Privacy-preserving Human Behaviour Monitoring Through Thermal Vision

Abdallah Fahmi Nemer Naser

Department of Computer Science

A thesis submitted in partial fulfilment of the requirements of
Nottingham Trent University for the degree of

Doctor of Philosophy

April 2022

This thesis is dedicated to those who have ground me to believe education is the passport to the future, my parents. I hope my success today gives you happiness after all the life hardships you have endured for the sake of raising my five siblings and me.

Acknowledgements

"Whoever is not grateful to the people, he is not grateful to Allah"
            -Prophet Muhammad (peace be upon him)

In the past couple of years, the research community has experienced the most severe living anomalies caused by the consequences of COVID-19 disease. From lockdown to virtual science communication, I have been fortunate to have a leader who significantly contributed to making my PhD research a success story. I owe my utmost sincere gratitude to the father of my PhD journey, Professor Ahmad Lotfi, for the continued supervision support, motivational words, inspiring actions and sage advice he offered with kindness during my PhD research. I believe the freedom of research he encouraged has improved my independent research skills that resulted in original contributions to knowledge.

I would also like to express my sincere gratitude to my external supervisor, Dr Junpei Zhong, for his support. Despite the time zone difference, he offered regular supervisory meetings, which was an authentic space to think out loud. Thanks to Dr Jun He and Dr Mufti Mahmud for providing constructive feedback during the annual review progress meetings.

What I am today is because of my father and my mother. I cannot express enough gratitude for what you sacrificed for me. My appreciation also goes to my other family members who have stood by me in difficult situations. To my uncle Mr Yousef Khanfar for his support during the financial difficulties, I faced just before starting my PhD.

I see the PhD degree as the capstone of a building. To achieve this crowning, there must be excellent builders at different stages of work. In this regard, I would like to thank all my teachers who taught me at different stages of my life in Palestine, Cyprus and the United Kingdom. Specifically, I want to express my sincere gratitude to Dr Sanaul Hoque, Dr Chee Siang Ang, Professor Farzin Deravi, and Dr Meryem Erbiklek for their motivation to pursue my PhD studies. Your confidence that I would be an excellent researcher was always in mind.

I want to thank all my friends and former colleagues who have kept in touch with me and made me not feel isolated. I would also like to appreciate my colleagues in the Computational Intelligence and Applications Research Group for the time we have shared in the past years. A special thanks to my good friend and neighbour, Dr Hamida Aljaridi, for sharing her PhD experience with me and her successful trials of cooking Palestinian dishes.

Last but not least, I have been fortunate to be awarded academic scholarships for all my higher education degrees. In particular, I am grateful to Nottingham Trent University for funding my PhD study through a fully-funded scholarship scheme. Without this scholarship, I would not be able to achieve this milestone in my academic life.

Abdallah Naser
April 2022

Abstract

Despite the abundance of human-centred research to support domestic human behaviour monitoring in various vital applications, there are still notable limitations to deploying such systems on a broader scale. The main challenge is the trade-off between privacy, performance, and cost of assistive technologies to support older adults to live independently in their own homes. For example, the traditional vision-based sensing approach provides excellent performance while violating human privacy in domestic environments. In contrast, the ambient sensing approach, e.g., employing Passive Infra-Red (PIR) sensors, maintains human privacy but suffers significant performance hindrances in realistic scenarios such as multi-occupancy environments.

This research proposes to utilise the Thermal Sensor Array (TSA) to adjust the trade-off between privacy and performance in domestic environment applications. The rationale behind proposing this sensor for human behaviour monitoring applications is its claimed advantages to perform well while maintaining human privacy, low-cost, and non-contact capabilities. Nevertheless, there has not been sufficient related work to empirically validate the hypothesis of using this low-resolution imager in domestic monitoring. Furthermore, most published works that use the TSA have not yet reached the deployment stage due to the TSA sensing constraints. In particular, TSA is sensitive to environmental thermal noise, and its Field of View (FoV) is not wide enough to cover a large inspection area. Intelligent algorithms should be employed in order to avoid these limitations.

The focus of this thesis is to investigate the human physiological and behavioural thermal patterns for privacy-preserving human

behaviour monitoring to support the independent living of older adults in a multi-occupancy environment by using TSA. This will be achieved through signal processing and machine learning techniques. To achieve this aim, the research methodology is drawn into two main directions. First, human physiological processing of the human thermal signal. Second, human behavioural processing of the human motion signal. This drawn methodology resulted in four main novel contributions.

The first novel contribution of this research is to propose an adaptive segmentation of the human physiological presence and count the number of people from different sensor placements, indoor environments, and human-to-sensor distance. The second contribution is to extract localisation knowledge of the human physiological appearance in terms of human-to-sensor distance and human-to-human distance. Extracting human localisation knowledge is also applicable in other applications such as caregivers and care time monitoring. The third contribution is to fuse multiple TSAs to cover a wide inspection area, e.g., private or care homes. Hence, objects that appear in the low-resolution thermal images acquired from TSA have low intra-class variations and high inter-class similarities, making the identification of the overlapping regions through matching a comparable template image in multiple images very difficult. This research proposes a motion-based approach to fuse multiple TSAs and learn the domestic environment layout with a privacy improvement of utilising TSA in potential monitoring applications running in the cloud. Inspired by the results from this stage of the research, the fourth contribution of the research presented in this thesis is a human-in-the-loop fall detection approach in the Activities of Daily Living (ADLs) that reduces the false-positive alerts while keeping the false-negative fall predictions as low as possible. The novel solutions and the results presented in this thesis demonstrate a significant contribution toward enabling privacy-preserving human behaviour monitoring.

Publications

As a result of the research presented in this thesis, the following publications have been published:

Refereed Journal Papers:
Abdallah Naser, Ahmad Lotfi, Maria Mwanje and Junpei Zhong. "Privacy-preserving, thermal vision with a human in the loop fall detection alert system" **IEEE Transactions on Human-Machine Systems** (2022).

Abdallah Naser, Ahmad Lotfi, and Junpei Zhong. "Multiple thermal sensor array fusion toward enabling privacy-preserving human monitoring applications." **IEEE Internet of Things Journal** (2022).

Abdallah Naser, Ahmad Lotfi, and Junpei Zhong. "Calibration of low-resolution thermal imaging for human monitoring applications" **IEEE Sensors Letters** (2022).

Abdallah Naser, Ahmad Lotfi, and Junpei Zhong. "Towards human distance estimation using a thermal sensor array." **Neural Computing and Applications** (2021).

Abdallah Naser, Ahmad Lotfi, and Junpei Zhong. "Adaptive thermal sensor array placement for human segmentation and occupancy estimation." **IEEE Sensors Journal** (2020).

Refereed Conference Papers:

Abdallah Naser, Ahmad Lotfi, and Junpei Zhong. "Contactless sleep quality monitoring through thermal vision." The 15th PErvasive Technologies Related to Assistive Environments Conference (submitted)

Abdallah Naser, Ahmad Lotfi, and Junpei Zhong. "A novel privacy-preserving approach for physical distancing measurement using thermal sensor array." The 14th PErvasive Technologies Related to Assistive Environments Conference (2021)

Abdallah Naser, Ahmad Lotfi, Junpei Zhong, and June He. "Heat-map based occupancy estimation using adaptive boosting." IEEE International Conference on Fuzzy Systems (FUZZ-IEEE) (2020)

Abdallah Naser, Ahmad Lotfi, Junpei Zhong and Jun He. "Human activity of daily living recognition in presence of an animal pet using thermal sensor array." The 13th PErvasive Technologies Related to Assistive Environments Conference (2020)

Refereed Poster:

Abdallah Naser, Ahmad Lotfi, and Junpei Zhong. "A Novel Privacy-Preserving Approach for Physical Distancing Measurement" NTU School of Science and Technology Annual Research Conference 2021 (selected as one of the three best posters)

Contents

Nomenclature

Acronyms

AmI	Ambient Intelligence
HCI	Human-Computer Interaction
ADL	Activity of Daily Living
IoT	Internet of Things
ROI	Region of Interest
FoV	Field of View
RGB	Red-Green-Blue
IRT	Infra-Red Thermography
TSA	Thermal Sensor Array
RF	Radio Frequency
FIR	Far Infrared Radiation
CSI	Channel State Information
PIR	Passive Infra-Red
WiFi	Wireless Fidelity
RFID	Radio Frequency Identification
mmWave	Millimeter-Wave
GPS	Global Positioning System
iMote2	Wireless Sensor Network Node
ToF	Time of Flight
ML	Machine Learning
MLP	Multilayer Perceptron
ANN	Artificial Neural Network

RNN	Recurrent Neural Network
CNN	Convolutional Neural Network
SIANN	Space Invariant Artificial Neural Network
LSTM	Long short-term memory
Bi-LSTM	Bidirectional long-short term memory
SVM	Support Vector Machine
SVR	Support Vector Regression
k-NN	K-Nearest Neighbour
HMM	Hidden Markov Model
ARHMM	Autoregressive HMM
LHMM	Layered HMM
ANFIS	Adaptive Neuro-Fuzzy Inference System
Adam	Adaptive Moment Estimation
SGDM	Stochastic Gradient Descent with momentum
LDA	Linear discriminant analysis
PCA	Principal Component Analysis

List of Figures

List of Tables

Chapter 1

Introduction

There has been an increase in the ageing population over recent years. According to the World Health Organisation (WHO), the older adult community aged 60+ years is expected to grow from 12% of the total population in 2015 to 22% in 2050 worldwide [1]. As a consequence, long-term care expenditures for older adults will increase. Moreover, the acceptability of care homes among older adults is low [2], and they often prefer to stay in their own homes. Therefore, there is a necessity to enhance the autonomy of older adults by finding new alternative human behaviour monitoring solutions that provide them with independent living in their own homes.

The sensor technologies which could be used to acquire information related to human behaviour in a domestic home environment can be classified into three main categories:

a) Wearable-based sensors usually require the users to wear or carry a device perpetually. This is inconvenient for older adults and could be even more challenging for older adults suffering from Dementia or other cognitive impairments, as there is a high probability of forgetting to carry these devices [3].

b) Ambient sensing devices such as Passive Infra-Red (PIR) sensors are installed in a home environment. Such devices preserve privacy but do not generally perform well in multi-occupancy home scenarios [4]. Other privacy-preserving device-free sensing methods, including Wireless Fidelity

(WiFi), Radar, and Radio Frequency Identification (RFID), suffer from notable limitations in domestic human monitoring applications such as vulnerability to environmental interference [5, 6].

c) Vision-based sensing, for example, cameras perform very well in real-world scenarios, although it violates users' privacy, clearly in domestic environments, e.g., homes, care homes etc.

Conversely, a deployable domestic human behaviour monitoring solution to meet the urgency of economic and societal requirements of older adults should be accepted by the older adults themselves as well as the care service providers. That is, human-centred systems for domestic monitoring should meet the following stakeholders' acceptability factors:

- Impacts, the proposed system should contain applicable solutions that have real economic and/or social impacts,

- Privacy-preservation, the system should maintain the privacy of its users,

- Reliability, the system should be reliable to perform its tasks in realistic domestic environments that may contain more than one occupant,

- Convenience, the system should operate autonomously without interfering with normal human activities at reasonable installation cost,

- Accountability, systems should be accountable to the users.

This research thesis undertakes these acceptability factors through utilising a privacy-preserving, non-contact, low-cost, high-performance, and low-resolution thermal imager, referred to as Thermal Sensor Array (TSA), for domestic human behaviour monitoring. Further, this research contributes to solving major limitations of TSA to operate in a different multi-occupancy environment with a multi-sensor signal processing approach. Important derived decisions from the proposed approach, such as abnormal behaviour alerts, are accountable to the human user by proposing a novel human-in-the-loop approach to significantly reduce false-positive alerts while keeping the false-negative fall predictions low.
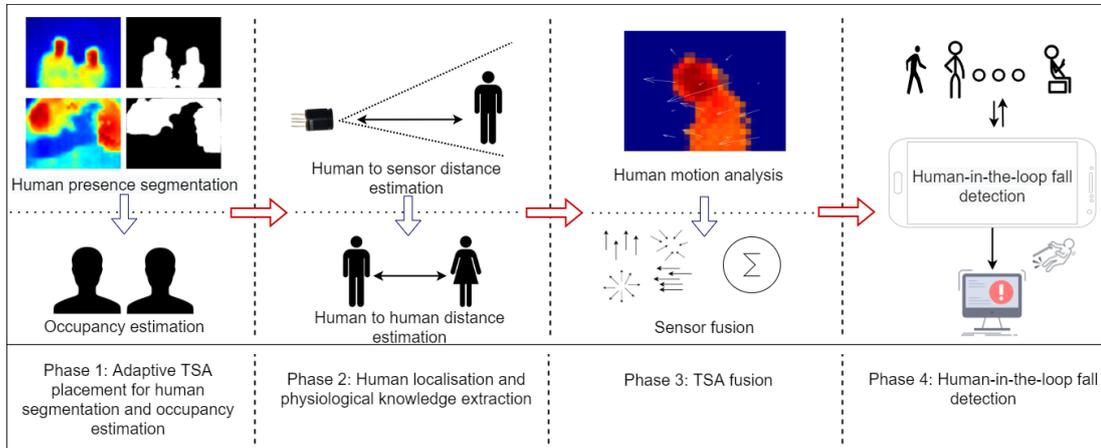
Figure 1.1: Schematic representation of the main functional phases design proposed for domestic human behaviour monitoring.

The rest of this chapter is structured as follows: an overview of this research is demonstrated in Section 1.1 followed by the research aim and objectives in Section 1.2. Section 1.3 introduces the major contribution of the thesis. Finally, the remaining thesis chapters are outlined in Section 1.4.

1.1 Overview of The Research

The prime motivation of this research is to enable human behaviour monitoring in applicable real-life scenarios. It is neither deployable nor impactful to propose a human monitoring system on the assumption that humans live in a single inhabitant environment. In fact, residential homes are occupied by an average of 3.14 people per household [7]. Nevertheless, some older adults may live alone, but they could still have visitors or care assistants visiting their homes at certain times. This causes the system to fail to operate or erroneously send abnormal alerts to the information support once more than one person has occupied the environment. On the other hand, abnormal human behaviours are unpredictable and may even be more challenging to collect actual abnormal behaviour data in a controlled lab environment. Therefore, it is essential to address a valid issue concerning the users' accountability to the system's decision in human behaviour monitoring applications.

This research methodology is drawn into two parts. The first part supports the monitoring of human behaviour through the analysis of human physiological thermal signals. The second part analyses human behaviours by processing human thermal motion signals. In total, four sequential functional phases have been proposed towards enabling potential human behaviour monitoring that applies to real-world scenarios. Figure 1.1 illustrates the proposed research design to enable an impactful human-centred monitoring approach. The design phases are defined as follows:

- Adaptive TSA placement for human segmentation and occupancy estimation: placing TSA in different locations and new domestic environments can pose a significant challenge due to the change in the human shape presence from different sensor placements and environment ambient temperature. In this thesis, a novel framework based on a deep convolutional encoder-decoder network is proposed to address this challenge in real-life deployment. The framework presents a semantic segmentation to segment the human presence and counts the number of people from different sensor locations, domestic environments, and human-to-sensor distance.

- Human localisation and physiological knowledge extraction: this thesis presents discrete and continuous distance estimators to extract human localisation knowledge based on their physiological presence in the TSA's Field of View (FoV) in terms of human-to-sensor distance and human-to-human distance. Also, it proposes a real-time distance-based field of view classification through a novel image-based feature.

- TSA fusion: a motion-based approach is proposed to fuse multiple TSAs and learn the domestic environment layout to enable further human behaviour monitoring applications to operate in different environment layouts. Besides, a privacy-improvement in utilising these TSAs in a centralised care service system is proposed.

- Human-in-the-loop fall detection: building on the previous steps, a novel human-in-the-loop fall detection approach in the Activities of Daily Living

(ADLs) is developed in this phase.

1.2 Research Aim and Objectives

The aim of this research is to enable TSA sensors to perform privacy-preserving human behaviour monitoring to support the independent living of older adults in a domestic multi-occupancy environment. This involves the combination of two concepts, human thermal physiological learning and human thermal motion learning to be used to overcome limitations of the utilised sensing technology, extract human-related localisation knowledge, detect abnormal behaviours and finally give people the autonomy of action through intelligence knowledge derived by computational intelligence techniques. To achieve the project aim, the following research objectives have been identified:

1. Investigate existing sensing technologies for domestic human behaviour monitoring and propose a suitable sensor to be utilised in the system's acquisition stage.

2. Calibrate between different thermal imagers to infer the opportunities and limitations of TSA in human-centred applications.

3. Propose a technique for human segmentation from different sensor placements and domestic environments.

4. Investigate how to identify the multi-occupancy environment and determine the number of people in an environment that could potentially contain an animal pet.

5. Explore human distance estimation techniques to localise multiple human subjects in the environment.

6. Propose a multi-TSA fusion approach to enable human behaviour monitoring to operate in large and different domestic environment layouts.

7. Investigate the capability of TSA to detect abnormal behaviours in ADLs and propose a human-in-the-loop approach to boost the system reliability and accountability.

1.3   Original Contributions

The major contributions of this thesis are summarised as follows:

- The utilisation of the TSA for human behaviour monitoring to adjust the trade-off between privacy, performance, and cost. This is achieved at the sensor level of the widely proposed human-centred applications in a domestic environment.

- An empirical calibration between low- and high-resolution thermal imagers with privacy assessment of low-resolution TSA sensors to observe the opportunities, limitations, and future trends of using TSA in human monitoring applications.

- A novel use of a deep encoder-decoder convolutional neural network to segment the human presence from a low-resolution TSA's output. The proposed human segmentation approach has the capability to segment the human presence from a distance of up to 9 meters.

- A robust and adaptive occupancy estimation framework, able to estimate the occupancy from different sensor placements, human-to-sensor distance, human-to-human distance, and in an unseen noisy domestic environment. Also, an investigation of the TSA operating distance for human presence acquisition has been conducted.

- A novel real-time feature to classify the sensor's Field of View (FoV) into distance-based regions.

- A novel continuous distance estimation approach to estimate the distance between the sensor placement and the human location using Artificial Neural Network (ANN) and a discrete distance estimation approach to predict human distance in a step of $0.5\,m$.

- A new human-to-human distance estimation is proposed. This is referred to as physical distance estimation. Besides, A transfer application to predict human height using the proposed continuous distance estimator has been developed.

- A novel approach to fusing multiple TSAs to cover a wide inspection area to enable further TSA-based human behaviour monitoring applications to run in a multi-sensor processing approach.

- A novel approach to identify overlapping regions between two or more low-resolution TSAs. The proposed approach is adaptive to work in different domestic layouts and sensor placements through proposing an environmental layout learner.

- An improvement for the TSA privacy feature in human behaviour monitoring applications to avoid the human image being reconstructed by a third party during data transmission and storage in the cloud.

- An exploration of the efficiency of optical flow features in ADL recognition and abnormal human behaviour detection using TSA. Besides, comprehensive experiments are conducted to validate the use of optical flow features with Long short-term memory (LSTM) and Bidirectional long-short term memory (Bi-LSTM) in the prediction stage.

- A new accountable human-in-the-loop fall detection approach is proposed. It maintains a high-performance fall detection and reduces human fall false-positive alerts reported to potential central information support by enabling a human-interaction interface.

The outlined contributions of the thesis are addressed in different chapters of this thesis. A summary of these chapters is presented in the following section.

1.4   Thesis Outline

This thesis consists of nine chapters. The organisation of the thesis structure is further illustrated in Figure 1.2. The contents of this thesis are summarised as follows:

Chapter 2: Literature Review - This chapter provides an overview of previous work in the field of human behaviour monitoring and its crucial aspects.
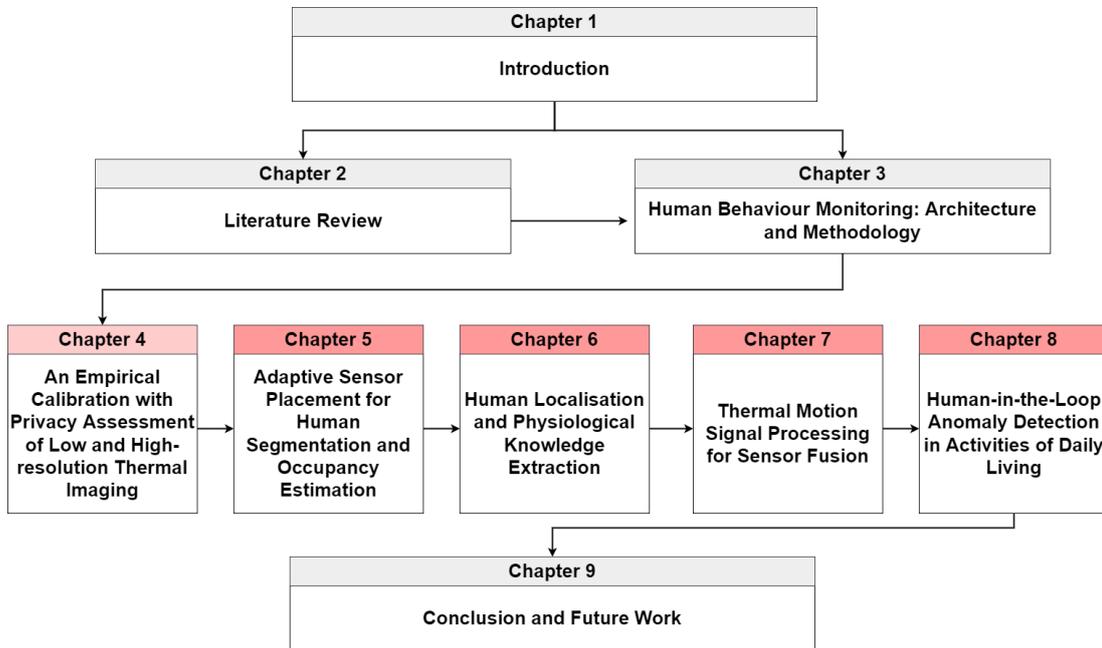
Figure 1.2: Thesis structure indicating the organisation of the chapters and their respective dependencies.

Chapter 3: Human Behaviour Monitoring: Architecture and Methodology - The proposed monitoring scheme architecture, the utilised sensing technology, and the data collection scenarios are provided in this chapter.

Chapter 4: An Empirical Calibration with Privacy Assessment of Low- and High-resolution Thermal Imaging - This chapter presents the calibration of various thermal imaging resolutions in order to provide a better understanding of the opportunities, limitations, and future trends of using TSA in human-centred applications. Precisely, this chapter provides fundamental empirical-driven knowledge to Chapters 5, 6, 7, and 8.

Chapter 5: Adaptive Sensor Placement for Human Segmentation and Occupancy Estimation - This chapter presents a framework to enable TSA to semantic segment the human presence from different TSA placements and determine the number of human subjects appearing in TSA's output accordingly.

Chapter 6: Human Localisation and Physiological Knowledge Extraction - This chapter introduces a novel localisation technique to extract physiological knowledge of human presence in the TSA's FoV. Precisely, to estimate human-to-sensor and human-to-human distances using a single TSA without relying on a second reference sensor.

Chapter 7: Thermal Motion Signal Processing for Sensor Fusion - This chapter presents a novel approach based on the apparent motion pattern of moving objects to fuse multiple TSAs and learn the domestic environment layout to enable further human behaviour applications to run in a multi-sensor processing approach. Besides, a privacy improvement of TSA output prior to integrating them in a central information support platform is introduced in this chapter.

Chapter 8: Human-in-the-Loop Anomaly Detection in Activities of Daily Living - This chapter proposes a novel human-in-the-loop fall detection approach in ADLs. The motivation for enabling a human interactive model, fall detection confirmation, is to influence resource efficiency by reducing false-positive alerts while keeping the false-negative fall predictions low.

Chapter 9: Conclusion and Future Work - This chapter presents the conclusions arising from the thesis and suggests directions for future work on monitoring human behaviour using TSA sensors.

Chapter 2

Literature Review

2.1   Introduction

As research into applications of human-centred monitoring has been an attractive and fruitful area of research, this has resulted in many published research works in recent years. The number of research work published in the field of occupancy monitoring has increased dramatically from less than 20 publications per year between 1981 and 2003 to more than 250 in 2021 [8]. This rapid growth indicates an increased demand for monitoring applications of human behaviour in domestic environments to meet socio-economic needs. In particular, to reduce the cost of long-term care for older adults and support their independent living [4, 9, 10, 11]. Therefore, it is essential to review the state-of-the-art that supports domestic human behaviour monitoring to justify the intent of this thesis research work. This chapter provides a comprehensive review of the previous work by critically analysing their proposed sensing technologies and data-driven methodologies.

The remainder of this chapter is organised as follows: Section 2.2 gives an overview of the utilised sensing technologies in human monitoring applications. To give a general understanding of data-driven methodologies to signal process the output of the sensors, Section 2.3 provides background on widely used learning and prediction models. In Section 2.4, human-centred applications related to the concept of this thesis have been critically analysed to conclude the research gaps in Section 2.5.

2.2   Sensing Technologies

Due to the rapid growth in the number of sensing technologies that have emerged in the context of monitoring human behaviour in domestic environments, this thesis categorises the used sensors in monitoring human behaviour into three different categories (ambient, vision, and wearable). Figure 2.1 visualises the sensor categories and their most prominent sensors. A summary of sensor categories is provided below.

2.2.1   Ambient Sensors

Ambient sensors are an example of a non-contact sensing approach that does not require the users to carry or wear any specific device. PIR sensors are one of the most widely used ambient sensors in domestic human monitoring. The working principle of the PIR sensor is based on using a pair of pyroelectric sensors to detect the heat energy in the ambient environment. The two pyroelectric sensors sit beside each other, and the change in the signal differential between them indicates a warm object motion. Thus, the output of PIR sensors is a digital (binary) signal that implies either a trigger of a new movement or not. The new movement term implies that PIR sensors cannot detect stationary subjects as they do not have motion. This leads to the failure of PIR sensors in human-centred applications that could have humans with inactive states, for example, sleep or rest. Nevertheless, PIR sensors has been proposed for various applications including ADLs recognition [4, 12], user localisation [13, 14], gait velocities analysis [15, 16], sleeping and night activities monitoring [17, 18]. Besides, PIR sensors are typically integrated with other sensors to detect the presence of human subjects, such as pressure sensor that is attached to beds or chairs, door sensor, and floor sensors [4].

Acoustic sensors are also employed in Ambient Intelligence (AmI) towards human detection and human-related action recognition through acoustic signal processing [19, 20, 21, 22]. The rationale behind utilising such sensors in human-centred applications is that human action usually generates a sound interaction that represents the performed human action, such as the sound produced while handling dishes or during a person's fall. However, there are serious privacy
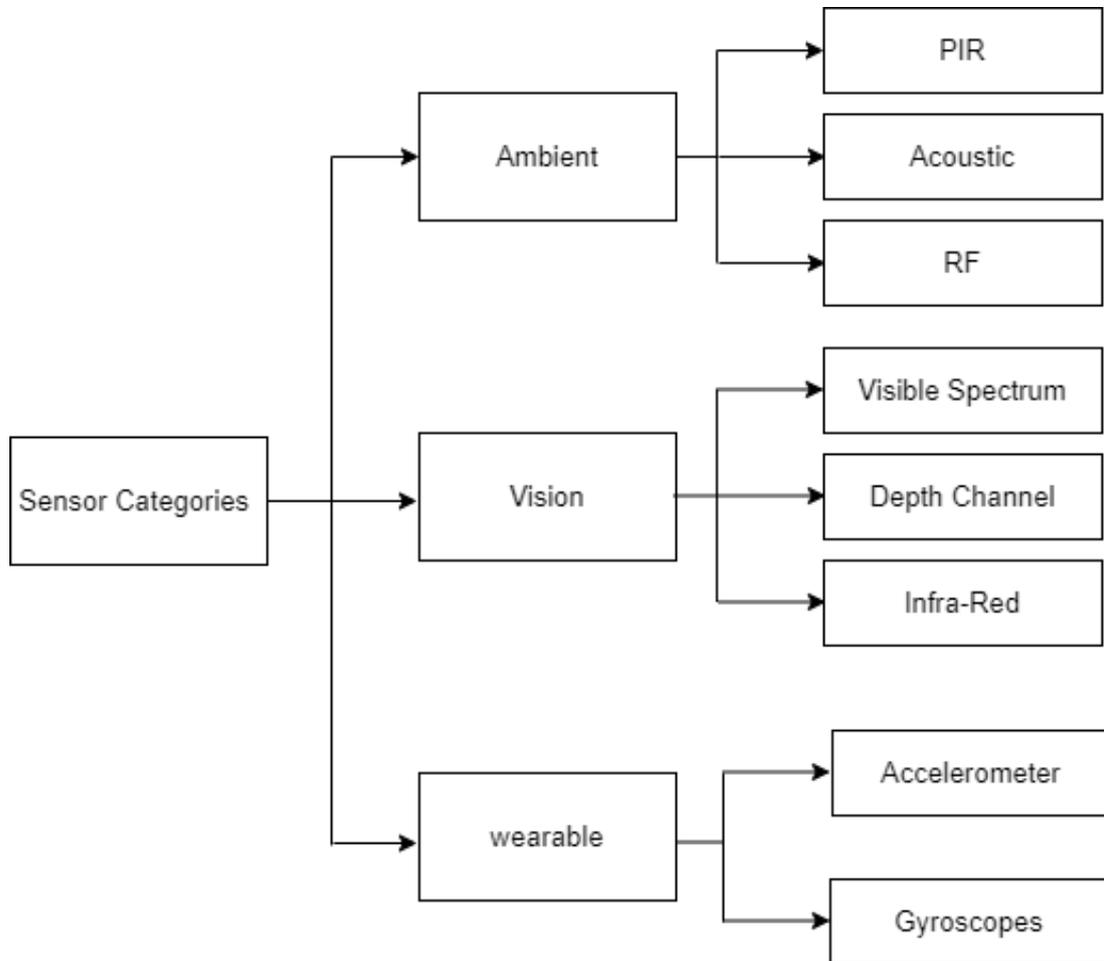
Figure 2.1: The categories of sensing technologies with examples of the most prominent used sensors in domestic human behaviour monitoring applications.

concerns regarding the installation of acoustic sensors in domestic environments, e.g., older adults' homes.

Most Radio Frequency (RF) sensors fall into the ambient sensing approach category and are getting popular due to their contact-less nature and privacy-preserving features. Some of the most commonly used RF technologies are Wireless Fidelity (WiFi), Radar, and Radio Frequency Identification (RFID). WiFi technology has emerged in the shift of human behaviour monitoring research from a device-bound approach to a device-free approach through exploring the properties of wireless networks, e.g., Channel State

Information (CSI) [23, 24, 25]. Furthermore, the radar is also an RF-based technology that monitors contact-less human behaviour by employing a signal reflection approach. Technically, it transmits a radio signal, which is reflected by the objects in the path before receiving it again from the radar to form an image of the path objects using the differences between the transmitted and reflected signals [26, 27, 28]. Cross and environmental interference are likely to occur between RFID systems and WiFi or personal area networks such as Bluetooth when they share common frequency bands.Table 2.1 summarises the advantages and disadvantages of RF technologies in domestic human behaviour monitoring applications.

2.2.2 Wearable Sensors

Wearable sensors can measure motion-related activities such as gait, ADL activity, and human fall. Nonetheless, the success in wearable sensors has been a mix of setbacks and progress [29]. One of the technical barriers when utilising wearable sensors in human behaviour monitoring is the obstruction of feature extraction from the signal due to artefacts, body movement or respiration that need to be resolved to obtain high-quality signals [30]. On the other hand, wearable-based biosensors rely on specific body postures or on-body placement to provide reliable measurements [31]. Some of the widely used sensors in wearable devices and smartphones are the Accelerometer and Gyroscopes sensors. These sensors, with other wearable sensors, have been widely used in various human behaviour monitoring for healthcare purposes [32, 33, 34]. Table 2.2 provides the requirements for monitoring wearable healthcare devices versus

Table 2.1: A summary of the advantages and disadvantages of RF sensing technologies.

Sensor	Advantages	Disadvantages
RFID	Passive, pervasive, low-cost	Environmental interference
WiFi	Low-cost, pervasive	Environmental interference, performance limitations
Radar	pervasive	Environmental interference, performance limitations

Table 2.2: Wearable health monitoring device requirements vs healthcare type.

Healthcare Type	Care Environment	Required Performance in Healthcare Use	Required Performance in Self-Monitoring	Requirements
Domiciliary care	Patient's home	High	Medium	Portable, robust, ease of use
Hospital care	Hospital	High	Medium	Portable within a hospital setting high accuracy
Wearable health monitoring	Anywhere, any time	Medium	Medium	Small and light, Highly portable and unobtrusive

health care type [35].

Although RFID technology uses RF technology, it is classified in this review as a wearable sensing approach because it requires tagging or carrying an electronic chip. RFID has been initially developed for military purposes to differentiate between friendly and hostile aircraft [36]. Since then, RFID has had momentous advancement in various human behaviour monitoring applications [37, 38, 39, 40]. The RFID sensing technology includes two main parts. The first part is the reader, an antenna-based device that emits radio waves to collect the information from the tags. These emitted radio waves are received and modulated by RFID tags before the reader captures them again through its antenna. The second part of RFID technology is the small electronic chips, which can be easily attached to any object, referred to as tags. These tags also have a chip and an antenna. The tag's antenna acquires the reader antenna's signal, passes it to the chip to introduce changes, and finally sends it back to the reader by the tag's antenna.

2.2.3 Vision Sensors

Vision-based human behaviour monitoring has been the basis for many applications, including healthcare, Human-Computer Interaction (HCI), and video surveillance [41]. However, conventional vision sensing technologies, such as cameras, have serious privacy concerns in domestic environments, which makes this sensing approach unacceptable for many people, including older adults.

Digital camera output images are in a 2D grid of pixels. The pixel is the smallest addressable element in an image with a variable intensity value. Typically, each pixel of the colour image has Red, Green, and Blue (RGB) values to form a colour image. The pixel value ranges from 0 to 255, which means the black image has RGB pixels of $(0, 0, 0)$, while the pure blue image would be $(0, 0, 255)$. On the other hand, the depth camera has pixels with a different value scale. That value is the camera to the acquired object distance, referred to as depth. The detection of the depth is typically computed using the Time of Flight (ToF) sensor, which floods the entire scene with light and calculates depth using the time it takes each photon to return to the sensor. Some depth cameras have both RGB and depth systems, which can output pixels with the RGB and depth values, or RGBD. Finally, an infrared camera (also known as a thermal imager) measures the infrared energy emitted by FoV's objects to form an electronic image showing the acquired objects' apparent surface temperature.

Several studies have been reported for human-centred applications using vision-based sensors [42, 43, 44, 45]. However, the privacy concerns of vision-based sensors in domestic environments are not the only hindrances of this approach. For example, traditional cameras are sensitive to light and cannot operate in a dark environment. In contrast, thermal cameras are light-independent but are a very costly approach. Similarly, the depth camera is robust to light variation, but it is hard to extract the image features. Also, the acquired objects' edges are very noisy.

2.3   Data-Driven Methods

The increase of sensing technologies means more signals are collected with heterogeneous statistical properties. Data-driven approaches are utilised for dealing with the randomness and complexity of human behaviour signals in intelligent domestic environments. To give a better understanding of the data-driven approaches discussed in this thesis, this section provides a critical and background review of some important data-driven methods used in the area of domestic human monitoring. This review grouped the data-driven methods

Table 2.3: A summative review of data-driven methods for domestic human monitoring applications.

Model Type	Algorithm	Advantage	Weakness
Deterministic Model	Poisson distribution	simple model construction	Hindrances in dealing with human behavioural signals The prediction accuracy is low
	Ordinal Logistic		
	Time series		
	Logistic regression		
	Bayesian probability		
	Support Vector Regression (SVR)		
Stochastic Model	Hidden Markov Model (HMM)	It can handle a degree of the randomness of human behaviour High prediction accuracy	The models' setting is complicated and therefore the applicability of the model to real-world problems is low.
	Autoregressive HMM (ARHMM)		
	Layered HMM (LHMM)		
	Dynamic Markov time-window inference		
	Markov Model		
Machine Learning	Adaptive Neuro-Fuzzy Inference System (ANFIS)	High prediction accuracy, applicable in real-time systems	Data should be of high quality and quantity Long training time
	Genitic Programming (GP)		
	Presence Sense (PS)		
	Decision tree		
	K-means		
	Adaptive Boosting		
	RNN		
	SVM		
	ANN		

into three main groups: Deterministic Methods, Stochastic Methods, and Machine Learning (ML) methods. A summary of the critical review on these three groups is provided in Table 2.3 and discussed below.

2.3.1 Deterministic Methods

The deterministic method is a fixed model that represents a low-complexity working principle. The requirement of this method's category is based on long-term human observation. Therefore, data collection should represent all probability distributions of particular human behaviours frequency. Precisely, the occupant behaviours can be modelled based on the human behaviours

probability distribution by finding a probabilistic model for a specific human subject. Several techniques can be used to find a probabilistic model for human behaviours, including Bayesian estimation [46], time series models [47] and Poisson distribution [48].

Several studies rely on the use of deterministic methods to monitor human behaviours in domestic environments, e.g. office and home environments. However, this method has many limitations when it comes to describing human behaviours. [49]. This is because human behaviours have a high degree of randomness during abnormal situations [50], and therefore, their behaviours can not be modelled on the basis of constant parameters, e.g., daytime.

