# Anonymous Indoor Positioning System using Depth Sensors for Context-aware Human-Building Interaction

Sergio Marlon Ballivian Amaya

Thesis submitted to the faculty of the Virginia Polytechnic Institute and State University in partial fulfillment of the requirements for the degree of

Master of Science
in
Civil Engineering

Farrokh Jazizadeh, Chair
Jesus M. de la Garza
Tanyel Bulbul

February 21st, 2019
Blacksburg, VA

Keywords: Sensing, machine learning, Human-technology frontier, Data modeling and computing, Sustainable urban and energy systems, Human-building interaction

Anonymous Indoor Positioning System using Depth Sensors for Context-aware Human-Building Interaction

Sergio Marlon Ballivian Amaya

**ABSTRACT**

Indoor Localization Systems (ILS), also known as Indoor Positioning Systems (IPS), has been created to determine the position of individuals and other assets inside facilities. Indoor Localization Systems have been implemented for monitoring individuals and objects in a variety of sectors. In addition, ILS could be used for energy and sustainability purposes. Energy management is a complex and important challenge in the Built Environment. The indoor localization market is expected to increase by 33.8 billion in the next 5 years based on the 2016 global survey report (Marketsandmarkets.com).

Therefore, this thesis focused on exploring and investigating “depth sensors” application in detecting occupants’ indoor positions to be used for smarter management of energy consumption in buildings. An interconnected passive depth-sensor-based system of occupants’ positioning was investigated for human-building interaction applications. This research investigates the fundamental requirements for depth-sensing technology to detect, identify and track subjects as they move across different spaces. This depth-based approach is capable of sensing and identifying individuals by accounting for the privacy concerns of users in an indoor environment. The proposed system relies on a fixed depth sensor that detects the skeleton, measures the depth, and further extracts multiple features from the characteristics of the human body to identify them through a classifier. An example application of such a system is to capture an individuals’ thermal preferences in an environment and deliver services (targeted air conditioning) accordingly while they move in the building.

The outcome of this study will enable the application of cost-effective depth sensors for identification and tracking purposes in indoor environments. This research will contribute to the feasibility of accurate detection of individuals and smarter energy management using depth sensing technologies by proposing new features and creating combinations with typical biometric features. The addition of features such as the area and volume of human body surface was shown to increase the accuracy of the identification of individuals. Depth-sensing imaging could be

combined with different ILS approaches and provide reliable information for service delivery in building spaces. The proposed sensing technology could enable the inference of people location and thermal preferences across different indoor spaces, as well as, sustainable operations by detecting unoccupied rooms in buildings.

Anonymous Indoor Positioning System using Depth Sensors for Context-aware Human-Building Interaction

Sergio Marlon Ballivian Amaya

## GENERAL AUDIENCE ABSTRACT

Although Global Positioning System (GPS) has a satisfactory performance navigating outdoors, it fails in indoor environments due to the line of sight requirements. Physical obstacles such as walls, overhead floors, and roofs weaken GPS functionality in closed environments. This limitation has opened a new direction of studies, technologies, and research efforts to create indoor location sensing capabilities. In this study, we have explored the feasibility of using an indoor positioning system that seeks to detect occupants' location and preferences accurately without raising privacy concerns. Context-aware systems were created to learn dynamics of interactions between human and buildings, examples are sensing, localizing, and distinguishing individuals. An example application is to enable a responsive air-conditioning system to adapt to personalized thermal preferences of occupants in an indoor environment as they move across spaces. To this end, we have proposed to leverage depth sensing technology, such as Microsoft Kinect sensor, that could provide information on human activities and unique skeletal attributes for identification.

The proposed sensing technology could enable the inference of people location and preferences at any time and their activity levels across different indoor spaces. This system could be used for sustainable operations in buildings by detecting unoccupied rooms in buildings to save energy and reduce the cost of heating, lighting or air conditioning equipment by delivering air-conditioning according to the preferences of occupants.

This thesis has explored the feasibility and challenges of using depth-sensing technology for the aforementioned objectives. In doing so, we have conducted experimental studies, as well as data analyses, using different scenarios for human-environment interactions. The results have shown that we could achieve an acceptable level of accuracy in detecting individuals across different spaces for different actions.

# ACKNOWLEDGEMENTS

I would like to thank God for being a part of each stage of my life; to my dad, mom and brother who trusted and supported each of my decisions; family, girlfriend and friends in and outside of the US, as well as teachers and labmates who helped on my education. Specially, Dr. Farrokh Jazizadeh, my advisor, teacher, and friend who believed in me from the beginning. He has been a source of inspiration in pursuing and achieving a Master's degree in the Civil and Environmental Engineering Department. Thank you to my committee members, Dr. Jesus de la Garza and Dr. Tanyel Bulbul who supported and approved my research. In addition, I would like to thank my assistantship program, New Horizons Scholars and GEM that supported my intellectual and personal learning.

# TABLE OF CONTENTS

# INTRODUCTION

Indoor Localization Systems demonstrated the importance of positioning for the clients and organizations inside building spaces. Promising results in health care, construction industry, and many service enterprises have lead ILS to grow remarkably in the near future. The depth-sensing technology could enable the understanding where people are located anytime and the activities performed between indoor spaces.

Indoor positioning technologies have focused on becoming more carrier-free, less intrusive, and avoiding privacy concerns to develop a system capable of localizing individuals for service delivery. Depth sensors present the capabilities of the vision-based solutions in being passive (i.e., independent of carrying a receiver device) while cost-effective without violating users' privacy comparing to RGB images. Depth sensors use the projection of irregular pattern dots by the IR camera to measure the distance from objects in their field of view rendering a silhouette depth image that enables image segmentation for object detection without revealing details that an RGB image could contain. The distance between the object and the sensor has limited the vision-based methods to determine its location. The depth information is an essential attribute to determine the location of subjects or objects. Therefore, by knowing the location of the sensor not only the occupied room is identified, but also the absolute location of the occupants could be inferred. Tracking the subjects calls for identification of the subjects moving across different spaces.

The need for tracking individuals or objects in indoor environments has created a spectrum of new indoor location technologies among interdisciplinary applications. Examples of tracking system applications are navigation, automation, robotics, management, and disaster response. ILS is widely applicable in a series of services such as retail, healthcare centers, airport terminals, enterprises, and the construction industry. Its' capability to facilitate navigation and thus, automation is the main reasons why ILS has become a widespread topic in the former category. Walking towards your desk in an office building, delivering information to personnel located on the construction site, finding the correct workroom at the new building are examples of such facilitation. The interest in positioning systems is well defined by both parties in the facilities operations– i.e., customers and service providers. Customers are more productive in their daily activities while firms provide better service-delivery strategies and increase their revenues. In addition, ILS has been applied in the built environment monitoring with promising results in the management sector, as well as, in the automation in the construction sector. Also, the application of depth based technology could further be applied to the sustainability; detecting unoccupied rooms in buildings to save energy in the built environment and reduce the cost of heating, lighting or air conditioning equipment. Positioning techniques were employed in buildings to create a smart environment for energy automation for indoor environments such by curtailing the unnecessary operation of lighting and air conditioning systems on unavailable meeting rooms in the building. The Department of Energy has invested almost $20 million on various projects involving sensor-based technology to improve building efficiency in the United States (Energy.gov 2018).

Localization refers to the one-time estimation of an individual or object position in space while tracking is the trajectory (location path over a duration of time) of an identified individual or object. The positioning and tracking of subjects also bring about additional potential applications in an environment such as activity monitoring, context-aware information delivery, and material tracking which could render a context-aware technology for service management in buildings.

This research work has been reflected in two studies which concentrate on the feasibility and the real-world implementation of the proposed depth-sensor-based system. Each study includes an experimental procedure of identification and/or tracking on several human subjects to show the applicability of the proposed method.

Given the importance of cross-space identification under privacy concerns, this study focuses on enabling depth sensors for cross-space and anonymous human subject identification. Anonymous identification refers to identification across sensors without the need for knowing the real-world identity of individuals. In other words, an anonymous identity will be assigned to an individual after observing them for the first time and other sensors will seek to infer that identity. Therefore, the main research question in this thesis was: How can we enable the depth-sensing technology to achieve passive data collection for identification and tracking individuals moving between indoor spaces?

## OBJECTIVES OF RESEARCH

The envisioned anonymous indoor positioning system will enable building systems to track individuals for service delivery (i.e., HVAC). For example, an application of such an IPS is to capture an individuals' thermal preferences in an indoor environment and deliver services accordingly while they move in the building and according to occupant's location.

The research goal is to study anonymous cross-space identification of occupants. The objectives in this thesis was to explore and investigate the fundamental requirements for feasibility of using "depth sensors" in detecting occupants' indoor positions to be used for smarter management of energy consumption in buildings. An interconnected passive depth-sensor-based system of occupants' position was investigated for human-building interaction.

Therefore, the two major objectives include:

1) Investigating the feasibility of occupants' anonymous identification in the same space by assessing different biometric features and classification algorithms under the common constraints in the indoor environment such as occlusions and variations in human posture.

2) Investigating the impact of varied interaction scenarios. In other words, how the performance could be affected when occupants move across spaces and the context of identification changes.

## LITERATURE REVIEW

In human-building interaction (HBI) studies, understanding the dynamics of interactions is a critical step towards context-aware service delivery. The capability of location identification in indoor environments is one of the critical information components that enable a variety of HBI applications. Beyond conventional services such as navigation and information delivery, examples of such applications include enabling adaptive and responsive ambient conditions, as well as supportive environments that boost the performance of occupants. Therefore, technologies that provide accurate and seamless occupants' location information have been the subject of research in the past few decades. Indoor positioning systems (IPS) is a well-established and well-studied field of research across different engineering disciplines with available commercial solutions.

A variety of technologies have been proposed with varying levels of trade-off between accuracy and instrumentation complexity. The resolution of location information and the requirements for acquiring that information could impact the use cases for location information. The systems for indoor location sensing could be divided into two generic categories of active and passive. Active systems are similar to GPS solutions that require a receiver, such as smart portable or wearable devices or tags, which are carried by the people or installed on assets. On the other hand, passive systems do not require a receiver and they leverage contact-free interaction for data acquisition without the subject awareness or active engagement.

The systems for localizing individuals are divided into two alternatives: active and passive systems. The most commonly used technologies in the active category use radio frequency (RF-) based methods (e.g., (Bolic et al. 2010; Koyuncu and Yang 2010; Ni et al. 2004)), which rely on transmitters and receivers to locate the objects. Examples include IPS techniques that use wireless networks (WLAN), RFIDs, or Ultrawideband RFs by leveraging Received Signal Strength Indicator (RSSI) (or other parameters such as time of flight) and triangulation or trilateration techniques. Survey on research studies of the RF-based methodologies show that these systems could achieve localization in indoor environments with accuracies ranging between 1 to 10 m (Koyuncu and Yang 2010; Mainetti et al. 2014), which depends on the contextual conditions such as the resolution of the instrumentation and therefore could show limitations in identification of subspaces in indoor environments. Similarly, acoustic ultrasound-based methods have been proposed that use a network of beacons in a space and a mobile device or tag for detecting the location of the objects. The ultrasonic methods have shown to have high accuracies, however, they still call for relatively dense transmitter implementation (Hazas and Ward 2002; Saad et al. 2012). Except for WLAN methods, active methods commonly call for specialized receivers that could pose limitations on their application for HBI. Moreover, these methods could be susceptible to signal interference in an environment (Bolic et al. 2010). The deployment of the equipment needed between rooms makes its implementation problematic and laborious (Hightower and Borriello 2001). Moreover, they call for transceivers that are carried by users. The wireless local area network (WLAN) which is a device-based technology categorized in the active system is reasonable to be implemented where room-level accuracy or less is appropriate such as airports and retail stores. The satisfactory performance of the system is a room-scale in these scenarios.

On the other hand, passive systems are independent of wearable tags or mobile devices and leverage contact-free interaction with the objects for data acquisition. Computer-vision-based and motion sensor-based technologies are examples of passive systems. The purpose of these technologies is to detect moving individuals without their active engagement while preserving the accuracy. Motion sensor-based techniques that use vibration or pressure sensors (Poston et al. 2015) are limited to specific use cases and their application for HBI purposes could be challenging. Vision-based methods use data acquisition technologies such as RGB (Red-Green-Blue) image sensors or infrared thermal cameras either as fixed or mobile camera systems (Mainetti et al. 2014). The fixed cameras track objects in the image and use feature analysis in different scenes for detecting the location information. The mobile cameras use targets in an environment or features from the scene to infer the location information. Despite the privacy concerns that fixed camera system could arise, the recent developments in high precision sensors and efficient image processing algorithms have brought about high accuracy location information.

