navigation Satellite Systems (GNSS) and Obtain coordinates. The GNSS devices were used to collect the coordinates of the selected locations. Two GNSS devices were used during the experiment. One GNSS device was used as the base station, while the other GNSS device was used as a rover to collect coordinates for anchors.

Step 3: Selecting landmarks and obtain landmark coordinates using GNSS. Five landmarks were selected along the crosswalk of the intersection. The coordinates of the landmarks were obtained using the GNSS devices, which were used as the ground truth for UWB sensor evaluation.

Step 4: Conduct the experiment. After collecting the coordinates using GNSS, four UWBs were placed at the anchor locations, and one UWB was connected to a laptop as the console. Meanwhile, a pedestrian or E-scooter with a UWB device (i.e., a tag) crossed the intersection and passed through the eight landmarks. Meanwhile, a video was recorded from the third floor of

the garage building to obtain the timestamp that the pedestrian arrived at each of the landmarks. Eight trials were conducted.

Ultra-Wide Band (UWB) is a short-range Radio Frequency (RF) technology for wireless communication that can be utilized for localization by transmitting data through radio waves (Figure 4). In this study, the data collection frequency for UWB is 50 Hz. Our proposed method consists of object detection and tracking, coordinate transformation, and sensor adjustment. Figure 17 displays the schematic procedure for extracting and calibrating pedestrians' trajectories from camera videos and the UWB sensor. First, during the object detection and tracking step, a pedestrian within the camera view is detected using the detection and tracking algorithms. The Yolov5 and DeepSORT algorithms were used in our study.

In the coordinate transformation step, it is necessary to convert points in a two-dimensional coordinate system into latitude and longitude coordinates. In the case of the camera, the authors marked seven points with known latitude and longitude from a two-dimensional video and gained a transform matrix. Then GPS coordinates were gained. Therefore, coordinate transformation can be achieved. In the case of the UWB sensor, the authors use the points of the four anchors as the reference points of the two-dimensional coordinate system and convert the two-dimensional coordinate points of the UWB tags into latitude and longitude coordinates.

In sensor adjustment, the camera and UWB sensors have been used through the outputs of the sensors. The camera is made to detect and track moving objects. The UWB sensor is used to calibrate object information obtained from cameras, such as pedestrians' trajectories. Then, from the UWB sensor's accurate location of objects, objects' trajectories are calibrated in the camera view. Finally, the accuracy of pedestrians' trajectories could be improved.

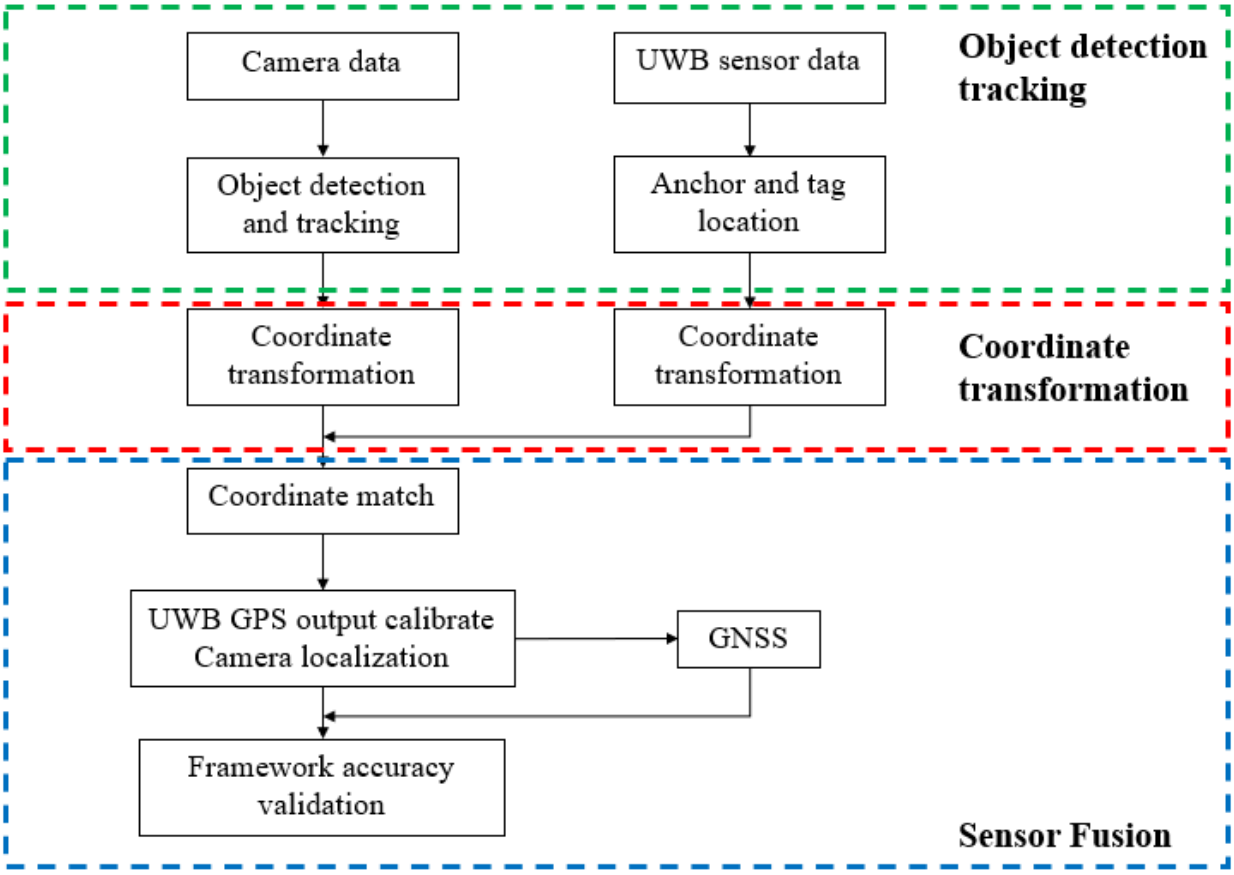

Figure 17 Proposed scheme for VRU using camera and UWB sensor adjustment

3.3.1 Objective Detection and Tracking

A. YOLOv5 object detection algorithm

YOLOv5 is used to detect objects in our study. As shown in Figure 18, the network structure of YOLOv5 is divided into four parts: Input, Backbone, Neck, and Prediction. Input, like YOLOv4, performs Mosaic data enhancement on the input image. Then, the backbone network mainly adopts a focused structure, the CSP structure. In the Neck of the network, the FPN and PAN structures, drawing on the

design ideas of the CSPNet, are used for rich feature fusion. Following the consistent practice of the YOLO series, the Bounding box loss function is adopted for the output of the network.