2.3.2 Stochastic Methods

Due to the uncertainty of human behaviours, the deterministic methods would generally fail to accurately describe stochastic human behaviours. Therefore, previous work has explored the correlation between human behaviours and their specific events in certain environments [8]. This has been achieved by considering human behaviour as a random variable, and the probability of their behaviour state at each time point is determined based on the previous state. Several techniques have been proposed on this concept, including Markov Chain Model [51], Hidden Markov Model (HMM) [52], and entropy measures [12].

In the context of human behaviour monitoring, the Markov chain is based on describing a sequence of possible human behaviour events in which the probability of each event depends only on the previously attained event [53]. This method assumes that an active human among different environmental zones creates a human profile and random mobility between other behavioural states. Therefore, the following behaviour state of the human only depends on their present state with some rules about the behaviour states. Building on top of this, the transition of human behaviour state should be defined in Markov metrics. Although this method has been one of the most frequently used occupancy models in recent years, it is strongly dependent on the time of human movements and their presence in the environment, which hinders the prediction of human activity from one area to another. On the other hand, the

HMM assumes the states of the human behaviours are connected in a Markov chain, but the state of each behaviour is not directly observed. Instead, each human activity state is associated with observable parameters through a probability distribution.

Continually, entropy is a randomness measure that has also been explored in monitoring human behaviour [12, 54] through the use of the randomness variance of single human activity at particular times. This stochastic method is a promising approach in human-centred applications due to its computational and time efficiency. However, the randomness measure of human behaviour can be analogous to different activities of the same person or a result of a multi-occupancy environment in the context of home monitoring.

2.3.3   Machine Learning Methods

Machine Learning (ML) is a data analysis method that automates analytical model building. ML is a branch of Artificial Intelligence (AI) established on the concept that systems can learn from data, identify patterns and make decisions with minimal human intervention. There are three subcategories of ML models. Supervised ML models are trained with ground truth data sets (labelled data), which allow the models to learn and grow more accurate over time. In contrast, the ground truth data sets are not provided in the Unsupervised ML subcategory, and the model is supposed to find patterns or trends by itself [55]. Finally, the working principle of the third subcategory, Reinforcement ML, is based on training the systems through trial and error to take the best action through establishing a reward system [56]. This section provides a fundamental review of essential machine learning techniques used in the following chapters of this thesis. In particular, classification and regression.

Classification is the problem of identifying the classes or categories of a set of observations. There are several classification techniques to deal with the problem of human behaviour recognition. Boosting algorithms are one of these possible methods that seek to boost the accuracy of a given learning algorithm by converting weak learners to strong learners [57]. In this context, a weak learner is a classifier that performs relatively poorly in classification and is
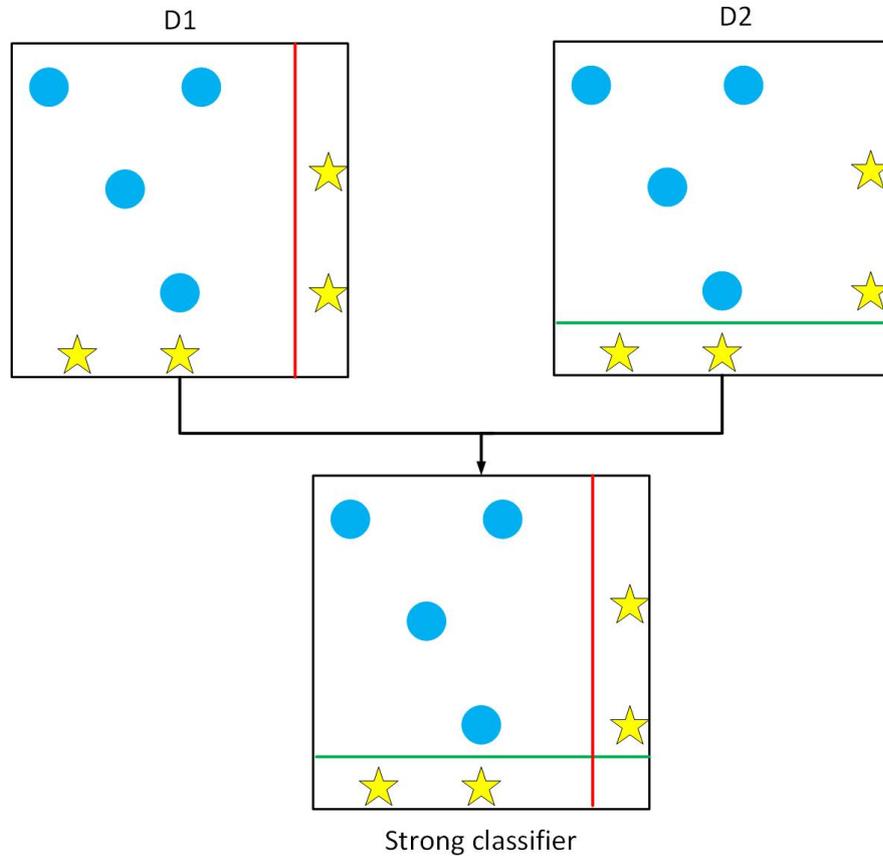
Figure 2.2: An illustration of adaptive boosting applied to a binary classification problem.

slightly better than a random guess. In contrast, a strong learner can label the testing examples (observations) more accurate than weak classifiers.

In Adaptive Boosting (AdaBoost) [58], the weak learners are decision trees with a single split, referred to as the decision stump. The prediction model in AdaBoost improved through training the weak learners sequentially. Each of these weak learners aims to correct its predecessor. The weights of the observations in the first decision stump are equal. In the next iteration, the incorrect observations that were inaccurately classified in the previous round carry more weight than the true classified observations to force the weak learner to focus on the hard samples in the training set. Figure 2.2 shows an illustration of a simple binary classification problem using AdaBoost. The first

decision stump ($D1$ and $D2$) separates stars from circles. In this separation, there are two misclassified stars. These incorrectly rated stars will carry more weight than others to feed the second learner. Combining these two learners leads to a strong final classifier that correctly classifies the objects.

Unlike the procedure of classification predictive modelling to approximate a mapping function from input variables to discrete output variables, the regression approach is the process of approximating a mapping function from input variables to a continuous output variable [59]. In other words, the regression output is a real-value, e.g., an integer or floating-point value. These are often quantities, such as human-to-sensor distance. One of the simple regression algorithms is linear regression, which aims to find a linear relationship between the dependent variable and one or more independent variables using a best-fit straight line [60]. Typically, the prediction of the linear model is simply computed by a weighted sum of input features and a constant value called the bias term. However, in many problems, the relationship between the input and output variables is not formed in a linear relationship. Thus, there are other techniques to deal with such problems, such as Artificial Neural Network (ANN), that can be used for both regression and classification [61].

Various types of ANN have been presented in the literature, including Feedforward Neural Network [62]. This network is the purest form of ANN that allows the data to travel in one direction. This type of ANN can have hidden layers, and the data enters through input nodes and exits through output nodes. The classifying activation function is also used in this type of ANN. However, there is no backpropagation as it only supports front propagated waves. Another important class of ANN is Convolutional Neural Network (CNN), commonly used in image processing and computer vision applications to analyse visual imagery [63]. This type of ANN is also known as a shift-invariant or Space Invariant Artificial Neural Network (SIANN). CNN models cannot effectively interpret temporal information. Thus, other types of ANN have been proposed to process sequential data, e.g., Recurrent Neural Network (RNN), that process sequence-based data [64]. Hence, CNN employ filters with convolutional layers to transform the data. In contrast, RNN models are predictive and reuse activation functions from other data points in the sequence

to generate the subsequent output in a series.

2.4 Domestic Human-centered Applications

The TSA sensors have recently become important through emerging them in responses to the global COVID-19 pandemic, for example, TSA-based occupancy estimation and temperature scanning systems [65, 66, 67, 68, 69]. Nevertheless, the usage of the TSA for human-centred applications is relatively low [70] compared to other sensing methods. This section includes a TSA-focused comprehensive critical review of related human-centred applications to identify previous research work gaps.

2.4.1 Occupancy Estimation

Research has been conducted to investigate the methods for counting the number of occupants, referred to as occupancy estimation, in domestic environments for different purposes using different sensing approaches [71, 72]. The authors in [73] have proposed a system for tracking the elderly using High-performance Wireless Sensor Network Node (iMote2) sensor with an Enalab camera board in smart homes. In their work, they were able to estimate the number of occupants by calculating the peaks within the histograms. The proposed system is based on PIR sensors to detect occupancy in some areas of the home. This makes their system complicated, and privacy concerns are raised here due to the use of the camera too. Other previous works have similar privacy concerns for using the camera to count the people in domestic environments [74, 75].

Other solutions based on wearable sensors have also been suggested. But the designs of wearable sensors are inconvenient to most users. For instance, [76] integrated the PIR sensor with active RFID tags to estimate the occupancy. The main limitation of this work is that users must carry these tags continuously.

The work reported in [12] utilised the PIR sensors to distinguish between the single and multi-occupancy environment to determine the visit time of the older adults in a single inhabitant environment by measuring the randomness of the PIR-based binary data using different entropy measures. Other previous

works were also able to use PIR sensors to identify multi-occupancy domestic environments [77, 78, 79, 80]. However, these works were only able to identify whether more than one person occupies the environment without providing an exact estimation of the number of people. Furthermore, their proposed methodologies relied heavily on the sensor layout and the ground-truth annotated sensor data.

Privacy-friendly sensing approaches reported in [81, 82, 83] have proposed multi-modal systems for counting the number of people in the home environment. They used multiple environmental sensors such as lighting, temperature, movement, CO, $CO2$, and humidity. Although the multi-modal approach increases these systems' performance, it raises serious questions about their applicability in real-use case scenarios. This is because these works have assumed that ventilation does not affect the performance of their proposed systems. However, ventilation may alter the level of humidity, CO, and CO2 in the home environment, resulting in a wrong estimate of the occupancy.

Recently, authors in [84] overcame the PIR sensor's deployability problem to count the number of people in a home environment by introducing a new algorithm, based on continually tracking each person's location in the home, without requiring other additional information such as the ground-truth annotated sensor data. However, it is unclear whether the algorithm will work when two people simultaneously walk alongside each other, for example, the caregiver who helps the older adult to walk. Other PIR-based sensing research has been shown for human monitoring [85, 86, 87, 88, 89]. However, none of these works has considered animal pets that may alter the output of any of these monitoring systems.

Authors in [70] have used a 4×16 thermal sensor array to estimate the occupancy. They removed the background infrared using the per-pixel and standard deviation values for a short occupancy period, and then the K* algorithm [90] was applied to estimate the occupancy. They were able to achieve 82.56% accuracy. It is reported that they could handle a prolonged period of occupancy by using a complex scaling algorithm. However, using per-pixel and standard deviation values to remove the background is not the best solution, because when a newer object with a higher temperature than the

human body enters the sensor environment, the system may view the human body as radiation from the background, which results in an error in estimation.

Beltran et al. [91] suggested a multi-modal system consisting of a PIR sensor and a TSA to estimate the occupancy. The purpose of using the PIR sensor was to detect the empty occupancy environment and the TSA to count the number of people in the environment. However, the proposed system may fail to estimate the occupancy when a person has been inactive for a long time - for example, sleeping as the probability of identifying human radiation as the background temperature increases. Furthermore, the TSA is placed in a fixed location, which results in the system not being worked at a different sensor location. Other works [70, 92, 93, 94, 95, 96, 97] that used the TSA to estimate the occupancy also contain similar sensor location adaptation problems.

2.4.2 Human Distance Estimation

Measuring human-to-human distance, which refers to human physical distancing, can be classified into either passive or active approaches. The passive systems usually rely on the use of vision sensors such as regular cameras or expensive high-resolution thermal cameras [98, 99, 100, 101, 102]. One of the advantages of using vision sensors is that users have no requirements to carry any particular device. They still perform reliably in measuring human physical distancing, thanks to the extensive research in computer vision. However, they raise concerns about people's privacy in domestic settings such as nursing homes, hospitals, or even smart homes to support independent living for older adults.

Recently, there has been a growing interest in active approaches due to their deployability feature, e.g. contact tracing apps, which use Bluetooth proximity sensors to track human's physical distancing. However, there exist problems in performance and reliability [103, 104, 105]. This raises a growing interest in more reliable proximity sensing solutions for sensitive environments [106].

Authors in [107] proposed an Ultra-sound based scanning approach to measure the human physical distance. The reason behind the Ultra-sound sensor is that it has higher accuracy than Bluetooth. However, it is prone to multi-path propagation errors and poses robustness concerns in real-use case

scenarios. Also, there exist other important viable methods, including cellular and Global Positioning System (GPS) [108], capacity body sensing [109], and magnetic field [110]. TSA sensors have been proposed for human positioning applications in several reports [111, 112, 113, 114, 115, 116]. However, none of the previously published works explicitly propose to measure the human physical distance or human-to-sensor distance.

2.4.3 Fall Detection

Conventional wearable devices use an accelerometer's sensor to measure the object acceleration [124, 125, 126]. The usage of wearable sensors for human fall detection utilises the change in motion, location, and posture of the monitored object. In general, wearable device approaches are cost-effective, easy to design, and commission. On the other hand, these need to be worn by the user for accurate results, and they could be inconvenient to use for many users[127]. In addition, individuals are prone to forgetting to wear the device, or situations such as the need to take a shower force users to take off the devices, hence the inability to detect falls accurately, leading to low performance.

In contrast to wearable devices, vision-based methods include the use of ordinary cameras[128, 129, 130], which solves the fixed body device location in the signal acquisition stage. The vision-based approach can be subdivided based on operation, including change in shape, inactivity, head change analysis, posture, and Spatio-temporal analysis [127]. For instance, in the inactivity technique, the user's period of inactivity on the floor contributes to detecting a

Table 2.4: A comparison of TSA type, resolution, placement, and sensor placement adaptability proposed for human fall detection systems.

Ref.	TSA Type and Resolution	TSA Placement	Adaptive Placement
[117]	Panasonic's Grid-EYE (8x8)	Ceiling	No
[118]	Melixis (16x4)	Wall	No
[119]	Panasonic's Grid-EYE (8x8)	Mini-robot	No
[120]	Panasonic's Grid-EYE (8x8)	Ceiling	No
[121]	Panasonic's Grid-EYE (8x8)	Cellinig	No
[122]	Melixis (16x4)	Ceiling, side by side	No
[123]	Melixis (16x4)	Ceiling, side by side	No
[121]	Panasonic's Grid-EYE (8x8)	Ceiling	No

Table 2.5: A comparison of the experimental setup, data-driven approach, and TSA-based fall detection state-of-art performance results.

Ref.	Performed Activities	Learning Algorithm	Accuracy (%)	Accountability
[117]	Standing Sitting Lying	CNN	-	Missing
[118]	Sitting Bending Squatting Walking Standing	k-NN	93%	Missing
[119]	Standing Sitting Picking up	SVM	88,7% - 94,7%	Missing
[120]	Sitting Walking	k-NN	94.3% - 95.8%	Missing
[121]	Walking Jogging Squatting Lying down Staying still	Random Forest	-	Missing
[122]	Sitting Bending Squatting Walking Standing	Voting classifier	97.75%	Missing
[123]	Standing Sitting Lying	Logistic regression	99.94%	Missing
[121]	Staying seated Staying up walking standing up	3D-ConvNet	97.22%	Missing

fall. However, the main drawbacks are light dependency and its violation of people's privacy. Finally, ambient sensors have also been used to detect falls, for example, pressure sensing [131], PIR sensors, and floor vibration [132]. Ambient sensing ensures user privacy, a critical issue in vision-based approaches and is more convenient than wearable-based approaches. However, its detection is affected by all variables within the environment, resulting in low performance. The use of TSA in fall detection has emerged in recent years to bridge the gap between performance and user privacy concerns [120, 133, 134, 135]. A critical comparison of TSA settings, data-driven methods, and performance of the state-of-art TSA-based fall detection systems has been provided in Table 2.4, and Table 2.5.

2.4.4 Sensor Fusion

Unlike PIR sensors, TSA sensors can detect motionless warm objects and moving objects with the direction of their movements within their FoV through employing conventional PIR motion detectors in what is called thermopile array [136]. Therefore, TSA has also been proposed for several human-centred applications [111, 112, 113, 114, 115, 116]. However, none of these works discovers the scenarios of using multiple TSAs. This is a critical issue, for example, in the occupancy estimation systems, which utilises TSA [70, 91, 92, 95, 137, 138]. This is because humans located in the overlapping FoV's of two or more sensors will be counted as two subjects in the prediction stage, which leads to a wrong occupancy estimate.

Similarly, there have been recent works on using the TSA on human activity recognition and abnormal behaviour detection [135, 139, 140, 141, 142]. The approach followed to process the TSA output is similar to image-processing approaches [143] while the analytical techniques on individual time intervals, frames, were different, for instance, Support Vector Machine (SVM) [119], K-Nearest Neighbour (k-NN) [91, 144], decision trees [97, 116], and Kalman filtering [145, 146]. One of the notable technical challenges reported in most human-based applications which use TSA is that external heat sources have a major negative impact on the system performance. On the other hand, TSA's

fusion has not yet been investigated in such applications. Considering falls as abnormal human behaviour that could happen in an overlapped area between two sensors (One sensor would not be sufficient to cover the entire environment, e.g. older adult home). The proposed systems may trigger two separate fall alerts incorrectly. Moreover, identifying overlapped regions between multiple TSAs significantly impacts other human monitoring applications, including occupancy estimation. The impact of this can be clearly demonstrated in situations where a person could be present in the overlapped region of two sensors, and thus the system might consider them as two people in the environment rather than one.

The fusion concepts have been applied in various applications [147], including biometric authentication systems [148, 149, 150], air pollution monitoring [151], COVID-19 non-remedial solutions [152], surveillance networks [153], vehicle accident detection and classification [154], and human emotion monitoring [155].

2.5   Discussion and Research Opportunity

The gaps identified from current research, as discussed in the review, are highlighted in this section. The fundamental research gap is with regards to the trade-off between the privacy, performance, and cost of the widely used sensors for human behaviour monitoring that hinders the deployability of such systems on a large scale. Although this research proposed to fill this gap by utilising a low-resolution TSA due to its low-cost and human privacy-preserving claims, there has not been an independent empirical calibration of low-resolution TSA and high-resolution imagers for human-centred applications. This means that the privacy-preserving feature of TSA sensors has not been experimentally validated to assess the potential for human identification from the raw TSA output. Besides, there has not been sufficient published research work on TSA-based human behaviour monitoring compared to typical sensing-based approaches.

In the context of data-driven methods, previous approaches to human-centred applications using the TSA usually rely on the use of a fixed sensor location to make the human-to-sensor distance and the human presence shape fixed. However, placing this sensor in different placements and new indoor environments can pose significant adaptability challenges. Such poor adaptability raises significant concerns about the deployability of TSA-based applications. For illustration, placing the sensor on the ceiling of the room reduces the sensor's FoV, which means that more sensors will be required to cover a wider area. Furthermore, most of the previous research work on assisted living to help older adults to live independently in their own homes assumed they live in a single residential environment [156]. Therefore, it is essential to add a new functional layer to distinguish between single and multi-occupancy status in smart home solutions to make them applicable to real-life scenarios.

Extracting localisation knowledge of human subjects in the domestic environment is vital to developing a robust human behaviour monitoring system. Previous approaches to estimating the distance between a reference sensing device and a human subject relied on ordinary or high-resolution thermal cameras. Also, an adaptive approach to human distance estimation

using the TSA has not been explored. Furthermore, the main limitation of deploying TSA-based systems on a large scale is the challenge of fusing multiple TSAs to cover a wide inspection area, e.g. residential homes and care homes. On the other hand, objects that appear in the low-resolution thermal images acquired from TSA have low intra-class variations and high inter-class similarities, making the identification of overlapping regions through matching a comparable template image in multiple images very difficult. Thus, there is a necessity to develop a fusion approach to enable multiple TSAs to cover a wide inspection area for an impactful monitoring human behaviour system.

Finally, most of the previous reported work that explored the use of TSA in human abnormal behaviour detection did not take into account the effect of false-positive detection scenarios on the consequent waste of emergency alerting and response resources. This leads to real concerns about the reliability of deploying such systems for multi-user environments with centralised information support. For instance, a human fall is abnormal human activity that can occur from standing, sitting or even unpredicted activity. Therefore, it is crucial to address a valid issue concerning the users' accountability to the system's decision in human behaviour monitoring schemes.

Chapter 3

Human Behaviour Monitoring: Architecture and Methodology

3.1 Introduction

An impactful domestic human behaviour monitoring should be flexible to operate in different domestic environment layouts, sensor placements and inclusion for multi-occupancy environment scenarios. It is also important that the system be reliable and able to solve its intended problems. For example, developing a health-related anomaly detection system to deploy in older adults' homes with centralised information support may create severe problems if the system is not sufficiently reliable. This can be demonstrated in the example of a fall detection system that triggers alerts to the information system upon detecting human falls. A large deployment of such systems without considering user accountability may lead to unreliable responses and actions, e.g., sending too many ambulances for false alarm falls. Similarly, a solution that has been designed on the assumption of a single-occupancy environment could erroneously report abnormal human behaviours in the presence of a visitor.

This chapter facilitates the thesis readability by introducing the proposed sensing technology and system architecture developed in this thesis for an effective domestic human behaviour monitoring scheme. This chapter is structured as follows: Section 3.2 presents the utilised sensing technology for

domestic human behaviour monitoring. Section 3.3 provides the architecture of the proposed monitoring scheme followed by the data collection scenarios in Section 3.4. Lastly, Section 3.5 summaries the chapter.

3.2   Sensing Technology

The TSA sensors are employed as a non-contact sensing approach to measuring the circumference temperature in a specific area. In this thesis, a commercial TSA[1] with the resolution of 32×24 IR array, which makes a total of 768 Far Infrared Radiation (FIR), has been used in the system acquisition stage. The justification behind choosing this TSA's resolution has been based on an empirical calibration of different thermal imagers presented in Chapter 4. Besides, the sensor can be accessed via the $I2C$ interface, and its electric current is less than $23\,mA$. This mA current makes it suitable for a battery-powered solution. The refresh rate of this sensor is between 0.5 and $64\,Hz$, which makes it capable of detecting swift human movements. Figure 3.1 shows a heat-map representation of the TSA output, which contains the acquired temperatures in the form of a matrix. This heat-map contains a single human presence as well as the acquired background temperatures. Based on this visual illustration, it is clear that it is not possible to obtain identifiable human information visually. This assumption has led to the claim that TSA is a privacy-preserving approach.

To determine the temperature of a specific region in the thermal image, this region has to illuminate at least one complete FIR (pixel). Otherwise, the pixel will represent a mixed temperature of the object and the adjacent background. For example, the human presence shown in Figure 3.1 shows a variation in human acquired temperatures. One possible reason for this variance is dressing. However, the facial region of the human presence has also suffered from temperature variance. This is because some of the acquired face temperatures were not fully illuminated by a full pixel.

On the other hand, the parameters used to describe a visible spectrum can also describe the IR-based optical system. The main difference apart from the

---

[1]The sensor details can be obtained from the Melexis website: https://www.melexis.com/en/product/MLX90640/
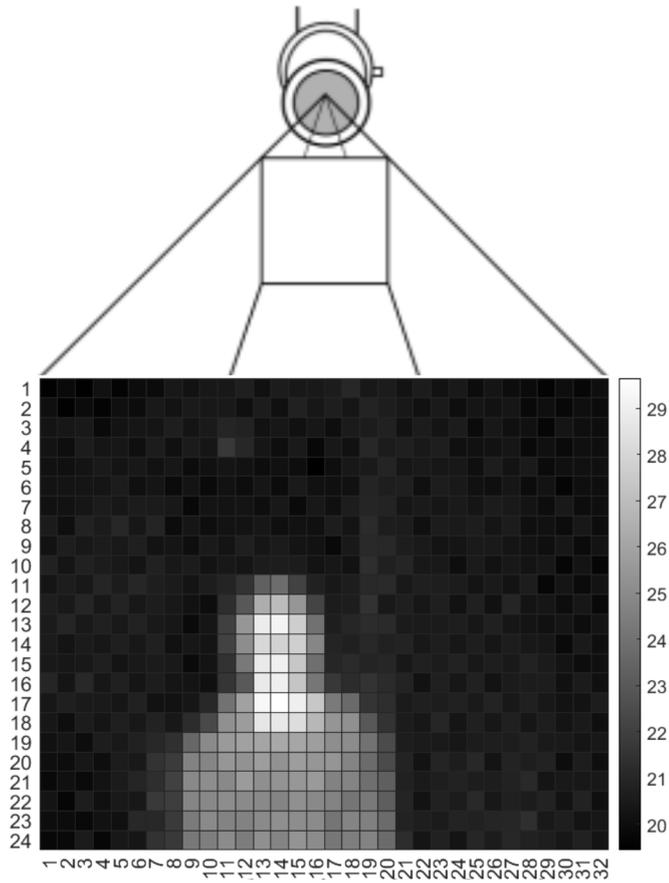
Figure 3.1: A visualisation of TSA spatial temperature matrix representing the presence of a human subject.

wavelength is the material of the lenses. In the context of IR optics, usually, Silicon, Germanium, Zinc Sulfide or Chalcogenide glass show good transparency for the IR spectrum. Thus, these materials are commonly used to produce lenses for infrared optical systems. Hence, ordinary glass is not transparent in the thermal IR spectrum. This explains why the IR optical system, e.g. TSA, fails to acquire human thermal signals blocked by a glass door.

3.3 Proposed Architecture

The focus of this research is to investigate the human physiological and behavioural thermal patterns for privacy-preserving human behaviour monitoring to support the independent living of older adults in a multi-occupancy environment. To achieve this, the research methodology is drawn into two main directions. First, human physiological processing of the human thermal signal. Second, human behavioural processing of the human motion signal. This drawn methodology consists of four novel main functional phases presented in the system architecture's Figure 3.2. An outline of these functional phases is provided below.

3.3.1 Phase 1

Employing TSA to obtain the data requires careful consideration of TSA-related constraints. One of these constraints is the sensor placement and coverage area. For instance, most of the early published TSA-based work proposes a ceiling placement to obtain a fixed human presence shape, and human temperature values [70]. However, this leads to severe engineering-related concerns regarding the cost and the required sensors to cover a large environment. In particular, relying on ceiling TSA placements means more sensors are demanded. Therefore, the first problem solved in this phase is to enable the TSA sensor to segment the human presence from an adaptive sensor placement using appropriate pre-processing, semantic segmentation, and post-processing techniques.

Multi-occupancy environments are also considered while designing the system architecture. Thus, this functional phase provides an approach to distinguish between empty, single and multi-occupancy environments through estimating the number of human subjects in the acquired low-resolution thermal images. The detailed technical concepts regarding this functional phase are provided in Chapter 5.
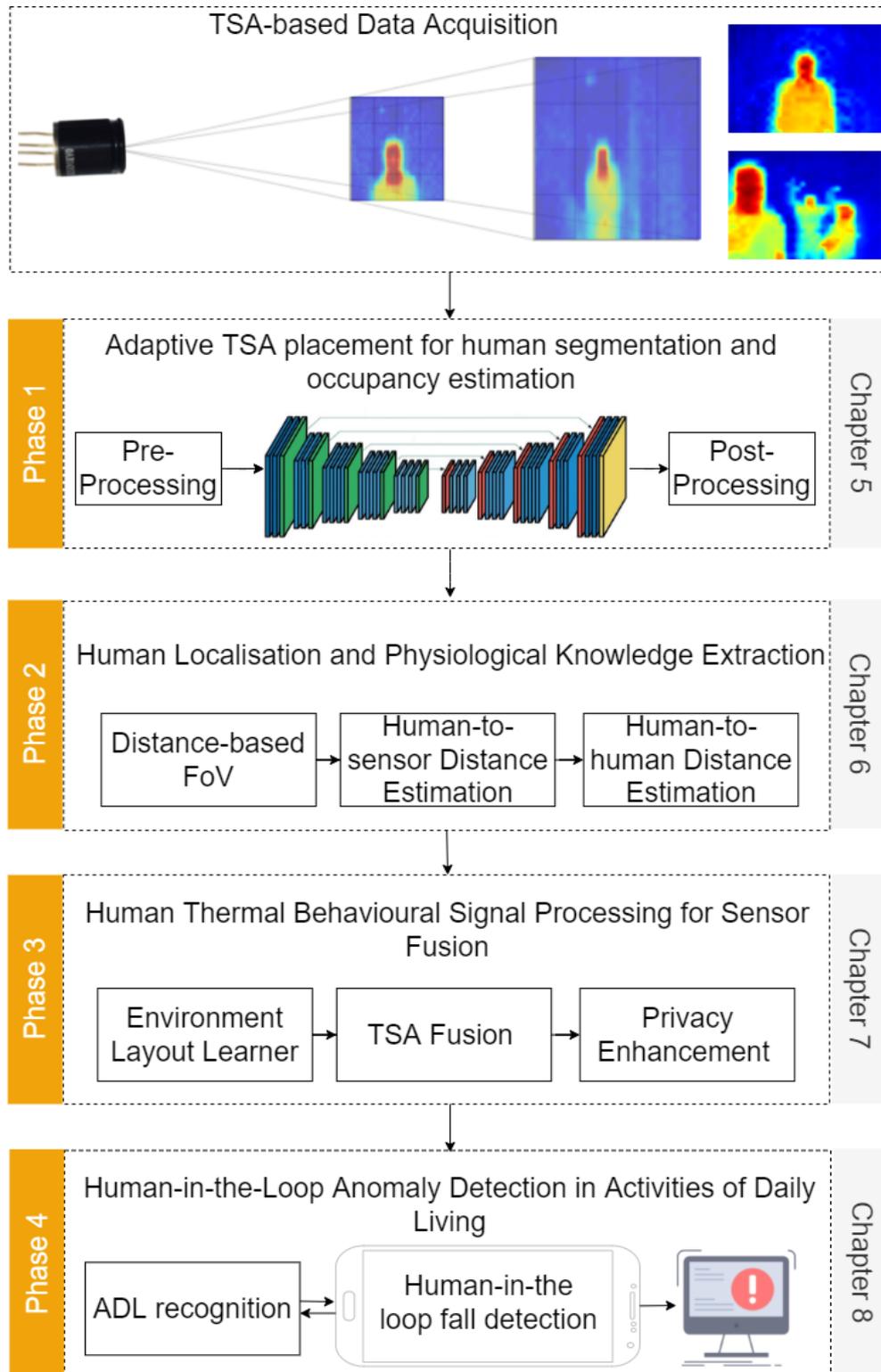
Figure 3.2: The proposed scheme architecture for domestic human behaviour monitoring.

3.3.2 Phase 2

The objective of the first phase was to segment and count the number of people using a flexible sensor placement. Nevertheless, the localisation and physiological knowledge of the human presence have not been extracted. Learning the exact location of the human presence supports tracking human movement in the multi-occupancy environment. Two important values of localisation information have been extracted from the TSA output in this phase: (1) the human-to-sensor distance, and (2) the human-to-human distance. Furthermore, this phase proposes a novel image processing feature to classify the TSA's FoV into depth-based regions to facilitate real-time human localisation.

Chapter 6 presents the detailed description and experiments conducted to evaluate this proposed phase.

3.3.3 Phase 3

Unlike the first two phases' methodology, which was based on a single frame processing, this phase explores human behavioural analysis through a time-series-based TSA signal processing to enable the proposed monitoring scheme to automatically learn the environment layout and identify the overlapping regions between multi-sensors' FoVs. Thus, this chapter is concerned with fusing multiple TSA sensors.

Identifying overlapped regions between multiple TSAs impacts various human monitoring applications, including occupancy estimation systems. The impact of this can be clearly demonstrated in situations where a person could be presented in the overlapped region of two sensors. Thus the system might consider two people in the environment rather than one. Besides, considering falls as abnormal human behaviour that could happen in an overlapped region between two sensors (One sensor would not be sufficient to cover the entire environment, e.g. older adult home). The proposed system may trigger two separate fall alerts incorrectly. More information regarding this phase is provided in Chapter 7.

3.3.4 Phase 4

Inspired by the results of the previous phase, this phase continues to explore human motion analysis in human ADL recognition and identifying human behaviour anomalies. Precisely, to detect human falls in ADLs and promote users' accountability through proposing a human-in-the-loop fall detection system. More information regarding this phase is provided in Chapter 8.

Several experiments and robustness analyses have been conducted to examine the proposed phases' generalisation ability and reliability, such as assessing the ability of the proposed human behaviour monitoring for human subjects during sleep or in the presence of an animal pet.

3.4 Data Collection Scenarios

One of the acceptability factors for good human behaviour monitoring is convenience, as discussed in Chapter 1. To achieve such a system, the system must be easy to use and not interfere with the users' daily activities. On the other hand, the system should also be easy to install in different environment layouts without restrictions on the sensor placements. Ideally, the proposed monitoring scheme should facilitate the system installers to choose where to place the TSA based on their on-site observations. Accordingly, the system should segment the human presence, learn the environment layout, and enable multi-sensor processing from different TSA placements and domestic environments.

A comprehensive data collection has been conducted with different TSA placement scenarios to evaluate the performance of the proposed approach functional phases. A summary of the TSA placement scenarios used to validate each functional phase in the proposed approach is provided below.

- Stage (i) - data collection scenarios: the first data collection scenario is implemented while placing the TSA on a vertical position as shown in Figure 3.3(a). Besides, a ceiling TSA placement, illustrated in Figure 3.3(b), is the second data collection scenario to evaluate the adaptability of the proposed
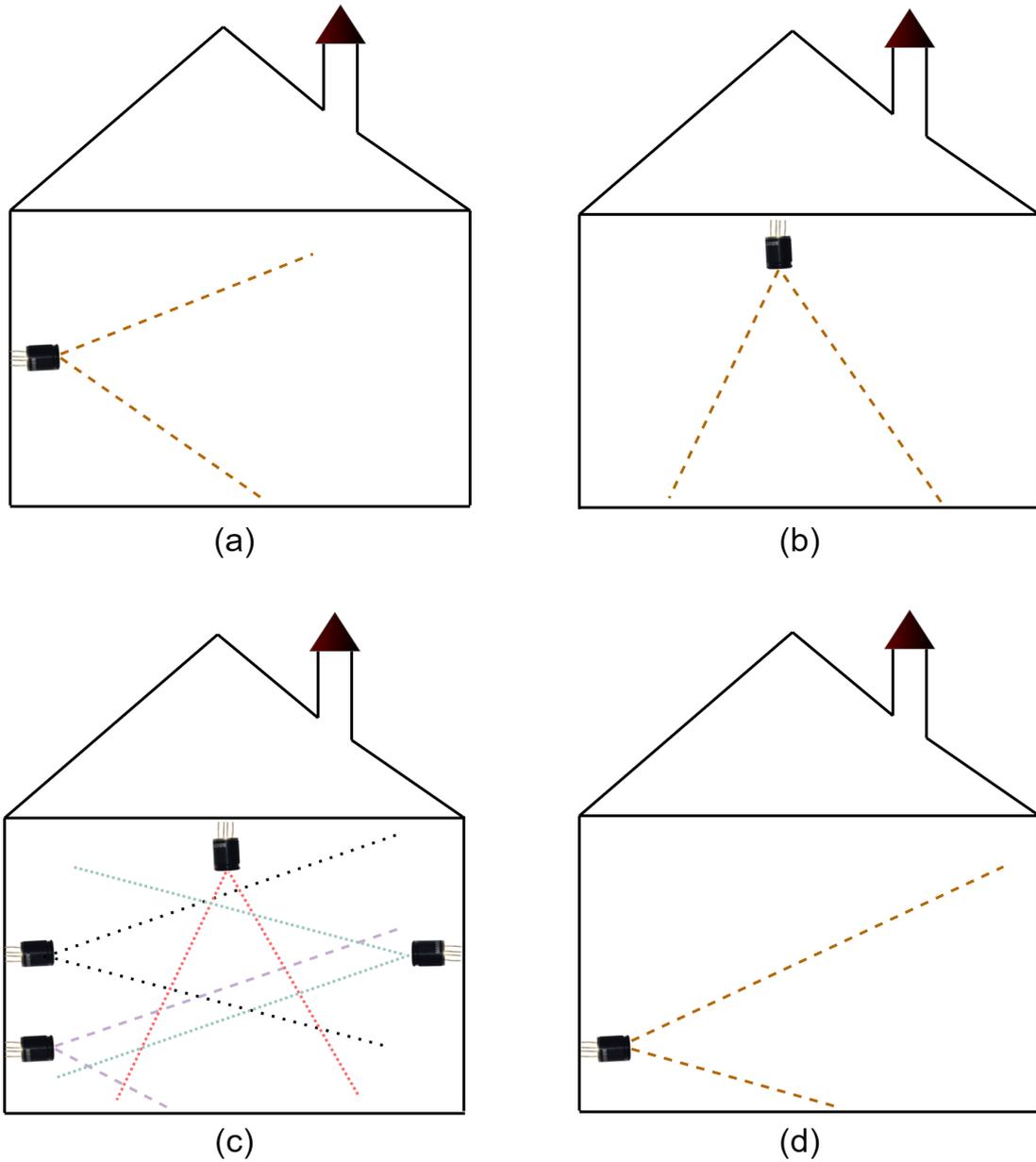
Figure 3.3: The scenarios of TSA placements to evaluate the performance of the proposed human behaviour approach.

human segmentation and occupancy estimation approach to operating with complete unseen data.