Vision-based approaches can retrieve valuable information from moving humans using different types of images (i.e. RGB, thermal). Physical characteristics (i.e. appearance, face recognition) are common features extracted for identification of subjects using vision-based methods. A number of studies have focused on using vision-based methods for human identification. In order to recognize individuals, different appearance features such as facial biometrics, clothing, hair color, and ethnicity have been employed (Chen et al. 2012; Dantcheva et al. 2010; Demirkus et al. 2010; Jaha and Nixon 2014). Computer vision systems could be installed in existing facilities, while some motion sensor-based techniques such as vibration and pressure (Poston et al. 2015) require walking force sensing components installation on the surface (tiles) during construction of the building.

Active technologies are susceptible to signal loss or interference with obstacles in their path and unreliable connections to the network. Although RF-based methodologies have successfully contributed to occupants tracking location accurately, they require rather dense deployment of transmitters in indoor environments, wearable devices, and attenuations at different locations in a space (Bolic et al. 2010). Due to the usage of RGB images, they could raise privacy concerns for the users. Privacy is considered a major concern for the selection of ILS or identification from user's point of view (Basiri et al. 2015). Extensive training, privacy issues, intrusiveness, and being cost-intensive (in case of thermal cameras) are common drawbacks to localize subjects in room spaces using passive technologies.

**Human-centered depth imaging technology**

Recent studies have demonstrated the feasibility of using depth sensors in a variety of applications such as surveillance (Albiol et al. 2012; del-Blanco et al. 2014), robotics (El-laithy et al. 2012; Stoyanov et al. 2013), indoor mapping (Khoshelham and Khoshelham 2012), occupancy detection (Diraco et al. 2015; Shih 2014), 3D pose recognition (Shotton et al. 2011), and human activity analysis in general (Han et al. 2013).

Depth sensors are categorized as passive methods, in which people are not actively involved in the localization process. In other words, depth sensors do not require people or objects to wear/carry devices. Depth sensors provide the third dimension missing from RGB images taken from the conventional video cameras without additional equipment, tags, or infrastructure. Depth sensors use the projection of irregular pattern dots by the IR camera to measure the distance from objects in their field of view rendering a silhouette depth image that enables image segmentation for object detection without revealing details that an RGB image could contain. Depth information will provide location information relative to the position of the depth sensors in an environment with sub-decimal accuracy (Khoshelham and Khoshelham 2012).

The identification process consists of extracting biometrics from individuals by the system and evaluate them using a supervised machine learning algorithm. The most common passive method to re-identify individuals is using human characteristics or features called biometrics. Biometric identification systems have been studied extensively in the computing disciplines. It refers to the measurements of the human body to determine certain unique characteristics or features of individuals. While depth sensors provide information, the data source is similar to point cloud and does not provide explicit identifying information (unlike RGB images) which alleviates the privacy concerns. The extracted information from the depth sensor in one space will be interconnected to every other sensor in adjacent spaces. The link between devices will assist in tracking individuals as they move through different areas.

**Depth-Sensor-Based Indoor Tracking**

Depth sensors have presented the potential to determine the location of individuals in a passive and cost-effective way. The capabilities of depth sensors in estimating the depth in a scene without the need for optimum ambient conditions renders them as a reliable tool for measuring the relative location of subjects in a scene. Knowing the location of the sensor enables of the identification of the occupied room as well as the absolute location of the occupants. The use of depth images for tracking individuals in indoor environments through range imaging has been explored in a number of research efforts, either through stereo camera or depth sensor technologies. The system in (Tao et al. 2005) required 12 pairs of stereo cameras for tracking individuals in 3 separate spaces. In (Bevilacqua et al. 2006) have proposed the Time of Flight depth sensors for people detection and tracking in different lighting conditions avoiding occlusion by placing them on the ceiling. The drawbacks found in these studies are the number of devices installed in the ceiling and tracking people in one single space. A robust system should be able to track individuals moving from one room to another without implementing dozens of cameras for this purpose.

Several visual modalities have been implemented for people tracking. In (Darrell et al. 2000) we see the fusion between depth sensing methods with video cameras to propose a more robust system. Stereo cameras detect independent blobs representing individuals. Then, the RGB image analyses the individual’s skin to track body parts. This method implies the detection of individuals face at all times which may be occluded in the room and the use of RGB images could violate user’s privacy.

An integrated (hybrid) approach found in (Duque Domingo et al. 2016) proposed a new ILS based on RGB-D sensor and WLAN using smartphones as a receiver to determine individuals’ location. Hybrid systems are the combination of different technologies to improve the performance of the overall system. As mentioned before, WLAN is categorized as an active system which requires a device, in this case, smartphone will be the receiver dependable on energy source. Also, WLAN requires implementation of additional equipment and creation of user's position database as part of the training stage to assist to the skeletal and depth map information from the RGB-D sensor.

**Depth-Sensor-Based Identification**

Using depth information from the depth sensor preserves individuals' privacy. A number of studies have focused on using skeletal information from depth sensors for human identification. These studies leveraged several attributes (i.e., features) that characterize the identity of an individual, coupled with supervised learning algorithms, specifically classifiers, to distinguish him/her from others for identification.

Static features have been comprehensively used in a number of research studies for human identification. In this context, static features refer to the measurement obtained by the distance between adjacent human joints (e.g., height; the length of torso, arms, and legs; the width of shoulders, and waist). Static features have been used with classification algorithms such as K-Nearest Neighbor (KNN), Multi-Layer Perceptron (MLP), Random Forests (Araujo et al. 2013), Support Vector Machine (SVM), Naïve Bayes (Sinha et al. 2013). In addition to static features, frontal pose features (i.e., area occupied in the upper body, and lower body) (Chakravarty and Chattopadhyay 2014) and gait-based identification (Sinha and Chakravarty 2013) have been used with promising results for human identification using SVM classifiers and Adaptive Neural Network (ANN) algorithm but obtaining a relatively lower accuracy (86%). Nonetheless, the aforementioned studies employed skeletal information to identify subjects and did not investigate other potential application such as localization and tracking between spaces. While alternative studies such as (Xia et al. 2011) have used depth sensor (such as Microsoft Kinect sensor) for both identification and tracking, they relied on partial data from human attributes such human head surface model and evaluated their model on two individuals. In Table 1, the characteristics of several related studies are summarized.

Table 1. Depth-sensor-based human identification background

| ***Feature Extraction*** | | ***No. of people*** | ***Classification Algorithm*** | ***Performance (%)*** [1]***Accuracy,*** [2]***F-Score*** | ***Reference*** |
|---|---|---|---|---|---|
| Static | *Height*<br>*Thighs and lower legs*<br>*Upper and forearms*<br>*Torso* | *8* | *MLP*<br>*C4.5*<br>*Random Forests*<br>*KNN* | *99.6*[1] | *(Araujo et al. 2013)* |
| Static<br><br>Area | *Height*<br>*Thighs and lower legs*<br>*Upper and forearms*<br>*Angles*<br>*Front torso area* | *10* | *SVM* | *95.8*[2] | *(Chakravarty and Chattopadhyay 2014)* |
| Area | *Upper and lower body areas* | *5*<br>*10* | *ANN*<br>*Naive Bayes* | *86.0*[2]<br>*55.0*[2] | *(Sinha et al. 2013)* |
| Surface | *3D head surface model* | *2* | *2D chamfer distance* | *98.4*[1] | *(Xia et al. 2011)* |

**STUDY 1: Anonymous Indoor Positioning of Occupants for Human-Building Interaction using Depth Sensors**

Authors:

Sergio Ballivian[1]

Farrokh Jazizadeh[2]

[1] Via Department of Civil and Environmental Engineering, Virginia Tech, 200 Patton Hall, 750 Drillfield Dr., Blacksburg, VA 24061. E-mail: sergiob5@vt.edu

[2] Via Department of Civil and Environmental Engineering, Virginia Tech, 200 Patton Hall, 750 Drillfield Dr., Blacksburg, VA 24061. E-mail: jazizadeh@vt.edu

# 1. INTRODUCTION

In this study, we have investigated the feasibility of an envisioned passive indoor positioning system based on depth sensors that facilitate adaptive and responsive indoor environment for HBI use cases. Due to their nature, HBI applications call for high resolution location information under the constraints of privacy. An example envisioned application is the provision of a local ambient condition to account for personalized thermal and visual preferences in a building as a service and for efficient energy management. In such application, a building system needs to be aware of the high-fidelity location information and preferences of the individuals and be able to track their movements in buildings. By using fixed depth sensors, installed across different sub-spaces (ideally one per room/sub-space) inside the building the IPS system acquire information from subjects moving between different spaces. In addition to measuring the distance of objects with respect to the sensor, depth sensors provide depth images that enable object recognition. The information is centrally shared across sensors to enable tracking occupants as they move through different areas. In the passive mode, tracking calls for identification of the subjects across different spaces. The identification and tracking of subjects also bring about additional potential applications in an environment such as activity monitoring, which could render this approach to a context-aware technology for service management in buildings. Figure 1 illustrates the schematic characteristics of the envisioned IPS system for HBI.

To this end, the rest of this paper is structured as follows. Section 2 presents an overview of existing studies on tracking and identification of objects via depth sensors. We display our methodology in a framework form in Section 3. In Section 4 the experimental set up is presented. Then, Section 5 covers the data analysis and results of the experimental study. Finally, we presented the limitations, conclusions and future directions in Section 6.

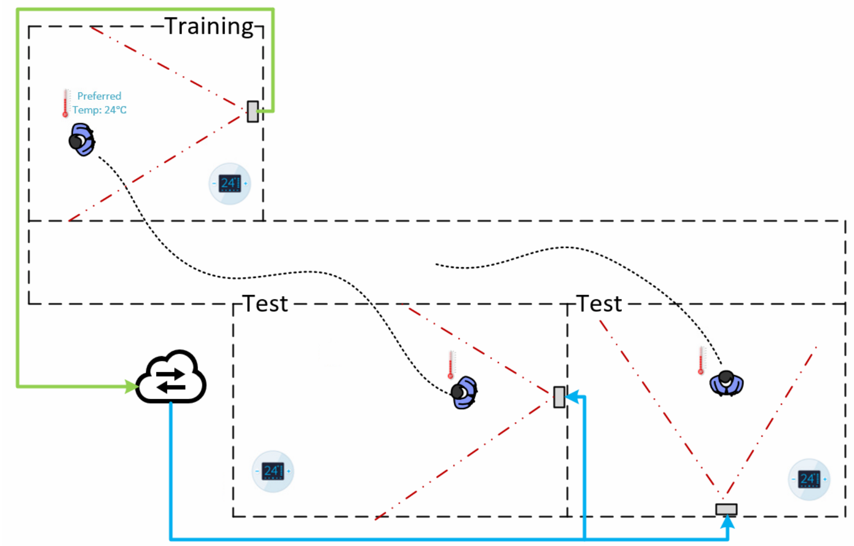



Figure 1. Schematic characteristics of the envisioned depth-sensor-based IPS for HBI applications

## 2. OBJECTIVES

Given that the location information is a direct output of the depth sensors, in this study we have focused on exploring the fundamental requirements of the cross-space anonymous identification of human subjects using depth sensors. Anonymous identification refers to identification across sensors without the need for knowing the real-world identity of individuals. In-depth images, humans could only be observed in the form of silhouettes and therefore the identification is considered anonymous, unlike RGB cameras that could reveal the actual identity of individuals. In other words, an anonymous identity (e.g., a key-value pair of index and preferences) will be assigned to an individual after observing them for the first time and sensors in other spaces will seek to infer that identity. Therefore, the main objective of the research is to see what features could be used, coupled with classifiers, to enable high accuracy anonymous identification of occupations. To this end, one of the research questions includes what features will result in improved performance. Due to naturalistic interactions between users and indoor environments, occlusions and varied posture in human body forms will be unavoidable. As a result, we have further investigated the impact of the occlusion and posture on the accuracy of the inference algorithms.

# 3. METHODOLOGY

## 3.1. Depth-Sensor-Based Indoor Localization

In this study, as noted, we are evaluating a depth-sensor based ILS system for tracking and identifying individuals in indoor environments. Several experiments were conducted on several human subjects as they walked in our data sensing set up. The envisioned system will be divided into different experimental scenarios to answer my research questions. The inference framework for identification and tracking individuals is depicted in Figure 2.