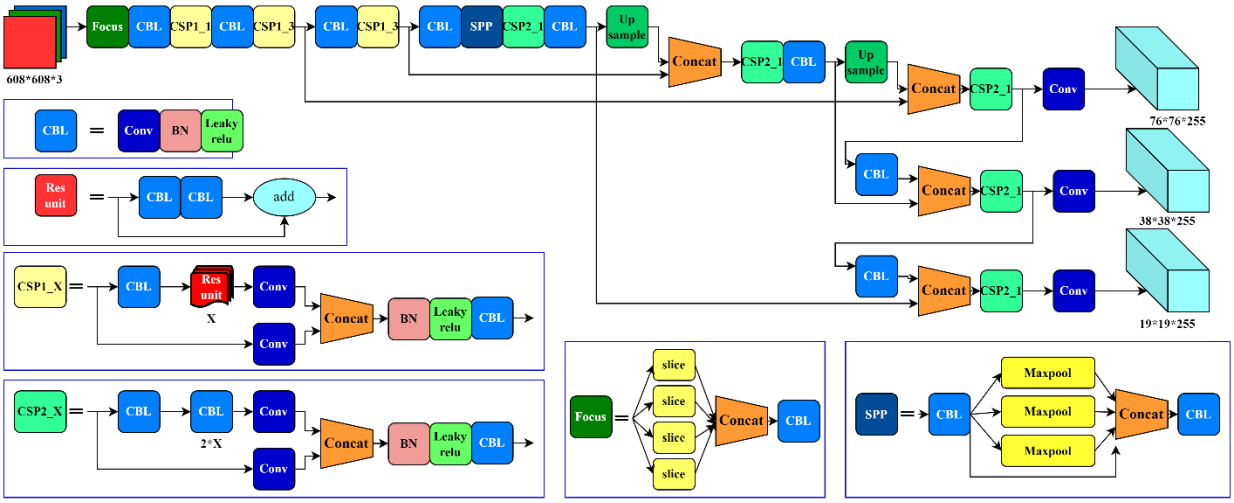

Figure 18 YOLOv5 network structure [55]

B. The DeepSORT object-tracking algorithm

The DeepSORT tracking algorithm is a detection-based tracking algorithm proposed in 2017, that has excellent performance. The DeepSORT algorithm adds Matching Cascade and a new trajectory confirmation to the Sort algorithm. Tracks are divided into Confirmed states and Unconfirmed states. Newly generated tracks are unconfirmed states. Tracks, which in the unconfirmed states must match Detections a certain number of times before they can be converted into Confirmed states. Confirmed Tracks are continuously mismatched with Detections for a certain number of times (30 times by default) before they are deleted. The overall process is shown in Figure 19.

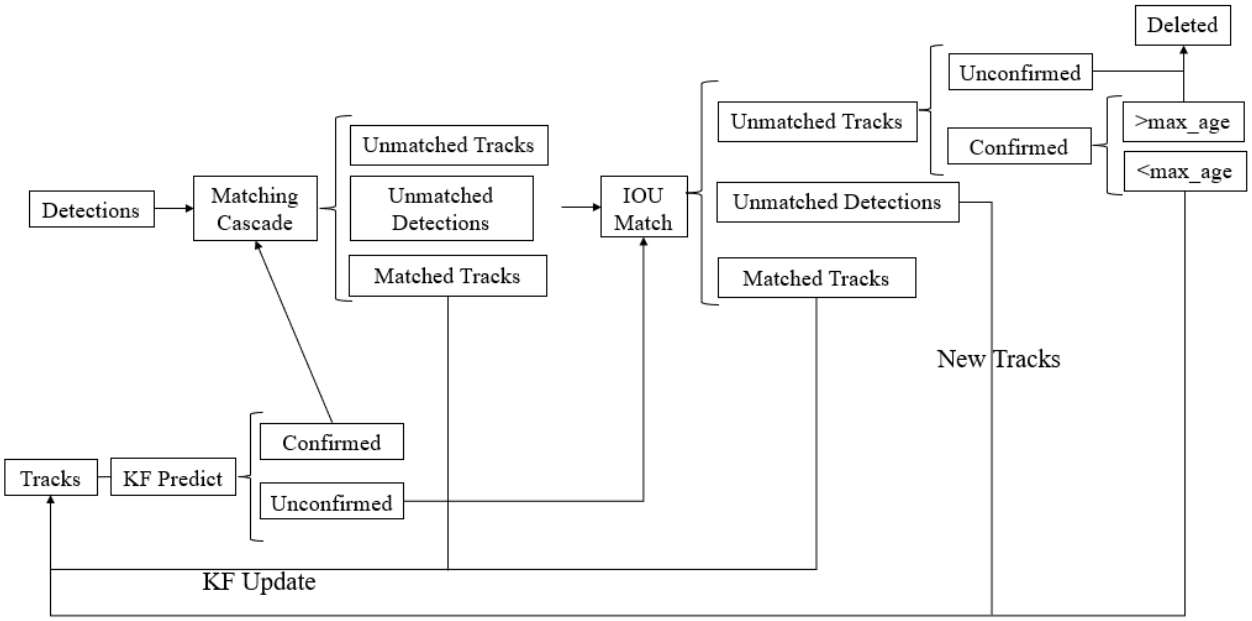

Figure 19 DeepSORT algorithm overall process

3.3.2 Coordinate Transformation

Figure 20 shows a flowchart of Coordinate Transformation. Firstly, find the new X and Y coordinates in the Plane Coordinate System according to the GPS reference points. A perspective transformation through the new X, Y and the original reference X, Y coordinates to get a matrix. Because perspective transformation is the process of projecting an image from one viewing plane to another, it is also known as projection mapping (Projection Mapping). The general process of perspective transformation includes obtained boundary points, defining target boundary points, obtaining transformation matrix, and transforming. The corresponding GPS coordinates are obtained by combining the X and Y coordinates, which are obtained from the pedestrian trajectory of the camera and UWB in different coordinate systems and inverse matrix, respectively.