- Stage (ii) - data collection scenarios: the data collection scenarios conducted for this stage is based on a wall TSA placement, which is illustrated in Figure 3.3(a). This scenario is more realistic than the ceiling placement scenario for human localisation and distance estimation. Hence, human-to-sensor distance from a wall placement is more challenging than ceiling placement since human-to-sensor distance could be predicted or generalised in-home environment settings. Furthermore, the proposed approach in this stage has also been tested with completely unseen data and from new TSA placement, which is the ceiling visualised in Figure 3.3(b).

- Stage (iii) - data collection scenarios: this thesis finds that the objects appearing in the low-resolution thermal images obtained from TSA have low intra-class variations and high inter-class similarities, making the identification of overlapping regions through matching a comparable template image in multiple images difficult. Thus, various sensor placements illustrated in Figure 3.3(c) has been employed to validate the proposed sensor fusion approach. Specifically, the first use case scenario to evaluate the feasibility of the proposed fusion approach to verifying if the movement of a person acquired from 2 sensors installed side by side at 90° can be determined in a multi-occupancy environment where there could be another person who may perform similar or different activities. The second case scenario is when sensing interference occurs between opposite sensor placements. Third, the interfering sensors are placed on the same wall at different heights. Finally, the interfering sensors are placed on the wall and ceiling.

- Stage (iv) - data collection scenarios: in the last data collection stage, the TSA is placed on a short height wall placement as illustrated in Figure 3.3(c) to evaluate the performance of the series-based signal processing for human ADL recognition and abnormal human behaviour detection. Furthermore, the ceiling placement has also been used to analyse the proposed approach's

robustness in predicting abnormal human behaviours during sleep.

The data collected has been conducted in different domestic environments in the summer and winter seasons of the United Kingdom. The reason for considering different seasons and different domestic environments is that the heating systems in the United Kingdom usually operate during the winter season. In the summer months, neither heating nor cooling is used. These evaluations ensure a high generalisation ability for the proposed human behaviour monitoring approach as the TSA sensor is sensitive to ambient temperature. Finally, all chapters in this thesis that contains data collection have provided a detailed description of the used TSA placement scenarios and how the data has been configured and analysed.

3.5   Chapter Summary

This chapter has presented the utilised sensing technology proposed to monitor human behaviour in domestic environment settings. Moreover, this chapter has abstractly discussed the architecture of the proposed approach phases to facilitate reading the thesis. The TSA placements scenarios used during the data collection have also been discussed.

The next chapter will provide a clear understanding of low-resolution thermal imaging by empirically calibrating different thermal imaging resolutions and evaluating the privacy-preserving claim of TSA imaging.

Chapter 4

An Empirical Calibration with Privacy Assessment of Low and High-resolution Thermal Imaging

4.1 Introduction

Thermal imaging (also referred to as Infra-Red Thermography (IRT)) is a method of using infrared radiation and thermal energy to gather information about objects. Thermal imaging is used in building diagnostics and maintenance [157], animal health check [158]. In the context of human-centred applications, it has been emerged in several human monitoring applications and recently came to light to measure high human body temperature (fever) in strategies to slow the spread of COVID-19 disease [159]. This is normally achieved by very expensive high-resolution thermal cameras. Lately, a new commercial low-resolution TSA sensor has gained growing interest in indoor human behaviour monitoring due to its low-cost and human privacy-preserving claims. However, there has not been sufficient independent empirical calibration of low-resolution TSA and high-resolution imagers for human behaviour monitoring applications.

This chapter provides empirical calibration of low- and high-resolution thermal imagers regarding their visible outputs, temperature accuracy and

stability. Besides, an assessment of the claimed privacy-preserving feature of
TSA has been conducted to experimentally validate the possibility of revoking
the human identity from the TSA's output. Thus, this chapter aims to
understand better the advantages, limitations, and research trends of using TSA
in domestic human monitoring schemes. The remaining parts of this chapter are
organised as follows: Section 4.2 provides a clear understanding of thermal
imaging for the thesis readers. An empirical calibration of thermal imaging is
provided in Section 4.3, and Section 4.4. An assessment of the
privacy-preserving feature of low-resolution imaging has been presented in
Section 4.5 followed by pertinent discussion and research trends drawn in
Section 4.6. A summary of this chapter is presented in Section 4.7

4.2 Understanding Thermal Imaging

The thermal imaging process relies on capturing the infrared radiations emitted
by the environmental objects to form a thermal image called thermograms. A
primary advantage of thermography over conventional imaging is its ability to
work with or without light since all objects with a temperature emit infrared
radiation. Recently, there has been a growing interest in utilising low-resolution
TSA in indoor human-centred applications [160, 161, 162]. The motivation behind
using TSA rather than high-resolution thermal imaging is due to several claimed
features, including privacy-preserving and low-cost capabilities.

Like any high-resolution thermographer, the spectrum of TSA radiation is
entirely determined by the temperature alone since no wavelength is selectively
emitted. Thus even a colourless object could still appear in TSA's thermograms.
On the other hand, the PIR sensor relies on a single IR sensing element to detect
warm objects as long as they have some degree of movement. The PIR sensors can
be considered a privacy-preserving approach in indoor environment applications.
Nevertheless, PIR can only detect temperature changes within the sensors' FoV
and, therefore, cannot be used reliably to detect, for example, different states
of humans. TSA overcomes the challenges of detecting stationary objects and
their orientation within the sensor's FoV by utilising multiple IR sensor elements,
referred to as IR array, that works together instead of a single sensing element.

The TSA's low-cost feature is evident compared to high-resolution thermographers. For instance, the price of a commercial TSA is about 0.125% of the price of the FLIR T6XX[1] camera, the high-end "gold standard" thermal camera. However, there has not been sufficient independent empirical calibration of the performance of TSA and high-resolution thermal imagers for human monitoring applications. Also, an assessment of TSA's privacy-preserving claim for human behaviour monitoring is missing from TSA-related works. This chapter aims to provide a fundamental understanding of the advantages, limitations and research trends of using TSA in human-centred applications by addressing the following specific objectives:

- to perform a visual thermal calibration of a high-resolution and high-cost imager with low-resolution and low-cost TSA;

- to perform temperature accuracy and stability calibration for various TSA sensors and high-resolution thermal imager;

- to validate the claimed privacy-preserving feature of TSA in cloud-based human monitoring applications.

The above study objectives, which have been visually illustrated in Figure 4.1, have been achieved using four thermal imagers, one TSA with the resolution of 16×12, two TSAs with the resolution of 32×24, and one high-resolution thermal imager with the resolution of 640×480. A detailed description of their experimental calibration is provided below.

4.3   Visual Thermal Calibration

Thermal imagers capture the thermal energy of objects in the FoVs and output as a temperature matrix. The temperature matrix visualisation is performed by applying a colour mapping scheme to create a visual image. Thus, thermal imagers could be considered as an image converter from the radiant thermal energy to the visible images. Therefore, there are specific attributes for

---

[1]More details about the camera is available from the FLIR website: https://www.flir.co.uk/support/products/t640/
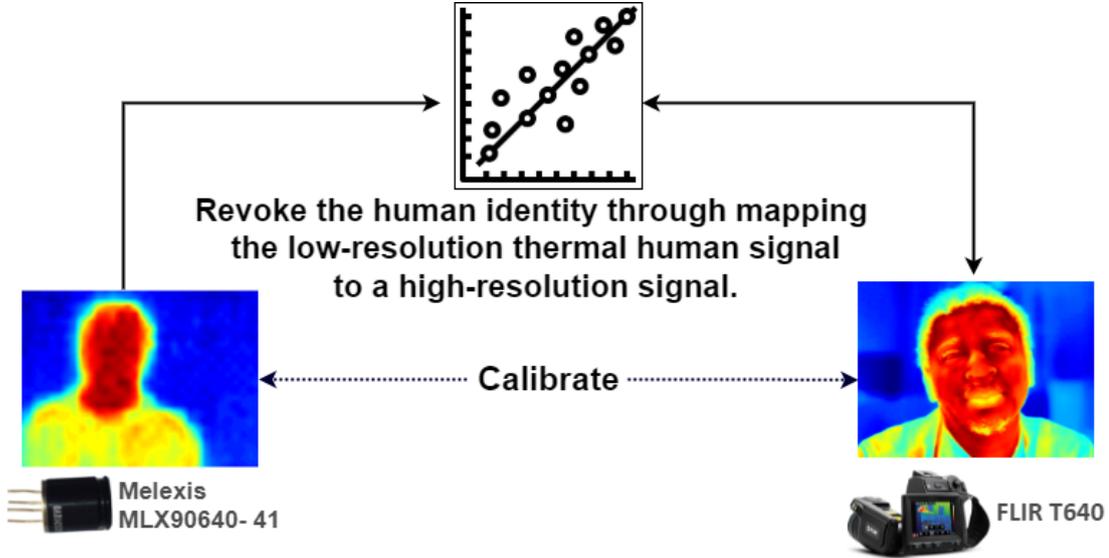
Figure 4.1: Graphical visualisation of these chapter objectives to calibrate between thermal imaging and assess the privacy of low-resolution TSA output.

determining the quality of the image: Accuracy, constant pattern noise and thermal sensitivity.

Figures 4.2(a), 4.2(b), and 4.2(c) show a visual calibration on the obtained temperature matrices post applying the same colour map scheme using TSA with 16×12 and 32×24 resolution versus high-resolution imager with 640×480 resolution, all captured at the same distance. Remarkable observations can be deduced from this visual calibration. First, the edges of the objects in the TSA output were not well preserved compared to the high-resolution imaging output. Second, the warmest region (maximum temperature values) that appears in all visual calibrations is the human face. However, the scale of temperature variations drops when the image resolution decreases. This is due to the fact that each temperature value (pixel) in the imaging output represents the average temperature of a wider inspection area for a lower resolution imager. This justifies the clear appearance of the human face in the high-resolution image and its high accuracy even in the detection of a small heat bump under the human lips in Figure 4.2(c).

The noise affecting thermography in indoor human-centred applications can be classified into two categories: (1) external noise, e.g. a cup of warm tea or ice
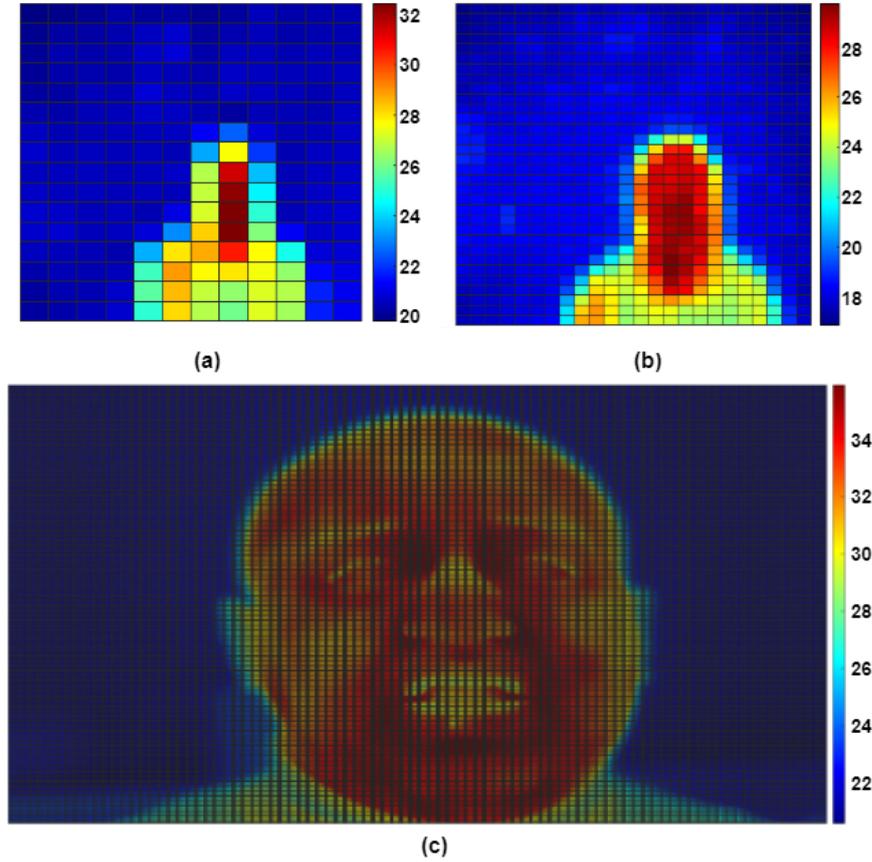
Figure 4.2: A visual calibration on colour mapped temperature matrices obtained using different thermal imaging resolutions, (a) TSA with 16×12 resolution, (b) TSA with 32×24 resolution, (c) high-resolution imager with 640×480.

cream, (2) thermal noise induced by human movement. Although external noises affect the colour map scale of both high and low thermal images, human-induced noise appears to be more serious in low-resolution thermal imaging. Figure 4.3 demonstrates two types of human-induced noise on the TSA output. The first is caused by a swift human movement that can be seen around the thermal human presence, while the second noise affects both low and high-resolution thermal images as it is caused by prolonged human contact with environmental objects such as a chair.
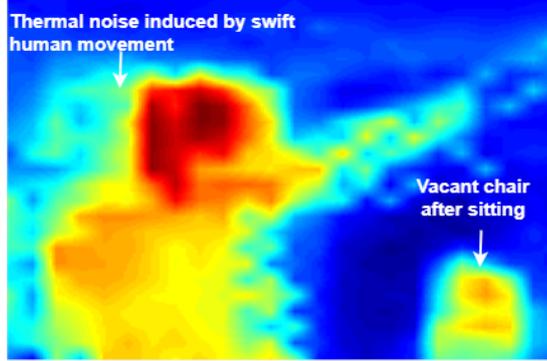
Figure 4.3: TSA is sensitive to thermal noise induced from a recent human movement and prolonged human contact with environmental objects such as a chair.

4.4   Temperature Value Calibration

The visual calibration demonstrated the potential of low and high-resolution thermal imaging in stationary and moving human detection. However, thermal imaging provides more useful information than conventional imaging, which is human temperature. This section investigates the reliability and accuracy of the human temperature acquired using different TSA resolutions. Figure 4.4 shows an experimental calibration of human skin temperature accuracy using different resolution thermal imagers placed at the same human-to-sensor distance. The first calibration shown in Figure 4.4(a) is concerned with assessing the stability of the same TSA resolution. In particular, two TSAs with the resolution of 32×24 have been used with a human moving in their FoVs. It can be observed from Figure 4.4(a) that the TSA is a well-stabilised sensor for acquiring the temperature value.

In the second calibration experiment, a high-resolution imager was also used. A human moved from a close human-to-sensor distance to a far distance in the sensors' FoVs and two different TSA resolutions. The result of this experiment is illustrated in Figure 4.4(b). It can be concluded from these results that there is a linear relationship between all of the thermal imagers regardless of their resolution. Specifically, the lower resolution TSA has a higher temperature value
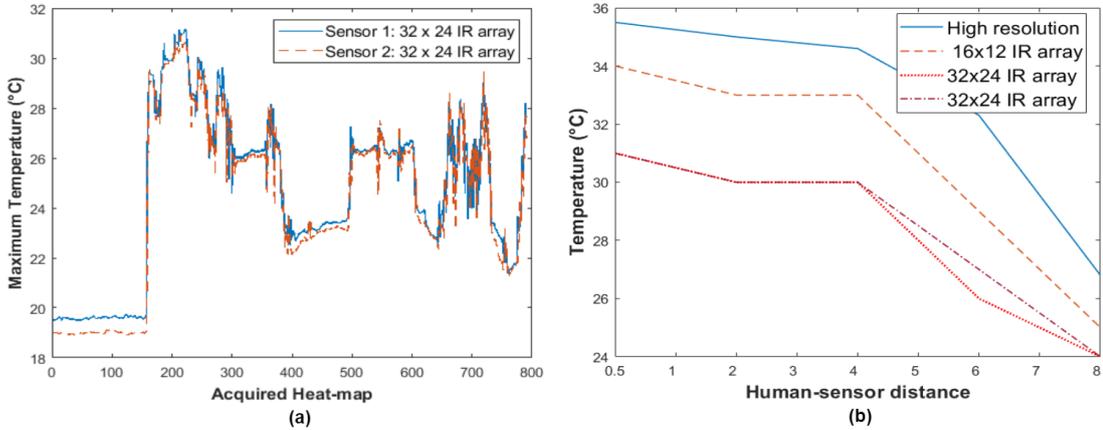
Figure 4.4: An empirical calibration of acquired human skin temperature using,
(a) same TSA resolution, (b) different TSA resolution with a high-resolution
thermal imager on a different human-to-sensor distance.

of $2°C$ in a linear relationship. Moreover, the lowest resolution TSA seems to be
more accurate than the higher resolution one with reference to the result of the
high-resolution thermal imager. Besides, the acquired temperature values vary
with the distance between the object and the sensor positions on all the used
thermal imagers.

4.5 Privacy Assessment of Low-Resolution Thermal Imaging

Unlike high-resolution thermal imagers, human information is not clear enough
to identify human identity in the TSA's output. Therefore, it has been claimed
that TSA is a privacy-preserving sensing approach [163]. An empirical privacy
assessment has been conducted to verify the possibility of reconstructing the
low-resolution thermal image to invoke identifiable human information from the
TSA's output.

The analysis performed to validate the privacy-preserving feature of TSA is
based on exploring if there is a relationship between low-resolution and
high-resolution imaging. Thus, low-resolution images can be converted using
this relationship to high-resolution images. Technically, to perform a regression

analysis to estimate the relationship between independent variables (low-resolution images) and dependent variables (high-resolution images). In this thesis, a two-layer feed-forward neural network is trained to solve this regression problem. The network input is the low-resolution data, while the output is the high-resolution data. The weight of the network is updated using Levenberg-Marquardt optimisation [164].

Given a set of m pairs (x_i, y_i) of low-resolution image and high resolution image, the primary goal of the optimisation algorithm is to find the parameters β of the network model $f(x, \beta)$ to minimise the sum of the squares of the deviations $S(\beta)$ as follow:

$$\hat{\boldsymbol{\beta}} \in \text{argmin}_{\boldsymbol{\beta}}\, S(\boldsymbol{\beta}) \equiv \text{argmin}_{\boldsymbol{\beta}} \sum_{i=1}^{m} \left[y_i - f\left(x_i, \boldsymbol{\beta}\right) \right]^2 \tag{4.1}$$

where $\hat{\beta}$ is the estimate of parameters β.

In the first experiment, a dataset of 916 low-resolution images and 916 high-resolution images were collected simultaneously for various indoor environmental thermal objects, including human subjects who move within the imagers' FoV while facing the imagers. Each subject presents in the imagers' FoV separately at the time of acquisition. To have the same size of input and output network layers, high-resolution images are resized to 32×24, the TSA output resolution. Further, the thermal images have also been converted from matrix to vector form. The data set was divided randomly into 70%, 15%, and 15% for training, validation, and testing. The R-value is used as a network evaluation matrix to report the extent to which the regression model can convert low-resolution images to high-resolution images and was 0.93691, 0.74869, 0.7128 and 0.85879 for training, validation, testing, and all of them, respectively.

The results above demonstrate the ability of the method to convert a low-resolution thermal signal into a high-resolution thermal signal. However, the result of the testing subset was 0.7128, which is not as good as the performance of the training subset. On the other hand, the aim of this experiment is to validate the claimed privacy-preserving feature of the TSA for human-centred applications. Therefore, a second data set was collected in the presence of a human in all the acquired scenes. The dataset contains 96 low-resolution images
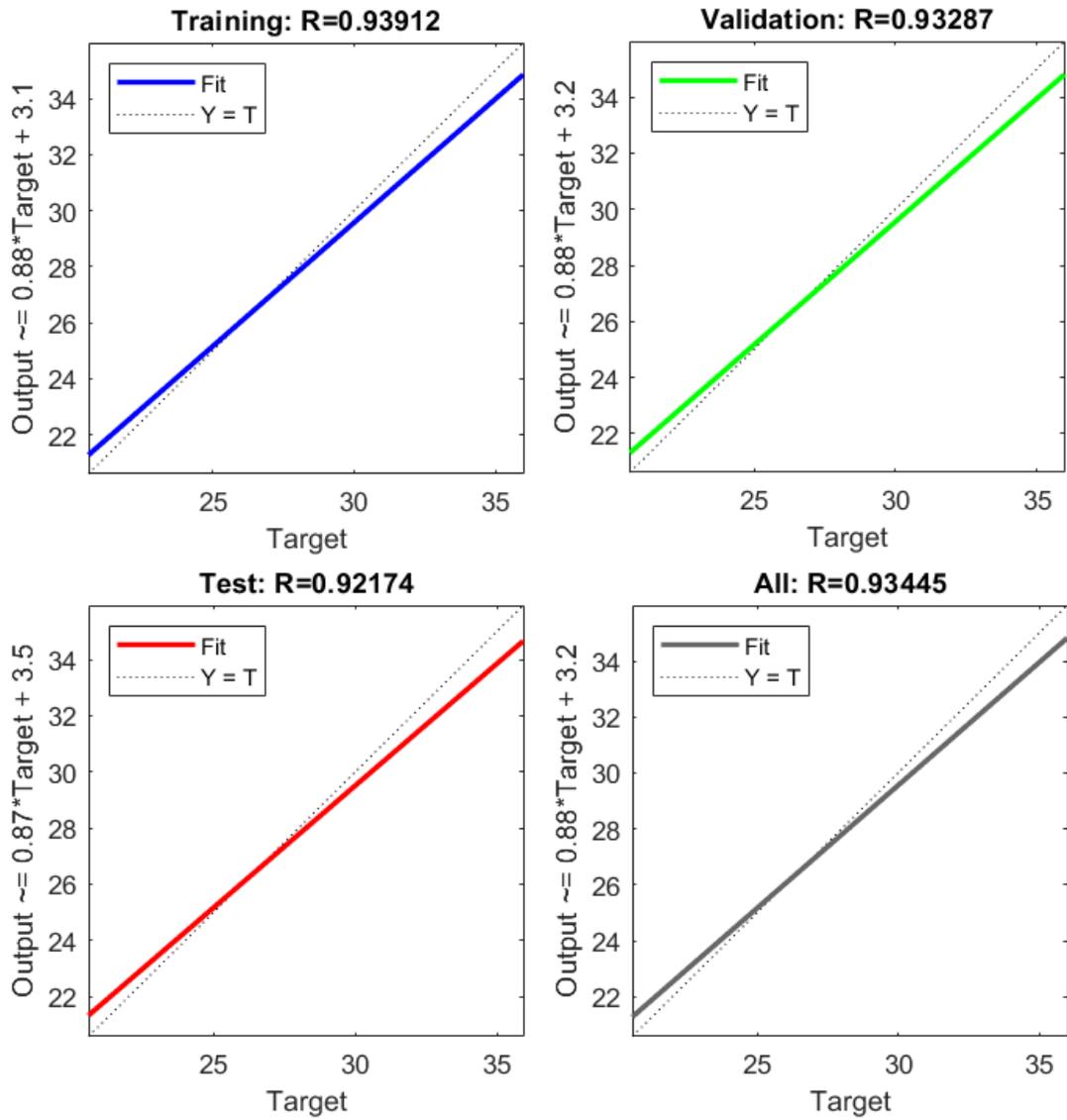
Figure 4.5: A visualisation of the regression model for the training, validation, and testing data sets shows the relationship between the low and high-resolution thermal images.

and 96 high-resolution images. The regression model achieves R-values of 0.93912, 0.93287, 0.92174 and 0.93287 for training, validation, testing, and all of them, respectively. Figure 4.5 shows the regression plots of the relationship between the acquired human presence in the low and high-resolution thermal images. Building on top of this, the identifiable human information may not be as private as claimed since it can be revoked after the low-resolution thermal signal is enhanced into a high-resolution signal.

4.6   Discussion and Research Trends

TSA shows a promising imaging approach for human-centred applications through overcoming challenges observed in other sensing approaches for human monitoring applications. For example, TSA does not require humans to wear or carry a device and thus could be more suitable for supporting older adults to live independently in their own homes. Also, the raw TSA output does not contain specific identifiable information compared to regular- or high-resolution thermal imagers.

The experimental calibrations of acquired human temperature show that human temperature values vary with imaging resolution in a linear relationship and human-to-sensor distance. Therefore, research work should take into account this variation in applications where the acquired human temperature is important for system decisions, e.g. human fever detection. Furthermore, the TSA output appears to be more sensitive to thermal noise, and thus it is very important to consider appropriate pre-processing techniques that are specifically suited to this sensing methodology.

The results of this chapter raise serious privacy concerns regarding TSA deployment in indoor human monitoring applications. Accordingly, TSA privacy should not be taken for granted since a third party could reconstruct the human thermal image from a low-resolution signal to a high-resolution signal. Thus identifiable human information could be revoked.

From an engineering point of view, TSA would be a better choice than high-resolution imagers due to the low cost and development integration for large-scale deployment of indoor human monitoring applications. Nevertheless, high-

resolution imagers provide richer thermal information than TSA and could be more useful in controlled-based applications such as human medical diagnostic systems or energy efficiency applications.

Based on the empirical calibrations presented in this chapter, the TSA with the resolution of 32×24 has been chosen to conclude the results of the next chapters.

4.7   Chapter Summary

This chapter has presented an experimental calibration of various thermal imaging resolutions to provide a clear understanding of the limitations and opportunities of low-resolution thermal imaging in the context of human behaviour monitoring. Furthermore, the privacy claim of low-resolution thermal imagers, TSA, has been examined practically.

The next chapter presents the proposed approach toward thermal human presence segmentation and occupancy estimation.

Chapter 5

Adaptive Sensor Placement for Human Segmentation and Occupancy Estimation

5.1 Introduction

Most research in the field of human behaviour monitoring in domestic environments, e.g., older adults' homes, are based on the assumption of a single inhabitant environment [156]. Homes, in reality, often contain more than one occupant. For instance, a reported study shows that the average number of individuals per household is more than 3.14 people per home [7]. Therefore, there is a need for a new functional layer to detect and determine the number of people in a given area, which is referred to as occupancy estimation. On the other hand, previous approaches to human-centred applications using the TSA usually relied on the use of a fixed sensor location to make the human-sensor distance and the human presence shape fixed. However, placing this sensor in different locations and new domestic environments can pose a significant challenge. In this chapter, a novel framework based on a deep convolutional encoder-decoder network is proposed to address this challenge in real-life deployment. The framework presents a semantic segmentation of the human presence and estimates the occupancy in the domestic environment. It can also
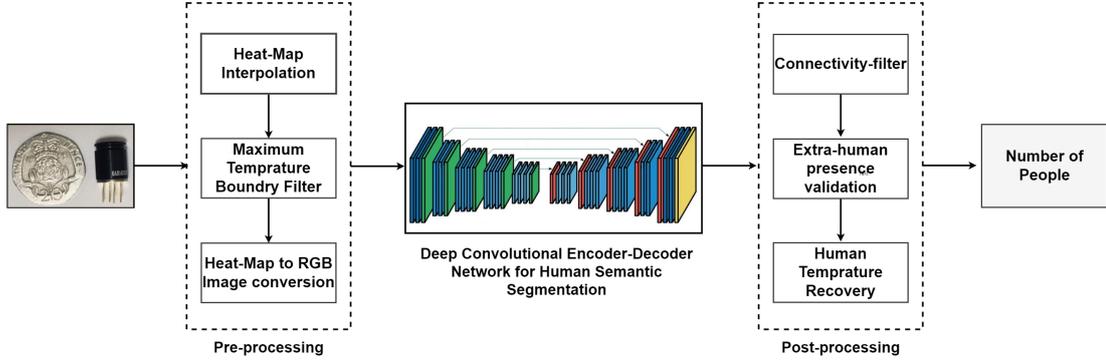
Figure 5.1: The proposed framework to estimate the number of people in the thermal scene obtained using the TSA after applying a set of pre-processing techniques, a deep convolutional encoder-decoder network to semantic segment the human presence, and post-processing techniques that consider the characteristics of the used sensor.

segment the human presence and count the number of people from different sensor locations, domestic environments, and human-to-sensor distance.

The remaining sections of this chapter are organised as follows: Section 5.2 explains the proposed framework architecture. The experimental results are presented and discussed in Section 5.3. A robust analysis of the proposed occupancy estimation phase of the human behaviour monitoring approach is provided in Section 5.4. Finally, Section 5.5 provides a summary of this chapter.

5.2 Human-Centred Occupancy Estimation

A schematic diagram of the proposed system designed to suit the characteristics of the TSA is shown in Figure 5.1. For example, the TSA is not light-sensitive compared to the camera sensor. However, the TSA is sensitive to environmental temperature and is of low resolution. Therefore, it is crucial to develop a systemic framework that depends on the type of the used sensor itself. Besides, the proposed framework segments the human presence from a noisy heat-map using a deep convolutional encoder-decoder network. A set of pre-processing and post-processing techniques are introduced to make the sensor output applicable to the proposed segmentation technique. A detailed description of
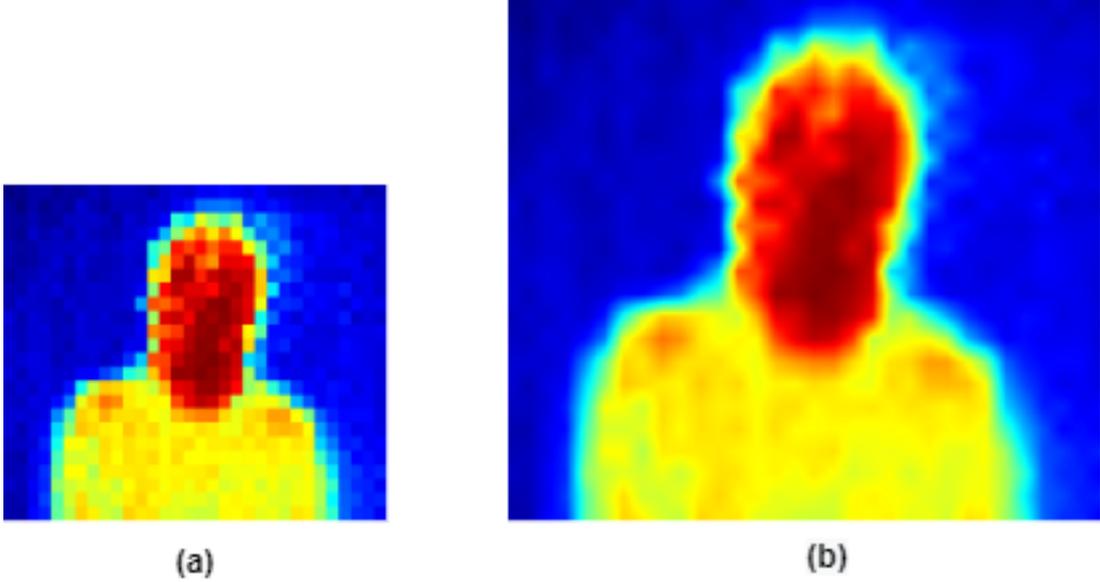
Figure 5.2: Heat-maps visualisation of (a) original heat-map, (b) interpolated heat-map.

the proposed framework stages is provided below.

5.2.1 Pre-possessing

The pre-processing stage consists of three sequential phases. The first phase is to increase the resolution of the TSA's original signals by interpolating the original 32×24 temperature matrix to 96×72 by repeating refined temperature values 3 times in each dimension. Figure 5.2(a) shows an example of the raw heat-map and Figure 5.2(b) shows the result of interpolating the heat-map by a factor of 3. This factor is chosen because it provides the best visual resolution of the TSA sensor-based heat-map and is proportional to the size of the input image of the used convolutional encoder-decoder network described in Section 5.2.2 to segment the human presence.

The detected human temperatures vary depending on the distance between the human and sensor locations. Also, the covered parts of the human body have a higher temperature than the uncovered parts. The normal maximum human temperature detected using the MLX90640 sensor from a nearby point is $33°C$.
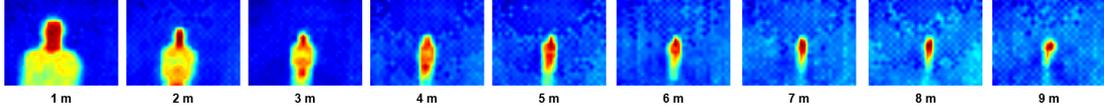
Figure 5.3: The Human presence using the thermal sensor array at different
distances after applying the proposed pre-processing techniques.

Thus, the second phase of the pre-processing stage is to filter any value with
a higher temperature than $33°C$ by converting it to the minimum temperature
value in the temperature matrix. By doing so, thermal noises such as a hot kettle
will be removed, and the human presence will remain as a foreground object in
the thermal scene.

It should be noted that Figure 5.2 shows a visualisation of the temperature
matrix using colour mapping to facilitate the reader's visibility of sensor data
visually. Moreover, the mentioned maximum filter converts any temperature
above the set limit to the minimum temperature in the acquired temperature
matrix to maintain the variance between the temperature values. If the filter
converts the high temperatures to zeros, this will cause the variance to be high,
which results in a different colour scheme. The third phase of the pre-processing
stage is to export the colour-mapped matrix to an RGB image.

5.2.2 Semantic Segmentation for the Human Heat-Map

Semantic segmentation is applied to separate a human subject from the RGB
image produced after the pre-processing step. Semantic segmentation aims to
classify each pixel in the image into a corresponding class. In contrast, object
detection classifies the regions of the image into a different class and draws a
bounding box around the object of interest. In order to make TSA adaptive to
different locations, object detection may not work well due to the high intra-class
variation of the human object in the thermal scene at different sensor locations
and human-sensor distances. Figure 5.3 shows the human presence in the thermal
scene after applying the pre-processing techniques at distances from 1 m to 9 m.
It can be observed from Figure 5.3 that the human presence changes its size and
topology with respect to the distance. Therefore, instead of detecting the human
object, this thesis proposes to use a deep convolutional encoder-decoder network

to classify each pixel in the thermal scene acquired by the TSA to either human or background classes.

The convolutional network architecture proposed in [165] is used here. The first path of this network, the encoder, is used to capture the context of the thermal image. The encoder consists of a typical stack of convolutional and max-pooling layers. The following part is the decoder part, which is the symmetric expanding part that enables precise localisation using transposed convolutions. In total, the architecture of this network has 23 convolutional layers. The reason for choosing this network architecture with the TSA is that it is designed for low-resolution images and does not require an extensive dataset as it performs excessive data-augmentation techniques.

The network is optimised using Adaptive Moment Estimation (Adam) [166] to compute the adaptive learning rates for each parameter using the gradient descent optimisation approach. This optimiser computes the first squared gradients m_t (the mean) and the second squared gradients v_t (the uncentered variance) as follow:

$$m_t = \beta_1 m_{t-1} + (1 - \beta_1) g_t \tag{5.1}$$

$$v_t = \beta_2 v_{t-1} + (1 - \beta_2) g_t^2 \tag{5.2}$$

m_t is the estimate of the first moment of the gradient, where v_t is the estimate of the second moment of the gradient. These estimates are biased toward zero, particularly during the initial time steps when the decay rates are low (i.e. β_1 and β_2 are close to 1). To compute the bias-corrected first and second-moment estimates:

$$\hat{m}_t = \frac{m_t}{1 - \beta_1^t} \tag{5.3}$$

$$\hat{v}_t = \frac{v_t}{1 - \beta_2^t} \tag{5.4}$$

Then, the network weight update as follow:

$$w_t = w_{t-1} - \eta \frac{\hat{m}_t}{\sqrt{\hat{v}_t} + \epsilon} \tag{5.5}$$

The initial default value for β_1 is 0.9, β_2 is 0.999, and 10^{-8} for ϵ.

The network is trained with a dataset containing 47 labelled thermal images acquired from one human object from a vertical position at distances from 0.5 m to 9 m. The output of this network is a matrix that shows the class (human or background) of each pixel, i.e. a binary mask that shows the human presence in the scene.

5.2.3 Post-possessing

The semantic segmentation technique proposed in the previous section has one drawback, which comes from the low-resolution thermal sensing methodology itself. Unlike the RGB camera, the TSA also senses the thermal noises left by humans even when they leave the sensor's field of view, which has a similar temperature to the human body. As a result, semantic segmentation may classify these noisy pixels belonging to a human. To overcome this drawback, a post-processing stage containing three phases is introduced.

The first phase is the connectivity filter to remove thermal noises that have a similar human temperature, such as noises generated by the human body or a warm object with a similar temperature to the humans, such as a warm cup of coffee. The connectivity filter is based on morphological operations. Specifically, the 8-connected algorithm [167] finds each connected component in the mask generated by the semantic segmentation network. The methodology behind this algorithm is to cluster each object based on the connectivity of its values. Each value in the mask mentioned above belongs to the same object if it has the same value (0 or 1) and is connected along the diagonal, horizontal, or vertical direction. Any connected component that is less than or equal to 30 pixels is considered a thermal noise and removed. This size is calculated based on finding the minimum human size acquired using the TSA from a maximum distance of 9 m. Next, the second image processing technique used is the Flood-Fill algorithm [168] to fill gaps in the human mask obtained from the last step. These holes may appear due to thick clothing, which reduces the temperature acquired for the human body
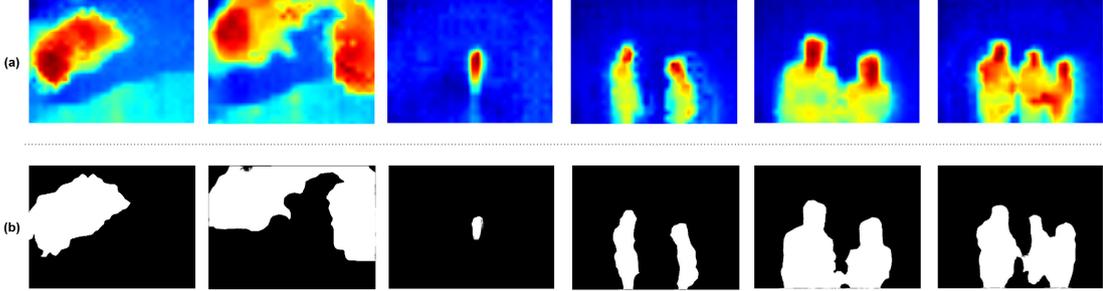
Figure 5.4: Illustrative results of the proposed framework, (a) the thermal images after applying the pre-processing techniques, (b) the human presence locations after using semantic segmentation, connectivity filter, and the extra human validation techniques.

by using the TSA. This decrease in the human temperature values can be seen as background pixels by the semantic segmentation network.