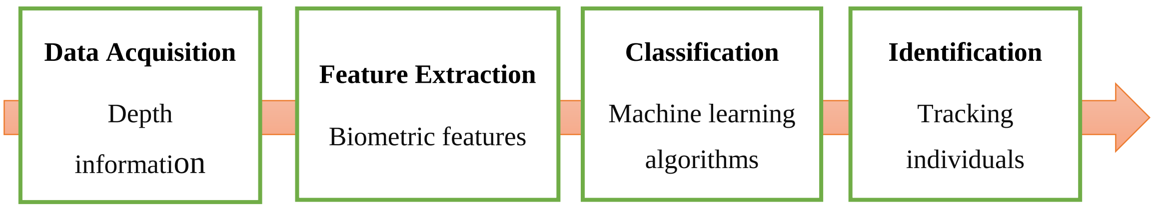



Figure 2. Process mapping of the proposed method

## 3.2. Data Acquisition

### 3.2.1. Kinect sensor

The Microsoft Kinect sensor V2 consists of a built-in RGB camera, set of 3D depth sensors, and microphones. The 3D depth sensors involve the IR projector and camera found in the middle of the device. In this study, the sensor commanded the data extraction and localization (horizontal and vertical coordinates from the depth sensor's reference as well as the depth information) using depth images to preserve user's privacy. The depth images, unlike the RGB video cameras, collected individual's characteristics using the biometric measurements from the IR camera. Therefore, the biometric feature extraction does not require RGB images. The Microsoft Kinect depth sensor can extract the skeletal data from the individual detected within 1.2 to 4.5 m. and depth images up to 8 meters from the sensor. The data extracted are the spatial coordinates of 25 joints of the human body.

# 4. EXPERIMENTAL DESIGN

## 4.1. Skeleton and Depth Information

The data acquisition captures depth frames from individuals walking in front of a depth sensor. In this work, I have used the Kinect SDK 2.0 for the data acquisition. The experimental testbed was a room 7 m long and 3.2 m wide, equipped with an off-the-shelf depth sensor (i.e., Kinect for Windows V2 sensor) located at one side of the room. Leveraging the Kinect Software Development Kit (SDK), the skeletal data (i.e., spatial coordinates of 25 joints of the human body) at 30 frames per second were collected. The Microsoft Kinect sensor was placed one meter above ground. Figure 3 shows the a) joint map from the Kinect sensor and b) diagram of experimental set up in this study.

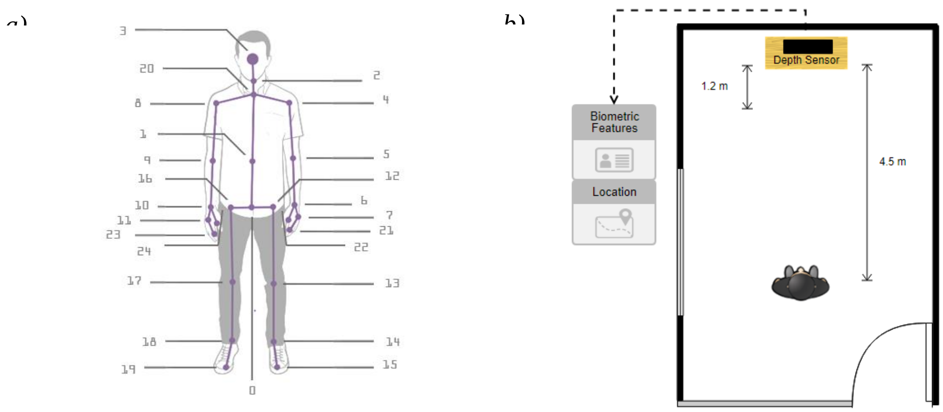



Figure 3. a) Skeleton showing 25 joints position ("Kinect for Windows SDK", 2018). b) Experimental setup

### 4.2. Participants

We selected 50 individuals in this experiment. The participants' age ranged from 20 to 30 years old. Each participant provided their height and weight shown in Figure 4. In order to ensure the performance of the system is evaluated in challenging cases, not only distinct individuals will be part of the selection, but I also consider individuals with similar physical features that result in analogous attributes as well. We want to evaluate the system considering the influence of individual and combined features.

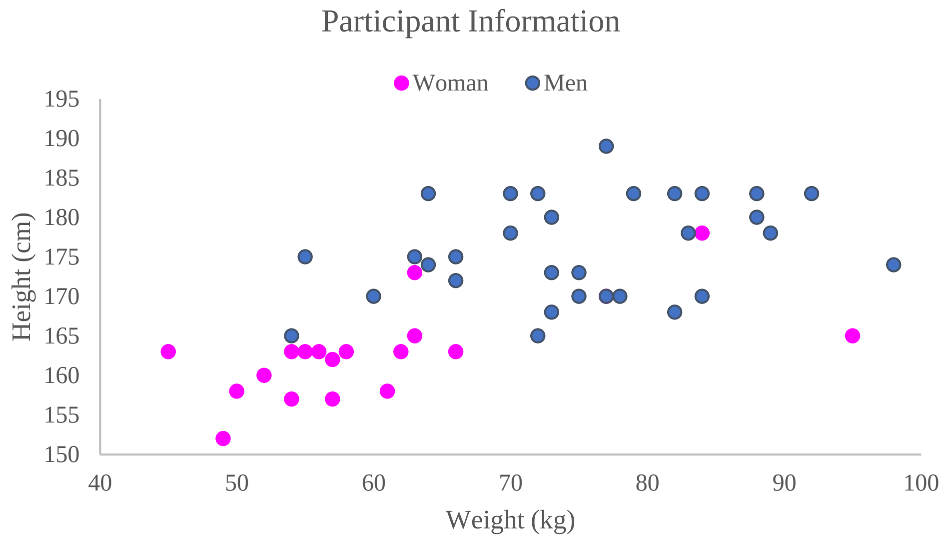



Figure 4. Height and weight from each participant.

### 4.3. Feature Extraction

The biometric attributes (i.e., features) were captured at each frame from depth and silhouette image, individual's skeleton. In order to obtain biometric features from the skeleton, using the Euclidean distance between the joints, the following attributes were calculated for each individual at each frame: height; the length of torso, arms, and legs; the width of shoulders, and waist. In addition, to benefit from unique features that depth images could provide, we proposed to use the following additional biometric features: the body surface volume as well as the silhouette area. Hereafter, we call these features static (representing skeletal information) and profile features (representing the body surface). Table 2 illustrates the biometric features collected in this study.

Table 2. Static and body surface biometric features extracted from individuals.

| ***Biometric features*** | | |
|---|---|---|
| Static | Length | Height, Arms, Torso, Legs |
| | Width | Waist, Shoulder to shoulder |
| Profile | Area | Silhouette |
| | Volume | Surface |
| Angles | Angle between arms and torso | |
| Position | Position of the individual from the sensor | |

The Microsoft Kinect sensor consists of a built-in RGB camera, set of 3D depth sensors, and microphones. The 3D depth sensors involve the IR projector and camera found in the middle of the device. In this study, the sensor commanded the data extraction and localization (horizontal and vertical coordinates from the depth sensor's reference as well as the depth information) using depth images to preserve user's privacy. The depth images, unlike the RGB video cameras, collected individual's characteristics using the biometric measurements from the IR camera. Therefore, the biometric feature extraction does not require RGB images. Figure 5 shows the biometric feature extraction of a detected individual using depth images, skeleton, and profile. The surface area and volume of the individuals' profile were calculated by depth data using MATLAB. A mesh was mapped onto the 3D information of each individual. An example of these surface area and volume of an individual's profile mesh graphs is found in Figure 5b.

a) b)

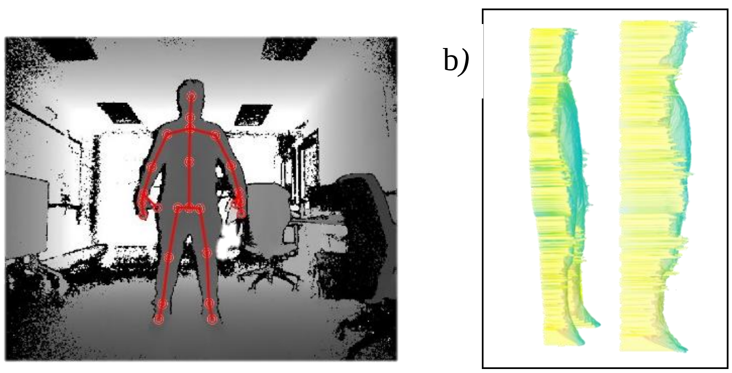


Figure 5. a) Detection of the individual using depth image. b) Individual's profile was shaped using depth data

We have extracted 13 biometric attributes using the data from 50 human participants. In order to remove the outliers from the depth sensor data, a pre-processing step was followed. As for the Kinect depth sensor, the outliers refer to the portion of collected data at timeframes (i.e., observations) when the distance between the human subject and the fixed sensor exceeds 4.5 meters. Therefore, the algorithm assigned zero values to the undetected joint locations of the subject at those observations, and they were removed prior to the analysis by the classifier.

### 4.4. Identity Inference Process

The identity inference process consists of extracting biometrics from individuals by the system and evaluate them using a machine learning algorithm. All the data extracted was passed to a supervised machine learning algorithms which were found repeatedly in the literature. Classification algorithms are supervised learning algorithms that make predictions based on labeled data. The classification using supervised algorithms consists of splitting the dataset into training set and were used: as K-Nearest Neighbor (KNN), Multi-Layer Perceptron (MLP), Random Forests (RF), and Naïve Bayes (NB) as they have generally provided promising performance for identification in the literature. One technique to evaluate the performance of the predictions made by the classification algorithm is known as cross-validation. Cross validation splits up the training set into two-thirds and the other third for forecasting evaluation metrics.

### 4.5. Occlusions

Indoor environments are generally found furnished (i.e., tables, chairs). Therefore, these objects could obstruct or occlude the sensor’s field of view leading to partial data collection. Important features from the individual might be obstructed in the extraction which alters the performance of the system. Also, occlusion could impact the data collection by extracting wrong attributes or measurements. However, the association of all features could improve the identification of individuals. In order to make a more realistic scenario, we have performed an experiment taking into account the partial information extracted from the skeleton and profile data. The occlusion consisted in the removal of inferior biometric information and only consider the upper body information. The labeling of each individual in the dataset remained the same for both situations. The purpose of this experiment is to evaluate the system’s capabilities to recognize or distinguish different individuals in the presence of tables and chairs in the environment. Moreover, the performance of the system to classify individuals with partial and total data information among other individuals. We have gathered a spectrum of indoor environments where the individuals might be occluded immediately as they walk in the room as shown in Figure 6.

Figure 6. Varied occluded indoor setting examples

### 4.6. Posture

Posture refers to the position and interaction of the body parts. We cannot generalize a posture when someone is walking in or out of a room, however the normal or baseline posture will be standing with both arms down as seen in Figure 7. It is not reliable to assume that each individual will react or act the same way as others or the normal or baseline posture. There are many postures that individuals take as they enter to a room; arms crossed, carrying a bag holding a cup, etc. Different postures could also affect the data collection and therefore, the identification of subjects as they walk into an indoor environment. Whether you are holding your coffee, carrying a bag or crossing your arms, the posture of your body has changed from the normal posture (baseline), and could affect the individual identification extracting false measurements. In Figure 7, we can observe the 3 different examples of an individual posture.

We decided to add some noise to the system and observe the accuracy of individual identification. The proposed system should be able to identify this individual even performing different postures other than the initial one. We evaluated the scenario of tracking the individual carrying a bag, holding a cup, and crossing arms. The experiment consists of individuals performing 3 different postures than the generic one; crossing arms, holding a cup and carrying a bag. We have extracted the skeleton and profile information. The purpose of this experiment is to evaluate the system's capabilities to recognize or distinguish individuals performing the same posture entering an indoor environment.

Figure 7. Example of an individual crossing arms, holding a cup and carrying a bag

## 5. RESULTS

In this study the performance of the supervised algorithms was carried out using a 10 fold cross-validation method. In order to provide quantified results, different metrics such as accuracy, precision, recall, and metrics were employed as follows in Equations (1) – (4):

$$Accuracy = \frac{TP+TN}{TP + FP + TN + FN} \tag{1}$$

$$Precision = \frac{TP}{TP + FP} \tag{2}$$

$$Recall = \frac{TP}{TP + FN} \tag{3}$$

$$F - Score = 2 * \frac{precision * recall}{precision + recall} \tag{4}$$

In which *TP* is the number of true positives, *TN* is the number of true negatives, *FP* is the number of false positives, and *FN* is the number of false negatives extracted from the confusion matrix.

We have obtained around 1000 frames per subject, and calculated 13 features each frame and indexed every subject. In order to avoid passing information from consecutive frames and have same number of frames per subject, we randomly pick 400 frames from our data set.