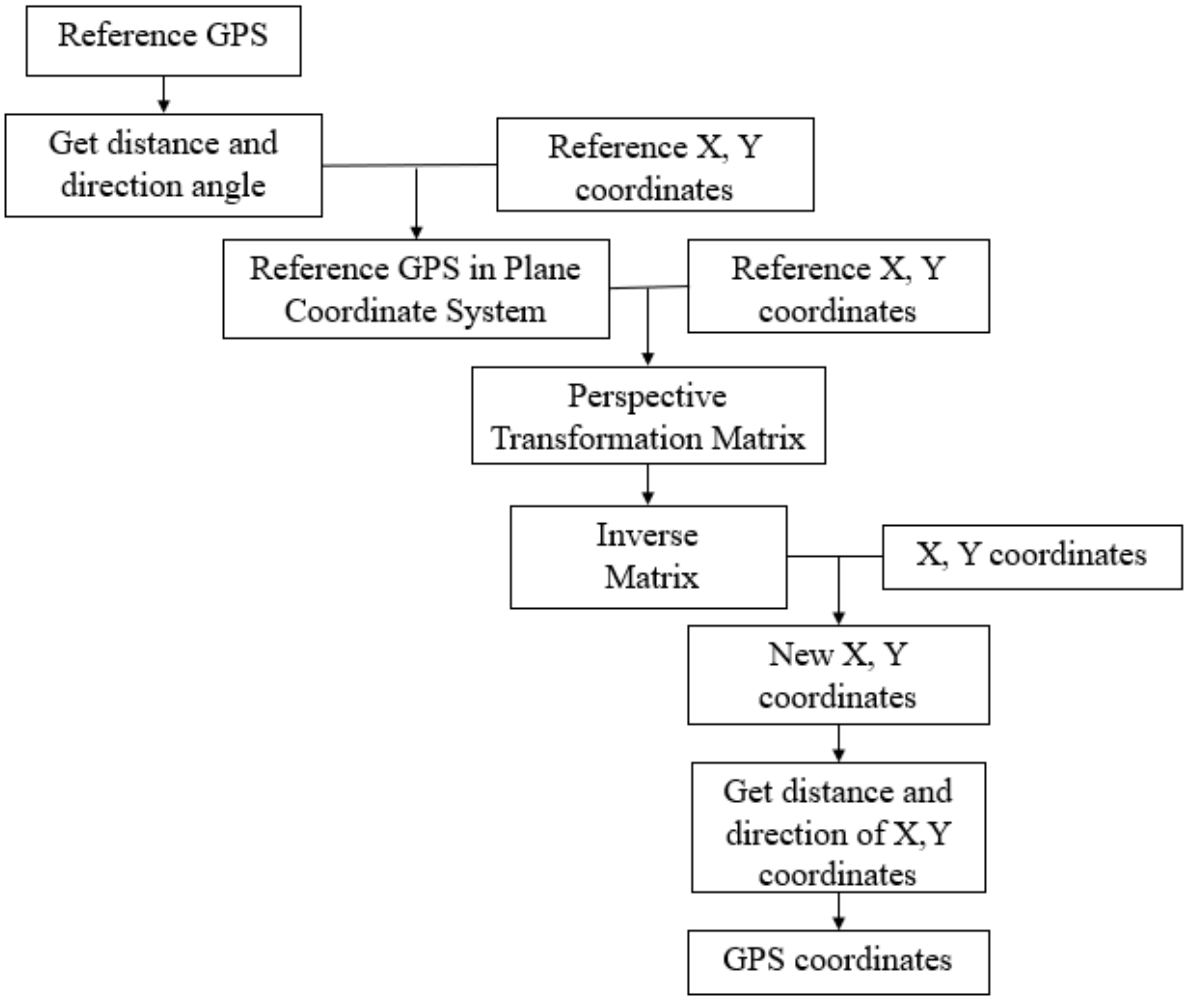

Figure 20 The flowchart of Coordinate Transformation

3.3.3 Sensor Adjustment Method

After the coordinate transformation in the previous step, the pedestrian trajectories of the camera video data and the UWB more accurate pedestrian trajectories are obtained. Authors fused the X and Y coordinates of the camera coordinate system with more accurate GPS coordinates (UWB latitude and UWB longitude), which are obtained from the UWB sensor.

Then a new transformation matrix was produced. Therefore, more accurate trajectories than the camera results individually were gained by experiments.

CHAPTER FOUR: RESULTS

4.1 UWB Sensor Data Analysis and Results

In order to evaluate the performance of the UWB sensors, data analysis was conducted. The data analysis process included the following steps:

- Step 1: Obtain the timestamps the pedestrian/E-scooter arrived at the ten landmarks (t_1, t_2, …, t_{10})
- Step 2: Obtain the coordinates ($UWB_{latitude}$, $UWB_{longitude}$) for those timestamps (t_1, t_2, …, t_{10}) from the UWB sensor.
- Step 3: For each landmark, calculate the distances between the GNSS outputs (i.e., ground truths) and the UWB sensor outputs ($UWB_{latitude}$, $UWB_{longitude}$).

Figure 21 illustrates an example for one of the pedestrian trials. The data collection frequency for UWB is 50 Hz. The corresponding UWB outputs for each landmark can be obtained using the timestamps collected from the video data. For example, in Figure 22, the pedestrian arrived at the landmark point at the timestamp 11:19:26.702. The corresponding UWB outputs (i.e., coordinates) for this timestamp is (28.60485125, -81.19634028The GNSS outputs of this landmark are (28.60485145, -81.19634048). Thus, the localization error for this timestamp is 0.029 cm by calculating the distance between the two points.

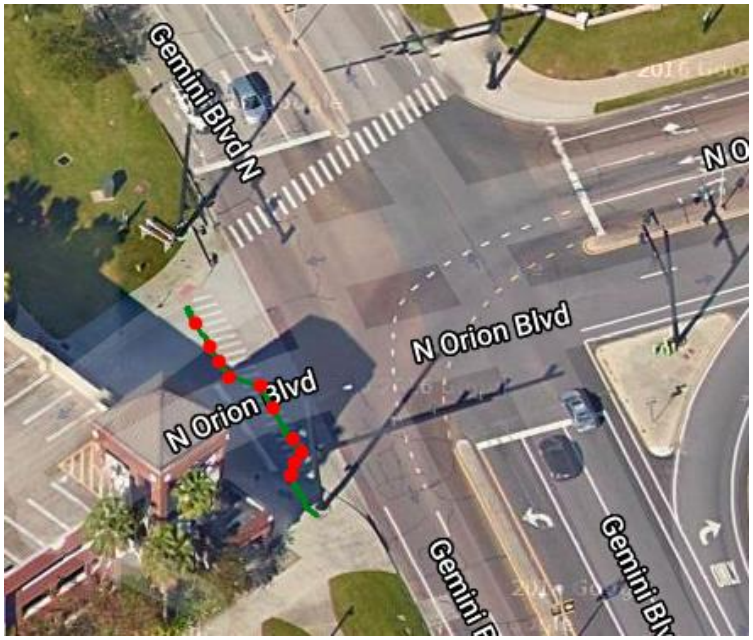

Figure 21 Ground Truth and UWB Reading Comparison for One Trial

*The green line is the pedestrian's trajectory from UWB. The red points are ground truth from GNSS.

Figure 22 Video Image from Roadside

Table 2 Validation Results for UWB (Pedestrian)Table 2 shows the results of the pedestrian experiments based on the ten (10) trials. The minimum localization error for the UWB is 0.39 centimeters, and the maximum error is 37.37 centimeters. The average error is 11.21 centimeters.

Table 2 Validation Results for UWB (Pedestrian)

Pedestrian	**Average error (cm)**	**Maximum error (cm)**	**Minimum error (cm)**
1	10.56	28.50	0.40
2	11.25	31.42	2.42
3	15.45	37.37	2.43
4	8.50	24.75	0.71
5	10.96	28.09	1.49
6	11.04	30.49	2.76
7	10.38	33.32	0.95
8	11.95	30.28	0.86
9	11.44	30.50	1.01
10	10.59	34.07	0.39
Average	11.21	30.88	1.34

Similar to the evaluation of pedestrians' localization, the distances between GNSS points and E-scooter points were calculated and summarized in Table 3. The average error between the GNSS outputs and UWB outputs is 11.82 centimeters. The minimum and maximum errors are 0.66 and 31.91 centimeters, respectively. The results indicate that UWB could provide acceptable position information for both pedestrians and E-scooters.