Since the colourmap is a colourful representation of the scene temperatures, humans at far long distances from each other lead to colour the distant human presence to background colours. Therefore, the second phase of the post-processing stage is to repeat the previous steps starting from obtaining a new RGB image of the interpolated heat-map without the locations of segmented human presence and adding the new segmented human, if found to the previously segmented mask. This extra-human presence validation phase repeats until the scene becomes an empty occupancy.

Figure 5.4 shows a few examples of applying the pre-processing, semantic segmentation, extra-human validation, and the connectivity filter to locate the human presence in the thermal images obtained by the TSA. Figure 5.4(a) shows the pre-processed heat-maps in different sensor location, human pose, human-human distance, and sensor-human distance, where Figure 5.4(b) shows the corresponding human presence in these thermal images after applying the semantic segmentation, extra-human validation, and the connectivity filter. As mentioned earlier, the encoder-decoder network for the semantic segment of the human heat-map was trained to detect the human presence with only one person from a vertical position. These illustrative examples are all unseen data for the network.

The human presence in the heat-maps converted to RGB images has been

determined at this stage. However, human temperature values are lost because they are converted to RGB values to focus on pixel intensity rather than the temperature value. Therefore, the third phase of the post-processing stage is to restore human temperatures through multiplying the masks obtained using the semantic segmentation, connectivity filter, and the extra human validation, which is shown in Figure 5.4(b), by the interpolated heat-maps to count the number of people as described in the following section.

5.2.4 Estimating the Occupancy Using Machine Learning Approach

The final stage of the proposed methodology is to count the number of people using the TSA. If all computed mask values from the semantic segmentation and the post-processing techniques are zero, it means no human presence in the thermal scene. Otherwise, two different machine learning approaches were evaluated to estimate the occupancy. In particular, a classification model using AdaBoost [58] and a regression model using shallow neural network [169] have been developed to count the number of people after segmenting the human presence from the thermal images. The primary difference between these two approaches is that classification deals with the problem of predicting a discrete class label, where the output of the regression is a continuous quantity.

5.2.4.1 Classification for Occupancy Estimation

In this thesis, an extension of the AdaBoost algorithm to a multi-class problem called AdaBoost.M2 described in Algorithm 1 is used as a holistic classification approach. The e-th training set for this algorithm includes the segmented human heat-maps x obtained after applying the pre-processing, semantic segmentation, and post-processing techniques, where y represents the class label (the number of people belongs to the set Y). The distribution $D_t(i, y)$ is maintained over the training set E and updated sequentially in each iteration c based on the output of that iteration. As mentioned earlier, misclassified training samples carry more weight than those correctly observed in the next iteration. By doing so, the update rule guarantees the upper bounds on training and generalisation error

---

Algorithm 1 The AdaBoost.M2 algorithm to classify each thermal scene into a class label, which represents the number of people in each scene.

Input: 1. Series of E of training samples $\{(x_1, y_1), \ldots, (x_e, y_e)\}$ with labels $y_e \in Y = \{1, \ldots, j\}$

       2. D represents the distribution over the E samples

       3. Weak learning algorithm **DecisionTree**

       4. Counter C for the number of iterations

1: **Initialize:** The weight vector: $w_{i,y}^1 = D(i)/(J-1)$, where $i = 1, \ldots, E, y \in Y - \{y_i\}$.

2: **for** $c = 1, 2, \ldots, C$ **do**

3:      $q_c(i, y) = \frac{w_{i,y}^c}{\sum_{y \neq y_i} w_{i,y}^c}$

4:      $D_c(i, y) = \frac{W_i^c}{\sum_{i=1}^E W_i^c}$ $(y \neq y_i)$

5:      Call `DecisionTree`.   ▷ Given the distribution D, and label weighting function q_c; return a hypothesis $G_c : X \times Y \to [0, 1]$

6:      $\epsilon_c = \frac{1}{2} \sum_{i=1}^E D_c(i, y) \left(1 - g_c(x_i, y_i) + \sum_{y \neq y_i} q_c(i, y) g_c(x_i, y) \right)$ ▷ Calculate the psudo-loss of g_c.

7:      $\beta_c = \epsilon_c/(1 - \epsilon_c)$

8:      $w_{i,y}^{c+1} = w_{i,y}^c \beta_c^{\frac{1}{2}(1 + g_c(x_i, y_i) - g_c)(x_i, y)}$ ▷ Update the new weights vector, for $i = 1, \ldots, E, y \in Y - \{y_i\}$

9: **end for**

Output: $g_f(x) = \arg\max_{y \in Y} \sum_{c=1}^C \log \frac{1}{\beta_c} g_c(x, y)$

---

rates.

5.2.4.2 Regression for Occupancy Estimation

The second step to estimate the occupancy is a regression through estimating the relationships between the segmented heat-maps and the number of people in the scene using an artificial neural network. In particular, a shallow neural network with only one hidden layer with sigmoid neurons and one output layer is used to determine the number of people in the scene. The input of the network is the segmented heat-maps, and the output is the number of people. The network is trained using the Levenberg-Marquardt backpropagation algorithm [170]. This
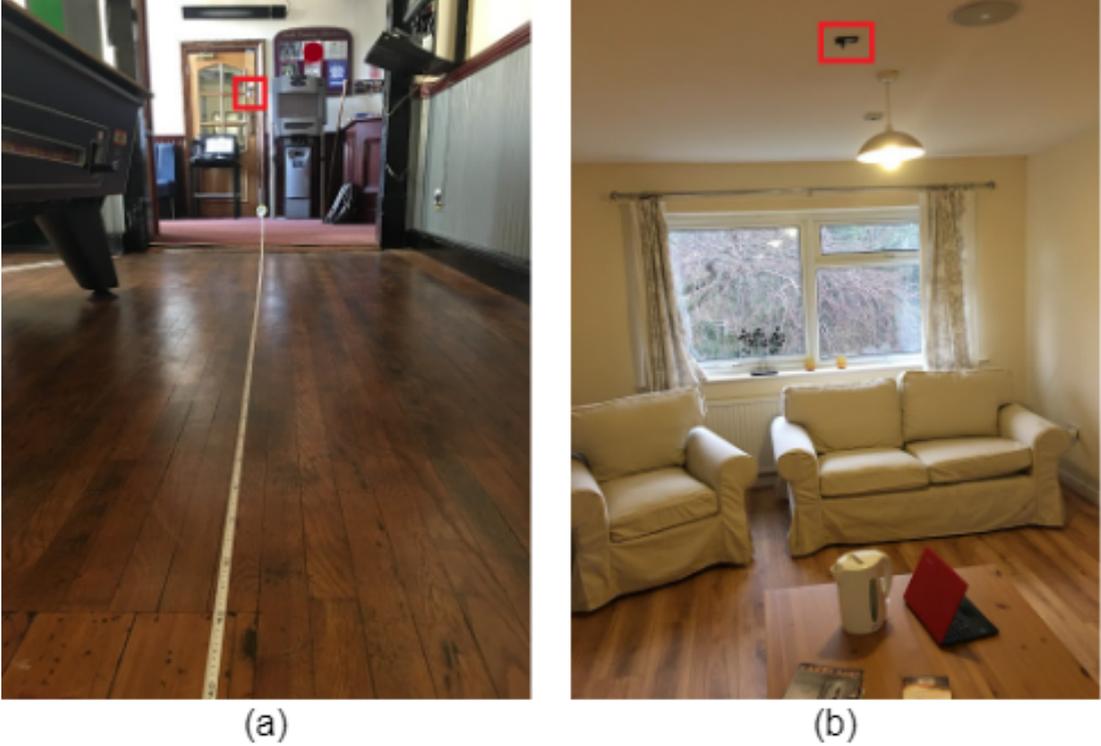
Figure 5.5: Data collection stages from two different domestic environments, (a) the sensor is placed on the wall, (b) the sensor is on the ceiling.

algorithm aims to minimise the sum of the squares of deviations $S(\beta)$ of a set of pair n $(x_i, \hat{y}_i)$ of input heat-maps x and the number of people $\hat{y}$ by finding the parameters β of the model output $f(x, \boldsymbol{\beta})$.

$$\hat{\boldsymbol{\beta}} \in \operatorname{argmin}_{\beta} S(\boldsymbol{\beta}) \equiv \operatorname{argmin}_{\boldsymbol{\beta}} \sum_{i=1}^{n} [\hat{y}_i - f(x_i, \boldsymbol{\beta})]^2 \tag{5.6}$$

The network training terminates when an increase in the mean square error of the validation dataset is detected. In this network, in contrast to classification, the result in the regression is a continuous value. Therefore, the output $\hat{y}$ is rounded to the nearest decimal point as our goal is to estimate the discrete number of people.

5.3 Experiments

To evaluate the performance of the proposed methodology, experiments were conducted with two different configurations of sensor locations. Correspondingly, two kinds of data were collected in different domestic environments and sensor locations.

In the first stage, the data was obtained while placing the sensor in a vertical position, as shown in Figure 5.5(a). Within this stage, two subsets of data were collected. The first subset consists of 47 thermal scenes in which only one person moves in the sensor field of view up to 9 meters long. The thermal objects in this subset are then labelled as either human or background objects. The second subset is used to analyse the effect of human distance on sensor performance. It is obtained in a human presence at distances of 0.5 m to 9 m away from the sensor. This subset is collected every 0.5 m, and its size is 325. The third subset was collected with one, two, and three different occupants moving in the sensor field of view to assess the performance of the occupancy estimation system. The size of this subset is 214.

The second stage aims to assess the adaptability of the proposed framework to work in a different domestic environment and sensor location. In this stage, the sensor is placed on the ceiling of the room, as shown in Figure 5.5(b). The dataset was collected in four different scenarios: one, two, three, and four occupants were moving in the scene. The size of this dataset is 203. In addition to the above dataset, 128 thermal scenes were collected from two empty-human environments and sensor locations with thermal noises such as a hot kettle, laptop, and heater when turned on to evaluate the proposed framework's ability to detect the empty occupancy environment. In total, 917 thermal scenes were collected to conclude the results of this chapter.

5.3.1 Human-to-Sensor Signal Analysis

To assess the impact of the distance on the human presence using TSA, a subset of the described dataset above is used. The dataset has different human heat-maps every 0.5 m and up to 9 m in length as described earlier, which makes a total of 18 distance steps. Further, an average human heat-map at every 0.5 m
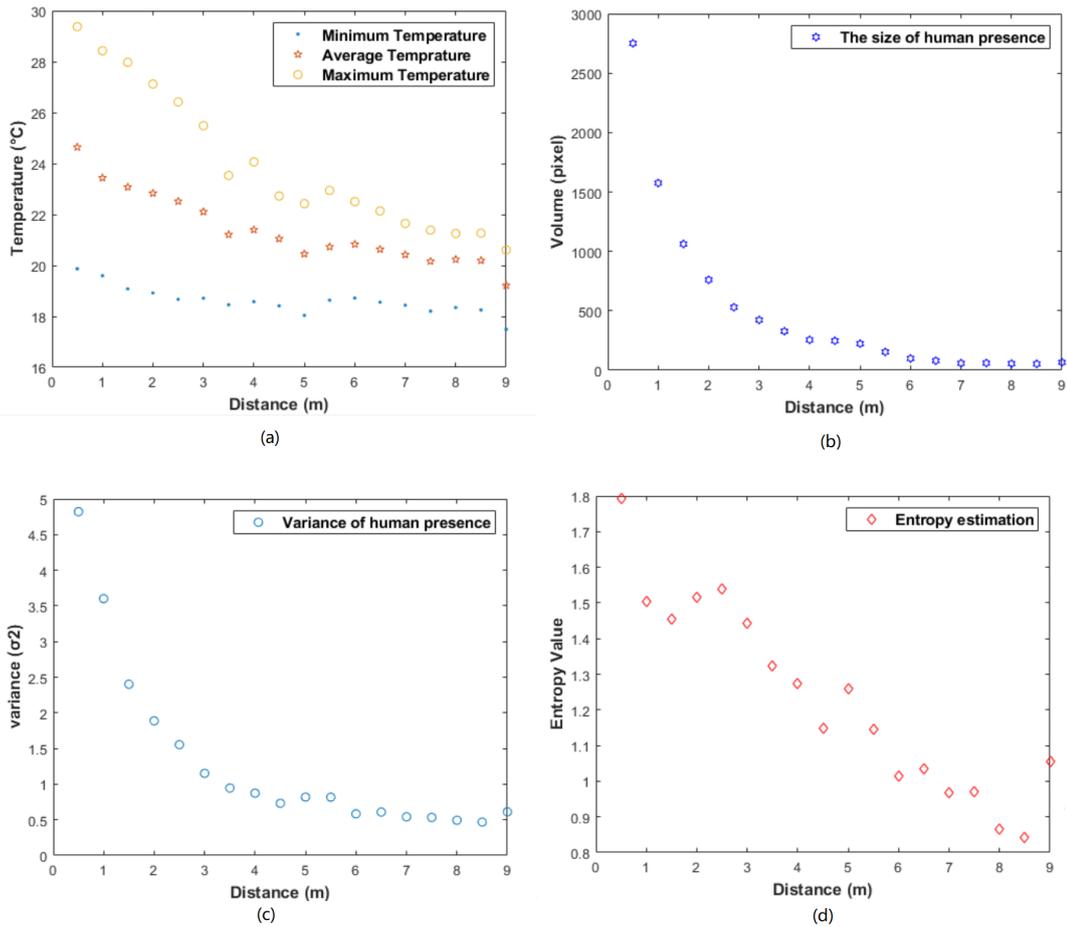
Figure 5.6: The effect of distance on human presence, (a) the minimum, maximum, and average temperatures, (b) the size of the human presence in the thermal scene, (c) the variance in human temperature, (d) the estimate of the entropy.

is computed. This average human heat-map for each distance step aims to avoid biased analysis of a random selection of a human heat-map.

Figure 5.6(a) shows the effect of the human-sensor distance on the value of the acquired temperature. Specifically, the minimum, maximum, and average temperatures of a human presence at different distances are shown. It can be seen the acquired human temperatures decrease when a person moves away from the sensor.

As shown in Figures 5.6(b), and 5.6(c) the sizes and the temperature variances of human presence vary depends on the relative location of the human subject to the sensor. One observation can be drawn from these figures; there is a significant decrease in the size of the human presence and temperature variance in the first 1.5 m. The decrease in the size of the human presence and the variance continues beyond 1.5 m, but with a smaller interval.

The common pattern of Figures 5.6(a), 5.6(b), and 5.6(c) is that they all have relatively stable values starting from a distance of 6 m meters and beyond. However, this stability does not exist in the entropy estimation of human presence at different distances, as shown in Figure 5.6(d). Besides, it can be identified that a linear relationship between the entropy metric for the human presence and the distance for every 1 m.

Based on these results, it can be concluded that the TSA and all the metrics calculated based on the thermal images are quite sensitive to distance. Furthermore, the calculated statistical metrics, in particular, the entropy point estimate, can be used to determine the human distance from the sensor using a suitable function approximator (e.g. our shallow ANN).

5.3.2   Occupancy Estimation Experimental Results

The first experiment was to use the collected empty occupancy dataset and evaluate the performance of the proposed framework to detect the empty human environment before proceeding to the classification or regression models as described in Section 5.2.4. The proposed framework was able to detect the empty occupancy with 100% accuracy. This performance validates the proposed pre-processing, semantic segmentation, and post-processing techniques in

segmenting only the human presence from the thermal scene.

The second experiment was to examine the performance of the classification approach using AdaBoost. The used dataset was obtained from the vertical position. In this experiment, the dataset is divided into 70% for model training and 30% for testing. The performance of this classification model in occupancy estimation is 98.43% achieved accuracy.

The third experiment is to validate the performance of the classification model with a different machine learning approach. In this experiment, the regression approach based on a shallow neural network is used. The same data used to train and test the classification model is also used to train and test the shallow network. This dataset is divided into 70% for training, validating and testing the shallow neural network during the network training stage, and 30% for testing the performance of the trained shallow neural network to count the number of people in the thermal scene. The performance of the regression model in occupancy estimation is 93.75%.

The results of the above experiments show that the classification model has better accuracy than the regression model. The assumed reason for the lower accuracy in the regression was due to the uncertainty of some of the regression outputs. Hence, as mentioned above, the output of the regression is a continuous numerical value. Since this milestone of the thesis is concerned with counting a discrete number of people, the output of the regression model is rounded to the nearest decimal number. Uncertainty occurs when the output has one-half (e.g. 1.5, 2.5, etc.).

5.4 Robust Analysis

This analysis has two main aims. The first aim is to validate the adaptability feature of the proposed framework for operating in a different domestic environment and sensor location. Therefore, the sensor is installed on the ceiling of the room in a different home. Hence, the trained deep convolutional encoder-decoder network from a vertical sensor location is used to segment the human presence from the overhead thermal scenes. The second aim of this analysis is to validate the possibility of parametrising the proposed framework

Table 5.1: A comparison of the experimental setup, sensor placement, occupancy estimation method and results of the proposed system with the state-of-the-art.

Paper/Method	Ref.	Sensor	Placement	Sensor Location	Estimation Method	Accuracy
Beltran et al.	[91]	TSA, PIR	Non-adaptive	Ceiling	K-NN	NA
Gomez et al.	[93]	TSA	Non-adaptive	Wall	CNN	53.7%
Tyndall et al.	[70]	TSA, PIR	Non-adaptive	Ceiling	K* algorithm	82.56%
Metwaly et al.	[96]	TSA	Non-adaptive	Ceiling	ANN	98.90%
Proposed		**TSA**	**Adaptive**	**Wall** **Wall** **Ceiling** **Ceiling**	**AdaBoost.M2** **ANN** **AdaBoost.M2** **ANN**	**98.43%** **93.75%** **100%** **58.33%**

to predict more people in the scene. Therefore, the dataset collected for this analysis contains more people compared to the previous experiments described in Section 5.3.2.

The configuration of this analysis dataset is divided in the same way as the vertical sensor dataset used in the previous experiment is divided with the same classification and regression approaches. Regarding the classification approach using the AdaBoost algorithm, the system achieves an accuracy of 100% in estimating the occupancy from 1 to 4 different occupants moving in the thermal scene. However, the system achieves 58.33% accuracy using a shallow neural network to estimate the occupancy.

The AdaBoost algorithm's classification approach shows high performance in estimating the occupancy in different sensor locations and environments using the proposed pre-processing, human segmentation, and post-processing techniques. However, the regression approach shows a lower performance in estimating the occupancy from the overhead thermal scenes. This was due to the high uncertainty of the output of the regression model in the overhead thermal scenes compared to vertical-based scenes. On the other hand, the regression approach is a promising approach to estimate the occupancy in an unsupervised learning problem in which the number of people is greater than the number used to train the model.

Furthermore, the proposed approach of this thesis has been compared with the state-of-the-art approaches. Table 5.1 shows a comparison of the proposed approach and other notable works in terms of the experimental setup, adaptability of the sensor placement, occupancy estimation technique, and

experimental results.

5.5   Chapter Summary

This chapter has presented a novel approach to segmenting the human presence from different sensor placements and estimates the occupancy of the environment as the first functional phase towards enabling human behaviour monitoring in a multi-occupancy domestic environment. The chapter has also presented robust pre-processing techniques that have considered the TSA characteristics and constraints. An empirical analysis of the human-to-sensor signal has also been presented.

Finally, this chapter provides a robust analysis of the proposed technique and assesses its generalisation ability. The next chapter presents the second sequential functional phase to localise human subjects and extract their physiological knowledge.

Chapter 6

Human Localisation and Physiological Knowledge Extraction

6.1 Introduction

Extracting human-to-sensor and human-to-human distances based on human physiological appearance is essential to developing a human behaviour monitoring system that supports domestic multi-occupancy environments. Human distance estimation is also vital in other applications, such as making places safe by preventing the transmission of contagious diseases through social distancing alert systems. This chapter proposes a novel approach to estimating human distance for indoor human-centred applications using a low-resolution TSA. The proposed system presents discrete and continuous human-to-sensor distance estimators using classification techniques and ANN, respectively. It also proposes a real-time distance-based FoV classification through a novel image-based feature. Inspired by the results from this research stage, a novel human-to-human distance estimator has also been explored in this chapter. Besides, this chapter proposes a transfer application to the proposed continuous distance estimator to measure human height.

The remaining parts of this chapter are organised as follows: Sections 6.2, 6.3,
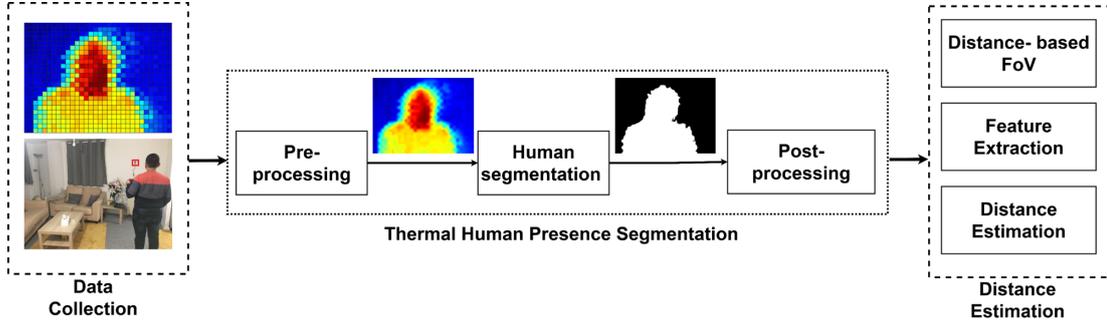
Figure 6.1: The proposed framework for estimating the distance between the human presence and the thermal sensor array placement after applying a set of techniques, which semantic segment the human presence, followed by a technique to classify the FoV into distance-based regions, and finally output the predicted human distance in the FoV.

and 6.4 explain the proposed framework architecture. Experimental results are presented and discussed in Sections 6.5, and 6.6 followed by pertinent chapter summary drawn in Section 6.7.

6.2 Enabling Human Distance Estimation

It is crucial to segment the human presence and count the number of people from the TSA outputs to extract human-related distances based on their physiological appearance. Following the proposed approach in Chapter 5 to segment and count the human subjects in the TSA's outputs, a schematic diagram of the proposed human distance estimation approach is shown in Figure 6.1. To recap, the proposed approach considers the TSA characteristics, which are different from regular cameras. In contrast with regular cameras, which are sensitive to light, TSA is not sensitive to light. Instead, the TSA is more sensitive to environmental radiation than the camera, resulting in much noise in the TSA images. For example, the edges of the human body in thermal images obtained from TSA are not sharp. Moreover, the moving body in thermal scenes changes the occupied area's temperature and surroundings. Therefore, although both the camera-based and TSA-based sensing generate images, their processing techniques are different.
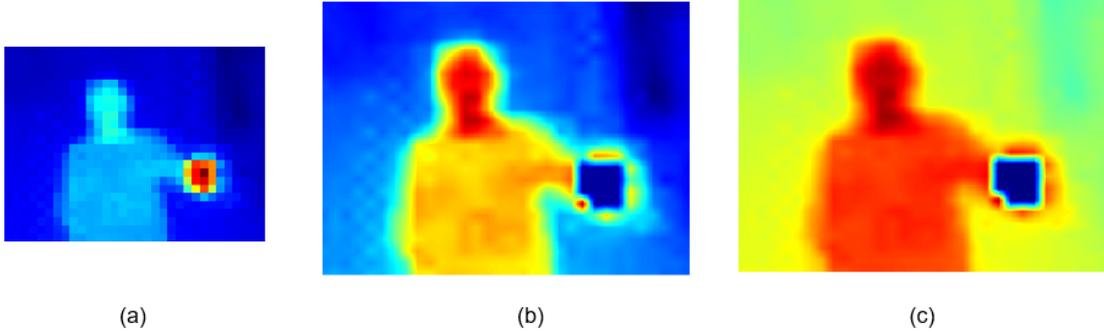
Figure 6.2: Illustrative results of the pre-processing techniques, (a) the original heat-map of a human holding a cup of coffee, (b) the heat-map after filtering and interpolating the original heat-map, (c) the effect of the faulty filter on the interpolated heat-map.

To enhance the resolution of TSA-based thermal images, an interpolation by the 3 factor of the original thermal images is applied. By doing so, the resolution of the obtained turns into 96×72 instead of its original size of 32×24. Concerning the distance estimation problem versus the TSA characteristics, the minimum captured human temperature varies depending on the sensor's distance and the human location. Conversely, the maximum human temperature can be determined from the closest point, which is $33°C$ using the MLX90640 sensor.

Based on this, any abnormally high temperatures, such as a hot kettle, can be filtered. On the other hand, it is important to maintain the variance between the minimum and maximum temperatures. So this proposed filter converts the detected high-temperature values to the minimum temperature in the thermal image itself rather than converting the abnormal human high-temperature values to zero. To give an impression, Figure 6.2 illustrates the results of applying the pre-processing techniques to TSA's output. Figure 6.2(a) shows the original heat map acquired while one person is holding a cup of coffee in the sensor's FoV. Figure 6.2(b) shows the result of applying interpolation and the maximum temperature filter. Figure 6.2(c) shows a negative example of a wrong, abnormal human temperature filter that converts high-temperature values to zero instead of minimum temperature value in the thermal scene. Although filtering the high-temperature values in the acquired heat-map to zero
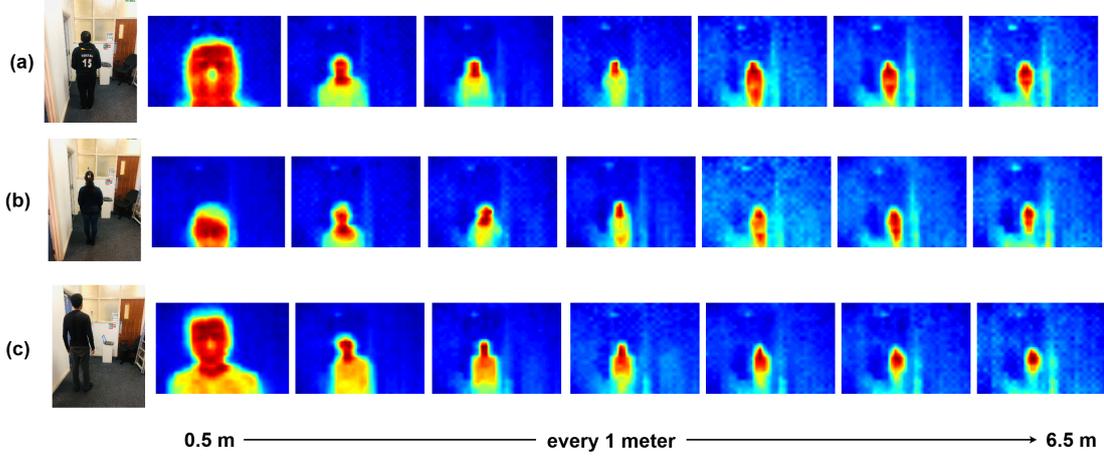
Figure 6.3: Distance aspect of thermal human presence at distances from $0.5\,m$ to $6.5\,m$ in a distance step of $1\,m$, (a) male participant, (b) short female participant, (c) a relatively tall male participant.

preserves the human presence in the foreground of the thermal image, it also increases the thermal noise in the background as well as a loss of visual, thermal information (e.g. the heat distribution within the human presence area). As a result, after the pre-processing, the resultant TSA output is a one-channel temperature matrix. These figures are generated by applying a colour mapping scheme to visualise the TSA output better. Thus the last step of the pre-processing is exporting the colour-mapped matrix into an RGB image to enable the proposed encoder-decoder convolutional neural network to segment human presence and estimate the occupancy as described earlier in Chapter 5.

6.3   Region Based Field of View

Based on geometry, it is possible to determine the distance, D, between the sensor and an object if the object's dimension, O, is known and the sensor's FoV covers the whole object. That is:

$$D = \frac{O}{2 \times \tan\left(\frac{FOV}{2}\right)} \tag{6.1}$$

However, this geometry does not apply to human-centred sensing applications by
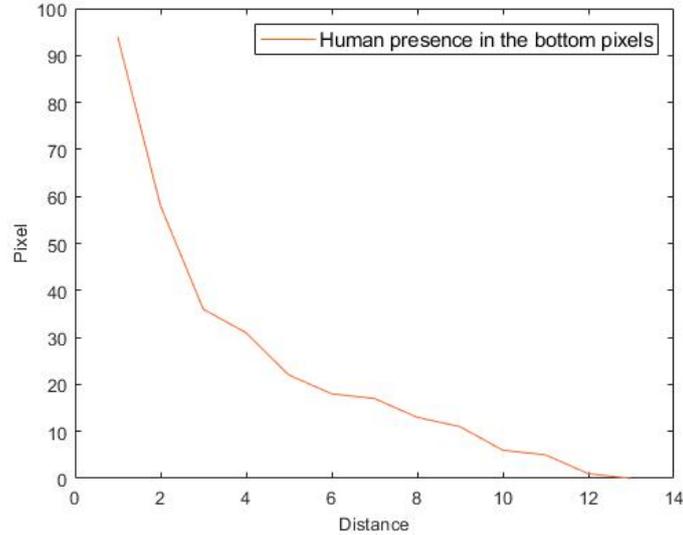
Figure 6.4: The number of occupied human presence pixels at the bottom of the image versus human to sensor distance.

TSA as humans vary in body shape in the output images. Figure 6.3 shows a visualisation of the TSA output used for three participants at distances ranging from $0.5\,m$ to $6.5\,m$ with a step of $1\,m$. From these three illustrative examples, it can be observed at the distance of $0.5\,m$ that the participant in Figure 6.3(a) had his head fully visible while this was not the case for the female participant in Figure 6.3(b). Continuously for a relatively tall participant, e.g. in Figure 6.3(c), the head and parts of the upper body are sensed from the same sensor placement. On the other hand, the human body begins to fully emerge in the TSA output at a distance of $3.5\,m$ and beyond. This means the distance for the first few meters is unpredictable using the above geometry, and to predict the distance after $3.5\,m$, the human dimension is required.

The human distance in the TSA field of view should be carefully estimated. To achieve this, a novel image-based feature to solve this problem is proposed. This feature is based on the observation that human presence diminishes in the bottom rows of the thermal image as the human goes further from the sensor location. Figure 6.4 shows an example of the bottom image rows of a human moving from a close point to a point far away from the location of the sensor. It

can be seen that the number of pixels belonging to the human presence located at the bottom rows of the thermal image decreases as the distance between the sensor and the human increases. Based on this, the sensor's FoV can be classified into distance-based regions, e.g. near, middle, and far regions depending on the human presence's location, using the number of occupied human pixels in the bottom rows of the thermal image. Hence, this feature's simplicity would allow real-time applications to quickly obtain the human location and reduce the processing time to compute the exact human distance estimate, as described in the next section.

The human presence mask, which is a binary mask that corresponds to the class (human or background) of each pixel in the obtained thermal image generated by the proposed encoder-decoder convolutional neural network, is used to determine the number of the occupied human pixels in the bottom rows of the thermal scene by counting the last non-zero values in the mask. This feature is then used to train a classification model to predict the region of the human location in the FoV as described in Section 6.5.

6.4 Human Distance Measurement

In this section, the exact estimate of human distance will be computed after finding the region of human presence in the sensor's FoV, as described in the previous section. Reducing the number of actual distance classes by categorising the FoV into regions reduces the processing time and increases the proposed estimation system's performance. Thus, this section provides a detailed description of the extracted features used to train and test the proposed estimation models.

6.4.1 Human Physiological Feature Extraction

A number of TSA-based human physiological features have been extracted to predict the exact human location and measure the human-to-sensor and human-to-human distances. Figure 6.5 shows an evaluative example of the effect of human-to-sensor distance on human temperature values captured by
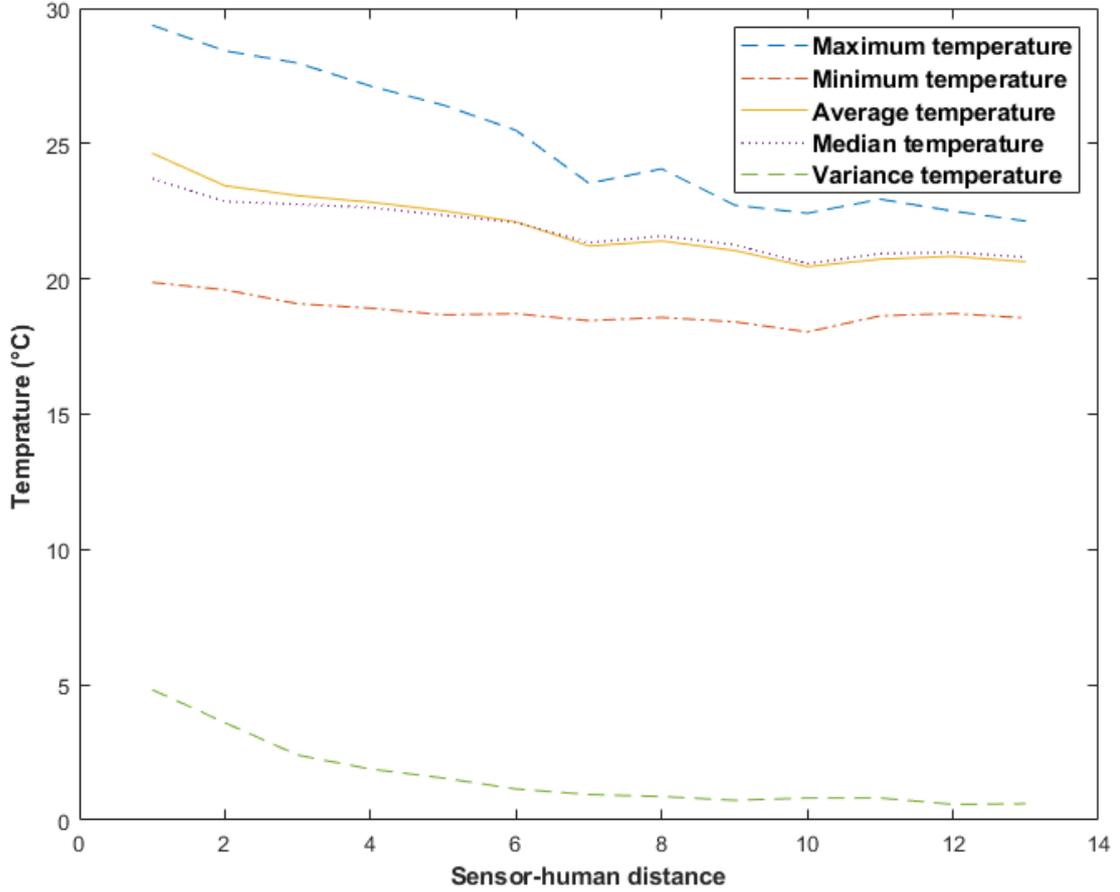
Figure 6.5: The effect of the distance on the acquired human temperature using the TSA.

TSA. Specifically, the minimum, maximum, average, mean, median and variance temperature of human presence from $0.5\,m$ to $6.5\,m$ with a distance step of $0.5\,m$. It can be seen that the overall trend of human temperature decreases with the increase of the human-to-sensor distance. To further evaluate the image, the entropy is extracted for each segmented human heat map histogram using the following equation:

$$H(X) = -\sum_{i=1}^{n} P\left(x_i\right) \log P\left(x_i\right) \quad where \quad n = histogram\ bins \qquad (6.2)$$

In addition to temperature-based features, human presence size was also

73

considered to feed the human distance estimation model. Hence, it has been previously shown that there is an inverse relationship between distance and the size of human existence.