The results for different performance metrics are provided in Table 3. As can be seen, identification of individuals is achieved with relatively high accuracy across all the methods and the Random Forests outperforms other classifiers. In addition, to evaluate the contribution of different classes of features, we have separated them into five groups: all (Static, profile, angles and position), static, profile, static and position, static and angles and static and profile.

Table 3: Accuracy of different classes of 50 people (400 frames)

| *Classifier* | *All* | *Static* | *Profile* | *Static + Position* | *Static + Angles* | *Static + Profile* |
|---|---|---|---|---|---|---|
| Random Forests | 96.13 | 90.00 | 25.66 | 92.47 | 94.54 | 93.09 |
| Multi-Layer Perceptron | 93.48 | 87.25 | 15.91 | 89.51 | 26.87 | 89.81 |
| K-Nearest Neighbor | 93.98 | 86.99 | 24.42 | 89.37 | 93.11 | 90.64 |
| Naïve Bayes | 86.43 | 79.08 | 5.73 | 79.87 | 85.96 | 79.94 |

Using static biometric features only, high accuracy is obtained. On the other hand, using only profile features, the accuracy is relatively lower compared to static features. Through combining both feature types (i.e., combined), the identification performance could be further improved. As the results shown in the second column, an improvement is achieved by the combination of both feature types. Accordingly, in our study, we used all features for identification of individuals.

Table 4 shows 10-fold cross-validation validation using 400 frames. Random Forest result is statistically significant compared to K-Nearest Neighbors, Naïve Bayes, and Multi-Layer Perceptron.

Table 4. Performance metrics of different classifiers for 50 individuals.

| *Classifier* | *Accuracy (%)* | *Precision (%)* | *Recall (%)* | *F-Score (%)* |
|---|---|---|---|---|
| Random Forests | 96.13 | 96.14 | 96.13 | 96.13 |
| Multi-Layer Perceptron | 93.48 | 93.49 | 93.48 | 93.48 |
| K-Nearest Neighbor | 93.98 | 94.06 | 93.98 | 94.02 |
| Naïve Bayes | 86.43 | 86.50 | 86.43 | 86.46 |

In Table 5a, the confusion sub-matrix using Random Forest for 50 subjects. We can observe that the system could distinguish subjects (2, 11, 18, 37, and 48) that have similar height and weight (see Figure 4). Another features or combination of them helped the system to classify people with similar physical attributes. However, the worst case that the system misclassified 2 people is shown in Table 5b.

Table 5. Normalized Confusion Sub-Matrix showing similar people well classified and misclassified.

a)

| | | Actual Class | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|
| | | **4** | **6** | **11** | **18** | **28** | **37** | **39** | **46** | **47** | **48** |
| **Predicted Class** | **4** | 381 | 0 | 2 | 0 | 0 | 0 | 0 | 4 | 0 | 12 |
| | **6** | 0 | 365 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 |
| | **11** | 1 | 0 | 376 | 2 | 2 | 1 | 0 | 9 | 0 | 2 |
| | **18** | 0 | 0 | 1 | 371 | 0 | 0 | 0 | 2 | 0 | 3 |
| | **28** | 0 | 0 | 0 | 0 | 387 | 0 | 0 | 0 | 1 | 0 |
| | **37** | 0 | 0 | 1 | 1 | 0 | 368 | 4 | 6 | 0 | 2 |
| | **39** | 0 | 0 | 0 | 1 | 0 | 1 | 390 | 3 | 0 | 0 |
| | **46** | 7 | 0 | 8 | 1 | 0 | 4 | 1 | 368 | 0 | 8 |
| | **47** | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 398 | 0 |
| | **48** | 25 | 0 | 0 | 1 | 0 | 0 | 0 | 10 | 0 | 355 |

b)

| | | Actual Class | |
|---|---|---|---|
| | | **31** | **43** |
| **Predicted Class** | **31** | 92.4 | 7.6 |
| | **43** | 4.4 | 95.6 |

### 5.1. Occlusion

The dataset contains features of 27 individuals with partial data information. In Table 6, the results of occlusion are presented across 10-fold cross-validation. Random Forests algorithm presents better performance than the rest of classifiers. As the table below shows, the approach is capable of identifying different individuals with high accuracy even with the partial data information.

Table 6. Performance metrics (%) for static, profile and all features using partial data information for 27 people.

| ***Classifier*** | ***Accuracy (%)*** | ***Precision (%)*** | ***Recall (%)*** | ***F-Score (%)*** |
|---|---|---|---|---|
| Random Forests | 97.49 | 97.51 | 97.49 | 97.50 |
| Multi-Layer Perceptron | 96.42 | 96.43 | 96.42 | 96.42 |
| K-Nearest Neighbor | 96.88 | 96.91 | 96.88 | 96.90 |
| Naïve Bayes | 92.46 | 92.57 | 92.46 | 92.51 |

### 5.2. Posture

An individual crossing arms, holding a cup or carrying a bag could affect the identification since the arms are in different position and could read false measurements of the torso or arms. The silhouette and the profile also get disturbed by this posture due to the absence of arms contour or addition of the bag. However, the system identifies different individuals with promising results.

The first posture dataset contains normal and arms crossed information for 42 individuals around 690 frames each. The second and third posture dataset (holding a cup and carrying a bag) has 28 individuals with 800 frames per individual, each dataset with their respective activity and normal posture. The labeling of each individual in the datasets remained the same for both activities. In Table 7, the results of different postures are presented using different algorithms. We have evaluated this impact in a subset of our data. Random Forests has the greatest accuracy distinguishing individuals performing the same posture.

Table 7. Performance metrics (%) for different postures and classifiers

| *Posture* | *Classifier* | *Accuracy (%)* | *Precision (%)* | *Recall (%)* | *F-Score (%)* |
|---|---|---|---|---|---|
| *Crossing Arms* | RF | 96.29 | 96.31 | 96.29 | 96.30 |
| | MLP | 90.03 | 90.15 | 90.03 | 90.09 |
| | KNN | 93.57 | 93.62 | 93.57 | 93.59 |
| | NB | 64.74 | 64.72 | 64.74 | 64.73 |
| *Holding Cup* | RF | 96.13 | 96.14 | 96.13 | 96.14 |
| | MLP | 93.11 | 93.10 | 93.11 | 93.11 |
| | KNN | 94.51 | 94.54 | 94.51 | 94.53 |
| | NB | 87.35 | 87.60 | 87.35 | 87.47 |
| *Carrying Bag* | RF | 93.09 | 93.13 | 93.09 | 93.11 |
| | MLP | 89.81 | 89.88 | 89.81 | 89.84 |
| | KNN | 90.64 | 90.74 | 90.64 | 90.69 |
| | NB | 79.94 | 79.84 | 79.94 | 79.89 |

In addition, we also created a dataset passing information for 50 individuals walking normally and 28 of those walking carrying a bag. We wanted to observe the performance of the system distinguishing individuals performing an action while other people are walking normally. The results are promising for all classifiers, however, Random forest obtained higher accuracy. In Table 8, the performance metrics of different classifiers for the former dataset.

Table 8. Performance metrics (%) of 50 people walking normally and 28 people carrying a bag and normal posture.

| ***Classifier*** | ***Accuracy (%)*** | ***Precision (%)*** | ***Recall (%)*** | ***F-Score (%)*** |
|---|---|---|---|---|
| Random Forests | 96.33 | 96.27 | 95.95 | 96.11 |
| Multi-Layer Perceptron | 92.24 | 92.25 | 91.95 | 92.10 |
| K-Nearest Neighbor | 94.74 | 94.55 | 94.50 | 94.53 |
| Naïve Bayes | 81.70 | 81.80 | 81.32 | 81.56 |

### 5.3. Combination of Human Posture

It is clear that there is not a unique human posture when performing different activities; in this case walking into an indoor environment. Moreover, individuals could operate multiple postures inside of an environment. We have evaluated the mix of different postures performed in the room as subjects walk into a room. For example, an individual carrying a bag enters to a furnished office, drops the bag, picks up the cup of coffee, and leaves the office.

The dataset contains data information of 25 individuals carrying a bag, holding a cup, normal posture in and occluded environment. The labeling of each individual in the dataset remained the same for all activities for the system to classify individuals executing diverse postures. The total of frames per activity was around 11000 frames. The features extracted and classifiers are those used in previous analysis respectively. The length of legs for partial data information was deleted as well as, the partial volume of the profile was taken into account instead of the whole profile information. In Table 9, the performance metrics of different human postures inside of the room.

Table 9: Classification 10-fold using Weka of 25 people walking performing diverse human postures inside of the environment.

| *Classifier* | *Accuracy (%)* | *Precision (%)* | *Recall (%)* | *F-Score (%)* |
|---|---|---|---|---|
| Random Forests | 97.95 | 97.95 | 97.95 | 97.95 |
| Multi-Layer Perceptron | 92.29 | 92.28 | 92.30 | 92.29 |
| K-Nearest Neighbor | 95.66 | 95.76 | 95.67 | 95.71 |
| Naïve Bayes | 83.36 | 83.88 | 83.36 | 83.62 |

### 5.4. Gender Classification

The previous results used all the data set using information from all participants and then, infer the subject's identity based on features extracted. Alternately, we can take this task into a binary classification problem; gender. Whole-body attributes could be used to determine the gender of participants. The system could infer the gender of the participants based on static and profile biometric features. Men and women could be distinguished using depth image since their body measurements are different.

In order to do this, we have created a dataset where the individuals were labelled as man or woman. We have 19 females and 31 males in our dataset. This dataset contained 400 frames per person. In Table 10, we present the normalized confusion matrix for 10-fold cross validation gender classification of 50 people.

Table 10. Normalized confusion matrix for 10-fold cross validation gender classification of 50 individuals

| | | **Actual Class** | |
|---|---|---|---|
| | | **Man** | **Woman** |
| **Predicted Class** | **Man** | 98.9 | 1.3 |
| | **Woman** | 1.1 | 98.7 |

### 5.5. Number of frames

The number of frames were evaluated for all classifiers as shown in Figure 8 for 50 individuals. Each subject contained the same number of frames. In order to achieve the identification of individual real time, we are aiming to obtain the less amount of valid frames. The number of frames plays an essential role in the classification of individuals in the training set as we have few seconds to capture features of the individual walking between adjacent rooms.

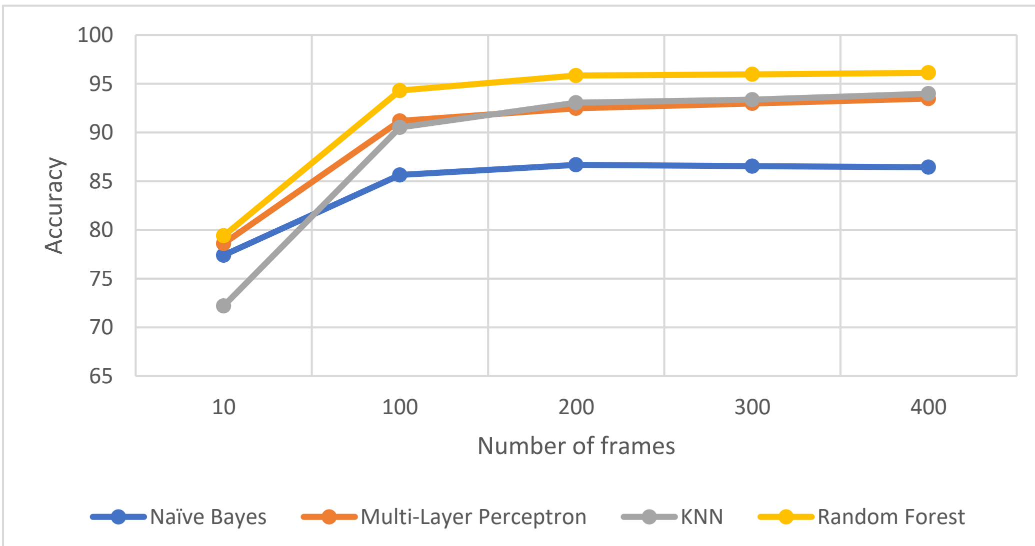



Figure 8. Accuracy of each classifier using different number of frames

The number of frames affects the accuracy of the classifier differently as they increase. In Figure 8, we can observe that around 400 frames we achieve 93% accuracy using Multilayer Perceptron, Random Forests, and K-Nearest Neighbor algorithms.