Table 3 Validation Results for UWB (E-scooter)

E-scooter	**Average error (cm)**	**Maximum error (cm)**	**Minimum error (cm)**
1	10.26	31.91	0.89
2	14.65	27.95	4.41
3	12.43	24.48	1.33
4	10.22	22.76	2.71
5	09.02	25.08	0.66
6	12.43	24.09	2.51
7	12.41	27.45	2.93
8	11.61	31.54	1.83
9	11.75	23.33	2.50
10	13.42	30.22	4.69
Average	11.82	26.88	2.45

As is shown in Figure 23, the latency, processing time, and response time of the UWB sensor are almost real-time. The three lines are shown for a 5-minute upload of real-time data from the field experiments. The average latency, processing time and response time are 0.15 s, 0.03 s, and 0.18 s, respectively. Therefore, the UWB could be used to detect and localize pedestrians and E-scooters.

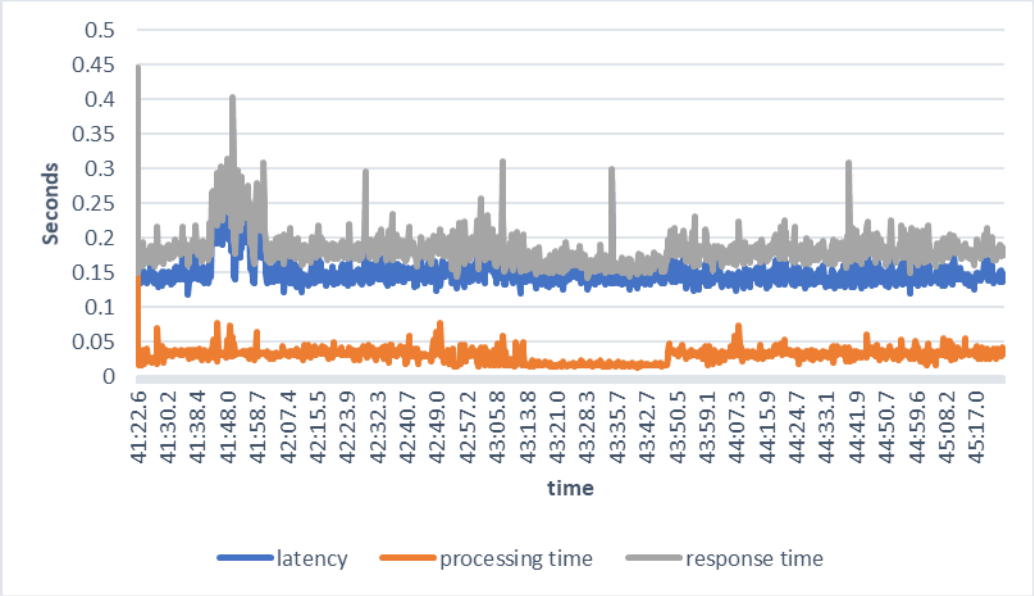

Figure 23 Test Results Between Latency, Processing Time, and Response Time

Using Kalman Filter, trajectory have been predicted for both the pedestrians and scooter riders. GPS points were predicted for 1 and 1.5 seconds ahead of time based on the trajectory of the previous one second. Then the predicted points were matched with the actual points given by UWB. Finally, deviation of the predicted points from the actual points were calculated in meters. This process was repeated each second predicting the next 1 and 1.5 seconds. Average deviation of prediction for the whole route is visualized in Figure 24 and Figure 25.

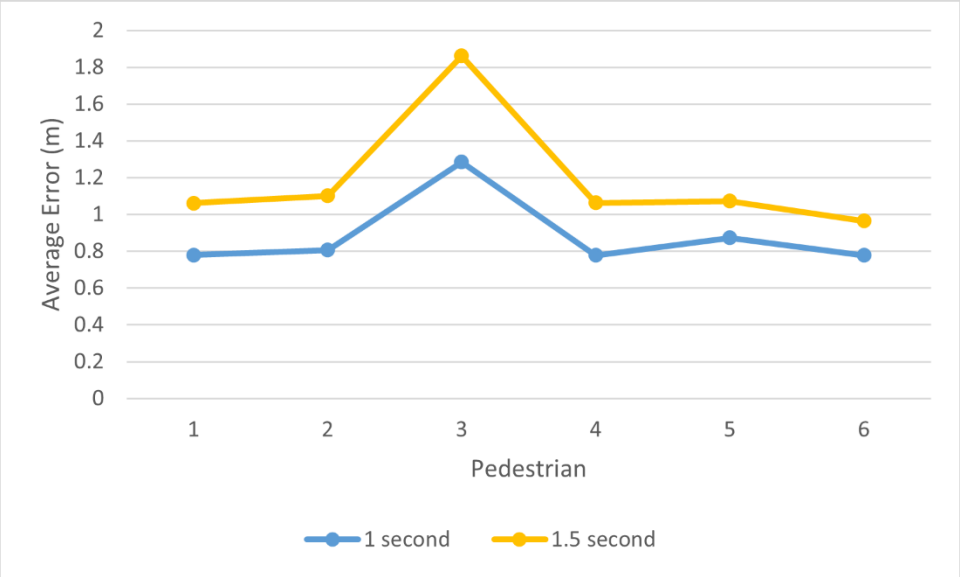

Figure 24 Pedestrian Trajectory Prediction Error

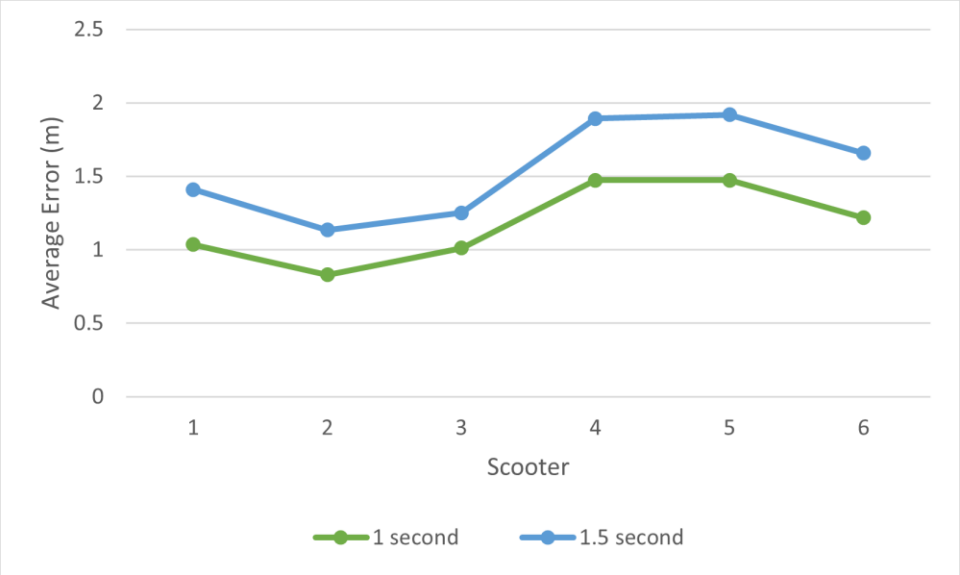

Figure 25 Scooter Trajectory Prediction Error

Table 4 Evaluation Results Summary

Sensor Type	**Equipment Name**	**Output Type**	**Accuracy**	**Latency**	**Processing time**	**Response time**
Ultra-Wideband(UWB)	NoopLoop LinkTrack	Vulnerable road user localization	0.39cm-37.37cm	0.15 s	0.03 s	0.18 s
Camera	ELP-USBFHD06H-KL36IR	Road user classification and localization	93%	0.29 s	0.24 s	0.53 s