6.4.2 Human-to-sensor Distance Estimation

The first proposed human distance estimation technique is a regression to map the extracted features x and the human-to-sensor distance using ANN. In particular, Multilayer Perceptron (MLP) artificial neural network with one input layer, one hidden layer with sigmoid neurons and one output layer is used. The weight updating Δw_{jk} can be written as:

$$\Delta w_{\mathrm{jk}}(\mathrm{p}) = \eta \times \mathrm{y_j}(\mathrm{p}) \times \delta_\mathrm{k}(\mathrm{p}) \tag{6.3}$$

where p refers to the number of iterations used to propagate the error signal from the output layer to the hidden layer. The gradient error $\delta_\mathrm{k}(\mathrm{p})$ in the output layer is determined from the derived activation function multiplied by the error in the output layer neuron. Hence, η refers to the learning rate. In this chapter, the network is trained using the Levenberg-Marquart backpropagation algorithm [170]. This algorithm tries to minimise the sum of the squares of deviations $S(\beta)$ of a set of pair n $(x_i, \hat{y}_i)$ of input heat-maps x and the sensor-human distance $\hat{y}$ by finding the parameters β of the model output $f(x, \boldsymbol{\beta})$.

$$\hat{\boldsymbol{\beta}} \in \mathrm{argmin}_\beta \, S(\boldsymbol{\beta}) \equiv \mathrm{argmin}_{\boldsymbol{\beta}} \sum_{i=1}^{n} \left[\hat{y}_i - f\left(x_i, \boldsymbol{\beta}\right) \right]^2 \tag{6.4}$$

The detection of the validation dataset's mean square error leads to the termination of the training process. In a real-life scenario, there is an infinite number of distance classes as one human could be at any distance in the sensor's FoV. Thus the aim of utilising this ANN architecture is to find a continuous-based sensor-human distance estimate. However, a discrete-based human distance estimation using the classification approach is also performed to evaluate the extracted TSA-based features' performance by having a specified number of classes for every $0.5\,m$ up to $6.5\,m$, making a total of 13 classes.
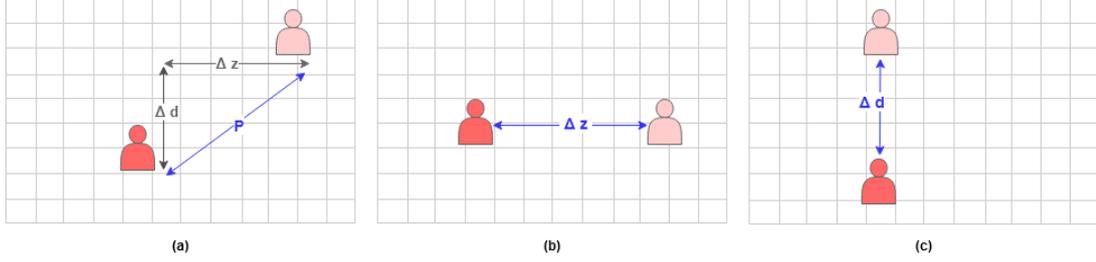
Figure 6.6: Illustrative use case scenarios for measuring the physical distance between two human subjects, (a) Δd is less than $2m$, (b) Δd is $0m$, (d) when Δd is greater or equal to $2m$.

6.4.3 Human-to-human Distance Estimation

The size of the inspection area captured in the FoV varies at different depth distances, which means that the TSA covers a wider area at greater distances than those at close distances. To measure the physical distance, the inspection area's size should be known at the specified depth of human presence. Therefore, the second step after calculating the human-to-sensor distance is to find the size of the inspection area on the depth of the human as follows:

$$a = 2 \times d \times \tan\left(\frac{FoV}{2}\right) \tag{6.5}$$

where d is the human-to-sensor distance and FoV for the used TSA is $55° \times 35°$. Assuming a human subject is located $2m$ far from the sensor position, the inspection area a at their location would be $2.08m \times 1.26m$. Since the measurement unit for images is in pixels, the horizontal a is converted into a pixel unit through dividing a by the horizontal image resolution, which is 32. In the example where a human is located at $2m$ far from the sensor, each pixel in the horizontal width of the image is equal to $0.065m$. This process is repeated to localise the human presence in the FoV for each human subject.

Figure 6.6 shows illustrative use case scenarios to measure the physical distance between two human subjects. In Figure 6.6(a) the difference between the sensor-human distances Δd to both subjects is less than $2m$. Also, Δz, which is calculated from the horizontal inspection area's size a and the human presence mask for each human subject, is not zero. Based on this, the physical

Figure 6.7: Data collection stages from three different indoor environments, (a) the sensor is placed on the wall to assess the performance of the proposed sensor-human distance methodology, (b) the sensor is also placed on the wall, (c) the sensor is on the ceiling to assess the generalisation of the proposed methodology.

distance p can be formulated as the hypotenuse in the Pythagorean theorem, and thus its value can be calculated as follows:

$$p = \sqrt{(\Delta z^2 + \Delta d^2)} \tag{6.6}$$

The second use case scenario is when Δd is zero as shown in Figure 6.6(b). In this case, the physical distance is Δz itself. On the other hand, when Δz is zero, the physical distance is Δd, as shown in Figure 6.6(c).

6.5 Experiments

To evaluate the performance of the proposed framework of human distance estimation, experiments were performed using two different configurations of the sensor's placements. They were also evaluated from three different indoor environments in the summer and winter seasons of the UK. The reason for considering different seasons and different indoor environments is that the indoor heating systems in the UK usually operate during the winter season. In the summer months, neither heating nor cooling is used. These evaluations

ensure a high generalisation ability for the proposed estimation system as the TSA sensor is sensitive to ambient temperature.

In the first data collection configuration, the sensor was placed in a vertical position with a height of $1.57\,m$ from the ground as shown in Figures 6.7(a) and 6.7(b). A total number of 703 thermal images were collected for 6 different human participants at distances from $0.5\,m$ to $6.5\,m$ every $0.5\,m$. During this data collection stage, participants were asked to stand on these 13 different distance classes to avoid the over-fitting problem during the algorithms' learning and testing phases.

The second data collection configuration aims to assess the proposed distance estimation system's adaptability versus sensor placement and human data bias. At this stage, the low-resolution thermal scenes of two new participants (male and female) were acquired from an overhead sensor placement, as shown in Figure 6.7(c). The size of this dataset is 90. In total, 793 thermal scenes were collected to conclude this chapter's results.

6.5.1 Region Based FoV Experimental Results

The first experiment examined the proposed image-based feature to categorise the sensor's FOV into three regions based on the human-to-sensor distance. The first defined region is from $0\,m$ to $2.5\,m$, the second region ranges from $3\,m$ to $4.5\,m$, and the last one is between $5\,m$ and $6.5\,m$. The used dataset was partitioned into 5 folds to protect against over-fitting, and the best overall achieved accuracy was 76.8% using decision trees. Further, focused experiments with the same data partition configuration were conducted on each user's data; Figure 6.8 illustrates the proposed image-based feature's performance on six different human participants. The confusion matrices shown in Figures 6.8(b), 6.8(e) are for female participants while Figures 6.8(a), 6.8(c), 6.8(d), and 6.8(f) for male participants.

It can be noted from these figures that the proposed feature works relatively better for male participants than females, with an overall accuracy of 91%, while for female participants, the accuracy was 74%. This observation does not necessarily imply that the heat signature differs based on human gender, but
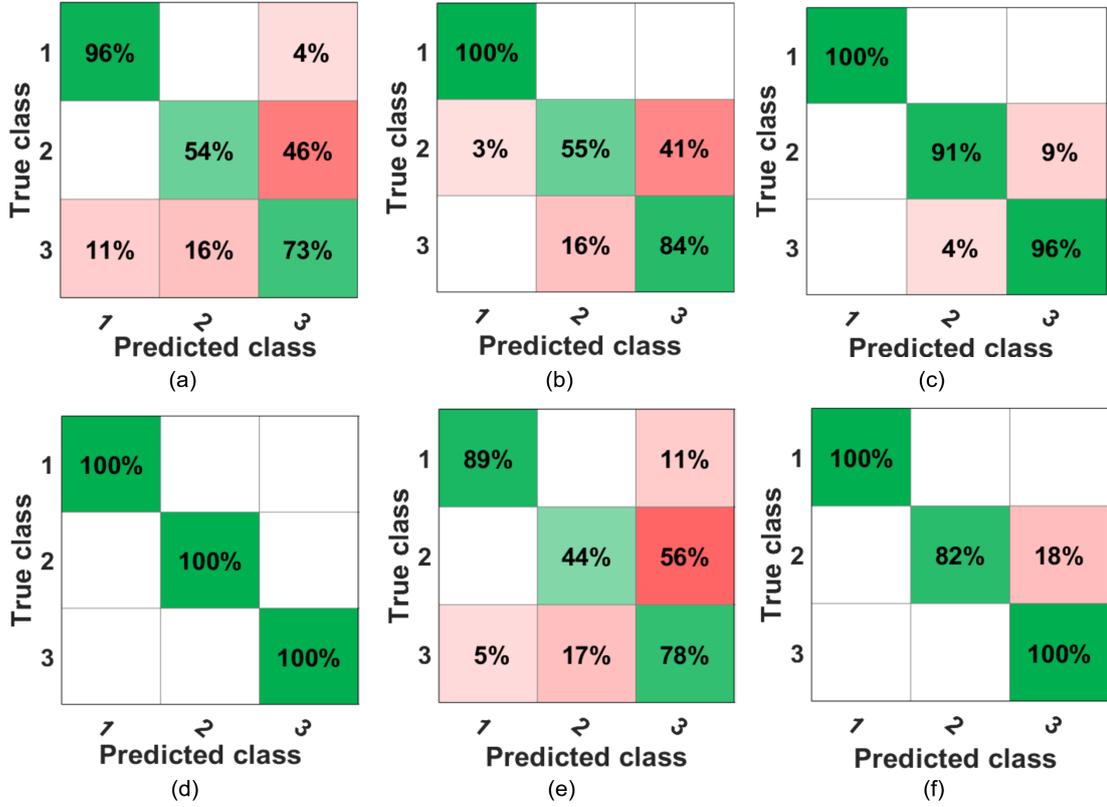
Figure 6.8: A visualisation of the participant-focused performance of the proposed image-based feature to classify the sensor's FoV into distance-based regions, where (a), (c), (d) and (f) are confusion matrices for different male participants while (b) and (e) are for female participants.

perhaps females tend to wear heavier clothing compared to males, and this reduces the temperature perceived by TSA. On the other hand, females are generally smaller in size than males, which means that their heat signature size will be smaller than that of males.

6.5.2 Human Distance Estimation Experimental Results

The first experiment is a continuous estimation of human distance using ANN from a vertical sensor placement described in Section 6.4.2. In this experiment, the collected dataset was divided into two subsets. The first subset is the thermal data obtained at decimal distances $(0.5, 1.5, 2.5, 3.5, 4.5, \text{ and } 6.5\,m)$. This subset is used to train the proposed neural network to predict the sensor-human distance

using the extracted feature vectors described in Section 6.4.1 as the network input and the corresponding distances as the output. This network is then tested with completely unseen data to predict the sensor-human distance. The data is from the second subset at integer distances (1, 2, 3, 4, 5, and $6\,m$). The median overall error in predicting the distances was $\pm 0.2\,m$. Hence, since the trained network's output is a continuous distance value (not a labelled class), this approach is called a continuous-based human distance estimation.

The same dataset is then used with 13 defined class labels (0.5, 1, 1.5,..., $6.5\,m$) for all data participants obtained from the vertical sensor position. In this experiment, various classification algorithms were used to evaluate the performance of the proposed features. The dataset is divided into the training and testing stages using cross-validation with 10 folds. Table 6.1 shows the performance of these classification algorithms. The best-achieved accuracy was 96.8% using Cubic SVM.

6.6 Robustness Analysis

The robustness analysis contains two main experiments. The first experiment evaluates the adaptability and performance of the proposed image-based feature of a distance-based FoV with a different number of regions. In this experiment, two regions were identified instead of the three suggested in Section 6.5.1. The first defined region ranges from $0\,m$ to $3\,m$, and the second region is from $3.5\,m$ to $6\,m$. Reducing the number of defined FoV regions increases the overall accuracy

Table 6.1: A comparison of different classification algorithms to classify the human-to-sensor distance with 10 cross-validation folds.

Classification Algorithm	Accuracy (%)
Naive Bayes	63.3%
Tree	83.4%
Ensemble - Bagged Trees	90.8%
Kernel Naive Bayes	91.6%
KNN	96.5%
Cubic SVM	**96.8%**
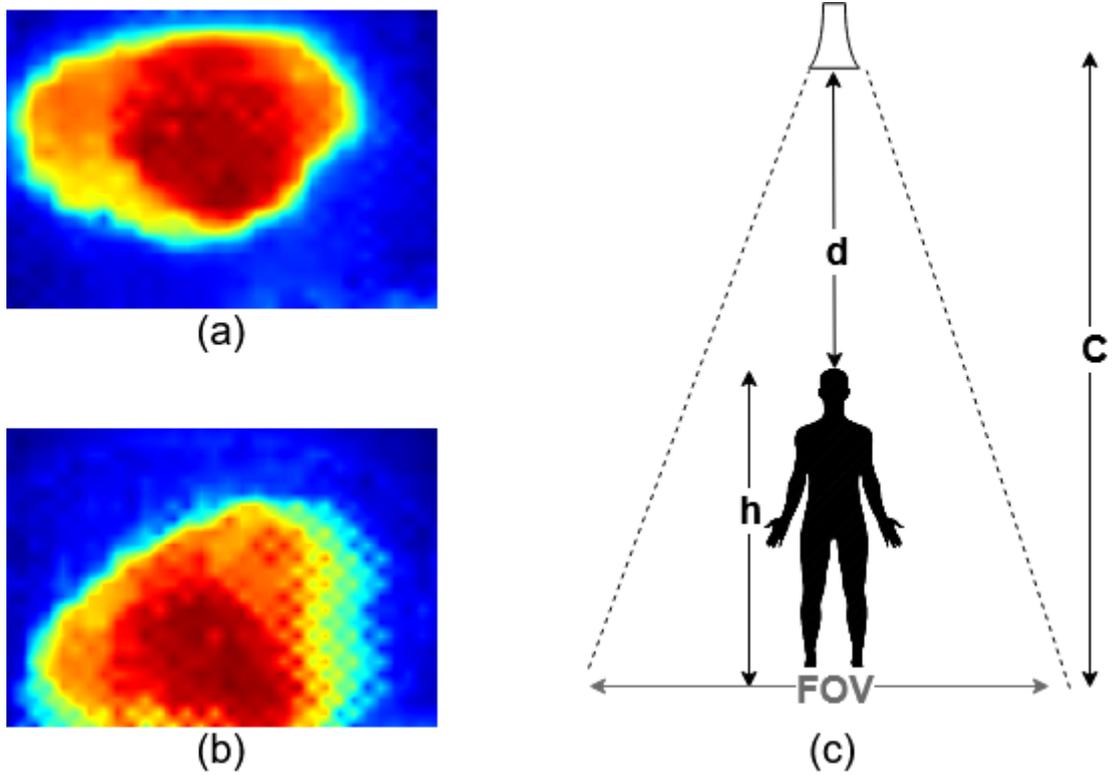
Figure 6.9: An overhead sensor placement, (a) the overhead image of a fixed-moving human presence, (b) the impact of movement on the thermal human presence, (c) a transfer application of the proposed distance estimator to predict the human height.

from 76.8% to 95.4%. This increase in performance underlines the robustness of the proposed real-time human localisation feature in terms of the FoV region occupied. Besides, it shows low inter-class variation within the second region between $3\,m$ and $4.5\,m$ with the other two defined regions in the previously defined three regions. Thus the performance was lower prior to the merging of the second region.

In the second experiment, the proposed ANN's generalisation ability to map between the extracted TSA features and the human-to-sensor distance was assessed. This was achieved through testing the ANN, which is already trained using data obtained from the vertical sensor placement, with completely unseen data obtained from the overhead sensor position and new human participants. The median error in predicting the male participant's distance was $\pm 0.07\,m$ and

$\pm 0.66\,m$ for a female participant. Hence, during the data collection phase, the female participant was wearing a headscarf, which reduced her acquired head temperature. Further experiments were performed on the collected data to analyse the impact of the thermal image quality on the performance of the continuous human distance estimator. The fixed human presence had a predictive error of $\pm 0.01\,m$ while moving humans decreases the robustness of the extracted features, resulting in a lower rate of prediction. Figure 6.9(a) shows a stable human presence from a sensor placed on the ceiling of the room, and Figure 6.9(b) shows the effects of human movements on the acquired thermal human presence of the same human participant. Importantly, the proposed approach for the estimation of human distance can be transferred to extract human physiological features such as human height. Given a user case scenario of overhead sensor placement, as shown in Figure 6.9(c). It is then possible to estimate human height h if room ceiling height c is known using the following simple geometry:

$$h = c - d, \quad \text{where } d \text{ is the predicted human-to-sensor distance} \tag{6.7}$$

The robust analyses concluded that the proposed human distance estimation using TSA has a high generalisation ability toward operating with different experimental configurations. Besides, the proposed transfer application to measure human height demonstrates the important impact of the proposed distance estimators on other human-centred applications.

6.7 Chapter Summary

Following the previous chapter on adaptive TSA placement to segment and estimate the occupancy. This chapter has presented a novel approach toward enabling TSA to extract physiological-based human localisation knowledge. In particular, human-to-sensor, followed by human-to-human distance estimation, has been presented. The chapter also provided a robust analysis to assess the generalisation ability of the proposed human distance estimators.

Finally, a transfer application of the proposed human-to-sensor distance estimator is explored to extract human physiological features, for example, human height, from the TSA output. The next chapter presents the third functional phase to fuse multiple TSA sensors toward enabling domestic multi-occupancy human behaviour monitoring.

Chapter 7

Thermal Motion Signal Processing for Sensor Fusion

7.1 Introduction

The main limitation of deploying TSA-based Internet of Things (IoT) systems on a large scale is the challenge of fusing multiple TSAs to cover a wide inspection area, e.g. smart homes, hospitals and many other domestic environments. On the other hand, objects that appear in the low-resolution thermal images acquired from TSA have low intra-class variations and high inter-class similarities, making the identification of overlapping regions through matching a comparable template image in multiple images very difficult. Following the work presented in Chapters 5 and 6 on occupancy and human distance estimation to identify the multi-occupancy environment, human-to-human distance, and human-to-sensor distance using single-based, frame-by-frame, processing. This chapter intends to address the challenge of enabling TSA to cover a wide inspection area by proposing a motion-based approach to fuse multiple TSAs and learn the domestic environment layout to enable further human behaviour monitoring applications to run in the cloud. Besides, a privacy improvement on utilising these sensors in IoT-based applications to meet the TSA privacy concerns raised in Chapter 5 is proposed. The proposed fusion approach is evaluated with comprehensive experiments on
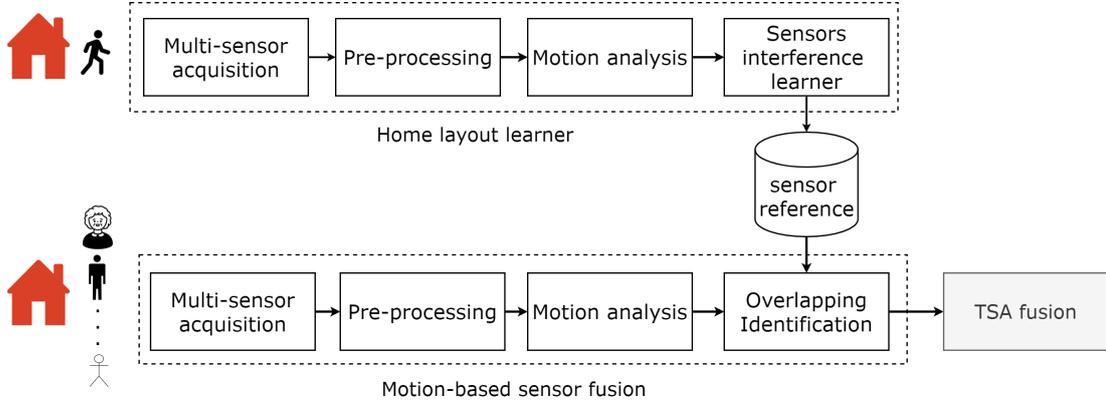
Figure 7.1: A schematic diagram of the proposed approach for home layout learning and multiple TSAs fusion through the analysis of apparent motion patterns of moving subjects in the acquired sensors' signals.

different sensor placements and domestic environment conditions.

The remaining parts of this chapter are organised as follows: Section 7.2 explains the proposed framework architecture. Comprehensive experiments and analysis are presented and discussed in Sections 7.3, and 7.4 followed by the chapter summary in Section 7.5.

7.2 TSA Fusion for Human Behaviour Monitoring

The proposed approach, which is illustrated in Figure 7.1 consists of two main stages. The first stage is called home layout learner, which enables the flexibility of the proposed approach to work with different and complicated environments by estimating the environment layout and the sensors' placement to find the flow of human movement in their domestic environment. This is achieved by finding a sensor reference, which indicates the typical human movement flow in multiple sensors' inspection areas and the overlap between the installed sensors' FoV by analysing the movement patterns of a single human subject moving in the environment. Hence, the home layout learner stage will only run once (the first time after installation) to identify which sensors are interfering with each other in their FoVs. By doing so, the proposed approach is not only flexible to work with
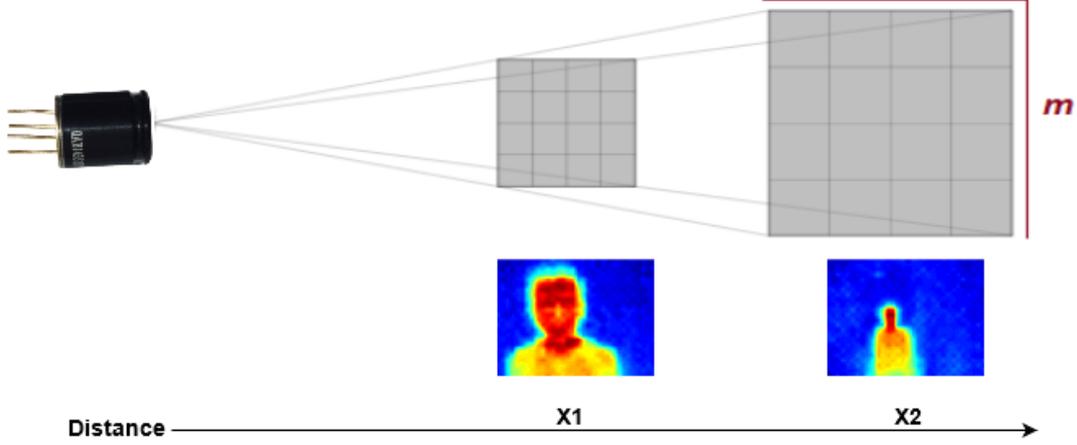
Figure 7.2: A visualisation of the relationship between the sensor FoV's depth and the inspection area.

different home layouts but also can reduce the computational time and resources needed to cover a large domestic environment. Besides, the environment layout learner can also determine the typical human movement flow in a simple domestic environment, including rooms and large open halls with more than one sensor to cover the inspection area. After the layout is learnt in the first stage, the second stage is used for sensor fusion and identification of overlapping regions between two or more TSA sensors.

A single TSA would not be sufficient to cover a wide inspection area. Figure 7.2 shows an illustrative diagram of the relationship between the sensor's FoV and the inspection area. It can be observed from this figure that the inspection area is larger from a distance. Considering the FoV of the TSA sensor used in our investigation, MLX90640, which is $55° \times 35°$. The size of the inspection area at a different distance can be calculated as follows:

$$m = 2 \times d \times \tan\left(\frac{FoV}{2}\right) \tag{7.1}$$

where d is the depth distance in the FoV and m is the inspection area. Assuming d is $2\,m$, then inspection area m would be $2.08\,m \times 1.26\,m$. The collected data, which represents the thermal signature in the FoV, at time t with size C is transformed from $1D$ linear vector $\mathbf{x}_t = [x_{t_1}, \ldots, x_{t_C}]^{\mathrm{T}}$ to $2D$ grid format of

85
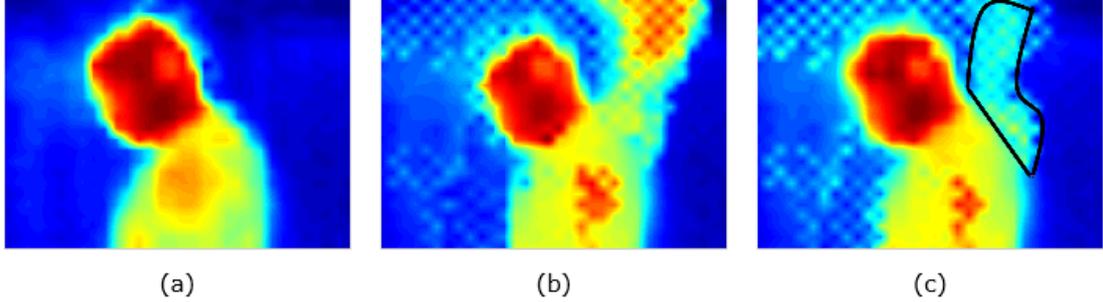
Figure 7.3: The effects of human movement on the acquired thermal scene, (a) a stationary human presence, (b) human hand movement, (c) thermal noise induced by human movement.

size 32×24 resolution. A colourmap scheme is then applied to each of the $2D$ temperature metrics for visualisation purposes. A detailed description of the remaining functional phases (pre-processing, motion analysis, sensors interference learner, and identifying the overlapped regions) of the two stages is described below.

7.2.1 TSA Signal Pre-processing for Human Motion Learning

There are notable limitations to proceeding with the human motion-based feature extraction on the output of this sensor. In particular, human movement generates a thermal noise in the background scene, which can bias human motion estimation by including unwanted background pixels. To illustrate this on a thermal image of a stationary human subject shown in Figure 7.3(a). Given the movement of the human hand shown in Figure 7.3(b), the acquired thermal image after movement is affected by thermal noise induced by human movement, as highlighted in Figure 7.3(c). This is a critical consideration in multiple TSA use case scenarios from different sensor positions, for example, the ceiling, where the human would be in direct contact with a background surface such as the floor, would result in prolonged thermal noise lingering in multiple acquired thermal images. To overcome this, an adaptive temperature level-based thresholding technique to separate the human presence from the thermal background scene is proposed on Ostu's method [171]. Mathematically,

the threshold k of separating each $2D$ temperature matrix x_t into foreground t_f and background t_b classes should minimise the intra-class variance $\sigma_w^2(k)$, defined as a weighted sum of variances of the two classes:

$$\sigma_w^2(k) = \omega_{t_f}(k)\sigma_{t_f}^2(k) + \omega_{t_b}(k)\sigma_{t_b}^2(k) \tag{7.2}$$

where ω_{t_f} and ω_{t_b} are the probabilities of the foreground and background classes, respectively separated by k, while σ_{t_f} and σ_{t_b} are the intra-class variances of these two classes. The class probability $\omega_{t_f}, \omega_{t_b}(k)$ is found from the L bins of thermal image histogram:

$$\omega_{t_f}(k) = \sum_{i=0}^{k-1} p(i) \tag{7.3}$$

$$\omega_{t_b}(k) = \sum_{i=k}^{L-1} p(i) \tag{7.4}$$

In a binary classification problem, minimising the intra-class variance (variation between multiple samples of a class) is equivalent to maximising the inter-class variance (variation between classes):

$$\begin{aligned}
\sigma_b^2(k) &= \sigma^2 - \sigma_w^2(k) = \omega_{t_f}\left(\mu_{t_f} - \mu_K\right)^2 + \omega_{t_b}\left(\mu_{t_b} - \mu_K\right)^2 \\
&= \omega_0(k)\omega_1(k)\left[\mu_0(t) - \mu_1(k)\right]^2
\end{aligned} \tag{7.5}$$

which is exposed in terms of class probabilities ω and class means μ, where $\mu_{t_f}(k)$, $\mu_{t_b}(k)$ and μ_K represents:

$$\mu_{t_f}(k) = \frac{\sum_{i=0}^{k-1} ip(i)}{\omega_{t_f}(k)} \tag{7.6}$$

$$\mu_{t_b}(k) = \frac{\sum_{i=k}^{L-1} ip(i)}{\omega_{t_b}(k)} \tag{7.7}$$

$$\mu_K = \sum_{i=0}^{L-1} ip(i) \tag{7.8}$$

By computing ω and μ iteratively, and k for each thermal scene, the algorithm
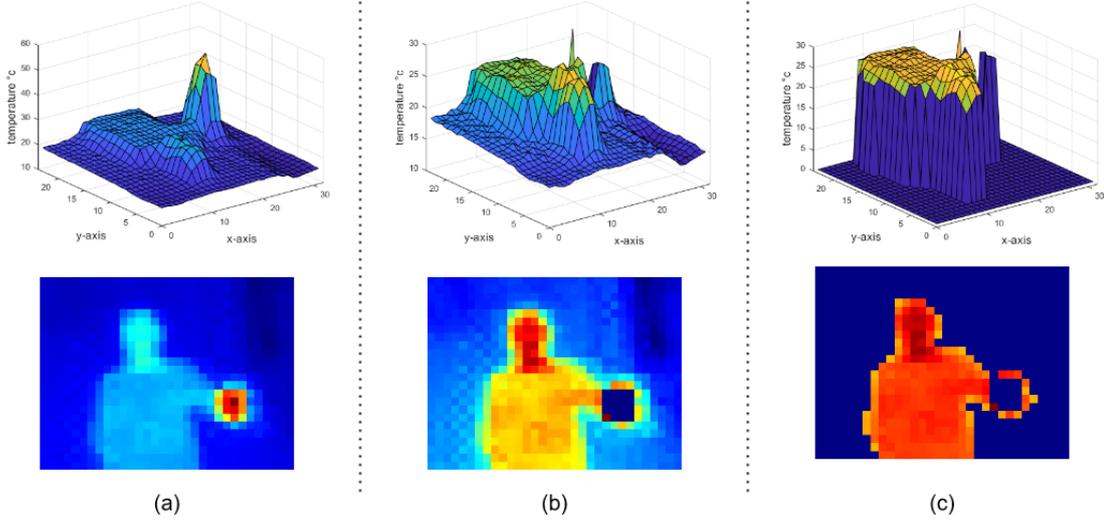
Figure 7.4: Illustrative results of the proposed pre-processing phase, (a) the original temperature surface plot and its corresponding heat-map, (b) the result of applying the temperature filter, (c) the result of separating the acquired temperatures into background and foreground categories.

would separate the scene into background and foreground classes regardless of the human-to-sensor distance. Hence, the larger human-to-sensor distances, the lower human-acquired temperatures. However, the drawback of this method is when the histogram is not bi-modal distribution and has, for example, three peaks like a temperature much higher than that of a human temperature, such as a hot cup of coffee, as shown in the representative surface diagram of Figure 7.4(a) and its corresponding thermal image. To avoid this, a modification before applying this approach is suggested by using a temperature filter to convert temperatures above normal human temperatures to the minimum temperature of the thermal scene (not zero, to maintain the variance of human and background temperatures). This empirical value is set to be $33°C$ by the designated TSA. The filtered value is, therefore, as follows:

$$x_t = \begin{cases} x_i & \text{for} \quad x_i <= 33 \\ min(x_t) & \text{otherwise} \end{cases} \quad where \quad x_i = FIR \qquad (7.9)$$

Figure 7.4(b) shows an illustrative result of applying this modification filter on the original acquired thermal scene shown in Figure 7.4(a), while Figure 7.4(c)
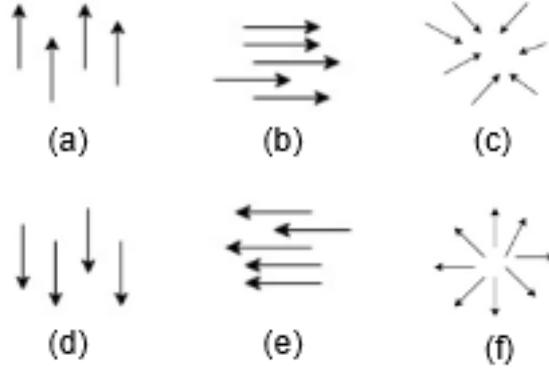
Figure 7.5: An illustration of the motion vector based on optical flow, (a) represents vertical movement upwards, (b) horizontal movement to the right, (c) backward movement, (d) vertical movement downwards, (e) horizontal movement to the left, (f) forward movement.

shows the result of the complete pre-processing phase.

7.2.2 Motion Analysis

The employed human motion analysis is based on extracting the apparent relative motion vector between a human observer and the background scene, referred to as the optical flow. The term optical flow can also refer to the distribution of the apparent velocities of movement of brightness pattern in the scene. Technically, the optical flow aims to find the motion pattern in terms of velocity and direction between a sequence of two image frames obtained at times t and $t + \Delta t$ at each temperature value in the acquired heat-map (at every pixel in each pre-processed thermal image). Figure 7.5 shows visible illustrations of extracted motion vectors on different human movements' directions. Accordingly, each movement has a corresponding flow representation. Therefore, it is possible to use optical flow to determine the direction of human movements.

In this thesis, Horn-Schunk algorithm [172] is used to estimate the flow for the output of TSA in the form of velocity and direction. The flow of this optical flow algorithm is formulated as a global energy functional E, which is then minimised. This function for the 2-D thermal image is given as:
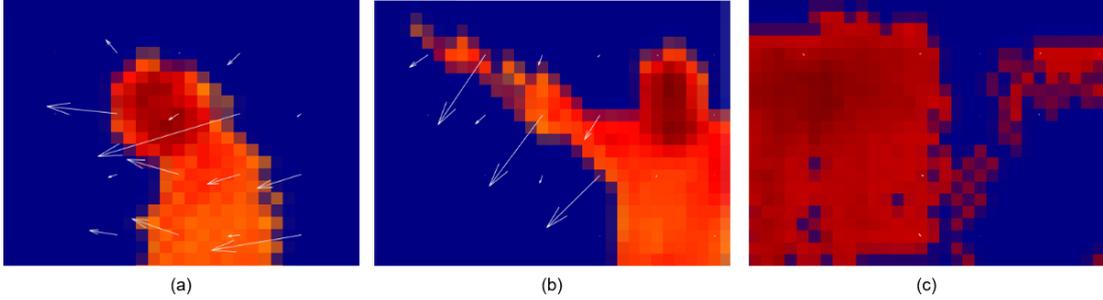
Figure 7.6: The estimate of optical flow on the pre-processed TSA output, (a) human moving horizontally, (b) human hand movement, (c) empty human scene.

$$E = \iint \left[(I_x u + I_y v + I_t)^2 + \alpha^2 \left(\|\nabla u\|^2 + \|\nabla v\|^2 \right) \right] \mathrm{d}x \; \mathrm{d}y \qquad (7.10)$$

where I_x, I_y, I_t refer to the derivatives of the thermal image temperature values along with x, y, and time dimensions, respectively, $\vec{V} = [u(x,y), v(x,y)]^\top$ is the motion vector, and the parameter α is a regularisation constant. Hence, Larger alpha leads to a smoother flow. The assumption behind this algorithm is that there is no significant change in the lighting between two consecutive frames, which refers to as Brightness Constancy Assumption. In other words, the colours of the moving thermal objects should remain the same, regardless of the change in temperature of these objects caused by the evolution of the sensor to human distance. This justifies our proposed adaptive thresholding in the pre-processing phase B to keep the human colours the same in the consecutive thermal image frames regardless of any potential changes in the human-acquired temperatures. Also, it is able to eliminate the background heat, which could be a moving thermal noise generated as a result of human movement.

Figure 7.6 shows sample images on applying the optical flow estimation on the pre-processed thermal images. Figures 7.6(a), 7.6(b) show the estimation of the optical flow on the pre-processed TSA output when the subject is moving horizontally and when there is a hand movement. Since the proposed pre-processing technique aims to find a threshold in the temperature values of the scene regardless of human presence, the relatively high-intensity pixels (temperature) from the background scene appear in the foreground scene of the image in non-human presence scenarios. However, this does not affect the

optical flow estimation because these pixels have no motion, as shown in Figure 7.6(c). Hence, the length of each arrow represents the magnitude of the velocity.

7.2.3 Sensors Interference, Overlap Learning and Fusion

The layout learner uses the motion vectors extracted from the previous phase to find the sensors that interfere with their FoV when enrolling on a new domestic environment (sensors installation). This thesis proposes to find the sensors that interfere with each other in their FoV and the time of the motion to find the order of typical human movement flow by a threshold-based motion trigger for a single occupancy movement. Considering a small home as an example, it requires four TSAs (S_1, S_2, S_3, S_4). If S_1 and S_3 have motion velocities more noticeable than the non-human presence scenes at the time of acquisition, this means that these two sensors are interfering with each other. Besides, the motion sequence order extracted from the outputs of the installed sensors collected at this stage indicates sufficient information regarding the layout of the home, which would be used to regulate the arrangement of motion vectors in the fusion phase for potential human activity recognition applications.

Additionally, there is another situation about the fusion of the sensors in the deployment stage: identifying multiple human subjects occupying the environment is based on matching the motion vectors between interference sensors at the time of acquisition to identify the subjects that may appear in multi-FoVs. This has been achieved by finding the Euclidean distance between moving objects in the interference sensors to be used as a feature in a binary classification problem. Referring to the previous example of a small home with four sensors (S_1, S_2, S_3, S_4), which has two interference sensors S_1 and S_3, the matching between the motion vectors will only be performed for these two sensors and not, e.g. S_1 and S_4, as these sensors do not interfere with each other as referenced in the enrolment stage. Doing this will reduce the required computational resources for the matching task and can overcome the situation when similar human movements are performed simultaneously by different human subjects, e.g. yoga, meditation, or prayers.