### 5.6. Relevant Features

The dataset contained a total of 13 features in this experiment. Using Correlation-based Feature Selection (CFS) (Hall 2000), we could understand the impact of each feature in our data. In Table 11a, the CFS results suggest the following ranked features for individual's identification, while Table 11b presents the results for gender's identification:

Table 11. a) Ranked Features using CFS for individual's classification and b) gender classification

| # | Feature | # | Feature |
|---|---|---|---|
| 1 | Waist | 8 | Left Leg |
| 2 | Shoulder Width | 9 | Angle right |
| 3 | Right Arm | 10 | Angle Left |
| 4 | Left Arm | 11 | Volume |
| 5 | Height | 12 | Area of Silhouette |
| 6 | Torso | 13 | Position |
| 7 | Right Leg | | |

| # | Feature | # | Feature |
|---|---|---|---|
| 1 | Shoulder Width | 8 | Right Arm |
| 2 | Waist | 9 | Volume |
| 3 | Height | 10 | Angle Left |
| 4 | Torso | 11 | Angle Right |
| 5 | Right Leg | 12 | Area of Silhouette |
| 6 | Left Leg | 13 | Position |
| 7 | Left Arm | | |

We can also observe that the majority of the top highest ranked features are static and the rest are related to the profile. As we had observed before, the profile features have a low accuracy identifying individuals by itself, however the combination of all features increased the accuracy overall. It is also valid to mention the importance of some features for subject and gender classification which could be used when only partial data information is available; Shoulder Width, Waist and Torso.

## 6. LIMITATIONS AND CHALLENGES

There are a number of limitations that should be taken into account in real-world implementation of depth-sensor-based ILS. According to the Kinect depth-sensing technology, the maximum depth range is about 8 meters, which will be limiting for larger areas. However, spaces with dimensions larger than 8 meters could be equipped with multiple sensors, which brings about further challenges in coordination between the sensors. Moreover, although individuals entering a room facing forward, others might pass the sensor sideways. Therefore, placement of the sensor is an important feature to ensure user identification. The assumptions made in this study were the incoming individual would cross the door facing forward and depth sensors would be placed facing the doors to proceed with the desirable biometric extraction.

## 7. CONCLUSIONS

In this paper, we have focused on the identification and localization of individuals using a depth-sensor-based ILS in indoor space. This system extracts multiple biometric features to identify individuals from human posture. We recruited individuals with different characteristics and gender to evaluate the system performance recognizing potential challenges in real-world implementations. The extracted features were static and profile. The following attributes were calculated for each individual: height; the length of torso, arms, and legs; the width of shoulders and waist; the body surface area, silhouette area and volume of the profile and the position from the sensor. The feasibility of the approach was evaluated in an experiment with 50 subjects walking in front of the depth sensor. We have included a more realistic scenario considering the partial data information due to tables and chairs in an indoor setting. To simulate this occluded scenario, the upper body information of 27 individuals was considered to infer their identification.

Several classification algorithms were used to evaluate the performance of the proposed system. Each algorithm demonstrated acceptable results, while all of them showed an accuracy above 86 %. The results show that static features are suitable for identification while adding profile biometric features could further increase the accuracy of the system.

In terms of different classifier performance, Random Forests outperformed other classifiers (96 % accuracy with both static and profile biometric features). The findings show that the proposed model can detect, localize, and also identify individuals with high accuracy in real-time from extracting biometric features attributed to human posture.

In addition, we investigated the innovative identification techniques in case of different posture in the indoor environment. The system obtained a high score classifying individuals performing the same posture (arms crossed, holding a cup, and a bag) entering an indoor environment with high accuracy (97%).

We decided to include a practical scenario the mix of different postures performed in the room as subjects walk into a room. For example, an individual carrying a bag enters to a furnished office, drops the bag, picks up the cup of coffee, and leaves the office. Additionally, the binary classification was performed to determine the gender of the participants. The system could infer with 98% accuracy the gender of the participants based on static and profile biometric features.

In the future studies, we plan to evaluate the performance on subjects walking between adjacent rooms. The experiments would allow us to provide an in-depth assessment of the profile feature contribution. Also, to investigate the identification techniques in case of different poses using the depth image as scene change effect will also be among the future directions in this study.

**STUDY 2: Anonymous Indoor Positioning System using Depth Sensors for Realistic Dynamics in Human-Building Interaction**

Authors:

Sergio Ballivian[1]

Farrokh Jazizadeh[2]

[1] Via Department of Civil and Environmental Engineering, Virginia Tech, 200 Patton Hall, 750 Drillfield Dr., Blacksburg, VA 24061. E-mail: sergiob5@vt.edu

[2] Via Department of Civil and Environmental Engineering, Virginia Tech, 200 Patton Hall, 750 Drillfield Dr., Blacksburg, VA 24061. E-mail: jazizadeh@vt.edu

# 1. INTRODUCTION

Context-aware systems were created to improve the interaction between human and buildings. Human-building interaction (HBI) has attracted significant research interest in the last two decades. HBI facilitates applications indoors such as activity monitoring, energy management, information delivery and personalized temperature conditions. Occupant presence modifies the temperature of indoor environments due to its external (seeking comfort) and internal (emission of heat) interaction with the building. All these behaviors affect the energy consumption in office buildings. Location and identification of occupants in contextual conditions are critical information to enable a naturalistic HBI.

Most of the recent indoor positioning technologies have focused on becoming less intrusive to increase the HBI. Indoor Positioning Systems (IPS) are methods to determine the location of a person or an object. Passive IPS does not require a receiver to locate individuals and leverages device-free interaction for data collection. Computer vision-based technologies use RGB image sensors and infrared cameras to detect subjects in the scene, however, these methods arise some issues for their invasion of privacy. In order to avoid privacy issues and boost HBI applications, a passive depth sensor based IPS could be capable of localization of subjects in an indoor environment. Depth sensors present the capabilities of the vision-based solutions in being passive while cost-effective without violating users' privacy by capturing RGB images. The lack of distance information between objects and sensor limits the vision-based methods since the depth information is an essential attribute to determine the location of subjects. Therefore, by knowing the location of the sensor not only the occupied room is identified, but also the absolute location of the occupants could be inferred. Tracking demands for identification of each individual moving across adjacent indoor spaces. Anonymous identification refers to detecting individuals and identifying them across sensors without the need for knowing real-world individuals' identity. In other words, an anonymous identity will be assigned to an individual after observing them for the first time and other sensors will seek to infer that identity. An example application of such a system is to capture an individuals' preferences in an environment and deliver services accordingly while they move in the office building.

In human-building interaction, understanding the dynamics of interactions is a critical step towards context-aware service delivery. The realistic dynamics in HBI are the natural interaction of occupants and environment that could affect the cross-space anonymous identification process. We have proposed a case study using the IPS system that simulates realistic scenarios in a complex environment. Anonymous identification brings some challenges that could affect the performance of the system. Change of environments and position of the sensor could affect the performance of the system. The schematic envisioned IPS diagram for human-building interaction applications is illustrated in Figure 5.

The rest of this chapter is structured as follows. Section 2 presents an overview of the methodology, experimental scenarios, and sensor parameters. In Section 3 the experimental design is presented. Then, Section 4 covers the data analysis and results of the experimental study. Finally, we presented the conclusions, limitations and future directions in Section 5.

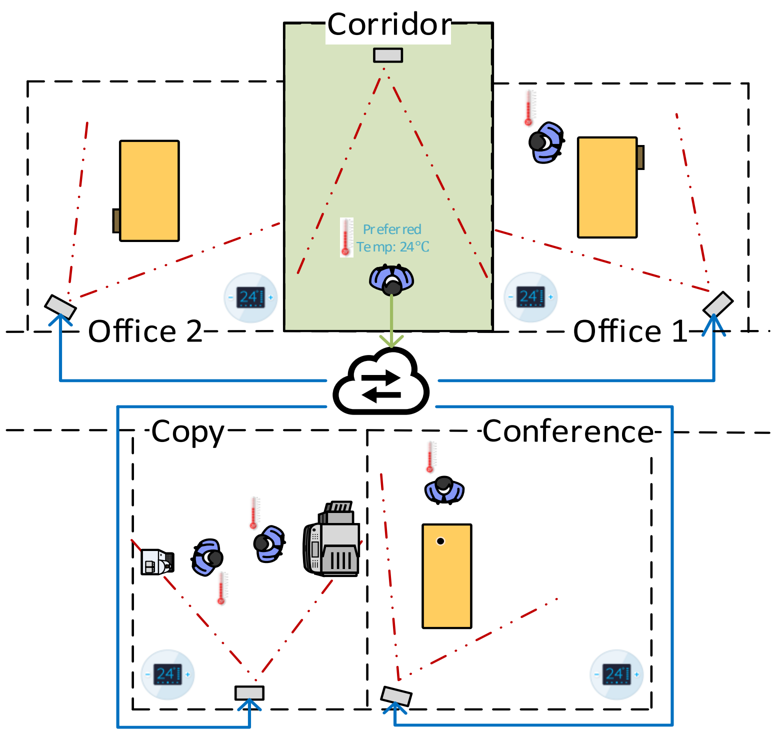



Figure 5. The schematic envisioned IPS diagram for Human-Building Interaction applications

## 2. OBJECTIVES

In this study, we have investigated the performance of the proposed a passive indoor positioning system, by accounting for realistic dynamics of human-building interaction. A case-study has been introduced to explore the feasibility of the depth-sensor based IPS and challenges of realistic dynamics in human-building interaction. The indoor positioning system leverages depth sensors to infer the location of occupants with high accuracy while seeking to address privacy concerns and training requirements enabling HBI.

Naturalistic interactions between occupants and buildings could result in noise and artifacts in data acquisition through depth sensors. Therefore, as occupants interact across different spaces, the changes in the environmental context and the angle between users and sensors could result in challenges for the inference algorithm. Therefore, this case study focuses on investigating the impact of naturalistic human-building interactions on cross-space and anonymous human subject identification by assessing sensor position, realistic scenarios, and heuristic analysis.

## 3. METHODOLOGY

### 3.1. Indoor Positioning System Framework

In this study, we are evaluating the performance of a developed passive indoor positioning system taking into account the contextual dynamics of human-building interaction. Several experiments were conducted on human subjects as they walked and completed general tasks in a testbed. The experiments were divided into different experimental scenarios to assess the feasibility of IPS. The inference framework for the identification and tracking process is depicted in Figure 6.

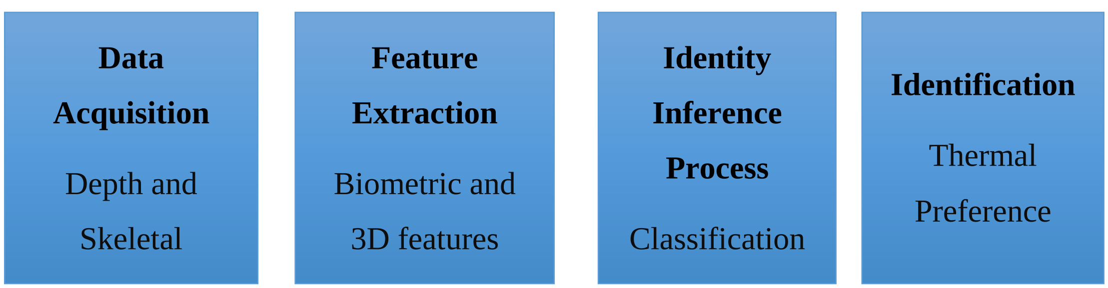



Figure 6. Inference framework for identification and tracking process

### 3.2. Identification Scenarios

The experimental scenarios for the identification process have been performed in a section on the second floor in Patton Hall. The experiments were divided into two main scenarios.

The first scenario consisted on determine the feasibility of the algorithm to identify individuals performing all and independent actions that account for realistic dynamics in HBI. The identity inference technique will be evaluated using Cross-Validation. Also, this technique validates the anonymous depth-sensor-based IPS considering the dynamics of the environment.

The sensor will capture the entire skeleton overall, extract the joint coordinates and depth data, and measure biometric attributes from the human body. The spatial coordinates were extracted as each individual walk at their usual pace. In order to simulate a realistic scenario, the human subjects will be asked to walk naturally in the room. The experiment consists of extracting biometrics from different individuals by walking while facing the depth sensor for few seconds. The individuals walking perpendicular and facing the sensor in a room simulated the location of the depth sensor placed facing the entrance door and captured the frontal view of the individual. Some limitations assuming individuals enter the room facing will be also evaluated in this experiment. Different postures (e.g., holding a cup, arms crossed) and poses were analyzed using supervised machine algorithms to quantify the effect of these activities.