4.2 UWB and Smartphone Fusion Results

Table 5 shows the trajectory prediction results from the 3-layer LSTM model. Three different data sources were used such as: fused data, only-UWB data and only smartphone data. Two different input windows were experimented with: 30 and 10 while the output window was 5. Mean Squared Error (MSE) as well as average error between the predicted points and true values were calculated.

From the results in the table, it can be seen that the result with an input window of 30 is better than the result with an input window of 10, and the effect of UWB and smartphone fusion is better than the individual sensors alone.

Table 5 Results for Trajectory Prediction

Dataset	Input window	Output window	test1		test2	
			MSE	Distance(m)	MSE	Distance(m)
UWB	10 (5s)	5	0.0314	0.8045	0.0303	0.5893
smartphone	10	5	0.0896	1.1212	0.1263	1.3331
UWB and smartphone	10	5	0.0520	0.5388	0.0297	0.5715
UWB	30 (15s)	5	0.0300	0.6603	0.0187	0.4627
smartphone	30	5	0.1458	0.5682	0.0924	0.6574
UWB and smartphone	30	5	0.0358	0.5262	0.0792	0.3858

Table 6 shows the transferability latency of UWB and smartphone data fusion. Transferability can be defined as the time taken to re-identify a road user in case of low accuracy of a sensor. For UWB-smartphone fusion, the transferability latency is essentially the time it takes to fuse the two datasets since the fusion steps consider which model to use based on data availability. Four different input windows were experimented with: 30 and 10 while the output window was 5. There were instances from the jaywalking scenario as well. From the results in the table, the results show no difference with the different change in window sizes. Moreover, the average latency of 0.05s and 0.02s are encouraging for real-time applications.

Table 6 UWB and Smartphone Transferability Latency

Scenario	Input Window	Output Window	Latency
Jaywalking	10	5	0.0530
	30	5	0.0530
Crosswalk	10	5	0.0223
	30	5	0.0223

4.3 UWB and Camera Adjustment Results

In order to compare the performance of the camera, the UWB sensors, and the results after camera calibration, data analysis was conducted. The data analysis process included the following steps:

Step 1: Determine the timestamps at which the pedestrian arrived at each of the ten landmarks (t1, t2, ... t5).

Step 2: Obtain the coordinates (latitude, longitude) for those timestamps (t1, t2, …, t5) from the camera, the UWB sensor, and the results after camera calibration.

Step 3: For each landmark, calculate the distances between the GNSS outputs (i.e., ground truths) and the camera, the UWB sensor, and after camera calibration outputs (latitude and longitude), respectively.

Figure 26 displays the results of the experiments based on the pedestrian trials. The data collection frequency for UWB is 50 Hz. The corresponding UWB outputs for each landmark can

be obtained using the timestamps collected from the video data. Figure 26 (a) to (c) show the pedestrian tracking results of the separate camera, separate UWB, and calibrated camera data, respectively. Four different experiments of pedestrian results were shown. The results demonstrate that after UWB and camera Adjustment more accurate position information for pedestrians were provided.

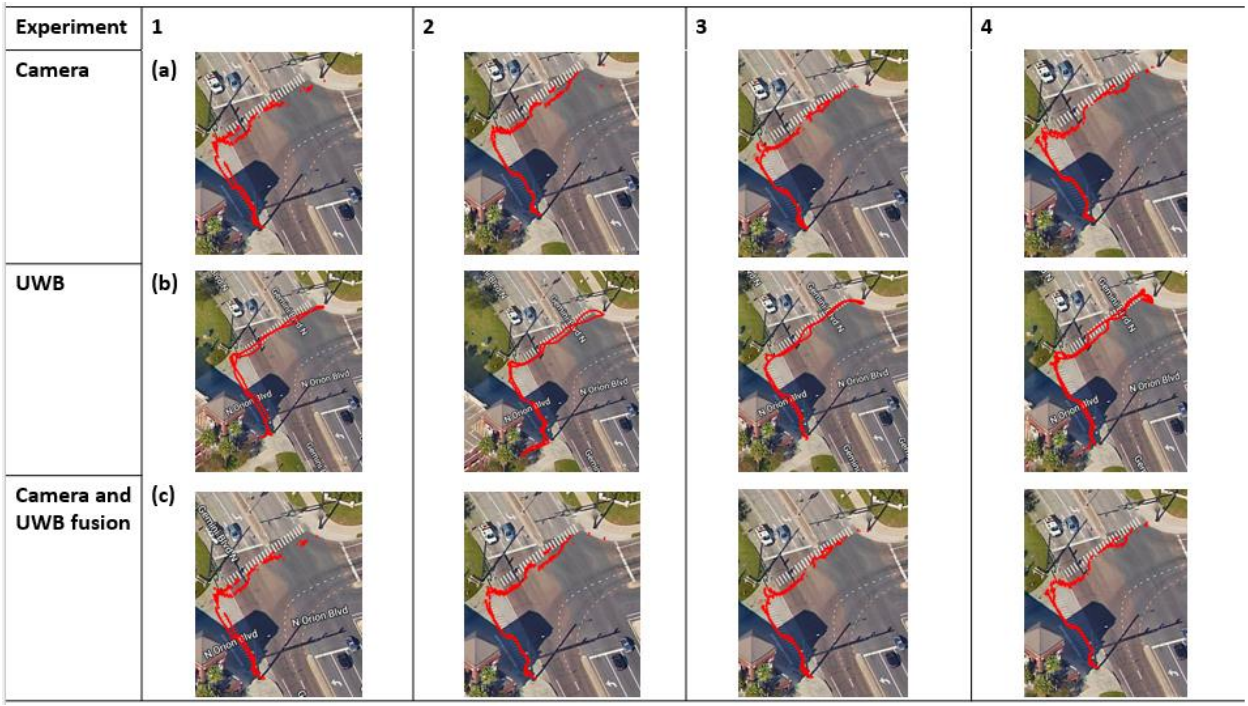

Figure 26 The results of the experiments

Table 7 shows the comparison of the pedestrian experiments based on the trials. The average error of the first test between the GNSS outputs and camera outputs is 0.37 meters. The average localization error for the UWB is 0.16 meters. However, if a UWB sensor is used to adjust camera data, the average error reaches 0.24 meters. Therefore, we are recommending correction and adjustment of the camera trajectories based on the UWB trajectories. In the future, we should focus on correcting the UWB's accuracy based on GNSS first.