Figure 7.7 illustrates the proposed TSA's fusion method of the extracted
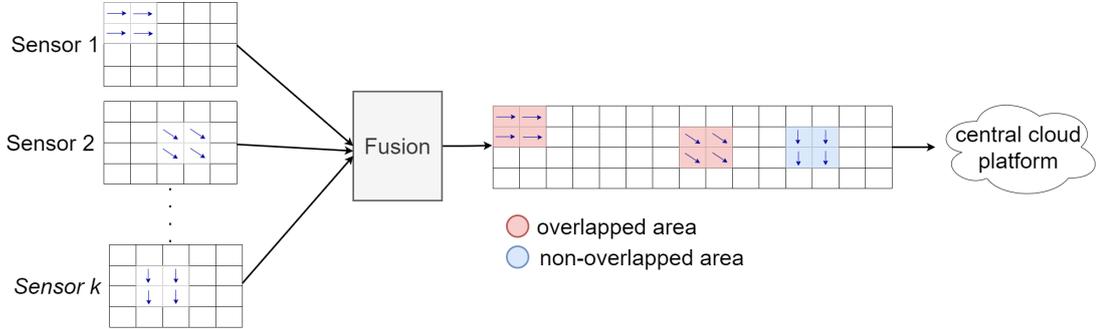
Figure 7.7: The fusion approach of multiple TSAs using the extracted motion vectors. The approach proposes a privacy transmission of the monitored human subject information to a central cloud platform by replacing the temperature values with the extracted motion features.

motion vectors, which has the exact size of the TSA heat-map. The output of this fusion method is a one commutative motion vector with the overlapping motion from different sensors identified. Hence, the order of these motion sequences is based on the obtained reference from the layout learner. Although the TSA sensors do not produce any identifiable information about people, it may still be a privacy risk to transfer their output to a centralised cloud platform to enable further human-centric applications. Therefore, this thesis proposes replacing the TSA output with the motion vector produced from the fusion method.

7.3 Experiments

To evaluate the performance of the proposed methodology, comprehensive experimental work was performed with different use cases and scenarios for sensor positions to confirm the validity of the proposed approach in different sensor placements. During data collection, the experimental home environment was in multi-occupancy mode, which means that a human subject was performing different activities of daily living, including walking, sitting, standing, and lying in bed in the overlapping region of the sensors shown in Figure 7.8 and another person was in a different sensor FoV. Therefore, this section is focused on experimentally verifying the capability of the proposed

(a)

(b)

(c)   (d)

Figure 7.8: Data collection stage, (a) sensors placed side by side at 90°, (b) sensors placed opposite each other, (c) sensors on the same wall, (d) sensors on wall and ceiling.

approach in identifying the sensor overlapped regions through matching different human presence acquired from different TSA placements and human-to-sensor distance on the basis of the moving object motion analysis. Building on top of this, there would not be any problem in the case of more than one human subject moving in a single sensor's FoVs as the problem arises when the same moving object (human) appears in two sensors' output due to overlapped FoVs. The size of the dataset collected to conduct these experiments was 1530 image frames. A detailed description of the experimental results is reported below.

7.3.1 Experiment 1

This experiment contains a human subject moving in the overlapping area of two sensors placed side by side at 90°, as shown in Figure 7.8(a). A third sensor was installed in another room with a different human participant performing normal activities. The extracted human motion vectors from these multiple sensors were analysed by finding the distance between a set of maximum motion velocities for each human participant in the outputs of the sensors. Figure 7.9(a) shows the distance between the overlapped and non-overlapped regions based on human movements.

Finally, the distance is then used as a feature in different classification algorithms, including Logistic Regression, SVM, k-NN, and Linear discriminant analysis (LDA). The performances of these algorithms were 87.9%, 89.6%, 85.2%, 87.9%, respectively. Hence, 10-fold cross-validation is used to avoid the over-fitting problem and guarantee the trained model's generalisation ability.

7.3.2 Experiment 2

Similar to the data collection settings of the first experiment, this experiment contains human participants in different rooms, and the aim is to assess the ability of the proposed methodology in distinguishing between overlapping and non-overlapping regions when sensing interference occurs between opposite sensor positions, as shown in Figure 7.8(b).
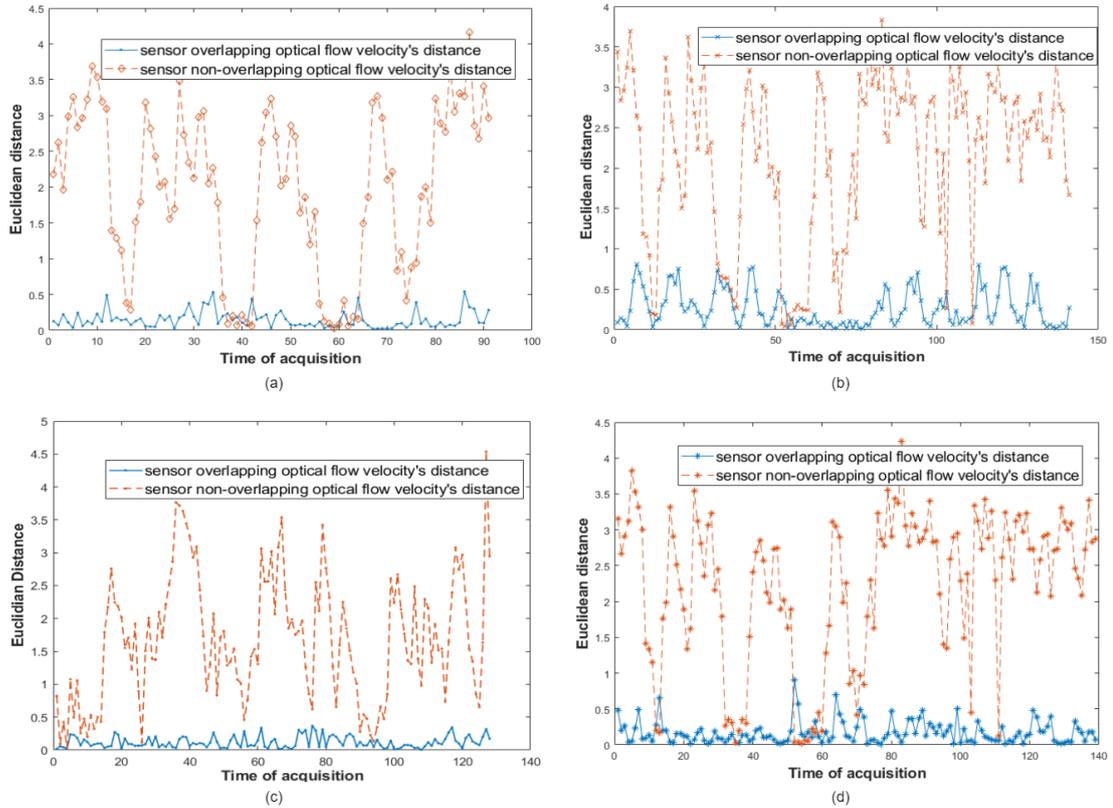
Figure 7.9: Motion analysis between overlapping and non-overlapping FoV's regions with different human participants using data obtained from different sensor placements, (a) sensors placed side by side at 90°, (b) sensors placed opposite each other, (c) sensors on the same wall, (d) sensor on wall and ceiling.

It can be observed from the corresponding visual analysis shown in Figure 7.9(b) is the distance between the magnitude of velocities for different human subject movements is significantly higher than those obtained for the same human subject at the opposite sensor positions. Utilising this distance measurement in the classification algorithms achieves a performance of 92.2%, 92.2%, 89.4%, 90.1% using logistic regression, SVM, k-NN, and LDA, respectively.

7.3.3 Experiment 3

The TSAs could be installed on the same wall but at different heights, as shown in Figure 7.8(c), for example, for human activity recognition for an adult occupancy with a child or to always acquire the upper body of an adult person at the different

sensor to human distance.

This experiment intends to assess the proposed methodology for such a use case scenario. The performance of the proposed methodology using logistic regression was 96.1%, SVM was 96.1%, k-NN was 87.9%, and LDA was 87.1% with 10-fold cross-validation. The justification behind obtaining a better performance on this sensor placement scenario is because the intra-class similarity between the classes is high, as shown in Figure 7.9(c).

7.3.4 Experiment 4

The last experiment was designed to assess the proposed methodology on completely different sensor positions, specifically on wall and ceiling sensor placements, as shown in Figure 7.8(d). The performance of using the distance measurement between the magnitude of velocities of the extracted motion vectors, which is shown in Figure 7.9(d), achieves 92.1% using logistic regression, 96.1% using SVM, 89.2% using k-NN, and 91.0% using LDA.

Table 7.1 provides a comparison of experimental results on identifying the overlapping regions using multiple sensors and different sensor placements. It can be concluded from these results that the proposed methodology provides excellent analytical performance with different sensor placements and human subjects. Besides, identifying the overlap regions from the same wall placement provides the best performance among others. In contrast, the worst performance result was reported in Experiment 1. The rationale behind this lower performance in the side-by-side sensor placements is that a large portion of human presence is missing compared to the other investigated cases, which

Table 7.1: A summary of experimental results on identifying overlapping regions between multiple sensors with different sensor placements.

	Logistic Regression	SVM	k-NN	Linear Discriminant
Experiment 1	87.9%	89.6%	85.2%	87.9%
Experiment 2	92.2%	92.2%	89.4%	90.1%
Experiment 3	96.1%	96.1%	87.9%	87.1%
Experiment 4	92.1%	91.4%	89.2%	91.0%

leads to a loss of motion features for the same human subject at a given time.

7.4 Robust Analysis

This section contains three additional experiments to validate the proposed approach's robustness to enable the fusion of multiple TSAs into human monitoring applications. The first experiment aims to validate the proposed technique to enhance the TSA privacy feature in IoT applications by replacing temperature values with motion vectors for further human-centric applications running on the cloud. Since TSA is proposed in this thesis for human-based applications, this experiment is intended to verify whether human presence can be determined from motion vectors without having to transfer temperature values to a central cloud platform. In this experiment, 218 non-human frame images were collected from noisy thermal scenes and scenes after the human subject left and another 209 human presence frame images. The performance of utilising the motion vectors in human detection using the classification approach was 94.9% achieved accuracy using SVM with 10-folds cross-validation. Hence, this result was based on the result of motion analysis of each frame image, and the performance could be boosted by using all the frame images at a specific time, e.g. every 1 second.

Figure 7.10(a) shows a sample of the distribution of the human temperature values on x and y-axis while Figure 7.10(b) shows the corresponding magnitude of velocities for the same pre-processed image frame. It can be observed from these sample distributions and experimental results that human detection and localisation can be computed using motion vectors without transferring the complete thermal information to the cloud. On the other hand, some human-centred applications may require the provision of human temperature values. However, TSA is a low-resolution sensor, and the acquired temperature varies depending on different conditions, e.g. human-to-sensor distance and wearing heavy clothes. Therefore, in such applications, it is sufficient and private to provide human temperature values from the acquired scene as features (maximum, minimum, median, and average) rather than the full picture of human temperature values to avoid human image construction on this
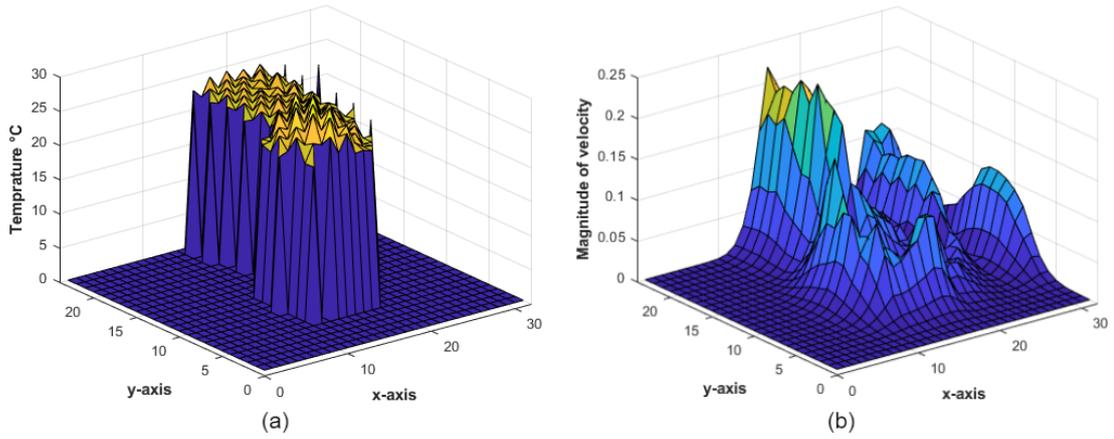
Figure 7.10: A comparison of human temperature distribution versus the corresponding motion vector for human detection and localisation application, (a) the pre-processed thermal scene temperature values, (b) the velocity movement magnitudes for the same acquired thermal scene.
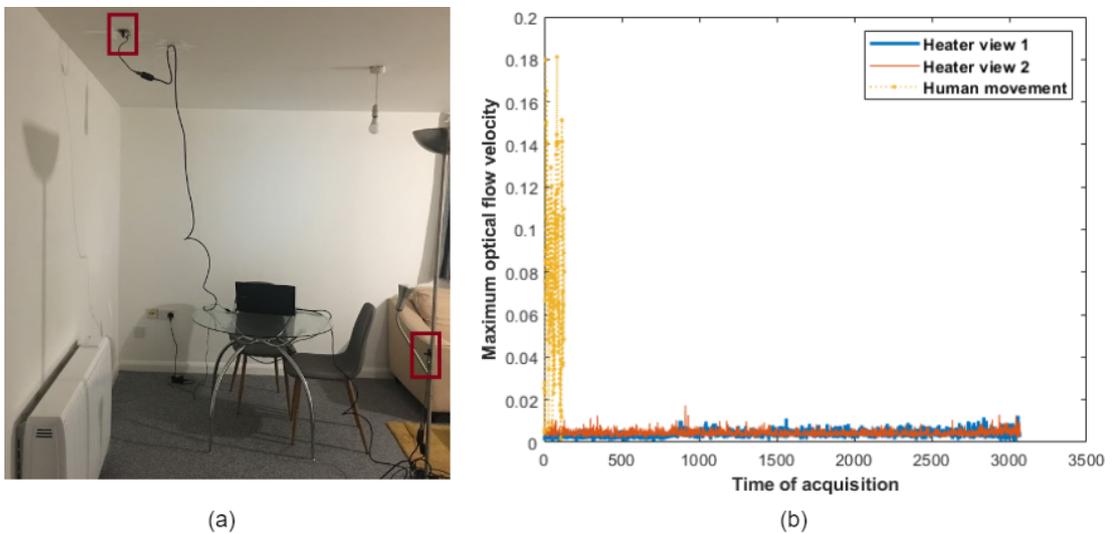
Figure 7.11: A robust analysis of the effects of a moving-based thermal noise generated by a domestic heater on the proposed sensor fusion approach, (a) a data collection stage consists of two simultaneous placements of TSAs, (b) the maximum velocities of the heat generated by a domestic heater from two different sensor placements versus a sample of human movement.

low-resolution sensing approach by a third party.

The second experiment aimed to further evaluate the proposed approach's practicality in a home environment where there could be circulating convection heat from heating systems. In this experiment, two sensors were installed on the top and opposite views of the heater, as shown in Figure 7.11(a). This experiment considered the most complex motion-based thermal noise scenario by collecting data from the cold-heater state to the hot state for more than six continuous hours. The total number of thermal images collected for this experiment is $145,776$ frames. Figure 7.11(b) shows the maximum motion velocities of a subset sample of the collected data from the two heater views versus a human movement.

Two important observations can be drawn from this experiment. First, motion-based thermal noises do not affect the applicability of the proposed approach to enable further human-centric IoT applications. Second, the proposed methodology for TSA overcomes considerable challenges that appear in temperature-based image processing techniques with this type of low-resolution thermal images.

7.5   Chapter Summary

Building on the top of the work presented in Chapters 5, and 6 to identify the multi-occupancy environment, human-to-human distance, and human-to-sensor distance using single-based TSA processing. This chapter has presented a novel approach to enable multi-TSA processing for human behaviour monitoring through motion-based TSA fusion. The proposed approach is capable of learning the domestic environment layout and identifying the overlap regions between multiple TSA's FoV.

To respond to TSA privacy concerns in domestic environments that have been raised in Chapter 4, this chapter proposed an improvement of TSA's privacy-preserving feature. The next chapter examines the proposed motion-based approach in a novel human-in-the-loop abnormal behaviour detection.

Chapter 8

Human-in-the-Loop Anomaly Detection in Activities of Daily Living

8.1 Introduction

To support the independent living of older adults in their own homes, it is essential to identify their abnormal behaviours before triggering an automated alert system. False alerts (false-positive) fall detection has not been addressed thoroughly in systems that report abnormal human behaviours as emergency alerts to the information support. Inspired by the result of Chapter 7, this chapter proposes a novel human-in-the-loop fall detection approach in ADLs through motion-based TSA series processing. The motivation for enabling a human interactive model, fall detection confirmation, is to influence resource efficiency by reducing false-positive alerts while keeping the false-negative fall predictions as low as possible.

The remaining parts of this chapter are organised as follows: Section 8.2 explains the proposed framework architecture. Comprehensive experiments and evaluations are presented and discussed in Sections 8.3, 8.4, 8.5, and 8.6 followed by a chapter summary in Section 8.7.
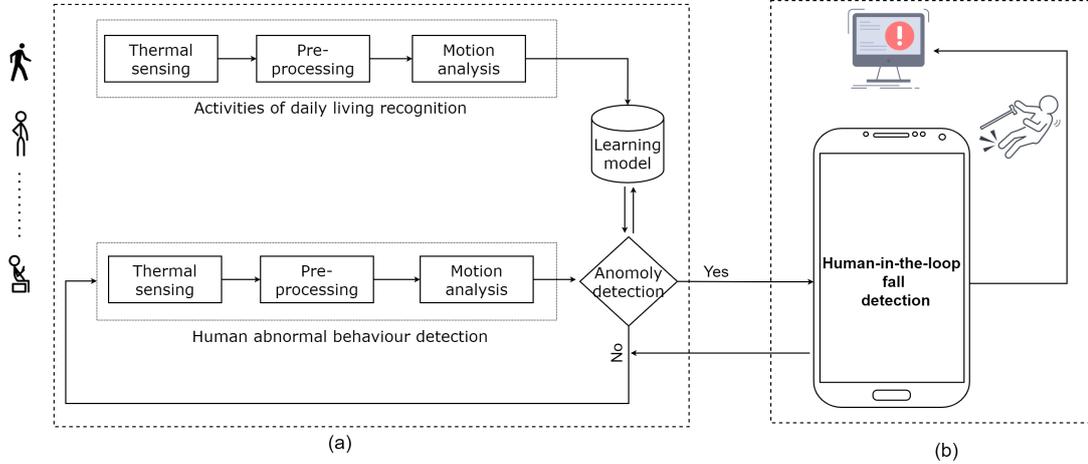
Figure 8.1: A schematic diagram of the proposed accountable human-in-the-loop fall detection system based on optical flow feature extraction, where (a) the ADL and human abnormal detection stage, (b) the human interactive interface to confirm the TSA-based human fall detection.

8.2 Human in the Loop Fall Detection Using a Low-resolution Thermal Sensor Array

The proposed approach, depicted in Figure 8.1, consists of two main stages: stage (a) abnormal behaviour recognition in ADL, and stage (b) human interaction interface. To detect abnormal human behaviours, the proposed system should be capable of distinguishing between normal and abnormal human activities. When a fall is detected, the proposed approach is accountable to the users to confirm the fall to the information support by means of an emergency alert, a notification to family members, etc. Human-based fall confirmation is provided by a mobile application interface model, which requires the user to confirm TSA-based fall detection to proceed to the alert phase. If the user does not confirm the fall, e.g. due to the consequences of the fall, the system will automatically consider this a human fall. Detailed descriptions of the functional phases of these two stages are provided below.
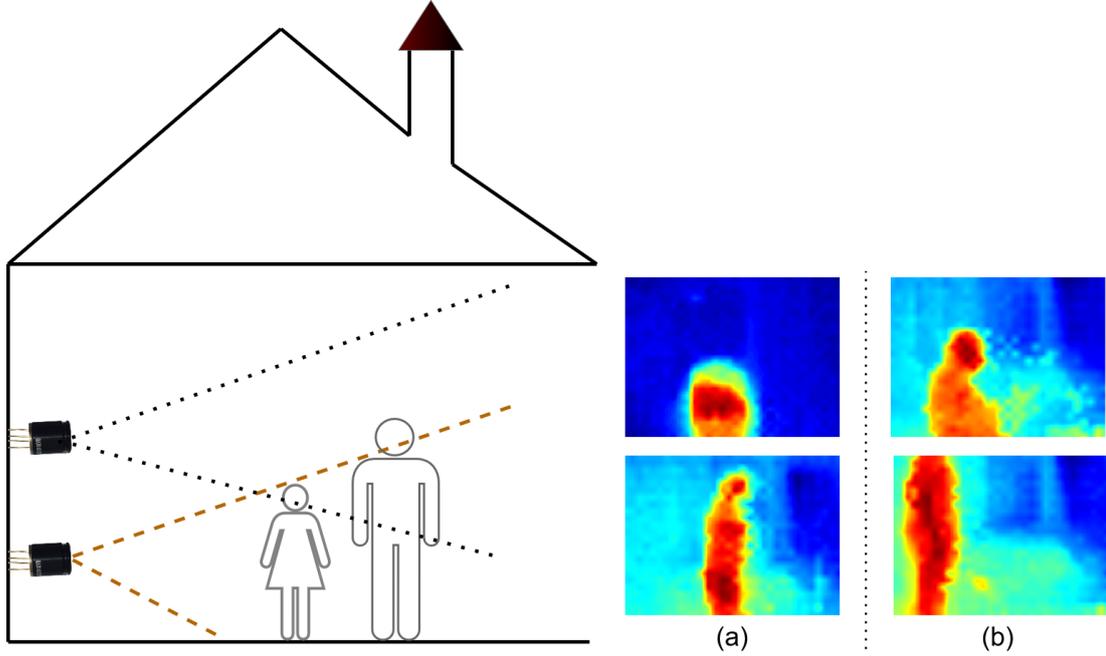
Figure 8.2: Effects of sensor position height on the acquired thermal human presence, (a) a short participant, (b) a tall participant.

8.2.1 Sensor Placement for Fall Detection

Placing the TSA in an elevated position on the wall would acquire the upper parts of the human presence in the sensor's FoV but not the lower parts. Figure 8.2(a) shows the effects of sensor position height on the acquired thermal human presence for a relativity short participant, while Figure 8.2(b) showing a taller participant. It can be seen from these figures that significant human presence has not been acquired for the short participant and is fully visible from short height sensor placement. On the other hand, short height sensor placement has observed some missing upper parts for tall human participants. Since this thesis deals with low-resolution thermal imaging for fall detection, we propose the sensor placement for human fall detection to be at a short height from the floor for a wider inspection area. This will positively affect the efficiency of the proposed approach by overcoming the human-sensor distance limitation. Hence, placing the sensor in the middle of the wall height would only acquire the full human presence at a relatively far human-sensor distance. Since this thesis deals with

fall detection, getting the human presence on the floor is more important than losing some parts of the upper human presence for relatively tall users.

8.2.2 Enabling Human Behaviour Recognition in the Presence of an Animal Pet

An animal pet filter is introduced to eliminate animal pets from the TSA signals since humans, particularly older adults, may own animal pets in single or multi-inhabitant environments. The target of this filter is to remove the thermal noise radiated by animal pets, whose bodies have similar temperatures to human body temperature and are relatively larger sizes than the threshold specified in the connected-component filter described in Chapter 5.

The input of this filter is the pre-processed TSA signal, which indicates the Region of Interest (ROI) after applying the segmentation techniques discussed in Chapter 5. Thus, this filter classifies each ROI into an animal pet or human based on two different feature extraction methods:

The first method uses Principal Component Analysis (PCA) with 95% of the variance. PCA is a linear technique for dimensional reduction based on a linear mapping of the data from a high-dimensional space to a lower-dimensional space. Mathematically, PCA is an orthogonal linear transformation that transforms the data to the greatest variance as a new coordinate system using the scalar projection of the data to lie on the first coordinate, which is referred to as the first principle component, the second greatest variance on the second coordinate, and so on. The transformation is described by a set vector of weights $w_{(k)} = (w_1, \ldots, w_p)_{(k)}$ that transform each row vector $x_{(i)}$ of a data matrix X with column-wise zero mean to a new vector of principal component scores $t_{(i)} = (t_1, \ldots, t_l)_{(i)}$, where each of the p columns represents a specific kind of feature (in this thesis, represents TSA signals), and each of the n shows a different iteration of the experiment. The individual variables $(t_1, \ldots, t_l)_{(i)}$ of t computed by:

$$t_{k(i)} = \mathbf{x}_{(i)} \cdot \mathbf{w}_{(k)}$$
$$\text{for} \quad i = 1, \ldots, n \quad \text{and} \quad k = 1, \ldots, l \tag{8.1}$$
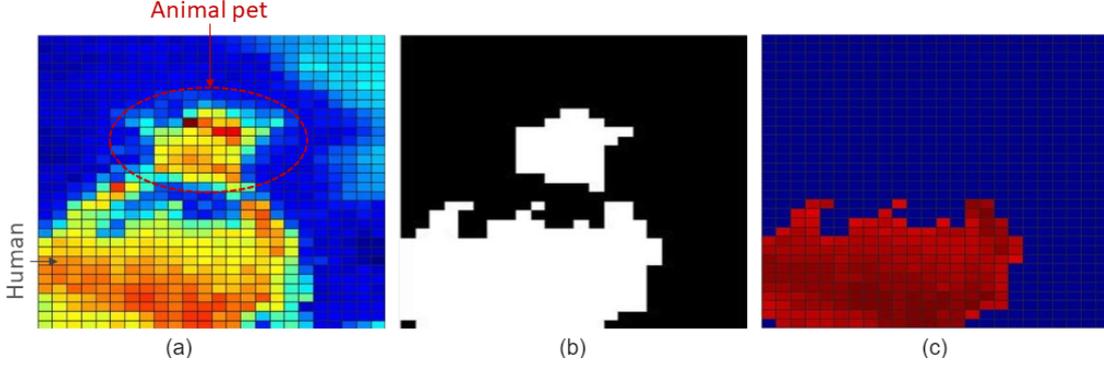
Figure 8.3: Animal pet filter, (a) original heat-map, (b) regions of interest, (c) heat-map after animal pet filter.

In order to maximise variance, the first weight vector has to satisfy:

$$\mathbf{w}_{(1)} = \underset{\|\mathbf{w}\|=1}{\arg\max}\left\{\sum_i (t_1)^2_{(i)}\right\} = \underset{\|\mathbf{w}\|=1}{\arg\max}\left\{\sum_i \left(\mathbf{x}_{(i)} \cdot \mathbf{w}\right)^2\right\} \qquad (8.2)$$

and

$$\mathbf{w}_{(1)} = \arg\max\left\{\frac{\mathbf{w}^T \mathbf{X^T X w}}{\mathbf{w}^T \mathbf{w}}\right\} \qquad (8.3)$$

The principal component Kth can be found, first, subtracting the first $k-1$ components from X

$$\hat{\mathbf{X}}_k = \mathbf{X} - \sum_{s=1}^{k-1} \mathbf{X}\mathbf{w}_{(s)}\mathbf{w}_{(s)}^{\mathrm{T}} \qquad (8.4)$$

Second, extract the maximum variance from the new matrix by finding the weight vector:

$$\mathbf{w}_{(k)} = \underset{\|\mathbf{w}\|=1}{\arg\max}\left\{\left\|\hat{\mathbf{X}}_k\mathbf{w}\right\|^2\right\} = \arg\max\left\{\frac{\mathbf{w}^T \hat{\mathbf{x}}_k^T \hat{\mathbf{x}}_k w}{\mathbf{w}^T \mathbf{w}}\right\} \qquad (8.5)$$

By doing so, the remaining eigenvectors of $X^T X$, with the most significant information, are determined by their corresponding eigenvalues. Hence, the weight vectors are eigenvectors of $X^T X$. The decomposition of X for full principal components can be given as:

$$\mathbf{T} = \mathbf{XW} \tag{8.6}$$

In the second approach, the entropy method is used. The entropy of the one-dimensional histogram of each ROI $\mathbf{X}$ is calculated:

$$H(X) = -\sum_{i=1}^{n} P(x_i) \log P(x_i) \quad where \quad n = histogram \ bins \tag{8.7}$$

The filter will remove the ROI (Figure 8.3(b)) from the heat-map, if the extracted feature vector belongs to the animal pet class. Figure 8.3(c) shows the result of applying this filter on the heat-map shown on Figure 8.3(a).

Finally, human subjects generate background thermal noise as a result of their movements. This generated noise creates biased motion estimation. Therefore, it is important to pre-process the TSA signals prior to human behaviour recognition. TSA signal pre-processing for human motion analysis is provided and discussed in Chapter 7.

8.2.3 Motion Feature Extraction

Two optical flow estimation methods are utilised to support motion analysis and model evaluation. The first method proposed by Horn and Schunck [172] is a sparse optical flow estimation algorithm that estimates the global optical flow. The second algorithm is proposed by Farneback [173], a dense optical flow estimation algorithm that computes the local optical flow of the movements in the acquired scene. Given a particular object in motion at time t, with pixel points x and y, displacement of the object by Δx and Δy over Δt forms a new image expressed as:

$$I(x, y, t) = I(x + \Delta x, y + \Delta y, t + \Delta t) \tag{8.8}$$

considering Taylor series method of approximation:

$$I(x + \Delta x, y + \Delta y, t + \Delta t) = I(x, y, t) + \frac{\partial I}{\partial x}\Delta x + \frac{\partial I}{\partial y}\Delta y + \frac{\partial I}{\partial t}\Delta t \tag{8.9}$$

hence:

$$\frac{\partial I}{\partial x}\Delta x + \frac{\partial I}{\partial y}\Delta y + \frac{\partial I}{\partial t}\Delta t = 0 \tag{8.10}$$

once it is dividing by Δt

$$\frac{\partial I}{\partial x}\frac{\Delta x}{\Delta t} + \frac{\partial I}{\partial y}\frac{\Delta y}{\Delta t} + \frac{\partial I}{\partial t}\frac{\Delta t}{\Delta t} = 0 \tag{8.11}$$

results in:

$$\frac{\partial I}{\partial x}u_x + \frac{\partial I}{\partial y}v_y + \frac{\partial I}{\partial t} = 0 \tag{8.12}$$

therefore:

$$I_x u_x + I_y v_y + I_t = 0 \tag{8.13}$$

where u_x, v_y indicate the velocity of the frame in the horizontal and vertical directions respectively and the derivatives of x and y over time t are represented by I_x, I_y, I_t. Hence, Equation 8.13 has two unknowns, u and v, which can be solved using various mathematical methods. For instance, in the Horn and Schunck (global optical flow estimation) algorithm [172], the velocity estimation between consecutive motion frames is based on two assumptions: brightness and smoothness. For consecutive pixels within two frames, shadows are neglected, and the direction of pixels is the same. Under the smoothness assumption, the derivative of u and v with reference to the x and y directions is calculated in Equation 8.14. Therefore, the expression of this global optical flow estimation can be expressed as follow:

$$E = \iint \left[(I_x u + I_y v + I_t)^2 + \alpha^2 \left(\|\nabla u\|^2 + \|\nabla v\|^2\right)\right] \mathrm{d}x\,\mathrm{d}y \tag{8.14}$$

where $\iint (I_x u + I_y v + I_t)^2 \mathrm{d}x\,\mathrm{d}y$ indicates brightness constancy along x, y and t

dimensions, while α indicates the weighing factor that determines the brightness and smoothness values. $(\|\nabla u\|^2 + \|\nabla v\|^2)\mathrm{d}x\,\mathrm{d}y$ indicates derivatives of u and v with reference to the x and y directions. The smoothness assumption is that all pixels in a particular neighbourhood of a thermal image are observed to move in the same direction.

In contrast to the global optical flow algorithm, the local estimation of the optical flow [173] does not aim to solve the optical flow equation expressed in Equation 8.13. Instead, it considers quadratic polynomial expansion. For each pixel within an image frame, there is a polynomial approximation of the neighbourhood. The velocity of a pixel upon displacement is determined by minimising the error function $e(X)$ of the neighbourhood area expressed in the equation below:

$$e(X) = \sum_{\Delta D \in D} \omega(\Delta D) \left\| A_M(X + \Delta D)\Delta X - \Delta b(X + \Delta D) \right\|^2 \tag{8.15}$$

where D indicates the neighbourhood area, ΔX signifies the pixel displacement, $\omega(\Delta D)$ indicates the Gaussian weighting function that calculates the degree of the neighbourhood area. The higher the Gaussian weighting value, the closer the target pixel. This thesis constructs three pyramid levels to support the motion estimation between pixels with larger displacement in the local optical flow estimation. A high pyramid level indicates high pixel displacement. A 0.5 pyramid scale is specified at each level, which defines the down-sampling rate with three iterations and a neighbourhood size of 5.

Both optical flow estimations return the magnitude and orientation of the movement in a sequence of frames. Since the proposed approach aims to analyse the velocity and direction of objects in motion, no motion is observed for all background static objects. It implies that the magnitude and orientation vectors are low for static objects in the background. Therefore, each activity obtains the highest k magnitude and its corresponding orientation during feature extraction. The elimination of unnecessary background optical flow vectors in the pre-processing phase makes the algorithm more robust because the potential ambient thermal noise is filtered. A representation of optical flow
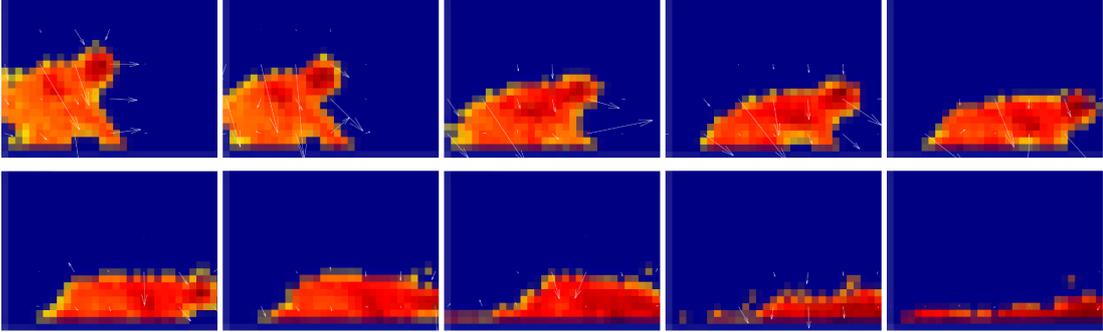
Figure 8.4: An illustrative result of optical flow features on pre-processed TSA output for a subset of the human fall motion sequence.

on a series of pre-processed low-resolution frames is shown in Figure 8.4. From these frames, it is apparent that the optical flow can represent the human motion pattern in terms of velocity and orientation, represented by the length and direction of the arrows in the frames shown in Figure 8.4 respectively. In this specific example, the pattern of human movements starts from up to down until it reaches the inactivity state, which is demonstrated by the decrease in human movement velocity in the last few frames.

8.2.4 Human Activities of Daily Living and Abnormal Behaviour Recognition

To detect an abnormal behaviour (human fall in this study), the proposed system should be capable of recognising associated normal ADL activities, e.g. walking, standing, sitting etc. Therefore, two recurrent neural network architectures are implemented separately, specifically, the LSTM [174] and Bi-LSTM [175] with ADAM [166] and Stochastic Gradient Descent with momentum (SGDM) [176] optimisation algorithms during the training phase to perform ADL recognition including the fall detection. Hence, these two networks can perform a sequence frame classification of the extracted motion vectors from the previous step, which was the justification behind this classification approach in this work.

Each network architecture consists of input, hidden and output layers. Unlike LSTMs, the Bi-LSTM network is designed to support both future and past sequence training. Technically, A Bi-LSTM supports two LSTM layers,

including forward and backwards, arranged parallel to each other. Therefore, its operation is similar to the LSTM, except that the motion sequence of human movement is passed in both directions during the training stage, described as follows:

$$\overrightarrow{h}_t = LSTM(x_t, \overrightarrow{h}_{t-1}; \overrightarrow{W}) \tag{8.16}$$

$$\overleftarrow{h}_t = LSTM(x_t, \overleftarrow{h}_{t-1}; \overleftarrow{W}) \tag{8.17}$$

$$y_t = [\overrightarrow{h}_t, \overleftarrow{h}_t] \tag{8.18}$$

where y_t specifies the output, $\overrightarrow{h}_t$, $\overleftarrow{h}_t$ signify the forward and backward LSTM layers, $\overrightarrow{W}$ and $\overleftarrow{W}$ signify the weights at the forward and backward LSTM layers.

Once a fall is detected, a human-interaction interface is triggered to ensure that the proposed approach is held accountable by human confirmation requirements before the fall is reported to the information support. In this thesis, a mobile interface is used as a human-interaction interface that asks the user in the form of an alert to cancel the fall during time t as shown in Figure 8.5. Suppose the fall is not confirmed within the specified time. In that case, the system will automatically report the fall to the information support in the mean of emergency alerts or notifications to family members, etc. This human-in-the-loop approach for fall detection would enable to overcome false-positive fall detection alerts, for example, to the emergency services while keeping the false-negative fall detection as low as possible.

8.3 Experiments

To evaluate the performance of the proposed methodology, comprehensive experimental work has been performed with different use classification networks, optimisers and optical flow algorithms. The data collection stage consists of eight participants (6 males and 2 females). The participants acted in fall incidents during three activities of daily living, including walking, sitting from a
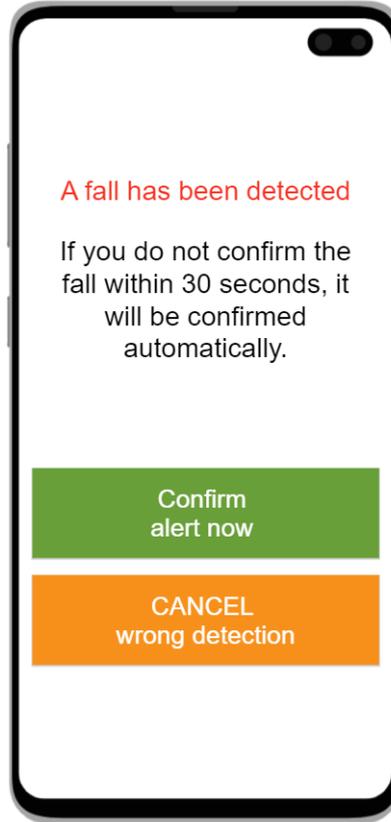
Figure 8.5: A mobile-human interaction interface to confirm the detected fall prior to reporting it to the information support to minimise the false-positive alerts while keeping low false-negative cases.

standing position, and standing from a sitting position. Each participant carried out a particular activity five times. A total of $1,282$ image frames were captured as part of the dataset used for the experiments, which consisted of 226 falling forward image frames, 321 sitting, 288 standing and 447 walking frames. A detailed description of the experimental results is reported below.