The second scenario analyses the gender classification and then proceeds tracking the subjects by detecting, classifying, and finally inferring their identity. “Train and Test” technique will be used in this scenario to infer the identity of the individuals. We have focused on analyzing the cross-space anonymous identification of human subjects. The identification evaluation was performed as human subjects move across adjacent rooms.

Whole-body attributes could be used to determine the gender of participants. The system could infer the gender of the participants based on static and profile biometric features. Men and women could be distinguished using depth image since their body characteristics are different. We have created a dataset where the 8 individuals were labelled as man or woman. To proceed with the anonymous identification, tracking was based on detecting individuals as they walked into some rooms to perform the action (training set), binary classification to determine how often the system recognize the subject and finally infer the identity of the individual in different rooms are part of the testing set. In Figure 3, the diagram of the identification scenarios and respective Identity Inference Process techniques.

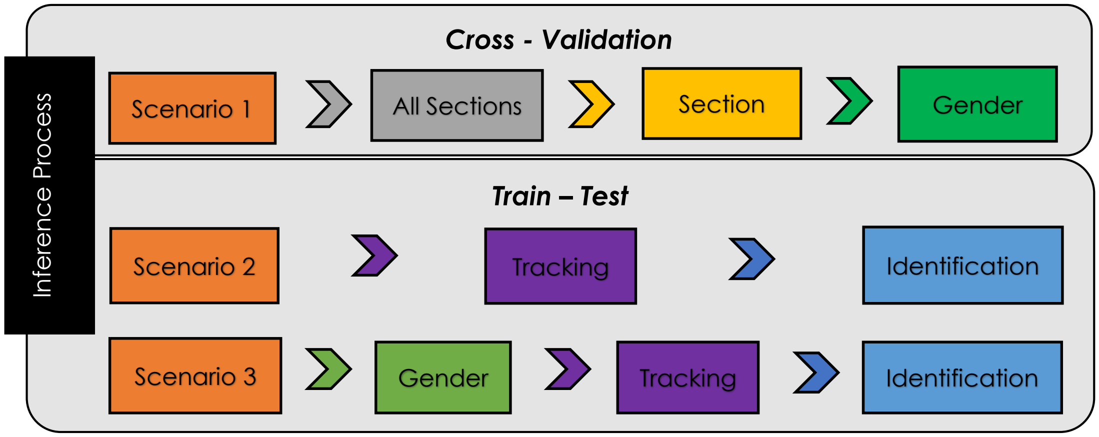



Figure 7. Diagram of experimental Identity Inference Process scenarios in this study

### 3.3. Sensor Position Parameter

The position of the sensor, placed in the room, is an important factor for the data acquisition. To avoid the challenges of occlusion, most studies install the sensors on ceilings for identification or detection of individuals since full skeleton or depth data information is not available all the time. In addition, the accuracy of people identification might get affected depending on the position of the sensor.

We decided to study the impact of sensor placement at different locations in the testbed. Experiments have been conducted by using 3 different positions. In Figure 8, the respective locations of the sensor in all the spaces have been shown.

| Level | Height (cm) |
|---|---|
| Ceiling | ~220 |
| Head | ~188 |
| Waist | ~109 |

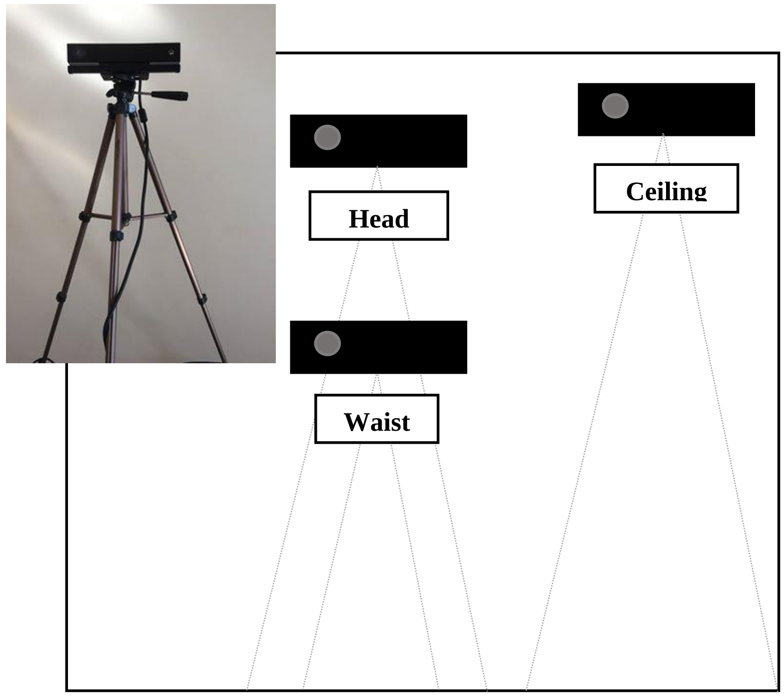



Figure 8. Ceiling, Head and Waist Level of the sensor located in all room-environments

## 4. Data Acquisition

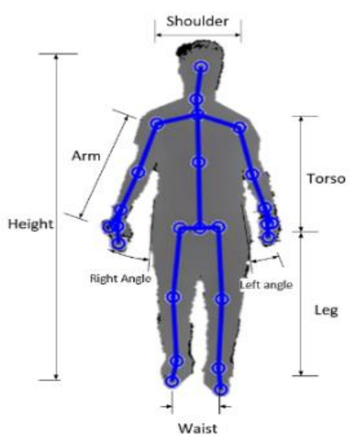
### 4.1. Skeletal and Depth Information

The Kinect Sensor V2 from Microsoft was used as the main depth sensing technology in this case study. This sensor consists of a built-in RGB camera, a system of 3D depth sensors, and microphones. The 3D depth sensor system includes a set of infrared (IR) projector and camera. The sensor collects the depth frames information using the IR camera to preserve user's privacy. The depth images, unlike the RGB video images, collected depth information of the whole environment. The depth sensor captures depth frames from objects and skeletal data from individuals in the scene. The skeletal data refers to the spatial coordinates of 25 joints of the human body. In this work, we have used the Kinect Software Development Kit (SDK) 2.0 for the data acquisition. The Kinect SDK is a component of software that processes the image data to extract further information including human detection, skeletal identification and measurements. The Kinect depth sensor could provide depth images up to 8 meters from the sensor.

### 4.2. Feature Extraction

As the previous study mentioned, the biometric attributes (i.e., features) were captured at each frame from individual’s skeleton and depth information. We have extracted 13 biometric attributes from the human body. Static features are based on the distance from the joints or length of bones such as height, length of torso, arms and legs; the width of waist and shoulders, and angles between arms and torso. To obtain biometric static features, the measurements were calculated for each individual using the Euclidean distance between the joints at each frame. In addition, the background removal of subjects could be observed in silhouette forms. The area of the silhouette and volume of the individuals’ profile were calculated by depth data information using MATLAB software. The system stores biometric features in the database and matches individuals in the future based on these characteristics. Hereafter, we will call these features static (representing skeletal information) and profile features (representing the body surface). Table 12 illustrates the biometric features collected in this study.

Table 12. Biometric Features and background subtraction collected from each individual




| ***Biometric Features*** | | |
|---|---|---|
| *Static* | Length | Height, Arms, Torso, Legs |
| | Width | Waist, Shoulder to shoulder |
| *Profile* | Area | Silhouette |
| | Volume | Surface |
| *Angles* | Angle between arms and torso | |
| *Position* | Position of the individual from the sensor | |

Background subtraction allowed us to identify the shape characteristics of the human body in the environment. The system could detect individuals from 0.8 to 3.5 meters ideally. The pixels on each depth image represents the distance from the sensor to the object. Pixels that are out of the sensor’s range were assigned zero values. The contour of the body is called human silhouette. In some cases, the system will detect region where the pixels do not belong to the human shape (outliers) but objects around the individuals (i.e., opening the door, grabbing a chair), specifically when occupants have contact with items as they walk in the room.

The human silhouette (depth image) was extracted as the individuals walk into the different spaces to perform different actions. The depth images were captured as the human subject entered the room. We were interested on identifying people when full data information is available. The depth sensors installed on different rooms inside of the building to capture skeletal, profile and position information from the individuals. given the identity of the individual is known, we can proceed with the localization by tracking the person based on the distance from the sensor inside of the room.

### 4.3. Identity Inference Process

The identity inference process consists of extracting biometrics from individuals by the system and evaluate them using a machine learning algorithm. Classification algorithms are supervised learning algorithms that make predictions based on labeled data. In the previous study, we have evaluated the system using various supervised machine learning algorithms (K-Nearest Neighbor (KNN), Multi-Layer Perceptron (MLP), Random Forests (RF), and Naïve Bayes (NB)). In this case study, we decided to pick the Random Forest algorithm based on its better performance in previous evaluation.

“Cross-validation” and “Train and Test” are two main techniques to evaluate the performance of the classification algorithms. Cross validation splits up the same dataset into training set and the rest is used to forecast prediction metrics.

The second technique uses two different datasets. “Training dataset” is used to create a model based on the data and “testing dataset” is used to predict incoming new data based on the training model.

### 4.4. Normal Data and Heuristic

After detecting subjects on the first few seconds, the system has selected reliable frames of human silhouettes and skeletal data. As previously mentioned, feature values were calculated using Euclidean distance algorithm between the physical joints at each frame. Method A consisted on the data collection coming from the field (normal dataset “A”) and apply the classification algorithm through Train and Test technique.

The realistic interactions between subjects and the environment exists noise and artifacts in the data. For example, some frames might show noisy frames. In order to address that problem, we have proposed a heuristic. In this heuristic case, we only considered frames that satisfy a predefined quantity threshold. We have selected the subjects' height static feature to leverage full body data acquisition as individuals walk into a room environment. Method B represents the analysis using the average height from all the subjects' data collected. As static features should not vary with time, we have also considered the fact that they will not vary as subjects move from one environment to another. Method B considers the dependency between the calculated average height from the training and testing set. The system calculated the average height per subject, and implemented the new subjects' height for future recognition.

$$h_c = \frac{\sum_0^n H_i}{n_i} = \vec{H}_i \,; Hj > \vec{H}_i \quad for\ all\ individuals \qquad (1)$$

The average height ($\vec{H}_i$) of all data points ($n$) from each subject ($i$) in the training stage and evaluated new incoming subjects frames from other rooms $\left(H_j\right)$ as part of the testing stage.

## 5. EXPERIMENTAL DESIGN

The research experiments were carried out in a section of the second floor in Patton Hall at Virginia Tech (Blacksburg, VA). The total area of the experimental testbed was approximately 60 meters squared. The section covers 4 rooms and a central hallway. The common entrance which divides both of the offices (211A & 211E) is a small corridor (211). Each sensing node was equipped with an off-the-shelf depth sensor (i.e., Kinect for Windows V2 sensor) located at the three different positions inside the rooms. Figure 9 shows the schematic design with multiple occupants of the section on the second floor.

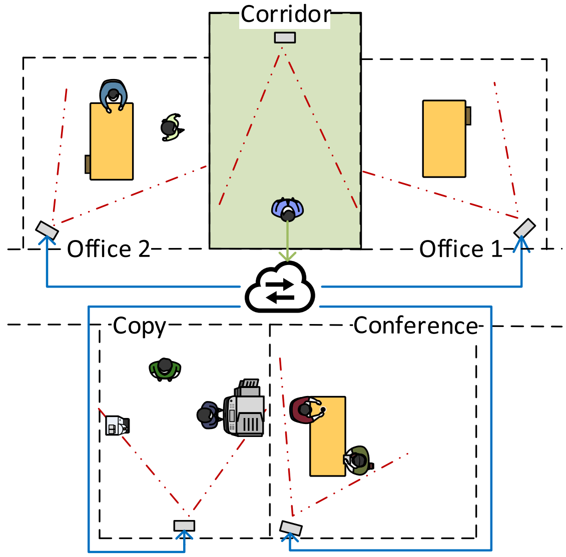



Figure 9. The schematic design with multiple occupants of the section on the second floor

The actions performed by each individual were designed to generate realistic scenarios in the office setting. The participants were not instructed to do any specific tasks besides the action found in Table 13. The purpose of these actions is to recreate actions naturally performed by individuals in a real world environment.

Table 13. Description of the actions performed on respective scenarios

| Action | Scenario | Description | Room |
|---|---|---|---|
| 1 | Entrance | Walk-in to the corridor | Corridor |
| | Office | Sit down in the office | Office 1 & 2 |
| 2 | Break time | Take a break using your phone | Corridor |
| 3 | Conference | Enter and sit down in the Conference room | Conference room |
| 4 | Photocopy | Get a photocopy | Copy room |
| 5 | Coffee | Pick up a cup of coffee | Copy room |

Human subjects were asked to simulate the entrance to their office (Action 1) and conference (Action 3) rooms. They opened the corridor door and walked in to any office or meeting room. They took a break using their phone in the corridor (Action 2). In the copy room, subjects were instructed to pick up a photocopy (Action 4) and a cup of coffee (Action 5).