Table 7 Comparison of the pedestrian experiments

Average error	**Camera(m)**	**UWB(m)**	**Camera+UWB(m)**
Experiment 1	0.36	0.15	0.23
Experiment 2	0.45	0.23	0.34
Experiment 3	0.34	0.12	0.20
Experiment 4	0.31	0.15	0.20
Total average error	0.37	0.16	0.24

According to the experimental results, we divided the sidewalk into four zones as Figure 27. As mentioned above, from the sidewalks of these four regions, UWB trajectories, Camera trajectories, and after-adjusted camera trajectories of a pedestrian were shown in Figure 26, respectively. Furthermore, we calculated the adjustment errors for different regions separately. The results from Table 8 showed the correction factor of every zone. Take zone 1 as an example, the correction distance between the original camera and after adjustment is 0.05 meters. In zone 2, the correction distance is 0.06 meters. Likewise, the correction distance in zone 3 and zone 4 is 0.13 meters and 0.17 meters. Therefore, the error of the trajectory adjustment using UWB was different depending on the distance from the camera. For example, in zone 1 and zone 2, the pedestrian trajectory of the camera will be more accurate. The average error of the pedestrian trajectory obtained by the camera before and after adjustment using UWB is 0.05 meters, while in zone 3 and zone 4, the average error of pedestrian trajectory obtained by the camera before

and after adjustment is 0.13 meters. The results indicate that UWB and camera adjustment could provide more accurate position information for pedestrians.

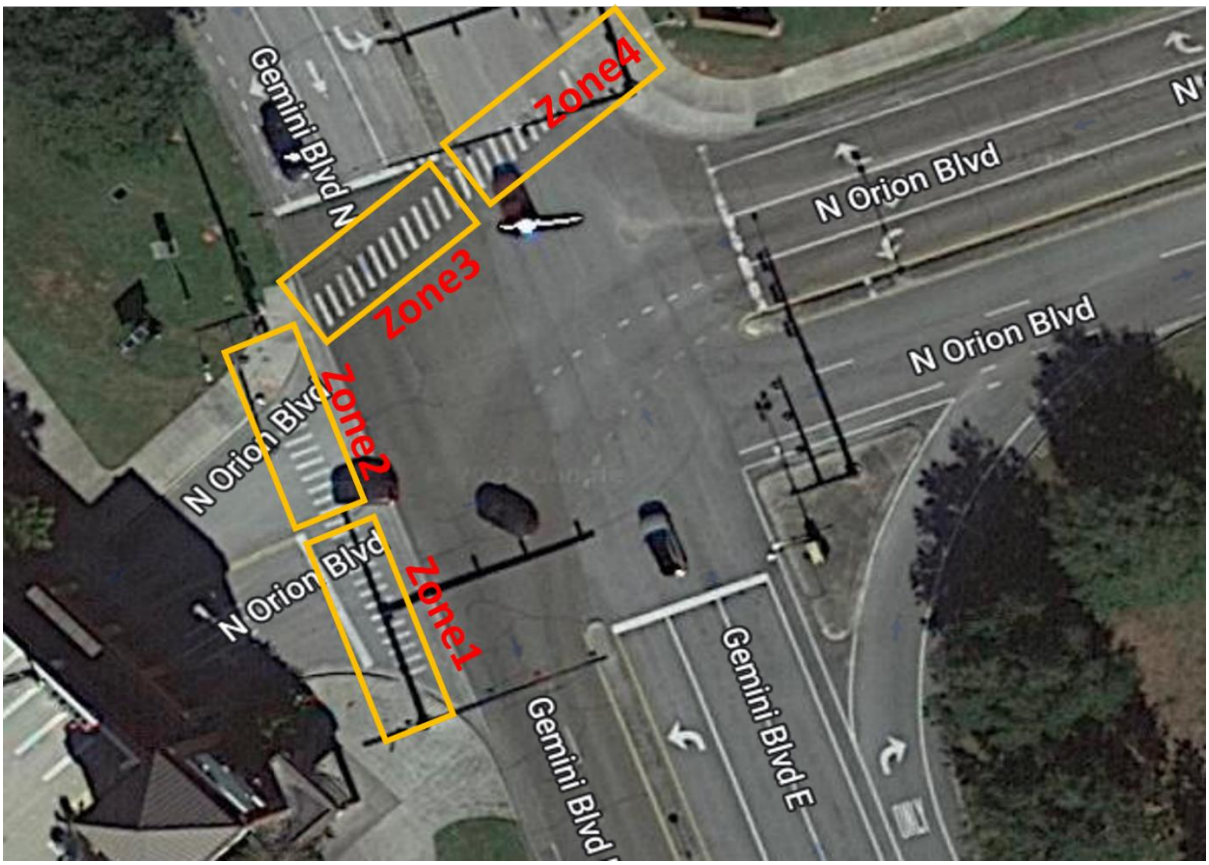

Figure 27 Sidewalk by area

Table 8 Adjustment results for trajectory by zone

Zone	**Correction Factor (m)**
1	0.05
2	0.06
3	0.13
4	0.17

According to adjustment method, after using UWB for the first time, the camera trajectory can be adjusted using its data. The error of the adjustment will be different in different regions.

CHAPTER FIVE: CONCLUSIONS

In this study, three types of sensors were evaluated, including Ultra-Wideband Band (UWB) and camera. A corridor within the University of Central Florida (UCF) campus was selected for field tests and demo purposes. Extensive field tests were conducted at the selected location in order to evaluate the sensor performance. The various outputs were provided by the different sensors: UWB camera sensors were used to provide localization information of road users, such as vehicles, pedestrians, E-scooters. The field tests results illustrate the above-mentioned sensors provide reliable outputs, which could be utilized for smart corridors. Future efforts could be conducted for road user localization or travel time estimation.

Next, sensor fusion and priority strategies were proposed based on the abovementioned sensor outputs to further improve data accuracy and reduce latency. Three types of sensor fusion strategies were implemented and evaluated. Extensive tests were conducted to fully evaluate the performance of the proposed algorithms. In the thesis, the smartphone data include speed, accelerometer and gyroscope have been used in this scenario to augment the UWB data. The two datasets were merged on the basis of closest timestamp and the frequency of UWB is reduced for data synchronization. The resulting dataset has precise latitude and longitude from UWB as well as the accelerometer, gyroscope and speed data from smartphone making the fused dataset accurate and rich in terms of parameters. The fused dataset was then used to predict GPS coordinates of pedestrian and scooter using LSTM. The results are more reliable than UWB individually because the accelerometer, gyroscope and speed data are fused. Furthermore, the authors propose a new scheme to calibrate the traffic cameras by fusing the camera and the UWB

sensor. Generally, the camera and UWB have been used for pedestrian detection and tracking, individually. However, the camera is sensitive to light intensity and environmental changes. Detection is also affected by the angle and height of the camera and the location of the trajectory with respect to the camera. The camera detection in image processing is more complex and needs more processing time. Meanwhile, the UWB sensor is an active detection, which means that it needs a sensor for each user.