8.3.1 Experiments 1

The first experiment involves the identification of falls in the presence of all other activities of daily living. The dataset is divided into 70% for training and 30% for testing. Classification of Local optical flow extracted features using LSTM with ADAM optimiser yields 91.6% compared to SGDM optimiser at
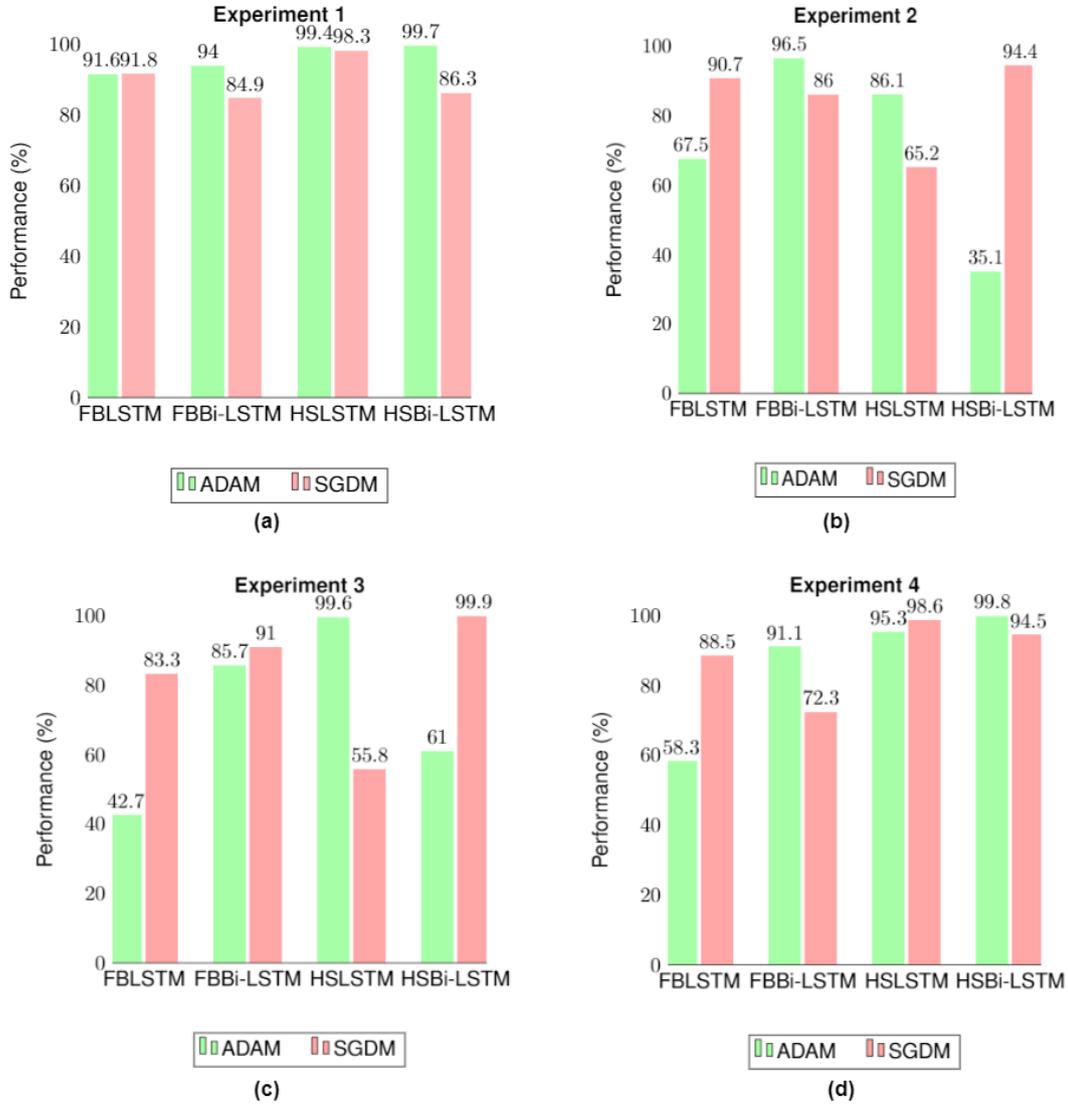
Figure 8.6: A summary of experimental results for (a) detecting human falls among all other ADLs, (b) fall and walk, (c) fall and sitting, (d) fall and stand where FBLSTM indicates the classification of Farneback (local) optical flow extracted features with LSTM, FBBi-LSTM indicates the classification of Farneback optical flow extracted features with Bi-LSTM, HSLSTM indicates the classification of Horn and Schunck (global) optical flow extracted features with LSTM and HSBi-LSTM indicates the classification of Horn and Schunck optical flow extracted features with Bi-LSTM.

91.8%. On the other hand, the classification of Horn and Schunck optical flow extracted features using LSTM with ADAM optimiser achieves performance at 99.4% compared to SGDM at 98.3%. Using a Bi-LSTM network to classify Local optical flow extracted features with ADAM optimiser displays performance at 94.0% compared to SGDM at 84.9%. Finally, classification of Global optical flow extracted features using ADAM optimiser yields performance at 99.7% compared to SGDM at 86.3%. A summary of the results is shown in Figure 8.6(a).

It can be observed from Figure 8.6(a) that the ADAM optimiser gives better performance than the SGDM for all sequence classification algorithms. Not to mention, the Global motion extracted vectors are observed to provide better motion features for classification. The ability of the Global algorithm to process selected pixels within an image during motion estimations is an attribute of its outstanding performance.

8.3.2   Experiments 2

To further evaluate the performance of the developed model, the second set of experiments is performed. It includes two sets of activities, namely, falling forward and walking. Similar to the first use case scenario, the performance of the pre-processing techniques is evaluated for each experiment. A dataset of 673 thermal frames was used, which consisted of 226 fall frames and 447 walk frames. Different optical flow methods are deployed and classified using LSTM and Bi-LSTM with different optimisers ADAM and SGDM. Classification of Local optical flow extracted features using Bi-LSTM with ADAM optimiser yields 96.5% compared to SGDM at 86.0%. On the other hand, the classification of Global optical flow extracted features using Bi-LSTM with ADAM optimiser achieves performance at 35.1% compared to SGDM at 94.4%. Using LSTM network to classify Local optical flow extracted features with ADAM optimiser displays performance at 67.5%, compared to SGDM at 90.7%. Finally, classification of Global optical flow extracted features using ADAM optimiser yields 86.1% compared to SGDM at 65.2%. A summary of the results is shown in Figure 8.6(b).

8.3.3 Experiments 3

This experiment aims at identifying fall frames from a human sitting position. A dataset of 547 thermal frames was used, which consisted of 226 fall and 321 sitting position frames. Similar to the previous scenarios, the performance of various parameters, including pre-processing techniques, classification algorithms, optimisers, and optical flow methods, is performed. Classification of Local optical flow extracted features using Bi-LSTM with ADAM optimiser yields 85.7% compared to SGDM at 91.0%. On the other hand, the classification of Global optical flow extracted features using Bi-LSTM with ADAM optimiser achieves performance at 61.0% compared to SGDM at 99.9%. Using LSTM network to classify Local optical flow extracted features with ADAM optimiser displays performance at 42.7%, compared to SGDM at 83.3%. Finally, classification of Global optical flow extracted features using ADAM optimiser yields 99.6% compared to SGDM at 55.8%. A summary of the results is shown in Figure 8.6(c).

8.3.4 Experiments 4

Finally, the performance of the proposed approach is evaluated using fall and stand frames. A dataset of 514 thermal frames was used, which consisted of 226 human fall and 288 standing frames. Classification of Local optical flow extracted features using Bi-LSTM with ADAM optimiser yields 91.1% compared to SGDM at 72.3%. On the other hand, the classification of Horn and Schunck optical flow extracted features using Bi-LSTM with ADAM optimiser achieves performance at 99.8% compared to SGDM at 94.5%. Using LSTM network to classify Local optical flow extracted features with ADAM optimiser displays performance at 58.3%, compared to SGDM at 88.5%. Finally, classification of Global optical flow extracted features using LSTM with ADAM optimiser yields 95.3% compared to SGDM at 98.6%.

The results shown in Figure 8.6(d) indicate Global motion extracted features with ADAM optimiser as the best performer during the classification of fall and stand activities.

Table 8.1: A comparison of evaluative experiments prior to applying the pre-processing techniques using LSTM for motion sequence classification.

Optical flow algorithm	optimiser	Evaluation 1	Evaluation 2	Evaluation 3	Evaluation 4
Farneback	SGDM	82.3%	57.8%	75.5%	59.6%
Franeback	ADAM	96.2%	99.9%	98.3%	87.7%
Horn-Schunck	SGDM	95.7%	98.7%	99.9%	45.2%
Horn-Schunck	ADAM	97.8%	98.5%	42.7%	45.2%

Table 8.2: A comparison of evaluative experiments prior to applying the pre-processing techniques using Bi-LSTM for motion sequence classification.

Optical flow algorithm	optimiser	Evaluation 5	Evaluation 6	Evaluation 7	Evaluation 8
Farneback	SGDM	92.3%	80.9%	75.2%	57.2%
Franeback	ADAM	91.6%	86.9%	69.7%	45.2%
Horn-Schunck	SGDM	91.4%	98.1%	99.7%	98.6%
Horn-Schunck	ADAM	84.5%	37.9%	51.0%	97.3%

8.4 Experimental Evaluation and Analysis

This section presents an evaluation of the conducted experiments versus the proposed approach to validate the importance of applying the proposed pre-processing techniques to the TSA's output for human motion-based applications. Therefore throughout these evaluative experiments, the suggested pre-processing techniques are not applied to compare them with the results described in Section 8.3. Hence, the same dataset settings were used to repeat each evaluative experiment.

Tables 8.1, and 8.2 show the results of evaluating the proposed approach prior applying the pre-processing techniques on conducted experiments described in Section 8.3 using LSTM and Bi-LSTM, respectively. It can be observed from these tables the application of the pre-processing techniques offers an increase in performance of Global motion extracted features with LSTM and ADAM optimiser from 97.8% to 99.4%. In addition, the performance of Global motion extracted features with Bi-LSTM and ADAM solver increased from 84.5% to 99.7%. Second, the performance of Local motion extracted features with Bi-LSTM and ADAM increased from 86.9% to 96.5% after applying the pre-processing techniques. Third, the classification of Global extracted features with LSTM and ADAM indicates increased performance

from 42.7% to 99.6%. Finally, the sequence motion classification of Global motion extracted features with LSTM and ADAM optimiser indicates an increase in performance from 45.2% to 95.3%. In addition, the classification of Global motion extracted vectors with Bi-LSTM and ADAM optimiser indicates increased performance from 97.3% to 99.8%.

It can be concluded from these empirical evaluations that using the proposed pre-processing technique on TSA outputs for human motion-based applications is necessary to achieve a high-performance and robust system against ambient thermal noise or noises induced by human movements. Besides, LSTM performs better than Bi-LSTM with ADAM optimiser in the conducted experiments.

8.5   Robust Analysis

The first robust analysis is with regard to the ability of the proposed approach to perform human behaviour monitoring in the presence of an animal pet. A dataset containing 81 ROIs for both the human heat-map and cat as a pet case study has been collected to examine the robustness of the two utilised features mentioned in Section 8.2.2 to distinguish between human and animal pet ROIs. Four different classifiers have been used to examine the performance of PCA and entropy features to distinguish between human and animal pet ROIs. The PCA features show a performance of 50%, 75%, 95%, and 100% using medium k-NN, cosine k-NN, Naive Bayes, and SVM, respectively. On the other hand, the performance of these classifiers using entropy estimate shows 100% accuracy in classifying heat maps into pet and human classes.

The second robust analysis is concerned with assessing the ability of the proposed human behaviour monitoring for inactive human subjects. For instance, abnormal sleep movements and behaviours are a subset of nocturnal events that may occur while sleeping, waking, or transitioning into or out of sleep. Nocturnal events can be demonstrated in single movements, repetitive movements, rhythmic movements, and/or complex behaviours such as sleepwalking. The visual output of the TSA shown in Figure 8.7 evident its ability to detect a person's sleeping position while lying in bed. This analysis utilises the extracted motion vectors from TSA outputs rather than the human
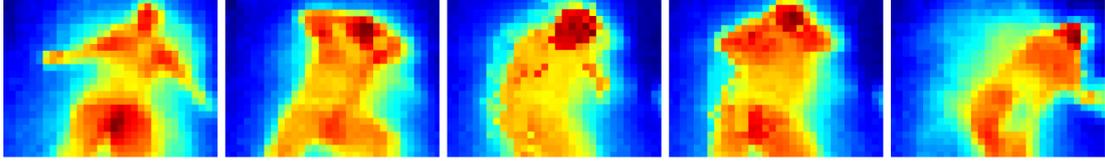
Figure 8.7: A visual demonstration of TSA's ability to acquire the human sleeping position.
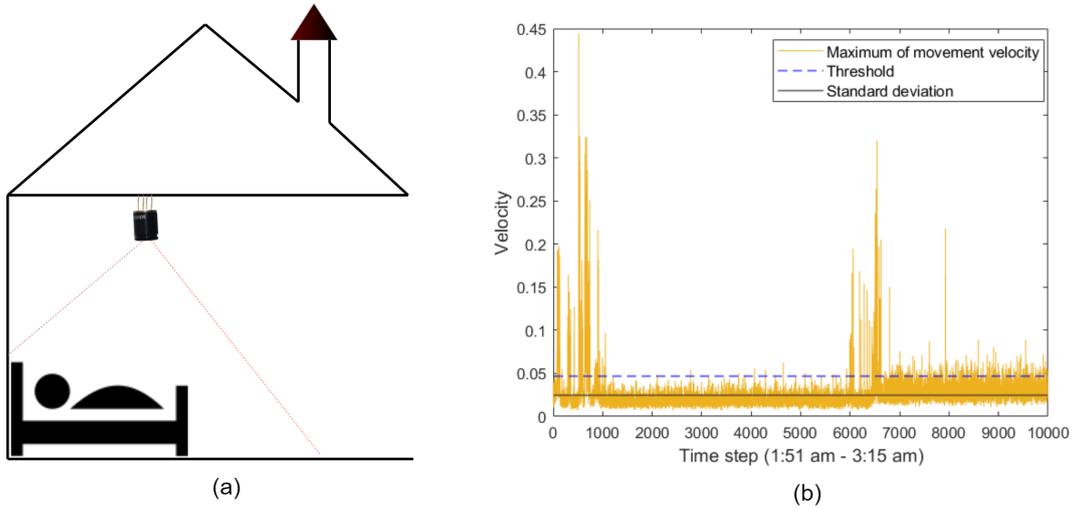
Figure 8.8: Human behaviour monitoring during sleep, (a) TSA placement, (b) A deterministic model to detect the interruption movement during sleep.

physical appearance to detect abnormal human movement during sleep. The justification for using the extracted motion from a series of TSA outputs rather than the TSA output per se is due to the fact that humans tend to cover themselves during sleep. Thus most of their body heat will not be acquired by TSA.

To evaluate the proposed approach to detect abnormal human behaviour movement during sleep, a data set of 9988 continues TSA singles have been collected. The sensor was mounted on the ceiling overhead the sleeping area as illustrated in Figure 8.8(a), while Figure 8.8(b) shows the maximum human movement velocity of the obtained TSA signals. Typically, low velocity indicates no human movement and can represent a good quality sleep pattern, while a higher velocity indicates human movement. In this line, the standard variation S of the data has been computed to find a threshold-based

deterministic model to detect abnormal human movement by adding a sensitivity of 0.9 to the standard variation ($S = S + 0.9$). Additionally, any velocity value above the threshold (the blue dashed line) can be considered as a human movement that could also mean abnormal human movement during sleep. It can be seen from Figure 8.8(b) that the first 1000 TSA signals represent continued human movement. However, the human subject tends to be stable between 1000 to 6000 before it suddenly increases in the sequence between 6000 and 6500. Besides, the signal returned stable, but with higher velocity movement that could represent soft human movement in acquired signals between 6500 and 9988.

8.6 Discussion

The acceptability factors for a deployable domestic human behaviour monitoring solution that can meet the urgency of economic and societal requirements of older adults have been identified in Chapter 1. In summary, the system should be impactful, privacy-friendly, reliable, convenient, and accountable. In this regard, fall detection is abnormal human activity that can occur from standing, sitting or even unpredicted activity. This chapter has addressed a valid issue concerning the users' accountability to the system's decision. On the other hand, this chapter continues the research efforts on TSA-based human behaviour monitoring to achieve the mentioned acceptability factors. In particular, the use of low-resolution TSA has been claimed as a privacy-preserving approach based on the assumption that no identifiable information can be extracted from the raw sensor outputs. However, Chapter 4 has presented a linear relationship between low- and high-resolution thermal imaging. Thus, low-resolution signals can be mapped to high-resolution signals to extract human-identifiable information. Replacing the TSA's signal output with the proposed motion features enhances the privacy-preserving quality of TSA-based human monitoring schemes.

Unlike conventional cameras, TSA is not sensitive to light, but it is sensitive to ambient temperature. This thesis ensures the system's reliability by proposing appropriate pre-processing techniques to enable the proposed

approach to operate in a noisy thermal environment. The proposed pre-processing methods for analysing the motion of TSA signals have been thoroughly evaluated to empirically validate their effectiveness through running the recognition experiments with and without applying the suggested pre-processing methods. Besides, the results of the used series-based learning algorithms show the reliability of the proposed solution in detecting human falls in various ADLs.

Although this chapter suggests placing the sensor at a short height from the floor for a wider inspection area, this placement may have its drawbacks, such as FoV being cluttered with home furniture. In this line, Chapters 5 and 6 have already contributed to resolving issues related to sensor placement flexibility, operating distance, and multi-occupancy environments. Furthermore, Chapter 7 has explored the ability of the proposed motion analysis to identify the motion of the same human subject obtained from multiple sensors' placements in a multi-occupancy environment.

The rationale behind suggesting the smartphone as a human interaction modality to confirm the detected falls is due to the fact that most people have smartphones and would therefore be more apt to use the technology they already have rather than adding an extra cost and effort with unfamiliar modality. Besides, the proposed modality could be a feature of an existing mobile healthcare tracking application, e.g., the UK National Health Service (NHS) mobile app. However, some older adults may not have smartphones to confirm the detection of abnormal behaviour cases that require urgent responses. In this case, there are no restrictions to switching the confirmation from a mobile notification to an automated landline call to confirm or cancel the detected fall.

8.7   Chapter Summary

This chapter has utilised the proposed techniques in the previous chapters to propose a novel human-in-the-loop human abnormal behaviour detection in the presence of an animal pet. Furthermore, this chapter provides a robust analysis to examine the proposed approach's potential to monitor inactive human subjects, e.g., humans during sleep.

The next chapter concludes the thesis and suggests directions for future work on monitoring human behaviour using TSA sensors.

Chapter 9

Conclusion and Future Work

9.1 Thesis Summary

The work undertaken in this thesis has presented a range of novel contributions for domestic human behaviour monitoring to enhance the usability of TSA sensors in the field of Ambient Assisted Living. The proposed monitoring scheme has overcome major challenges associated with TSA sensors, such as static sensor placement, thermal noise, privacy, and limited inspection area, by proposing a chain of novel frameworks toward a user-accountable behaviour monitoring scheme. The privacy-preserving and the ability of the proposed monitoring scheme to operate in a multi-occupancy with the presence of animal pets confirms its potentially significant impact on various sectors, including the healthcare sector, on the example of supporting the independent living of older adults in their own homes.

The following sections in this chapter summarise the concluding remarks drawn from the thesis and suggest directions for future research work.

9.2 Concluding Remarks

The research presented in this thesis has demonstrated the plausibility of developing an accurate framework for monitoring human behaviour using low-resolution thermal imaging. The research effort was carried out in five

sequential phases. First, to understand the opportunities and limitations of TSA by experimentally calibrating between low-resolution TSA and a high-resolution thermal imager. The second phase focused on developing an adaptive sensor placement scheme to segment and identify the number of people in TSA signals. The third focus of this thesis has been to extract physiological-based human localisation knowledge, such as human-to-sensor and human-to-human distances, while the fourth phase has been to shift from single TSA processing to multiple TSA processing through a motion-based sensor fusion approach. Finally, a human-in-the-loop abnormal behaviour detection system has been the capstone of the research effort. The following sub-section outlined the findings of the thesis.

9.2.1 An Empirical Calibration with Privacy Assessment of Low and High-resolution Thermal Imaging

The calibration results reported in Chapter 4 confirm the potential of TSA in human behaviour monitoring applications due to its low cost and development integration compared to commercial high-resolution thermal imagers. However, potential thermography-based human monitoring applications should consider the variation of human temperature acquired using different imaging resolutions and human-to-imager distance. On the other hand, TSA is sensitive to thermal noises, and thus, it is essential to perform robust pre-processing techniques suited explicitly to this sensing methodology.

A regression model that uses an ANN to fit the low-resolution thermal signal to the high-resolution signal has also been proposed to ensure the TSA's privacy-preserving capability. It can be concluded that identifiable human information can be extracted from TSA's output, and therefore security measures should be in place to ensure users' privacy in cloud-based applications. The proposed regression model is also valid for enhancing the low-resolution thermal signal in potential human-centred applications.

9.2.2 Adaptive Sensor Placement for Human Segmentation and Occupancy Estimation

A novel framework has been presented in Chapter 5 for adaptive TSA placement to segment the human presence and estimates the occupancy from TSA signals. The proposed framework is evaluated in different domestic environments, sensor placements, human subjects, and learning techniques. From the obtained results, it can be concluded that using a deep convolutional encoder-decoder network with appropriate pre-processing and post-processing methods that consider the characteristics of the low-resolution TSA is an objective approach for human segmentation. Besides, the adaptive boosting classification algorithm provides accurate results to estimate the occupancy in the proposed framework.

This chapter confirms that TSA's human-centred applications should focus on segmenting human presence rather than human shape detection. This is because the intra-class variations in the human presence using the TSA are relatively high compared to the normal camera with respect to the human-to-sensor distance and sensor placement.

9.2.3 Human Localisation and Physiological Knowledge Extraction

Chapter 6 presents novel human localisation techniques for human-to-sensor and human-to-human distance estimators. Furthermore, A novel feature has been introduced to classify the FoV into depth-based regions depending on the human location in the FoV. The proposed approach has considered the concluding remarks from Chapter 5 regarding the high intra-class variation in the human shape by utilising the suggested human segmentation technique that enables the proposed distance estimators to operate from adaptive sensor placement.

A transfer application using the proposed human-to-sensor distance estimator is presented to extract human physiological features (human height). It can be concluded from the results obtained that the use of TSA could be an approach for human-centred indoor applications.

9.2.4 Human Thermal Behavioural Signal Processing for Sensor Fusion

TSA sensor provides low-resolution thermal imaging, making it an excellent approach for low-cost, privacy-preserving, and passive human-centred applications. In contrast to high-resolution imaging, it is difficult to find a comparable image-based reference to incorporate multiple TSAs to cover a wide inspection area. Chapter 7 proposes a motion-based approach to integrate multiple TSAs and identify overlapping regions in the sensors' FOV. The efficiency of computing time and resources is achieved through proposing an environmental layout learner. Furthermore, this chapter proposes to replace the temperature values acquired by the sensors with the extracted motion vectors for further human activity recognition operating in a centralised cloud platform to avoid re-configuring the human image by a third party.

Extensive experiments were performed to validate the proposed approach on different sensor placement and motion-based thermal noise and a transfer application of the proposed approach to detect human presence using motion vectors without any temperature values.

9.2.5 Human-in-the-Loop Anomaly Detection in Activities of Daily Living

The results of this thesis show that the TSA is capable of bridging the gap between privacy and performance in human-centred applications, including fall detection and ADL recognition. Chapter 8 proposes a motion-based approach for human-in-the-loop fall detection using a low-resolution TSA output. Including a human interactive model to confirm the fall detection results in a very positive impact on reducing the false positive fall detection cases while keeping the false-negative predictions as low as possible. By doing so, the potential deployability and reliability of the system are enhanced as fewer false fall alerts are reported to emergency or information support, resulting in a significant efficiency utilisation of the resources.

Comprehensive experiments and evaluations were conducted to validate the

performance of the proposed approach using local and global optical flow methods, different motion sequence classification algorithms and optimisers. The achieved results indicate that, in general, global optical flow estimation has better performance results with the low-resolution TSA output. The use of optical flow with this type of sensor overcomes the challenges of such sensors to operate in a noisy thermal environment as ambient thermal noises have a shallow motion velocity compared to human motions.

9.3 Future Work and Recommendations

Following the work undertaken in this thesis, this section outlines the potential directions for future work:

- **Inactive human behaviour monitoring**. The detection of abnormal human behaviour was considered in this thesis with human subjects who are awake. Inspired by the results of this thesis to explore the possibility of utilising the proposed scheme to monitor humans during sleep, future work can be undertaken to predict the stage of sleep and sleep behaviour disorder using Electroencephalography (EEG) as a calibrated reference.

- **Human robot collaborative intelligence** . This thesis has presented a human-in-the-loop abnormal behaviour detection to enhance user accountability and boost system performance. It would be an interesting future research direction to build a human-robot collaborative intelligence to confirm the detected abnormal behaviours.

- **User Profiling**. Future work could be undertaken to explore the possibility of utilising the TSA to profile human subjects in a multi-occupancy environment. The discussed motion analysis approach in Chapters 7 and 8 could be the starting point to study the human gait patterns with the aim of classifying human subjects into different age groups or users of interest and users out of interest. This future research direction will be a promising approach to achieving a ranking-based anomaly detection system that categorises detected cases based on their

associated risks or interests. For instance, a young human fall may not be as serious as an older adult fall.

- **Optimising the number of TSA sensors** . One interesting future work direction is to minimise the number of TSA sensors required to cover the domestic environment. Technically, this is a set cover problem under the umbrella of approximation algorithms to identify the smallest sub-collection whose union equals the universe. Implementing such a solution in a user-friendly application will help the system installers identify the number of sensors required for a given environment, which significantly reduces the installation cost.

- **Pilot Testing**. Comprehensive experiments with robustness analysis were conducted to evaluate the performance of the proposed approach. Nevertheless, this thesis has significant social and economic impacts on various sectors, including the older adults' community. In this line, it is important to conduct a pilot testing of the proposed approach under a real-time operating condition through the actual deployment of the proposed approach in older adults' homes for a period of time to evaluate the feasibility, time, cost, risk, and performance of the proposed approach before the widespread deployment.

Appendix A - Further Information Related to Experimental Setup

Table 9.1: A comparison of relevant thermal imagers costs.

Manufacturer name / Model	Price per unit
Omron/ D6T-44L-06	£32.59
Melexis/ MLX90640	£44.49
FLIR/ T1020	£46,114.80

Figure 9.1: The interface, evaluation board, used to connect between the TSA sensors and the PC during the data collection stage.
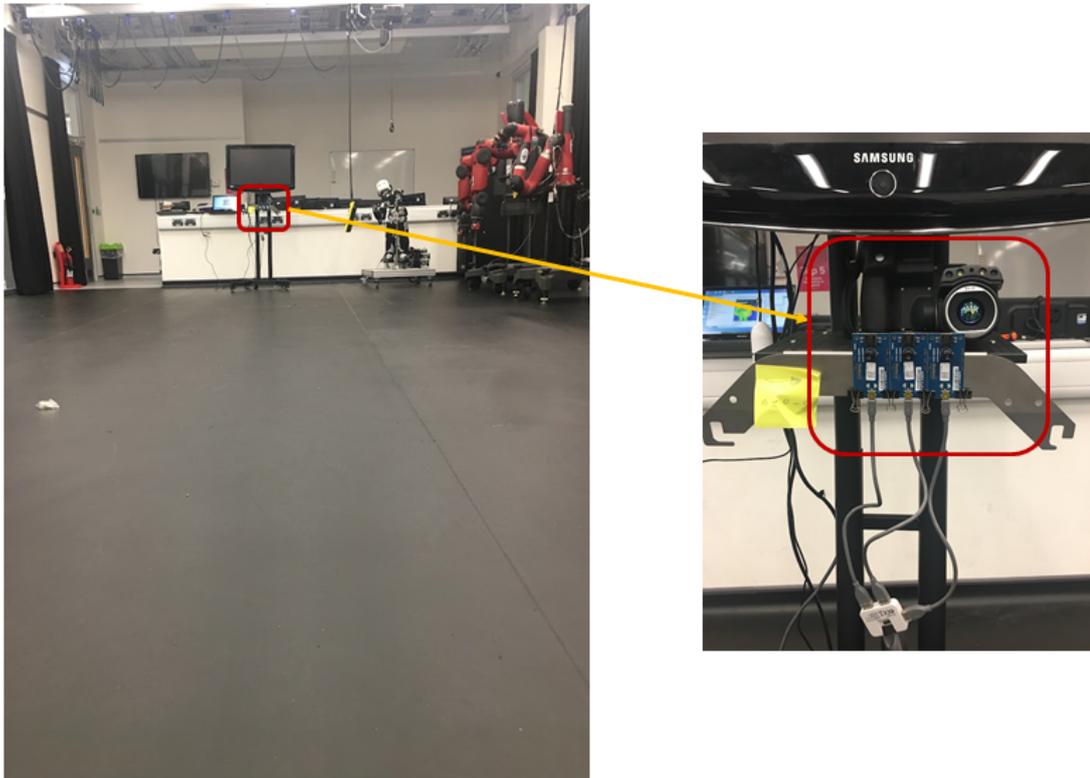
Figure 9.2: The experimental setup to calibrate between different thermal imagers discussed in Chapter 4.

Appendix B - Further Elaborations on the human fall detection results

Figure 9.3: A visual representation of the experimental results of human fall detection among all other ADLs in the form of a confusion matrix, (a) using Bi-LSTM with local optical flow with ADAM optimiser, (b) Bi-LSTM with the local optical flow with SGDM optimiser, (c) LSTM with Global optical flow estimation with Adam optimiser, and (d) LSTM with Global optical flow estimation with SGDM optimiser.

References

[1] United Nations. Department of Economic and Social Affairs. Population Division, *World population ageing 2009*, UN, UK distributor: Stationery Office, 2010. 1

[2] Susan Quine and Stephen Morrell, "Fear of loss of independence and nursing home admission in older australians", *Health & social care in the community*, vol. 15, no. 3, pp. 212–220, 2007. 1

[3] Subhas Chandra Mukhopadhyay, "Wearable sensors for human activity monitoring: A review", *IEEE sensors journal*, vol. 15, no. 3, pp. 1321–1330, 2014. 1

[4] Ahmad Lotfi, Caroline Langensiepen, Sawsan M Mahmoud, and Mohammad Javad Akhlaghinia, "Smart homes for the elderly dementia sufferers: identification and prediction of abnormal behaviour", *Journal of ambient intelligence and humanized computing*, vol. 3, no. 3, pp. 205–218, 2012. 1, 10, 11

[5] Zawar Hussain, Quan Z Sheng, and Wei Emma Zhang, "A review and categorization of techniques on device-free human activity recognition", *Journal of Network and Computer Applications*, vol. 167, pp. 102738, 2020. 2

[6] Han Cui and Naim Dahnoun, "High precision human detection and tracking using millimeter-wave radars", *IEEE Aerospace and Electronic Systems Magazine*, vol. 36, no. 1, pp. 22–32, 2021. 2

[7] Jun Han, Shijia Pan, Manal Kumar Sinha, Hae Young Noh, Pei Zhang, and Patrick Tague, "Smart home occupant identification via sensor fusion

across on-object devices", *ACM Transactions on Sensor Networks (TOSN)*, vol. 14, no. 3-4, pp. 23, 2018. 3, 51

[8] Yan Ding, Shuxue Han, Zhe Tian, Jian Yao, Wanyue Chen, and Qiang Zhang, "Review on occupancy detection and prediction in building simulation", in *Building Simulation*. Springer, 2021, pp. 1–24. 10, 17

[9] Enrique Dorronzoro-Zubiete, Octavio Rivera-Romero, Guido Giunti, and José Luis Sevillano, "Smart home applications for cognitive health of older adults", in *Smart Home Technologies and Services for Geriatric Rehabilitation*, pp. 123–140. Elsevier, 2022. 10

[10] Mirza Mansoor Baig, Shereen Afifi, Hamid GholamHosseini, and Farhaan Mirza, "A systematic review of wearable sensors and iot-based monitoring applications for older adults–a focus on ageing population and independent living", *Journal of medical systems*, vol. 43, no. 8, pp. 1–11, 2019. 10

[11] Majid Shishehgar, Donald Kerr, and Jacqueline Blake, "A systematic review of research into how robotic technology can help older people", *Smart Health*, vol. 7, pp. 1–18, 2018. 10

[12] Aadel Howedi, Ahmad Lotfi, and Amir Pourabdollah, "Exploring entropy measurements to identify multi-occupancy in activities of daily living", *Entropy*, vol. 21, no. 4, pp. 416, 2019. 11, 17, 18, 21

[13] Chia-Ming Wu, Xuan-Ying Chen, Chih-Yu Wen, and William A Sethares, "Cooperative networked pir detection system for indoor human localization", *Sensors*, vol. 21, no. 18, pp. 6180, 2021. 11

[14] Xinyao Tang and Soumyajit Mandal, "Indoor occupancy awareness and localization using passive electric field sensing", *IEEE Transactions on Instrumentation and Measurement*, vol. 68, no. 11, pp. 4535–4549, 2019. 11

[15] Björn Friedrich, Enno-Edzard Steen, Sandra Hellmers, Jürgen M Bauer, and Andreas Hein, "Estimating the gait speed of older adults in smart home environments", *SN Computer Science*, vol. 3, no. 2, pp. 1–14, 2022. 11

[16] A Nait Aicha, Gwenn Englebienne, and B Kröse, "Continuous measuring of the indoor walking speed of older adults living alone", *Journal of ambient intelligence and humanized computing*, vol. 9, no. 3, pp. 589–599, 2018. 11

[17] Jung-Yoon Kim, Chao-Hsien Chu, and Mi-Sun Kang, "Iot-based unobtrusive sensing for sleep quality monitoring and assessment", *IEEE Sensors Journal*, vol. 21, no. 3, pp. 3799–3809, 2020. 11

[18] Sara Casaccia, Eleonora Braccili, Lorenzo Scalise, and Gian Marco Revel, "Experimental assessment of sleep-related parameters by passive infrared sensors: Measurement setup, feature extraction, and uncertainty analysis", *Sensors*, vol. 19, no. 17, pp. 3773, 2019. 11

[19] Diego Droghini, Emanuele Principi, Stefano Squartini, Paolo Olivetti, and Francesco Piazza, "Human fall detection by using an innovative floor acoustic sensor", in *Multidisciplinary Approaches to Neural Computing*, pp. 97–107. Springer, 2018. 11

[20] Diego Droghini, Stefano Squartini, Emanuele Principi, Leonardo Gabrielli, and Francesco Piazza, "Audio metric learning by using siamese autoencoders for one-shot human fall detection", *IEEE Transactions on Emerging Topics in Computational Intelligence*, vol. 5, no. 1, pp. 108–118, 2019. 11

[21] Yingxue Wang, Yanan Chen, Md Zakirul Alam Bhuiyan, Yu Han, Shenghui Zhao, and Jianxin Li, "Gait-based human identification using acoustic sensor and deep neural network", *Future Generation Computer Systems*, vol. 86, pp. 1228–1237, 2018. 11

[22] Sharnil Pandya and Hemant Ghayvat, "Ambient acoustic event assistive framework for identification, detection, and recognition of unknown acoustic events of a residence", *Advanced Engineering Informatics*, vol. 47, pp. 101238, 2021. 11

[23] Eman Shalaby, Nada ElShennawy, and Amany Sarhan, "Utilizing deep learning models in csi-based human activity recognition", *Neural Computing and Applications*, pp. 1–18, 2022. 13

[24] Santosh Kumar Yadav, Siva Sai, Akshay Gundewar, Heena Rathore, Kamlesh Tiwari, Hari Mohan Pandey, and Mohit Mathur, "Csitime: Privacy-preserving human activity recognition using wifi channel state information", *Neural Networks*, vol. 146, pp. 11–21, 2022. 13

[25] Huan Yan, Yong Zhang, Yujie Wang, and Kangle Xu, "Wiact: A passive wifi-based human activity recognition system", *IEEE Sensors Journal*, vol. 20, no. 1, pp. 296–305, 2019. 13

[26] Panos P Markopoulos, Sivan Zlotnikov, and Fauzia Ahmad, "Adaptive radar-based human activity recognition with l1-norm linear discriminant analysis", *IEEE Journal of Electromagnetics, RF and Microwaves in Medicine and Biology*, vol. 3, no. 2, pp. 120–126, 2019. 13