### 5.1. Participants

In these experiments, we have selected 8 participants ranging from 20 to 40 years old. Each participant provided their information (height and weight) shown in Figure 10. In order to ensure the performance of the system is evaluated in challenging cases, not only distinct individuals will be part of the selection, but we also considered individuals with similar physical features that result in analogous attributes as well.

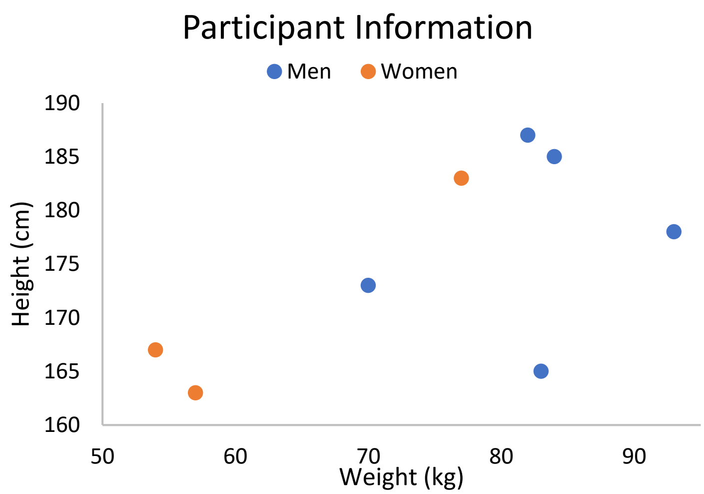



Figure 10. Height and weight information from each participant

### 5.2. Room Environments

Room environments are generally found with furniture such as cabinets, tables, chairs. Therefore, these objects could occlude the sensor’s field of view leading to partial data collection and wrong measurements. Several human attributes might be obstructed in the data acquisition which alters the system’s performance. Nevertheless, the association of all attributes could improve the subject’s identification. Figure 11 displays the pictures from the rooms on the second floor in Patton Hall.

Figure 11. Rooms (Corridor, Office 1, Copy room, Conference room) on the second floor in Patton Hall

## 6. DATA ANALYSIS AND RESULTS

In this study the performance of the supervised algorithms was carried out using methods described in Section 3. In order to provide quantified results, different metrics such as accuracy, precision, recall, and metrics were employed as follows in Equations (2) – (5):

$$Accuracy = \frac{TP+TN}{TP + FP + TN + FN} \tag{2}$$

$$Precision = \frac{TP}{TP + FP} \tag{3}$$

$$Recall = \frac{TP}{TP + FN} \tag{4}$$

$$F - Score = 2 * \frac{precision * recall}{precision + recall} \tag{5}$$

In which *TP* is the number of true positives, *TN* is the number of true negatives, *FP* is the number of false positives, and *FN* is the number of false negatives extracted from the confusion matrix.

### 6.1. Cross-validation Evaluations

Cross validation technique was used for identification of individuals performing all actions. The recognition of individuals performing specific action was evaluated using Cross-validation. This procedure will help us understand if it is feasible to distinguish individuals who were doing the same action. After the IPS distinguishes subjects, the system could adjust room temperature to deliver individual's thermal preference.

#### 6.1.1. All Actions and Subjects

We have performed the analysis of people identification performing all and independent actions. As noted, the data was collected from different levels (ceiling, head, waist). The purpose of this experiment is to evaluate the system's capabilities to recognize or distinguish different individuals. Table 14 shows the performance metrics for identification of 8 individuals using normal data (A) and height heuristic (B).

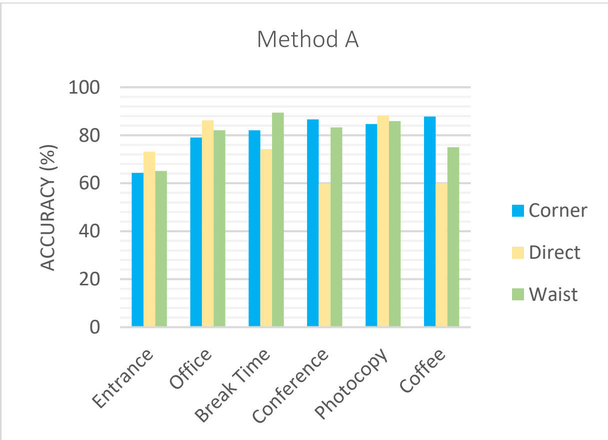
Table 14. Performance Metrics of cross-validating normal data (A) and height heuristic (B) from the identification of 8 subjects performing all actions

| | Ceiling | | Head | | Waist | |
|---|---|---|---|---|---|---|
| **Performance Metrics** | A | B | A | B | A | B |
| **Accuracy (%)** | 73.2 | 86.7 | 71.2 | 88.5 | 78.4 | 88.0 |
| **Precision (%)** | 74.3 | 87.3 | 72.5 | 89.1 | 78.1 | 88.1 |
| **Recall (%)** | 71.1 | 85.5 | 70.5 | 87.6 | 75.9 | 87.2 |
| **F-Score (%)** | 72.2 | 86.2 | 70.9 | 88.3 | 76.6 | 87.5 |

The application of heuristic improved the performance of the system in all cases since the average height from the training data points is accountable to the evaluation of the incoming testing data points. The minimum percentage change was 20% in the head level between normal data and heuristic. The normal classification F-score in the waist level case, has 77% performance which is better than ceiling and waist level. In addition, the average F-score heuristic achieves 88% using cross-validation technique for anonymous people identification across sub-spaces.

#### 6.1.2. Isolated Actions

The analysis of subjects performing the same action in the room evaluates the system's performance in recognition of individuals within the same environment. Each dataset has the information of different individuals collected from the same level performing the same action in the same environment. Figure 12 shows the accuracy for identification of 8 individuals using normal classification and heuristic between individual actions at each sensor location.



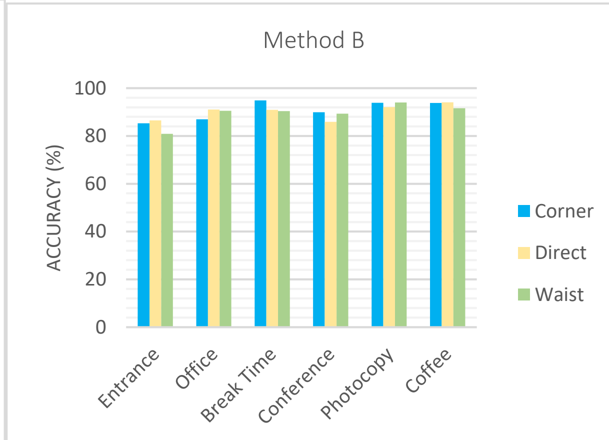




Figure 12. Accuracy of people identification using normal classification (A) and heuristic (B) for each action at each sensor location.

The results for each action have been visually identified the improvement using heuristic compared to normal classification. The analysis results in each action were expected to be high since individuals are on the same physical environment. Since indoor environment and sensor position is not a variable, the system focuses on the identification of the people performing the same action and performs superior than diverse sub-spaces. The average F-score analyzing with normal and heuristic classification in all sensor locations achieved 78% and 90% using cross-validation technique, respectively.

#### 6.1.3. Gender Classification

We have 3 females and 5 males occupants in our dataset. As discussed previously, the system was trained in labelled rooms and evaluated in others using Cross-Validation. In Table 15, we have presented the accuracy of training and testing gender classification of 8 people using normal classification (A) versus height heuristic (B).

Table 15. Performance metrics for cross-validation using normal classification (A) and height heuristic (B) for detecting the gender of 8 subjects

| | **Ceiling** | | **Head** | | **Waist** | |
|---|---|---|---|---|---|---|
| **Performance Metrics** | A | B | A | B | A | B |
| **Accuracy (%)** | 82.6 | 93.2 | 86.1 | 94.1 | 90.1 | 93.9 |
| **Precision (%)** | 83.0 | 93.6 | 85.8 | 94.3 | 90.4 | 94.0 |
| **Recall (%)** | 82.9 | 91.0 | 81.1 | 92.1 | 88.4 | 92.8 |
| **F-Score (%)** | 82.2 | 95.1 | 90.4 | 95.8 | 92.4 | 95.2 |

There is at least a 4% percentage improvement using heuristic data compared to normal data from all sensor positions. In the case of the ceiling level, the percentage change of 13% shows that there is a significant increment between the data collected and height heuristic. The F-score of normal classification has increased almost 9% from ceiling to head and waist level, while the average F-score of height heuristic is constant through the 3 level positions and around 95%. Head and waist level F-scores have shown promising results using Cross-Validation technique for gender identity of subjects tracked in other room environments.

### 6.2. Train and Test Evaluations

The depth sensors installed on different rooms collected information from the individuals once (minimal training) and evaluated in other room environments. This approach simulates the realistic scenarios of cross-space identification. We collected the subject's data on specific sub-spaces for the first time as part of the training set. The envisioned system of interconnected sensors in other sub-spaces will receive collected information and infer the identity of individuals which is part of the testing dataset. Training and testing datasets for the respective actions in the scenario shown in Table 16.

Table 16. Training and testing sets for respective actions

| | Actions | | | | | |
|---|---|---|---|---|---|---|
| **Run** | **Entrance** | **Office** | **Break time** | **Conference** | **Photocopy** | **Coffee** |
| **1** | Training | Training | Test | Test | Test | Test |
| **2** | Training | Training | Training | Test | Test | Test |
| **3** | Training | Training | Test | Training | Test | Test |
| **4** | Training | Training | Test | Test | Training | Test |
| **5** | Training | Training | Test | Test | Test | Training |

### 6.2.1. Gender Classification

The gender identity of all subjects is achieved with relatively high accuracy (81% F-score) across all actions performed in different rooms using the Train and Test inference process. for anonymous people identification in subsequent indoor environments. We have determined gender of occupants in the system, moreover, this information could be passed as another feature for the identity inference algorithm. Table 17 shows the accuracy gender classification of 8 people using normal classification (A) versus height heuristic (B) evaluating training and testing inference process when the sensor was placed in the head level.

Table 17. Average accuracy of gender identity of Train and Test inference process using normal classification (A) and heuristic (B) of 8 subjects

| | Actions | | | | | | Head | |
|---|---|---|---|---|---|---|---|---|
| **Run** | **Entrance** | **Office** | **Break time** | **Conference** | **Photocopy** | **Coffee** | **A** | **B** |
| **1** | Training | Training | Test | Test | Test | Test | 73.8 | 77.8 |
| **2** | Training | Training | Training | Test | Test | Test | 74.7 | 74.9 |
| **3** | Training | Training | Test | Training | Test | Test | 80.0 | 83.3 |
| **4** | Training | Training | Test | Test | Training | Test | 77.3 | 79.9 |
| **5** | Training | Training | Test | Test | Test | Training | 80.3 | 80.8 |
| **Average** | | | | | | | **78.2** | **81.4** |

### 6.2.2. Identification

Table 18 summarizes the F-score results for different Train and Test action runs for all level positions using normal data and heuristic. In addition, we have separated into sensor positions to evaluate the contribution of different types of data points. An overview of the results shows that using the dependent average height value between training and testing dataset (Method B) demonstrated a better performance compared to the normal approach (Method A).

Table 18. Average accuracy of people identification of Train and Test inference process using normal data and heuristic for ceiling, head and waist level

| | **Ceiling** | | **Head** | | **Waist** | |
|---|---|---|---|---|---|---|
| **Actions** | **A** | **B** | **A** | **B** | **A** | **B** |
| **1** | 43.5 | 45.1 | 35.2 | 75.6 | 53.7 | 57.5 |
| **2** | 30.6 | 33.8 | 51.1 | 68.4 | 44.0 | 47.0 |
| **3** | 50.8 | 50.9 | 38.9 | 67.4 | 41.2 | 57.7 |
| **4** | 49.1 | 54.9 | 60.4 | 67.4 | 50.1 | 55.0 |
| **5** | 49.9 | 57.4 | 44.6 | 73.2 | 51.5 | 62.9 |
| **Average** | **44.8** | **48.4** | **46.0** | **70.4** | **48.1** | **56.0** |

## 7. CONCLUSIONS

In this study, we have investigated a developed passive indoor positioning system which accounts for realistic dynamics of human-building interaction. The proposed approach relies on fixed depth sensors installed on multiple spaces inside a building to acquire information from subjects moving between different spaces. The identification of individuals plays an important role in localizing them moving between indoor spaces since interconnected devices have to recognize individuals previously seen in a passive manner. This study focuses on enabling depth sensors for cross-space and anonymous human subject identification. A depth-sensor-based IPS, passive and cost-effective solution, could be capable of localization of human subjects on indoor environments avoiding privacy concerns. We focused on the use of commercially available, off-the-shelf vision-based technologies, Microsoft Kinect V2.