This thesis proposes a new concept of camera calibration using the UWB sensor in order to improve the accuracy of VRU's trajectories. To verify the performance of our proposed method, various experiments were performed in real road environments. According to experimental results, proposed method reaches an average error of 0.24 meters. It has better detection performance compared to the individual camera detection. For the same intersection and location of the camera, the use of UWB is needed only once to fuse the data and determine the correct trajectories. Other trajectories collected by the camera latter can be corrected using the same adjustment.

However, there are still some improvements to be made. The proposed method does not complete lost frames, which means it is more likely to lose some jaywalking VRUs. This problem may be overcome by finding the relationship between UWB data and camera data. For the results of the pedestrian and E-scooters experiments from UWB sensor, we also should take the speed and time latency from GNSS and UWB comparison. In addition, for the prediction of pedestrian trajectory in fusing smartphone and UWB algorithms, we should add the pedestrian direction and angle to improve the prediction accuracy.

This camera calibration scheme will be very meaningful in the autonomous vehicle field because VRUs can be detected more accurately and avoid collisions by using our method. Also, this VRU detection and tracking scheme could be adopted to prevent accidents in safety systems of autonomous vehicles, such as active emergency braking systems.

LIST OF REFERENCES

1. Al-Mahameed, F.J., X. Qin, R.J. Schneider, and M.R.R. Shaon, Analyzing pedestrian and bicyclist crashes at the corridor level: structural equation modeling approach. Transportation research record, 2019. **2673**(7): p. 308-318.
2. Izquierdo, R., A. Quintanar, I. Parra, D. Fernández-Llorca, and M. Sotelo. The prevention dataset: a novel benchmark for prediction of vehicles intentions. in 2019 IEEE Intelligent Transportation Systems Conference (ITSC). 2019. IEEE.
3. Maurya, S.K. and A. Choudhary. Deep learning based vulnerable road user detection and collision avoidance. in 2018 IEEE International Conference on Vehicular Electronics and Safety (ICVES). 2018. IEEE.
4. Association, A.A., Automatic emergency braking with pedestrian detection. Retrieved January, 2019. **20**: p. 2020.
5. Klewe, T., C. Strangfeld, T. Ritzer, and S. Kruschwitz, Combining signal features of ground-penetrating radar to classify moisture damage in layered building floors. Applied Sciences, 2021. **11**(19): p. 8820.
6. Dannheim, C., C. Icking, M. Mäder, and P. Sallis. Weather detection in vehicles by means of camera and LIDAR systems. in 2014 Sixth International Conference on Computational Intelligence, Communication Systems and Networks. 2014. IEEE.
7. Wang, J., F. Shi, J. Zhang, and Y. Liu, A new calibration model of camera lens distortion. Pattern recognition, 2008. **41**(2): p. 607-615.
8. Jiang, J., H. Zheng, X. Ji, T. Cheng, Y. Tian, Y. Zhu, W. Cao, R. Ehsani, and X. Yao, Analysis and evaluation of the image preprocessing process of a six-band multispectral

camera mounted on an unmanned aerial vehicle for winter wheat monitoring. Sensors, 2019. **19**(3): p. 747.

9. Lee, S., D. Lee, P. Choi, and D. Park, Accuracy–power controllable LiDAR sensor system with 3D object recognition for autonomous vehicle. Sensors, 2020. **20**(19): p. 5706.
10. https://www.inpixon.com/technology/standards/ultra-wideband.
11. Oppermann, I., M. Hämäläinen, and J. Iinatti, UWB: theory and applications. 2004: John Wiley & Sons.
12. Lydon, D., M. Lydon, S. Taylor, J.M. Del Rincon, D. Hester, and J. Brownjohn, Development and field testing of a vision-based displacement system using a low cost wireless action camera. Mechanical Systems and Signal Processing, 2019. **121**: p. 343-358.
13. Melotti, G., C. Premebida, N.M.d.S. Gonçalves, U.J. Nunes, and D.R. Faria. Multimodal CNN pedestrian classification: a study on combining LIDAR and camera data. in 2018 21st International Conference on Intelligent Transportation Systems (ITSC). 2018. IEEE.
14. Buch, N., S.A. Velastin, and J. Orwell, A review of computer vision techniques for the analysis of urban traffic. IEEE Transactions on intelligent transportation systems, 2011. **12**(3): p. 920-939.
15. De Raeve, N., M. de Schepper, J. Verhaevert, P. Van Torre, and H. Rogier, A bluetooth-low-energy-based detection and warning system for vulnerable road users in the blind spot of vehicles. Sensors, 2020. **20**(9): p. 2727.

16. Campbell, J., R. Sukthankar, and I. Nourbakhsh. Techniques for evaluating optical flow for visual odometry in extreme terrain. in 2004 IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS)(IEEE Cat. No. 04CH37566). 2004. IEEE.

17. Su, Z., X. Zhong, G. Zhang, Y. Li, X. He, Q. Wang, Z. Wei, C. He, and D. Li, High sensitive night-time light imaging camera design and in-orbit test of Luojia1-01 satellite. Sensors, 2019. **19**(4): p. 797.

18. Lim, J., J. Seo, and Y. Baek. Camthings: Iot camera with energy-efficient communication by edge computing based on deep learning. in 2018 28th International Telecommunication Networks and Applications Conference (ITNAC). 2018. IEEE.

19. Muehlfellner, P., P. Furgale, W. Derendarz, and R. Philippsen. Evaluation of fisheye-camera based visual multi-session localization in a real-world scenario. in 2013 IEEE Intelligent Vehicles Symposium (IV). 2013. IEEE.

20. Zhu, J., J. Zhu, X. Wan, and C. Xu. Downside Hemisphere Object Detection and Localization of MAV by Fisheye Camera. in 2018 15th International Conference on Control, Automation, Robotics and Vision (ICARCV). 2018. IEEE.

21. Cui, Z., L. Heng, Y.C. Yeo, A. Geiger, M. Pollefeys, and T. Sattler. Real-time dense mapping for self-driving vehicles using fisheye cameras. in 2019 International Conference on Robotics and Automation (ICRA). 2019. IEEE.

22. Zhang, Y., D. Yao, T.Z. Qiu, L. Peng, and Y. Zhang, Pedestrian safety analysis in mixed traffic conditions using video data. IEEE Transactions on Intelligent Transportation Systems, 2012. **13**(4): p. 1832-1844.

23. Hashmi, M.F. and A.G. Keskar. Analysis and monitoring of a high density traffic flow at T-intersection using statistical computer vision based approach. in 2012 12th International Conference on Intelligent Systems Design and Applications (ISDA). 2012. IEEE.

24. Sadruddin, H., A. Mahmoud, and M. Atia. An Indoor Navigation System using Stereo Vision, IMU and UWB Sensor Fusion. in 2019 IEEE SENSORS. 2019. IEEE.