[27] Wen Ding, Xuemei Guo, and Guoli Wang, "Radar-based human activity recognition using hybrid neural network model with multidomain fusion", *IEEE Transactions on Aerospace and Electronic Systems*, vol. 57, no. 5, pp. 2889–2898, 2021. 13

[28] Zhe Chen, Chao Cai, Tianyue Zheng, Jun Luo, Jie Xiong, and Xin Wang, "Rf-based human activity recognition using signal adapted convolutional neural network", *IEEE Transactions on Mobile Computing*, 2021. 13

[29] Jajack Heikenfeld, Andrew Jajack, Jim Rogers, Philipp Gutruf, Lei Tian, Tingrui Pan, Ruya Li, Michelle Khine, Jintae Kim, and Juanhong Wang, "Wearable sensors: modalities, challenges, and prospects", *Lab on a Chip*, vol. 18, no. 2, pp. 217–248, 2018. 13

[30] Min Chen, Yujun Ma, Jeungeun Song, Chin-Feng Lai, and Bin Hu, "Smart clothing: Connecting human with clouds and big data for sustainable health monitoring", *Mobile Networks and Applications*, vol. 21, no. 5, pp. 825–845, 2016. 13

[31] Edward T Chen, "The internet of things: Opportunities, issues, and challenges", in *The internet of things in the modern business environment*, pp. 167–187. IGI global, 2017. 13

[32] Yin Tang, Lei Zhang, Qi Teng, Fuhong Min, and Aiguo Song, "Triple cross-domain attention on human activity recognition using wearable sensors", *IEEE Transactions on Emerging Topics in Computational Intelligence*, 2022. 13

[33] Nidhi Dua, Shiva Nand Singh, and Vijay Bhaskar Semwal, "Multi-input cnn-gru based human activity recognition using wearable sensors", *Computing*, vol. 103, no. 7, pp. 1461–1478, 2021. 13

[34] Rahul Jain, Vijay Bhaskar Semwal, and Praveen Kaushik, "Deep ensemble learning approach for lower extremity activities recognition using wearable sensors", *Expert Systems*, p. e12743, 2021. 13

[35] Ichiro Yamada and Guillaume Lopez, "Wearable sensing systems for healthcare monitoring", in *2012 Symposium on VLSI Technology (VLSIT)*. IEEE, 2012, pp. 5–10. 14

[36] Jeremy Landt, "The history of rfid", *IEEE potentials*, vol. 24, no. 4, pp. 8–11, 2005. 14

[37] Chao Yang, Lingxiao Wang, Xuyu Wang, and Shiwen Mao, "Environment adaptive rfid based 3d human pose tracking with a meta-learning approach", *IEEE Journal of Radio Frequency Identification*, 2022. 14

[38] Yuanqing Zheng and Yanwen Wang, "Rfid-based vital signs monitoring", *Contactless Vital Signs Monitoring*, pp. 257–279, 2022. 14

[39] Shruti S Wakchoure, Priyanka S Shewale, Janhavi G Rajput, Sayali A Gaupal, Mohan P Thakre, and Minal R Rade, "Multiple approach of rfid-based attendance system using iot", in *Soft Computing for Security Applications*, pp. 487–499. Springer, 2022. 14

[40] Amrutha Subrahmannian and Santanu Kumar Behera, "Chipless rfid: A unique technology for mankind", *IEEE Journal of Radio Frequency Identification*, 2022. 14

[41] Junfeng Gao, Yong Yang, Pan Lin, and Dong Sun Park, "Computer vision in healthcare applications", 2018. 14

[42] Pushpajit Khaire, Praveen Kumar, and Javed Imran, "Combining cnn streams of rgb-d and skeletal data for human activity recognition", *Pattern Recognition Letters*, vol. 115, pp. 107–116, 2018. 15

[43] Muhammad Ehatisham-Ul-Haq, Ali Javed, Muhammad Awais Azam, Hafiz MA Malik, Aun Irtaza, Ik Hyun Lee, and Muhammad Tariq Mahmood, "Robust human activity recognition using multimodal feature-level fusion", *IEEE Access*, vol. 7, pp. 60736–60751, 2019. 15

[44] Ibrahim Alrashdi, Muhammad Hameed Siddiqi, Yousef Alhwaiti, Madallah Alruwaili, and Mohammad Azad, "Maximum entropy markov model for human activity recognition using depth camera", *IEEE Access*, vol. 9, pp. 160635–160645, 2021. 15

[45] Yair A Andrade-Ambriz, Sergio Ledesma, Mario-Alberto Ibarra-Manzano, Marvella I Oros-Flores, and Dora-Luz Almanza-Ojeda, "Human activity recognition using temporal convolutional neural network architecture", *Expert Systems with Applications*, vol. 191, pp. 116287, 2022. 15

[46] Seymour Geisser, "Bayesian estimation in multivariate analysis", *The Annals of Mathematical Statistics*, vol. 36, no. 1, pp. 150–159, 1965. 17

[47] Mohamed Chaouch, "Clustering-based improvement of nonparametric functional time series forecasting: Application to intra-day household-level load curves", *IEEE Transactions on Smart Grid*, vol. 5, no. 1, pp. 411–419, 2013. 17

[48] Fan Zhou, Xovee Xu, Goce Trajcevski, and Kunpeng Zhang, "A survey of information cascade analysis: Models, predictions, and recent advances", *ACM Computing Surveys (CSUR)*, vol. 54, no. 2, pp. 1–36, 2021. 17

[49] Wei Zeng, Cong Wang, and Yuanqing Li, "Model-based human gait recognition via deterministic learning", *Cognitive Computation*, vol. 6, no. 2, pp. 218–229, 2014. 17

[50] Chin-Poo Lee, Kian-Ming Lim, and Wei-Lee Woon, "Statistical and entropy based multi purpose human motion analysis", in *2010 2nd International*

Conference on Signal Processing Systems, 2010, vol. 1, pp. V1–734–V1–738. 17

[51] Nong Ye et al., "A markov chain model of temporal behavior for anomaly detection", in *Proceedings of the 2000 IEEE Systems, Man, and Cybernetics information assurance and security workshop*. West Point, NY, 2000, vol. 166, p. 169. 17

[52] Bhavya Mor, Sunita Garhwal, and Ajay Kumar, "A systematic review of hidden markov models and their applications", *Archives of computational methods in engineering*, vol. 28, no. 3, pp. 1429–1448, 2021. 17

[53] Paul A Gagniuc, *Markov chains: from theory to implementation and experimentation*, John Wiley & Sons, 2017. 17

[54] Aadel Howedi, Ahmad Lotfi, and Amir Pourabdollah, "Employing entropy measures to identify visitors in multi-occupancy environments", *Journal of Ambient Intelligence and Humanized Computing*, pp. 1–14, 2020. 18

[55] Michael W Berry, Azlinah Mohamed, and Bee Wah Yap, *Supervised and unsupervised learning for data science*, Springer, 2019. 18

[56] Leslie Pack Kaelbling, Michael L Littman, and Andrew W Moore, "Reinforcement learning: A survey", *Journal of artificial intelligence research*, vol. 4, pp. 237–285, 1996. 18

[57] Robert E Schapire, "A brief introduction to boosting", in *Ijcai*, 1999, vol. 99, pp. 1401–1406. 18

[58] Yoav Freund and Robert E Schapire, "A decision-theoretic generalization of on-line learning and an application to boosting", *Journal of computer and system sciences*, vol. 55, no. 1, pp. 119–139, 1997. 19, 58

[59] Iqbal H Sarker, "Machine learning: Algorithms, real-world applications and research directions", *SN Computer Science*, vol. 2, no. 3, pp. 1–21, 2021. 20

[60] Xin Yan and Xiaogang Su, *Linear regression analysis: theory and computing*, world scientific, 2009. 20

[61] Oludare Isaac Abiodun, Aman Jantan, Abiodun Esther Omolara, Kemi Victoria Dada, Nachaat AbdElatif Mohamed, and Humaira Arshad, "State-of-the-art in artificial neural network applications: A survey", *Heliyon*, vol. 4, no. 11, pp. e00938, 2018. 20

[62] Ganesh Arulampalam and Abdesselam Bouzerdoum, "A generalized feedforward neural network architecture for classification and regression", *Neural networks*, vol. 16, no. 5-6, pp. 561–568, 2003. 20

[63] Saad Albawi, Tareq Abed Mohammed, and Saad Al-Zawi, "Understanding of a convolutional neural network", in *2017 international conference on engineering and technology (ICET)*. Ieee, 2017, pp. 1–6. 20

[64] Joseph M Ackerson, Rushit Dave, and Naeem Seliya, "Applications of recurrent neural network for biometric authentication & anomaly detection", *Information*, vol. 12, no. 7, pp. 272, 2021. 20

[65] Valeriu Manuel Ionescu and Florentina Magda Enescu, "Low cost thermal sensor array for wide area monitoring", in *2020 12th International conference on electronics, computers and artificial intelligence (ECAI)*. IEEE, 2020, pp. 1–4. 21

[66] Elisabetta Moisello, Piero Malcovati, and Edoardo Bonizzoni, "Thermal sensors for contactless temperature measurements, occupancy detection, and automatic operation of appliances during the covid-19 pandemic: A review", *Micromachines*, vol. 12, no. 2, pp. 148, 2021. 21

[67] M Arun, E Baraneetharan, A Kanchana, S Prabu, et al., "Detection and monitoring of the asymptotic covid-19 patients using iot devices and sensors", *International Journal of Pervasive Computing and Communications*, 2020. 21

[68] K Logu, T Devi, N Deepa, S Rakesh Kumar, and N Gayathri, "Real-time mild and moderate covid-19 human body temperature detection using artificial intelligence", *AI and IoT-Based Intelligent Automation in Robotics*, pp. 189–204, 2021. 21

[69] Aya Hossam, Ahmed Magdy, Ahmed Fawzy, and Shriene M Abd El-Kader, "An integrated iot system to control the spread of covid-19 in egypt", in *International Conference on Advanced Intelligent Systems and Informatics*. Springer, 2020, pp. 336–346. 21

[70] Ash Tyndall, Rachel Cardell-Oliver, and Adrian Keating, "Occupancy estimation using a low-pixel count thermal imager", *IEEE Sensors Journal*, vol. 16, no. 10, pp. 3784–3791, 2016. 21, 22, 23, 26, 33, 65

[71] Homagni Saha, Anthony R Florita, Gregor P Henze, and Soumik Sarkar, "Occupancy sensing in buildings: A review of data analytics approaches", *Energy and Buildings*, vol. 188, pp. 278–285, 2019. 21

[72] Zhenghua Chen, Chaoyang Jiang, and Lihua Xie, "Building occupancy estimation and detection: A review", *Energy and Buildings*, vol. 169, pp. 260–270, 2018. 21

[73] Athanasios Bamis, Dimitrios Lymberopoulos, Thiago Teixeira, and Andreas Savvides, "The behaviorscope framework for enabling ambient assisted living", *Personal and Ubiquitous Computing*, vol. 14, no. 6, pp. 473–487, 2010. 21

[74] Juan Serrano-Cuerda, José Carlos Castillo, Marina V Sokolova, and Antonio Fernández-Caballero, "Efficient people counting from indoor overhead video camera", in *Trends in Practical Applications of Agents and Multiagent Systems*, pp. 129–137. Springer, 2013. 21

[75] Varick L Erickson, Stefan Achleitner, and Alberto E Cerpa, "Poem: Power-efficient occupancy-based energy management system", in *Proceedings of the 12th international conference on Information processing in sensor networks*. ACM, 2013, pp. 203–216. 21

[76] James Scott, AJ Bernheim Brush, John Krumm, Brian Meyers, Michael Hazas, Stephen Hodges, and Nicolas Villar, "Preheat: controlling home heating using occupancy prediction", in *Proceedings of the 13th international conference on Ubiquitous computing*. ACM, 2011, pp. 281–290. 21

[77] Azkario Rizky Pratama, Alexander Lazovik, and Marco Aiello, "Office multi-occupancy detection using ble beacons and power meters", in *2019 IEEE 10th Annual Ubiquitous Computing, Electronics & Mobile Communication Conference (UEMCON)*. IEEE, 2019, pp. 0440–0448. 22

[78] Aaron S Crandall and Diane J Cook, "Coping with multiple residents in a smart environment", *Journal of Ambient Intelligence and Smart Environments*, vol. 1, no. 4, pp. 323–334, 2009. 22

[79] Qingjuan Li, Wei Huangfu, Fadi Farha, Tao Zhu, Shunkun Yang, Liming Chen, and Huansheng Ning, "Multi-resident type recognition based on ambient sensors activity", *Future Generation Computer Systems*, 2020. 22

[80] Jie Yin, Meng Fang, Ghassem Mokhtari, and Qing Zhang, "Multi-resident location tracking in smart home through non-wearable unobtrusive sensors", in *International Conference on Smart Homes and Health Telematics*. Springer, 2016, pp. 3–13. 22

[81] Zhenghua Chen, Chaoyang Jiang, Mustafa K Masood, Yeng Chai Soh, Min Wu, and Xiaoli Li, "Deep learning for building occupancy estimation using environmental sensors", in *Deep Learning: Algorithms and Applications*, pp. 335–357. Springer, 2020. 22

[82] S Mamidi, R Maheswaran, and Y Chang, "Smart sensing, estimation, and prediction for efficient building energy management", in *Multi-agent Smart Computing Workshop*, 2011. 22

[83] Khee Poh Lam, Michael Höynck, Bing Dong, Burton Andrews, Yun-Shang Chiou, Rui Zhang, Diego Benitez, Joonho Choi, et al., "Occupancy detection through an extensive environmental sensor network in an open-plan office building", *IBPSA Building Simulation*, vol. 145, pp. 1452–1459, 2009. 22

[84] T Wang and DJ Cook, "smrt: Multi-resident tracking in smart homes with sensor vectorization.", *IEEE transactions on pattern analysis and machine intelligence*, 2020. 22

[85] Munkhjargal Gochoo, Tan-Hsu Tan, Shing-Hong Liu, Fu-Rong Jean, Fady S Alnajjar, and Shih-Chia Huang, "Unobtrusive activity recognition of elderly people living alone using anonymous binary sensors and dcnn", *IEEE journal of biomedical and health informatics*, vol. 23, no. 2, pp. 693–702, 2018. 22

[86] Diane J Cook, "Learning setting-generalized activity models for smart spaces", *IEEE intelligent systems*, vol. 2010, no. 99, pp. 1, 2010. 22

[87] Narayanan C Krishnan and Diane J Cook, "Activity recognition on streaming sensor data", *Pervasive and mobile computing*, vol. 10, pp. 138–154, 2014. 22

[88] KS Gayathri, KS Easwarakumar, and Susan Elias, "Probabilistic ontology based activity recognition in smart homes using markov logic network", *Knowledge-Based Systems*, vol. 121, pp. 173–184, 2017. 22

[89] Joseph Rafferty, Chris D Nugent, Jun Liu, and Liming Chen, "From activity recognition to intention recognition for assisted living within smart homes", *IEEE Transactions on Human-Machine Systems*, vol. 47, no. 3, pp. 368–379, 2017. 22

[90] John G Cleary and Leonard E Trigg, "K*: An instance-based learner using an entropic distance measure", in *Machine Learning Proceedings 1995*, pp. 108–114. Elsevier, 1995. 22

[91] Alex Beltran, Varick L Erickson, and Alberto E Cerpa, "Thermosense: Occupancy thermal based sensing for hvac control", in *Proceedings of the 5th ACM Workshop on Embedded Systems For Energy-Efficient Buildings*. ACM, 2013, pp. 1–8. 23, 26, 65

[92] Yazhou Yuan, Xin Li, Zhixin Liu, and Xinping Guan, "Occupancy estimation in buildings based on infrared array sensors detection", *IEEE Sensors Journal*, vol. 20, no. 2, pp. 1043–1053, 2019. 23, 26

[93] Andres Gomez, Francesco Conti, and Luca Benini, "Thermal image-based cnn's for ultra-low power people recognition", in *Proceedings of the 15th*

ACM International Conference on Computing Frontiers, 2018, pp. 326–331. 23, 65

[94] Hessam Mohammadmoradi, Sirajum Munir, Omprakash Gnawali, and Charles Shelton, "Measuring people-flow through doorways using easy-to-install ir array sensors", in *2017 13th International Conference on Distributed Computing in Sensor Systems (DCOSS)*. IEEE, 2017, pp. 35–43. 23

[95] Mertcan Cokbas, Prakash Ishwar, and Janusz Konrad, "Low-resolution overhead thermal tripwire for occupancy estimation", in *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition Workshops*, 2020, pp. 88–89. 23, 26

[96] Aly Metwaly, Jorge Peña Queralta, Victor Kathan Sarker, Tuan Nguyen Gia, Omar Nasir, and Tomi Westerlund, "Edge computing with embedded ai: Thermal image analysis for occupancy estimation in intelligent buildings", in *Proceedings of the INTelligent Embedded Systems Architectures and Applications Workshop 2019*, 2019, pp. 1–6. 23, 65

[97] Liam Walmsley-Eyre and Rachel Cardell-Oliver, "Hierarchical classification of low resolution thermal images for occupancy estimation", in *2017 IEEE 42nd conference on local computer networks workshops (LCN Workshops)*. IEEE, 2017, pp. 9–17. 23, 26

[98] Dongfang Yang, Ekim Yurtsever, Vishnu Renganathan, Keith A Redmill, and Ümit Özgüner, "A vision-based social distancing and critical density detection system for covid-19", *arXiv preprint arXiv:2007.03578*, pp. 24–25, 2020. 23

[99] Maya Aghaei, Matteo Bustreo, Yiming Wang, Gianluca Bailo, Pietro Morerio, and Alessio Del Bue, "Single image human proxemics estimation for visual social distancing", in *Proceedings of the IEEE/CVF Winter Conference on Applications of Computer Vision*, 2020, pp. 2785–2795. 23

[100] Adarsh Jagan Sathyamoorthy, Utsav Patel, Yash Ajay Savle, Moumita Paul, and Dinesh Manocha, "Covid-robot: Monitoring social distancing

constraints in crowded scenarios", *arXiv preprint arXiv:2008.06585*, 2020. 23

[101] Rinkal Keniya and Ninad Mehendale, "Real-time social distancing detector using socialdistancingnet-19 deep learning network", *Available at SSRN 3669311*, 2020. 23

[102] Narinder Singh Punn, Sanjay Kumar Sonbhadra, and Sonali Agarwal, "Monitoring covid-19 social distancing with person detection and tracking via fine-tuned yolo v3 and deepsort techniques", *arXiv preprint arXiv:2005.01385*, 2020. 23

[103] Manish Shukla, Sachin Lodha, Gautam Shroff, Ramesh Raskar, et al., "Privacy guidelines for contact tracing applications", *arXiv preprint arXiv:2004.13328*, 2020. 23

[104] Scott McLachlan, Peter Lucas, Kudakwashe Dube, Graham A Hitman, Magda Osman, Evangelia Kyrimi, Martin Neil, and Norman E Fenton, "Bluetooth smartphone apps: Are they the most private and effective solution for covid-19 contact tracing?", *arXiv preprint arXiv:2005.06621*, 2020. 23

[105] Hossein Gorji, Markus Arnoldini, David F Jenny, Wolf-Dietrich Hardt, and Patrick Jenny, "Stecc: Smart testing with contact counting enhances covid-19 mitigation by bluetooth app based contact tracing", *medRxiv*, 2020. 23

[106] Sizhen Bian, Bo Zhou, and Paul Lukowicz, "Social distance monitor with a wearable magnetic field proximity sensor", *Sensors*, vol. 20, no. 18, pp. 5101, 2020. 23

[107] Mojtaba Akbari, Jay Carriere, Tyler Meyer, Ron Sloboda, Siraj Husain, Nawaid Usmani, and Mahdi Tavakoli, "Robotic ultrasound scanning with real-time image-based force adjustment: quick response for enabling physical distancing during the covid-19 pandemic", *Frontiers in Robotics and AI*, vol. 8, pp. 62, 2021. 23

[108] Spencer Woody, Mauricio Garcia Tec, Maytal Dahan, Kelly Gaither, Michael Lachmann, Spencer Fox, Lauren Ancel Meyers, and James G

Scott, "Projections for first-wave covid-19 deaths across the us using social-distancing measures derived from mobile phones", *medRxiv*, 2020. 24

[109] Sizhen Bian, Vitor F Rey, Junaid Younas, and Paul Lukowicz, "Wrist-worn capacitive sensor for activity and physical collaboration recognition", in *2019 IEEE International Conference on Pervasive Computing and Communications Workshops (PerCom Workshops)*. IEEE, 2019, pp. 261–266. 24

[110] Sizhen Bian, Bo Zhou, Hymalai Bello, and Paul Lukowicz, "A wearable magnetic field based proximity sensing system for monitoring covid-19 social distancing", in *Proceedings of the 2020 International Symposium on Wearable Computers*, 2020, pp. 22–26. 24

[111] Jingyi Zhang, Jianxin Chen, Qingyu Lin, and Liming Cheng, "Moving object distance estimation method based on target extraction with a stereo camera", in *2019 IEEE 4th International Conference on Image, Vision and Computing (ICIVC)*. IEEE, 2019, pp. 572–577. 24, 26

[112] Jürgen Kemper and Daniel Hauschildt, "Passive infrared localization with a probability hypothesis density filter", in *2010 7th Workshop on Positioning, Navigation and Communication*. IEEE, 2010, pp. 68–76. 24, 26

[113] Hock M Ng, "Human localization and activity detection using thermopile sensors", in *2013 ACM/IEEE International Conference on Information Processing in Sensor Networks (IPSN)*. IEEE, 2013, pp. 337–338. 24, 26

[114] Masato Kuki, Hiroshi Nakajima, Naoki Tsuchiya, and Yutaka Hata, "Human movement trajectory recording for home alone by thermopile array sensor", in *2012 IEEE International Conference on Systems, Man, and Cybernetics (SMC)*. IEEE, 2012, pp. 2042–2047. 24, 26

[115] Zhangjie Chen, Ya Wang, and Hanwei Liu, "Unobtrusive sensor-based occupancy facing direction detection and tracking using advanced machine learning algorithms", *IEEE Sensors Journal*, vol. 18, no. 15, pp. 6360–6368, 2018. 24, 26

[116] Peter Hevesi, Sebastian Wille, Gerald Pirkl, Norbert Wehn, and Paul Lukowicz, "Monitoring household activities and user location with a cheap, unobtrusive thermal sensor array", in *Proceedings of the 2014 ACM international joint conference on pervasive and ubiquitous computing*, 2014, pp. 141–145. 24, 26

[117] Jindrich Adolf, Martin Macas, Lenka Lhotska, and Jaromir Dolezal, "Deep neural network based body posture recognitions and fall detection from low resolution infrared array sensor", in *2018 IEEE International Conference on Bioinformatics and Biomedicine (BIBM)*. IEEE, 2018, pp. 2394–2399. 24, 25

[118] Wei-Han Chen and Hsi-Pin Ma, "A fall detection system based on infrared array sensors with tracking capability for the elderly at home", in *2015 17th International Conference on E-health Networking, Application & Services (HealthCom)*. IEEE, 2015, pp. 428–434. 24, 25

[119] Zhangjie Chen and Ya Wang, "Infrared–ultrasonic sensor fusion for support vector machine–based fall detection", *Journal of Intelligent Material Systems and Structures*, vol. 29, no. 9, pp. 2027–2039, 2018. 24, 25, 26

[120] Shota Mashiyama, Jihoon Hong, and Tomoaki Ohtsuki, "A fall detection system using low resolution infrared array sensor", in *2014 IEEE 25th Annual International Symposium on Personal, Indoor, and Mobile Radio Communication (PIMRC)*. IEEE, 2014, pp. 2109–2113. 24, 25, 26

[121] Zhixin Liu, Ming Yang, Yazhou Yuan, and Kit Yan Kan, "Fall detection and personnel tracking system using infrared array sensors", *IEEE Sensors Journal*, 2020. 24, 25

[122] Yuya Ogawa and Katsuhiro Naito, "Fall detection scheme based on temperature distribution with ir array sensor", in *2020 IEEE International Conference on Consumer Electronics (ICCE)*. IEEE, 2020, pp. 1–5. 24, 25

[123] Sagar Shelke and Baris Aksanli, "Static and dynamic activity detection with ambient sensors in smart spaces", *Sensors*, vol. 19, no. 4, pp. 804, 2019. 24, 25

[124] Yoosuf Nizam, Mohd Norzali Haji Mohd, and M Mahadi Abdul Jamil, "A study on human fall detection systems: Daily activity classification and sensing techniques", *International Journal of Integrated Engineering*, vol. 8, no. 1, 2016. 24

[125] Samad Barri Khojasteh, José R Villar, Camelia Chira, Víctor M González, and Enrique De la Cal, "Improving fall detection using an on-wrist wearable accelerometer", *Sensors*, vol. 18, no. 5, pp. 1350, 2018. 24

[126] Faisal Hussain, Fawad Hussain, Muhammad Ehatisham-ul Haq, and Muhammad Awais Azam, "Activity-aware fall detection and recognition based on wearable sensors", *IEEE Sensors Journal*, vol. 19, no. 12, pp. 4528–4536, 2019. 24

[127] Muhammad Mubashir, Ling Shao, and Luke Seed, "A survey on fall detection: Principles and approaches", *Neurocomputing*, vol. 100, pp. 144–152, 2013. 24

[128] Ahmad Lotfi, Suad Albawendi, Heather Powell, Kofi Appiah, and Caroline Langensiepen, "Supporting independent living for older adults; employing a visual based fall detection through analysing the motion and shape of the human body", *IEEE Access*, vol. 6, pp. 70272–70282, 2018. 24

[129] Fouzi Lezzar, Djamel Benmerzoug, and Ilham Kitouni, "Camera-based fall detection system for the elderly with occlusion recognition", *Applied Medical Informatics.*, vol. 42, no. 3, pp. 169–179, 2020. 24

[130] Nuno Ferrete Ribeiro and Cristina P Santos, "Two fall-related and kinematic data-based approaches for an instrumented conventional cane", *IEEE Transactions on Human-Machine Systems*, vol. 51, no. 5, pp. 554–563, 2021. 24

[131] Seung-Bae Jeon, Young-Hoon Nho, Sang-Jae Park, Weon-Guk Kim, Il-Woong Tcho, Daewon Kim, Dong-Soo Kwon, and Yang-Kyu Choi, "Self-powered fall detection system using pressure sensing triboelectric nanogenerators", *Nano Energy*, vol. 41, pp. 139–147, 2017. 26

[132] Ahmet Yazar, Furkan Keskin, B Ugur Töreyin, and A Enis Çetin, "Fall detection using single-tree complex wavelet transform", *Pattern Recognition Letters*, vol. 34, no. 15, pp. 1945–1952, 2013. 26

[133] Akira Hayashida, Vasily Moshnyaga, and Koji Hashimoto, "The use of thermal ir array sensor for indoor fall detection", in *2017 IEEE International Conference on Systems, Man, and Cybernetics (SMC)*. IEEE, 2017, pp. 594–599. 26

[134] Carla Taramasco, Tomas Rodenas, Felipe Martinez, Paola Fuentes, Roberto Munoz, Rodrigo Olivares, Victor Hugo C De Albuquerque, and Jacques Demongeot, "A novel monitoring system for fall detection in older people", *Ieee Access*, vol. 6, pp. 43563–43574, 2018. 26

[135] Shigeyuki Tateno, Fanxing Meng, Renzhong Qian, and Yuriko Hachiya, "Privacy-preserved fall detection method with three-dimensional convolutional neural network using low-resolution infrared array sensor", *Sensors*, vol. 20, no. 20, pp. 5957, 2020. 26

[136] José Luis Honorato, Ignacio Spiniak, and Miguel Torres-Torriti, "Human detection using thermopiles", in *2008 IEEE Latin American Robotic Symposium*. IEEE, 2008, pp. 151–157. 26

[137] Veena Chidurala and Xinrong Li, "Occupancy estimation using thermal imaging sensors and machine learning algorithms", *IEEE Sensors Journal*, vol. 21, no. 6, pp. 8627–8638, 2021. 26

[138] Cristian Perra, Amit Kumar, Michele Losito, Paolo Pirino, Milad Moradpour, and Gianluca Gatto, "Monitoring indoor people presence in buildings using low-cost infrared sensor array in doorways", *Sensors*, vol. 21, no. 12, pp. 4062, 2021. 26

[139] Cunyi Yin, Jing Chen, Xiren Miao, Hao Jiang, and Deying Chen, "Device-free human activity recognition with low-resolution infrared array sensor using long short-term memory neural network", *Sensors*, vol. 21, no. 10, pp. 3551, 2021. 26

[140] Kshirasagar Naik, Tejas Pandit, Nitin Naik, and Parth Shah, "Activity recognition in residential spaces with internet of things devices and thermal imaging", *Sensors*, vol. 21, no. 3, pp. 988, 2021. 26

[141] Luis Ignacio Lopera Gonzalez, Marc Troost, and Oliver Amft, "Using a thermopile matrix sensor to recognize energy-related activities in offices", *Procedia Computer Science*, vol. 19, pp. 678–685, 2013. 26

[142] Zhangjie Chen and Ya Wang, "Sleep monitoring using an infrared thermal array sensor", in *Sensors and Smart Structures Technologies for Civil, Mechanical, and Aerospace Systems 2019*. International Society for Optics and Photonics, 2019, vol. 10970, p. 109701D. 26

[143] Rikke Gade and Thomas B Moeslund, "Thermal cameras and applications: a survey", *Machine vision and applications*, vol. 25, no. 1, pp. 245–262, 2014. 26

[144] Junichi Tanaka, Hiroshi Imamoto, Tomonori Seki, and Masatoshi Oba, "Low power wireless human detector utilizing thermopile infrared array sensor", in *SENSORS, 2014 IEEE*. IEEE, 2014, pp. 462–465. 26

[145] Suk Jin Lee, Gaurav Shah, Arka Aloke Bhattacharya, and Yuichi Motai, "Human tracking with an infrared camera using a curve matching framework", *EURASIP Journal on Advances in Signal Processing*, vol. 2012, no. 1, pp. 99, 2012. 26

[146] Piero Zappi, Elisabetta Farella, and Luca Benini, "Tracking motion direction and distance with pyroelectric ir sensors", *IEEE Sensors Journal*, vol. 10, no. 9, pp. 1486–1494, 2010. 26

[147] Ren C Luo, Chih-Chen Yih, and Kuo Lan Su, "Multisensor fusion and integration: approaches, applications, and future research directions", *IEEE Sensors journal*, vol. 2, no. 2, pp. 107–119, 2002. 27

[148] Sanaul Hoque, Konstantinos Sirlantzis, and Michael C Fairhurst, "A new chain-code quantization approach enabling high performance handwriting recognition based on multi-classifier schemes.", in *ICDAR*, 2003, pp. 834–838. 27

[149] Konstantinos Sirlantzis, Michael C Fairhurst, and MS Hoque, "Genetic algorithms for multi-classifier system configuration: a case study in character recognition", in *International Workshop on Multiple Classifier Systems*. Springer, 2001, pp. 99–108. 27

[150] Konstantinos Sirlantzis, Sanaul Hoque, and Michael C Fairhurst, "Diversity in multiple classifier ensembles based on binary feature quantisation with application to face recognition", *Applied Soft Computing*, vol. 8, no. 1, pp. 437–445, 2008. 27

[151] Pau Ferrer-Cid, Jose M Barcelo-Ordinas, Jorge Garcia-Vidal, Anna Ripoll, and Mar Viana, "Multisensor data fusion calibration in iot air pollution platforms", *IEEE Internet of Things Journal*, vol. 7, no. 4, pp. 3124–3132, 2020. 27

[152] Hui Lin, Sahil Garg, Jia Hu, Xiaoding Wang, Md Jalil Piran, and M Shamim Hossain, "Privacy-enhanced data fusion for covid-19 applications in intelligent internet of medical things", *IEEE Internet of Things Journal*, 2020. 27

[153] Khan Muhammad, Tanveer Hussain, Muhammad Tanveer, Giovanna Sannino, and Victor Hugo C de Albuquerque, "Cost-effective video summarization using deep cnn with hierarchical weighted fusion for iot surveillance networks", *IEEE Internet of Things Journal*, vol. 7, no. 5, pp. 4455–4463, 2019. 27

[154] Nikhil Kumar, Debopam Acharya, and Divya Lohani, "An iot-based vehicle accident detection and classification system using sensor fusion", *IEEE Internet of Things Journal*, vol. 8, no. 2, pp. 869–880, 2020. 27

[155] Eiman Kanjo, Eman MG Younis, and Chee Siang Ang, "Deep learning analysis of mobile physiological, environmental and location sensor data for emotion detection", *Information Fusion*, vol. 49, pp. 46–56, 2019. 27

[156] Asma Benmansour, Abdelhamid Bouchachia, and Mohammed Feham, "Multioccupant activity recognition in pervasive smart home environments", *ACM Computing Surveys (CSUR)*, vol. 48, no. 3, pp. 34, 2016. 28, 51

[157] Angeliki Kylili, Paris A Fokaides, Petros Christou, and Soteris A Kalogirou, "Infrared thermography (irt) applications for building diagnostics: A review", *Applied Energy*, vol. 134, pp. 531–549, 2014. 40

[158] Daniel Mota-Rojas, Dehua Wang, Cristiane Gonçalves Titto, Jocelyn Gómez-Prado, Verónica Carvajal-de la Fuente, Marcelo Ghezzi, Luciano Boscato-Funes, Hugo Barrios-García, Fabiola Torres-Bernal, Alejandro Casas-Alvarado, et al., "Pathophysiology of fever and application of infrared thermography (irt) in the detection of sick domestic animals: Recent advances", *Animals*, vol. 11, no. 8, pp. 2316, 2021. 40

[159] Abdul Majeed and Seong Oun Hwang, "A comprehensive analysis of privacy protection techniques developed for covid-19 pandemic", *IEEE Access*, 2021. 40

[160] Stefano Savazzi, Vittorio Rampa, Leonardo Costa, Sanaz Kianoush, and Denis Tolochenko, "Processing of body-induced thermal signatures for physical distancing and temperature screening", *IEEE Sensors Journal*, 2020. 41

[161] Munkhjargal Gochoo, Tan-Hsu Tan, Shih-Chia Huang, Tsedevdorj Batjargal, Jun-Wei Hsieh, Fady S Alnajjar, and Yung-Fu Chen, "Novel iot-based privacy-preserving yoga posture recognition system using low-resolution infrared sensors and deep learning", *IEEE Internet of Things Journal*, vol. 6, no. 4, pp. 7192–7200, 2019. 41

[162] Dongning Qu, Bo Yang, and Nanhao Gu, "Indoor multiple human targets localization and tracking using thermopile sensor", *Infrared Physics & Technology*, vol. 97, pp. 349–359, 2019. 41

[163] Zhangjie Chen and Ya Wang, "Remote recognition of in-bed postures using a thermopile array sensor with machine learning", *IEEE Sensors Journal*, vol. 21, no. 9, pp. 10428–10436, 2021. 46

[164] Martin T Hagan and Mohammad B Menhaj, "Training feedforward networks with the marquardt algorithm", *IEEE transactions on Neural Networks*, vol. 5, no. 6, pp. 989–993, 1994. 47

[165] Olaf Ronneberger, Philipp Fischer, and Thomas Brox, "U-net: Convolutional networks for biomedical image segmentation", in *International Conference on Medical image computing and computer-assisted intervention.* Springer, 2015, pp. 234–241. 55

[166] Diederik P Kingma and Jimmy Ba, "Adam: A method for stochastic optimization", *arXiv preprint arXiv:1412.6980*, 2014. 55, 108

[167] HJAM Heijmans, "Introduction to connected operators", *Nonlinear filters for image processing*, pp. 207–235, 1999. 56

[168] Youjun Lu, Jikai Huang, and Pengfei Zheng, "A cfd–dem study of bubble dynamics in fluidized bed using flood fill method", *Chemical Engineering Journal*, vol. 274, pp. 123–131, 2015. 56

[169] Monica Bianchini and Franco Scarselli, "On the complexity of neural network classifiers: A comparison between shallow and deep architectures", *IEEE transactions on neural networks and learning systems*, vol. 25, no. 8, pp. 1553–1565, 2014. 58

[170] S Sapna, A Tamilarasi, M Pravin Kumar, et al., "Backpropagation learning algorithm based on levenberg marquardt algorithm", *Comp Sci Inform Technol (CS and IT)*, vol. 2, pp. 393–398, 2012. 59, 74

[171] Nobuyuki Otsu, "A threshold selection method from gray-level histograms", *IEEE transactions on systems, man, and cybernetics*, vol. 9, no. 1, pp. 62–66, 1979. 86

[172] Berthold KP Horn and Brian G Schunck, "Determining optical flow", *Artificial intelligence*, vol. 17, no. 1-3, pp. 185–203, 1981. 89, 105, 106

[173] Gunnar Farnebäck, "Two-frame motion estimation based on polynomial expansion", in *Scandinavian conference on Image analysis.* Springer, 2003, pp. 363–370. 105, 107

[174] Sepp Hochreiter and Jürgen Schmidhuber, "Long short-term memory", *Neural computation*, vol. 9, no. 8, pp. 1735–1780, 1997. 108

[175] Mike Schuster and Kuldip K Paliwal, "Bidirectional recurrent neural networks", *IEEE transactions on Signal Processing*, vol. 45, no. 11, pp. 2673–2681, 1997. 108

[176] Ning Qian, "On the momentum term in gradient descent learning algorithms", *Neural networks*, vol. 12, no. 1, pp. 145–151, 1999. 108