We proposed a study of a real-world implementation in a hand-full of rooms in Patton Hall, Virginia Tech. It simulates 8 human subjects performing realistic actions in an office building setting that accounts for dynamics and event sequence path.

Given the challenges that occlusion and human data availability could impact the performance of the approach, we decided to study this phenomenon by placing the Kinect sensor in three different locations in the room to have a more realistic scenario. The proposed framework was evaluated for different locations: ceiling, head and waist level. Sensors were placed on all room spaces resulting on better performance in the head level position.

The passive IPS has been evaluated through experimental scenarios using “Cross-validation” and “Train-Test” techniques for anonymous occupant identification across sub-spaces. The first scenario consisted on determine the feasibility of the classification algorithm to identify individuals performing set of actions on the same and other room environments that account for realistic dynamics in HBI. The results of the heuristic for human subjects realizing individual actions in the same room achieved 90% (F-Score) while human subjects performed the actions in different room environments was 88% (F-Score). We could infer that the room environment context plays an important role in people identification.

The second scenario assesses realistic actions in an office building setting that accounts for dynamics and event sequence path. As human subjects walk between adjacent rooms, the system determines gender and then proceeds tracking the subjects by detecting, classifying, and finally inferring their identity. The interconnected sensors in sub-spaces receive collected information and infer the identity of individuals which is part of the testing dataset. The analysis of the results revealed that heuristic in the head level show a better performance (70% F-Score) for human identification across indoor spaces.

The system has inferred the gender of the participants based on static and profile biometric features as body attributes could be used to determine the gender of participants. The average of cross-space anonymous identification using height heuristic is around 95% (F-Score) and constant through all sensor positions.

There are a number of limitations that should be taken into account of depth-sensor-based IPS in a larger real-world implementation. We have achieved great performance assuming the room environment will have a unique door which is the entrance as same as the exit way and a sensor per space, some spaces could be designed an individual door for each flow-traffic and will be equipped with multiple sensors, which brings about further challenges in coordination between the sensors.

Another important limiting factor is the training data collection. Although the system captured a minimal amount of human data points entering a room for the training set, different posture and room environments affect the user identification even though they are performing the same action. Deep Learning could facilitate the feature extraction process from the users and environmental context.

The depth-sensing technology could enable the understanding where human subjects are located anytime and the activities performed between indoor spaces. In addition, the depth sensing ILS technology could assist not only monitoring personnel in a particular office but increase productivity in the company. The outcome of this study will enable the application of cost-effective depth sensor system for identification and tracking purposes in office setting, and responsive to individuals' comfort preferences as they move across adjacent indoor environments. The application of depth based positioning system could further be applied to the sustainability; detecting unoccupied rooms in buildings to save energy in the built environment, and reduce cost of heating, lighting or air conditioning equipment levering the Human-Building interaction.

## CONTRIBUTIONS TO THE BODY OF KNOWLEDGE

This thesis investigated the fundamental requirements for the depth-sensing-based occupant tracking for the interaction between humans and buildings. The first component of the research focused on investigating the feasibility of the passive and anonymous indoor occupant positioning systems using depth sensors. The depth sensors could help eliminate the need for carrying a receiver by users and could improve the accuracy of the location information. However, for tracking, they require mechanisms for detecting the identity of occupants as they move across different spaces. Therefore, in this thesis, we have proposed to extract multiple biometric features to identify individuals from human posture. To this end, this thesis has investigated the impact of new features and postures considering natural dynamics of occupant behavior. Different postures were evaluated to simulate natural occupant behavior in an office setting such as partial information, carrying a bag, holding a cup, and crossing arms. This study not only evaluated the binary and multi-class classification through a large scale experiment (50 individuals) but also investigated realistic scenarios such as occlusion and varied posture of occupants as they enter the room. The proposed system performed with a high accuracy in detecting, localizing, and also identifying individuals. The second component of research contributions focused on the study of tracking occupants in real-world scenarios in which occupants conduct their daily activities across different spaces. Tracking demands for identification of each individual moving across adjacent indoor spaces. Anonymous identification for IPS was evaluated under the constraints of realistic scenarios in an indoor space. To this end, the second component contributes by introducing a data analysis framework that seeks to tackle the challenges of real-world data artifacts.

## CONCLUSIONS

Studies on Indoor Localization Systems demonstrated the importance of positioning for the clients and organizations inside building spaces. Promising results in health care, construction industry, and many service enterprises have led ILS to grow remarkably in the near future. The depth-sensing technology could enable the inference of occupants’ locations inside building and the activities performed between indoor spaces. Such information could be leveraged in sustainable operation of buildings by detecting unoccupied rooms in buildings for conservation of energy and associated costs through reducing unnecessary use of heating, lighting and air conditioning equipment.

Our envisioned system is an ILS system based on the depth-sensing technology that could enable buildings to adapt and respond to occupants' dynamics. An example application of such a system is the provision of personalized thermal preference to specific occupants anonymously detected. The system based on depth sensors needs to be aware of the accurate location of the occupants to deliver such services. While localization refers to the one-time estimation of individual position in space, tracking is the trajectory (location path over a duration of time) of an identified individual. Occupants do not perform tasks statically all the time in an office setting, instead, they do have a dynamic behavior during the day. Individuals perform different actions and move between indoor spaces. Therefore, this thesis focused on the tracking of identified occupants using a depth-sensor-based system while they move across different spaces accounting for realistic dynamics of the indoor environment.

Based on observed limitations in the literature, the experimental design and scenarios in this thesis accounted for occlusions in the space, different postures of the users and different positions of the depth sensors. In addition, this thesis has evaluated the feasibility of employing interconnected sensing systems for cross-space occupant's identification by using minimal training in rooms, extracting biometric features and inferring their identity across spaces. Experimental research scenarios were evaluated to address the tracking challenges in real-world environments. In order to provide quantified results, different performance metrics were employed. The cross-space anonymous individuals and gender identification for the optimum sensor location and data analytics algorithm showed a performance of 70% and 95% F-Score, respectively.

The positioning and tracking of subjects provided additional insights, in particular, certain limitations and challenges were provided by the office space environment implementation. For example, Kinect depth-sensing technology has the maximum depth range of 8 meters, and larger spaces may require being equipped with multiple sensors, which brings about further challenges regarding the coordination between the sensors. Moreover, more advanced indoor tracking algorithms could be used to improve the performance of the algorithms for identification.

## REFERENCES


Albiol, A., Albiol, A., Oliver, J., and Mossi, J. M. (2012). "Who is who at different cameras: people re-identification using depth cameras." *IET Computer Vision*, 6(5), 378-387.

Araujo, R. M., Graña, G., and Andersson, V. "Towards skeleton biometric identification using the microsoft kinect sensor." *Proc., Proceedings of the 28th Annual ACM Symposium on Applied Computing*, ACM, 21-26.

Basiri, A., Peltola, P., Silva, P. F. e., Lohan, E. S., Moore, T., and Hill, C. "Indoor positioning technology assessment using analytic hierarchy process for pedestrian navigation services." *Proc., 2015 International Conference on Location and GNSS (ICL-GNSS)*, 1-6.

Bolic, M., Simplot-Ryl, D., and Stojmenovic, I. (2010). *RFID systems: research trends and challenges*, John Wiley & Sons.

Chakravarty, K., and Chattopadhyay, T. "Frontal-standing pose based person identification using kinect." *Proc., International Conference on Human-Computer Interaction*, Springer, 215-223.

Chen, H., Gallagher, A., and Girod, B. "Describing Clothing by Semantic Attributes." Springer Berlin Heidelberg, 609-623.

Dantcheva, A., Dugelay, J.-L., and Elia, P. "Person recognition using a bag of facial soft biometrics (BoFSB)." *Proc., Multimedia Signal Processing (MMSP), 2010 IEEE International Workshop on*, IEEE, 511-516.

del-Blanco, C., Mantecón, T., Camplani, M., Jaureguizar, F., Salgado, L., and García, N. (2014). "Foreground Segmentation in Depth Imagery Using Depth and Spatial Dynamic Models for Video Surveillance Applications." *Sensors*, 14(2), 1961.

Demirkus, M., Garg, K., and Guler, S. "Automated person categorization for video surveillance using soft biometrics." *Proc., SPIE Defense, Security, and Sensing*, SPIE, 12.

Diraco, G., Leone, A., and Siciliano, P. (2015). "People occupancy detection and profiling with 3D depth sensors for building energy management." *Energy and Buildings*, 92(Supplement C), 246-266.

El-laithy, R. A., Huang, J., and Yeh, M. "Study on the use of Microsoft Kinect for robotics applications." *Proc., Proceedings of the 2012 IEEE/ION Position, Location and Navigation Symposium*, 1280-1288.

Hall, M. A. (2000). "Correlation-based feature selection of discrete and numeric class machine learning."

Han, J., Shao, L., Xu, D., and Shotton, J. (2013). "Enhanced Computer Vision With Microsoft Kinect Sensor: A Review." *IEEE Transactions on Cybernetics*, 43(5), 1318-1334.

Hazas, M., and Ward, A. "A novel broadband ultrasonic location system." *Proc., International Conference on Ubiquitous Computing*, Springer, 264-280.

Hightower, J., and Borriello, G. (2001). "Location systems for ubiquitous computing." *Computer*, 34(8), 57-66.

Jaha, E. S., and Nixon, M. S. "Soft biometrics for subject identification using clothing attributes." *Proc., IEEE International Joint Conference on Biometrics*, 1-6.

Khoshelham, K., and Khoshelham, K. (2012). "Accuracy and Resolution of Kinect Depth Data for Indoor Mapping Applications." *Sensors (Basel, Switzerland)*, 12(12), 1437-1454.

Koyuncu, H., and Yang, S. H. (2010). "A survey of indoor positioning and object locating systems." *IJCSNS International Journal of Computer Science and Network Security*, 10(5), 121-128.

Mainetti, L., Patrono, L., and Sergi, I. "A survey on indoor positioning systems." *Proc., Software, Telecommunications and Computer Networks (SoftCOM), 2014 22nd International Conference on*, IEEE, 111-120.

Ni, L. M., Liu, Y., Lau, Y. C., and Patil, A. P. (2004). "LANDMARC: indoor location sensing using active RFID." *Wireless networks*, 10(6), 701-710.

Poston, J. D., Schloemann, J., Buehrer, R. M., Malladi, V. V. N. S., Woolard, A. G., and Tarazaga, P. A. "Towards indoor localization of pedestrians via smart building vibration sensing." *Proc., 2015 International Conference on Location and GNSS (ICL-GNSS)*, 1-6.

Saad, M. M., Bleakley, C. J., Ballal, T., and Dobson, S. (2012). "High-Accuracy Reference-Free Ultrasonic Location Estimation." *IEEE Transactions on Instrumentation and Measurement*, 61(6), 1561-1570.

Shih, H.-C. (2014). "A robust occupancy detection and tracking algorithm for the automatic monitoring and commissioning of a building." *Energy and Buildings*, 77(Supplement C), 270-280.

Shotton, J., Fitzgibbon, A., Cook, M., Sharp, T., Finocchio, M., Moore, R., Kipman, A., and Blake, A. "Real-time human pose recognition in parts from single depth images." *Proc., CVPR 2011*, 1297-1304.

Sinha, A., and Chakravarty, K. "Pose Based Person Identification Using Kinect." *Proc., 2013 IEEE International Conference on Systems, Man, and Cybernetics*, 497-503.

Sinha, A., Chakravarty, K., and Bhowmick, B. "Person identification using skeleton information from kinect." *Proc., Proc. Intl. Conf. on Advances in Computer-Human Interactions*, 101-108.

Stoyanov, T., Mojtahedzadeh, R., Andreasson, H., and Lilienthal, A. J. (2013). "Comparative evaluation of range sensor accuracy for indoor mobile robotics and automated logistics applications." *Robotics and Autonomous Systems*, 61(10), 1094-1105.

Xia, L., Chen, C. C., and Aggarwal, J. K. "Human detection using depth information by Kinect." *Proc., CVPR 2011 WORKSHOPS*, 15-22.