25. Wörmann, J., U. Jagdhold, E.R. Bammidi, and I. Kallfass. Receiver Synchronization of Ultra-Wideband Phase Modulated Signals with a Fully Analog QPSK Costas Loop. in 2022 14th German Microwave Conference (GeMiC). 2022. IEEE.

26. Cheng, Y. and T. Zhou. UWB indoor positioning algorithm based on TDOA technology. in 2019 10th international conference on information technology in medicine and education (ITME). 2019. IEEE.

27. Alarifi, A., A. Al-Salman, M. Alsaleh, A. Alnafessah, S. Al-Hadhrami, M.A. Al-Ammar, and H.S. Al-Khalifa, Ultra wideband indoor positioning technologies: Analysis and recent advances. Sensors, 2016. **16**(5): p. 707.

28. Zhao, M., T. Chang, A. Arun, R. Ayyalasomayajula, C. Zhang, and D. Bharadia, ULoc: Low-power, scalable and cm-accurate UWB-tag localization and tracking for indoor applications. Proceedings of the ACM on Interactive, Mobile, Wearable and Ubiquitous Technologies, 2021. **5**(3): p. 1-31.

29. Djahel, S., R. Doolan, G.-M. Muntean, and J. Murphy, A communications-oriented perspective on traffic management systems for smart cities: Challenges and innovative approaches. IEEE Communications Surveys & Tutorials, 2014. **17**(1): p. 125-151.

30. Liu, F., Z. Zeng, and R. Jiang, A video-based real-time adaptive vehicle-counting system for urban roads. PloS one, 2017. **12**(11): p. e0186098.

31. Kwon, S.K., E. Hyun, J.-H. Lee, J. Lee, and S.H. Son. A low-complexity scheme for partially occluded pedestrian detection using LiDAR-radar sensor fusion. in 2016 IEEE 22nd International Conference on Embedded and Real-Time Computing Systems and Applications (RTCSA). 2016. IEEE.

32. Wu, T.-E., C.-C. Tsai, and J.-I. Guo. LiDAR/camera sensor fusion technology for pedestrian detection. in 2017 Asia-Pacific Signal and Information Processing Association Annual Summit and Conference (APSIPA ASC). 2017. IEEE.

33. Ishikawa, R., T. Oishi, and K. Ikeuchi. Lidar and camera calibration using motions estimated by sensor fusion odometry. in 2018 IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS). 2018. IEEE.

34. Han, X., J. Lu, Y. Tai, and C. Zhao. A real-time lidar and vision based pedestrian detection system for unmanned ground vehicles. in 2015 3rd IAPR Asian Conference on Pattern Recognition (ACPR). 2015. IEEE.

35. Kidono, K., T. Naito, and J. Miura. Reliable pedestrian recognition combining high-definition lidar and vision data. in 2012 15th international IEEE conference on intelligent transportation systems. 2012. IEEE.

36. Streubel, R. and B. Yang. Fusion of stereo camera and MIMO-FMCW radar for pedestrian tracking in indoor environments. in 2016 19th International Conference on Information Fusion (Fusion). 2016. IEEE.

37. Siegwart, R., I.R. Nourbakhsh, and D. Scaramuzza, Introduction to autonomous mobile robots. 2011: MIT press.

38. Gentili, M. and P.B. Mirchandani, Locating active sensors on traffic networks. Annals of Operations Research, 2005. **136**(1): p. 229-257.

39. Kamal, M.A.S., J.-i. Imura, T. Hayakawa, A. Ohata, and K. Aihara, A vehicle-intersection coordination scheme for smooth flows of traffic without using traffic lights. IEEE Transactions on Intelligent Transportation Systems, 2014. **16**(3): p. 1136-1147.

40. Wang, X., W. Zhang, X. Wu, L. Xiao, Y. Qian, and Z. Fang, Real-time vehicle type classification with deep convolutional neural networks. Journal of Real-Time Image Processing, 2019. **16**(1): p. 5-14.

41. Fayyad, J., M.A. Jaradat, D. Gruyer, and H. Najjaran, Deep learning sensor fusion for autonomous vehicle perception and localization: A review. Sensors, 2020. **20**(15): p. 4220.

42. Sasiadek, J.Z., Sensor fusion. Annual Reviews in Control, 2002. **26**(2): p. 203-228.

43. Kim, B. and K. Yi, Probabilistic and holistic prediction of vehicle states using sensor fusion for application to integrated vehicle safety systems. IEEE Transactions on Intelligent Transportation Systems, 2014. **15**(5): p. 2178-2190.

44. Bresson, G., Z. Alsayed, L. Yu, and S. Glaser, Simultaneous localization and mapping: A survey of current trends in autonomous driving. IEEE Transactions on Intelligent Vehicles, 2017. **2**(3): p. 194-220.

45. Eckelmann, S., T. Trautmann, H. Ußler, B. Reichelt, and O. Michler, V2v-communication, lidar system and positioning sensors for future fusion algorithms in connected vehicles. Transportation research procedia, 2017. **27**: p. 69-76.

46. Chang, T.-H. and G.-Y. Sun, Modeling and optimization of an oversaturated signalized network. Transportation Research Part B: Methodological, 2004. **38**(8): p. 687-707.

47. Ran, C.-J. and Z.-L. Deng. Two average weighted measurement fusion Kalman filtering algorithms in sensor networks. in 2008 7th World Congress on Intelligent Control and Automation. 2008. IEEE.

48. Wu, D., L. Xia, and J. Geng, Heading estimation for pedestrian dead reckoning based on robust adaptive Kalman filtering. Sensors, 2018. **18**(6): p. 1970.

49. Olfati-Saber, R. and J.S. Shamma. Consensus filters for sensor networks and distributed sensor fusion. in Proceedings of the 44th IEEE Conference on Decision and Control. 2005. IEEE.

50. Das, T.K., P.D. Harischandra, and A.H.S. Abeykoon. Extended Kalman Filter based fusion of reliable sensors using fuzzy logic. in 2017 Moratuwa Engineering Research Conference (MERCon). 2017. IEEE.

51. Pires, I.M., N.M. Garcia, N. Pombo, and F. Flórez-Revuelta, From data acquisition to data fusion: a comprehensive review and a roadmap for the identification of activities of daily living using mobile devices. Sensors, 2016. **16**(2): p. 184.

52. Yang, H., Q. Ma, Z. Wang, Q. Cai, K. Xie, and D. Yang, Safety of micro-mobility: Analysis of E-Scooter crashes by mining news reports. Accident Analysis & Prevention, 2020. **143**: p. 105608.

53. Kalman, R.E., A new approach to linear filtering and prediction problems. 1960.

54. Welch, G.F., Kalman filter. Computer Vision: A Reference Guide, 2020: p. 1-3.

55. Gai, Y., W. He, and Z. Zhou. Pedestrian Target Tracking Based On DeepSORT With YOLOv5. in 2021 2nd International Conference on Computer Engineering and Intelligent Control (ICCEIC). 2021. IEEE.